Simo Puntanen ja Kimmo Vehkalahti

Matriiseja tilastotieteilijälle

TAMPEREEN YLIOPISTO

INFORMAATIOTIETEIDEN YKSIKÖN RAPORTTEJA 56/2017

TAMPERE 2017

TAMPEREEN YLIOPISTO
INFORMAATIOTIETEIDEN YKSIKÖN RAPORTTEJA 56/2017
ELOKUU 2017

Simo Puntanen ja Kimmo Vehkalahti

Matriiseja tilastotieteilijälle

ISBN 978-952-03-0535-2 (pdf)

ISSN-L 1799-8158
ISSN 1799-8158

Matriiseja tilastotieteilijälle

Simo Puntanen & Kimmo Vehkalahti

Alkusanat 1. painokseen

Tämä kirja on tarkoitettu nimensä mukaiseen toimintaan, eikä siis siinä suhteessa eroa muista kirjoista suuremmin. Toisin sanoen: tarkoituksena on esitellä lukijalle matriisilaskennasta ja lineaarialgebrasta sellaisia piirteitä, joita erityisesti tilastotieteessä tarvitaan. Lukijalta edellytetään varsin vähän etukäteistietoja lineaarialgebrasta; pyrimme lähtemään liikkeelle alkeiskäsitteistä. Tilastotieteen ensimmäisen vuoden yliopisto-opinnot on hyvä olla takana (menestyksellisesti).

Esitystavassa pyritään korostamaan geometrista ajattelutapaa aina kun se on mahdollista. Erityisesti projektio-käsite on aivan avainasemassa. Kirjassa on tehty silloin tällöin epäortodoksisesti niin, että sama käsite on määritelty uudelleen toistuvasti (samalla tavalla, luojan kiitos!). Kertaus olkoon oppimisen isä. Kuten näille aloille suuntautunut jo opiskelujensa tässä vaiheessa tietää, ei näitä asioita varsinaisesti lukemalla opi – harjoitustehtävien aikaa vievä pohdiskelu on oleellisen oppimisen aa ja oo.

Numeeriset esimerkit on tehty Survon avulla. On monesti hyvin opettavaista ja konkretisoivaa käydä tietokoneella läpi joitakin tarkasteluja. Survo sopii erittäin hyvin siihen lähestymistapaan, jota kirjassa on käytetty.

Tämä versio on alustava testiversio, josta puuttuvat mm. harjoitustehtävät ja hakemisto. Myös loppuluvut 9–11 ovat keskeneräisiä. Lopullinen versio on tarkoitus tehdä tämän vuoden loppuun mennessä.

Simo Puntanen
18. helmikuuta 1998

Alkusanat 2. painokseen

Ei valmistunut ihan kuin edellä luvattiin, ei valmistunut.

Tässä tämä nyt kuitenkin on, sen pidemmittä puheitta. Suurkiitokset Jarmo Niemelälle erinomaisesta LaTeX-avusta. Kiitokset menevät myös Tampereen yliopistolle kirjan viimeistelyn tukemisessa.

Lukijaa kehotetaan vilkaisemaan seuraavia kirjoja: Puntanen, Styan & Isotalo (2011, 2013). *Matrix Tricks* -kirjassa on liuta harjoitustehtäviä, joista osaan on ratkaisut sivulla http://www.sis.uta.fi/tilasto/matrixtricks/.

Simo Puntanen
Luonnontieteiden tiedekunta
33014 Tampereen yliopisto
simo.puntanen@uta.fi

Kimmo Vehkalahti
Valtiotieteellinen tiedekunta
Menetelmäkeskus
00014 Helsingin yliopisto
kimmo.vehkalahti@helsinki.fi

Sisältö

Luku 1

Johdanto

Jos lukija kuuluu niihin virheettömiin ja synnittömiin ihmisiin, jotka eivät lähimmäisissäänkään siedä omahyväisyyttä eivätkä itsekehua missään muodossa, lukija voi jättää tämän johdantoluvun lukematta, niin jännittävä kuin se tekijäin mielestä onkin.

Tauno Nurmela (1975): *Vanha varis eli tuntematon Boccaccio.*

1.1 Matriisin määrittely

Sana *matriisi* on lukijalle tuttu jo tilastotieteen peruskurssilta, jossa käsiteltiin tilastotieteen kannalta mm. kolmea hyvin keskeistä matriisia: havaintomatriisia, kovarianssimatriisia ja korrelaatiomatriisia. Havaintomatriisi muodostaa lähtokohdan monille tilastotieteellisille tarkasteluille ja aloitammekin tämän ensimmäisen luvun – lämmittelyluvun – palauttamalla mieleen mitä havaintomatriisilla tarkoitetaan.

1.1.1 Havaintomatriisi

Havaintomatriisi on tietyn sopimuksen mukainen havaintoaineiston esitys. Jos esimerkiksi neljältä henkilöltä on mitattu kolme ominaisuutta, on aineiston havaintomatriisiesitys seuraava:

$$
\begin{array}{c c}
 & \begin{array}{ccc} \text{mja 1} & \text{mja 2} & \text{mja 3} \end{array} \\
\mathbf{U} = & \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \end{pmatrix} \begin{array}{l} \text{henkilö 1} \\ \text{henkilö 2} \\ \text{henkilö 3} \\ \text{henkilö 4} \end{array}
\end{array}
\tag{1.1}
$$

Luku u_{ij} kuvaa i. henkilön saamaa arvoa j. muuttujalla; ensimmäinen alaindeksi viittaa siis havaintomatriisin vaakariviin ja toinen indeksi kertoo pystyrivin. Havaintomatriisissa on siten n (4) vaakariviä ja p (3) pystyriviä, jos aineistossa on n havaintoyksikköä ja p muuttujaa. Tässä tilanteessa voimme ajatella, että tarkasteltavat muuttujat ovat u_1, u_2 ja u_3, mutta jos muuttujat olisivatkin x, y ja z, niin voisimme merkitä

$$\mathbf{U} = \begin{array}{c} \begin{array}{ccc} \text{mja } x & \text{mja } y & \text{mja } z \end{array} \\ \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{pmatrix} \end{array} . \tag{1.2}$$

Jos havaintomatriisin $\mathbf{U}$ kukin pystyrivi korvataan vastaavalla vaakarivillä saadaan sama informaatio esitetyksi muodossa

$$\mathbf{U}' = \begin{array}{cc} \begin{array}{cccc} \text{hlö 1} & \dots & & \text{hlö 4} \end{array} & \\ \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{pmatrix} & \begin{array}{l} \text{muuttuja } x \\ \text{muuttuja } y \\ \text{muuttuja } z \end{array} \end{array} \tag{1.3}$$

Matriisi $\mathbf{U}'$ on $\mathbf{U}$:n *transpoosi*. Matriisin $\mathbf{A}$ $(n \times m)$ transpoosilla tarkoitetaan matriisia $\mathbf{A}'$ $(m \times n)$, joka saadaan vaihtamalla $\mathbf{A}$:n kukin pystyrivi vastaavaksi vaakariviksi. Joissakin tilastotieteen oppikirjoissa käytetään havaintomatriisi-nimitystä esityksestä (1.3), mutta tässä esityksessä tarkoitamme havaintomatriisilla nimenomaan (1.1):n mukaista aineiston esitystapaa:

- pystyrivit vastaavat muuttujia,
- vaakarivit vastaavat havaintoja.

Esimerkiksi Seber (1984) käyttää (1.1):n mukaista merkintätapaa, kun taas mm. Anderson (2003) ja Mustonen (1995) käyttävät havaintomatriisi-nimitystä (1.3):sta.

1.1.2 Korrelaatiomatriisi

Kolmen muuttujan tapauksessa voimme merkitä havaintomatriisista $\mathbf{U}$ $(n \times 3)$ laskettua korrelaatiomatriisia seuraavasti:

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix} = \operatorname{cor_d}(\mathbf{U}) \,, \tag{1.4}$$

missä r_{ij} tarkoittaa i. ja j. muuttujan välistä korrelaatiokerrointa. Tietenkin kaikki r_{ii}:t (lävistäjäelementit) ovat ykkösiä ja r_{ij}:t ovat itseisarvoltaan enintään ykkösiä. On kuitenkin huomattava, että r_{ij}:t eivät voi olla mitä tahansa itseisarvoltaan ykköstä pienempiä tai yhtäsuuria lukuja, jos $\mathbf{R}$ on korrelaatiomatriisi; tähän palataan myöhemmin ei-negatiivisesti definiittien matriisien yhteydessä – $\mathbf{R}$ nimittäin on ei-negatiivisesti definiitti matriisi. (Ks. myös HT 1.2, s. 62.) Matriisi $\mathbf{R}$ on *neliömatriisi*, koska siinä on yhtä monta vaakariviä kuin saraketta. Lisäksi

$$\mathbf{R} \text{ on } \textit{symmetrinen} \text{ eli } r_{ij} = r_{ji}, \; i, j = 1, 2, 3 .$$

Muita keskeisiä matriiseja tilastotieteessä ovat mm. muuttujien *tulosummamatriisi* ja *kovarianssimatriisi*:

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & r_{32} & t_{33} \end{pmatrix} = \operatorname{ssp}(\mathbf{U}), \text{ tulosummamatriisi,} \tag{1.5a}$$

$$\mathbf{S} = \frac{1}{n-1}\mathbf{T} = \begin{pmatrix} s_1^2 & s_{12} & s_{13} \\ s_{21} & s_2^2 & s_{23} \\ s_{31} & s_{32} & s_3^2 \end{pmatrix} = \operatorname{cov_d}(\mathbf{U}), \text{ kovarianssimatriisi,} \tag{1.5b}$$

tai toisin indeksoiden (kun muuttujat ovat x, y ja z)

$$\mathbf{T} = \begin{pmatrix} t_{xx} & t_{xy} & t_{xz} \\ t_{yx} & t_{yy} & t_{yz} \\ t_{zx} & r_{zy} & t_{zz} \end{pmatrix}, \quad \mathbf{S} = \frac{1}{n-1}\mathbf{T} = \begin{pmatrix} s_x^2 & s_{xy} & s_{xz} \\ s_{yx} & s_y^2 & s_{yz} \\ s_{zx} & s_{zy} & s_z^2 \end{pmatrix}, \tag{1.5c}$$

missä

$$t_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = \mathrm{SS}_y , \tag{1.6a}$$

$$t_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = \mathrm{SP}_{xy} , \tag{1.6b}$$

$$s_y^2 = \tfrac{1}{n-1} t_{yy} = \operatorname{var_s}(y) = \operatorname{var_d}(\mathbf{y}), \; y\text{:n (otos)varianssi,} \tag{1.6c}$$

$$s_{xy} = \tfrac{1}{n-1} t_{xy} = r_{xy} s_x s_y = \operatorname{cov_s}(x, y) = \operatorname{cov_d}(\mathbf{x}, \mathbf{y}), \; \text{(otos)kovarianssi,} \tag{1.6d}$$

$$r_{xy} = \frac{\mathrm{SP}_{xy}}{\sqrt{\mathrm{SS}_x \, \mathrm{SS}_y}} = \frac{s_{xy}}{s_x s_y} = \frac{t_{xy}}{\sqrt{t_{xx} t_{yy}}} \tag{1.6e}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \operatorname{cor_s}(x, y) = \operatorname{cor_d}(\mathbf{x}, \mathbf{y})$$

$$= \text{(otos)korrelaatiokerroin.} \tag{1.6f}$$

Merkintä $\mathrm{var_s}(y)$ viittaa otosvarianssiin, kun argumenttina on muuttuja y ja $\mathrm{var_d}(\mathbf{y})$ tarkoittaa otosvarianssia, kun argumenttina on vektori $\mathbf{y} \in \mathbb{R}^n$, joka sisältää y:n havaitut arvot. Lauseke $t_{yy} = \mathrm{SS}_y$ tarkoittaa y:n neliösummaa (sum of squares) tai itse asiassa y:n keskiarvosta laskettujen erotusten neliöiden summaa eli ns. ”korjattua” neliösummaa ja $t_{xy} = \mathrm{SP}_{xy}$ on x:n ja y:n (korjattu) tulosumma (sum of products of deviations). ”Korjaus” tarkoittaa sitä, että termit, joiden neliöitä ja tuloja tarkastellaan, ovat alkuperäisten havaintoarvojen erotuksia keskiarvosta.

Jos $s_x = 0$ tai $s_y = 0$, niin välttämättä myös $s_{xy} = 0$, jolloin $r_{xy} = 0/0$ eli korrelaatiokerroin on määritelty vain jos muuttujien varianssit eivät ole nollia.

Lisäksi lukija muistaa regressiokertoimien (estimaattien) lausekkeet

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}\,, \tag{1.7a}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\mathrm{SP}_{xy}}{\mathrm{SS}_x} = \frac{s_{xy}}{s_x^2} = \frac{s_{xy}}{s_x s_y}\frac{s_y}{s_x} = r_{xy}\frac{s_y}{s_x}. \tag{1.7b}$$

Regressiosuoran yhtälö on

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = (\bar{y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 x = \bar{y} + \hat{\beta}_1 (x - \bar{x})\,, \tag{1.7c}$$

eli regressiosuoran kulmakerroin on $\frac{s_{xy}}{s_x^2}$ ja se kulkee keskiarvopisteen $(\bar{x}, \bar{y})$ kautta. Kertoimet $\hat{\beta}_0$ ja $\hat{\beta}_1$ ovat todellisten regressiokertoimien β_0 ja β_1 pienimmän neliösumman estimaatteja (Ordinary Least Squares Estimates, OLSE), kun tarkasteltava lineaarinen tilastollinen malli on

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,, \qquad i = 1, \dots, n\,. \tag{1.7d}$$

Satunnaismuuttujat ε_i ovat mallin *virhetermejä*. Datasta laskettuja mallin antamia arvoja $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ sanotaan *sovitteiksi* (fitted values) ja erotuksia $y_i - \hat{y}_i$ *residuaaleiksi* tai jäännöksiksi. Tavanomainen oletus on, että ε_i:t ovat riippumattomia ja kukin on normaalisti jakautunut odotusarvona 0 ja varianssina σ^2 eli $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$. Regressiokertoimien estimaatit voidaan esittää pystyvektorina

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}. \tag{1.7e}$$

Regressioanalyysissä merkitään ns. selitettävää muuttujaa y:llä ja sitä sanotaan *vasteeksi*. Selittäjiä, joista jokaisesta on n havaintoa, on tapana merkitä $x_1, \dots, x_k$ sekä lisäksi mukana on ns. ”vakio”: vakio on muuttuja, jonka arvot ovat kaikki samoja, yleensä vakion arvot ovat ykkösiä. Kaavat (1.7) sopivat tilanteeseen, jossa selittäjinä ovat vain vakio ja x. Yksi regressioanalyysin johtavista ideoista on löytää selitys (parasta kirjoittaa ”selitys”) sille, miksi vasteen arvot vaihtelevat ts. purkaa kokonaisvaihtelua kuvaava neliösumma

$$\sum_{i=1}^n (y_i - \bar{y})^2 = t_{yy} = \mathrm{SST} = \text{ kokonaisneliösumma} \tag{1.8}$$

sopivasti osiin. Toiveena on selittää y:n vaihtelun johtuvan joistakin muista tekijöistä, ns. selittäjistä. Tämä on vähän löysä kuvaus regressioanalyysistä, mutta on tuossa jotakin perää.

Termillä *regressio* on poikkeuksellisen mielenkiintoinen historia: ks. lukua entitled *Regression towards Mean* Stieglerin kirjassa 1999, missä (s. 177) hän sanoo, että tarina Francis Galtonin (1822–1911) regression löytämisestä on "an exciting one, involving science, experiment, mathematics, simulation, and one of the great thought experiments of all time". Galton tutki isien ja poikien pituuksien välistä suhdetta. Hän havaitsi pitkien isien saavan yleensä pitkiä poikia mutta erityisen pitkät isät saivatkin poikia, joiden pituudella oli taipumus olla lähempänä keskiarvoa (keskihajonnan avulla skaalattuna). Mielenkiintoisia kommentteja regressioanalyysistä on esittänyt myös mm. Speed (2012a,b, 2013).

Havaintomatriisi $\mathbf{U}$ on (1.1):n (s. 8) tapauksessa 4×3-matriisi ja (1.4):n (s. 9) mukainen $\mathbf{R}$ on puolestaan 3×3-matriisi. Matriiseja merkitsemme isoilla lihavoitetuilla kirjaimilla $\mathbf{A}$, $\mathbf{B}$, ... Pystyvektoreita, jotka siis ovat $n \times 1$-matriiseja, merkitsemme pienillä lihavilla kirjaimilla $\mathbf{a}$, $\mathbf{b}$, ... Vaakavektoreista käytetään merkintöjä $\mathbf{a}'$, $\mathbf{b}'$, ..., missä symboli $'$ tarkoittaa transponointia:

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \quad \mathbf{a}' = (a_1, a_2, \dots, a_n)\,. \tag{1.9}$$

Huomattakoon, että kovarianssimatriisin kaavassa (1.5b) käytämme sääntöä, jonka mukaan matriisin $\mathbf{A}$ kertominen reaaliluvulla λ tarkoittaa $\mathbf{A}$:n jokaisen elementin kertomista λ:lla: $\lambda\mathbf{A} = \{\lambda a_{ij}\}$.

Merkinnällä $\mathbb{R}^{n\times m}$ tarkoitetaan kaikkien reaalielementtisten (jollaisten tarkasteluun vain rajoitutaan) $n \times m$-matriisien joukkoa eli

$$\mathbb{R}^{n\times m} = \{\, \mathbf{A} : \mathbf{A} \text{ on reaalielementtinen } n \times m\text{-matriisi} \,\}. \tag{1.10}$$

Myös merkintää $\mathbf{A}_{n\times m}$ käytetään haluttaessa ilmaista matriisin rivi- ja sarakelukumäärä eli matriisin dimensio. Erityisesti merkitään $\mathbb{R}^{n\times 1} = \mathbb{R}^n$, ja $\mathbb{R}^{1\times 1} = \mathbb{R}$. Merkinnällä $\mathbb{R}^n$ lukija on tottunut tarkoittamaan n-ulotteista euklidista reaaliavaruutta; nyt itse asiassa sovitaan vain, että $\mathbb{R}^n$:n alkiot esitetään pystyvektoreina. Joskus käytämme myös merkintää

$$\mathrm{Sym}(n) = \{\, \mathbf{A} : \mathbf{A} \text{ on symmetrinen reaalielementtinen } n \times n\text{-matriisi} \,\}.$$

1.1.3 Muuttujavektori ja havaintovektori

Jos merkitsemme havaintomatriisin $\mathbf{U}_{n\times 2}$ sarakkeita symboleilla $\mathbf{x}$ ja $\mathbf{y}$, voimme esittää $\mathbf{U}$:n ositetussa muodossa seuraavasti:

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix}. \tag{1.11}$$

Jos ositamme $\mathbf{U}$:n vaakariveittäin siten että vaakarivit ovat $\mathbf{u}'_{(1)}$, $\mathbf{u}'_{(2)}$, $\dots$, $\mathbf{u}'_{(n)}$, missä

$$\mathbf{u}_{(i)} = \begin{pmatrix} x_i \\ y_i \end{pmatrix}, \quad \text{ts. } \mathbf{u}'_{(i)} = (x_i, y_i)\,, \tag{1.12}$$

saamme havaintomatriisille $\mathbf{U}$ esityksen

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix}. \tag{1.13}$$

Havaintomatriisin sarakkeita $\mathbf{x}$ ja $\mathbf{y}$ sanomme *muuttujavektoreiksi* ja vaakarivien transpooseja $\mathbf{u}_{(1)}$, $\mathbf{u}_{(2)}$, $\dots$, $\mathbf{u}_{(n)}$, *havaintovektoreiksi* tai yksinkertaisesti vain *havainnoiksi*. On erityisesti korostettava, että tässä tilanteessa

- muuttujavektorit $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,
- havaintovektorit $\mathbf{u}_{(1)}, \dots, \mathbf{u}_{(n)} \in \mathbb{R}^2$.

Pari kommenttia merkinnöistä (1.11)–(1.13):ssä on paikallaan. Ensinnäkin selitys sille, miksi käytimme suluissa olevaa indeksointimerkintää $\mathbf{u}_{(i)}$ on se, että pyrimme varaamaan suluttoman merkinnän $\mathbf{u}_i$ matriisin $\mathbf{U}$ pystyriveille:

$$\mathbf{U} = (\mathbf{u}_1 : \mathbf{u}_2) = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix}. \tag{1.14}$$

Aina emme kuitenkaan käytä (1.14):n mukaista merkintää; näinhän on esimerkiksi (1.11):ssa, jossa $\mathbf{U}$:n sarakkeita on $\mathbf{u}_1$:n ja $\mathbf{u}_2$:n sijasta merkitty $\mathbf{x}$:llä ja $\mathbf{y}$:llä. On selvää, että jos haluamme merkitä $\mathbf{U}$:n sarakkeita ja vaakarivejä samalla kirjaimella (esim. $\mathbf{u}$:lla), niin indeksointi sarakkeille ja vaakariveille on tehtävä eri tavalla. Joissakin tarkasteluissa (erityisesti monimuuttujamenetelmissä) nimenomaan havaintovektoreilla (= vaakarivien transpooseilla) on aivan erityisen keskeinen rooli, jolloin on perusteltua käyttää ”kevyempää” sulutonta indeksointia havaintovektoreille.

Mainittakoon vielä, että useimmiten käytämme symbolia n kuvaamaan havaintojen lukumäärää ja muuttujien lukumäärän merkintänä on usein p. Seber (1984, s. 3) käyttää d:tä kuvaamaan muuttujien lukumäärää ja tähän on luonteva syy se, että d viittaa moniulotteisen muuttujan $d$imensioon. Lukija varmaan huomaa myös, että tässä kirjassa matriiseilla (ei ainoastaan havaintomatriiseilla) on usein vaakarivien lukumäärä nimenomaan n.

1.1.4 Osittaiskorrelaatiomatriisi

Muuttujien x ja y *osittaiskorrelaatiokerroin*, kun muuttujan z vaikutus on eliminoitu, määritellään lausekkeena

$$\mathrm{pcor_d}(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = r_{xy\cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \,. \tag{1.15}$$

Tällöin vastaava osittaiskorrelaatiomatriisi on

$$\mathrm{pcor_d}(\mathbf{x} : \mathbf{y} \mid \mathbf{z}) = \begin{pmatrix} 1 & r_{xy\cdot z} \\ r_{yx\cdot z} & 1 \end{pmatrix}. \tag{1.16}$$

On huomattava, että merkintä $\mathrm{cor_d}(\mathbf{x}, \mathbf{y})$ tarkoittaa korrelaatiokerrointa, mutta merkintä $\mathrm{cor_d}(\mathbf{x} : \mathbf{y})$ tarkoittaakin korrelaatiomatriisia:

- $\mathrm{cor_d}(\mathbf{x}, \mathbf{y}) = \text{korrelaatiokerroin} = r_{xy} \in \mathbb{R}$,
- $\mathrm{cor_d}(\mathbf{x} : \mathbf{y}) = \text{korrelaatiomatriisi} \in \mathbb{R}^{2\times 2}$.

Vastaava tärkeä ero on merkinnöillä $\mathrm{pcor_d}(\mathbf{x}, \mathbf{y} \mid \mathbf{z})$ ja $\mathrm{pcor_d}(\mathbf{x} : \mathbf{y} \mid \mathbf{z})$. Muuttujien x_1, x_2 ja x_3 osittaiskorrelaatiomatriisi kun x_4 on eliminoitu on tietenkin

$$\mathrm{pcor_d}(\mathbf{x}_1 : \mathbf{x}_2 : \mathbf{x}_3 \mid \mathbf{x}_4) = \begin{pmatrix} 1 & r_{12\cdot 4} & r_{13\cdot 4} \\ r_{21\cdot 4} & 1 & r_{23\cdot 4} \\ r_{31\cdot 4} & r_{32\cdot 4} & 1 \end{pmatrix}. \tag{1.17}$$

Kommentti 1.1. Osittaiskorrelaation tulkinta saattaa olla mutkikasta. Jos muuttujien x, y, z yhteisjakauma (populaation jakauma) on multinormaalijakauma, niin $r_{xy\cdot z}$ on muuttujien x ja y ehdollinen korrelaatiokerroin (itse asiassa sen estimaatti), mutta mikäli kyseessä ei ole multinormaalijakauma, näin ei välttämättä ole. Usein käytetään sanontaa: y:stä ja x:stä eliminoidaan z:n vaikutus ts. pidetään z vakiona ja lasketaan tässä tilanteessa r_{xy}, tulos on $r_{xy\cdot z}$. Tälläinen tulkinta ei aina ole tietenkään oikeutettua. □

1.1.5 Siirtymätodennäköisyysmatriisi

Tarkastellaan sitten yhtä, edellisistä luonteeltaan hiukan poikkeavaa matriisia, ns. *siirtymätodennäköisyysmatriisia*, josta yksinkertaisena esimerkkinä olkoon

$$\mathbf{P} = \begin{pmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{pmatrix}. \tag{1.18}$$

Olkoon $\mathbf{P}$:n elementeillä sellainen tulkinta, että

$$p_{ij} = \text{todennäköisyys että puolueen } i \text{ tämänhetkinen kannattaja äänestää seuraavissa vaaleissa puoluetta } j.$$

Täten todennäköisyys, että puolueen 1 kannattaja äänestää puoluetta 1 ensi vaaleissa on 0.3. Matriisin $\mathbf{P}$ luvut ovat siis tulkittavissa siirtymätodennäköisyyksinä tilasta E_i tilaan E_j: puolue-esimerkissä $E_i = \{\text{henkilö äänestää puoluetta } i\}$. Tällöin täytyy tietenkin kunkin vaakarivin lukujen summan olla 1. Jos tiloja on n kappaletta, on siirtymätodennäköisyysmatriisi seuraava:

$$\begin{array}{cc} & \begin{array}{cccc} \text{uusi tila } E_1 & E_2 & \ldots & E_n \end{array} \\ \mathbf{P} = & \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \begin{array}{c} E_1 \\ E_2 \\ \vdots \\ E_n \end{array} \text{ vanha tila} \end{array} \tag{1.19}$$

Neliömatriisia $\mathbf{P}_{n\times n} = \{p_{ij}\}$, jonka jokaisen vaakarivin elementtien summa on 1 (ja jokainen elementti $p_{ij} \geq 0$),

$$p_{i1} + p_{i2} + \cdots + p_{in} = 1\,, \quad i = 1, 2, \ldots, n\,, \tag{1.20}$$

sanotaan *stokastiseksi* matriisiksi. Matriisien $\mathbf{A}_{n\times m}$ ja $\mathbf{B}_{n\times m}$ *summa* määritellään siten, että se on matriisi, joka saadaan laskemalla $\mathbf{A}$:n ja $\mathbf{B}$:n "päällekkäiset" elementit yhteen: $\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\}$. Täten matriisin

$$\mathbf{P} = (\mathbf{p}_1 : \mathbf{p}_2 : \ldots : \mathbf{p}_n) = \begin{pmatrix} \mathbf{p}'_{(1)} \\ \mathbf{p}'_{(2)} \\ \vdots \\ \mathbf{p}'_{(n)} \end{pmatrix} \tag{1.21}$$

sarakkeiden summa on

$$\mathbf{p}_1 + \mathbf{p}_2 + \cdots + \mathbf{p}_n = \begin{pmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{n1} \end{pmatrix} + \begin{pmatrix} p_{12} \\ p_{22} \\ \vdots \\ p_{n2} \end{pmatrix} + \cdots + \begin{pmatrix} p_{1n} \\ p_{2n} \\ \vdots \\ p_{nn} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} := \mathbf{1}_n\,. \tag{1.22}$$

Jos myös vaakariveillä on sama ominaisuus eli

$$\mathbf{p}_{(1)} + \mathbf{p}_{(2)} + \cdots + \mathbf{p}_{(n)} = \mathbf{1}_n \,, \tag{1.23}$$

sanotaan matriisia *kaksois-stokastiseksi*; ks. myös taikaneliöt ja taikamatriisit, s. 71.

Esimerkki 1.1. Tutkittaessa kolmen valmennusmenetelmän V_1, V_2 ja V_3 vaikutusta urheilusuorituksiin saatiin seuraava aineisto, kun menetelmiä sovelsi 7 urheilijaa:

$$V_1 : 1, 3, \quad V_2 : 4, 5, 6, \quad V_3 : 7, 9. \tag{1.24}$$

Tarkastellaaan kahta tilannetta:

(a) Aineistoon tehdään yksisuuntainen varianssianalyysi, ts. aineisto luokitellaan kolmeen ryhmään ja vertailun kohteena ovat ryhmäkeskiarvot.

(b) Oletetaan, että V_i tarkoittaa, että henkilö juoksee i tuntia joka päivä; aineistoon tehdään regressioanalyysi selittäjän ollessa juoksumäärä.

Ongelma: Muodosta havaintomatriisit kummastakin tilanteesta.

Ratkaisu: (a) Merkitään y:llä selitettävää muuttujaa, y = urheilutulos. Havaintojen luokittelun voimme ilmaista esimerkiksi seuraavasti määriteltyjen muuttujien t_1, t_2 ja t_3 (joiden arvot ovat muuttujavektoreissa $\mathbf{t}_1$, $\mathbf{t}_2$ ja $\mathbf{t}_3$) avulla:

$$\mathbf{T} = (\mathbf{t}_1 : \mathbf{t}_2 : \mathbf{t}_3) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \{t_{ij}\} \,. \tag{1.25}$$

Havaintojen luokittelu käy siis ilmi t_i-muuttujista siten että

$$t_{ij} = \begin{cases} 1, & \text{jos } i\text{. henkilö saa ”käsittelyn” } j \ (V_j), \\ 0, & \text{muuten.} \end{cases}$$

Koko aineisto (luokittelu & valmennustulokset) on tällöin esitettävissä matriisina

$$\mathbf{U} = (\mathbf{T} : \mathbf{y}) = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 4 \\ 0 & 1 & 0 & 5 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 1 & 9 \end{pmatrix} . \tag{1.26}$$

Ratkaisu: (b) Olkoon x = juoksumäärä. Tällöin:

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 2 & 4 \\ 2 & 5 \\ 2 & 6 \\ 3 & 7 \\ 3 & 9 \end{pmatrix}. \tag{1.27}$$

Tilanteessa (b) meillä on siis 2 muuttujavektoria ja 7 havaintovektoria. On huomattava, että (b)-tilanteessa havaintomatriisi on yksikäsitteinen, mutta (a)-tilanteessa on monta tapaa muodostaa $\mathbf{U}$.

Jatketaan vielä (a)-tilanteen tarkastelua olettamalla, että kolme ensimmäistä henkilöä (vastaten kolmea ensimmäistä vaakariviä matriisissa $\mathbf{T}$) ovat naisia ja loput miehiä; sukupuolen perusteella muodostetaan kaksi *lohkoa* (blocks). Tämä tilanne voidaan kuvata seuraavan matriisin avulla:

$$\mathbf{Z} = (\mathbf{t}_1 : \mathbf{t}_2 : \mathbf{t}_3 : \mathbf{b}_1 : \mathbf{b}_2) = (\mathbf{T} : \mathbf{B}) = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}. \tag{1.28}$$

Matriisista $\mathbf{T}$ voidaan käyttää nimitystä ”design matrix for treatment effects” ja $\mathbf{B}$:stä ”design matrix for block effects”. Nyt siis käsittelyjen lukumäärä on 3 ja lohkojen lukumäärä on 2. □

1.2 Havaintomatriisin geometrinen tulkinta

Tarkastellaan havaintomatriisia

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} 1 & 1 \\ 4 & 1 \\ 4 & 4 \end{pmatrix} \begin{array}{l} \text{Kalle} \\ \text{Ville} \\ \text{Maija} \end{array} \tag{1.29}$$

Siis kolmelta henkilöltä on mitattu kaksi ominaisuutta. Geometrisesti voimme käsitellä havaintomatriisin aineistoa kahdella tavalla:

(i) vaakariveittäin eli havainnoittain,

(ii) sarakkeittain eli muuttujittain.

Tapauksessa (i) meillä on kolme $\mathbb{R}^2$:n vektoria ja (ii):ssä kaksi $\mathbb{R}^3$:n vektoria. Voimme piirtää tilanteista kuviot 1.1a (s. 19) ja 1.1b (s. 19).

Kuviossa 1.1a on siis aivan tavanomainen kahden muuttujan pisteparvi, jossa kutakin havaintoa (havaintomatriisin vaakariviä) vastaa yksi piste ○. Kuviossa 1.1b puolestaan on havaintomatriisia tarkasteltu muuttujittain eli sarakkeittain. Siinä jokaista muuttujaa (havaintomatriisin saraketta) vastaa nyt yksi origosta lähtevä vektori (suunnattu jana), jonka päätepisteen koordinaatit ovat kyseisen muuttujan havaintoarvot. Voisimme tietysti merkitä havaintojakin geometrisesti vektoreina, mutta yleensä havainnoittain tarkastellessa käytämme pisteitä (tai muita sopivia symboleja) ja muuttujittain tarkastellen käytämme vektoreita.

Vektoriavaruutta, johon on piirretty muuttujien x ja y pisteparvi kuvion 1.1a tapaan, voimme sanoa

havaintoavaruudeksi.

Vastaavasti voimme sanoa kuvion 1.1b kolmiulotteista avaruutta

muuttuja-avaruudeksi.

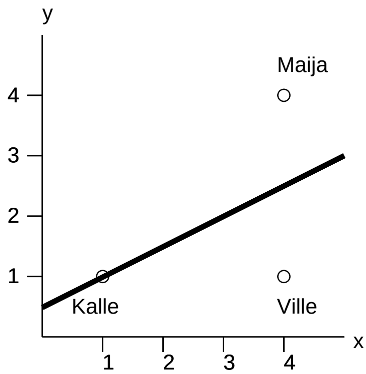
Molemmilla esitystavoilla on etunsa.

On hyvä havaita, että tilastollisen muuttujan (tai sitä vastaavan havaitun muuttujavektorin) ja puhtaasti matemaattisen vektorin välillä on tiettyjä eroja. Muuttujilla on yleensä jokin tilanteesta riippuva tausta, mittayksikkö ja tulkinta jne. Toisaalta voimme tietenkin käsitellä vektoreita puhtaasti matemaattisesti ottamatta mitenkään huomioon sitä tilastollista empiiristä tilannetta, missä kyseiset vektorit ovat syntyneet.

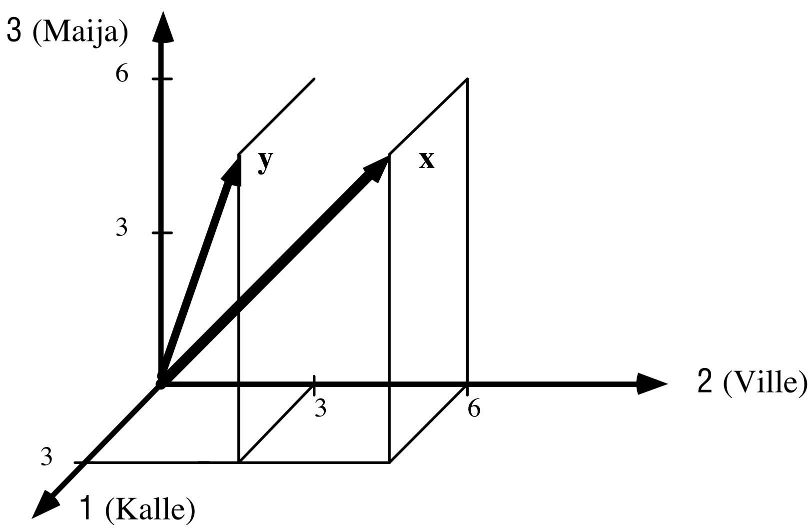
Jos lisäämme aineistoon (1.29) yhden havainnon (esim. Liisan) lisää, kuvion 1.1a pisteparveen lisättäisiin vain yksi piste. Sen sijaan kuvion 1.1b muuttujavektoriesityksessä meidän olisi lisättävä yksi koordinaattiakseli lisää, mutta muuttujavektoreita olisi edelleen vain kaksi. Koordinaattiakseleita on tässä tapauksessa ($\mathbb{R}^4$) hyvin hankala piirtää ja niinpä ne tälläisissa tapauksissa usein jätetään kuviosta kokonaan pois.

Monien tilastollisten menetelmien johtoajatuksen ymmärtäminen käy parhaiten päinsä sopivalla geometrisella tarkastelulla. Esimerkikiksi korrelaatiokertoimen r_{xy} geometrinen tulkinta on aivan välttämätön taito tilastotieteilijälle. Tunnetusti pisteparven tapauksessa $|r_{xy}|$ on sitä lähempänä ykköstä mitä suoranomaisemmin parvi on asettunut. Toisaalta tulemme havaitsemaan, että r_{xy} on keskistettyjen (keskiarvot nollia) muuttujavektoreiden välisen kulman kosini: suuri kosini (pieni kulma) kertoo suuresta korrelaatiosta.

Esimerkki 1.2. Korrelaatiokertoimeen liittyvät laskutoimitukset voidaan or-



Kuvio 1.1a. Pisteparvi aineistosta (1.29), koordinaattiakseleina muuttujat x ja y. Kuvioon on piirretty myös regressiosuora.



Kuvio 1.1b. Aineiston (1.29), kun $\mathbf{x} = (3, 6, 6)'$ ja $\mathbf{y} = (3, 3, 6)'$, geometrinen esitys muuttujittain, koordinaattiakseleina havainnot. Kuvioon on piirretty myös joitakin koordinaatteja selventäviä janoja.

ganisoida esim. seuraavasti:

	x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
	1	1	-2	4	-1	1	2
	4	1	1	1	-1	1	-1
	4	4	1	1	2	4	2
yht	9	6	0	$\text{SS}_x = 6$	0	$\text{SS}_y = 6$	$\text{SP}_{xy} = 3$

Tällöin $\bar{x} = 9/3 = 3$, $\bar{y} = 6/3 = 2$, ja

$$r_{xy} = \frac{3}{\sqrt{6 \cdot 6}} = 0.5, \quad \hat{\beta}_1 = \frac{\text{SP}_{xy}}{\text{SS}_x} = \frac{3}{6} = 0.5, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2 - 0.5 \cdot 3 = 0.5,$$

joten regressiosuoran yhtälö on

$$\hat{y} = 0.5 + 0.5 \cdot x \,. \tag{1.30}$$

Oheisessa kuviossa 1.2 (s. 21) on vielä ote Survon toimituskentästä, jossa on tehty regressioanalyysi em. aineistolle. Kuten näkyy, kolmestakin havainnosta voi laskea tuhdisti tunnuslukuja. □

Valmennusesimerkin 1.1 aineistosta (1.27) (s. 17) saamme kuviot 1.4a (s. 22) ja 1.4b (s. 23) Piirtämistaidot alkavat siis rakoilla, kun vektoriavaruuden dimensio kasvaa yli kolmen. Siksi kuviossa 1.4b on tyydytty piirtämään vain muuttujavektorit sinänsä näkyviin. Mikään ei estä meitä kuitenkaan kuvittelemasta niiden lojuvan 7-ulotteisessa vektoriavaruudessa $\mathbb{R}^7$.

Kuvioon 1.5 (s. 23) on piirretty kolmiulotteinen pisteparvi, joka on ryhmittynyt ellipsoidin sisälle. Siinä on siis kyseessä useita havaintoja muuttujista x, y ja z ja yhtä havaintoa vastaa yksi piste. Kaksiulotteisen pisteparven tulkinta on suoraviivaista, mutta lukija voi helposti kuvitella, että kolmiulotteista pisteparvea voisikin olla hyvä katsella eri suunnista.

1.2.1 Vektoreiden normit, kulmien kosinit ...

Muuttuja-avaruudessa meitä meitä voivat erityisesti kiinnostaa muuttujavektoreiden *pituudet* eli *normit*, keskinäiset *kulmat* ja vektorien *etäisyydet* toisistaan. Näillä matemaattisilla termeillä on luonnolliset tulkintansa tilastotieteen kannalta.

Vektoreiden

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{ja} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \tag{1.31}$$

välisen kulman kosini on tunnetusti (kun $\mathbf{x}, \mathbf{y} \neq \mathbf{0}$)

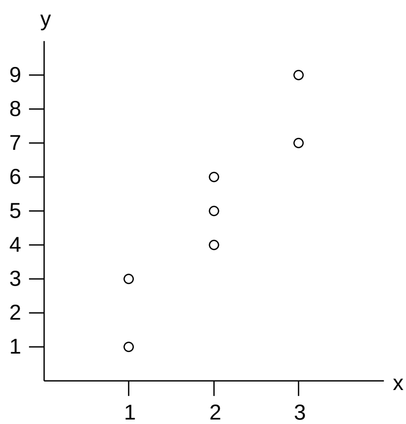
$$\cos(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i^2}} = \frac{\mathbf{x}'\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x} \cdot \mathbf{y}'\mathbf{y}}} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} \,. \tag{1.32}$$

```
 -   - SURVO MM  Thu Sep 23 11:26:21 2010                C:\SJP\D\   2000  100 0
*
*DATA KOE
*   Name     x   y
*   Kalle    1   1
*   Ville    4   1
*   Maija    4   4
*
*LINREG KOE,CUR+1  /   VARS=x(X),y(Y)
*Means, std.devs and correlations of KOE  N=3
*Variable  Mean        Std.dev.
*x         3.000000    1.732051
*y         2.000000    1.732051
*Correlations:
*             x       y
* x            1.0000  0.5000
* y            0.5000  1.0000
*
*Linear regression analysis: Data KOE, Regressand y         N=3
*Variable Regr.coeff.    Std.dev.    t      beta
*x         0.500000       0.866025  0.577  0.500
*constant  0.500000       2.872281  0.174
*Variance of regressand y=3.000000000 df=2
*Residual variance=4.500000000 df=1
*R=0.5000 R^2=0.2500
*
*/RANEW      /  tämä on sukro, joka muokkaa edellä olevaa tulostusta
*..............
*  Modifying regression output based on LINREG
*
*
*  REGRESSION ANALYSIS
*             Data: KOE   n=3   k=1
*       Y-variable: y
*      X-variables: x
*
* SS                   df       MS=SS/df               F=MSR/MSE
* SSR=1.5              dfr=1    MSR=1.5                F=0.3333333333333
* SSE=4.5              dfe=1    MSE=4.500000000
* SST=6                dft=2    MST=3.000000000
*   R^2=0.25      R=0.5                        P-value for F: Fp=0.66666
* 1-R^2=0.75
*
*Variable  Regr.coeff.      Std.dev.    t-value  P-value  Stcoef
*const.    b0=0.500000   se0=2.872281   t0=0.174  0.8903
*x         b1=0.500000   se1=0.866025   t1=0.577  0.6668   0.500
*                                                 (MK10-001, MKT-01-01)
```

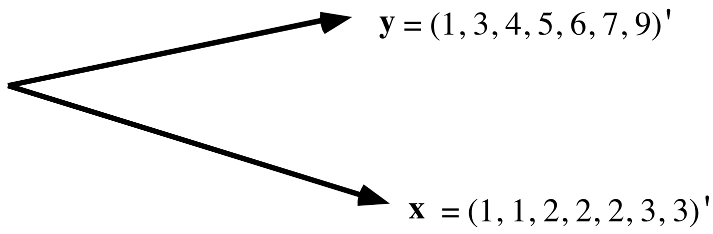
Kuvio 1.2. Survon toimituskenttä, jossa tehdään regressioanalyysi (1.29):n aineistolle.

```
 -  - SURVO MM  Thu Sep 23 11:32:14 2010            C:\SJP\D\   2000  100 0
*
*MATRIX U
*///     x   y
*Kalle   1   1
*Ville   4   1
*Maija   4   4
*
*MAT SAVE U
*MAT X=U(*,x)
*MAT Y=U(*,y)
*MAT A!=CENTER(X)    / *A~CENTER(U(*,x)) 3*1
*MAT B!=CENTER(Y)    / *B~CENTER(U(*,y)) 3*1
*MAT LOAD A          /  A = keskistetty muuttujavektori X
*MATRIX A
*///             x
*Kalle          -2
*Ville           1
*Maija           1
*
*MAT R=A'*B/(A'*A*B'*B)^(1/2)    / rxy = keskistettyjen mja-vektoreiden
*MAT LOAD R                      /       välisen kulman kosini
*MATRIX R
*A'*B/(A'*A*B'*B)^(1/2)
*///             y
*x        0.500000
*
*MAT SSX=A'*A        / *SSX~A'*A D1*1
*MAT SSY=B'*B        / *SSY~B'*B D1*1
*MAT SPXY=A'*B       / *SPXY~A'*B D1*1
*MAT_SSX(1,1)=6    MAT_SSY(1,1)=6        MAT_SPXY(1,1)=3
*r=MAT_SPXY(1,1)/sqrt(MAT_SSX(1,1)*MAT_SSX(1,1))
*r=0.5
*                                                    (MK10-001, MKT-01-02)
```

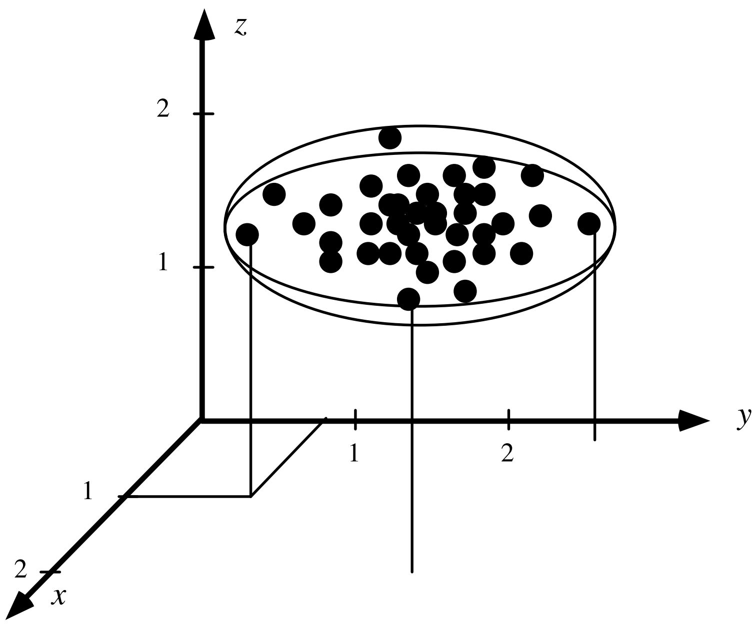
Kuvio 1.3. Matriisioperaatioita Survossa: korrelaatio on keskistettyjen muuttujavektoreiden välisen kulman kosini.



Kuvio 1.4a. Valmennusaineiston (1.27) pisteparvi.



Kuvio 1.4b. Valmennusaineiston (1.27) esitys muuttujavektoreina.



Kuvio 1.5. Kolmiulotteinen pisteparvi havaintoavaruudessa; kolme muuttujaa, 33 havaintoa.

Merkintä $\mathbf{x}'\mathbf{y}$ tarkoittaa vaakavektorin $\mathbf{x}'$ ja pystyvektorin $\mathbf{y}$ kertolaskua:

$$\mathbf{x}'\mathbf{y} = (x_1, x_2, \ldots, x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \,. \tag{1.33}$$

Merkinnällä $\langle \mathbf{x}, \mathbf{y} \rangle$ tarkoitamme vektorien $\mathbf{x}$ ja $\mathbf{y}$ *sisätuloa*:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = \langle \mathbf{y}, \mathbf{x} \rangle \,, \tag{1.34}$$

ja $\|\mathbf{x}\|$ tarkoittaa $\mathbf{x}$:n *pituutta* eli *normia*:

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \,. \tag{1.35}$$

Tietenkin on voimassa

$$-1 \leq \cos(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i^2}} \leq 1 \,, \tag{1.36}$$

josta vektorimerkinnöin saamme *Cauchyn–Schwarzin epäyhtälön*, (ks. luku 6.7 s. 207),

$$(\mathbf{x}'\mathbf{y})^2 \leq \mathbf{x}'\mathbf{x} \cdot \mathbf{y}'\mathbf{y}. \tag{CS}$$

Kosini $\cos(\mathbf{x}, \mathbf{y})$ on tietenkin 1 eli (CS):ssä on voimassa yhtäsuuruus, jos on olemassa sellainen luku $\lambda > 0$, että $\mathbf{x} = \lambda \mathbf{y}$, ja $\cos(\mathbf{x}, \mathbf{y}) = -1$, jos $\mathbf{x} = \mu \mathbf{y}$, missä $\mu < 0$.

Vektorit $\mathbf{x}$ ja $\mathbf{y}$ ovat kohtisuorassa toisiaan vastaan eli *ortogonaalisia*, jos $\mathbf{x}'\mathbf{y} = 0$; voimme merkitä lyhyesti $\mathbf{x} \perp \mathbf{y}$.

Jos vektorit $\mathbf{x}$ ja $\mathbf{y}$ ovat ykkösen pituisia, niin niiden välisen kulman kosini on yksinkertaisesti $\mathbf{x}'\mathbf{y}$:

$$\|\mathbf{x}\| = \|\mathbf{y}\| = 1 \implies \cos(\mathbf{x}, \mathbf{y}) = \mathbf{x}'\mathbf{y} \,. \tag{1.37}$$

Vektorien $\mathbf{x}$ ja $\mathbf{y}$ *etäisyyden* neliö $\mathrm{d}^2(\mathbf{x}, \mathbf{y})$ on erotuksen

$$\mathbf{x} - \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{pmatrix} \tag{1.38}$$

normin neliö:

$$\begin{aligned} \mathrm{d}^2(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\|^2 = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle \\ &= (\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y}) = \mathbf{x}'\mathbf{x} + \mathbf{y}'\mathbf{y} - \mathbf{x}'\mathbf{y} - \mathbf{y}'\mathbf{x} \\ &= \mathbf{x}'\mathbf{x} + \mathbf{y}'\mathbf{y} - 2\mathbf{x}'\mathbf{y} = \sum_{i=1}^n (x_i - y_i)^2. \end{aligned} \tag{1.39}$$

1.2.2 Sisätulon ja normin yleistys

Normi ja muut em. käsitteet voidaan määritellä myös muilla tavoin. Edellä määriteltyä normia sanotaan *euklidiseksi* normiksi tai l_2-normiksi, ja siitä käytetään usein merkintää $\|\mathbf{x}\|_2$, vastaavaa etäisyyttä euklidiseksi etäisyydeksi ja sisätuloa $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y}$ sanotaan tavanomaiseksi eli euklidiseksi sisätuloksi. Ellei erikseen mainita, tarkoitamme normilla nimenomaan euklidista normia.

Matriisi $\mathbf{V}_{n\times n}$ on ei-negatiivisesti definiitti eli lyhyesti merkiten $\mathbf{V} \in \mathrm{NND}_n$, jos on olemassa matriisi $\mathbf{L}$ siten, että $\mathbf{V} = \mathbf{L}'\mathbf{L}$; matriisien definiittisyyteen palataan täsmällisemmin luvussa 6.3 (s. 193). (Huomaa, että $\mathbf{V}$ on symmetrinen!) Mikäli $\mathbf{V}$ on lisäksi kääntyvä eli sen determinatti on nollasta poikkeava, on $\mathbf{V}$ positiivisesti definiitti: $\mathbf{V} \in \mathrm{PD}_n$. Nyt vektorien sisätulo voidaan määritellä myös siten, että

$$\begin{aligned}\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{V}} &= \mathbf{x}'\mathbf{V}\mathbf{y} = \mathbf{x}'\mathbf{L}'\mathbf{L}\mathbf{y} = (\mathbf{L}\mathbf{x})'\mathbf{L}\mathbf{y} := \mathbf{t}'\mathbf{u} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} v_{ij}x_i y_j = \sum_{i=1}^{n} t_i u_i \,, \end{aligned} \tag{1.40}$$

missä $\mathbf{V}$ on positiivisesti definiitti; (1.40):ssa on käytetty joitakin matriisitulon ominaisuuksia [kuten esim. $(\mathbf{L}\mathbf{x})' = \mathbf{x}'\mathbf{L}')$] joihin palataan tuonnempana. Näin määritelty funktio $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{V}}$ toteuttaa sisätulon yleisen määritelmän. Tällöin matriisia $\mathbf{V}$ sanotaan *sisätulomatriisiksi*. Vektorit $\mathbf{x}$ ja $\mathbf{y}$ ovat tämän sisätulon suhteen kohtisuorassa, jos

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{V}} = \mathbf{x}'\mathbf{V}\mathbf{y} = 0 \,. \tag{1.41}$$

Vastaava normi neliöitynä on ($\mathbf{V}$:n symmetrisyyden nojalla) lauseke

$$\begin{aligned}\|\mathbf{x}\|_{\mathbf{V}}^2 &= \langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{V}} = \mathbf{x}'\mathbf{V}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} v_{ij}x_i x_j \\ &= \sum_{i=1}^{n} v_{ii}x_i^2 + 2\sum_{i\neq j} v_{ij}x_i x_j \,. \end{aligned} \tag{1.42}$$

Esimerkiksi jos $\mathbf{V} \in \mathrm{NND}_2$, saamme (käyttäen matriisitulon ominaisuuksia)

$$\begin{aligned}\|\mathbf{x}\|_{\mathbf{V}}^2 &= (x_1, x_2)\begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2)\begin{pmatrix} v_{11}x_1 + v_{12}x_2 \\ v_{21}x_1 + v_{22}x_2 \end{pmatrix} \\ &= v_{11}x_1^2 + v_{12}x_1x_2 + v_{21}x_1x_2 + v_{22}x_2^2 \\ &= v_{11}x_1^2 + v_{22}x_2^2 + 2v_{12}x_1x_2 \,. \end{aligned} \tag{1.43}$$

Jos $\mathbf{V} \in \mathrm{NND}_3$, niin

$$\begin{aligned}\|\mathbf{x}\|_{\mathbf{V}}^2 &= (x_1, x_2, x_3)\begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= v_{11}x_1^2 + v_{22}x_2^2 + v_{33}x_3^2 + 2v_{12}x_1x_2 + 2v_{13}x_1x_3 + 2v_{23}x_2x_3 \,. \end{aligned} \tag{1.44}$$

Erityisesti jos $\mathbf{V}$ on lävistäjämatriisi, niin

$$\begin{aligned}\|\mathbf{x}\|^2_{\mathbf{V}} &= (x_1, x_2, x_3)\begin{pmatrix} v_{11} & 0 & 0 \\ 0 & v_{22} & 0 \\ 0 & 0 & v_{33}\end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3\end{pmatrix} \\ &= v_{11}x_1^2 + v_{22}x_2^2 + v_{33}x_3^2\,. \end{aligned} \tag{1.45}$$

Yleisesti tarkastellen sisätulolla $\mathbb{R}^n$:ssä tarkoitetaan reaaliarvoista funktiota $\langle \mathbf{x}, \mathbf{y}\rangle$, joka toteuttaa seuraavat ehdot:

ST1. $\langle \mathbf{x}, \mathbf{x}\rangle \geq 0$ ja $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ jos ja vain jos $\mathbf{x} = \mathbf{0}$,

ST2. $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle$,

ST3. $\langle \mathbf{x} + \mathbf{z}, \mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{z}, \mathbf{y}\rangle$,

ST4. $\langle \lambda\mathbf{x}, \mathbf{y}\rangle = \lambda\langle \mathbf{x}, \mathbf{y}\rangle, \lambda \in \mathbb{R}\,.$

Vektorinormin $\|\mathbf{x}\|$ yleiset perusominaisuudet ovat vastaavasti:

VN1. $\|\mathbf{x}\| \geq 0$ ja $\|\mathbf{x}\| = 0$ jos ja vain jos $\mathbf{x} = \mathbf{0}$,

VN2. $\|\lambda\mathbf{x}\| = |\lambda|\,\|\mathbf{x}\|$, $\lambda \in \mathbb{R}$,

VN3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Mainittakoon ohimennen, että jos normi määritellään sisätulon avulla siten, että $\|\mathbf{x}\| = +\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$, niin normin sanotaan olevan sisätulon indusoima. Normin ei kuitenkaan ole välttämätöntä olla sisätulon indusoima.

Tarkastellaan esimerkkinä kahden muuttujan havaintomatriisia

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix}, \quad \bar{\mathbf{u}} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, \quad \bar{\bar{\mathbf{x}}} = \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix} \in \mathbb{R}^n, \tag{1.46}$$

missä $\bar{\mathbf{u}} \in \mathbb{R}^2$ on muuttujien x ja y keskiarvovektori havaintoavaruudessa ja $\bar{\bar{\mathbf{x}}} \in \mathbb{R}^n$ on x:n keskiarvovektori muuttuja-avaruudessa. Tällöin tavanomaisen euklidisen etäisyyden mielessä

$$\begin{aligned} \|\mathbf{x} - \bar{\bar{\mathbf{x}}}\|^2 &= \sum_{i=1}^{n}(x_i - \bar{x})^2 = \text{muuttujavektorin } \mathbf{x} \text{ etäisyys } \bar{\bar{\mathbf{x}}}\text{:stä (neliöitynä)}, \\ \|\mathbf{u}_{(i)} - \bar{\mathbf{u}}\|^2 &= (x_i - \bar{x})^2 + (y_i - \bar{y})^2 \\ &= \text{havaintovektorin } \mathbf{u}_{(i)} \text{ etäisyys } \bar{\mathbf{u}}\text{:stä (neliöitynä)}. \end{aligned}$$

Tilastollisesti saattaisi olla mielenkiintoista tietää, kuinka monen hajonnan päässä i. havainnon x-arvo on $\bar{x}$:stä ja y-arvo on $\bar{y}$:stä. Näin voisi olla perusteltua laskea lauseke

$$\begin{aligned} D_i^2 &= \frac{(x_i-\bar{x})^2}{s_x^2} + \frac{(y_i-\bar{y})^2}{s_y^2} \\ &= (x_i-\bar{x}, y_i-\bar{y}) \begin{pmatrix} 1/s_x^2 & 0 \\ 0 & 1/s_y^2 \end{pmatrix} \begin{pmatrix} x_i-\bar{x} \\ y_i-\bar{y} \end{pmatrix} \\ &= (\mathbf{u}_{(i)}-\bar{\mathbf{u}})'\mathbf{S}_\delta^{-1}(\mathbf{u}_{(i)}-\bar{\mathbf{u}}) \\ &= \left\| \begin{matrix} (x_i-\bar{x})/s_x \\ (y_i-\bar{y})/s_y \end{matrix} \right\|^2, \end{aligned} \tag{1.47}$$

missä

$$\mathbf{S}_\delta = \operatorname{diag}(\mathbf{S}) = \begin{pmatrix} s_x^2 & 0 \\ 0 & s_y^2 \end{pmatrix}, \quad \mathbf{S}_\delta^{-1} = \begin{pmatrix} 1/s_x^2 & 0 \\ 0 & 1/s_y^2 \end{pmatrix}. \tag{1.48}$$

Lauseke (1.47) ei kuitenkaan mitenkään ota huomioon muuttujien x ja y välistä korrelaatiota. On helppo kuvitella pisteparvi, jossa D_i^2 on verraten pieni, mutta kyseinen havainto on kuitenkin jotenkin ns. ”pihalla parvesta”.

Havaintovektorin $\mathbf{u}_{(i)}$ *Mahalanobisin etäisyys* (neliöitynä) keskiarvovektorista $\bar{\mathbf{u}}$ määritellään lausekkeena

$$\begin{aligned} \mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S}) &= (\mathbf{u}_{(i)}-\bar{\mathbf{u}})'\mathbf{S}^{-1}(\mathbf{u}_{(i)}-\bar{\mathbf{u}}) = \|\mathbf{u}_{(i)}-\bar{\mathbf{u}}\|^2_{\mathbf{S}^{-1}} \\ &= \|\mathbf{S}^{-1/2}(\mathbf{u}_{(i)}-\bar{\mathbf{u}})\|^2, \end{aligned} \tag{1.49}$$

missä $\mathbf{S}$ on aineiston kovarianssimatriisi, $\mathbf{S}^{-1}$ on $\mathbf{S}$:n käänteismatriisi ja $\mathbf{S}^{-1/2}$ on $\mathbf{S}^{-1}$:n (positiivisesti definiitti symmetrinen) neliöjuuri (ks. s. 198). Mahalanobisin etäisyys ottaa muuttujien korrelaation huomioon. Osoittautuu, että joissakin tilanteissa Mahalanobisin etäisyys on paljon informatiivisempi tunnusluku kuin tavanomainen euklidinen etäisyys.

Esimerkki 1.3. Tarkastellaan tilannetta, jossa $s_x = s_y = 1$ eli kovarianssimatriisi $\mathbf{S}$ on korrelaatiomatriisi $\mathbf{R}$. Tällöin osoittautuu, että

$$\mathbf{R}^{-1} = \begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix}^{-1} = \frac{1}{1-r^2} \begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix}, \tag{1.50}$$

missä $r = \mathrm{cor_s}(x, y)$. Olkoon $\bar{x} = \bar{y} = 0$. Tällöin

$$\begin{aligned} \mathrm{MHLN}^2(\mathbf{u}_{(i)}, \mathbf{0}, \mathbf{R}) &= \mathbf{u}'_{(i)}\mathbf{R}^{-1}\mathbf{u}_{(i)} \\ &= \frac{1}{1-r^2}(x_i, y_i) \begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} \\ &= \frac{1}{1-r^2}(x_i^2 + y_i^2 - 2rx_iy_i). \end{aligned} \tag{1.51}$$

Olkoon $r = 0.9$. Nyt havaintovektoreilla $\mathbf{u}_{(1)} = \binom{-1}{1}$ ja $\mathbf{u}_{(2)} = \binom{1}{1}$ on sama eukklidinen etäisyys keskiarvovektorista $\mathbf{0}$, mutta

$$\mathrm{MHLN}^2(\mathbf{u}_{(1)}, \mathbf{0}, \mathbf{R}) = 380, \quad \mathrm{MHLN}^2(\mathbf{u}_{(2)}, \mathbf{0}, \mathbf{R}) = 20. \tag{1.52}$$

Piirrä tilannetta havainnollistava kuvio. □

Tarkastellaan joukkoa $\mathcal{A}$:

$$\mathcal{A} = \{ \mathbf{u} \in \mathbb{R}^2 : (\mathbf{u} - \bar{\mathbf{u}})'\mathbf{S}^{-1}(\mathbf{u} - \bar{\mathbf{u}}) = c^2 \} , \tag{1.53}$$

missä c on jokin annettu nollasta poikkeava reaaliluku. Joukko $\mathcal{A}$ on ellipsi keskipisteenä $\bar{\mathbf{u}}$. Tästä on pääteltävissä yksi Mahalanobisin etäisyyden tärkeä ominaisuus:

♠ Ne havainnot, joilla on sama Mahalanobisin etäisyys keskiarvosta $\bar{\mathbf{u}}$, sijaitsevat saman $\bar{\mathbf{u}}$-keskisen ellipsin kehällä. (1.54)

Mahalanobisin etäisyyden historiasta kertoo Das Gupta (1993); aiheesta ensimmäisen artikkelin laati Mahalanobis (1936).

Kommentti 1.2. Mainittakoon tässä (ennen aikojaan), että jos keskistetyille havainnoille $\mathbf{u}_{(i)} - \bar{\mathbf{u}}$ tehdään muunnos $\mathbf{S}^{-1/2}(\mathbf{u}_{(i)} - \bar{\mathbf{u}})$, niin uudet muuttujat ovat kaikki varianssiltaan 1:iä ja lisäksi korreloimattomia. Voimme kuvitella, että pisteparvi on siirretty origoon (= uusi keskiarvo) ja parvea on pyöräytetty ja kutistettu. Tässä uudessa parvessa havainnon etäisyys origosta on täsmälleen sama kuin (alkuperäisen) havainnon Mahalanobisin etäisyys keskiarvosta. Jos $\mathbf{C}$ on keskistäjämatriisi, ks. s. 93, niin matriisi $\underset{\sim}{\mathbf{U}} = \mathbf{CUS}^{-1/2}$ on juuri tällainen keskistetty havaintomatriisi, jossa kaikilla muuttujilla on varianssina 1 ja muuttujat ovat korreloimattomia. Lisäksi

$$\mathrm{diag}(\underset{\sim}{\mathbf{U}}\underset{\sim}{\mathbf{U}}') = \mathrm{diag}(d_1^2, \ldots, d_n^2) , \quad \text{missä } d_i^2 = \mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S}) . \tag{1.55}$$

□

1.2.3 Matriisien normit, sisätulot ...

Matriisien $\mathbf{A}_{n\times m}$ ja $\mathbf{B}_{n\times m}$ välinen (tavanomainen) sisätulo on

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ij} ; \tag{1.56}$$

ts. "asetetaan matriisit päällekkäin, kerrotaan päällekkäiset elementit keskenään ja lasketaan tulot yhteen". Vastaavasti matriisin $\mathbf{A}_{n\times m}$ tavanomaisella normilla eli *Frobeniuksen* normilla tarkoitetaan $\mathbf{A}$:n elementtien neliöiden summan neliöjuurta:

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2 . \tag{1.57}$$

Käytämme merkintää $\|\mathbf{A}\|_F$ kuvaamaan, että kyseessä on nimenomaan Frobeniuksen normi. Yleisesti tarkastellen matriisin $\mathbf{A} \in \mathbb{R}^{n\times m}$ normi määritellään reaaliarvoisena funktiona, joka toteuttaa seuraavat ehdot; ks. esim. Ben-Israel & Greville (2003, s. 13) ja Meyer (2000, §5.2).

MN1. $\|\mathbf{A}\| \geq 0$ ja $\|\mathbf{A}\| = 0$ jos ja vain jos $\mathbf{A} = \mathbf{0}$,

MN2. $\|\lambda\mathbf{A}\| = |\lambda|\,\|\mathbf{A}\|$, $\lambda \in \mathbb{R}$,

MN3. $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|,\ \mathbf{B} \in \mathbb{R}^{n\times m}$,

MN4. $\|\mathbf{AC}\| \leq \|\mathbf{A}\|\|\mathbf{C}\|,\ \mathbf{C} \in \mathbb{R}^{m\times p}$.

Yksi tapa määritellä matriisinormi on seuraava:

$$\begin{aligned}\|\mathbf{A}\|_2 &= \max_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 = \max_{\mathbf{x}\neq\mathbf{0}} \left(\frac{\mathbf{x}'\mathbf{A}'\mathbf{Ax}}{\mathbf{x}'\mathbf{x}}\right)^{1/2} \\ &= \sqrt{\mathrm{ch}_1(\mathbf{A}'\mathbf{A})} = \mathrm{sg}_1(\mathbf{A}), \end{aligned} \tag{1.58}$$

missä $\mathrm{ch}_i(\cdot)$ ja $\mathrm{sg}_i(\cdot)$ viittaavat matriisin $\mathbf{A}$ i. suurimpaan ominais- ja singulaariarvoon. Lauseketta (1.58) sanotaan euklidisen vektorinormin $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}'\mathbf{x}}$ indusoimaksi matriisinormiksi: *matriisin 2-normiksi* eli spektraalinormiksi. Vektori $\mathbf{x}$, joka johtaa (1.58):een, on siis se vektori, jota $\mathbf{A}$ ”venyttää” eniten ja siksi osoittautuu matriisin $\mathbf{A}'\mathbf{A}$ suurinta ominaisarvoa vastaava ominaisvektori. Ominaisarvoihin ja singulaariarvoihin palataan tuonnempana vielä lukuisia kertoja.

Matriisien $\mathbf{A}_{n\times m}$ ja $\mathbf{B}_{n\times m}$ euklidisen etäisyyden neliö on erotuksen

$$\mathbf{A}-\mathbf{B} = \begin{pmatrix} a_{11}-b_{11} & a_{12}-b_{12} & \dots & a_{1m}-b_{1m} \\ a_{21}-b_{21} & a_{22}-b_{22} & \dots & a_{2m}-b_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1}-b_{n1} & a_{n2}-b_{n2} & \dots & a_{nm}-b_{nm} \end{pmatrix} \tag{1.59}$$

Frobeniuksen normin neliö:

$$\mathrm{d}^2(\mathbf{A},\mathbf{B}) = \sum_{i=1}^{n}\sum_{j=1}^{m}(a_{ij}-b_{ij})^2 = \|\mathbf{A}-\mathbf{B}\|_F^2\,. \tag{1.60}$$

Ns. *vec-operaatiolla* tarkoitetaan matriisin sarakkeiden kirjoittamista alakkain siten että esimerkiksi

$$\mathrm{vec}(\mathbf{A}_{n\times m}) = \mathrm{vec}(\mathbf{a}_1 : \dots : \mathbf{a}_m) = \begin{pmatrix}\mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m\end{pmatrix} \in \mathbb{R}^{nm}. \tag{1.61}$$

Tämän merkintätavan nojalla voimme kirjoittaa yhtälön

$$\begin{aligned}\langle \mathbf{A}, \mathbf{B} \rangle &= \langle \mathrm{vec}(\mathbf{A}), \mathrm{vec}(\mathbf{B}) \rangle = [\mathrm{vec}(\mathbf{A})]' \,\mathrm{vec}(\mathbf{B}) \\ &= (\mathbf{a}_1', \ldots, \mathbf{a}_m') \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_m \end{pmatrix} = \mathbf{a}_1'\mathbf{b}_1 + \cdots + \mathbf{a}_m'\mathbf{b}_m \,, \qquad (1.62)\end{aligned}$$

missä $\langle \mathbf{A}, \mathbf{B} \rangle$ tarkoittaa matriisien $\mathbf{A}$ ja $\mathbf{B}$ välistä tavanomaista sisätuloa ja $\langle \mathrm{vec}(\mathbf{A}), \mathrm{vec}(\mathbf{B}) \rangle$ tarkoittaa vektorien $\mathrm{vec}(\mathbf{A})$ ja $\mathrm{vec}(\mathbf{B})$ välistä tavanomaista sisätuloa.

1.2.4 Matriisinormi matriisin jäljen lausekkeena

Osoitamme, että sisätulo $\langle \mathbf{A}, \mathbf{B} \rangle$ on esitettävissä lausekkeena

$$\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{tr}(\mathbf{A}'\mathbf{B}) = c_{11} + c_{22} + \cdots + c_{mm} \,, \qquad (1.63)$$

missä $\mathbf{A}'\mathbf{B} = \mathbf{C}$ ja $\mathrm{tr}(\cdot)$ tarkoittaa neliömatriisin *jälkeä* eli lävistäjäelementtien summaa. Olkoon $\mathbf{C}$ matriisien $\mathbf{A}' \in \mathbb{R}^{m\times n}$ ja $\mathbf{B} \in \mathbb{R}^{n\times m}$ tulo:

$$\mathbf{C}_{m\times m} = \mathbf{A}'\mathbf{B}, \quad \mathbf{A} = (\mathbf{a}_1 : \ldots : \mathbf{a}_m), \quad \mathbf{B} = (\mathbf{b}_1 : \ldots : \mathbf{b}_m) \,, \qquad (1.64a)$$

$$\mathbf{C} = \begin{pmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_m' \end{pmatrix} (\mathbf{b}_1 : \ldots : \mathbf{b}_m) = \begin{pmatrix} \mathbf{a}_1'\mathbf{b}_1 & \mathbf{a}_1'\mathbf{b}_2 & \ldots & \mathbf{a}_1'\mathbf{b}_m \\ \mathbf{a}_2'\mathbf{b}_1 & \mathbf{a}_2'\mathbf{b}_2 & \ldots & \mathbf{a}_2'\mathbf{b}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_m'\mathbf{b}_1 & \mathbf{a}_m'\mathbf{b}_2 & \ldots & \mathbf{a}_m'\mathbf{b}_m \end{pmatrix} . \qquad (1.64b)$$

Koska $c_{ii} = \mathbf{a}_i'\mathbf{b}_i$, saamme

$$\mathrm{tr}(\mathbf{A}'\mathbf{B}) = c_{11} + c_{22} + \cdots + c_{mm} = \mathbf{a}_1'\mathbf{b}_1 + \cdots + \mathbf{a}_m'\mathbf{b}_m = \langle \mathbf{A}, \mathbf{B} \rangle \,. \qquad (1.65)$$

Matriisin $\mathbf{A}$ Frobeniuksen normin neliö on näillä merkinnöillä

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2 = \langle \mathbf{A}, \mathbf{A} \rangle = \mathrm{tr}(\mathbf{A}'\mathbf{A}) \,, \qquad (1.66)$$

ja

$$\begin{aligned}\mathrm{d}^2(\mathbf{A}, \mathbf{B}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} (a_{ij} - b_{ij})^2 = \|\mathbf{A} - \mathbf{B}\|_F^2 \\ &= \langle \mathbf{A} - \mathbf{B}, \mathbf{A} - \mathbf{B} \rangle = \mathrm{tr}(\mathbf{A} - \mathbf{B})'(\mathbf{A} - \mathbf{B}) \\ &= \mathrm{tr}(\mathbf{A}'\mathbf{A} + \mathbf{B}'\mathbf{B} - \mathbf{A}'\mathbf{B} - \mathbf{B}'\mathbf{A}) \\ &= \mathrm{tr}(\mathbf{A}'\mathbf{A}) + \mathrm{tr}(\mathbf{B}'\mathbf{B}) - 2\,\mathrm{tr}(\mathbf{A}'\mathbf{B}) \,, \qquad (1.67)\end{aligned}$$

jossa on käytetty matriisin jäljen ominaisuutta

$$\operatorname{tr}(\mathbf{A}'\mathbf{B}) = \operatorname{tr}[(\mathbf{A}'\mathbf{B})'] = \operatorname{tr}(\mathbf{B}'\mathbf{A})\,. \tag{1.68}$$

Yllä on käytetty myös tulon transponointisääntöä

$$(\mathbf{A}'\mathbf{B})' = \mathbf{B}'\mathbf{A}\,. \tag{1.69}$$

1.2.5 Ortogonaaliprojektio suoralle

Kuvio 1.6 (s. 33) havainnollistaa vektoreiden $\mathbf{x} = \binom{4}{2}$ ja $\mathbf{y} = \binom{1}{2}$ välisen kulman kosinia. Kuvioon on myös merkitty vektori $\hat{\mathbf{y}}$, joka on $\mathbf{y}$:n *ortogonaaliprojektio* vektorin $\mathbf{x}$ virittämälle suoralle (ts. vektorin $\mathbf{x}$ kautta kulkevalle suoralle); tästä suorasta käytetään merkintää $\mathscr{C}(\mathbf{x})$:

$$\mathscr{C}(\mathbf{x}) = \{\, \lambda\mathbf{x} : \lambda \in \mathbb{R} \,\} = \text{vektorin } \mathbf{x} \text{ kautta kulkeva suora.} \tag{1.70}$$

Ortogonaaliprojektio $\hat{\mathbf{y}}$ voidaan määritellä siten, että

$$\hat{\mathbf{y}} = \hat{\lambda}\mathbf{x} = \text{se vektorin } \mathbf{x} \text{ monikerta, jonka etäisyys } \mathbf{y}\text{:stä on pienin,} \tag{1.71}$$

eli

$$\min_{\lambda} \|\mathbf{y} - \lambda\mathbf{x}\|^2 = \|\mathbf{y} - \hat{\lambda}\mathbf{x}\|^2, \tag{1.72}$$

tai yhtäpitävästi:

$$\hat{\mathbf{y}} = \mathbf{y}\text{:n ortogonaaliprojektio } \mathscr{C}(\mathbf{x})\text{:lle} \iff \mathbf{y} - \hat{\mathbf{y}} \perp \mathbf{x}\,. \tag{1.73}$$

Palaamme projisointiin tuonnempana moneen kertaan; varmaankaan (1.71) ja (1.73) eivät tulleet yllätyksenä.

Voimme ratkaista $\hat{\mathbf{y}}$:n (1.73):n nojalla:

$$(\mathbf{y} - \hat{\lambda}\mathbf{x})'\mathbf{x} = 0 \iff \mathbf{x}'\mathbf{x}\hat{\lambda} = \mathbf{y}'\mathbf{x} \iff \hat{\lambda} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}\,, \tag{1.74}$$

joten

$$\hat{\mathbf{y}} = \hat{\lambda}\mathbf{x} = \mathbf{x}\hat{\lambda} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} = \mathbf{P}_{\mathbf{x}}\mathbf{y}\,, \tag{1.75}$$

missä

$$\mathbf{P}_{\mathbf{x}} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}' = \frac{1}{\mathbf{x}'\mathbf{x}}\mathbf{x}\mathbf{x}' = \text{ ortogonaaliprojektori } \mathscr{C}(\mathbf{x})\text{:lle.} \tag{1.76}$$

Kannattaa panna merkille, että

$$\mathbf{P}_{\mathbf{x}} \in \mathbb{R}^{n\times n}, \quad \mathbf{x}\mathbf{x}' \in \mathbb{R}^{n\times n}, \quad \mathbf{x}'\mathbf{x} \in \mathbb{R}\,, \tag{1.77}$$

ja lisäksi $\mathbf{P_x}$ on symmetrinen ja idempotentti eli

$$\mathbf{P}'_{\mathbf{x}} = \mathbf{P_x} \quad \text{ja} \quad \mathbf{P_x}\mathbf{P_x} = \mathbf{P_x}\,. \tag{1.78}$$

Kuvion 1.6 tapauksessa

$$\hat{\lambda} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} = \frac{4+4}{16+4} = \frac{4}{10}\,, \qquad \hat{\mathbf{y}} = \hat{\lambda}\begin{pmatrix}4\\2\end{pmatrix} = \begin{pmatrix}1.6\\0.8\end{pmatrix}. \tag{1.79}$$

Kommentti 1.3. (Ortogonaaliprojektori $\mathscr{C}(\mathbf{A})$:lle.) Yleisemmässä tapauksessa ortogonaaliprojektoriksi (tavanomaisen sisätulon suhteen) matriisin $\mathbf{A}$ sarakkeiden virittämälle sarakeavaruudelle $\mathscr{C}(\mathbf{A})$ osoittautuu matriisi

$$\mathbf{P_A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}', \tag{1.80}$$

missä oletetaan, että $\mathbf{A}'\mathbf{A}$ on kääntyvä matriisi. Tallöin vektorin $\mathbf{y}$ ortogonaaliprojektio $\mathscr{C}(\mathbf{A})$:lle on

$$\hat{\mathbf{y}} = \mathbf{P_A}\mathbf{y} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} := \mathbf{A}\mathbf{b}_*\,. \tag{1.81}$$

Projektiolla $\hat{\mathbf{y}}$ on ominaisuus

$$\min_{\mathbf{b}} \|\mathbf{y} - \mathbf{A}\mathbf{b}\|^2 = \|\mathbf{y} - \mathbf{P_A}\mathbf{y}\|^2 = \|\mathbf{y} - \mathbf{A}\mathbf{b}_*\|^2. \tag{1.82}$$

Matriisilla $\mathbf{P_A}$ on ominaisuudet

$$\mathbf{P_A} = \mathbf{P}_{\mathbf{A}}^2 = \mathbf{P}'_{\mathbf{A}}. \tag{1.83}$$

Osoittautuu, että jos neliömatriisi $\mathbf{P}$ toteuttaa ehdot

$$\mathbf{P} = \mathbf{P}^2 = \mathbf{P}', \tag{1.84}$$

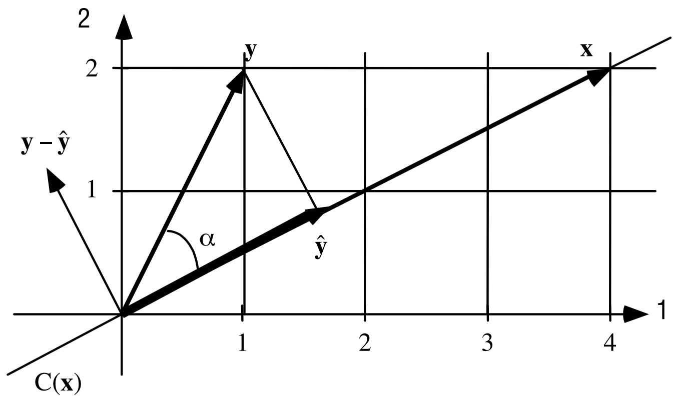
niin $\mathbf{P}$ on ortogonaaliprojektori $\mathscr{C}(\mathbf{P})$:lle. Mikäli $\mathbf{P}^2 = \mathbf{P}$, niin $\mathbf{P}$ on projektori suuntaan $\mathscr{C}(\mathbf{I}_n - \mathbf{P})$. Jos $\mathbf{P}$ on lisäksi symmetrinen, $\mathbf{P}$ on ortogonaaliprojektori. Täsmällisempi ortogonaaliprojektorin käsittely on luvuissa 8.1.1 (s. 252) ja 9.7 (s. 318). □

Tarkastellaan toisena esimerkkinä (kuviot 1.7a, s. 34, ja 1.7b, s. 34) muuttujavektoreita

$$\mathbf{x} = (-1, 1, 0)', \quad \mathbf{y} = (0, 1, -1)'. \tag{1.85}$$

Tällöin $\mathbf{x}'\mathbf{y} = 1$, $\|\mathbf{x}\| = \|\mathbf{y}\| = \sqrt{2}$, joten $\cos(\mathbf{x},\mathbf{y}) = 0.5$. Tässä tapauksessa muuttujien $\mathbf{x}$ ja $\mathbf{y}$ keskiarvot ovat nollia, $\bar{x} = \bar{y} = 0$, ja siksi

$$\mathrm{cor_d}(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}} = \cos(\mathbf{x},\mathbf{y}) = 0.5\,. \tag{1.86}$$



Kuvio 1.6. Vektoreiden $\mathbf{x}$ ja $\mathbf{y}$ välisen kulman kosini geometrisesti: $\cos(\mathbf{x}, \mathbf{y}) = \cos(\alpha) = 0.8$. Vektori $\hat{\mathbf{y}}$ on $\mathbf{y}$:n ortogonaaliprojektio $\mathbf{x}$:n kautta kulkevalle suoralle $\mathscr{C}(\mathbf{x})$.

Sanomme, että

$$\text{muuttuja on keskistetty, jos sen keskiarvo on nolla.} \tag{1.87}$$

Havaitsemme (1.86):sta välittömästi, että jos muuttujien keskiarvot ovat nollia, niin muuttujavektoreiden välisen kulman kosini on sama kuin muuttujien välinen korrelaatiokerroin ts.

$$\text{keskistettyjen muuttujien tapauksessa } \operatorname{cor_d}(\mathbf{x}, \mathbf{y}) = \cos(\mathbf{x}, \mathbf{y}). \tag{1.88}$$

Täten on erityisesti voimassa seuraava tulos:

$$\begin{gathered}\text{Keskistettyjen muuttujien tapauksessa} \\ \text{korreloimattomuus ja ortogonaalisuus ovat yhtäpitäviä asioita.}\end{gathered} \tag{1.89}$$

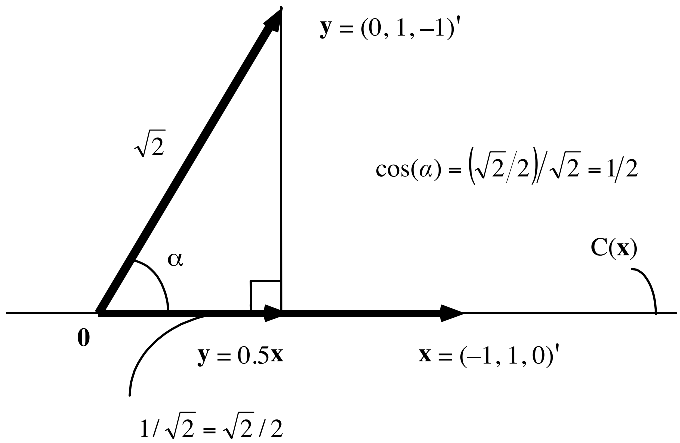
Kuvio 1.7a havainnollistaa aineistoa (1.85). Kuvioon on myös merkitty vektori $\hat{\mathbf{y}}$, joka on $\mathbf{y}$:n ortogonaaliprojektio vektorin $\mathbf{x}$ virittämälle suoralle. Luonnollisesti myös kaava $\|\hat{\mathbf{y}}\|/\|\mathbf{y}\|$ antaa kosiniksi 0.5. Kuviossa 1.7b on sama tilanne kuin kuviossa 1.7a, mutta katsoja ”näkee” vektorit eri kulmasta ja näin saadaan paremmin aikaan vaikutelma, että vektorit $\mathbf{x}$ ja $\mathbf{y}$ todella virittävät (origon kautta kulkevan) tason; tätä tasoa merkitään symbolilla $\mathscr{C}(\mathbf{x} : \mathbf{y})$.

Tarkastellaan vielä muuttujaparia

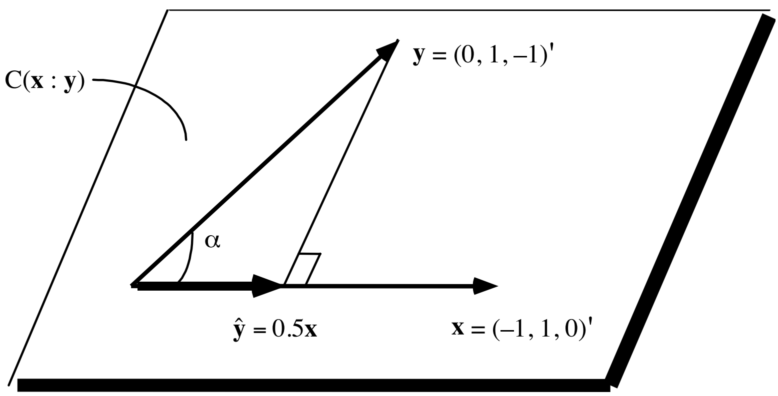
$$\mathbf{x} = (1, 1, -\sqrt{2})', \quad \mathbf{y} = (1, 1, \sqrt{2})'. \tag{1.90}$$

Tällöin $\mathbf{x}$ ja $\mathbf{y}$ ovat kohtisuorassa toisiaan vastaan: $\mathbf{x}'\mathbf{y} = 0$. Ovatko $\mathbf{x}$ ja $\mathbf{y}$ myös korreloimattomia? Eivät suinkaan, sillä havaitsemme (miksi?) heti, että $\operatorname{cor_d}(\mathbf{x}, \mathbf{y}) = -1$. Korrelaatiokertoimen geometriseen tulkintaan palaamme uudestaan myöhemmin.

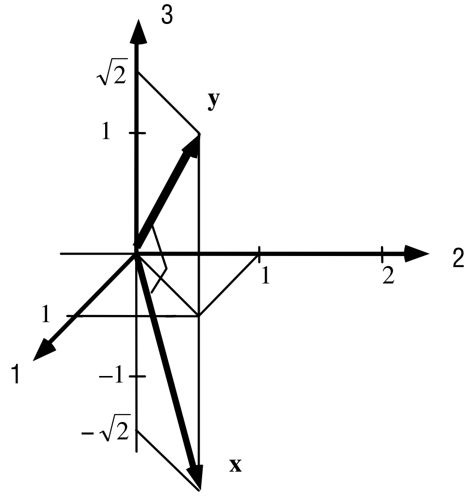
Lukija voi piirtää kuvioon 1.8 myös muuttujavektorit $\mathbf{x} = (0, 1, 0)'$ ja $\mathbf{y} = (1, 0, 1)'$, ja tutkia niiden ortogonaalisuutta ja korrelaatiota.



Kuvio 1.7a. Kolmiulotteisten keskistettyjen muuttujavektoreiden $\mathbf{x}$ ja $\mathbf{y}$ välisen kulman kosini geometrisesti: $\cos(\mathbf{x}, \mathbf{y}) = \cos(\alpha) = \mathrm{cor_d}(\mathbf{x}, \mathbf{y}) = 0.5$. Vektori $\hat{\mathbf{y}}$ on $\mathbf{y}$:n ortogonaaliprojektio $\mathbf{x}$:n kautta kulkevalle suoralle $\mathscr{C}(\mathbf{x})$. [Kuviossa $\hat{\mathbf{y}} = 0.5\mathbf{x}$.]



Kuvio 1.7b. Sama kuin kuvio 1.7a, mutta tilanteen katselukulmaa on vaihdettu. Sarakeavaruus $\mathscr{C}(\mathbf{x} : \mathbf{y})$ on taso, jonka $\mathbf{x}$ ja $\mathbf{y}$ virittävät eli $\mathscr{C}(\mathbf{x} : \mathbf{y})$ on $\mathbf{x}$:n ja $\mathbf{y}$:n kaikkien lineaarikombinaatioiden joukko.



Kuvio 1.8. Muuttujavektorit $\mathbf{x} = (1, 1, -\sqrt{2})'$ ja $\mathbf{y} = (1, 1, \sqrt{2})'$ ovat ortogonaalisia eli $\cos(\mathbf{x}, \mathbf{y}) = 0$, mutta $\mathrm{cor}_\mathrm{d}(\mathbf{x}, \mathbf{y}) = -1$.

1.3 Satunnaisvektorit

Edellä esitetyt tunnusluvut cor_d (korrelaatio), var_d (varianssi) ja cov_d (kovarianssi) ovat kaikki otossuureita, ts. ne on laskettu otoksen perusteella. Niiden teoreettisia vastineita merkitään symbolein cor, var ja cov. Palautettakoon mieleen että jos z_1 ja z_2 ovat satunnaismuuttujia, niin:

$$\mathrm{var}(z_i) = \mathrm{E}(z_i - \mu_i)^2 = \mathrm{E}(z_i^2) - \mu_i^2 = \sigma_i^2\,, \;\; \mu_i = \mathrm{E}(z_i)\,, \tag{1.91a}$$

$$\begin{aligned}\mathrm{cov}(z_1, z_2) &= \mathrm{E}(z_1 - \mu_1)(z_2 - \mu_2) = \mathrm{E}(z_1 z_2) - \mu_1 \mu_2 \\ &= \sigma_{12} = \varrho_{12}\sigma_1\sigma_2\,,\end{aligned} \tag{1.91b}$$

$$\mathrm{cor}(z_1, z_2) = \frac{\mathrm{cov}(z_1, z_2)}{\sqrt{\mathrm{var}(z_1)\,\mathrm{var}(z_2)}} = \frac{\sigma_{12}}{\sigma_1\sigma_2} = \varrho_{12}\,, \tag{1.91c}$$

missä $\mathrm{E}(y)$ viittaa satunnaismuuttujan y odotusarvoon. Yleensä matemaattisessa tilastotieteessä satunnaismuuttujia merkitään isoin kirjaimin ja niiden arvoja vastaavin pienin kirjaimin. Emme kuitenkaan tässä esityksessä noudata tätä sääntöä; syynä on se, että matriisien kanssa pelatessa on hyvin vaikea olla tässä suhteessa johdonmukainen. Pyrimme kuitenkin merkitsemään satunnaismuuttujia aakkosten loppupään kirjaimilla. Täten esimerkiksi

- x ja X voivat tarkoittaa reaalilukua tai 1-ulotteista satunnaismuuttujaa,
- merkintä $\mathbf{x}$ voi viitata $\mathbb{R}^p$:n vektoriin tai p-ulotteiseen satunnaisvektoriin,
- $\mathbf{X}_{n\times p}$ voi tarkoittaa reaalielementtistä $n \times p$-matriisia tai satunnaismatriisia, jonka jokainen elementti on satunnaismuuttuja.

Korrelaatiokerroin ϱ_{12} on hyvin määritelty vain jos σ_1 ja σ_2 ovat nollaa suurempia. Mikäli $\sigma_1 = 0$, niin välttämättä myös $\sigma_{12} = 0$, ja korrelaatioksi saataisiin $\frac{0}{0}$. Lisäksi luonnollisesti oletamme, että $\sigma_1 < \infty$ ja $\sigma_2 < \infty$.

Kommentti 1.4. *Odotusarvo, varianssi ja kovarianssi diskreetin todennäköisyysjakauman tapauksessa.* Olkoot x ja y ovat diskreettejä satunnaismuuttujia siten että x:n mahdolliset arvot ovat $x_1, x_2, \ldots, x_k$ ja vastaavat todennäköisyydet $p_1, p_2, \ldots, p_k$, ja y:n mahdolliset arvot ovat $y_1, y_2, \ldots, y_\ell$ ja vastaavat todennäköisyydet $q_1, q_2, \ldots, q_\ell$. Olkoon lisäksi $\mathrm{P}\{x = x_i, y = y_j\} = p_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, \ell$. Tällöin

$$\mathrm{E}(x) = \mu_x = p_1 x_1 + p_2 x_2 + \cdots + p_k x_k \,, \tag{1.92a}$$

$$\begin{aligned} \mathrm{var}(x) &= \mathrm{E}(x - \mu_x)^2 = \mathrm{E}(x^2) - \mu_x^2 \\ &= p_1(x_1 - \mu_x)^2 + p_2(x_2 - \mu_x)^2 + \cdots + p_k(x_k - \mu_x)^2 \\ &= p_1 x_1^2 + p_2 x_2^2 + \cdots + p_k x_k^2 - \mu_x^2 \,, \end{aligned} \tag{1.92b}$$

$$\begin{aligned} \mathrm{cov}(x, y) &= \mathrm{E}(x - \mu_x)(y - \mu_y) = \mathrm{E}(xy) - \mu_x \mu_y \\ &= \sum_{i=1}^{k} \sum_{j=1}^{\ell} p_{ij}(x_i - \mu_x)(y_j - \mu_y) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} p_{ij} x_i y_j - \mu_x \mu_y \,. \end{aligned} \tag{1.92c}$$

□

Esimerkki 1.4. Tarkastellaan kuvion 1.9 mukaista kaksiulotteinen diskreettiä todennäköisyysjakaumaa, jossa on 6 xy-arvoparia, joilla kaikilla on sama todennäköisyys $\frac{1}{6}$. Saamme

$$\begin{aligned} \mathrm{E}(x) = \mu_x &= \tfrac{3}{6}\,1 + \tfrac{2}{6}\,2 + \tfrac{1}{6}\,3 \\ &= \tfrac{1}{6}\,(3 \cdot 1 + 2 \cdot 2 + 1 \cdot 3) = \tfrac{10}{6} = \tfrac{5}{3} \,, \end{aligned} \tag{1.93a}$$

$$\mathrm{E}(y) = \mu_y = \tfrac{14}{6} = \tfrac{7}{3} \,, \tag{1.93b}$$

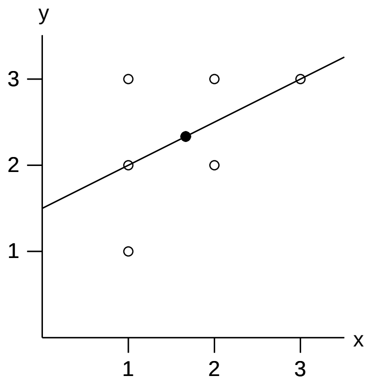
$$\begin{aligned} \mathrm{var}(x) &= \tfrac{3}{6}\,(1 - \mu_x)^2 + \tfrac{2}{6}\,(2 - \mu_x)^2 + \tfrac{1}{6}\,(3 - \mu_x)^2 \\ &= \tfrac{1}{6}[3(1 - \mu_x)^2 + 2(2 - \mu_x)^2 + 1(3 - \mu_x)^2] \\ &= \tfrac{1}{6}[3 \cdot 1^2 + 2 \cdot 2^2 + 1 \cdot 3^2] - \mu_x^2 \\ &= \tfrac{5}{9} = \sigma_x^2 = \sigma_y^2 = \mathrm{var}(y) \,. \end{aligned} \tag{1.93c}$$

Laskemme kovarianssin $\mathrm{cov}(x, y) = \sigma_{xy}$ hieman kiertotietä. Nimittäin on osoitettavissa seuraava tulos:

$$\mathrm{E}(y \mid x = \underline{x}) = \alpha + \beta \underline{x} \tag{1.94a}$$

$$\Longrightarrow$$

$$\alpha = \mu_y - \beta \cdot \mu_x \,, \qquad \beta = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \frac{\sigma_y}{\sigma_x} = \varrho_{xy} \frac{\sigma_y}{\sigma_x} \,. \tag{1.94b}$$



Kuvio 1.9. (Ks. esimerkki 1.4.) Kaksiulotteinen diskreetti todennäköisyysjakauma: kunkin arvoparin todennäköisyys sama eli $\frac{1}{6}$. Kuvioon piiretty myös keskiarvopiste sekä regressiosuora – kun käsitellään pisteparvea datana. Huomaa, että ehdolliset keskiarvot ovat suoralla $y = 1.5 + 0.5x$.

Toisin sanoen: mikäli y:n ehdollinen odotusarvo kun x:llä on arvo $\underline{x}$ on muotoa $\alpha + \beta\underline{x}$, niin $\alpha = \mu_y - \beta \cdot \mu_x$ ja $\beta = \varrho_{xy}\frac{\sigma_y}{\sigma_x}$. Kuviosta 1.9 näemme välittömästi, että $\mathrm{E}(y \mid x = \underline{x}) = 1.5 + 0.5\underline{x}$, joten σ_x:n ja σ_y:n yhtäsuuruuden takia

$$\mathrm{cor}(x, y) = \varrho_{xy} = 0.5\,, \quad \mathrm{cov}(x, y) = \sigma_{xy} = \varrho_{xy}\sigma_x^2 = \tfrac{1}{2} \cdot \tfrac{5}{9} = \tfrac{5}{18}\,. \tag{1.95}$$

Mainittakoon, että (1.94) perustuu sellaiseen faktaan, että jos ehdollisten odotusarvojen ura on suora, niin kyseinen suora on nimenomaan regressiosuora. Satunnaismuuttujien tapauksessa regressiosuoran kulmakerroin on $\frac{\sigma_{xy}}{\sigma_x^2}$ ja se kulkee pisteen (μ_x, μ_y) kautta; ja empiirisestä datasta laskettu regressiosuoran kulmakerroin on $\frac{s_{xy}}{s_x^2}$ ja se kulkee pisteen $(\bar{x}, \bar{y})$ kautta.

On selvää, että ehdollisten odotusarvojen (tai keskiarvojen) ei tarvitse välttämättä sijaita samalla suoralla.

Täten satunnaisvektorin $\binom{x}{y}$ odotusarvo $\boldsymbol{\mu}$ ja kovarianssimatriisi $\boldsymbol{\Sigma}$ ovat

$$\mathrm{E}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mathrm{E}(x) \\ \mathrm{E}(y) \end{pmatrix} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} = \begin{pmatrix} 5/3 \\ 7/3 \end{pmatrix} = \boldsymbol{\mu}\,, \tag{1.96a}$$

$$\mathrm{cov}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} 5/9 & 5/18 \\ 5/18 & 5/9 \end{pmatrix} = \frac{10}{18}\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} = \boldsymbol{\Sigma}\,. \tag{1.96b}$$

□

Kommentti 1.5. Entäpä jos kuvio 1.9 kuvaa yksinkertaisesti vain otoksesta muodostettua pisteparvea? Merkitään

$$\mathbf{U} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 2 & 2 \\ 2 & 3 \\ 3 & 3 \end{pmatrix} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \mathbf{u}'_{(3)} \\ \mathbf{u}'_{(4)} \\ \mathbf{u}'_{(5)} \\ \mathbf{u}'_{(6)} \end{pmatrix}, \quad \bar{\mathbf{u}} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} 5/3 \\ 7/3 \end{pmatrix}, \tag{1.97}$$

$$\widetilde{\mathbf{U}} = \begin{pmatrix} 1-\bar{x} & 1-\bar{y} \\ 1-\bar{x} & 2-\bar{y} \\ 1-\bar{x} & 3-\bar{y} \\ 2-\bar{x} & 2-\bar{y} \\ 2-\bar{x} & 3-\bar{y} \\ 3-\bar{x} & 3-\bar{y} \end{pmatrix} = (\tilde{\mathbf{x}} : \tilde{\mathbf{y}}) = \begin{pmatrix} (\mathbf{u}_{(1)} - \bar{\mathbf{u}})' \\ (\mathbf{u}_{(2)} - \bar{\mathbf{u}})' \\ (\mathbf{u}_{(3)} - \bar{\mathbf{u}})' \\ (\mathbf{u}_{(4)} - \bar{\mathbf{u}})' \\ (\mathbf{u}_{(5)} - \bar{\mathbf{u}})' \\ (\mathbf{u}_{(6)} - \bar{\mathbf{u}})' \end{pmatrix}. \tag{1.98}$$

Tällöin esim. x:n otosvarianssi on

$$\begin{aligned} s_x^2 &= \tfrac{1}{6-1}[3\cdot(1-\bar{x})^2 + 2\cdot(2-\bar{x})^2 + (3-\bar{x})^2] \\ &= \tfrac{6}{5}\tfrac{1}{6}[3\cdot(1-\mu_x)^2 + 2\cdot(2-\mu_x)^2 + (3-\mu_x)^2] \\ &= \tfrac{6}{5}\sigma_x^2, \end{aligned} \tag{1.99}$$

sillä

$$\sigma_x^2 = \tfrac{1}{6}[3\cdot(1-\mu_x)^2 + 2\cdot(2-\mu_x)^2 + (3-\mu_x)^2]. \tag{1.100}$$

Kaiken kaikkiaan on voimassa

$$\bar{x} = \mu_x, \quad \bar{y} = \mu_y, \quad s_x^2 = s_y^2 = \tfrac{6}{5}\sigma_x^2, \quad s_{xy} = \tfrac{6}{5}\sigma_{xy}, \quad r_{xy} = \varrho_{xy}, \tag{1.101}$$

eli otoksesta laskettu $\bar{\mathbf{u}}$ ja $\mathbf{S}$ ovat

$$\bar{\mathbf{u}} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} 5/3 \\ 7/3 \end{pmatrix} = \boldsymbol{\mu}, \quad \mathbf{S} = \tfrac{6}{5}\boldsymbol{\Sigma}. \tag{1.102}$$

Korrelaatiomatriisit ovat kuitenkin täsmälleen samat olipa kyseessä otos tai teoreettinen todennäköisyysjakauma. Tästä on pääteltävissä, että korrelaatioihin perustuvat tulokset ovat havaintoaineiston ja vastaavan teoreettisen jakauman (jossa jokaisella havainnolla sama todennäköisyys) tapauksessa yhtäpitäviä. Seberin sanoin (2008, s. 433): voimme "siirtää" otosominaisuudet poulaatio-ominaisuuksiksi käyttämällä sopivaa diskreettiä populaatiota.

Todettakoon vielä [ks. (2.119), s. 101], että $\mathbf{U}$:hun perustuva otoskovarianssimatriisi $\mathbf{S}$ on

$$\begin{aligned}
\operatorname{cov_d}(\mathbf{U}) = \mathbf{S} &= \tfrac{1}{6-1}\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}} = \tfrac{1}{6-1}\begin{pmatrix} \tilde{\mathbf{x}}'\tilde{\mathbf{x}} & \tilde{\mathbf{x}}'\tilde{\mathbf{y}} \\ \tilde{\mathbf{y}}'\tilde{\mathbf{x}} & \tilde{\mathbf{y}}'\tilde{\mathbf{y}} \end{pmatrix} \\
&= \tfrac{1}{6-1}(\mathbf{u}_{(1)} - \bar{\mathbf{u}} : \ldots : \mathbf{u}_{(6)} - \bar{\mathbf{u}}) \begin{pmatrix} (\mathbf{u}_{(1)} - \bar{\mathbf{u}})' \\ \vdots \\ (\mathbf{u}_{(6)} - \bar{\mathbf{u}})' \end{pmatrix} \\
&= \tfrac{1}{6-1}\sum_{i=1}^{6}(\mathbf{u}_{(i)} - \bar{\mathbf{u}})(\mathbf{u}_{(i)} - \bar{\mathbf{u}})' := \tfrac{1}{6-1}\mathbf{T}\,. \qquad (1.103)
\end{aligned}$$

□

Satunnaisvektorin $\mathbf{z} = \binom{z_1}{z_2}$ odotusarvovektori on

$$\mathrm{E}(\mathbf{z}) = \begin{pmatrix} \mathrm{E}(z_1) \\ \mathrm{E}(z_2) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \boldsymbol{\mu} \qquad (1.104)$$

ja $\mathbf{z}$:n kovarianssimatriisi on

$$\begin{aligned}
\operatorname{cov}(\mathbf{z}) = \operatorname{cov}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \boldsymbol{\Sigma} &= \begin{pmatrix} \operatorname{var}(z_1) & \operatorname{cov}(z_1, z_2) \\ \operatorname{cov}(z_2, z_1) & \operatorname{var}(z_2) \end{pmatrix} \qquad (1.105) \\
&= \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho_{12} \\ \sigma_1\sigma_2\varrho_{12} & \sigma_2^2 \end{pmatrix}.
\end{aligned}$$

Vastaava korrelaatiomatriisi on

$$\operatorname{cor}(\mathbf{z}) = \begin{pmatrix} 1 & \varrho_{12} \\ \varrho_{21} & 1 \end{pmatrix}. \qquad (1.106)$$

Voimme käyttää lyhennettyä merkintää $\mathbf{z} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ilmaisemaan, että satunnaisvektorin $\mathbf{z}$ odotusarvo on $\boldsymbol{\mu}$, ja kovarianssimatriisi on $\boldsymbol{\Sigma}$.

Jos $\mathbf{x}$ on 2-ulotteinen ja $\mathbf{y}$ on 3-ulotteinen satunnaisvektori,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \qquad (1.107)$$

niin niiden keskinäinen kovarianssimatriisi on

$$\operatorname{cov}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} \operatorname{cov}(x_1, y_1) & \operatorname{cov}(x_1, y_2) & \operatorname{cov}(x_1, y_3) \\ \operatorname{cov}(x_2, y_1) & \operatorname{cov}(x_2, y_2) & \operatorname{cov}(x_2, y_3) \end{pmatrix} := \boldsymbol{\Sigma}_{\mathbf{xy}}\,. \qquad (1.108)$$

Tällöin tietenkin

$$\operatorname{cov}(\mathbf{x}, \mathbf{x}) = \operatorname{cov}(\mathbf{x}) = \boldsymbol{\Sigma}_{\mathbf{xx}}\,, \qquad (1.109a)$$

$$\operatorname{cov}(\mathbf{y}, \mathbf{x}) = [\operatorname{cov}(\mathbf{x}, \mathbf{y})]' = \boldsymbol{\Sigma}_{\mathbf{yx}}\,. \qquad (1.109b)$$

Olkoon $\mathbf{a} \in \mathbb{R}^n$ ja $\mathbf{b} \in \mathbb{R}^p$. Tällöin kertolasku $\mathbf{ab}'$ määritellään siten, että

$$\begin{aligned}\mathbf{ab}' &= \begin{pmatrix} a_1\mathbf{b}' \\ a_2\mathbf{b}' \\ \vdots \\ a_n\mathbf{b}' \end{pmatrix} = (b_1\mathbf{a} : b_2\mathbf{a} : \ldots : b_p\mathbf{a}) \\ &= \begin{pmatrix} a_1b_1 & a_1b_2 & \ldots & a_1b_p \\ a_2b_1 & a_2b_2 & \ldots & a_2b_p \\ \vdots & \vdots & & \vdots \\ a_nb_1 & a_nb_2 & \ldots & a_nb_p \end{pmatrix} = \{a_ib_j\} \in \mathbb{R}^{n\times p}. \end{aligned} \tag{1.110}$$

Jos muodostamme 2×2-matriisin

$$\begin{aligned}(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})' &= \begin{pmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{pmatrix} (z_1-\mu_1, z_2-\mu_2) \\ &= \begin{pmatrix} (z_1-\mu_1)^2 & (z_1-\mu_1)(z_2-\mu_2) \\ (z_2-\mu_2)(z_1-\mu_1) & (z_2-\mu_2)^2 \end{pmatrix}, \end{aligned} \tag{1.111}$$

niin odotusarvo tästä satunnaismatriisista on tietenkin $\mathbf{z}$:n kovarianssimatriisi:

$$\mathrm{E}(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})' = \begin{pmatrix} \mathrm{var}(z_1) & \mathrm{cov}(z_1,z_2) \\ \mathrm{cov}(z_2,z_1) & \mathrm{var}(z_2) \end{pmatrix} = \mathrm{cov}(\mathbf{z}) = \boldsymbol{\Sigma}\,. \tag{1.112}$$

Aivan vastaavasti voimme esittää p-ulotteisen satunnaisvektorin $\mathbf{z}$ kovarianssimatriisin lausekkeena

$$\mathrm{cov}(\mathbf{z}) = \mathrm{E}(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})'. \tag{1.113}$$

Esimerkiksi, jos satunnaisvektorin $\mathbf{z}$ mahdolliset arvot ovat $\mathbf{z}_{(1)}, \ldots, \mathbf{z}_{(n)}$ ja jokainen näistä arvoista on yhtä todennäköinen, niin

$$\mathrm{cov}(\mathbf{z}) = \tfrac{1}{n}\sum_{i=1}^{n}(\mathbf{z}_{(i)}-\boldsymbol{\mu})(\mathbf{z}_{(i)}-\boldsymbol{\mu})' = \boldsymbol{\Sigma}\,, \quad \boldsymbol{\mu} = \tfrac{1}{n}\sum_{i=1}^{n}\mathbf{z}_{(i)}\,. \tag{1.114}$$

Jos pidämmekin vektoreita $\mathbf{z}_{(1)}, \ldots, \mathbf{z}_{(n)}$ havaintoaineistona, niin otoskovarianssimatriisiksi $\mathbf{S}$ saadaan kommentin 1.5 (s. 37) mukaisesti $\mathbf{S} = \frac{n}{n-1}\,\boldsymbol{\Sigma}$.

Jos $\mathbf{x}$ on 2-ulotteinen ja $\mathbf{y}$ on 3-ulotteinen satunnaisvektori, joiden odotusarvot ovat $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$ ja $\mathrm{E}(\mathbf{y}) = \boldsymbol{\nu}$, niin

$$\begin{aligned}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\nu})' &= \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} (y_1-\nu_1, y_2-\nu_2, y_3-\nu_3) \\ &= \begin{pmatrix} (x_1-\mu_1)(y_1-\nu_1) & (x_1-\mu_1)(y_2-\nu_2) & (x_1-\mu_1)(y_3-\nu_3) \\ (x_2-\mu_2)(y_1-\nu_1) & (x_2-\mu_2)(y_2-\nu_2) & (x_2-\mu_2)(y_3-\nu_3) \end{pmatrix}. \end{aligned} \tag{1.115}$$

Täten $\mathbf{x}$:n ja $\mathbf{y}$:n keskinäinen kovarianssimatriisi on matriisimerkinnöin

$$\begin{aligned}\operatorname{cov}(\mathbf{x},\mathbf{y}) &= \mathrm{E}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\nu})' \\ &= \begin{pmatrix} \operatorname{cov}(x_1,y_1) & \operatorname{cov}(x_1,y_2) & \operatorname{cov}(x_1,y_3) \\ \operatorname{cov}(x_2,y_1) & \operatorname{cov}(x_2,y_2) & \operatorname{cov}(x_2,y_3) \end{pmatrix}. \end{aligned} \tag{1.116}$$

Tällöin tietenkin $\operatorname{cov}(\mathbf{x},\mathbf{x}) = \operatorname{cov}(\mathbf{x})$.

On paikallaan korostaa merkintöjen

$$\operatorname{cov}(\mathbf{x},\mathbf{y}) = \boldsymbol{\Sigma}_{\mathbf{xy}} \in \mathbb{R}^{p\times q} \quad \text{ja} \quad \operatorname{cov}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} := \boldsymbol{\Sigma} \in \operatorname{Sym}(p+q) \tag{1.117}$$

välistä eroa. Nyt siis $\operatorname{cov}(\mathbf{x},\mathbf{y}) = \boldsymbol{\Sigma}_{\mathbf{xy}}$ tarkoittaa satunnaisvektorien $\mathbf{x}$ (p elementtiä) ja $\mathbf{y}$ (q elementtiä) keskinäistä kovarianssimatriisia (cross-covariance matrix), mikä on $p\times q$-matriisi eikä siis ole välttämättä neliömatriisi. Sen sijaan $\boldsymbol{\Sigma}$ on $(p+q)$-ulotteisen satunnaisvektorin $\binom{\mathbf{x}}{\mathbf{y}}$ kovarianssimatriisi (symmetrinen neliömatriisi), joka voidaan esittää ositetussa muodossa

$$\operatorname{cov}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \begin{pmatrix} \operatorname{cov}(\mathbf{x}) & \operatorname{cov}(\mathbf{x},\mathbf{y}) \\ \operatorname{cov}(\mathbf{y},\mathbf{x}) & \operatorname{cov}(\mathbf{y}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}} \\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \boldsymbol{\Sigma}_{\mathbf{yy}} \end{pmatrix} = \boldsymbol{\Sigma}\,. \tag{1.118}$$

Erityisesti jos $\mathbf{z} = \binom{\mathbf{x}}{y}$ on $(p+1)$:n elementin satunnaisvektori, niin

$$\operatorname{cov}(\mathbf{z}) = \operatorname{cov}\begin{pmatrix}\mathbf{x}\\ y\end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}'_{\mathbf{x}y} & \sigma_y^2 \end{pmatrix}, \tag{1.119a}$$

missä

$$\boldsymbol{\sigma}_{\mathbf{x}y} = \operatorname{cov}(\mathbf{x},y) = \begin{pmatrix} \operatorname{cov}(x_1,y) \\ \operatorname{cov}(x_2,y) \\ \vdots \\ \operatorname{cov}(x_p,y) \end{pmatrix} = \begin{pmatrix} \sigma_{1y} \\ \sigma_{2y} \\ \vdots \\ \sigma_{py} \end{pmatrix}. \tag{1.119b}$$

1.3.1 Esimerkkejä

Esimerkki 1.5. *Yksinkertainen satunnaisotos palauttaen/palauttamatta.* Tarkastellaan ns. *tasakorrelaatiomatriisia* (*intraclass*-correlation)

$$\operatorname{cor}(\mathbf{y}) = \operatorname{cor}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & \varrho & \dots & \varrho \\ \varrho & 1 & \dots & \varrho \\ \vdots & \vdots & \ddots & \vdots \\ \varrho & \varrho & \dots & 1 \end{pmatrix} := \boldsymbol{\Omega}(\varrho), \tag{1.120}$$

missä siis jokaisen (y_i, y_j)-parin korrelaatio on sama. Tälläinen korrelaatiomatriisi esiintyy esimerkiksi seuraavassa tilanteessa.

Olkoon $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ yksinkertainen satunnaisotos (YSO) kokonaisluvuista $1, 2, \dots, N$, kun $n \le N$. Mikä on satunnaisvektorin $\mathbf{y}$ korrelaatiomatriisi, kun otos valitaan

(a) palauttaen,

(b) palauttamatta?

On selvää, että (a)-tapauksessa satunnaismuuttujat y_i ovat korreloimattomia ja siten $\mathbf{y}$:n korrelaatiomatriisi on yksikkömatriisi $\mathbf{I}_n$:

$$\operatorname{cor}(\mathbf{y}) = \mathbf{I}_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}, \tag{1.121}$$

eli $\operatorname{cor}(\mathbf{y}) = \mathbf{\Omega}(0)$. Koska jokainen y_i noudattaa diskreettiä tasajakaumaa $\operatorname{Tasd}(1, N)$, on

$$\mathrm{E}(y_i) = \frac{1}{N}(1 + 2 + \cdots + N) = \frac{N+1}{2} := \mu_y\,. \tag{1.122}$$

Lisäksi on osoitettavissa, että

$$\operatorname{var}(y_i) = \frac{1}{N}\sum_{i=1}^{N}(i-\mu_y)^2 = \frac{1}{N}\sum_{i=1}^{N} i^2 - \mu_y^2 = \frac{N^2-1}{12} := \sigma_y^2\,, \tag{1.123}$$

sillä

$$\sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}\,. \tag{1.124}$$

Täten $\operatorname{cov}(\mathbf{y}) = \frac{N^2-1}{12}\mathbf{I}_n = \sigma_y^2\mathbf{I}_n$.

Merkitään (b)-tilanteessa saatua otosta $\mathbf{z} = (z_1, z_2, \dots, z_n)'$ eli kyseessä on siis yksinkertainen satunnaisotos kokonaisluvuista $1, 2, \dots, N$, kun otos valitaan palauttamatta. Nyt $\operatorname{var}(z_i) = \operatorname{var}(y_i) = \sigma_y^2$, mutta satunnaismuuttujat z_i ovat korreloituneita. Intuitiivisesti on selvää, että jokaisen (z_i, z_j)-parin korrelaatio on kuitenkin sama. Osoitamme (ks. myös harjoitustehtävä 1.20, s. 66), että

$$\operatorname{cov}(z_i, z_j) = -\frac{1}{N-1}\frac{N^2-1}{12} = -\frac{1}{N-1}\sigma_y^2 = -\frac{N+1}{12}\,, \tag{1.125a}$$

$$\operatorname{cor}(z_i, z_j) = \frac{-\frac{1}{N-1}\sigma_y^2}{\sigma_y\cdot\sigma_y} = -\frac{1}{N-1}\,, \quad i \neq j\,. \tag{1.125b}$$

Koska y_i ja y_j $(i \neq j)$ ovat korreloimattomia, on voimassa

$$\operatorname{cov}(y_i, y_j) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{\ell=1}^{N}(k-\mu_y)(\ell-\mu_y) := \frac{1}{N^2}\,\mathrm{SP}_{y_iy_j} = 0\,. \tag{1.126}$$

Koska

$$\operatorname{cov}(z_i, z_j) = \frac{1}{N(N-1)} \sum_{k=1}^{N} \sum_{\substack{\ell=1 \\ k\neq\ell}}^{N} (k-\mu_y)(\ell-\mu_y) := \frac{1}{N(N-1)} \mathrm{SP}_{z_i z_j} \,, \quad (1.127)$$

saamme

$$\mathrm{SP}_{z_i z_j} = \mathrm{SP}_{y_i y_j} - \sum_{k=1}^{N} (k-\mu_y)^2 = 0 - N\sigma_y^2 \,, \quad (1.128a)$$

$$\operatorname{cov}(z_i, z_j) = -\frac{1}{N(N-1)} N \frac{N^2-1}{12} = -\frac{N+1}{12} \,. \quad (1.128b)$$

Aivan vastaavalla tavalla on osoitettavissa, että jos luvuista $\{a_1, a_2, \dots, a_N\}$ valitaan yksinkertainen satunnaisotos palauttamatta, niin (1.125) on voimassa, kun $\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(a_i - \bar{a})^2$. □

Esimerkki 1.6. *Ehdollinen odotusarvo satunnaismuuttujana.* Tarkastellaan satunnaismuuttujien x ja y yhteisjakaumaa ja olkoon x yksinkertaisuuden vuoksi diskreetti satunnaismuuttuja, jonka arvot ovat $x_1, \dots, x_k$ ja vastaavat todennäköisyydet $p_1, \dots, p_k$. Merkitään $m(\underline{x}) = \mathrm{E}(y \mid x = \underline{x})$ ja $v(\underline{x}) = \operatorname{var}(y \mid x = \underline{x})$. Toisin sanoen $m(\underline{x})$:n arvo kertoo satunnaismuuttujan y ehdollisen odotusarvon kun x saa arvon $\underline{x}$ ja $v(\underline{x})$:n arvo on vastaavasti y ehdollinen varianssi kun x saa arvon $\underline{x}$. Merkinnällä $m(x) = \mathrm{E}(y \mid x)$ tarkoitamme tällöin satunnaismuuttujaa, jonka arvot ovat y:n ehdolliset odotusarvot

$$m(\underline{x}_1) = \mathrm{E}(y \mid x = \underline{x}_1), \dots, m(\underline{x}_k) = \mathrm{E}(y \mid x = \underline{x}_k),$$

ja näiden arvojen todennäköisyydet ovat $p_1, \dots, p_k$. Satunnaismuuttuja $\operatorname{var}(y \mid x)$ on määritelty vastaavalla tavalla.

Oletetaan, että kuvio 1.10 esittää satunnaismuuttujien x ja y yhteisjakaumaa, jossa jokainen 7:stä arvoparista $\binom{1}{y_{11}}, \binom{1}{y_{12}}, \dots, \binom{3}{y_{31}}, \binom{3}{y_{32}}$ on yhtä todennäköinen; ks. myös esimerkki 1.1 (s. 16). Osoita, että

$$\mathrm{E}(y) = \mathrm{E}[\mathrm{E}(y \mid x)] = \mathrm{E}[m(x)] \,, \quad (1.129a)$$

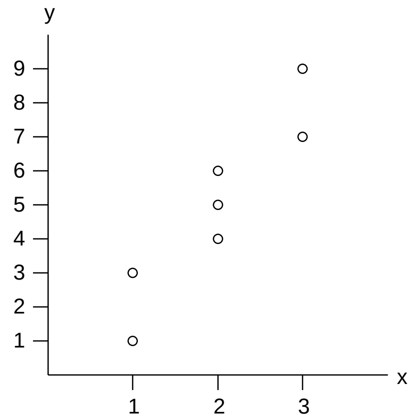
$$\operatorname{var}(y) = \operatorname{var}[\mathrm{E}(y \mid x)] + \mathrm{E}[\operatorname{var}(y \mid x)] = \operatorname{var}[m(x)] + \mathrm{E}[v(x)] \,. \quad (1.129b)$$

Varmista (1.129b):n ja varianssianalyysin neliösummahajotelman

$$\sum_{i=1}^{3}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{3} n_i(\bar{y}_i - \bar{y})^2 + \sum_{i=1}^{3}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2, \quad (1.130a)$$

$$\mathrm{SSTotal} = \mathrm{SSBetween} + \mathrm{SSError} \,, \quad (1.130b)$$

$$\mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y} = \mathbf{y}'(\mathbf{H} - \mathbf{J})\mathbf{y} + \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} \,. \quad (1.130c)$$



Kuvio 1.10. Esimerkkiin 1.6 liittyvä todennäköisyysjakauma: kaikki 7 arvoparia yhtä todennäköisiä.

välinen yhteys. Kaavassa (1.130a) n_i viittaa i:nnen ryhmän havaintojen lukumäärään ja $\bar{y}_i$ tarkoittaa y:n ehdollista odotusarvoa, kun x:llä on arvo i eli i:nnen ryhmän keskiarvoa. Kaava (1.130c) on lukijalle tässä vaiheessa luultavasti outo. Ilman sen kummempia selityksiä voimme mainita, että ko. matriisit ovat seuraavat:

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{J} = \mathbf{P}_{\mathbf{1}_n} = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' = \mathbf{J}_n\,, \tag{1.131a}$$

$$\mathbf{H} = \mathbf{P}_{\mathbf{X}} = \begin{pmatrix} \mathbf{J}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_2 \end{pmatrix} = \begin{pmatrix} \frac12 & \frac12 & 0 & 0 & 0 & 0 & 0 \\ \frac12 & \frac12 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac13 & \frac13 & \frac13 & 0 & 0 \\ 0 & 0 & \frac13 & \frac13 & \frac13 & 0 & 0 \\ 0 & 0 & \frac13 & \frac13 & \frac13 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac12 & \frac12 \\ 0 & 0 & 0 & 0 & 0 & \frac12 & \frac12 \end{pmatrix}. \tag{1.131b}$$

Mainittakoon, että Casella (2008, s. 9) kutsuu yhtälöä

$$\operatorname{var}(y) = \operatorname{var}[\mathrm{E}(y \mid x)] + \mathrm{E}[\operatorname{var}(y \mid x)] \tag{1.132}$$

”a most important equality” kirjassaan *Statistical Design*. O’Hagan (2012) puolestaan käyttää siitä nimitystä ”my thing of beauty”. □

Esimerkki 1.7. Arthur Argunot pysäytettiin ylinopeudesta tiellä, jossa oli 90 km/h nopeusrajoitus. Käräjillä Arthur väitti ajaneensa vain 85 km/h. Tuomari Eidelburger, joka käsitteli tapausta, oli kutsunut todistajaksi Ms. Polly Nomialin, joka kertoi ajaneensa nopeudella 80 km/h, kun Arthur häpeämättömästi oli ohittanut hänet. Lisäksi hän väitti Arthurin ajaneen 10 km/h lujempaa kuin hän itse. Ylinopeuden havainnut poliisi Carlo Cannelloni väitti Arthurin nopeuden olleen 100 km/h ja Pollyn 60 km/h.

Muodostamme seuraavan muuttujavektorin $\mathbf{y}$:

$$\mathbf{y} = \begin{pmatrix} 85 \\ 80 \\ 10 \\ 100 \\ 60 \end{pmatrix} \begin{matrix} \text{Arthur} \\ \text{Polly1} \\ \text{Polly2} \\ \text{Carlo1} \\ \text{Carlo2} \end{matrix} \tag{1.133}$$

Tällöin muuttuja y siis ilmaisee annetut viisi lausuntoa nopeuksista. Kolmas y-arvo poikkeaa muista, sillä se ei ole arvio Arthurin tai Pollyn yksittäiselle nopeudelle, vaan niiden erotukselle.

Voimme ajatella, että vektori $\mathbf{y}$ on yksi reaalisaatio tietystä satunnaisvektorista. (Lisää realisaatioita saataisiin, kun kysyttäisiin uudestaan samat kysymykset samassa tilanteessa.) Oletetaan, että Pollyn antamien arvioiden välinen korrelaatio on ϱ samoin kuin Carlon antamien arvioiden. Oletetaan lisäksi, että eri henkilöiden antamien arvioiden välillä ei ole korrelaatiota. Tällöin saamme satunnaisvektorin $\mathbf{y}$ korrelaatiomatriisiksi

$$\operatorname{cor}(\mathbf{y}) = \operatorname{cor}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \varrho & 0 & 0 \\ 0 & \varrho & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & \varrho \\ 0 & 0 & 0 & \varrho & 1 \end{pmatrix} := \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_{2\times 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Omega}_{2\times 2} \end{pmatrix} := \mathbf{\Sigma}\,. \tag{1.134}$$

On tähdennettävä, että $\mathbf{\Sigma}$ on teoreettinen kovarianssimatriisi teoreettiselle satunnaisvektorille; vektori $\mathbf{y}$ (1.133):ssa puolestaan on tämän satunnaisvektorin yksi realisaatio. □

Esimerkki 1.8. 1. *kertaluvun autoregressiivinen prosessi:* AR(1). Olkoon

$$\begin{aligned} \operatorname{cor}(\mathbf{y}) = \operatorname{cor}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} &= \begin{pmatrix} 1 & \varrho & \varrho^2 & \dots & \varrho^{n-1} \\ \varrho & 1 & \varrho & \dots & \varrho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \varrho^{n-1} & \varrho^{n-2} & \varrho^{n-3} & \dots & 1 \end{pmatrix} \\ &= \{\varrho^{|i-j|}\} := \mathbf{\Psi}(\varrho). \end{aligned} \tag{1.135}$$

Tällöin siis peräkkäisten y_i-arvojen korrelaatio on

$$\operatorname{cor}(y_i, y_{i+1}) = \varrho, \;\operatorname{cor}(y_i, y_{i+2}) = \varrho^2, \;\text{jne.} \tag{1.136}$$

Havaintojen välinen korrelaatio heikkenee erotuksen $|i-j|$ kasvaessa, koska luonnollisesti oletamme että $|\varrho| < 1$. Tälläinen tilanne voi syntyä (aikasarjojen yhteydessä) esim. seuraavasti.

Määritellään y_i siten että

$$y_i = \varrho y_{i-1} + u_i\,, \quad i = 1, \dots, n\,, \quad |\varrho| < 1\,, \tag{1.137}$$

missä u_i:t ($i = \dots, -2, -1, 0, 1, 2, \dots$) ovat riippumattomia satunnaismuuttujia, joilla on kaikilla sama odotusarvo 0 ja varianssi σ_u^2. Tällöin voidaan näyttää, että

$$\operatorname{var}(y_i) = \frac{\sigma_u^2}{1-\varrho^2}\,, \quad i = 1, \dots, n\,, \quad \operatorname{cor}(y_i, y_j) = \varrho^{|i-j|}, \quad i \neq j\,. \tag{1.138}$$

□

Esimerkki 1.9. *Multinomijakauma.* Multinomijakauma on binomijakauman yleistys. Binomijakauman taustallahan voidaan ajatella olevan sellainen tilanne, jossa populaatio muodostuu kahdesta ryhmästä E_1 ja E_2 (arvosta, luokasta tms). Kun populaatiosta valitaan satunnaisesti yksi havainto, niin se kuuluu ryhmään E_1 todennäköisyydellä p ja ryhmään E_2 todennäköisyydellä $q = 1-p$. Voimme toistaa tätä menettelyä n kertaa (satunnaisotos palauttaen) ja määritellä

$$z_i = \begin{cases} 1, & \text{jos } i. \text{ havainto kuuluu luokkaan } E_1, \\ 0 & \text{muuten.} \end{cases}$$

Satunnaismuuttuja z_i noudattaa *Bernoulli*-jakaumaa:

$$z_i \sim \operatorname{Ber}(p)\,,\ \operatorname{E}(z_i) = p\,,\ \operatorname{var}(z_i) = p(1-p),\ i = 1, 2, \dots, n\,, \tag{1.139}$$

ja satunnaismuuttujien z_i summa z noudattaa binomijakaumaa:

$$z = z_1 + z_2 + \cdots + z_n \sim \operatorname{Bin}(n, p)\,,\ \operatorname{E}(z) = np\,,\ \operatorname{var}(z) = np(1-p)\,. \tag{1.140}$$

Multinomijakaumassa on k (toisensa poissulkevaa) ryhmää, joihin kuulumistodennäköisyydet ovat $p_1, p_2, \dots, p_k$, ja joiden summa on 1:

$$p_1 + p_2 + \cdots + p_k = 1\,. \tag{1.141}$$

Poimitaan tästä populaatiosta satunnaisesti yksi havainto ja määritellään k elementin satunnaisvektori $\mathbf{x}$ siten että

$$x_1 = \begin{cases} 1, & \text{jos havainto kuuluu luokkaan } E_1;\ \operatorname{P}(x_1 = 1) = p_1, \\ 0 & \text{muuten,} \end{cases}$$

$$\vdots$$

$$x_k = \begin{cases} 1, & \text{jos havainto kuuluu luokkaan } E_k;\ \operatorname{P}(x_k = 1) = p_k, \\ 0 & \text{muuten.} \end{cases}$$

```
 -  - SURVO MM  Thu Sep 23 14:51:58 2010          C:\SJP\D\   2000  100 0
*
*  KOKEILUJA INTRACLASS-KORRELAATIOMATRIISILLA
*
*MAT ONE=CON(n,1)    /  ONE = tolppa-1     n=4  r=0.7
*MAT R!=(1-r)*IDN(n,n)+r*ONE*ONE'
*MAT LOAD R,##.###,CUR+1
*MATRIX R
*///          1      2      3      4
*  1      1.000  0.700  0.700  0.700
*  2      0.700  1.000  0.700  0.700
*  3      0.700  0.700  1.000  0.700
*  4      0.700  0.700  0.700  1.000
*
*MATRIX R1  ///      /  Toinen tapa määritellä R
*1  r   r   r
*r  1   r   r
*r  r   1   r
*r  r   r   1
*
*MAT SAVE R1   /  r=0.9
*MAT INVR!=INV(R,det)          / *INVR~INV(R) det=0.0837 4*4
*MAT LOAD INVR,##.###,CUR+1    /  INV(A,det) laskee myös determinantin
*MATRIX INVR
*///          1      2      3      4
*  1      2.581 -0.753 -0.753 -0.753
*  2     -0.753  2.581 -0.753 -0.753
*  3     -0.753 -0.753  2.581 -0.753
*  4     -0.753 -0.753 -0.753  2.581
*
*MAT B!=INV(R(1:3,1:3))*R(1:3,4)  / B:ssä ovat regressiokertoimet
*MAT LOAD B                       / kun 4. muuttujaa selitetään muilla
*MATRIX B
*///            4
*  1     0.291667
*  2     0.291667
*  3     0.291667
*
*MAT YHTKOR2!=B'*R(1:3,4)  / selitysaste
*MAT_YHTKOR2(1,1)=0.63     / kun 4. muuttujaa selitetään muilla
*                                          (MK10-008, MKT-01-03)
```

Kuvio 1.11. Kokeiluja tasakorrelaatiomatriisilla. Toimituskentässä on laskettu myös standardoidut regressiokertoimet kun 4. muuttujaa selitetään muilla.

```
 -   - SURVO MM  Thu Sep 23 15:32:11 2010            C:\SJP\D\   2000  100 0
57 *
58 *            KORRELAATIOMATRIISIN ELEMENTIT r:n potensseja
59 *  n=5   r=0.9
60 *MAT R=CON(n,n)                     /  kaikki elementit ensin ykkösiksi
61 *MAT TRANSFORM R BY r^abs(I#-J#)    /   (1,2)-elementti = r
62 *MAT NAME R AS R                    /   (1,3)-elementti = r^2  jne
63 *MAT LOAD R,#.###,CUR+1             /   r^4=0.6561
64 *MATRIX R
65 *///          1     2     3     4     5
66 *  1      1.000 0.900 0.810 0.729 0.656
67 *  2      0.900 1.000 0.900 0.810 0.729
68 *  3      0.810 0.900 1.000 0.900 0.810
69 *  4      0.729 0.810 0.900 1.000 0.900
70 *  5      0.656 0.729 0.810 0.900 1.000
71 *
72 *MAT INVR!=INV(R,det)          / *INVR~INV(R) det=0.00130321 5*5
73 *MAT LOAD INVR,##.###,CUR+1    / huomaa determinantin pienuus!
74 *MATRIX INVR
75 *///           1      2      3      4      5
76 *  1       5.263 -4.737  0.000 -0.000 -0.000
77 *  2      -4.737  9.526 -4.737  0.000  0.000
78 *  3       0.000 -4.737  9.526 -4.737 -0.000
79 *  4      -0.000  0.000 -4.737  9.526 -4.737
80 *  5      -0.000 -0.000 -0.000 -4.737  5.263
81 *
82 *MAT B!=INV(R(1:4,1:4))*R(1:4,5)   / B:ssä ovat regressiokertoimet
83 *MAT LOAD B                        / kun 5. muuttujaa selitetään muilla
84 *MATRIX B
85 *///             5
86 *  1       0.00000
87 *  2      -0.00000
88 *  3       0.00000
89 *  4       0.90000
90 *
91 *MAT YHTKOR2!=B'*R(1:4,5)  / selitysaste
92 *MAT_YHTKOR2(1,1)=0.81     / kun 5. muuttujaa selitetään muilla
93 *                                          (MK10-008, MKT-01-04)
```

Kuvio 1.12. Korrelaatiot r:n potensseja.

Vektorin $\mathbf{x}$ elementeistä siis yksi on 1 ja muut nollia. Tällöin on näytettävissä (HT), että

$$\mathrm{E}(x_i) = p_i\,, \quad \mathrm{var}(x_i) = p_i(1-p_i)\,, \quad i = 1, 2, \dots, k\,, \tag{1.142a}$$

$$\mathrm{cov}(x_i, x_j) = -p_i p_j\,, \quad i \neq j = 1, 2, \dots, k\,. \tag{1.142b}$$

Satunnaisvektorin $\mathbf{x}$ kovarianssimatriisi on siten

$$\mathrm{cov}(\mathbf{x}) = \mathbf{\Sigma} = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_k \\ -p_2p_1 & p_2(1-p_2) & \dots & -p_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -p_kp_1 & -p_kp_2 & \dots & p_k(1-p_k) \end{pmatrix}. \tag{1.143}$$

Merkitään

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_k \end{pmatrix}. \tag{1.144}$$

Matriisikertolasku $\mathbf{pp}'$ tarkoittaa seuraavaa:

$$\mathbf{pp}' = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix} (p_1, p_2, \dots, p_k) = \begin{pmatrix} p_1^2 & p_1p_2 & \dots & p_1p_k \\ p_2p_1 & p_2^2 & \dots & p_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ p_kp_1 & p_kp_2 & \dots & p_k^2 \end{pmatrix}. \tag{1.145}$$

Kovarianssimatriisi $\mathbf{\Sigma}$ voidaan tällöin esittää muodossa

$$\mathrm{cov}(\mathbf{x}) = \mathbf{\Sigma} = \mathbf{D} - \mathbf{pp}'. \tag{1.146}$$

Jos havaintoyksiköitä valitaan n kappaletta – riippumattomasti eli satunnaisotos palauttaen – ja merkitään saatuja k elementin vektoreita $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$, niin näiden vektoreiden summavektorin

$$\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n \tag{1.147}$$

elementit kertovat kuhunkin luokkaan kuuluvien havaintojen lukumäärän. Vektorin $\mathbf{y}$ elementtien summa on tietenkin n:

$$\mathbf{1}_k'\mathbf{y} = \mathbf{1}_k'\mathbf{x}_1 + \mathbf{1}_k'\mathbf{x}_2 + \dots + \mathbf{1}_k'\mathbf{x}_n = n\,. \tag{1.148}$$

On helppo osoittaa satunnaisvektorien $\mathbf{x}_i$ riippumattomuuden perusteella, että

$$\mathrm{E}(\mathbf{y}) = n\mathbf{p}, \quad \mathrm{cov}(\mathbf{y}) = n\mathbf{\Sigma}\,. \tag{1.149}$$

Vektorin $\mathbf{y}$ arvojen todennäköisyydet ovat juuri multinomijakauman todennäköisyyksiä ja sanomme että $\mathbf{y}$ noudattaa multinomijakaumaa parametrein n ja $\mathbf{p}$ eli $\mathbf{y} \sim \mathrm{Mult}(n, \mathbf{p})$.

Harjoitustehtävänä on osoittaa seuraavat $\mathbf{\Sigma}$:n ominaisuudet (tässä tarvitaan joitakin myöhemmin esitettäviä käsitteitä):

(a) $\mathbf{\Sigma}$ on kaksoiskeskistetty (sekä rivi- että sarakesummat nollia), singulaarinen, $\operatorname{rank}(\mathbf{\Sigma}) = 2$.

(b) $\mathbf{\Sigma}\mathbf{D}^{-1}\mathbf{\Sigma} = \mathbf{\Sigma}$, ts. $\mathbf{D}^{-1}$ on $\mathbf{\Sigma}$:n yleistetty käänteismatriisi. Vakuuttaudu, että $\mathbf{D}^{-1}$ ei välttämättä toteuta kaikkia Moore–Penrose -ehtoja. □

1.3.2 Merkintöjä

Edellä merkitsimme (1.6c):ssa (s. 10) muuttujan y otosvarianssia kahdella tavalla: $\operatorname{var}(\mathbf{y})$ ja $\operatorname{var}(y)$. Oleellisinta on huomata, että lihavoitettu pieni kirjain, esim. $\mathbf{y}$, erityisesti laskutoimituksissa viittaa pystyvektoriin (sarakkeeseen), jonka tulkinnalle on kaksi mahdollisuutta:

- $\mathbf{y}$ sisältää kyseisen muuttujan n havaintoarvoa

tai

- $\mathbf{y}$ on n-ulotteinen satunnaisvektori.

Yleensä emme merkinnällisesti tee eroa satunnaisvektorin (tai satunnaismuuttujan) ja sen havaitun arvon välillä, vaikka näiden käsitteiden välillä on tietenkin sinänsä suuri ero. Asiayhteydestä käy aina ilmi kumpi tapaus on kyseessä.

Kerrataanpa vielä merkintöjen eroja:

- $\operatorname{cor}_{\mathrm{d}}(\mathbf{U}) = \operatorname{cor}_{\mathrm{d}}(\mathbf{u}_1 : \ldots : \mathbf{u}_p)$, havaintomatriisista $\mathbf{U}_{n\times p}$ laskettu otoskovarianssimatriisi ($p \times p$; p muuttujaa ja n havaintoa)
- $\operatorname{cov}(\mathbf{z})$, p-ulotteisen satunnaisvektorin $\mathbf{z}$ kovarianssimatriisi ($p \times p$)
- $\operatorname{cor}_{\mathrm{d}}(\mathbf{U}) = \operatorname{cor}_{\mathrm{d}}(\mathbf{u}_1 : \ldots : \mathbf{u}_p)$, havaintomatriisista $\mathbf{U}_{n\times p}$ laskettu otoskorrelaatiomatriisi ($p \times p$; p muuttujaa ja n havaintoa)
- $\operatorname{cor}(\mathbf{z})$, p-ulotteisen satunnaisvektorin $\mathbf{z}$ korrelaatiomatriisi ($p \times p$)
- $\operatorname{cor}_{\mathrm{d}}(\mathbf{t}, \mathbf{v})$, muuttujien $\mathbf{t}$ ja $\mathbf{v}$ otoskorrelaatiokerroin
- $\operatorname{cor}_{\mathrm{d}}(\mathbf{t} : \mathbf{v})$, muuttujien $\mathbf{t}$ ja $\mathbf{v}$ otoskorrelaatiomatriisi (2×2)
- $\operatorname{cor}(\mathbf{z}, \mathbf{w})$, satunnaisvektorien $\mathbf{z}$ ($p \times 1$) ja $\mathbf{w}$ ($q \times 1$) välinen korrelaatiomatriisi ($p \times q$).

1.4 Satunnaisotos matriisimerkinnöin

Tilastotieteen peruskurssilta lukija muistaa usein käytetyn sanonnan:

> ”Poimitaan n havainnon satunnaisotos $u_1, u_2, \ldots, u_n$ populaatiosta, joka noudattaa normaalijakaumaa parametrein μ ja σ^2.” (1.150)

Tämähän tarkoittaa matemaattisesti silloin sitä, että meillä on itse asiassa

n riippumatonta satunnaismuuttujaa u_i, joilla jokaisella on sama jakauma $\mathrm{N}(\mu, \sigma^2)$. (1.151)

Kyseinen satunnaisotos voidaan esittää pystyvektorina

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad \mathbf{u}' = (u_1, u_2, \dots, u_n)\,, \quad u_i \sim \mathrm{N}(\mu, \sigma^2)\,. \tag{1.152}$$

Vektori $\mathbf{u}$ on nyt siis satunnaisvektori, jonka havaittu arvo on $\mathbb{R}^n$:n vektori. Koska satunnaisvektorin $\mathbf{u}$ komponentit ovat riippumattomia, noudattaa $\mathbf{u}$ n-ulotteista normaalijakaumaa odotusarvovektorina $\mu\mathbf{1}_n$ ja kovarianssimatriisina $\sigma^2\mathbf{I}_n$ eli

$$\mathbf{u} \sim \mathrm{N}_n(\mu\mathbf{1}_n, \sigma^2\mathbf{I}_n)\,, \tag{1.153}$$

missä $\mathbf{I}_n$ on yksikkömatriisi ja $\mathbf{1}_n$ on ns. ”tolppaykkönen”:

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}, \quad \sigma^2\mathbf{I}_n = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}, \tag{1.154a}$$

$$\mathbf{1}_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n, \quad \mu\mathbf{1}_n = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix}. \tag{1.154b}$$

Tilanteessa (1.150) meillä on otos yksiulotteisesta normaalijakaumasta. Erityisesti monimuuttujamenetelmissä tarkastellaan satunnaisotoksia useampiulotteisesta jakaumasta:

”Olkoon $\mathbf{u}_{(1)}, \mathbf{u}_{(2)}, \dots, \mathbf{u}_{(n)}$ satunnaisotos populaatiosta, joka noudattaa multinormaalijakaumaa $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.” (1.155)

Tarkastellaan ensin yksinkertaisuuden vuoksi tilannetta, missä $\mathbf{u}_{(1)}, \mathbf{u}_{(2)}, \dots, \mathbf{u}_{(n)}$ on satunnaisotos 2-ulotteisesta normaalijakaumasta $\mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Tällöin voimme merkitä

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix}. \tag{1.156}$$

Merkintä (1.156) on aivan identtinen (1.11):n (s. 13) kanssa, mutta (1.156):ssä $\mathbf{U}$ onkin satunnaismatriisi kun taas (1.11):n $\mathbf{U}$ voidaan tulkita tämän satunnaismatriisin havaituksi arvoksi. Havaintovektorit $\mathbf{u}_{(1)}$, $\mathbf{u}_{(2)}$, ..., $\mathbf{u}_{(n)}$ ovat (1.156):ssä riippumattomia 2-ulotteisia satunnaisvektoreita, joilla jokaisella on sama jakauma $\mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, missä

$$\mathrm{E}(\mathbf{u}_{(i)}) = \mathrm{E}\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} = \boldsymbol{\mu}\,, \quad i = 1, 2, \dots, n\,, \tag{1.157a}$$

$$\begin{aligned} \operatorname{cov}(\mathbf{u}_{(i)}) &= \operatorname{cov}\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} \operatorname{var}(x_i) & \operatorname{cov}(x_i, y_i) \\ \operatorname{cov}(y_i, x_i) & \operatorname{var}(y_i) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} = \boldsymbol{\Sigma}\,, \quad i = 1, 2, \dots, n\,. \end{aligned} \tag{1.157b}$$

Sarakkeittain tarkasteltuna

$$\mathrm{E}(\mathbf{x}) = \begin{pmatrix} \mu_x \\ \mu_x \\ \vdots \\ \mu_x \end{pmatrix}, \quad \operatorname{cov}(\mathbf{x}) = \begin{pmatrix} \sigma_x^2 & 0 & \dots & 0 \\ 0 & \sigma_x^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_x^2 \end{pmatrix} = \{\operatorname{cov}(x_i, x_j)\}\,, \tag{1.158a}$$

$$\mathrm{E}(\mathbf{y}) = \begin{pmatrix} \mu_y \\ \mu_y \\ \vdots \\ \mu_y \end{pmatrix}, \quad \operatorname{cov}(\mathbf{y}) = \begin{pmatrix} \sigma_y^2 & 0 & \dots & 0 \\ 0 & \sigma_y^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_y^2 \end{pmatrix} = \{\operatorname{cov}(y_i, y_j)\}\,, \tag{1.158b}$$

$$\operatorname{cov}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} \sigma_{xy} & 0 & \dots & 0 \\ 0 & \sigma_{xy} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{xy} \end{pmatrix} = \{\operatorname{cov}(x_i, y_j)\}\,. \tag{1.158c}$$

Kaavat (1.158) seuraavat suoraan satunnaisvektorien (havaintojen) $\mathbf{u}_{(1)}$, $\mathbf{u}_{(2)}$, ..., $\mathbf{u}_{(n)}$ riippumattomuudesta, mikä siis merkitsee että $\operatorname{cov}(\mathbf{u}_{(i)}, \mathbf{u}_{(j)}) = \mathbf{0}$ $(i \neq j)$ eli

$$\begin{aligned} \operatorname{cov}(\mathbf{u}_{(i)}, \mathbf{u}_{(j)}) = \operatorname{cov}\left[\begin{pmatrix} x_i \\ y_i \end{pmatrix}, \begin{pmatrix} x_j \\ y_j \end{pmatrix}\right] &= \begin{pmatrix} \operatorname{cov}(x_i, x_j) & \operatorname{cov}(x_i, y_j) \\ \operatorname{cov}(y_i, x_j) & \operatorname{cov}(y_i, y_j) \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \; i \neq j\,. \end{aligned} \tag{1.159}$$

Merkintöjen (1.154a)–(1.154b) mukaan voimme kirjoittaa (1.158):n lyhyemmin seuraavasti:

$$\mathrm{E}(\mathbf{x}) = \mu_x \mathbf{1}_n\,, \quad \operatorname{cov}(\mathbf{x}) = \sigma_x^2 \mathbf{I}_n\,, \quad \operatorname{cov}(\mathbf{x}, \mathbf{y}) = \sigma_{xy} \mathbf{I}_n\,, \tag{1.160}$$

ja siten siis

$$\operatorname{cov}[\operatorname{vec}(\mathbf{x}:\mathbf{y})] = \operatorname{cov}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \begin{pmatrix}\operatorname{cov}(\mathbf{x}) & \operatorname{cov}(\mathbf{x},\mathbf{y})\\ \operatorname{cov}(\mathbf{y},\mathbf{x}) & \operatorname{cov}(\mathbf{y})\end{pmatrix} = \begin{pmatrix}\sigma_x^2\mathbf{I}_n & \sigma_{xy}\mathbf{I}_n\\ \sigma_{xy}\mathbf{I}_n & \sigma_y^2\mathbf{I}_n\end{pmatrix} \in \mathbb{R}^{2n\times 2n}, \tag{1.161}$$

missä $\operatorname{vec}(\cdot)$ on ns. *vec-operaatio*, sillä tarkoitetaan matriisin sarakkeiden kirjoittamista alakkain siten että esimerkiksi

$$\operatorname{vec}(\mathbf{x}:\mathbf{y}) = \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}. \tag{1.162}$$

On siis huomattava, että

- havaintovektorit ($\mathbf{U}$:n vaakarivit) ovat keskenään korreloimattomia,

mutta

- muuttujavektorit ($\mathbf{U}$:n sarakkeet) ovat korreloimattomia vain jos $\sigma_{xy} = 0$.

Matriisien $\mathbf{\Sigma}$ ja $\mathbf{I}_n$ *Kroneckerin tulo* on juuri (1.161):ssä esiintyvä lauseke

$$\begin{pmatrix}\sigma_x^2\mathbf{I}_n & \sigma_{xy}\mathbf{I}_n\\ \sigma_{yx}\mathbf{I}_n & \sigma_y^2\mathbf{I}_n\end{pmatrix} = \mathbf{\Sigma}\otimes\mathbf{I}_n \in \mathbb{R}^{2n\times 2n}. \tag{1.163}$$

Yleisesti matriisien $\mathbf{A}_{n\times m}$ ja $\mathbf{B}_{p\times q}$ Kroneckerin tulo on määritelty seuraavasti:

$$\mathbf{A}\otimes\mathbf{B} = \begin{pmatrix}a_{11}\mathbf{B} & \dots & a_{1m}\mathbf{B}\\ \vdots & \vdots & \vdots\\ a_{n1}\mathbf{B} & \dots & a_{nm}\mathbf{B}\end{pmatrix} \in \mathbb{R}^{np\times mq}. \tag{1.164}$$

Oletetaan, että meillä on n riippumatonta satunnaisvektoria $\mathbf{u}_{(i)}$, joilla jokaisella on sama p-ulotteinen jakauma $\mathrm{N}_p(\boldsymbol{\mu},\mathbf{\Sigma})$; edellä $p = 2$. Koko satunnaisotos voidaan tällöin esittää matriisina

$$\begin{pmatrix}\mathbf{u}'_{(1)}\\ \mathbf{u}'_{(2)}\\ \vdots\\ \mathbf{u}'_{(n)}\end{pmatrix} = (\mathbf{u}_1 : \mathbf{u}_2 : \dots : \mathbf{u}_p) = \mathbf{U}_{n\times p}\,. \tag{1.165}$$

Näin johdettu matriisi $\mathbf{U}$ on siis itse asiassa havaintomatriisin teoreettinen vastine; kun otos poimitaan, satunnaismatriisille realisoituu jokin arvo.

Sanonnan (1.155) voimme esittää täten yhtäpitävästi:

> ”Olkoon $\mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \dots : \mathbf{u}_{(n)})$ satunnaisotos populaatiosta, joka noudattaa multinormaalijakaumaa $\mathrm{N}_p(\boldsymbol{\mu},\mathbf{\Sigma})$.” (1.166)

Toistettakoon vielä, että (1.166) tarkoittaa että

- $\mathbf{u}_{(1)}, \mathbf{u}_{(2)}, \dots, \mathbf{u}_{(n)}$ ovat riippumattomia p-ulotteisia satunnaisvektoreita, joilla jokaisella on sama jakauma, tässä tapauksessa $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,
- $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$ ovat n-ulotteisia satunnaisvektoreita ja

$$\mathbf{u}_i \sim \mathrm{N}_n(\mu_i \mathbf{1}_n, \sigma_i^2 \mathbf{I}_n)\,, \quad \mathrm{cov}(\mathbf{u}_i, \mathbf{u}_j) = \sigma_{ij} \mathbf{I}_n\,. \tag{1.167}$$

Havaitsemme edelleen, että

$$\mathbf{u}_* = \mathrm{vec}(\mathbf{U}) = \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_p \end{pmatrix}, \quad \mathrm{E}(\mathbf{u}_*) = \begin{pmatrix} \mu_1 \mathbf{1}_n \\ \vdots \\ \mu_p \mathbf{1}_n \end{pmatrix} = \boldsymbol{\mu} \otimes \mathbf{1}_n \in \mathbb{R}^{pn}, \tag{1.168}$$

ja

$$\mathrm{cov}(\mathbf{u}_*) = \begin{pmatrix} \sigma_1^2 \mathbf{I}_n & \sigma_{12} \mathbf{I}_n & \dots & \sigma_{1p} \mathbf{I}_n \\ \sigma_{21} \mathbf{I}_n & \sigma_2^2 \mathbf{I}_n & \dots & \sigma_{2p} \mathbf{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} \mathbf{I}_n & \sigma_{p2} \mathbf{I}_n & \dots & \sigma_p^2 \mathbf{I}_n \end{pmatrix} = \boldsymbol{\Sigma} \otimes \mathbf{I}_n \in \mathrm{NND}_{pn}\,. \tag{1.169}$$

1.4.1 Multinormaalijakaumasta

Pari sanaa multinormaalijakaumasta. – Jos p-ulotteinen satunnaisvektori $\mathbf{z}$ (saaden arvoja $\mathbb{R}^p$:ssä) noudattaa multinormaalijakaumaa odotusarvona $\boldsymbol{\mu} = \mathrm{E}(\mathbf{z})$ ja $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{z})$, niin merkitsemme

$$\mathbf{z} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})\,. \tag{1.170}$$

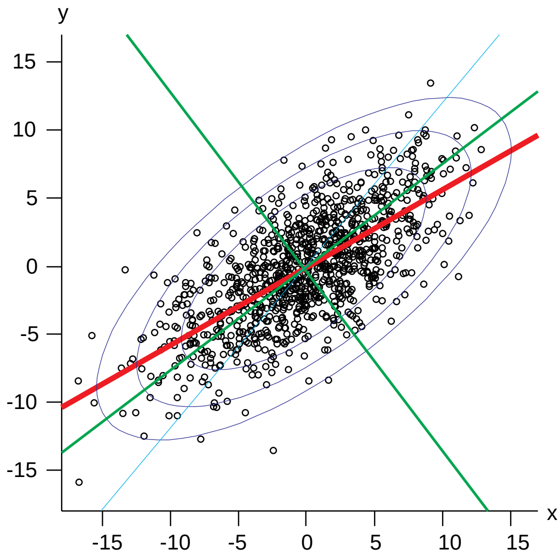
Yksi tapa määritellä multinormaalijakauma on seuraava: p-ulotteinen satunnaisvektori $\mathbf{z}$ noudattaa p-ulotteista multinormaalijakaumaa N_p, jos jokainen lineaarikombinaatio $\mathbf{a}'\mathbf{z}$, missä $\mathbf{a} \in \mathbb{R}^p$, noudattaa yksiulotteista normaalijakaumaa. Mikäli $\mathbf{a}'\mathbf{z} = b$, missä b on vakio, merkitään $\mathbf{a}'\mathbf{z} \sim \mathrm{N}(b, 0)$.

Toinen yhtäpitävä määritelmä on seuraava: p-ulotteinen satunnaisvektori $\mathbf{z}$, jolla on $\mathrm{E}(\mathbf{z}) = \boldsymbol{\mu}$ ja $\mathrm{cov}(\mathbf{z}) = \boldsymbol{\Sigma}$, noudattaa p-ulotteista multinormaalijakaumaa N_p, jos se voidaan esittää muodossa $\mathbf{z} = \boldsymbol{\mu} + \mathbf{F}\mathbf{u}$, missä $\mathbf{F}$ on $p \times r$-matriisi, $\mathrm{r}(\mathbf{F}) = r$, ja $\mathbf{u}$ on r-dimensioinen satunnaisvektori, jonka jokainen elementti noudattaa yksiulotteista normaalijakaumaa.

Jos $\mathbf{z} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, missä $\boldsymbol{\Sigma}$ positiivisesti definiitti, niin $\mathbf{z}$:n tiheysfunktio on

$$n(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \mathrm{e}^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{z}-\boldsymbol{\mu})}. \tag{1.171}$$

Lähteinä multinormaalijakaumaan mainittakoon mm. Rao (1973, ss. 525–528), ja Seber (2008, §20.5).



Kuvio 1.13. Havaintoja normaalijakaumasta $N_2(\mathbf{0}, \mathbf{\Sigma})$; $\sigma_x = 5$, $\sigma_y = 4$, $\varrho_{xy} = 0.7$. Regressiosuoran kulmakerroin $\hat{\beta}_1 \approx \varrho_{xy}\sigma_y/\sigma_x$. Kuviossa on myös x:n regressiosuora y:n suhteen. Tasa-arvokäyrän (ellipsin) ensimmäisen pääakseli on vektorin $\mathbf{t}_1$ suuntainen; $\mathbf{t}_1$ on $\mathbf{\Sigma}$:n suurinta ominaisarvoa vastaava ominaisvektori.

1.4.2 Satunnaislukujen generointi Survossa

Olkoon `KOE` Survon toimituskentässä oleva 100 havainnon havaintoaineisto, ja olkoon toimituskentässä seuraavat määritteeet:

```
1   U=PROBIT(RND(0)) V=PROBIT(RND(0))              (TMK A)
2   VAR U,V TO KOE
```

Tämä tarkoittaa, että muuttujan `U` arvoiksi generoituu 100 riippumatonta havaintoa $(0, 1)$-normaalista jakaumasta ja samoin `V`:n arvoiksi saadaan 100 riippumatonta havaintoa $N(0, 1)$:sta. Toisin sanoen:

> `KOE`-aineiston sarakkeet `U` ja `V` sisältävät 100 havainnon satunnaisotoksen kaksiulotteisesta normaalijakaumasta $N_2(\mathbf{0}, \mathbf{I}_2)$.

Generoituneesta aineistosta laskettu muuttujien `U` ja `V` välinen korrelaatio on jotakuinkin 0 (seurauksena näiden muuttujien välisestä riippumattomuudesta).

Uusien (`U`, `V`)-sarjojen generointi käy yksinkertaisesti siten, että aktivoidaan rivin 2 käsky uudestaan; näin saadaan toistuvasti 100 havainnon satunnaisotoksia $N_2(\mathbf{0}, \mathbf{I}_2)$:sta.

Tarkastellaan seuraavaa toimituskenttää:

(TMK B)

```
1  U=PROBIT(RND(0))   V=PROBIT(RND(0))
2  X=a*U
3  Y=b*(r*U+SQRT(1-r*r)*V)
4  VAR U,V TO KOE
5  VAR X,Y TO KOE / a=1 b=2 r=0.6
```

Kun aktivoidaan rivin 5 käsky (kun ensin rivi 4 on aktivoitu), laskee Survo muuttujien `X` ja `Y` arvot rivien 2 ja 3 kaavojen mukaan. Muuttujien `U` ja `V` arvot otetaan havaintomatriisista `KOE` – niitä ei nyt generoida uudestaan. Jos rivin 5 operaatio aktivoidaan uudestaan, tulos on täsmälleen sama kuin ensimmäisellä aktivoinnilla

Tarkastellaan sitten seuraavaa toimituskenttää, jossa muuttujia `U` ja `V` ei ole lainkaan havaintomatriisissa `KOE`:

(TMK C)

```
1  U=PROBIT(RND(0))  V=PROBIT(RND(0))
2  X=a*U
3  Y=b*(r*U+SQRT(1-r*r)*V)
4  VAR X,Y TO KOE / a=1 b=2 r=0.6
```

Rivin 4 käskyn aktivointi aiheuttaa sen, että Survo laskee muuttujien `X` ja `Y` arvot rivien 2 ja 3 kaavojen mukaan. Rivin 4 toistuva aktivointi generoi uusia (`X`, `Y`)-sarjoja.

Muuttujat `X` ja `Y` eivät enää olekaan riippumattomia satunnaismuuttujia. Nyt siis

$$X = aU\,, \tag{1.172a}$$

$$Y = b(rU + \sqrt{1-r^2}\,V)\,, \tag{1.172b}$$

missä $\mathrm{E}(\mathbf{z}) = \mathrm{E}\binom{U}{V} = \mathbf{0}$, $\mathrm{cov}(\mathbf{z}) = \mathrm{cov}\binom{U}{V} = \mathbf{I}_2$. Tällöin

$$\mathrm{E}(X) = \mathrm{E}(aU) = a\,\mathrm{E}(U) = 0, \tag{1.172c}$$

$$\mathrm{E}(Y) = \mathrm{E}[b(rU + \sqrt{1-r^2}\,V)] = br\,\mathrm{E}(U) + b\sqrt{1-r^2}\,\mathrm{E}(V) = 0, \tag{1.172d}$$

$$\mathrm{var}(X) = \mathrm{var}(aU) = a^2\,\mathrm{var}(U) = a^2, \tag{1.172e}$$

$$\begin{aligned}\mathrm{var}(Y) &= \mathrm{var}[b(rU + \sqrt{1-r^2}V)] \\ &= \mathrm{var}(brU) + \mathrm{var}(b\sqrt{1-r^2}V) \qquad [\mathrm{cov}(U,V)=0] \\ &= b^2r^2\,\mathrm{var}(U) + b^2(1-r^2)\,\mathrm{var}(V) \\ &= b^2r^2 + b^2(1-r^2) = b^2,\end{aligned} \tag{1.172f}$$

$$\begin{aligned}\mathrm{cov}(X,Y) &= \mathrm{cov}[aU, b(rU + \sqrt{1-r^2}\,V)] \\ &= \mathrm{cov}(aU, brU) + \mathrm{cov}(aU, b\sqrt{1-r^2}\,V) \\ &= \mathrm{cov}(aU, brU) = abr\,\mathrm{cov}(U,U) \\ &= abr\,\mathrm{var}(U) = abr\,.\end{aligned} \tag{1.172g}$$

[Huom: $\operatorname{cov}(U, U+V) = \operatorname{cov}(U,U) + \operatorname{cov}(U,V) = \operatorname{var}(U)$.] Täten

$$\operatorname{cov}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} a^2 & abr \\ abr & b^2 \end{pmatrix} = \mathbf{\Sigma}\,, \tag{1.172h}$$

ja korrelaatiomatriisi

$$\operatorname{cor}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} = \boldsymbol{\rho}\,. \tag{1.172i}$$

Matriisimerkinnöin (palaamme tähän myöhemmin) voimme kirjoittaa yhtälön

$$\mathbf{T} = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} a & 0 \\ br & b\sqrt{1-r^2} \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} := \mathbf{Az}, \tag{1.172j}$$

jolloin

$$\begin{aligned} \operatorname{cov}(\mathbf{Az}) &= \mathbf{A}\operatorname{cov}(\mathbf{z})\mathbf{A}' = \mathbf{AA}' \\ &= \begin{pmatrix} a & 0 \\ br & b\sqrt{1-r^2} \end{pmatrix} \begin{pmatrix} a & br \\ 0 & b\sqrt{1-r^2} \end{pmatrix} = \begin{pmatrix} a^2 & abr \\ abr & b^2 \end{pmatrix}. \end{aligned} \tag{1.172k}$$

Kuviossa 1.14 (s. 58) on 100 havainnon generointeja 2-ulotteisesta normaalijakaumasta.

1.5 Joitakin erityisiä matriiseja

Matriisin $\mathbf{A}$ *osittamisella* tarkoitetaan – kuten aiemmin jo on käynyt ilmi – $\mathbf{A}$:n tiettyjen vierekkäisten elementtien merkitsemistä matriisiksi. Olemme jo useamman kerran tarkastelleet sarakkeittain tai vaakariveittäin ositettuja matriiseja. Jos siis $\mathbf{A} \in \mathbb{R}^{3\times 4}$, niin $\mathbf{A}$ voidaan osittaa esim. siten, että

$$\mathbf{A} = (\mathbf{a}_1 : \mathbf{a}_2 : \mathbf{a}_3 : \mathbf{a}_4)\,, \tag{1.173}$$

missä $\mathbf{a}_i$:t ovat $\mathbf{A}$:n sarakkeita ja kukin $\mathbf{a}_i \in \mathbb{R}^3$. Jos $\mathbf{A}$ halutaan kirjoittaa ositetussa asussa vaakariveittäin, voidaan merkitä esim.

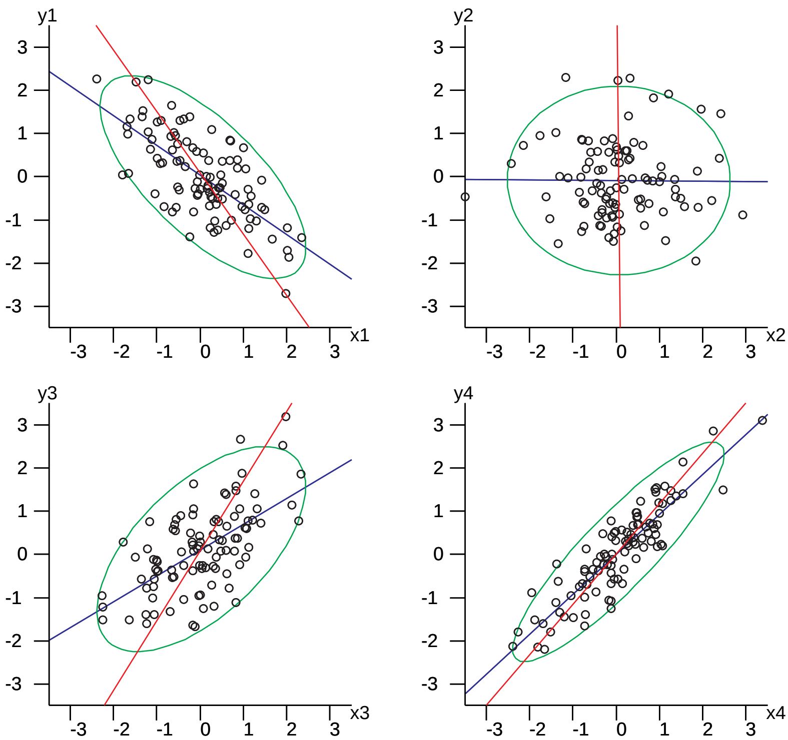
$$\mathbf{A} = \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \mathbf{a}'_{(3)} \end{pmatrix}, \tag{1.174}$$

missä $\mathbf{a}'_{(i)}$ on siis $\mathbf{A}$:n i. vaakarivi, $\mathbf{a}_{(i)} \in \mathbb{R}^4$.

Esimerkki 1.10. Olkoon

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \tag{1.175a}$$

```
 -   - SURVO MM  Thu Sep 23 18:03:15 2010            C:\SJP\D\   2000  100 0
*
*   2-ULOTTEISESTA NORMAALIJAKAUMASTA 100 havainnon otoksia
*
*U=PROBIT(RND(0))    V=PROBIT(RND(0))
*   x1=U   y1=r1*U+sqrt(1-r1*r1)*V     r1=-0.7
*   x2=U   y2=r2*U+sqrt(1-r2*r2)*V     r2=0
*   x3=U   y3=r3*U+sqrt(1-r3*r3)*V     r3=0.7
*   x4=U   y4=r4*U+sqrt(1-r4*r4)*V     r4=0.9
*VAR x1,y1 TO NOR100   / jokaisessa (x,y)-sarjassa eri U- ja V-arvot
*VAR x2,y2 TO NOR100
*VAR x3,y3 TO NOR100
*VAR x4,y4 TO NOR100
*CORR NOR100,CUR+1   /  VARS=x1,y1
*Means, std.devs and correlations of NOR100  N=100
*Variable  Mean         Std.dev.
*x1         0.067819    0.971943
*y1        -0.005665    0.959302
*Correlations:
*              x1       y1
* x1           1.0000 -0.6967
* y1          -0.6967  1.0000
*                                                  (MK10-003, MKT-01-NJ4)
```



Kuvio 1.14. Generointeja normaalijakaumasta.

missä $\mathbf{A}_{ij} \in \mathbb{R}^{2\times 2}$ ja siten siis esim.

$$\mathbf{A}_{11} = \begin{pmatrix} 1 & 2 \\ 5 & 6 \end{pmatrix}. \tag{1.175b}$$

Edelleen voidaan matriisi $\mathbf{A}$ osittaa esim. seuraavasti:

$$\mathbf{A} = (\mathbf{A}_1 : \mathbf{A}_2),\ \ \mathbf{A}_i \in \mathbb{R}^{4\times 2}; \tag{1.175c}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{b}' \\ \mathbf{B} \end{pmatrix},\ \ \mathbf{b}' = (1,2,3,4) \in \mathbb{R}^{1\times 4},\ \ \mathbf{B} \in \mathbb{R}^{3\times 4}; \tag{1.175d}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{C} & \mathbf{d} \\ \mathbf{e}' & f \end{pmatrix},\ \ f = 16 \in \mathbb{R}. \tag{1.175e}$$

□

Tarkastelemme vain sellaisia matriisien osituksia, joissa alakkaisten ositteiden sarakemäärät ovat samat ja samoin rinnakkaisten ositteiden vaakarivien lukumäärät ovat samat.

Esimerkki 1.11. Regressioanalyysissä käytetään usein seuraavanlaista merkintätapaa:

$$\begin{aligned} \mathbf{X} &= (\mathbf{1}_n : \mathbf{X}_0) = (\mathbf{1}_n : \mathbf{x}_1 : \ldots : \mathbf{x}_k) \\ &= \begin{pmatrix} 1 & \mathbf{x}'_{(1)} \\ 1 & \mathbf{x}'_{(2)} \\ \vdots & \vdots \\ 1 & \mathbf{x}'_{(n)} \end{pmatrix} \in \mathbb{R}^{n\times(k+1)},\ \ \mathbf{X} \text{ on ns. } \textit{mallimatriisi,} \end{aligned} \tag{1.176}$$

missä $\mathbf{1}_n$ (tai lyhyesti $\mathbf{1}$) on n elementin pystyvektori ”tolppaykkönen”, ja $\mathbf{x}_i$ sisältää i. selittäjämuuttujan x_i saamat n havaintoarvoa. Tällöin tietenkin

$$\mathbf{x}_i \in \mathbb{R}^n,\ \ i = 1,\ldots,k, \qquad \mathbf{x}_{(i)} \in \mathbb{R}^k,\ \ i = 1,\ldots,n. \tag{1.177}$$

Matriisi $\mathbf{X}_0$ sisältää siis varsinaisten selittäjien saamat arvot ja $\mathbf{X}$:n ensimmäinen sarake $\mathbf{1}_n$ saa perustelunsa regressiomallin vakiotermistä. Mainittakoon vielä, että jos merkitään $\boldsymbol{\beta} = \binom{\beta_0}{\boldsymbol{\beta}_{\mathbf{x}}}$ missä $\boldsymbol{\beta}_{\mathbf{x}} \in \mathbb{R}^k$ ja $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, niin

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} = (\mathbf{1}_n : \mathbf{X}_0)\begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_{\mathbf{x}} \end{pmatrix} = \beta_0\mathbf{1}_n + \mathbf{X}_0\boldsymbol{\beta}_{\mathbf{x}}, \tag{1.178a}$$

$$\mu_i = \beta_0 + \mathbf{x}'_{(i)}\boldsymbol{\beta}_{\mathbf{x}}, \quad i = 1,\ldots,n. \tag{1.178b}$$

□

Jos $\mathbf{A} \in \mathbb{R}^{n\times n}$, niin $\mathbf{A}$:ta sanotaan *neliömatriisiksi*. Neliömatriisin $\mathbf{A}$ elementit a_{11}, a_{22}, a_{33}, ..., a_{nn} muodostavat $\mathbf{A}$:n *lävistäjän* eli *diagonaalin*. Merkintä diag($\mathbf{A}$) tarkoittaa matriisia

$$\operatorname{diag}(\mathbf{A}) = \mathbf{A}_\delta = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}. \tag{1.179}$$

Voimme käyttää myös merkintää

$$\operatorname{diag}(a_{11}, a_{22}, a_{33}, \dots, a_{nn}) = \operatorname{diag}(\mathbf{A}). \tag{1.180}$$

Lävistäjämatriisiksi sanotaan neliömatriisia, jonka kaikki ei-lävistäjäelementit ovat nollia.

Neliömatriisin $\mathbf{A}_{n\times n}$ *jäljellä* eli *tracella* tarkoitetaan $\mathbf{A}$:n lävistäjäelementtien summaa:

$$\operatorname{tr}(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{nn}. \tag{1.181}$$

Havaitsemme välittömästi, että esim. kovarianssimatriisin jälki on varianssien summa:

$$\operatorname{tr}[\operatorname{cor_d}(\mathbf{U}_{n\times p})] = \operatorname{tr}(\mathbf{S}_{p\times p}) = s_1^2 + s_2^2 + \cdots + s_p^2, \tag{1.182a}$$

$$\operatorname{tr}[\operatorname{ssp}(\mathbf{U}_{n\times p})] = \operatorname{tr}(\mathbf{T}_{p\times p}) = t_{11} + t_{22} + \cdots + t_{pp}, \tag{1.182b}$$

$$\operatorname{tr}[\operatorname{cor_d}(\mathbf{U}_{n\times p})] = \operatorname{tr}(\mathbf{R}_{p\times p}) = p, \tag{1.182c}$$

$$\operatorname{tr}[\operatorname{cov}(\mathbf{z}_{p\times 1})] = \operatorname{tr}(\mathbf{\Sigma}_{p\times p}) = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_p^2. \tag{1.182d}$$

Yksinkertaisesta määritelmästään huolimatta trace-operaattori osoittautuu yllättävän käyttökelpoiseksi monissa yhteyksissä.

Edelleen merkitään

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = (\mathbf{i}_1 : \mathbf{i}_2 : \dots : \mathbf{i}_n), \text{ yksikkömatriisi,}$$

$$\mathbf{i}_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ j\text{. elementti on 1, muut nollia, yksikkövektori.}$$

Kuten aiemmin on jo todettu, matriisin $\mathbf{A}$ $(n \times m)$ *transpoosilla* tarkoitetaan matriisia $\mathbf{A}'$ $(m \times n)$, joka saadaan vaihtamalla $\mathbf{A}$:n kukin sarake vastaavaksi vaakariviksi. Siten esim.

$$\mathbf{A}' = (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m)' = \begin{pmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_m' \end{pmatrix}. \tag{1.183}$$

On helppo havaita, että jos

$$\mathbf{A} = \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{pmatrix}, \quad \text{niin} \quad \mathbf{A}' = \begin{pmatrix} \mathbf{B}' & \mathbf{D}' \\ \mathbf{C}' & \mathbf{E}' \end{pmatrix}. \tag{1.184}$$

Tarkastellaan esimerkkinä matriisia

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 10 \end{pmatrix} = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{C} & \mathbf{E} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}. \tag{1.185a}$$

Tällöin

$$\begin{pmatrix} \mathbf{B}' & \mathbf{C}' \\ \mathbf{0} & \mathbf{E}' \end{pmatrix} = \begin{pmatrix} 1 & 2 & 4 & 7 \\ 0 & 3 & 5 & 8 \\ 0 & 0 & 6 & 9 \\ 0 & 0 & 0 & 10 \end{pmatrix} = \mathbf{A}'. \tag{1.185b}$$

Matriisi $\mathbf{A}$ (1.185a):ssa on esimerkki *yläkolmiomatriisista*: sen lävistäjän yläpuoliset elementit ovat kaikki nollia. Vastaavasti määritellään alakolmiomatriisi.

Neliömatriisi $\mathbf{A}$ on *symmetrinen*, jos $\mathbf{A}' = \mathbf{A}$. Luonnollisesti satunnaisvektorin $\mathbf{x}$ kovarianssimatriisi $\operatorname{cov}(\mathbf{x})$ on symmetrinen, mutta satunnaisvektorien $\mathbf{x}$ ja $\mathbf{y}$ keskinäinen kovarianssimatriisi $\operatorname{cov}(\mathbf{x}, \mathbf{y})$ ei tietenkään välttämättä ole symmetrinen eikä edes neliömatriisi. Jos $\mathbf{A}' = -\mathbf{A}$, niin $\mathbf{A}$ on *vinosymmetrinen*. Esimerkiksi

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{1.186}$$

on vinosymmetrinen.

Mainittakoon vielä tämän luvun lopuksi, että matriisien ja lineaarialgebran roolia lineaarisissa malleissa ja monimuuttujamenetelmissä ovat lyhyesti käsitelleet arikkeleissaan Puntanen, Seber & Styan (2013a,b).

Harjoitustehtäviä

1.1. Tarkastellaan muuttujia $\mathbf{y}_1$, $\mathbf{y}_2$ ja $\mathbf{y}_3$, joiden korrelaatiomatriisi on

$$\mathbf{R} = \begin{pmatrix} 1 & r & r_{13} \\ r & 1 & r \\ r_{31} & r & 1 \end{pmatrix} = \operatorname{cor_d}(\mathbf{y}_1 : \mathbf{y}_2 : \mathbf{y}_3)\,.$$

Määritä r_{13} siten että osittaiskorrelaatiokerroin $r_{13\cdot 2} = 0$. Osittaiskorrelaatiokertoimen lauseke on

$$r_{xy\cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}\,.$$

1.2. Tarkastellaan symmetristä matriisia

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix},$$

jossa $|r_{12}|$, $|r_{13}|$ ja $|r_{23}| \leq 1$. Tällöin $\mathbf{R}$ on siis tavallaan"ehdokas" korrelaatiomatriisiksi eli se "näyttää" korrelaatiomatriisilta. (On osoitettavissa, että $\mathbf{R}$ on korrelaatiomatriisi jos ja vain jos jokin seuraavista yhtäpitävistä ehdoista on voimassa.) Osoita seuraavien viiden väitteen keskinäinen yhtäpitävyys:

(a) $|r_{12\cdot 3}| \leq 1$, (b) $|r_{13\cdot 2}| \leq 1$, (c) $|r_{23\cdot 1}| \leq 1$,
(d) $r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1$,
(e) $\det(\mathbf{R}) \geq 0$ [det = determinantti].

Huom: Osittaiskorrelaatiota $r_{ij\cdot k}$ tarkasteltaessa on oletettava, että $r_{ik}^2 < 1$ ja $r_{jk}^2 < 1$. Ks. determinantti: (4.22) (s. 132).

1.3. Laske osittaiskorrelaatiokertoimet $\operatorname{pcor}(y_i, y_{i+2} \mid y_{i+1}) = \varrho_{i,i+2\cdot i+1}$ seuraavasta korrelaatiomatriisista:

$$\operatorname{cor}(\mathbf{y}) = \operatorname{cor}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & \varrho & \varrho^2 & \dots & \varrho^{n-1} \\ \varrho & 1 & \varrho & \dots & \varrho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \varrho^{n-1} & \varrho^{n-2} & \varrho^{n-3} & \dots & 1 \end{pmatrix} = \{\varrho^{|i-j|}\}.$$

1.4. Kuinka monella eri tavalla matriisi

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix}$$

voidaan osittaa? Huomaa, että tarkastelemme vain sellaisia osituksia, joissa aina kahden alakkaisen ositteen sarakemäärät ovat samat ja samoin kahden rinnakkaisen ositteen vaakarivien lukumäärät ovat samat.

1.5. Tarkastellaan vektoreita

$$\mathbf{x} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} 1 \\ 1 \\ -\sqrt{2} \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ \sqrt{2} \end{pmatrix} \quad \text{ja} \quad \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

(a) Piirrä muuttuja-avaruuteen ($\mathbb{R}^3$) vektoreiden $\mathbf{x}$ ja $\mathbf{y}$ virittämä taso $\mathscr{C}(\mathbf{x} : \mathbf{y})$. Havaitse, että $\mathbf{x}$ ja $\mathbf{y}$ ovat kohtisuorassa toisiaan vastaan: $\mathbf{x}'\mathbf{y} = 0$.

(b) Määritä reaaliluvut a ja b siten että $\mathbf{y} = a\mathbf{1} + b\mathbf{x}$.

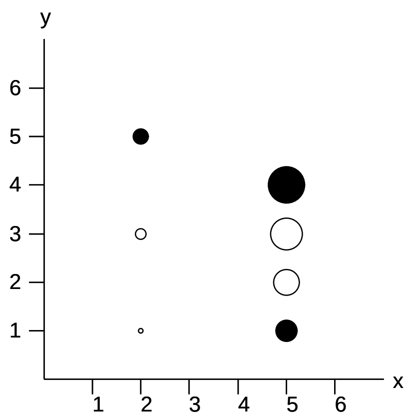
(c) Todista täsmällisesti: $\mathscr{C}(\mathbf{x} : \mathbf{y}) = \mathscr{C}(\mathbf{x} : \mathbf{1}) = \mathscr{C}(\mathbf{y} : \mathbf{1})$.

(d) Piirrä aineiston $(\mathbf{x} : \mathbf{y})$ pisteparvi havaintoavaruuteen $\mathbb{R}^2$. Regressiosuora?

(e) Piirrä aineiston $(\mathbf{x} : \mathbf{y} : \mathbf{1})$ pisteparvi havaintoavaruuteen $\mathbb{R}^3$.

(f) Tee kohdat (a)–(e) vektoreille $\mathbf{u}$ ja $\mathbf{v}$.

1.6. Oheisessa kuvion 1.15 pisteparvessa täytetty ympyrä ● merkitsee, että kyseessä on mies ja avoin ympyrä ○ kertoo kyseessä on nainen. Lisäksi ympyrän koko kertoo havainnon saaman arvon muuttujalla z (suuri ympyrä = suuri arvo). Muodosta alkuperäinen havaintomatriisi $\mathbf{U}$ (suurinpiirtein). Määritä lisäksi x:n ja y:n välinen korrelaatio (a) naisten ryhmässä, (b) miesten keskuudessa, (c) koko aineistossa.



Kuvio 1.15. Tehtävään 1.6 liittyvä pisteparvi.

1.7. Määritä alkuperäiset havaintomatriisit seuraavista frekvenssitaulukoista (jokaisessa on rivimuuttujana x). Laske myös kunkin muuttujaparin korrelaatiot. Mitä tapahtuu, jos nollafrekvenssi vaihtaa paikkaa muiden frekvenssien pysyessä keskenään yhtäsuurina? Mikä on se lineaarinen yhteys, joka vallitsee muuttujien y ja u välillä (kun siis y:n arvoa 0 vastaa u:n arvo 2 ja y:n arvoa 1 vastaa u:n arvo 5)?

$$
\begin{array}{cc|cc|}
 & & \multicolumn{2}{c}{y} \\
 & & 0 & 1 \\
\cline{3-4}
x & 0 & 1 & 1 \\
 & 1 & 0 & 1 \\
\cline{3-4}
\end{array}
\qquad
\begin{array}{c|cc|}
 & \multicolumn{2}{c}{z} \\
 & 0 & 1 \\
\cline{2-3}
0 & 2 & 2 \\
1 & 0 & 2 \\
\cline{2-3}
\end{array}
\qquad
\begin{array}{c|cc|}
 & \multicolumn{2}{c}{u} \\
 & 2 & 5 \\
\cline{2-3}
0 & 1 & 1 \\
1 & 0 & 1 \\
\cline{2-3}
\end{array}
$$

1.8. Tarkastellaan seuraavien taulukkojen mukaisten satunnaismuuttujaparien jakaumia: kunkin arvoparin todennäköisyys saadaan jakamalla solussa oleva luku kaikkien solujen lukujen summalla. Laske kunkin muuttujaparin korrelaatiot.

$$
\text{(a)}\quad
\begin{array}{cc|cc|}
 & & \multicolumn{2}{c}{y} \\
 & & 0 & 1 \\
\cline{3-4}
x & 0 & 1 & 1 \\
 & 1 & 1 & 0 \\
\cline{3-4}
\end{array}
\qquad
\text{(b)}\quad
\begin{array}{c|cc|}
 & \multicolumn{2}{c}{z} \\
 & 0 & 1 \\
\cline{2-3}
0 & 1 & 2 \\
1 & 2 & 0 \\
\cline{2-3}
\end{array}
\qquad
\text{(c)}\quad
\begin{array}{c|cc|}
 & \multicolumn{2}{c}{u} \\
 & 0 & 1 \\
\cline{2-3}
0 & a & b \\
1 & c & d \\
\cline{2-3}
\end{array}
$$

1.9. Tarkastellaan taulukkoa (c) sekä sen esitystä todennäköisyyksien avulla (kaikki taulukon luvut jaetaan n:llä):

$$
\begin{array}{cc ccc}
 & & \multicolumn{3}{c}{y} \\
 & & 0 & 1 & \text{yht} \\
\hline
 & 0 & a & b & \alpha \\
\text{x} & 1 & c & d & \beta \\
\hline
 & \text{yht} & \gamma & \delta & n
\end{array}
\qquad\qquad
\begin{array}{cc ccc}
 & & \multicolumn{3}{c}{y} \\
 & & 0 & 1 & \text{yht} \\
\hline
 & 0 & p_{11} & p_{12} & p_{1\cdot} \\
x & 1 & p_{21} & p_{22} & p_{2\cdot} \\
\hline
 & \text{yht} & p_{\cdot 1} & p_{\cdot 2} & 1
\end{array}
$$

Osoita että

$$
\begin{aligned}
\mathrm{E}(y) &= \frac{\delta}{n} = \frac{b+d}{n}, & \mathrm{E}(x) &= \frac{\beta}{n} = \frac{c+d}{n}, \\
\mathrm{var}(y) &= \frac{\delta}{n}\left(1-\frac{\delta}{n}\right) = \frac{\delta}{n}\frac{\gamma}{n}, & \mathrm{var}(x) &= \frac{\beta}{n}\left(1-\frac{\beta}{n}\right) = \frac{\beta}{n}\frac{\alpha}{n}, \\
\mathrm{cov}(x,y) &= \frac{ad-bc}{n^2}, & \mathrm{cor}(x,y) &= \varrho = \frac{ad-bc}{\sqrt{\alpha\beta\gamma\delta}}\,.
\end{aligned}
$$

Ilmaise em. lausekkeet myös todennäköisyyksien avulla.

1.10. Jos tulkitsemme taulukon (c) aivan tavalliseksi frekvenssitaulukoksi, niin siitä lasketut otossuureet ovat seuraavat (miksi?):

$$\bar{x} = \frac{\delta}{n}, \quad \bar{y} = \frac{\beta}{n}, \quad s_x^2 = \frac{1}{n-1}\frac{\gamma\delta}{n}, \quad s_y^2 = \frac{1}{n-1}\frac{\alpha\beta}{n},$$
$$s_{xy} = \frac{1}{n-1}\frac{ad-bc}{n}, \quad r_{xy} = r = \varrho.$$

Huom: Taulukosta (c) laskettu khiin neliö on $\chi^2 = \frac{n(ad-bc)^2}{\alpha\beta\gamma\delta} = nr^2$.

1.11. Satunnaismuuttujien x ja y yhteisjakauma ilmenee oheisista taulukoista: kunkin arvoparin todennäköisyys on solussa oleva luku. Merkitään $\mathbf{u} = \binom{x}{y}$. Määritä $\mathrm{E}(\mathbf{u})$, $\mathrm{cov}(\mathbf{u})$ ja $\mathrm{cor}(\mathbf{u})$.

		y		
		1	2	3
x	1	1/9	1/9	1/9
	2	1/9	1/9	1/9
	3	1/9	1/9	1/9

		y		
		1	2	3
x	1	0	0	1/3
	2	0	1/3	0
	3	1/3	0	0

1.12. Tarkastellaan tilannetta, jossa aineistoon halutaan sovittaa regressiomalliksi 3. asteen polynomi $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3 + \varepsilon$. Selittäjät ovat tällöin siis x:n potensseja: $x_1 = x$, $x_2 = x^2$, $x_3 = x^3$. Aineisto on kerätty siten, että selitettävä muuttuja y on havainnoitu x:n arvoilla 1, 2, 3, ..., 10. Muodosta mallimatriisi $\mathbf{X}_{10\times 4} = (\mathbf{1} : \mathbf{x}_1 : \mathbf{x}_2 : \mathbf{x}_3)$. Entä jos jokaisella x:n arvolla on havainnoitu y kolme kertaa?

1.13. Oletetaan, että meillä on kaksi postipakettia, A ja B, joiden todellisista painoista olemme kiinnostuneita. Käytössä on vaaka, joka antaa A:n painoksi 300 g ja B:n painoksi 500 g. Kun ne punnitaan yhdessä, vaaka näyttää 1100 g. Kuten huomaat, mittaustuloksissa on satunnaisvirhettä: vaaka ei samaa esinettä punnitessa anna välttämättä samaa tulosta. Muodosta aineistosta havaintomatriisi. Entä jos A punnitaan a kertaa, B b kertaa ja yhteispunnitus toistetaan c kertaa?

1.14. Mikä on seuraavan siirtymämatriisin tulkinta? Ks. Feller (1957, s. 341).

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}, \quad p + q = 1.$$

1.15. Tarkastellaan kolmen tilan siirtymätodennäköisyysmatriisia

$$\mathbf{P} = \begin{pmatrix} 19/20 & 1/20 & 0 \\ 0 & 2/3 & 1/3 \\ 1/5 & 0 & 4/5 \end{pmatrix},$$

missä tilat ovat E_1 =”olio nukkuu”, E_2 =”liikkuu”, E_3 = ”syö”. Siis jos olio on alunperin tilassa E_1, niin todennäköisyys että se on seuraavana ajanhetkenä tilassa E_j on 1. vaakarivin luku p_{1j}. Sisältäköön vektori $\mathbf{x}_0$ alkutilan (hetki 0) todennäköisyysjakauman. Tällöin hetken 1 todennäköisyydet saadaan lausekkeesta $\mathbf{x}_1 = \mathbf{P}'\mathbf{x}_0$. Vastaavasti saadaan $\mathbf{x}_2 = \mathbf{P}'\mathbf{x}_1, \dots$. Tutki Survolla tilatodennäköisyyksien mahdollista suppenemista, kun $\mathbf{x}_0 = (1, 0, 0)'$ eli alussa olio nukkuu. Toisin sanoen: tutki miten käy lausekkeelle $(\mathbf{P}')^n\mathbf{x}_0'$, kun n kasvaa. Kokeile myös jollakin muulla $\mathbf{x}_0$:n arvolla.

Mustonen (1996, ss. 199–201)

1.16. Kuinka monta erilaista lukua enintään voi olla p muuttujan (a) korrelaatiomatriisissa, (b) kovarianssimatriisissa?

1.17. Osoita, että matriisi $\mathbf{A} - \mathbf{A}'$ on vinosymmetrinen.

1.18. Olkoon $\mathbf{C} = \mathbf{I}_n - \mathbf{J}$, missä $\mathbf{J} = \left\{\frac{1}{n}\right\}$. Määritä $\operatorname{tr}(\mathbf{J})$ ja $\operatorname{tr}(\mathbf{C})$.

1.19. Olkoon $\mathbf{U}' = (\mathbf{u}_1 : \mathbf{u}_2)' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \dots : \mathbf{u}_{(n)})$ satunnaisotos 2-ulotteisesta jakaumasta, jonka odotusarvo on $\boldsymbol{\mu}$, kovarianssimatriisi on $\boldsymbol{\Sigma}$, ja korrelaatiomatriisi on $\boldsymbol{\rho}$. Vakuuttaudu, että

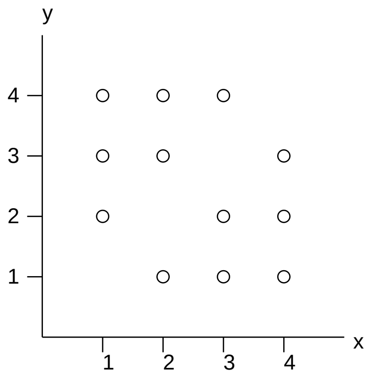
(a) $$\operatorname{cov}[\operatorname{vec}(\mathbf{u}_1 : \mathbf{u}_2)] = \operatorname{cov}\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2\mathbf{I}_n & \sigma_{12}\mathbf{I}_n \\ \sigma_{21}\mathbf{I}_n & \sigma_2^2\mathbf{I}_n \end{pmatrix} = \boldsymbol{\Sigma} \otimes \mathbf{I}_n\,,$$

$$\operatorname{cor}[\operatorname{vec}(\mathbf{u}_1 : \mathbf{u}_2)] = \operatorname{cor}\begin{pmatrix} \mathbf{u}_i \\ \mathbf{u}_j \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n & \varrho_{12}\mathbf{I}_n \\ \varrho_{21}\mathbf{I}_n & \mathbf{I}_n \end{pmatrix} = \boldsymbol{\rho} \otimes \mathbf{I}_n\,,$$

(b) $$\operatorname{cov}\big[\operatorname{vec}\big(\mathbf{u}_{(i)} : \mathbf{u}_{(j)}\big)\big] = \operatorname{cov}\begin{pmatrix} \mathbf{u}_{(i)} \\ \mathbf{u}_{(j)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{pmatrix} = \mathbf{I}_2 \otimes \boldsymbol{\Sigma}\,,$$

$$\operatorname{cor}\big[\operatorname{vec}\big(\mathbf{u}_{(i)} : \mathbf{u}_{(j)}\big)\big] = \operatorname{cor}\begin{pmatrix} \mathbf{u}_{(i)} \\ \mathbf{u}_{(j)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\rho} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\rho} \end{pmatrix} = \mathbf{I}_2 \otimes \boldsymbol{\rho}\,.$$

1.20. Olkoon $\mathbf{y} = \binom{y_1}{y_2}$ palauttamatta poimittu yksinkertainen satunnaisotos luvuista $\{1, 2, 3, 4\}$. Osoita, että $\mathbf{y}$:n todennäköisyysjakauma on kuvion 1.16 mukainen. (Vrt myös esimerkki 1.5, s. 41.) Kaikki 12 mahdollista arvoparia ovat nyt yhtä todennäköisiä. Määritä $\mathbf{y}$:n odotusarvo $\mathrm{E}(\mathbf{y})$, $\mathbf{y}$:n kovarianssimatriisi $\operatorname{cov}(\mathbf{y})$ ja $\mathbf{y}$:n korrelaatiomatriisi $\operatorname{cor}(\mathbf{y})$.



Kuvio 1.16. (Ks. tehtävä 1.20.) Kaksiulotteinen diskreetti todennäköisyysjakauma: kunkin arvoparin todennäköisyys $\frac{1}{12}$.

1.21. Oletetaan että muuttujista y_1 ja y_2 on saatu 12 havaintoa, jotka ryhmittyvät edellisen tehtävä kuvion 2 tapaan. Muodosta (Survon toimituskenttään) alkuperäinen havaintomatriisi $\mathbf{Y} = (\mathbf{y}_1 : \mathbf{y}_2)$. Laske muuttujien keskiarvot, kovarianssimatriisi $\mathrm{cov_d}(\mathbf{Y})$ sekä korrelaatiomatriisi $\mathrm{cor_d}(\mathbf{Y})$. Vertaa tuloksia edellisen tehtävän tuloksiin. Ks. myös kuvion 1.17 (s. 68) mukainen tulostus.

1.22. Tarkastellaan seuraavaa aineistoa:

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{array}{cl} \begin{pmatrix} 2 & 2 \\ 5 & 2 \\ 5 & 5 \end{pmatrix} & \begin{array}{l} \text{Antti} \\ \text{Liisa} \\ \text{Leena} \end{array} \end{array}$$

Oheisena on kaksi kuviota odottamassa täydennystä: toinen on havaintoavaruus ja toinen muuttuja-avaruus, joudut valitsemaan kumpaan kuvioon milloinkin piirrät.

(a) Piirrä havainnot havaintoavaruuteen.

(b) Piirrä muuttujien $\mathbf{x}$ ja $\mathbf{y}$ keskiarvovektori $\bar{\mathbf{u}}$.

(c) Piirrä keskistettyä 1. havaintoa vastaava piste $\tilde{\mathbf{u}}_{(1)}$.

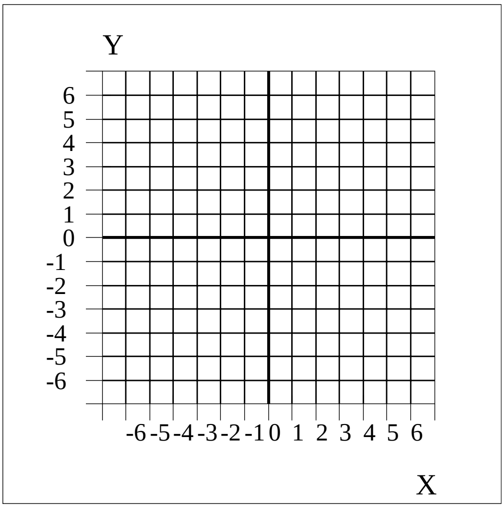
(d) Piirrä muuttujavektori $\mathbf{y}$.

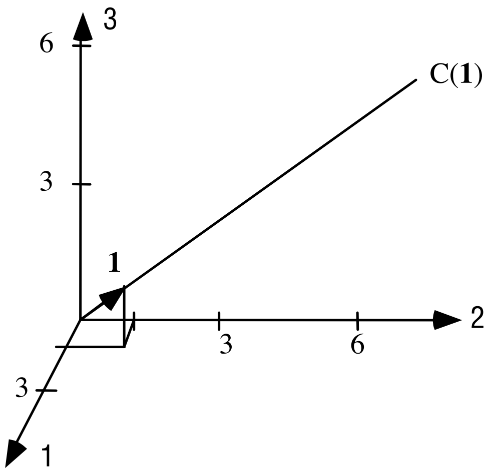
(e) Piirrä $\mathbf{y}$:n keskiarvovektori $\bar{\bar{\mathbf{y}}}$.

(f) Piirrä keskistetty $\tilde{\mathbf{y}}$.

```
   -   - SURVO MM  Thu Sep 23 18:58:46 2010            C:\SJP\D\   2000  100 0
*
*CORR TWELVE,CUR+1  /  VARS=x,y
*Means, std.devs and correlations of TWELVE  N=12
*Variable  Mean        Std.dev.
*x         2.500000    1.167748
*y         2.500000    1.167748
*Correlations:
*              x       y
* x             1.0000 -0.3333
* y            -0.3333  1.0000
*
*   sx=1.167748          sx2=sx^2     sy=sx
*   sx2=1.363635391504     = x:n otosvarianssi
*   sxy=sx*sy*rxy = otoskovarianssi   sxy=-0.45454513050133
*   rxy=-1/3  datasta
*
*   my=mx  mx=2.5 = x:n odotusarvo kun sitä pidetään sat.mjana
*          = x:n keskiarvo kun aineisto on empiirinen data
*   sigmax2=((1-mx)^2+(2-mx)^2+(3-mx)^2+(4-mx)^2)/4
*   sigmax2=1.25    teoreettinen  varianssi
*   sigmax=SQRT(sigmax2)    sigmax=1.1180339887499
*   sigmay2=sigmax2    sigmay=sigmax
*   Satunnaismuuttuja X on Tasd(1,4), joten diskreetin tasajakauman
*   ominaisuuksien nojalla VARX=(K^2-1)/12   K=4
*   VARX=1.25 (= teoreettinen varianssi; eroaa otosvarianssista!)
*
*   vakio=11/12   Jos tällä vakiolla kerrotaan sx2, sy2 ja sxy, niin
*   saadaan teoreettiset varianssit ja kovarianssi  (MIKSI?)
*   Teoreettinen varianssi on   vakio*sx2=1.2499991088787
*   Teoreettinen kovarianssi on cov=vakio*sxy
*
*   cov=-0.41666636962622        teoreettinen kovarianssi:   -(N+1)/12
*   cor=cov/SQRT(sigmax*sigmay)
*   cor=-0.37267773056891
*       YLEISESTI:  COR=-1/(N-1)      COR=-0.33333333333333
*                   COV=-(N+1)/12     COV=-0.41666666666667
*                                              (MK10-002, MKT-01-TW)
```

Kuvio 1.17. Harjoitustehtävään 1.21 liittyviä laskelmia.





1.23. Vektorit $\mathbf{x}$ ja $\mathbf{y}$ ovat ortonormaaleja, jos $\mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{y} = 1$ ja $\mathbf{x}'\mathbf{y} = 0$. Matriisi $\mathbf{A}$ on ortogonaalinen, jos se on neliömatriisi ja sen sarakkeet ovat keskenään ortonormaaleja eli kaikki sarakkeet ovat 1:n pituisia ja keskenään kohtisuorassa (jolloin myös sen vaakariveillä on vastaava ominaisuus). Mitkä seuraavista matriiseista ovat

(1) sarakkeittain ortogonaalisia,	A	B	C	D	E	F	G	H
(2) vaakariveittäin ortogonaalisia,	A	B	C	D	E	F	G	H
(3) sarakkeittain ortonormaaleja,	A	B	C	D	E	F	G	H
(4) vaakariveittäin ortonormaaleja,	A	B	C	D	E	F	G	H
(5) ortogonaalisia?	A	B	C	D	E	F	G	H

$$\mathbf{A} = \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \qquad \mathbf{D} = \frac{1}{3} \begin{pmatrix} 2 & -2 \\ 1 & 2 \\ 2 & 1 \end{pmatrix}$$

$$\mathbf{E} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix} \qquad \mathbf{F} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

$$\mathbf{G} = \frac{1}{2} \begin{pmatrix} 1 & \sqrt{3} \\ 0 & 0 \\ -\sqrt{3} & 1 \end{pmatrix} \qquad \mathbf{H} = \frac{1}{2} \begin{pmatrix} 1 & -1 & 1 \\ -2 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

1.24 (Kontingessitaulukko). Tarkastellaan dikotomisia muuttujia x (arvot A_1, A_2) ja y (arvot B_1, B_2). Oletetaan että meillä on n havaintoa näistä muuttujista. Määritellään uudet muuttujat x_1 ja x_2 seuraavasti:

$$x_1 = 1 \text{ jos } x \text{ saa arvon } A_1, \text{ ja } x_1 = 0 \text{ muuten,}$$
$$x_2 = 1 \text{ jos } x \text{ saa arvon } A_2, \text{ ja } x_2 = 0 \text{ muuten,}$$

ja olkoot y_1 ja y_2 määritelty vastaavalla tavalla arvojen B_1 ja B_2 suhteen. Merkitään saatua $n \times 4$-havaintomatriisia

$$\mathbf{U} = (\mathbf{x}_1 : \mathbf{x}_2 : \mathbf{y}_1 : \mathbf{y}_2) = (\mathbf{X} : \mathbf{Y}).$$

Olemme kiinnostuneita muuttujien x ja y välisestä tilastollisesta riippuvuudesta ja sen vuoksi muodostamme seuraavan frekvenssitaulukon (kontingenssitaulukon):

		y		
		B_1	B_2	yht
x	A_1	f_{11}	f_{12}	r_1
	A_2	f_{21}	f_{22}	r_2
	yht	c_1	c_2	n

Olkoon e_{ij} solun (i,j) odotettu frekvenssi (χ^2-testisuuretta varten laskettu) ja merkitään

$$e_{ij} = \frac{r_i c_j}{n}, \quad \mathbf{E} = (\mathbf{e}_1 : \mathbf{e}_2), \quad \mathbf{F} = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} = (\mathbf{f}_1 : \mathbf{f}_2),$$

ja $\mathbf{c} = \binom{c_1}{c_2}$, $\mathbf{r} = \binom{r_1}{r_2}$. Oletetaan, että vektorien $\mathbf{c}$ ja $\mathbf{r}$ kaikki elementit ovat nollasta eroavia. Osoita:

(a) $\mathrm{cor_d}(\mathbf{x}_1, \mathbf{x}_2) = -1$, $\mathrm{r}(\mathbf{X} : \mathbf{Y}) \leq 3$, $\mathrm{r}[\mathrm{cor_d}(\mathbf{X} : \mathbf{Y})] \leq 2$,

(b) $\mathbf{X}'\mathbf{1}_n = \mathbf{R}$, $\mathbf{y}'\mathbf{1}_n = \mathbf{r}$, $\mathbf{X}'\mathbf{X} = \mathrm{diag}(\mathbf{R}) = \mathbf{D_r}$, $\mathbf{Y}'\mathbf{Y} = \mathrm{diag}(\mathbf{c}) = \mathbf{D_c}$,

(c) $\mathbf{X}'\mathbf{Y} = \mathbf{F}$, $\mathbf{E} = \mathbf{rc}'/n = \mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{Y}/n = \mathbf{X}'\mathbf{J}\mathbf{Y}$.

(d) Matriisin $\mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}$ sarakkeet muodostavat muuttujan x ehdollisten jakaumien suhteelliset frekvenssit.

(e) $\mathbf{F} - \mathbf{E} = \mathbf{X}'\mathbf{C}\mathbf{Y}$, missä $\mathbf{C}$ on keskistäjämatriisi ja $\frac{1}{n-1}(\mathbf{F} - \mathbf{E})$ on x- ja y-muuttujien välinen kovarianssimatriisi.

1.25. (Jatkoa ...)

(f) Osoita että $\mathbf{X}'\mathbf{C}\mathbf{Y}$, $\mathbf{X}'\mathbf{C}\mathbf{X}$ ja $\mathbf{Y}'\mathbf{C}\mathbf{Y}$ ovat kaksoiskeskitettyjä; $\mathbf{A}_{n\times p}$ on kaksoiskeskitetty jos $\mathbf{A}\mathbf{1}_p = \mathbf{0}_n$ ja $\mathbf{A}'\mathbf{1}_n = \mathbf{0}_p$.

(g) Osoita että $\mathbf{1}_n \in \mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{Y})$ ja että on mahdollista että $\dim \mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{Y}) > 1$.

(h) Osoita:

$$\mathrm{r}(\mathbf{Y}'\mathbf{C}\mathbf{Y}) = \mathrm{r}(\mathbf{C}\mathbf{Y}) = c - 1 \quad \text{ja} \quad \mathrm{r}(\mathbf{X}'\mathbf{C}\mathbf{X}) = \mathrm{r}(\mathbf{C}\mathbf{X}) = r - 1\,,$$

missä c ja r viittaavat muuttujien y ja x eri kategorioiden lukumääriin; tässä tilanteessa tietenkin $c = r = 2$.

(i) Osoita, että $(\mathbf{Y}'\mathbf{Y})^{-1}$ on matriisin $\mathbf{Y}'\mathbf{C}\mathbf{Y}$ yleistetty käänteismatriisi, ts.

$$\mathbf{Y}'\mathbf{C}\mathbf{Y} \cdot (\mathbf{Y}'\mathbf{Y})^{-1} \cdot \mathbf{Y}'\mathbf{C}\mathbf{Y} = \mathbf{Y}'\mathbf{C}\mathbf{Y}\,,$$
$$(\mathbf{D_c} - \tfrac{1}{n}\mathbf{cc}') \cdot \mathbf{D_c}^{-1} \cdot (\mathbf{D_c} - \tfrac{1}{n}\mathbf{cc}') = \mathbf{D_c} - \tfrac{1}{n}\mathbf{cc}'.$$

(j) Mikä on matriisin $\mathbf{G}$ tulkinta:

$$\mathbf{G} = \sqrt{n}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'\mathbf{C}\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1/2} = \sqrt{n}\,\mathbf{D_R}^{-1/2}(\mathbf{F} - \mathbf{E})\mathbf{D_c}^{-1/2}\,.$$

1.26 (Taikaneliö). Matriisi

$$\mathbf{A} = \begin{pmatrix} 16 & 3 & 2 & 13 \\ 5 & 10 & 11 & 8 \\ 9 & 6 & 7 & 12 \\ 4 & 15 & 14 & 1 \end{pmatrix}$$

määrittelee taikaneliön ts. $k \times k$-taulukon, jossa sarake-, rivi- sekä kummatkin diagonaalisummat ovat samoja (taikasumma, magic sum), 34 tässä tapauksessa; nyt $k = 4$. Matriisi $\mathbf{A}$ määrittelee klassisen taikaneliön, koska $\mathbf{A}$:n elementit ovat kokonaisluvut $1, 2, \ldots, k^2$. Voimme kutsua $\mathbf{A}$:ta taikamatriisiksi.

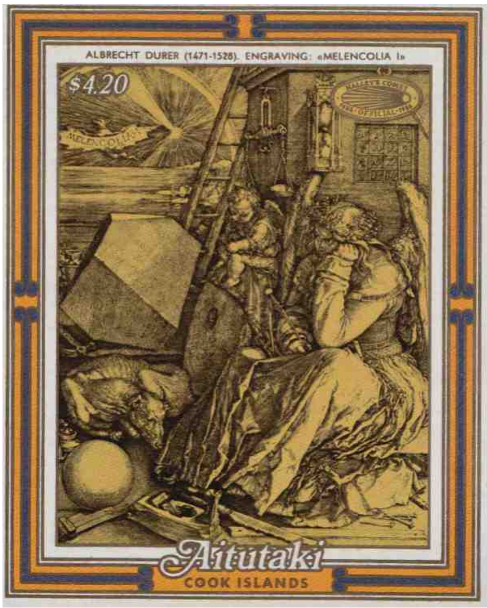
Osoita, että 34 on on yksi $\mathbf{A}$:n ominaisarvo eli on olemassa nollasta poikkeava vektori $\mathbf{t} \in \mathbb{R}^4$ siten että $\mathbf{At} = 34\mathbf{t}$. Vakuuttaudu (vaikkapa tietokoneen avulla) että 0 on myös $\mathbf{A}$:n ominaisarvo.

Osoita, että Moore–Penrose -inverssi

$$\mathbf{A}^+ = \frac{1}{34 \cdot 80} \begin{pmatrix} 275 & -201 & -167 & 173 \\ 37 & -31 & -65 & 139 \\ -99 & 105 & 71 & 3 \\ -133 & 207 & 241 & -235 \end{pmatrix}$$

on myös taikaneliö ja että sen taikasumma on $1/34$.

Trenkler & Trenkler (2001).



Kuvio 1.18. (Ks. tehtävä 1.26.) Matriisin $\mathbf{A}$ määrittelemä taikaneliö esiintyy Albrecht Dürerin kuparikaiverruksessa *Melencolia I*. Postimerkit Aitutakilta (Cook Islands) 1986 ja Mongoliasta 1978.

1.27. Tutustu Vehkalahden & Sundin (2015) artikkeliin Survo-ristikoista.

1.28. Olkoon

$$\mathbf{A} = \begin{pmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 6 & 1 & 8 \\ 7 & 5 & 3 \\ 2 & 9 & 4 \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

(a) Mikä yhteys on matriisien $\mathbf{A}, \mathbf{B}$ ja $\mathbf{F}$ välillä? Mikä on $\mathbf{F}^{-1}$?

(b) Määritä matriisien $\frac{1}{15}\mathbf{A}$, $\frac{1}{15}\mathbf{B}$ ja $\mathbf{F}$ ominaisarvot.

1.29. Matriisia $\mathbf{A}_{n\times n}$ sanotaan stokastiseksi, jos kaikki sen elementit $a_{ij} \geq 0$ ja $\mathbf{A1}_n = \mathbf{1}_n$ ja kaksoisstokastiseksi, jos lisäksi $\mathbf{A}'\mathbf{1}_n = \mathbf{1}_n$. Jos vielä lisäksi $\mathbf{A}$:n molemmat diagonaalisummat ovat ykkösiä, on $\mathbf{A}$ superstokastinen. Tällöin tietenkin $\frac{1}{m}\mathbf{M}$ on superstokastinen, jos $\mathbf{M}$ on taikamatriisi ja m vastaava taikasumma. Kokeile tietokoneella mitä tapahtuu lausekkeelle $(\frac{1}{m}\mathbf{M})^k$, kun $k = 1, 2, \ldots$

1.30. Selitä muuttujaa y muuttujilla x_1 ja x_2 kun

$$(\mathbf{x}_1 : \mathbf{x}_2 : \mathbf{y}) = \begin{pmatrix} 0.4 & 19.7 & 19.7 \\ 2.8 & 19.1 & 19.3 \\ 4.0 & 18.2 & 18.6 \\ 6.0 & 5.2 & 7.9 \\ 1.1 & 4.3 & 4.4 \\ 2.6 & 9.3 & 9.6 \\ 7.1 & 3.6 & 8.0 \\ 5.3 & 14.8 & 15.7 \\ 9.7 & 11.9 & 15.4 \\ 3.1 & 9.3 & 9.8 \\ 9.9 & 2.8 & 10.3 \\ 5.3 & 9.9 & 11.2 \\ 6.7 & 15.4 & 16.8 \\ 4.3 & 2.7 & 5.1 \\ 6.1 & 10.6 & 12.2 \\ 9.0 & 16.6 & 18.9 \\ 4.2 & 11.4 & 12.2 \\ 4.5 & 18.8 & 19.3 \\ 5.2 & 15.6 & 16.5 \\ 4.3 & 17.9 & 18.4 \end{pmatrix}.$$

Mitäs tuumit siitä, että $y_i^2 = x_i^2 + x_2^2$?

Katso http://andrewgelman.com/2004/11/22/a_linear_regres/

Gelman (2004).

Luku 2

Peruslaskutoimituksia ja geometrisia tarkasteluja

Jalassani on kuolio. Näinkö kauhealla tavalla menetän hyvän jalkani, vaikka se on tähän asti uskollisesti palvellut minua vaeltavan elämäni aikana vain harvoin hervoten altani, milloin saatana on vietellyt minut nauttimaan liiaksi viiniä.

Mika Waltari (1949): *Mikael Hakim.*

Käsittelemme tässä ja seuraavassa luvussa matriisilaskennan perusoperaatioita: summaa ja kertolaskua. Näitä on jo käsitelty luvussa 1 mutta tehdään se nyt hieman perustellisemmin. Matriisien $\mathbf{A}_{n\times m}$ ja $\mathbf{B}_{n\times m}$ *summa* määritellään siten, että se on matriisi, joka saadaan laskemalla $\mathbf{A}$:n ja $\mathbf{B}$:n ”päällekkäiset” elementit yhteen. Siten jos $\mathbf{C} = \mathbf{A} + \mathbf{B}$, niin $c_{ij} = a_{ij} + b_{ij}$. Erityisesti jos $\mathbf{a}$ ja $\mathbf{b}$ ovat n elementin pystyvektoreita, niin niiden summa $\mathbf{c}$ on

$$\mathbf{a} + \mathbf{b} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \mathbf{c}\,. \tag{2.1}$$

Seuraavat yhteenlaskun ominaisuudet on helppo vahvistaa:

Y1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}\,,$

Y2. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})\,,$

Y3. $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B},\ \lambda \in \mathbb{R}\,,$

Y4. $(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}'$.

Olkoon $\mathbf{A} \in \mathbb{R}^{n\times m}$ ja $\mathbf{B} \in \mathbb{R}^{m\times p}$, ja olkoon $\mathbf{C}$ matriisien $\mathbf{A}$ ja $\mathbf{B}$ *tulo*:

$$\mathbf{A}_{n\times m}\mathbf{B}_{m\times p} = \mathbf{C}_{n\times p}\,. \tag{2.2}$$

Tällöin $\mathbf{C}$:n elementit lasketaan seuraavasti:

$$c_{ij} = \sum_{k=1}^{m} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj}\,. \tag{2.3}$$

Jos $\mathbf{A}$ on ositettu vaakariveittäin ja $\mathbf{B}$ sarakkeittain, niin

$$\mathbf{AB} = \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix} (\mathbf{b}_1 : \mathbf{b}_2 : \ldots : \mathbf{b}_p) = \begin{pmatrix} \mathbf{a}'_{(1)}\mathbf{b}_1 & \mathbf{a}'_{(1)}\mathbf{b}_2 & \ldots & \mathbf{a}'_{(1)}\mathbf{b}_p \\ \mathbf{a}'_{(2)}\mathbf{b}_1 & \mathbf{a}'_{(2)}\mathbf{b}_2 & \ldots & \mathbf{a}'_{(2)}\mathbf{b}_p \\ \vdots & \vdots & & \vdots \\ \mathbf{a}'_{(n)}\mathbf{b}_1 & \mathbf{a}'_{(n)}\mathbf{b}_2 & \ldots & \mathbf{a}'_{(n)}\mathbf{b}_p \end{pmatrix} \tag{2.4}$$

eli tulon $\mathbf{AB}$ elementit ovat $\mathbf{A}$:n vaakarivien transpoosien $\mathbf{a}_{(i)}$ ja $\mathbf{B}$:n sarakkeiden $\mathbf{b}_j$ sisätuloja:

$$\boxed{\mathbf{A}_{n\times m}\mathbf{B}_{m\times p} = \{\mathbf{a}'_{(i)}\mathbf{b}_j\} = \{\langle \mathbf{a}_{(i)}, \mathbf{b}_j\rangle\} = \{c_{ij}\} \in \mathbb{R}^{n\times p}. \qquad (2.5)}$$

Huomaa, että tulossa $\mathbf{A}_{n\times m}\mathbf{B}_{m\times p}$ on ”keskimmäisten” matriisidimensioiden (m) oltava samat.

Tulon $\mathbf{A}'\mathbf{A}$ elementit ovat matriisin $\mathbf{A}_{n\times m}$ sarakkeiden sisätuloja:

$$\mathbf{A}'\mathbf{A} = \{\mathbf{a}'_i\mathbf{a}_j\} = \{\langle \mathbf{a}_i, \mathbf{a}_j\rangle\} \in \mathbb{R}^{m\times m}. \tag{2.6}$$

Pystyvektorin $\mathbf{a} \in \mathbb{R}^n$ ja vaakavektorin $\mathbf{b}'$ ($\mathbf{b} \in \mathbb{R}^m$) tulo on

$$\mathbf{ab}' = \begin{pmatrix} a_1b_1 & a_1b_2 & \ldots & a_1b_m \\ a_2b_1 & a_2b_2 & \ldots & a_2b_m \\ \vdots & \vdots & & \vdots \\ a_nb_1 & a_nb_2 & \ldots & a_nb_m \end{pmatrix} = \{a_ib_j\} \in \mathbb{R}^{n\times m}. \tag{2.7}$$

Täten esimerkiksi $\mathbf{1}_n\mathbf{1}'_n = \{1\} \in \mathbb{R}^{n\times n}$. Kertolaskun ominaisuuksiin palaamme yksityiskohtaisemmin luvussa 3, mutta sitä ennen on vielä syytä mainita seuraava tärkeä kertolaskun ominaisuus (todistus HT):

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'. \tag{2.8}$$

2.1 Geometrinen tulkinta

Vektoreiden yhteenlaskulle on erittäin hyödyllistä antaa geometrinen tulkinta. Olkoon esimerkiksi

$$\mathbf{a} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{c} = \mathbf{a} + \mathbf{b} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}. \tag{2.9}$$

Tällöin kuvio 2.1 (s. 78) kertoo vektorisumman geometrisesta tulkinnasta, mikä on varmaankin lukijattarelle entuudestaan tuttua. Vakiolla kertomisen eli säännön

$$\lambda\mathbf{a} = \lambda \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{pmatrix} \tag{2.10}$$

geometrinen merkitys selviää kuviosta 2.3 (s. 78). Vakiolla $\lambda \neq 0$ kertominen antaa tulokseksi alkuperäisen vektorin $\mathbf{a}$ monikerran eli vektorin, joka on $\mathbf{a}$:n suuntainen (tai vastakkaissuuntainen jos $\lambda < 0$) mutta pituudeltaan erilainen (kun $|\lambda| \neq 1$) vektori. Vektorin $\mathbf{a}$ pituus eli (euklidinen) normi määriteltiin jo johdanto-luvussa lausekkeena

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}. \tag{2.11}$$

Vektorien erotuksen $\mathbf{a}-\mathbf{b}$ ja etäisyyden $\mathrm{d}(\mathbf{a},\mathbf{b}) = \|\mathbf{a}-\mathbf{b}\|$ geometrinen tulkinta on esitetty kuviossa 2.2 (s. 78).

2.2 Vektorien lineaarikombinaatiot & sarakeavaruus

On helppo päätellä, että mikä hyvänsä $\mathbb{R}^2$:n vektori voidaan ilmaista (2.9):n vektorien $\mathbf{a}$ ja $\mathbf{b}$ lineaarikombinaationa ts. jos $\mathbf{f} \in \mathbb{R}^2$, niin on olemassa sellaiset reaaliluvut α ja β, että

$$\mathbf{f} = \alpha\mathbf{a} + \beta\mathbf{b}\,. \tag{2.12}$$

Vektorien $\mathbf{a}$ ja $\mathbf{b}$ kaikkien mahdollisten lineaarikombinaatioiden joukkoa sanotaan matriisin $\mathbf{A}_{2\times 2} = (\mathbf{a} : \mathbf{b})$ *sarakeavaruudeksi* ja siitä käytetään merkintää $\mathscr{C}(\mathbf{A})$:

$$\begin{aligned} \mathscr{C}(\mathbf{A}) &= \text{matriisin } \mathbf{A} = (\mathbf{a} : \mathbf{b}) \text{ sarakeavaruus} \\ &= \{\, \mathbf{f} \in \mathbb{R}^2 : \mathbf{f} = \alpha\mathbf{a} + \beta\mathbf{b},\ \alpha,\beta \in \mathbb{R}\,\}\,. \end{aligned} \tag{2.13}$$

Sanomme, että $\mathbf{a}$ ja $\mathbf{b}$ *virittävät* $\mathscr{C}(\mathbf{A})$:n. Kuvion 2.1 (s. 78) tilanteessa siis $\mathscr{C}(\mathbf{A}) = \mathbb{R}^2$.

Kuvion 2.4 (s. 79) tapauksessa vektorit $\mathbf{x}$ ja $\mathbf{y}$ virittävät origon kautta kulkevan tason $\mathscr{C}(\mathbf{x} : \mathbf{y})$. Jos $\mathbf{x}$ olisikin $\mathbf{y}$:n monikerta, niin $\mathscr{C}(\mathbf{x} : \mathbf{y})$ olisi origon kautta kulkeva suora. Tietenkin jos $\mathbf{x} = \mathbf{y} = \mathbf{0}$, niin $\mathscr{C}(\mathbf{x} : \mathbf{y}) = \{\mathbf{0}\}$ eli pelkkä origo. Sarakeavaruus $\mathscr{C}(\mathbf{x} : \mathbf{y}) := \mathcal{A}$ on vektoriavaruuden $\mathbb{R}^3$ *aliavaruus* sillä se toteuttaa ehdot:

AA1. $\mathbf{u}, \mathbf{v} \in \mathcal{A} \implies \mathbf{u} + \mathbf{v} \in \mathcal{A}\,,$

AA2. $\lambda \in \mathbb{R},\ \mathbf{u} \in \mathcal{A} \implies \lambda\mathbf{u} \in \mathcal{A}\,,$

AA3. $\mathbf{0} \in \mathcal{A}\,.$

Vektorit $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ovat *ortogonaalisia* eli keskenään kohtisuorassa jos

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y} = 0, \quad \text{jota merkitään } \mathbf{x} \perp \mathbf{y}\,. \tag{2.14}$$

Kannattaa panna merkille, että kohtisuoruus riippuu annetusta sisätulosta; (2.14):ssa on kyseessä tavanomainen sisätulo. Jos $\mathcal{U}$ on jokin $\mathbb{R}^n$:n aliavaruus, niin $\mathcal{U}$:n ortogonaalikomplementti $\mathcal{U}^\perp$ on kaikkien niiden $\mathbb{R}^n$:n vektoreitten joukko, jotka ovat kohtisuorassa jokaisen $\mathcal{U}$:n vektoreiden kanssa:

$$\mathcal{U}^\perp = \{\, \mathbf{z} \in \mathbb{R}^n : \mathbf{z}'\mathbf{x} = 0 \text{ kaikilla } \mathbf{x} \in \mathcal{U} \,\}\,. \tag{2.15}$$

Jos $\mathcal{U}$ ja $\mathcal{V}$ ovat $\mathbb{R}^n$:n aliavaruuksia, niin niiden summa $\mathcal{Z}$ on seuraava aliavaruus:

$$\mathcal{Z} = \mathcal{U} + \mathcal{V} = \{\, \mathbf{z} : \mathbf{z} = \mathbf{u} + \mathbf{v},\ \mathbf{u} \in \mathcal{U}, \mathbf{v} \in \mathcal{V} \,\}\,. \tag{2.16}$$

Erityisesti jos $\mathcal{U} \cap \mathcal{V} = \{\mathbf{0}\}$, merkitään

$$\mathcal{Z} = \mathcal{U} \oplus \mathcal{V} = \mathcal{U}\text{:n ja } \mathcal{V}\text{:n } \textit{suora summa.} \tag{2.17}$$

Jos lisäksi $\mathcal{U}$ ja $\mathcal{V}$ ovat kohtisuorassa, käytämme merkintää $\mathcal{Z} = \mathcal{U} \boxplus \mathcal{V}$. Lukija varmaan allekirjoittaa seuraavan faktan:

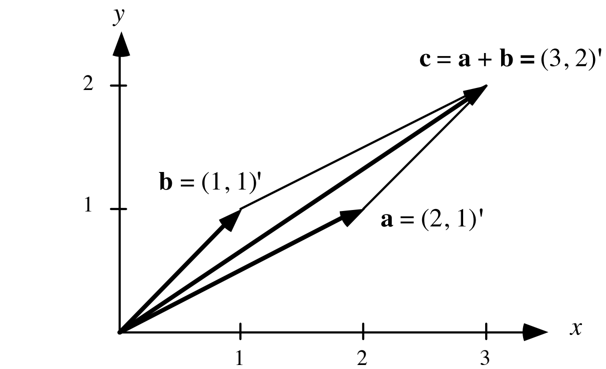
$$\mathbb{R}^n = \mathcal{U} \boxplus \mathcal{U}^\perp. \tag{2.18}$$

2.2.1 Matriisin ja pystyvektorin tulo

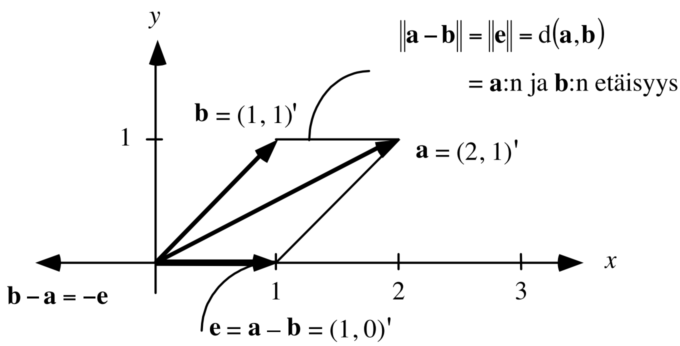
Matriisin $\mathbf{A}_{n\times 2} = (\mathbf{a}_1 : \mathbf{a}_2)$ sarakkeiden lineaarikombinaatio $\mathbf{y} = x_1\mathbf{a}_1 + x_1\mathbf{a}_2$ voidaan esittää matriisin $\mathbf{A}$ ja pystyvektorin $\mathbf{x} = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$ *matriisitulona*:

$$\mathbf{y} = \mathbf{A}\mathbf{x} = (\mathbf{a}_1 : \mathbf{a}_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \begin{pmatrix} x_1 a_{11} \\ x_1 a_{21} \\ \vdots \\ x_1 a_{n1} \end{pmatrix} + \begin{pmatrix} x_2 a_{12} \\ x_2 a_{22} \\ \vdots \\ x_2 a_{n2} \end{pmatrix}. \tag{2.19}$$

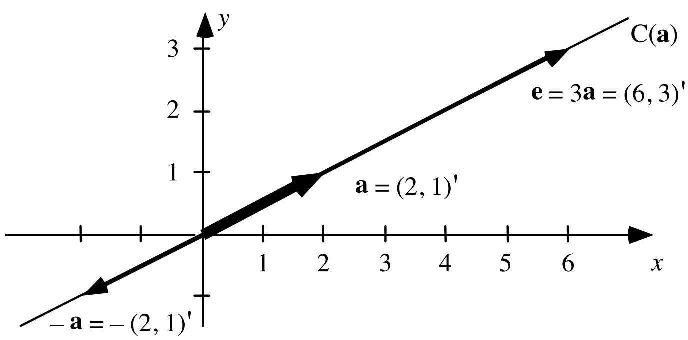
Matriisin $\mathbf{A}$ ja pystyvektorin $\mathbf{x}$ tulo $\mathbf{A}\mathbf{x}$ tarkoittaa siis sellaista $\mathbf{A}$:n sarakkeiden lineaarikombinaatiota, jossa $\mathbf{A}$:n sarakkeiden kertoimina ovat $\mathbf{x}$:n elementit.



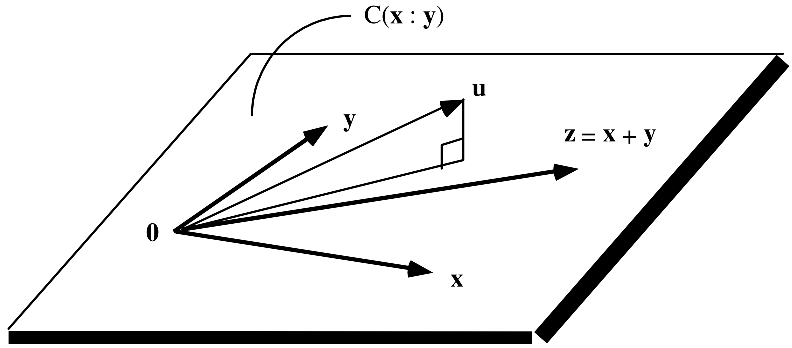
Kuvio 2.1. Vektoreiden $\mathbf{a}$ ja $\mathbf{b}$ summa $\mathbf{a}+\mathbf{b}=\mathbf{c}$ geometrisesti.



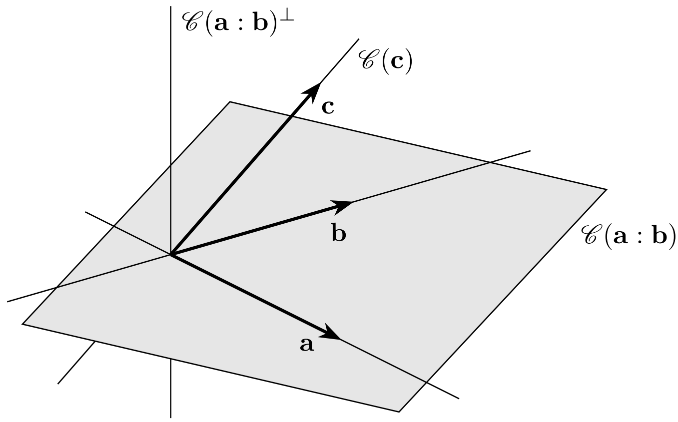
Kuvio 2.2. Vektoreiden $\mathbf{a}$ ja $\mathbf{b}$ erotuksen $\mathbf{e}=\mathbf{a}-\mathbf{b}$ ja etäisyyden $\mathrm{d}(\mathbf{a},\mathbf{b}) = \|\mathbf{e}\| = \|\mathbf{a}-\mathbf{b}\|$ geometrinen tulkinta.



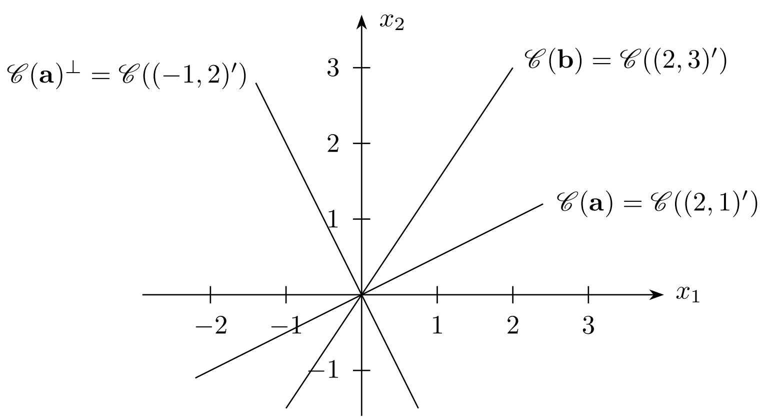
Kuvio 2.3. Vektorin $\mathbf{a}$ monikerran $3\mathbf{a}$ geometrinen tulkinta. Kuviossa myös vektori $-\mathbf{a}$. $\mathscr{C}(\mathbf{a})$ on $\mathbf{a}$:n virittämä suora: $\mathscr{C}(\mathbf{a}) = \{\, \lambda\mathbf{a} : \lambda \in \mathbb{R} \,\}$.



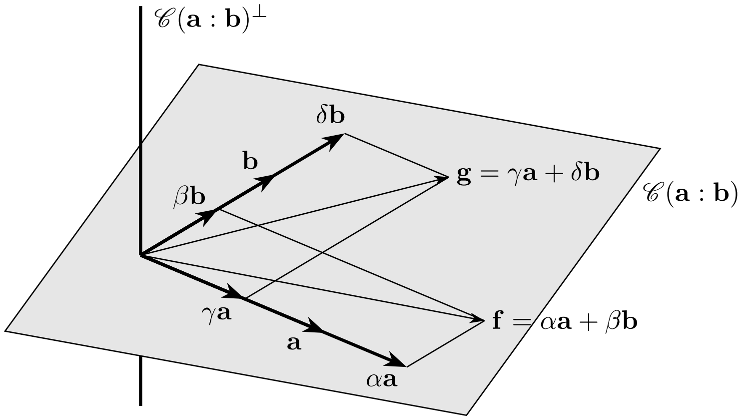
Kuvio 2.4. Vektorien $\mathbf{x}$ ja $\mathbf{y}$ summa $\mathbf{x} + \mathbf{y} = \mathbf{z}$ geometrisesti; kaikki kolme vektoria kuuluvat samaan tasoon $\mathscr{C}(\mathbf{x} : \mathbf{y})$. Vektori $\mathbf{u}$ sen sijaan on hiukan irti tästä tasosta.



Kuvio 2.5. Vektorit $\mathbf{a}$ ja $\mathbf{b}$ virittävät tason $\mathscr{C}(\mathbf{a} : \mathbf{b})$; $\mathbf{c} \notin \mathscr{C}(\mathbf{a} : \mathbf{b})$.



Kuvio 2.6. Tässä $\mathbb{R}^2 = \mathscr{C}(\mathbf{a}) \oplus \mathscr{C}(\mathbf{b})$, $\mathbb{R}^2 = \mathscr{C}(\mathbf{a}) \boxplus \mathscr{C}(\mathbf{a})^\perp$.



Kuvio 2.7. Vektorit $\mathbf{f} = \alpha\mathbf{a} + \beta\mathbf{b}$ ja $\mathbf{g} = \gamma\mathbf{a} + \delta\mathbf{b}$ geometrisesti, $\alpha > 1$, $\beta < 1$, $\gamma < 1$, $\delta > 1$.

Voimme myös ajatella matriisituloa $\mathbf{y} = \mathbf{Ax}$ *lineaarisena kuvauksena*, jossa vektori $\mathbf{x} \in \mathbb{R}^2$ kuvautuu vektoriksi $\mathbf{y} \in \mathbb{R}^n$. Sanonta ”vektorille $\mathbf{x}$ on tehty *lineaarinen muunnos*” tarkoittaa juuri kertolaskua $\mathbf{Ax}$.

Jos ositamme $\mathbf{A}$:n vaakariveittäin, voimme esittää (2.19):n muodossa

$$\mathbf{y} = \mathbf{Ax} = \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{a}'_{(1)}\mathbf{x} \\ \mathbf{a}'_{(2)}\mathbf{x} \\ \vdots \\ \mathbf{a}'_{(n)}\mathbf{x} \end{pmatrix}. \tag{2.20}$$

Matriisitulon merkintöjä käyttäen matriisin $\mathbf{A}_{n\times 2}$ sarakeavaruus on tällöin:

$$\begin{aligned} \mathscr{C}(\mathbf{A}) &= \text{matriisin } \mathbf{A}_{n\times 2} = (\mathbf{a}_1 : \mathbf{a}_2) \text{ sarakeavaruus} \\ &= \{\, \mathbf{y} \in \mathbb{R}^n : \exists\, \mathbf{x} \in \mathbb{R}^2 \text{ s.e. } \mathbf{y} = \mathbf{Ax} \,\} \\ &= \{\, \mathbf{Ax} : \mathbf{x} \in \mathbb{R}^2 \,\}. \end{aligned} \tag{2.21}$$

On selvää, että jos tarkastellaan vektoreiden $\mathbf{a}_1$ ja $\mathbf{a}_2$ kaikkien lineaarikombinaatioiden joukkoa $\mathscr{C}(\mathbf{a}_1 : \mathbf{a}_2)$, niin tämä joukko muodostaa geometrisesti joko

(i) origon kautta kulkevan tason $[\dim \mathscr{C}(\mathbf{A}) = 2 = \operatorname{rank}(\mathbf{A})]$ tai

(ii) origon kautta kulkevan suoran $[\dim \mathscr{C}(\mathbf{A}) = 1 = \operatorname{rank}(\mathbf{A})]$ tai

(iii) pelkän origon $[\dim \mathscr{C}(\mathbf{A}) = 0 = \operatorname{rank}(\mathbf{A})]$.

Tapauksessa (iii) sekä $\mathbf{a}_1$ että $\mathbf{a}_2$ ovat nollavektoreita.

Matriisin $\mathbf{A}_{n\times 2} = (\mathbf{a}_1 : \mathbf{a}_2)$ *aste*, josta käytämme merkintää $\mathrm{r}(\mathbf{A})$ tai $\operatorname{rank}(\mathbf{A})$, on joko 2, 1 tai 0. Aste on

- 0 täsmälleen silloin kun $\mathbf{A} = \mathbf{0}$ eli $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{0}$,
- 1, mikäli $\mathbf{a}_1$ on $\mathbf{a}_2$:n monikerta eli on olemassa sellainen $\alpha \in \mathbb{R}$, että $\mathbf{a}_1 = \alpha\mathbf{a}_2$ (ja ainakin toinen vektoreista $\mathbf{a}_1$ ja $\mathbf{a}_2$ eroaa nollavektorista),
- 2, jos $\mathbf{a}_1$ ei ole $\mathbf{a}_2$:n monikerta (ja kumpikin $\neq \mathbf{0}$).

Jos $\mathbf{A}$:n aste on sama kuin sen sarakkeiden lukumäärä sanomme, että $\mathbf{A}$:n sarakkeet ovat *vapaat* eli *lineaarisesti riippumattomat*; ks. §2.3 (s. 86). Voimme myös sanoa, että tällöin

> $\mathbf{A}$:lla on *täysi sarakeaste*.

Yleisesti:

> $\mathbf{A}$:n aste määritellään $\mathbf{A}$:n lineaarisesti riippumattomien sarakkeiden maksimaalisena lukumääränä; tämä osoittautuu samaksi kuin lineaarisesti riippumattomien vaakarivien maksimaalinen lukumäärä.

Jos $\mathbf{A}$ on $n \times m$-matriisi ja $\mathbf{x} \in \mathbb{R}^m$ niin

$$\begin{aligned} \mathbf{Ax} &= (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \\ &= \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \cdots + \mathbf{a}_m x_m := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{y} \in \mathbb{R}^n. \end{aligned} \tag{2.22}$$

Täten siis:

> $\mathbf{Ax}$ on matriisin $\mathbf{A}$ sarakkeiden lineaarikombinaatio.

Näillä merkinnöillä voimme esittää yleisessä tapauksessa matriisin $\mathbf{A}_{n\times m}$ sarakeavaruuden määritelmän seuraavasti:

$$\begin{aligned} \mathscr{C}(\mathbf{A}) &= \text{matriisin } \mathbf{A}_{n\times m} = (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) \text{ sarakeavaruus} \\ &= \text{matriisin } \mathbf{A}_{n\times m} \text{ sarakkeiden lineaarikombinaatioiden joukko} \\ &= \{\, \mathbf{y} \in \mathbb{R}^n : \exists\, \mathbf{x} \in \mathbb{R}^m \text{ s.e. } \mathbf{y} = \mathbf{Ax} = \mathbf{a}_1 x_1 + \cdots + \mathbf{a}_m x_m \,\} \\ &= \{\, \mathbf{Ax} : \mathbf{x} \in \mathbb{R}^m \,\}. \end{aligned} \tag{2.23}$$

2.2.2 Vaakavektorin ja matriisin tulo

Vaakavektorin $\mathbf{x}'$ ja matriisin $\mathbf{A}_{n\times m}$ tulo on

$$\mathbf{y}' = \mathbf{x}'\mathbf{A} = (x_1, x_2, \dots, x_n) \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix} = x_1\mathbf{a}'_{(1)} + x_2\mathbf{a}'_{(2)} \cdots + x_n\mathbf{a}'_{(n)}\,, \tag{2.24}$$

eli $\mathbf{y}' = \mathbf{x}'\mathbf{A}$ on $\mathbf{A}$:n vaakarivien lineaarikombinaatio ja on siis tietenkin vaakavektori. Esimerkiksi $\mathbf{A}$:n 1. vaakarivi saadaan seuraavasti:

$$\mathbf{a}'_{(1)} = \mathbf{i}'_1\mathbf{A} = (1, 0, \dots, 0) \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix}. \tag{2.25}$$

2.2.3 Esimerkkejä

Esimerkiksi matriisin $\mathbf{A}_{n\times m}$ kaikkien sarakkeiden summa on

$$(\mathbf{a}_1 : \mathbf{a}_2 : \dots : \mathbf{a}_m) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \mathbf{A}\mathbf{1}_m = \mathbf{a}_1 + \mathbf{a}_2 + \dots + \mathbf{a}_m := \mathbf{s} \in \mathbb{R}^n, \tag{2.26}$$

mistä seuraa, että matriisin $\mathbf{A}_{n\times m}$ kaikkien elementtien summa on

$$\mathbf{1}'_n\mathbf{s} = \mathbf{1}'_n\mathbf{A}\mathbf{1}_m = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}\,. \tag{2.27}$$

Matriisin $\mathbf{A}_{n\times m}$ vaakarivien summa on vaakavektori

$$\mathbf{1}'_n\mathbf{A} = (1, 1, \dots, 1) \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix} = \mathbf{a}'_{(1)} + \mathbf{a}'_{(2)} \cdots + \mathbf{a}'_{(n)} \in \mathbb{R}^{1\times m}. \tag{2.28}$$

Neliömatriisin $\mathbf{A}_{n\times n}$ elementtien keskiarvo $\bar{a}$ voidaan matriisimerkinnöin kirjoittaa muodossa

$$\bar{a} = \frac{\mathbf{1}'_n\mathbf{A}\mathbf{1}_n}{(\mathbf{1}'_n\mathbf{1}_n)^2} = \frac{\mathbf{1}'_n\mathbf{A}\mathbf{1}_n}{n^2}\,. \tag{2.29}$$

Luvussa 1 (s. 15) käsiteltiin ns. stokastista matriisia $\mathbf{P}_{n\times n} = (\mathbf{p}_1 : \mathbf{p}_2 : \ldots : \mathbf{p}_n)$, jolla on sellainen ominaisuus, että sen sarakkeiden summa on

$$\mathbf{p}_1 + \mathbf{p}_2 + \cdots + \mathbf{p}_n = \begin{pmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{n1} \end{pmatrix} + \begin{pmatrix} p_{12} \\ p_{22} \\ \vdots \\ p_{n2} \end{pmatrix} + \cdots + \begin{pmatrix} p_{1n} \\ p_{2n} \\ \vdots \\ p_{nn} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \mathbf{1}_n\,. \tag{2.30a}$$

Yhtälön (2.30a) voimme esittää lyhyesti muodossa

$$\mathbf{P1}_n = \mathbf{1}_n\,. \tag{2.30b}$$

Matriisin $\mathbf{A}$ ensimmäinen sarake $\mathbf{a}_1$ saadaan kertolaskulla

$$(\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{Ai}_1 = \mathbf{a}_1\,. \tag{2.31}$$

Matriisin $\mathbf{A}$ 1. ja 2. sarakkeen erotus on

$$(\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{a}_1 - \mathbf{a}_2 = \mathbf{Ai}_1 - \mathbf{Ai}_2\,. \tag{2.32}$$

Jos $\mathbf{U}_{n\times p}$ on havaintomatriisi, jossa on siis n havaintoa p:stä muuttujasta:

$$\mathbf{U} = (\mathbf{u}_1 : \mathbf{u}_2 : \ldots : \mathbf{u}_p) = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix}, \quad \mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)})\,, \tag{2.33}$$

niin sellainen pystyvektori $\bar{\mathbf{u}} = (\bar{u}_1, \ldots, \bar{u}_p)'$, joka sisältää kaikkien muuttujien keskiarvot, saadaan jakamalla $\mathbf{U}$:n havaintovektoreiden $\mathbf{u}_{(i)}$ summa n:llä:

$$\begin{aligned} \bar{\mathbf{u}} = \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \vdots \\ \bar{u}_p \end{pmatrix} &= \frac{1}{n}(\mathbf{u}_{(1)} + \mathbf{u}_{(2)} + \cdots + \mathbf{u}_{(n)}) \\ &= \frac{1}{n}(\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)})\mathbf{1}_n = \frac{1}{n}\mathbf{U}'\mathbf{1}_n \in \mathbb{R}^p. \end{aligned} \tag{2.34}$$

2.2.4 Kertolasku ositetussa muodossa

Jos $\mathbf{A}_{n\times m}$ on ositettu vaakariveittäin ja $\mathbf{B}_{m\times p} = (\mathbf{b}_1 : \mathbf{b}_2 : \ldots : \mathbf{b}_p)$ sarakkeittain ja $\mathbf{AB} = \mathbf{C}_{n\times p} = (\mathbf{c}_1 : \mathbf{c}_2 : \ldots : \mathbf{c}_p)$, niin

$$\mathbf{Ab}_i = \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix} \mathbf{b}_i = \begin{pmatrix} \mathbf{a}'_{(1)}\mathbf{b}_i \\ \mathbf{a}'_{(2)}\mathbf{b}_i \\ \vdots \\ \mathbf{a}'_{(n)}\mathbf{b}_i \end{pmatrix} = \mathbf{c}_i\,, \tag{2.35}$$

mistä saamme välittömästi kertolaskun tärkeän ominaisuuden:

$$\mathbf{AB} = \mathbf{A}(\mathbf{b}_1 : \mathbf{b}_2 : \ldots : \mathbf{b}_p) = (\mathbf{Ab}_1 : \mathbf{Ab}_2 : \ldots : \mathbf{Ab}_p)\,. \tag{2.36}$$

2.2.5 Muuttujien lineaarikombinaatio

Tarkastellaan muuttujien x ja y havaintomatriisia $\mathbf{U}_{n\times 2}$:

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix}. \tag{2.37}$$

Voimme ajatella, että x ja y ovat satunnaismuuttujia, joiden muodostama satunnaisvektori on

$$\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2.38}$$

ja että $\mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)})$ on satunnaisotos populaatiosta, jonka odotusarvo on $\boldsymbol{\mu}$ ja kovarianssimatriisi on $\boldsymbol{\Sigma}$.

Olkoot uudet satunnaismuuttujat $\underline{x}$ ja $\underline{y}$ satunnaismuuttujien x ja y lineaarikombinaatioita:

$$\underline{x} = a_1x + a_2y\,, \quad \underline{y} = b_1x + b_2y\,, \tag{2.39}$$

eli

$$\underline{x} = (a_1, a_2)\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{a}'\mathbf{u}\,, \quad \underline{y} = (b_1, b_2)\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{b}'\mathbf{u}\,. \tag{2.40}$$

Kun merkitsemme $\underline{\mathbf{u}}$:lla muuttujien $\underline{x}$ ja $\underline{y}$ muodostamaa pystyvektoria, saamme yhtälön

$$\underline{\mathbf{u}} = \begin{pmatrix} \underline{x} \\ \underline{y} \end{pmatrix} = \begin{pmatrix} \mathbf{a}'\mathbf{u} \\ \mathbf{b}'\mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{a}' \\ \mathbf{b}' \end{pmatrix}\mathbf{u} = \mathbf{A}'\mathbf{u}\,, \tag{2.41}$$

missä siis

$$\mathbf{A} = (\mathbf{a} : \mathbf{b}) = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}, \quad \mathbf{A}' = \begin{pmatrix} \mathbf{a}' \\ \mathbf{b}' \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}. \tag{2.42}$$

Satunnaisvektori $\underline{\mathbf{u}}$ saadaan siis tekemällä lineaarinen muunnos satunnaisvektorille $\mathbf{u}$.

Muuttujia x ja y vastaavat empiiriset havaitut muuttujavektorit olkoot $\mathbf{x}$ ja $\mathbf{y}$. Tällöin uusia muuttujia $\underline{x}$ ja $\underline{y}$ vastaavat muuttujavektorit ovat

$$\underline{\mathbf{x}} = a_1\mathbf{x} + a_2\mathbf{y} = (\mathbf{x} : \mathbf{y}) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \mathbf{Ua}\,, \tag{2.43a}$$

$$\underline{\mathbf{y}} = b_1\mathbf{x} + b_2\mathbf{y} = (\mathbf{x} : \mathbf{y}) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{Ub}\,, \tag{2.43b}$$

joten uusi havaintomatriisi $\underline{\mathbf{U}} = (\underline{\mathbf{x}} : \underline{\mathbf{y}})$ on

$$\underline{\mathbf{U}} = (\underline{\mathbf{x}} : \underline{\mathbf{y}}) = (\mathbf{Ua} : \mathbf{Ub}) = \mathbf{U}(\mathbf{a} : \mathbf{b}) = \mathbf{UA}\,. \tag{2.44}$$

Huomattakoon, että (2.44):ssä on käytetty seuraavaa matriisitulon yleistä ominaisuutta (2.36):

$$\mathbf{F}(\mathbf{g}_1 : \mathbf{g}_2 : \ldots : \mathbf{g}_p) = (\mathbf{F}\mathbf{g}_1 : \mathbf{F}\mathbf{g}_2 : \ldots : \mathbf{F}\mathbf{g}_p)\,. \tag{2.45}$$

Tulon transponointisäännön $(\mathbf{FG})' = \mathbf{G}'\mathbf{F}'$ perusteella (2.44):stä seuraa, että $\underline{\mathbf{U}}' = \mathbf{A}'\mathbf{U}'$ eli

$$\begin{aligned}(\underline{\mathbf{u}}_{(1)} : \underline{\mathbf{u}}_{(2)} : \ldots : \underline{\mathbf{u}}_{(n)}) &= \mathbf{A}'(\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)}) \\ &= (\mathbf{A}'\mathbf{u}_{(1)} : \mathbf{A}'\mathbf{u}_{(2)} : \ldots : \mathbf{A}'\mathbf{u}_{(n)})\,. \end{aligned} \tag{2.46}$$

Täten uusien ja vanhojen havaintovektoreiden välillä on yhteys

$$\underline{\mathbf{u}}_{(i)} = \mathbf{A}'\mathbf{u}_{(i)}, \quad i = 1, 2, \ldots, n\,, \tag{2.47}$$

kun taas muuttujavektoreiden välinen yhteys on

$$\underline{\mathbf{x}} = \mathbf{Ua} = (\mathbf{x} : \mathbf{y})\mathbf{a}\,, \quad \underline{\mathbf{y}} = \mathbf{Ub} = (\mathbf{x} : \mathbf{y})\mathbf{b}\,. \tag{2.48}$$

Toistettakoon vielä, että lauseke (2.47) tarkoittaa, että

> havaintovektoreille $\mathbf{u}_{(i)}$ tehdään *lineaarinen muunnos*

eli

> alkuperäisten muuttujien korvaaminen lineaarikombinaatioillaan voidaan tulkita havaintovektoreiden lineaariseksi muunnokseksi.

Samoin on perusteltua (vielä kerran) korostaa havainnoille tehtävän lineaarisen muunnoksen $\underline{\mathbf{u}}_{(i)} = \mathbf{A}'\mathbf{u}_{(i)}$ vaikutusta havaintomatriisiin:

$$\boxed{\underline{\mathbf{u}}_{(i)} = \mathbf{A}'\mathbf{u}_{(i)} \implies \underline{\mathbf{U}}' = \mathbf{A}'\mathbf{U}', \quad \underline{\mathbf{U}} = \mathbf{U}\mathbf{A}. \qquad (2.49)}$$

Kuvioissa 2.8 ja 2.9 on otteita Survon toimituskentästä, jossa

$$\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} := \mathbf{A}' \begin{pmatrix} x \\ y \end{pmatrix} \implies \underline{\mathbf{U}} = \mathbf{U}\mathbf{A}.$$

2.3 Lineaarinen riippumattomuus

On olemassa yksi erityisen mielenkiintoinen $\mathbf{A}$:n sarakkeiden lineaarikombinaatio: nollavektori $\mathbf{0}$. Voimme kysyä mikä on se vektori $\mathbf{x}$, jolla on ominaisuus

$$\mathbf{A}\mathbf{x} = \mathbf{0}\,. \tag{2.50}$$

Jos esim. $\mathbf{A} = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, niin mm. seuraavat vektorit toteuttavat ehdon (2.50):

$$\mathbf{b} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \mathbf{d} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{e} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \tag{2.51}$$

Kaikkien niiden vektoreiden $\mathbf{x}$ joukkoa, jotka toteuttavat ehdon $\mathbf{A}\mathbf{x} = \mathbf{0}$, sanotaan $\mathbf{A}$:n *nolla-avaruudeksi* eli *ytimeksi* ja sitä merkitään $\mathscr{N}(\mathbf{A})$:lla:

$$\mathscr{N}(\mathbf{A}) = \text{matriisin } \mathbf{A}_{n\times m} \text{ nolla-avaruus} = \{\, \mathbf{x} \in \mathbb{R}^m : \mathbf{A}\mathbf{x} = \mathbf{0} \,\}\,. \tag{2.52}$$

On erityisesti huomattava, että

$$\mathscr{C}(\mathbf{A}_{n\times m}) \subset \mathbb{R}^n, \quad \mathscr{N}(\mathbf{A}_{n\times m}) \subset \mathbb{R}^m. \tag{2.53}$$

Olemme jo aikaisemmin käsitelleet ohimennen lineaarisen riippuvuuden käsitettä. Täsmällinen määritelmä on seuraava:

matriisin $\mathbf{A}_{n\times m}$ sarakkeet $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^n$ ovat *lineaarisesti riippumattomat* eli *vapaat*, jos

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m = \mathbf{0} \iff x_1 = x_2 = \cdots = x_m = 0\,. \tag{2.54}$$

Matriisimerkinnöin ehto (2.54) voidaan kirjoittaa

$$\mathbf{A}\mathbf{x} = \mathbf{0} \iff \mathbf{x} = \mathbf{0}\,, \tag{2.55}$$

eli nolla-avaruuden avulla esitettynä

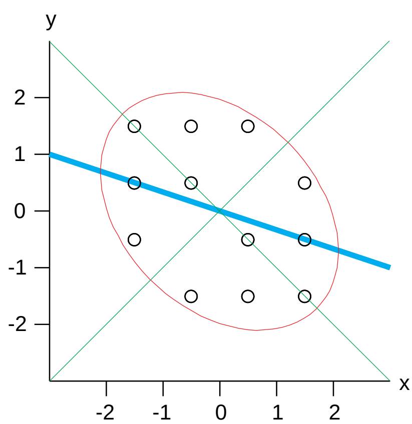
$$\mathbf{A}\text{:n sarakkeet vapaat} \iff \mathscr{N}(\mathbf{A}) = \{\mathbf{0}\}\,. \tag{2.56}$$

```
  -   - SURVO MM  Thu Sep 23 20:20:52 2010            C:\SJP\D\   2000  100 0
*
*DATA KOED,a,b,c,d
d 11            11.1   11.1     11.11111  11.11111
*MATRIX KOE1
c ///            x        y        w         z
a  1            -1.5   -0.5     -0.70711  -1.41421
*  2            -1.5    0.5     -1.41421  -0.70711
*  3            -1.5    1.5     -2.12132   0.00000
*  4            -0.5   -1.5      0.70711  -1.41421
*  5            -0.5    0.5     -0.70711   0.00000
*  6            -0.5    1.5     -1.41421   0.70711
*  7             0.5   -1.5      1.41421  -0.70711
*  8             0.5   -0.5      0.70711   0.00000
*  9             0.5    1.5     -0.70711   1.41421
* 10             1.5   -1.5      2.12132   0.00000
* 11             1.5   -0.5      1.41421   0.70711
b 12             1.5    0.5      0.70711   1.41421
*
*VAR w=q*x-q*y TO KOED   / viedään uudet muutujat KOED:hen
*VAR z=q*x+q*y TO KOED   / q=1/SQRT(2)
*MAT SAVE KOE1           / HUOM: riveillä a-b on sekä MATRIX että DATA
*MATRIX   A  ///
*   q   q
*  -q   q
*
*MAT SAVE A          /  pi=3.1415926535893  cos(pi/4)=0.70710678118663
*MAT LOAD A,CUR+1    /  A on ortogonaalinen matriisi: 45 asteen rotaatio
*MATRIX A
*///           1        2
*  1       0.70711  0.70711
*  2      -0.70711  0.70711
*
*MAT KOE2=KOE1(*,1:2)  / KOE2 muodostuu KOE1:n kahdesta ens. sarakk.
*MAT KOE3!=KOE2*A     / *KOE3~KOE1(*,1:2)*A 12*2
*MAT KOE3(0,1)="w"     / annetaan KOE3:n sarakkeille nimet
*MAT KOE3(0,2)="z"
*MAT LOAD KOE3         / KOE3:ssa ovat samat w ja z KOE1:ssä (ja KOED:ss
*MATRIX KOE3
*///           w        z
*  1      -0.70711 -1.41421
*  2      -1.41421 -0.70711
           ...
* 12       0.70711  1.41421
*                                         (MK10-004, MKT-02-ROT1)
```

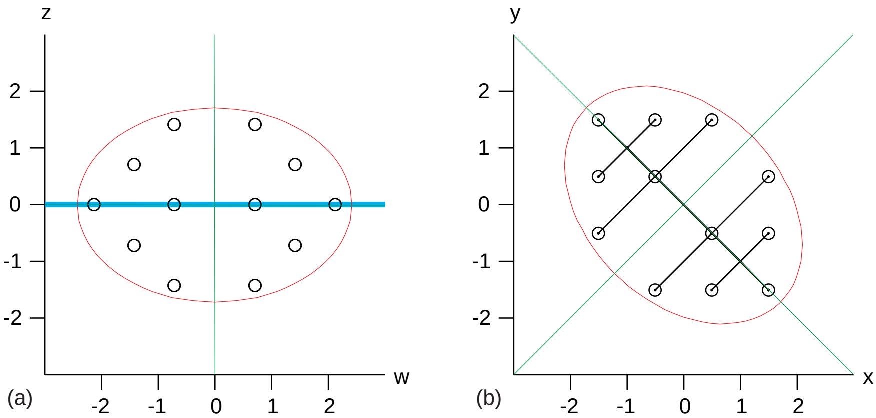
Kuvio 2.8. Toimituskenttä, jossa muodostetaan uusi havaintomatriisi $\underline{\mathbf{U}} = \mathbf{UA}$.

```
 -  - SURVO MM  Thu Sep 23 20:24:33 2010          C:\SJP\D\    2000  100 0
*
*LINREG KOED,CUR+1  / VARS=x(X),y(Y)
*Means, std.devs and correlations of KOED  N=12
*Variable  Mean         Std.dev.
*x         0.000000     1.167748
*y         0.000000     1.167748
*Correlations:
*             x       y
* x            1.0000 -0.3333
* y           -0.3333  1.0000
*
*Linear regression analysis: Data KOED, Regressand y         N=12
*Variable Regr.coeff.    Std.dev.    t      beta
*x        -0.333333       0.298142 -1.118 -0.333
*constant  0.000000       0.333333  0.000
*Variance of regressand y=1.363636364 df=11
*Residual variance=1.333333333 df=10
*R=0.3333 R^2=0.1111
*
*CORR KOED,CUR+1 / VARS=w,z  / uudet muuttujat w ja z korreloimattomia
*Means, std.devs and correlations of KOED  N=12
*Variable  Mean         Std.dev.
*w         0.000000     1.348399
*z         0.000000     0.953462
*Correlations:
*             w       z
* w            1.0000 -0.0000
* z           -0.0000  1.0000
*                                          (MK10-004, MKT-02-ROT1)
```

Kuvio 2.9. Jatkoa kuvioon 2.8.



Kuvio 2.10. Pisteparvi toimituskentän 2.8 aineistosta KOE2.



Kuvio 2.11. (Jatkoa kuvioon 2.10.) (a) Havaintomatriisi KOE2 on kerrottu oikealta rotaatiomatriisilla $\mathbf{A}$.(b) Havaintojen ortogonaaliset etäisyydet suorasta $\mathscr{C}(\mathbf{t}_1)$ on merkitty näkyviin; mikä on $\mathbf{t}_1$?

Vastaavasti määritellään, että vektorit $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ ovat *lineaarisesti riippuvat* eli *sidotut*, jos ne eivät ole vapaat eli

$$\text{on olemassa vektori } \mathbf{x} \neq \mathbf{0} \text{ siten että } \mathbf{Ax} = \mathbf{0}\,, \tag{2.57a}$$

ts.

$$\mathbf{A}\text{:n sarakkeet sidotut} \iff \mathscr{N}(\mathbf{A}) \neq \{\mathbf{0}\}\,. \tag{2.57b}$$

Lineaarisen riippuvuuden käsitteeseen palataan myöhemmin. Esitämme kuitenkin tässä yhteydessä vielä seuraavan hyödyllisen tuloksen:

$$\boxed{\mathscr{N}(\mathbf{A}) = \mathscr{N}(\mathbf{A}'\mathbf{A})\,. \tag{2.58}}$$

Tulos (2.58) voidaan päätellä seuraavasti. Ensinnäkin

$$\mathbf{x} \in \mathscr{N}(\mathbf{A}) \implies \mathbf{Ax} = \mathbf{0} \implies \mathbf{A}'\mathbf{Ax} = \mathbf{0} \implies \mathbf{x} \in \mathscr{N}(\mathbf{A}'\mathbf{A})\,, \tag{2.59}$$

joten $\mathscr{N}(\mathbf{A}) \subset \mathscr{N}(\mathbf{A}'\mathbf{A})$. Toisaalta

$$\begin{aligned} \mathbf{x} \in \mathscr{N}(\mathbf{A}'\mathbf{A}) &\implies \mathbf{A}'\mathbf{Ax} = \mathbf{0} \implies \mathbf{x}'\mathbf{A}'\mathbf{Ax} = 0 \\ &\implies \mathbf{Ax} = \mathbf{0} \implies \mathbf{x} \in \mathscr{N}(\mathbf{A})\,, \end{aligned} \tag{2.60}$$

joten myös $\mathscr{N}(\mathbf{A}'\mathbf{A}) \subset \mathscr{N}(\mathbf{A})$ ja näin (2.58) on todistettu.

Tuloksen (2.58) perusteella voimme tehdä seuraavat päätelmät:

$$\mathbf{A}\text{:n sarakkeet vapaat} \iff \mathbf{A}'\mathbf{A}\text{:n sarakkeet vapaat}, \tag{2.61a}$$

$$\operatorname{rank}(\mathbf{A}_{n\times m}) = m \iff \operatorname{rank}(\mathbf{A}'\mathbf{A}) = m\,. \tag{2.61b}$$

Itse asiassa osoittautuu, että aina on voimassa

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}'\mathbf{A}) = \operatorname{rank}(\mathbf{A}\mathbf{A}')\,. \tag{2.62}$$

Tuloksen (2.62) todistamiseksi tarkastellaan sarakeavaruuden $\mathscr{C}(\mathbf{A}_{n\times m})$ ortogonaalikomplementtia $\mathscr{C}(\mathbf{A})^{\perp}$. Se määritellään kaikkien niiden $\mathbb{R}^n$:n vektoreitten joukkona, jotka ovat kohtisuorassa $\mathscr{C}(\mathbf{A})$:n jokaisen vektorin kanssa ts.

$$\begin{aligned}
\mathscr{C}(\mathbf{A})^{\perp} &= \{\, \mathbf{y} \in \mathbb{R}^n : \mathbf{y}'\mathbf{u} = 0 \text{ kaikilla } \mathbf{u} \in \mathscr{C}(\mathbf{A}) \,\} \\
&= \{\, \mathbf{y} \in \mathbb{R}^n : \mathbf{y}'\mathbf{A}\mathbf{z} = 0 \text{ kaikilla } \mathbf{z} \in \mathbb{R}^m \,\} \\
&= \{\, \mathbf{y} \in \mathbb{R}^n : \mathbf{A}'\mathbf{y} = \mathbf{0} \,\} = \mathscr{N}(\mathbf{A}')\,,
\end{aligned} \tag{2.63}$$

ja näin on johdettu tärkeä tulos

$$\boxed{\mathscr{C}(\mathbf{A})^{\perp} = \mathscr{N}(\mathbf{A}')\,. \tag{2.64}}$$

”Nastoittamalla” (2.58) molemmin puolin saadaan $\mathscr{N}(\mathbf{A})^{\perp} = \mathscr{N}(\mathbf{A}'\mathbf{A})^{\perp}$ eli

$$\mathscr{C}(\mathbf{A}') = \mathscr{C}(\mathbf{A}'\mathbf{A})\,. \tag{2.65}$$

Koska $\operatorname{rank}(\mathbf{U}) = \dim \mathscr{C}(\mathbf{U})$, saadaan (2.62) välittömästi (2.65):sta.

Matriisia $\mathbf{A}^{\perp}$ ovat käsitelleet laajahkosti Markiewicz & Puntanen (2015) artikkelissa *All about the $\perp$ with its applications in the linear statistical models.*

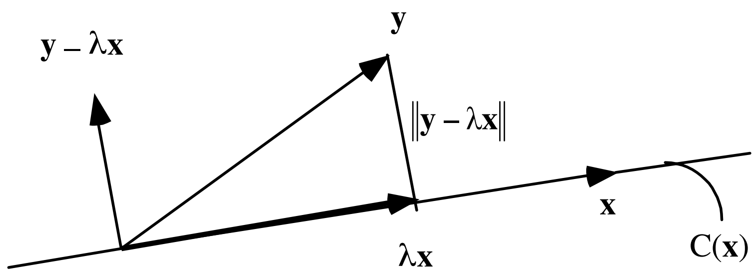
2.3.1 Ortogonaaliprojektio suoralle

Luvussa 1.2 (s. 31) jo tarkasteltiin vektorien välisten kulmien yhteydessä vektorin $\mathbf{y}$ projisointia vektorin $\mathbf{x}$ virittämälle suoralle $\mathscr{C}(\mathbf{x})$. Kuten totesimme, vektorin $\mathbf{y}$ ortogonaaliprojektio $\mathscr{C}(\mathbf{x})$:lle, $\hat{\mathbf{y}}$, on se vektorin $\mathbf{x}$ monikerta [eli $\mathscr{C}(\mathbf{x})$:n vektori], jonka etäisyys $\mathbf{y}$:stä on pienin.

$$\begin{aligned}
\|\mathbf{y} - \hat{\mathbf{y}}\|^2 &= \|\mathbf{y} - \hat{\lambda}\mathbf{x}\|^2 \\
&= \min_{\lambda} \|\mathbf{y} - \lambda\mathbf{x}\|^2 = \min_{\mathbf{u}\in\mathscr{C}(\mathbf{x})} \|\mathbf{y} - \mathbf{u}\|^2 \\
&= \min_{\lambda} \operatorname{SSE}(\mathbf{y}; \lambda\mathbf{x})\,.
\end{aligned} \tag{2.66}$$

Merkintä $\operatorname{SSE}(\mathbf{y}; \lambda\mathbf{x})$ saa selityksensä regressioanalyysistä: $\operatorname{SSE}(\mathbf{y}; \lambda\mathbf{x})$ vastaa jäännösneliösummaa regressiomallissa, kun $\mathbf{y}$:tä selitetään (vain) $\mathbf{x}$:llä. Minimoitava funktio on siis

$$\begin{aligned}
\operatorname{SSE}(\mathbf{y}; \lambda\mathbf{x}) = f(\lambda) &= \|\mathbf{y} - \lambda\mathbf{x}\|^2 = (\mathbf{y} - \lambda\mathbf{x})'(\mathbf{y} - \lambda\mathbf{x}) \\
&= \mathbf{y}'\mathbf{y} - \mathbf{y}'\lambda\mathbf{x} - \lambda\mathbf{x}'\mathbf{y} + \lambda^2\mathbf{x}'\mathbf{x} \\
&= \lambda^2\mathbf{x}'\mathbf{x} - 2\lambda\mathbf{y}'\mathbf{x} + \mathbf{y}'\mathbf{y}\,.
\end{aligned} \tag{2.67}$$



Kuvio 2.12. Vakion λ määrittäminen siten että etäisyys $\|\mathbf{y}-\lambda\mathbf{x}\|$ minimoituu; tuloksena vektorin $\mathbf{y}$ ortogonaaliprojektio suoralle $\mathscr{C}(\mathbf{x})$.

Funktion $f(\lambda)$ kuvaaja on ylöspäin aukeava paraabeli. Kun $f(\lambda)$ derivoidaan λ:n suhteen ja asetetaan derivaatta nollaksi, saadaan yhtälö

$$\mathbf{x}'\mathbf{x}\lambda = \mathbf{x}'\mathbf{y}\,. \tag{2.68}$$

Yhtälö 2.68 on itse asiassa erikoistapaus ns. *normaaliyhtälöstä*, johon päädytään *pienimmän neliösumman menetelmässä*. Funktion $f(\lambda)$ minimi saavutetaan kun λ saa arvon

$$\hat{\lambda} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}\,, \tag{2.69}$$

jolloin (1.75):n (s. 31) tapaan saamme

$$\hat{\mathbf{y}} = \hat{\lambda}\mathbf{x} = \mathbf{x}\hat{\lambda} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} = \mathbf{P}_{\mathbf{x}}\mathbf{y}\,, \tag{2.70}$$

missä neliömatriisi

$$\boxed{\mathbf{P}_{\mathbf{x}} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}' = \frac{1}{\mathbf{x}'\mathbf{x}}\mathbf{x}\mathbf{x}'} \tag{2.71}$$

on *ortogonaaliprojektori* suoralle $\mathscr{C}(\mathbf{x})$. Matriisi $\mathbf{P}_{\mathbf{x}}$ on symmetrinen ja idempotentti:

$$\mathbf{P}_{\mathbf{x}}\mathbf{P}_{\mathbf{x}} = \mathbf{P}_{\mathbf{x}} = \mathbf{P}_{\mathbf{x}}'\,, \tag{2.72}$$

sillä $\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}' = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'$; samoin on

$$(\mathbf{I}-\mathbf{P}_{\mathbf{x}})(\mathbf{I}-\mathbf{P}_{\mathbf{x}}) = \mathbf{I}-\mathbf{P}_{\mathbf{x}} = (\mathbf{I}-\mathbf{P}_{\mathbf{x}})'. \tag{2.73}$$

Minimoitavan funktion arvoksi saadaan

$$\begin{aligned}\mathrm{SSE}(\mathbf{y};\hat{\lambda}\mathbf{x}) &= \|\mathbf{y}-\hat{\mathbf{y}}\|^2 = \|\mathbf{y}-\mathbf{P}_{\mathbf{x}}\mathbf{y}\|^2 = \|(\mathbf{I}-\mathbf{P}_{\mathbf{x}})\mathbf{y}\|^2 \\ &= \mathbf{y}'(\mathbf{I}-\mathbf{P}_{\mathbf{x}})'(\mathbf{I}-\mathbf{P}_{\mathbf{x}})\mathbf{y} \\ &= \mathbf{y}'(\mathbf{I}-\mathbf{P}_{\mathbf{x}})\mathbf{y}\,. \end{aligned} \tag{2.74}$$

Voisimme johtaa $\hat{\lambda}$:n lausekkeen yksinkertaisesti myös siten, että määritämme vektorin $\hat{\lambda}\mathbf{x}$, jolla on sellainen ominaisuus, että $\mathbf{y} - \hat{\lambda}\mathbf{x}$ on kohtisuorassa vektoria $\mathbf{x}$ vastaan [ks. (1.73), s. 31]:

$$\mathbf{x}'(\mathbf{y} - \hat{\lambda}\mathbf{x}) = 0 \quad \text{ts.} \quad \mathbf{x}'\mathbf{x}\hat{\lambda} = \mathbf{x}'\mathbf{y}. \tag{2.75}$$

Ehto (2.75) on tietysti identtinen lausekkeen (2.68) eli "normaaliyhtälön" kanssa.

Täsmällinen todistus sille, että (2.75):n toteuttava $\hat{\lambda}$ todella minimoi jäännösneliösumman $\|\mathbf{y} - \lambda\mathbf{x}\|^2$ on seuraava:

$$\begin{aligned}
\|\mathbf{y} - \lambda\mathbf{x}\|^2 &= \|(\mathbf{y} - \hat{\lambda}\mathbf{x}) - (\lambda\mathbf{x} - \hat{\lambda}\mathbf{x})\|^2 \\
&= \|\mathbf{y} - \hat{\lambda}\mathbf{x}\|^2 + \|\lambda\mathbf{x} - \hat{\lambda}\mathbf{x}\|^2 - 2(\mathbf{y} - \hat{\lambda}\mathbf{x})'(\lambda\mathbf{x} - \hat{\lambda}\mathbf{x}) \\
&= \|\mathbf{y} - \hat{\lambda}\mathbf{x}\|^2 + \|(\lambda - \hat{\lambda})\mathbf{x}\|^2 - 2\,\underbrace{(\mathbf{y} - \hat{\lambda}\mathbf{x})'\mathbf{x}}_{=0}(\lambda - \hat{\lambda}) \\
&= \|\mathbf{y} - \hat{\lambda}\mathbf{x}\|^2 + \|(\lambda - \hat{\lambda})\mathbf{x}\|^2 \\
&\geq \|\mathbf{y} - \hat{\lambda}\mathbf{x}\|^2 \ \forall\, \lambda \in \mathbb{R}\,,
\end{aligned} \tag{2.76}$$

missä yhtäsuuruus on voimassa tietenkin vain jos $\lambda = \hat{\lambda}$.

2.4 Keskiarvovektori projektiona

Mikä on muuttujavektorin $\mathbf{y} \in \mathbb{R}^n$ ortogonaaliprojektio vektorin $\mathbf{1}$ kautta kulkevalle suoralle $\mathscr{C}(\mathbf{1})$? Kaavasta (2.69) seuraa välittömästi, että

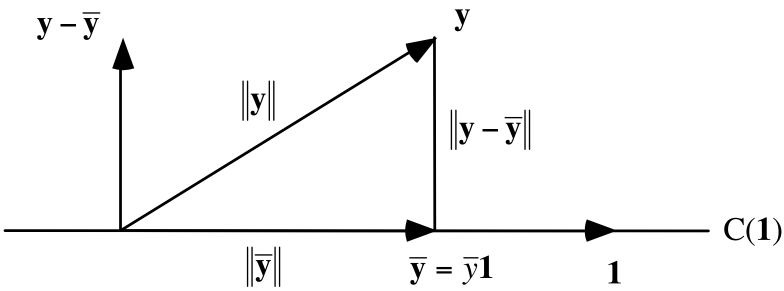
$$\hat{\lambda} = \frac{\mathbf{1}'\mathbf{y}}{\mathbf{1}'\mathbf{1}} = \frac{1}{n}\mathbf{1}'\mathbf{y} = \frac{y_1 + y_2 + \cdots + y_n}{n} = \bar{y}\,, \tag{2.77}$$

joten haettu ortogonaaliprojektio on

$$\hat{\mathbf{y}} = \hat{\lambda}\mathbf{1} = \bar{y}\mathbf{1} = \begin{pmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix} := \bar{\bar{\mathbf{y}}} \in \mathbb{R}^n\,. \tag{2.78}$$

Tämä merkitsee sitä, että lauseke $\sum_{i=1}^n (y_i - \alpha)^2 = \|\mathbf{y} - \alpha\mathbf{1}\|^2$ saavuttaa miniminsä, kun $\alpha = \bar{y}$, ts.

$$\begin{aligned}
\min_\alpha \sum_{i=1}^n (y_i - \alpha)^2 &= \min_\alpha \|\mathbf{y} - \alpha\mathbf{1}\|^2 = \min_\alpha \mathrm{SSE}(\mathbf{y}; \alpha\mathbf{1}) \\
&= \sum_{i=1}^n (y_i - \bar{y})^2 = \mathrm{SSE}(\mathbf{y}; \bar{y}\mathbf{1})\,.
\end{aligned} \tag{2.79}$$



Kuvio 2.13. Vektorin $\mathbf{y}$ projisointi suoralle $\mathscr{C}(\mathbf{1})$. (Kuviossa pitäisi $\bar{\mathbf{y}}$ korvata $\bar{\bar{\mathbf{y}}}$:llä.)

Käytämme merkintää $\bar{\bar{\mathbf{y}}}$, $\bar{\bar{\mathbf{x}}}$ jne vektoreista, joiden jokainen elementti on kyseisen vektorin komponenttien keskiarvo eli vastaavan muuttujan keskiarvo.

Havaitsemme välittömästi, että $\bar{\bar{\mathbf{y}}}$ on esitettävissä matriisitulona

$$\bar{\bar{\mathbf{y}}} = \bar{y}\mathbf{1} = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y} = \mathbf{P_1 y} = \mathbf{J}\mathbf{y}\,, \tag{2.80}$$

missä $\mathbf{J}$ on ortogonaaliprojektori:

$$\mathbf{J} = \frac{1}{n}\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} = \frac{1}{n}\begin{pmatrix} \mathbf{1}' \\ \mathbf{1}' \\ \vdots \\ \mathbf{1}' \end{pmatrix} = \frac{1}{n}\mathbf{1}\mathbf{1}'. \tag{2.81}$$

Koska

$$\mathbf{y} - \bar{\bar{\mathbf{y}}} = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} = \tilde{\mathbf{y}} = \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_n \end{pmatrix} = \text{keskistetty } \mathbf{y}\,, \tag{2.82}$$

niin

$$\mathbf{y} - \bar{\bar{\mathbf{y}}} = \mathbf{y} - \mathbf{J}\mathbf{y} = (\mathbf{I} - \mathbf{J})\mathbf{y} := \mathbf{C}\mathbf{y}\,, \tag{2.83}$$

missä matriisi $\mathbf{C} = \mathbf{I} - \mathbf{J}$ on *keskistäjämatriisi*:

$$\mathbf{C} = \mathbf{I} - \mathbf{J} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}' = \begin{pmatrix} 1 - 1/n & -1/n & \dots & -1/n \\ -1/n & 1 - 1/n & \dots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \dots & 1 - 1/n \end{pmatrix}. \tag{2.84}$$

Sekä $\mathbf{J}$ että $\mathbf{C} = \mathbf{I} - \mathbf{J}$ ovat idempotentteja ja symmetrisiä:

$$\mathbf{J}\mathbf{J} = \mathbf{J} = \mathbf{J}', \quad (\mathbf{I} - \mathbf{J})(\mathbf{I} - \mathbf{J}) = \mathbf{I} - \mathbf{J} = (\mathbf{I} - \mathbf{J})'. \tag{2.85}$$

Tällöin keskistetyn muuttujavektorin pituuden neliö on

$$\|\mathbf{y} - \bar{\bar{\mathbf{y}}}\|^2 = \|\tilde{\mathbf{y}}\|^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}\tilde{y}_i^2 = \mathrm{SS}_y = t_{yy}\,, \tag{2.86a}$$

eli matriisimerkinnöin

$$t_{yy} = \|\mathbf{y} - \mathbf{J}\mathbf{y}\|^2 = \|(\mathbf{I} - \mathbf{J})\mathbf{y}\|^2 = \mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y}\,. \tag{2.86b}$$

Muuttujan y otosvarianssi on suoraan verrannollinen keskistetyn $\mathbf{y}$:n normin neliöön:

$$\mathrm{var}_{\mathrm{d}}(\mathbf{y}) = \frac{1}{n-1}\|\tilde{\mathbf{y}}\|^2 = \frac{1}{n-1}\tilde{\mathbf{y}}'\tilde{\mathbf{y}} = \frac{1}{n-1}\mathbf{y}'\mathbf{C}\mathbf{y}\,. \tag{2.86c}$$

Muuttujien x:n ja y:n välinen tulosumma on

$$t_{xy} = \tilde{\mathbf{x}}'\tilde{\mathbf{y}} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \mathbf{x}'\mathbf{C}\mathbf{y}\,, \tag{2.87}$$

joten kovarianssi on

$$\mathrm{cov}_{\mathrm{d}}(\mathbf{x}, \mathbf{y}) = s_{xy} = \frac{1}{n-1}\tilde{\mathbf{x}}'\tilde{\mathbf{y}} = \frac{1}{n-1}\mathbf{x}'\mathbf{C}\mathbf{y}\,, \tag{2.88}$$

ja korrelaatio on

$$r_{xy} = \mathrm{cor}_{\mathrm{d}}(\mathbf{x}, \mathbf{y}) = \cos(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \frac{\tilde{\mathbf{x}}'\tilde{\mathbf{y}}}{\|\tilde{\mathbf{x}}\|\|\tilde{\mathbf{y}}\|} = \frac{t_{xy}}{\sqrt{t_{xx}t_{yy}}} = \frac{\mathbf{x}'\mathbf{C}\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{C}\mathbf{x} \cdot \mathbf{y}'\mathbf{C}\mathbf{y}}}\,. \tag{2.89}$$

Olkoon vektori $\tilde{\tilde{\mathbf{y}}}$ määritelty seuraavasti:

$$\tilde{\tilde{\mathbf{y}}} = \frac{1}{\|\tilde{\mathbf{y}}\|}\tilde{\mathbf{y}} = \frac{1}{\sqrt{t_{yy}}}\tilde{\mathbf{y}} = \frac{1}{\sqrt{t_{yy}}}\begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}, \tag{2.90}$$

eli $\tilde{\tilde{\mathbf{y}}}$ on ykkösen pituiseksi skaalattu keskistetty $\mathbf{y}$. Olkoon $\tilde{\tilde{\mathbf{x}}}$ vastaavalla tavalla skaalattu x-muuttujaa vastaava muuttujavektori. Vektoreiden $\tilde{\tilde{\mathbf{x}}}$ ja $\tilde{\tilde{\mathbf{y}}}$ avulla lausuttuna r_{xy} on yksinkertaisesti $\tilde{\tilde{\mathbf{x}}}$:n ja $\tilde{\tilde{\mathbf{y}}}$:n sisätulo

$$r_{xy} = \tilde{\tilde{\mathbf{x}}}'\tilde{\tilde{\mathbf{y}}} = \cos(\tilde{\tilde{\mathbf{x}}}, \tilde{\tilde{\mathbf{y}}})\,. \tag{2.91}$$

Kommentti 2.1. Korrelaatiokerroin on keskistettyjen muuttujavektoreiden välisen kulman kosini – se ei ole välttämättä sama kuin alkuperäisten muuttujavektoreiden välisen kulman kosini.

Kommentti 2.2. Kuten olemme todenneet, n elementin muuttujavektori $\mathbf{z}$ on keskistetty jos $z_1 + z_2 + \cdots + z_n = 0$ eli matriisimerkinnöin:

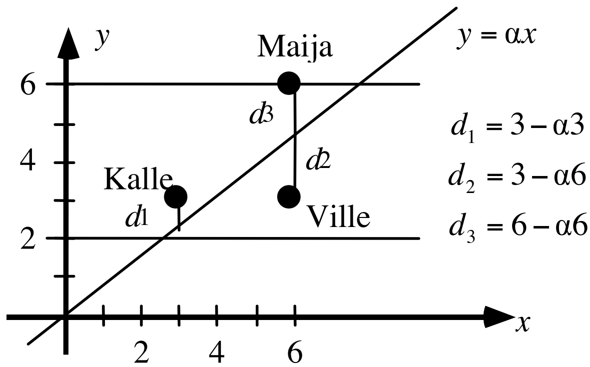
$$\mathbf{z} \text{ on keskistetty} \iff \mathbf{1}'\mathbf{z} = 0 \iff \mathbf{z} \in \mathscr{N}(\mathbf{1}') := \mathscr{C}(\mathbf{1})^{\perp}. \tag{2.92}$$

Toisin sanoen: $\mathbf{z}$ on keskistetty $\iff$ $\mathbf{z}$ on kohtisuorassa vektoria $\mathbf{1}$ vastaan.

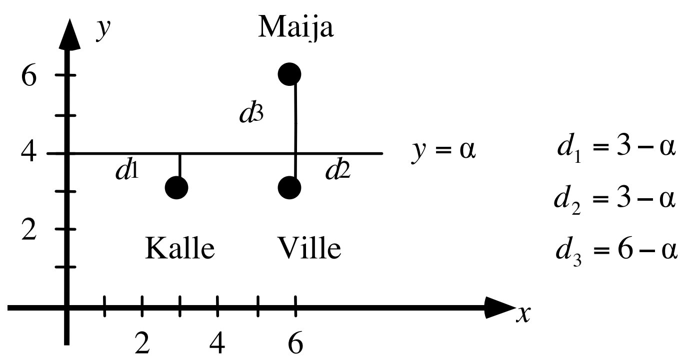
2.5 Projisointi muuttuja-avaruudessa vs. minimointi havaintoavaruudessa

Kahdessa edellisessä alaluvussa on tarkasteltu muuttuja-avaruutta: muuttujavektori $\mathbf{y}$ on projisoitu suoralle, jonka virittää muuttujavektori $\mathbf{x}$ (tai $\mathbf{1}$). Mikähän toimenpide havaintoavaruudessa vastaisi samaa asiaa?

Piirretäänpä tätä varten kahden muuttujan pisteparvi kuvion 2.14 tapaan.



Kuvio 2.14. Vektorin $\mathbf{y}$ projisointia suoralle $\mathscr{C}(\mathbf{x})$ vastaava minimointi havaintoavaruudessa: minimoidaan havaintojen y-akselin suuntaisten etäisyyksien neliösummaa suorasta $y = \alpha x$. (Tulos: kulmakerroin $\hat{\alpha} = 63/81 = 7/9$.)



Kuvio 2.15. Vektorin $\mathbf{y}$ projisointia suoralle $\mathscr{C}(\mathbf{1})$ vastaava minimointi havaintoavaruudessa: minimoidaan havaintojen y-akselin suuntaisten etäisyyksien neliösummaa suorasta $y = \alpha$. (Tulos: $\hat{\alpha} = \bar{y} = 4$.)

Oletetaan, että tehtävänä on sijoittaa kuvioon 2.14 origon kautta kulkeva suora $y = \alpha x$. Olkoon $|d_i|$ = havaintopisteen y-akselin suuntainen etäisyys suorasta $y = \alpha x$: $d_i = |y_i - \alpha x_i|$. Näiden etäisyyksien neliöiden summa on n

havainnon tapauksessa

$$\begin{aligned} g(\alpha) &= d_1^2 + d_2^2 + \cdots + d_n^2 \\ &= (y_1 - \alpha x_1)^2 + (y_2 - \alpha x_2)^2 + \cdots + (y_n - \alpha x_n)^2 \\ &= \|\mathbf{y} - \alpha \mathbf{x}\|^2. \end{aligned} \tag{2.93}$$

Havaitsemme välittömästi, että funktion $g(\alpha)$ minimointi on yhtäpitävää aikaisemmin tarkastellun funktion $f(\lambda) = \mathrm{SSE}(\mathbf{y}; \lambda\mathbf{x})$ minimoinnin kanssa. Minimin antava α:n arvo on siten

$$\hat{\alpha} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}\,. \tag{2.94}$$

Kuvion 2.14 tilanteessa $\hat{\alpha} = 63/81 = 7/9$ ja jäännösneliösumma $\mathrm{SSE}(\mathbf{y}; \hat{\alpha}\mathbf{x}) = 5$.

Toisin sanoen:

> muuttuja-avaruudessa tapahtuvaa $\mathbf{y}$:n projisointia suoralle $\mathscr{C}(\mathbf{x})$ vastaa havaintoavaruudessa origon kautta kulkevan suoran sijoittaminen siten, että havaintojen suorasta y-akselin suunnassa laskettujen etäisyyksien neliöitten summa on minimissään.

On oleellista huomata, että havaintoavaruudessa etäisyydet nyt lasketaan nimenomaan y-akselin suunnassa. Jos lasketaan havaintopisteiden kohtisuorat etäisyydet suorasta, tilanne mutkistuu; palataan tähän myöhemmin.

Itse asiassa origon kautta kulkevan suoran määrittämisessä edellä tarkastellulla tavalla on kyse regressiosuoran määrittämisestä (*pienimmän neliösumman menetelmällä*) sellaisessa tilanteessa, jossa mallissa on vain yksi selittäjä x ja vakiotermi puuttuu eli regressiosuora pakotetaan kulkemaan origon kautta. Malli on siis

$$y = \alpha x + \text{satunnaisvirhe } \varepsilon\,. \tag{2.95}$$

Entä mikä on suoralle $\mathscr{C}(\mathbf{1})$ tapahtuvaa projisointia vastaava operaatio havaintoavaruudessa? Vastaus ilmenee kuviosta 2.15. Siinä sijoitetaan pisteparveen suora $y = \alpha$ siten, että havaintopisteiden suorasta laskettujen etäisyyksien neliösumma on minimissään. Tulokseksi tulee $\hat{\alpha} = \bar{y}$:

$$\begin{aligned} \min_{\alpha} \sum_{i=1}^{n} (y_i - \alpha)^2 &= \min_{\alpha} \|\mathbf{y} - \alpha\mathbf{1}\|^2 = \min_{\alpha} \mathrm{SSE}(\mathbf{y}; \alpha\mathbf{1}) \\ &= \sum_{i=1}^{n} (y_i - \bar{y})^2 = \mathrm{SSE}(\mathbf{y}; \bar{y}\mathbf{1})\,. \end{aligned} \tag{2.96}$$

Regressioanalyysin näkökulmasta tarkasteltuna sijoitamme dataan mallin

$$y = \alpha + \text{satunnaisvirhe } \varepsilon\,, \tag{2.97}$$

eli x:llä ei ole mitään merkitystä, selittäjänä on vain vakiotermi ja malli on todella yksinkertainen. Pienimmän neliösumman estimaatiksi parametrille α tulee siis $\bar{y}$: Kuvion 2.15 tilanteessa $\bar{y} = 4$ ja jäännösneliösumma $\mathrm{SSE}(\mathbf{y}; \bar{y}\mathbf{1}) = 6$.

Tavanomainen yhden selittäjän regressiomalli, jossa on vakiotermi mukana, on

$$y = \beta_0 + \beta_1 x + \text{satunnaisvirhe } \varepsilon. \tag{2.98}$$

Pienimmän neliösumman menetelmä tarkoittaa, että aineistoon sijoitetaan suora $y = \beta_0 + \beta_1 x$ siten että havaintopisteiden suorasta (y-akselin suunnassa) laskettujen etäisyyksien neliösumma $\mathrm{SSE}(\mathbf{y}; \mathbf{1}\beta_0 + \mathbf{x}\beta_1)$ on minimissään. Tätä minimointia vastaa muuttuja-avaruudessa vektorin $\mathbf{y}$ projisointi matriisin $(\mathbf{1} : \mathbf{x})$ sarakeavaruudelle – palataan asiaan!

2.6 Keskiarvovektorit & tulosummamatriisi havaintomatriisista

Tarkastellaan n havainnon ja p muuttujan havaintomatriisia $\mathbf{U}_{n \times p}$ ja merkitään

$$\mathbf{U} = (\mathbf{u}_1 : \mathbf{u}_2 : \ldots : \mathbf{u}_p) = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix}, \tag{2.99}$$

missä siis pystyvektori $\mathbf{u}_1$ sisältää ensimmäisen muuttujan – merkittäköön sitä u_1:llä – saamat n havaintoarvoa ja p elementin vaakavektori $\mathbf{u}'_{(1)}$ sisältää ensimmäisen havaintoyksikön saamat arvot eri muuttujilla eli

- $\mathbf{u}_{(1)}$ on 1. havaintovektori, $\mathbf{u}_{(1)} \in \mathbb{R}^p$,
- $\mathbf{u}_1$ on 1. muuttujavektori, $\mathbf{u}_1 \in \mathbb{R}^n$.

Muuttujan u_1 keskiarvo $\bar{u}_1$ saadaan laskemalla $\mathbf{U}$:n 1. sarakkeen $\mathbf{u}_1$ luvut yhteen ja jakamalla summa n:llä:

$$\bar{u}_1 = \frac{1}{n}\mathbf{1}'_n\mathbf{u}_1 = (\mathbf{1}'_n\mathbf{1}_n)^{-1}\mathbf{1}'_n\mathbf{u}_1\,. \tag{2.100}$$

Tällöin

$$\bar{\bar{\mathbf{u}}}_1 = \bar{u}_1\mathbf{1}_n = \begin{pmatrix} \bar{u}_1 \\ \bar{u}_1 \\ \vdots \\ \bar{u}_1 \end{pmatrix} = \mathbf{J}\mathbf{u}_1 = \mathbf{u}_1\text{:n ortogonaaliprojektio } \mathscr{C}(\mathbf{1})\text{:lle.} \tag{2.101}$$

Vastaavasti sellainen pystyvektori $\bar{\mathbf{u}} = (\bar{u}_1, \dots, \bar{u}_p)'$, joka sisältää kaikkien muuttujien keskiarvot, saadaan jakamalla $\mathbf{U}$:n havaintovektoreiden summa n:llä:

$$\bar{\mathbf{u}} = \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \vdots \\ \bar{u}_p \end{pmatrix} = \frac{1}{n}(\mathbf{u}_{(1)} + \mathbf{u}_{(2)} + \cdots + \mathbf{u}_{(n)}) \in \mathbb{R}^p. \tag{2.102}$$

Matriisin $\mathbf{U}$ vaakarivien summa transponoituna eli siis lauseke $\mathbf{u}_{(1)} + \cdots + \mathbf{u}_{(n)}$ voidaan esittää matriisitulona

$$(\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)})\mathbf{1}_n = \mathbf{U}'\mathbf{1}_n = \mathbf{u}_{(1)} + \mathbf{u}_{(2)} + \cdots + \mathbf{u}_{(n)}\,, \tag{2.103}$$

joten

$$\bar{\mathbf{u}} = \frac{1}{n}\mathbf{U}'\mathbf{1}_n \in \mathbb{R}^p. \tag{2.104}$$

Jos keskistämme kolmen muuttujan havaintomatriisin $\mathbf{U} = (\mathbf{x} : \mathbf{y} : \mathbf{z})$, niin tulos on

$$\begin{aligned}
\widetilde{\mathbf{U}} &= (\mathbf{x} : \mathbf{y} : \mathbf{z}) - (\bar{\bar{\mathbf{x}}} : \bar{\bar{\mathbf{y}}} : \bar{\bar{\mathbf{z}}}) = (\mathbf{x} - \bar{\bar{\mathbf{x}}} : \mathbf{y} - \bar{\bar{\mathbf{y}}} : \mathbf{z} - \bar{\bar{\mathbf{z}}}) = (\tilde{\mathbf{x}} : \tilde{\mathbf{y}} : \tilde{\mathbf{z}}) \\
&= \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix} - \begin{pmatrix} \bar{x} & \bar{y} & \bar{z} \\ \bar{x} & \bar{y} & \bar{z} \\ \vdots & \vdots & \vdots \\ \bar{x} & \bar{y} & \bar{z} \end{pmatrix} = \begin{pmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ x_2 - \bar{x} & y_2 - \bar{y} & z_2 - \bar{z} \\ \vdots & \vdots & \vdots \\ x_n - \bar{x} & y_n - \bar{y} & z_n - \bar{z} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix} - \begin{pmatrix} \bar{\mathbf{u}}' \\ \bar{\mathbf{u}}' \\ \vdots \\ \bar{\mathbf{u}}' \end{pmatrix} = \begin{pmatrix} \mathbf{u}'_{(1)} - \bar{\mathbf{u}}' \\ \mathbf{u}'_{(2)} - \bar{\mathbf{u}}' \\ \vdots \\ \mathbf{u}'_{(n)} - \bar{\mathbf{u}}' \end{pmatrix} \\
&= \begin{pmatrix} (\mathbf{u}_{(1)} - \bar{\mathbf{u}})' \\ (\mathbf{u}_{(2)} - \bar{\mathbf{u}})' \\ \vdots \\ (\mathbf{u}_{(n)} - \bar{\mathbf{u}})' \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{u}}'_{(1)} \\ \tilde{\mathbf{u}}'_{(2)} \\ \vdots \\ \tilde{\mathbf{u}}'_{(n)} \end{pmatrix},
\end{aligned} \tag{2.105}$$

missä $\bar{\mathbf{u}} = (\bar{x}, \bar{y}, \bar{z})'$. On huomattava että

$$\tilde{\mathbf{u}}_{(i)} = \mathbf{u}_{(i)} - \bar{\mathbf{u}} = \begin{pmatrix} x_i - \bar{x} \\ y_i - \bar{y} \\ z_i - \bar{z} \end{pmatrix} \tag{2.106}$$

on *keskistetty i. havaintovektori*: i. havainnon tapauksessa kunkin muuttujan havaitusta arvosta on vähennetty vastaavan muuttujan keskiarvo. On hyvä tehdä selväksi ero keskistetyn muuttujan ja keskistetyn havainnon välillä:

- keskistetty muuttuja $\tilde{\mathbf{y}}$ on muuttujavektorin $\mathbf{y}$ ja sen keskiarvovektorin $\bar{\bar{\mathbf{y}}} = \bar{y}\mathbf{1}_n$ erotus $\mathbf{y} - \bar{\bar{\mathbf{y}}}$ ($\in \mathbb{R}^n$ kun havaintojen lukumäärä on n)

- keskistetty havainto $\tilde{\mathbf{u}}_{(i)}$ on havaintovektorin $\mathbf{u}_{(i)}$ ja kaikkien muuttujien keskiarvovektorin $\bar{\mathbf{u}}$ erotus $\mathbf{u}_{(i)} - \bar{\mathbf{u}}$ ($\in \mathbb{R}^p$ kun muuttujien lukumäärä on p).

Jos tarkasteltavat muuttujat ovat u_1, u_2, u_3, joista jokaisesta on n havaintoa, niin (2.105):n mukainen esitys on tietenkin

$$\begin{aligned}
\widetilde{\mathbf{U}} &= \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ \vdots & \vdots & \vdots \\ u_{n1} & u_{n2} & u_{n3} \end{pmatrix} - \begin{pmatrix} \bar{u_1} & \bar{u_2} & \bar{u_3} \\ \bar{u_1} & \bar{u_2} & \bar{u_3} \\ \vdots & \vdots & \vdots \\ \bar{u_1} & \bar{u_2} & \bar{u_3} \end{pmatrix} \\
&= \begin{pmatrix} u_{11} - \bar{u}_1 & u_{12} - \bar{u}_2 & u_{13} - \bar{u}_3 \\ u_{21} - \bar{u}_1 & u_{22} - \bar{u}_2 & u_{23} - \bar{u}_3 \\ \vdots & \vdots & \vdots \\ u_{n1} - \bar{u}_1 & u_{n2} - \bar{u}_2 & u_{n3} - \bar{u}_3 \end{pmatrix} = \{u_{ij} - \bar{u}_j\} = \{\tilde{u}_{ij}\}\,. \qquad (2.107)
\end{aligned}$$

Joissakin tilanteissa, esim. haettaessa ”poikkeavaa” havaintoa, voi olla kiinnostavaa tietää havaintovektorin $\mathbf{u}_{(i)}$ euklidinen etäisyys keskiarvovektorista $\bar{\mathbf{u}}$. Tämän etäisyyden neliö on tietenkin keskistetyn havaintovektorin pituuden neliö:

$$\begin{aligned}
\|\mathbf{u}_{(i)} - \bar{\mathbf{u}}\|^2 &= \|\tilde{\mathbf{u}}_{(i)}\|^2 \\
&= (\mathbf{u}_{(i)} - \bar{\mathbf{u}})'(\mathbf{u}_{(i)} - \bar{\mathbf{u}}) \\
&= \tilde{\mathbf{u}}_{(i)}'\tilde{\mathbf{u}}_{(i)}\,, \qquad (2.108)
\end{aligned}$$

mistä kolmen muuttujan (x, y, z) tapauksessa tulee siis

$$\|\mathbf{u}_{(i)} - \bar{\mathbf{u}}\|^2 = (x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2. \qquad (2.109)$$

Vastaavasti kahden havaintovektorin välinen etäisyys on

$$\|\mathbf{u}_{(i)} - \mathbf{u}_{(j)}\|\,, \qquad (2.110)$$

jonka neliön arvo kolmen muuttujan tapauksessa on tietenkin

$$(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2. \qquad (2.111)$$

Jos pisteparvi on ryhmittynyt jollakin erityisellä tavalla muuttujien keskinäisestä korrelaatiosta tai erilaisista varianseista johtuen, on luonnollista ajatella, että euklidinen etäisyys ei kuvaakaan kovin hyvin havainnon poikkeavuutta. *Mahalanobisin etäisyyden* yhteydessä palataan tähän asiaan; havaintovektorin $\mathbf{u}_{(i)}$ Mahalanobisin etäisyys (neliöitynä) keskiarvovektorista $\bar{\mathbf{u}}$ on lauseke

$$\mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S}) = (\mathbf{u}_{(i)} - \bar{\mathbf{u}})'\mathbf{S}^{-1}(\mathbf{u}_{(i)} - \bar{\mathbf{u}}) , \tag{2.112}$$

missä $\mathbf{S}$ on aineiston kovarianssimatriisi. Jos muuttujat ovat korreloimattomia eli $\mathbf{S}$ on lävistäjämatriisi $\mathbf{S} = \mathrm{diag}(s_x^2, s_y^2, s_z^2)$, saamme

$$\mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S}) = \frac{(x_i - \bar{x})^2}{s_x^2} + \frac{(y_i - \bar{y})^2}{s_y^2} + \frac{(z_i - \bar{z})^2}{s_z^2} . \tag{2.113}$$

2.6.1 Tulosummamatriisi

Tarkastellaan havaintomatriisia $\mathbf{U} = (\mathbf{u}_1 : \mathbf{u}_2 : \mathbf{u}_3)$ ja olkoon sen keskistetty versio

$$\widetilde{\mathbf{U}} = (\mathbf{u}_1 - \bar{\bar{\mathbf{u}}}_1 : \mathbf{u}_2 - \bar{\bar{\mathbf{u}}}_2 : \mathbf{u}_3 - \bar{\bar{\mathbf{u}}}_3) = (\tilde{\mathbf{u}}_1 : \tilde{\mathbf{u}}_2 : \tilde{\mathbf{u}}_3) . \tag{2.114}$$

Tällöin muuttujien $\mathbf{u}_i$ ja $\mathbf{u}_i$ välinen tulosumma on

$$t_{ij} = (\mathbf{u}_i - \bar{\bar{\mathbf{u}}}_i)'(\mathbf{u}_j - \bar{\bar{\mathbf{u}}}_j) = \tilde{\mathbf{u}}_i'\tilde{\mathbf{u}}_j = \mathbf{u}_i'(\mathbf{I} - \mathbf{J})\mathbf{u}_j = \mathbf{u}_i'\mathbf{C}\mathbf{u}_j . \tag{2.115}$$

Matriisitulon määritelmän mukaan tulon $\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}$ elementit ovat $\widetilde{\mathbf{U}}$:n sarakkeiden sisätuloja:

$$\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}} = \begin{pmatrix} \tilde{\mathbf{u}}_1' \\ \tilde{\mathbf{u}}_2' \\ \tilde{\mathbf{u}}_3' \end{pmatrix} (\tilde{\mathbf{u}}_1 : \tilde{\mathbf{u}}_2 : \tilde{\mathbf{u}}_3) = \{\tilde{\mathbf{u}}_i'\tilde{\mathbf{u}}_j\} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} , \tag{2.116}$$

eli tulosummamatriisi $\mathbf{T} = \widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}$. Koska $\widetilde{\mathbf{U}} = \mathbf{CU}$, saamme

$$\mathbf{T} = \widetilde{\mathbf{U}}'\widetilde{\mathbf{U}} = (\mathbf{CU})'\mathbf{CU} = \mathbf{U}'\mathbf{CU} = \mathbf{U}'(\mathbf{I} - \mathbf{J})\mathbf{U} . \tag{2.117}$$

Havaintovektorien avulla voimme esittää tulosummamatriisin seuraavasti:

$$\begin{aligned} \mathbf{T} &= \mathbf{U}'(\mathbf{I} - \mathbf{J})\mathbf{U} = \mathbf{U}'\mathbf{U} - (\mathbf{JU})'\mathbf{JU} \\ &= (\mathbf{u}_{(1)} : \ldots : \mathbf{u}_{(n)})(\mathbf{u}_{(1)} : \ldots : \mathbf{u}_{(n)})' - (\bar{\mathbf{u}} : \ldots : \bar{\mathbf{u}})(\bar{\mathbf{u}} : \ldots : \bar{\mathbf{u}})' \\ &= \sum_{i=1}^{n} \mathbf{u}_{(i)}\mathbf{u}_{(i)}' - n\bar{\mathbf{u}}\bar{\mathbf{u}}' \\ &= (\mathbf{CU})'\mathbf{CU} = (\mathbf{u}_{(1)} - \bar{\mathbf{u}} : \ldots : \mathbf{u}_{(n)} - \bar{\mathbf{u}}) \begin{pmatrix} (\mathbf{u}_{(1)} - \bar{\mathbf{u}})' \\ \vdots \\ (\mathbf{u}_{(n)} - \bar{\mathbf{u}})' \end{pmatrix} \\ &= \sum_{i=1}^{n} (\mathbf{u}_{(i)} - \bar{\mathbf{u}})(\mathbf{u}_{(i)} - \bar{\mathbf{u}})'. \end{aligned} \tag{2.118}$$

eli vielä kerran:

$$\begin{aligned}\mathbf{T} = \mathbf{U}'(\mathbf{I}-\mathbf{J})\mathbf{U} &= \{\mathbf{u}_i'\mathbf{C}\mathbf{u}_j\} = \{(\mathbf{u}_i - \bar{\bar{\mathbf{u}}}_i)'(\mathbf{u}_j - \bar{\bar{\mathbf{u}}}_j)\} \\ &= \sum_{i=1}^{n}(\mathbf{u}_{(i)} - \bar{\mathbf{u}})(\mathbf{u}_{(i)} - \bar{\mathbf{u}})' = \sum_{i=1}^{n}\mathbf{u}_{(i)}\mathbf{u}_{(i)}' - n\bar{\mathbf{u}}\bar{\mathbf{u}}'. \qquad (2.119)\end{aligned}$$

Kommentti 2.3. (Mahalanobisin etäisyys projektorin avulla.) Havaintovektorin $\mathbf{u}_{(i)}$ Mahalanobisin etäisyys (neliöitynä) keskiarvovektorista $\bar{\mathbf{u}}$ on siis lauseke

$$\mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S}) = (\mathbf{u}_{(i)} - \bar{\mathbf{u}})'\mathbf{S}^{-1}(\mathbf{u}_{(i)} - \bar{\mathbf{u}}) = d_i^2\,, \qquad (2.120)$$

missä $\mathbf{S}$ on aineiston kovarianssimatriisi. Havaintomatriisi $\mathbf{U}$ transpoosi on $\mathbf{U}' = (\mathbf{u}_{(1)} : \ldots : \mathbf{u}_{(n)})$ ja sen keskistetty versio

$$\widetilde{\mathbf{U}}' = (\mathbf{u}_{(1)} - \bar{\mathbf{u}} : \ldots : \mathbf{u}_{(n)} - \bar{\mathbf{u}}) = (\tilde{\mathbf{u}}_{(1)} : \ldots : \tilde{\mathbf{u}}_{(n)})\,. \qquad (2.121)$$

Tällöin i. keskistetty havaintovektori saadaan laskutoimituksella

$$\tilde{\mathbf{u}}_{(i)} = \widetilde{\mathbf{U}}'\mathbf{i}_i\,, \qquad (2.122)$$

missä

$$\mathbf{i}_i = (0, \ldots, 1\,(= i.\ \text{elementti}), \ldots, 0)'. \qquad (2.123)$$

Täten

$$\begin{aligned} d_i^2 &= \tilde{\mathbf{u}}_{(i)}'\mathbf{S}^{-1}\tilde{\mathbf{u}}_{(i)} \\ &= \mathbf{i}_i'\widetilde{\mathbf{U}}(\tfrac{1}{n-1}\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}})^{-1}\widetilde{\mathbf{U}}'\mathbf{i}_i \\ &= (n-1)\,\mathbf{i}_i'\widetilde{\mathbf{U}}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}})^{-1}\widetilde{\mathbf{U}}'\mathbf{i}_i \\ &= (n-1)\,\mathbf{i}_i'\mathbf{P}_{\widetilde{\mathbf{U}}}\mathbf{i}_i \\ &= (n-1)\,p_{ii}\,, \end{aligned} \qquad (2.124)$$

missä

$$\mathbf{P}_{\widetilde{\mathbf{U}}} = \widetilde{\mathbf{U}}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}})^{-1}\widetilde{\mathbf{U}}' = \{p_{ij}\} \qquad (2.125)$$

on ortogonaaliprojektori $\mathscr{C}(\widetilde{\mathbf{U}})$:lle. Kaava (2.124) on varsin kätevä tapa laskea Mahalanobisin etäisyydet. Ortogonaaliprojektorin lävistäjäelementeillä on ominaisuus $0 \leq p_{ii} \leq 1$, joten $d_i^2 \leq n-1$, $i = 1, \ldots, n$. □

Esimerkki 2.1. Tarkastellaan seuraavaa jo aiemmin käsiteltyä aineistoa; kuviot 2.16 and 2.17 havainnollistavat tilannetta.

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} 1 & 1 \\ 4 & 1 \\ 4 & 4 \end{pmatrix} \begin{matrix} \text{Kalle} \\ \text{Ville} \\ \text{Maija} \end{matrix}\,, \tag{2.126a}$$

$$\bar{\mathbf{u}} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \frac{1}{3}\left[\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 4 \end{pmatrix}\right] = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \tag{2.126b}$$

$$\bar{\bar{\mathbf{x}}} = \bar{x}\mathbf{1} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}, \quad \bar{\bar{\mathbf{y}}} = \bar{y}\mathbf{1} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \tag{2.126c}$$

$$\widetilde{\mathbf{U}} = (\tilde{\mathbf{x}} : \tilde{\mathbf{y}}) = \begin{pmatrix} 1 & 1 \\ 4 & 1 \\ 4 & 4 \end{pmatrix} - \begin{pmatrix} 3 & 2 \\ 3 & 2 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} -2 & -1 \\ 1 & -1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{u}}'_{(1)} \\ \tilde{\mathbf{u}}'_{(2)} \\ \tilde{\mathbf{u}}'_{(3)} \end{pmatrix}, \tag{2.126d}$$

$$\mathbf{S} = \begin{pmatrix} 3 & 3/2 \\ 3/2 & 3 \end{pmatrix}, \quad \mathbf{S}^{-1} = \frac{4}{9}\begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}, \tag{2.126e}$$

missä esim. 1. keskistetty havaintovektori on

$$\tilde{\mathbf{u}}_{(1)} = \mathbf{u}_{(1)} - \bar{\mathbf{u}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}. \tag{2.126f}$$

Lisäksi

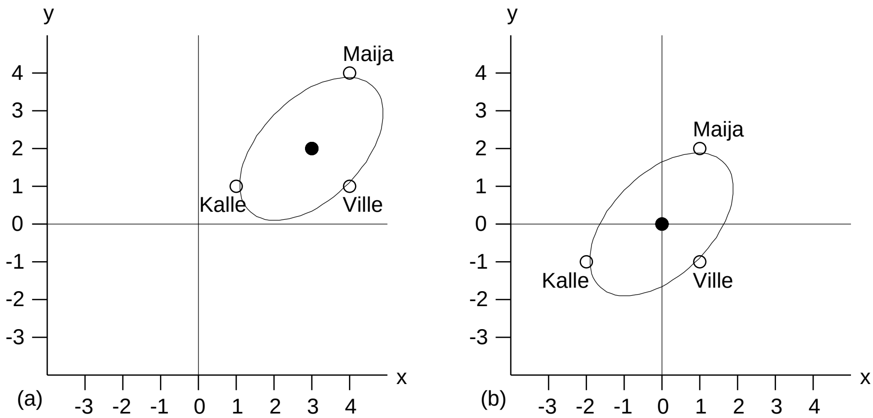
$$\|\mathbf{u}_{(1)} - \bar{\mathbf{u}}\|^2 = \|\mathbf{u}_{(3)} - \bar{\mathbf{u}}\|^2 = 5\,, \quad \|\mathbf{u}_{(2)} - \bar{\mathbf{u}}\|^2 = 2\,, \tag{2.126g}$$

$$\mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S}) = 4/3\,, \quad i = 1,2,3\,. \tag{2.126h}$$

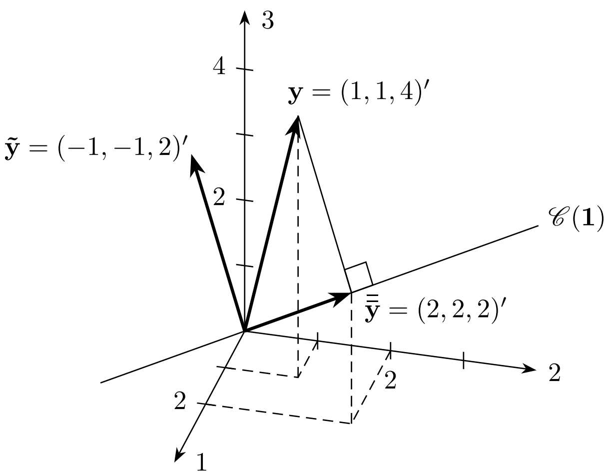
Korostettakoon vielä kerran, että $\bar{\mathbf{u}}$ ja $\tilde{\mathbf{u}}_{(i)}$ ovat havaintoavaruuden vektoreita, mutta $\bar{\bar{\mathbf{y}}}$ ja $\bar{\bar{\mathbf{x}}}$ ovat muuttuja-avaruuden vektoreita. □

Siis:

- havaintoaineiston keskistäminen merkitsee havaintoavaruudessa vain origon siirtoa: pisteparvi vain vaihtaa paikkaa,
- muuttuja-avaruudessa keskistäminen merkitsee muuttujavektoreiden projisointia $\mathscr{C}(\mathbf{1})$:n ortogonaalikomplementille – tähän palataan myöhemmin,
- muuttujan y kertominen vakiolla $a \neq 1$, $a > 0$, merkitsee muuttuja-avaruudessa muuttujavektorin $\mathbf{y}$ pituuden muutosta, havaintoavaruuden pisteparvessa se merkitsee y-akselin skaalauksen muutosta (esim. senteistä milleihin).



Kuvio 2.16. (Esimerkki 2.1) (a) Alkuperäinen aineisto, (b) keskistetty aineisto. Keskiarvopiste $\bar{\mathbf{u}} = (\bar{x}, \bar{y})'$ merkitty •:lla. Kuviossa on myös konfidenssiellipsi (normaalijakauman mukaisesti). Kaikilla havainnoilla sama Mahalanobisin etäisyys keskiarvosta. Kyseessä havaintoavaruus.



Kuvio 2.17. (Esimerkki 2.1) Muuttujavektori $\mathbf{y}$ havaintoarvaruudessa $\mathbb{R}^3$; $\tilde{\mathbf{y}}$ on keskistetty $\mathbf{y}$ ja $\bar{\bar{\mathbf{y}}} = \mathbf{J}\mathbf{y}$.

Katsotaanpa vielä Kalle-havainnon etäisyyttä (neliöön korotettuna) keskiarvosta $\bar{\mathbf{u}}$ eli keskistetyn havaintovektorin pituuden neliötä

$$\|\tilde{\mathbf{u}}_{(1)}\|^2 = \tilde{\mathbf{u}}'_{(1)}\tilde{\mathbf{u}}_{(1)} = (-2)^2 + (-1)^2 = 5\,, \tag{2.127}$$

mikä siis on keskistetyn havaintomatriisin $\widetilde{\mathbf{U}}$ ensimmäisen vaakarivin lukujen neliöiden summa. Matriisin $\widetilde{\mathbf{U}}$ kaikkien elementtien neliöiden summa on

$$\tilde{\mathbf{u}}'_{(1)}\tilde{\mathbf{u}}_{(1)} + \tilde{\mathbf{u}}'_{(2)}\tilde{\mathbf{u}}_{(2)} + \tilde{\mathbf{u}}'_{(3)}\tilde{\mathbf{u}}_{(3)} = \|\tilde{\mathbf{u}}_{(1)}\|^2 + \|\tilde{\mathbf{u}}_{(2)}\|^2 + \|\tilde{\mathbf{u}}_{(3)}\|^2 := \mathrm{SS}_{\mathbf{U}}\,, \tag{2.128}$$

joka siis tietenkin on sarakkeittain tarkasteltuna

$$\mathrm{SS}_{\mathbf{U}} = \mathrm{SS}_x + \mathrm{SS}_y = \|\tilde{\mathbf{x}}\|^2 + \|\tilde{\mathbf{y}}\|^2. \tag{2.129}$$

Luku $\mathrm{SS}_{\mathbf{U}}$ on matriisin $\widetilde{\mathbf{U}}$ Frobeniuksen normin neliö:

$$\|\widetilde{\mathbf{U}}\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{p} \tilde{u}_{ij}^2 = \mathrm{SS}_{\mathbf{U}}\ . \tag{2.130}$$

Luku $\mathrm{SS}_{\mathbf{U}}$ kertoo tavallaan aineiston kokonaisvaihtelusta. Jakamalla $\mathrm{SS}_{\mathbf{U}}$ luvulla $(n-1)$ saadaan tulokseksi muuttujien varianssien summa:

$$\frac{1}{n-1}\,\mathrm{SS}_{\mathbf{U}} = \mathrm{var_d}(\mathbf{x}) + \mathrm{var_d}(\mathbf{y}) = \mathrm{tr}[\mathrm{cov_d}(\mathbf{U})] = \mathrm{tr}(\mathbf{S})\,. \tag{2.131}$$

Havaintoavaruudessa tulkittuna $\mathrm{SS}_{\mathbf{U}}$ on havaintojen etäisyyksien neliöitten summa, etäisyydet laskettuna keskiarvosta $\bar{\mathbf{u}}$. Muuttuja-avaruudessa se on keskistettyjen muuttujavektorien normien neliöiden summa. Voimme ilmaista $\mathrm{SS}_{\mathbf{U}}$:n myös matriisin jäljen avulla seuraavasti:

$$\mathrm{SS}_{\mathbf{U}} = \mathrm{tr}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}) = \|\widetilde{\mathbf{U}}\|_F^2 = \mathrm{tr}(\mathbf{U}'\mathbf{C}\mathbf{U}) = \mathrm{tr}(\mathbf{T})\,. \tag{2.132}$$

Luku 3

Kertolasku

Kiitos viimeisestä ja anteeksi, jos käyttäydyin sopimattomasti, mutta käyttäydyn yleensä sopimattomasti ja kaikki ihmiset tietävät jo sen, niin ettei siinä pitäisi olla mitään yllättävää. Kuule, taisin unohtaa sydämeni sinne luoksesi.

Mika Waltari (1949): *Neljä päivänlaskua.*

3.1 Määritelmä

Olemme jo aiemmin käsitelleet vaakavektorin $\mathbf{a}'$ ja pystyvektorin $\mathbf{b}$ tuloa (eli $\mathbf{a}$:n ja $\mathbf{b}$:n sisätuloa)

$$\mathbf{a}'\mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{b}'\mathbf{a} = \langle \mathbf{b}, \mathbf{a} \rangle \,. \tag{3.1}$$

Samoin on määritelty matriisin $\mathbf{A}_{n\times m}$ ja pystyvektorin $\mathbf{b} \in \mathbb{R}^m$ tuloksi lauseke

$$\mathbf{Ab} = (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \mathbf{a}_1 b_1 + \mathbf{a}_2 b_2 + \cdots + \mathbf{a}_m b_m \,. \tag{3.2}$$

Vaakavektorin $\mathbf{x}'$ ja matriisin $\mathbf{A}_{n\times m}$ tulo on $\mathbf{A}$:n vaakarivien lineaarikombinaatio

$$\mathbf{x}'\mathbf{A} = (x_1, x_2, \ldots, x_n) \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix} = x_1 \mathbf{a}'_{(1)} + x_2 \mathbf{a}'_{(2)} \cdots + x_n \mathbf{a}'_{(n)} \,. \tag{3.3}$$

Nämä lausekkeet ovat erikoistapauksia yleisestä matriisien kertolaskusäännöstä. Olkoon $\mathbf{A} \in \mathbb{R}^{n\times m}$ ja $\mathbf{B} \in \mathbb{R}^{m\times p}$, ja olkoon $\mathbf{C}$ matriisien $\mathbf{A}$ ja $\mathbf{B}$ tulo:

$$\mathbf{A}_{n\times m}\mathbf{B}_{m\times p} = \mathbf{C}_{n\times p} \,. \tag{3.4}$$

Tällöin $\mathbf{C}$:n elementit lasketaan seuraavasti:

$$c_{ij} = \sum_{k=1}^{m} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj}\,. \tag{3.5}$$

Esimerkiksi

$$c_{23} = \sum_{k=1}^{m} a_{2k}b_{k3} = a_{21}b_{13} + a_{32}b_{23} + \cdots + a_{3m}b_{m3}\,. \tag{3.6}$$

Yksinkertaisin muistisääntö kertolaskulle on seuraava: Jos $\mathbf{A}_{n\times m}$ kirjoitetaan vaakariveittäin ja $\mathbf{B}_{m\times p}$ sarakkeittain ositettuna, niin [ks. (2.4), s. 75]

$$\begin{aligned}
\mathbf{AB} &= \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix} (\mathbf{b}_1 : \mathbf{b}_2 : \ldots : \mathbf{b}_p) = \begin{pmatrix} \mathbf{a}'_{(1)}\mathbf{b}_1 & \mathbf{a}'_{(1)}\mathbf{b}_2 & \ldots & \mathbf{a}'_{(1)}\mathbf{b}_p \\ \mathbf{a}'_{(2)}\mathbf{b}_1 & \mathbf{a}'_{(2)}\mathbf{b}_2 & \ldots & \mathbf{a}'_{(2)}\mathbf{b}_p \\ \vdots & \vdots & & \vdots \\ \mathbf{a}'_{(n)}\mathbf{b}_1 & \mathbf{a}'_{(n)}\mathbf{b}_2 & \ldots & \mathbf{a}'_{(n)}\mathbf{b}_p \end{pmatrix} \\
&= \{\mathbf{a}'_{(i)}\mathbf{b}_j\} = \{\langle \mathbf{a}_{(i)}, \mathbf{b}_j \rangle\} = \{c_{ij}\} \\
&= (\mathbf{c}_1 : \mathbf{c}_2 : \ldots : \mathbf{c}_p) \in \mathbb{R}^{n\times p}.
\end{aligned} \tag{3.7}$$

Lausekkeesta (3.7) havaitaan, että $\mathbf{A}$:n i. vaakarivin transpoosi $\mathbf{a}'_{(i)}$ asetetaan $\mathbf{B}$:n j. sarakkeen $\mathbf{b}_j$ päälle, päällekkäiset elementit kerrotaan ja tulot lasketaan yhteen; näin saadaan tulomatriisin i. rivin j. elementti. Jotta kertolasku $\mathbf{A}$:n ja $\mathbf{B}$:n välillä olisi määritelty, on

> $\mathbf{A}$:n sarakkeiden lukumäärän oltava sama kuin $\mathbf{B}$:n vaakarivien lukumäärän.

Vektoreiden $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ sisätulo määriteltiin lausekkeena $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}'\mathbf{v}$. Täten matriisitulon $\mathbf{AB}$ $(= \mathbf{C})$ elementit ovat tiettyjä sisätuloja: $\mathbf{AB} = \{c_{ij}\} = \{\langle \mathbf{a}_{(i)}, \mathbf{b}_j \rangle\}$, ts. c_{ij} on $\mathbf{A}$:n i. vaakavektorin (transpoosin) ja $\mathbf{B}$:n j. sarakkeen sisätulo. Esimerkiksi $\mathbf{A}'\mathbf{A}$ on tällöin

$$\begin{aligned}
\mathbf{A}'\mathbf{A} &= (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m)'(\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) \\
&= \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_m \end{pmatrix} (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) = \{\mathbf{a}'_i\mathbf{a}_j\} = \{\langle \mathbf{a}_i, \mathbf{a}_j \rangle\}\,.
\end{aligned} \tag{3.8}$$

Esimerkki 3.1.

(a) $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 6 & \ldots & \ldots \\ 15 & \ldots & \ldots \end{pmatrix}$

(b) $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ 8 & 15 \\ 12 & 23 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$

(d) $(0, 0, 1) \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = (5, 6)$

(e) $\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \\ 11 \end{pmatrix}$

(f) $(1, 1, 1) \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = (9, 12)$

(g) $(1, 1, 1) \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 21$

(h) $\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

(i) $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

(j) $\begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix} \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix} = \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix}$

(k) $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} (4, 5, 6) = \begin{pmatrix} 1 \cdot 4 & 1 \cdot 5 & 1 \cdot 6 \\ 2 \cdot 4 & 2 \cdot 5 & 2 \cdot 6 \\ 3 \cdot 4 & 3 \cdot 5 & 3 \cdot 6 \end{pmatrix}$

(l) $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (1, 2, 3) = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$

(m) $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (1, 1, 1) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

(n) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 0 & 0 \end{pmatrix}$

(o) $\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

(p) $\begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} b-a \\ c-a \\ d-a \end{pmatrix}$

(q) $\mathbf{A} = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix} \implies \mathbf{A}^2 = \mathbf{A}', \quad \mathbf{A}^4 = \mathbf{A}, \quad \mathbf{A}\mathbf{A}' = \mathbf{I}_2.$

3.2 Kertolaskun ominaisuuksia

Seuraavassa on lueteltu joitakin helposti todistettavia kertolaskun ominaisuuksia:

K1. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) = \mathbf{ABC}$,

K2. $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$,

K3. $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$,

K4. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$,

K5. $\mathbf{A}'\mathbf{A}$ ja $\mathbf{AA}'$ ovat aina symmetrisiä,

K6. $\mathbf{AB} = \mathbf{0}$, mutta silti voi olla $\mathbf{A} \neq \mathbf{0}$, $\mathbf{B} \neq \mathbf{0}$,

K7. $\mathbf{A}^2 = \mathbf{A} \iff (\mathbf{I}-\mathbf{A})^2 = \mathbf{I} - \mathbf{A}$,

K8. $\mathbf{A}^2 = \mathbf{A} \implies \mathbf{A}^k = \mathbf{A},\ k = 1, 2, \dots$

K9. $\mathbf{AB} \neq \mathbf{BA}$ (yleensä),

K10. $\mathbf{LAY} = \mathbf{MAY} \not\Rightarrow \mathbf{LA} = \mathbf{MA}$ (yleensä),

K11. $\mathbf{A}'\mathbf{A} = \mathbf{0} \iff \mathbf{A} = \mathbf{0}$,

K12. $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$ (kun $\mathbf{A} \in \mathbb{R}^{n\times m}$ ja $\mathbf{B} \in \mathbb{R}^{m\times n}$),

K13. $\mathbf{A} = \mathbf{0} \iff \operatorname{tr}(\mathbf{A}'\mathbf{A}) = 0$,

K14. $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} \in \mathbb{R}$, kun $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, *K15.* $\mathbf{x}'\mathbf{A}_{n\times m}\mathbf{y} = \mathbf{y}'\mathbf{A}'\mathbf{x} = \mathrm{tr}(\mathbf{x}'\mathbf{A}\mathbf{y}) = \mathrm{tr}(\mathbf{A}\mathbf{y}\mathbf{x}') \in \mathbb{R}$, kun $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$,

K16. $(\mathbf{x}'\mathbf{A}_{n\times m}\mathbf{y})^2 = \mathbf{y}'\mathbf{A}'\mathbf{x}\mathbf{x}'\mathbf{A}\mathbf{y}$, kun $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$.

On erityisesti huomattava, että em. ominaisuuksia tarkastellessa on matriisien dimensioiden oltava sopivia ts. rivi- ja sarakelukumäärien on oltava sellaiset, että kyseiset laskutoimitukset ovat määriteltyjä. Esimerkiksi matriisin jäljen kommutatiivisuus $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ edellyttää, että tulot $\mathbf{AB}$ ja $\mathbf{BA}$ ovat olemassa; tämähän toteutuu vain jos $\mathbf{A}$ on $n \times m$-matriisi ja $\mathbf{B}$:n on $m \times n$-matriisi. Mainittakoon, että jäljen kommutatiivisuus on aivan erityisen hyödyllinen tulos ja lukijan on hyvä käydä sen todistus yksityiskohtaisesti läpi.

Reaaliluvuillahan on voimassa

$$ab = 0 \iff a = 0 \text{ ja/tai } b = 0. \tag{3.9}$$

Vastaava tulos ei kuitenkaan päde matriiseille. Samoin nollasta poikkeavilla reaaliluvuilla a, y, ℓ, m on tietenkin voimassa

$$\ell a y = m a y \implies \ell a = m a \implies \ell = m. \tag{3.10}$$

Matriiseilla ei kuitenkaan ole vastaavaa supistussääntöä – vain erikoistilanteessa voimme matriisiyhtälöstä

$$\mathbf{LA} = \mathbf{MA} \tag{3.11}$$

supistaa matriisin $\mathbf{A}$ pois ja päätyä yhtälöön $\mathbf{L} = \mathbf{M}$.

Jos $\mathbf{A}$ on $n \times n$-neliömatriisi ja $\mathbf{B}_{n\times n}$ toteuttaa yhtälön

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n\,, \tag{3.12}$$

niin kertomalla (3.11) oikealta $\mathbf{B}$:llä saadaan $\mathbf{A}$ supistetuksi. Ehdon (3.12) toteuttava matriisi $\mathbf{B}$ on $\mathbf{A}$:n *käänteismatriisi* ja merkitsemme sitä symbolilla $\mathbf{A}^{-1}$. Jos $\mathbf{A}^{-1}$ on olemassa, sanomme $\mathbf{A}$:ta *epäsingulaariseksi* matriisiksi. Havaitsemme välittömästi mm., että jos $d_i \neq 0$, $i = 1, 2, \dots, n$, niin

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix} \implies \mathbf{D}^{-1} = \begin{pmatrix} 1/d_1 & 0 & \dots & 0 \\ 0 & 1/d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/d_n \end{pmatrix}. \tag{3.13}$$

Seuraavassa on muutamia hyödyllisiä supistussääntöjä:

3.2.1 Supistussääntöjä

SS1. $\mathbf{A}'\mathbf{A} = \mathbf{0} \implies \mathbf{A} = \mathbf{0}$

SS2. $\mathbf{A}'\mathbf{AB} = \mathbf{0} \implies \mathbf{AB} = \mathbf{0}$

SS3. $\mathbf{A}'\mathbf{AC} = \mathbf{A}'\mathbf{AD} \implies \mathbf{AC} = \mathbf{AD}$

SS4. $\mathbf{A}'\mathbf{Ax} = \mathbf{0} \implies \mathbf{Ax} = \mathbf{0}$

SS5. $\mathbf{x}'\mathbf{A}'\mathbf{Ax} = 0 \implies \mathbf{Ax} = \mathbf{0}$

SS6. $\mathbf{Ax} = \mathbf{0}\ \forall\ \mathbf{x} \implies \mathbf{A} = \mathbf{0}$

SS7. $\mathbf{Ax} = \mathbf{Bx}\ \forall\ \mathbf{x} \implies \mathbf{A} = \mathbf{B}$

SS8. Olkoot $\mathbf{A}$:n vaakarivit vapaat. Tällöin $\mathbf{LA} = \mathbf{MA} \implies \mathbf{L} = \mathbf{M}$.

SS9. Olkoot $\mathbf{B}$:n sarakkeet vapaat. Tällöin $\mathbf{BF} = \mathbf{BG} \implies \mathbf{F} = \mathbf{G}$.

SS10. Olkoon $\mathrm{r}(\mathbf{AY}) = \mathrm{r}(\mathbf{A})$. Tällöin $\mathbf{LAY} = \mathbf{MAY} \implies \mathbf{LA} = \mathbf{MA}$.

SS11. Olkoon $\mathrm{r}(\mathbf{DA}) = \mathrm{r}(\mathbf{A})$. Tällöin $\mathbf{DAM} = \mathbf{DAN} \implies \mathbf{AM} = \mathbf{AN}$.

SS12. Olkoon $\mathbf{y}$ jokin annettu $\mathbb{R}^n$:n vektori, $\mathbf{y} \neq \mathbf{0}$. Tällöin

$$\mathbf{y}'\mathbf{Qx} = 0\ \forall\ \mathbf{Q}_{n\times p} \implies \mathbf{x} = \mathbf{0}\,.$$

Sääntö *SS1* seuraa siitä, että matriisin $\mathbf{A}'\mathbf{A}$ i. lävistäelementti on (3.8):n mukaan $\mathbf{a}_i'\mathbf{a}_i$, mikä on tietenkin 0 vain jos $\mathbf{a}_i = \mathbf{0}$. Koska näin on oltava kaikilla i:n arvoilla, on $\mathbf{A}$ välttämättä nollamatriisi.

Tulos *SS2* saadaan kertomalla yhtälö $\mathbf{A}'\mathbf{AB} = \mathbf{0}$ vasemmalta $\mathbf{B}'$:lla ja merkitsemällä $\mathbf{AB} = \mathbf{F}$. Nyt $\mathbf{F}'\mathbf{F} = \mathbf{0}$ toteutuu vain jos $\mathbf{F} = \mathbf{AB} = \mathbf{0}$.

Säännön *SS3* osoittamiseksi kirjoitetaan se muotoon

$$\mathbf{A}'\mathbf{A}(\mathbf{C} - \mathbf{D}) = \mathbf{0}\,. \tag{3.14}$$

Kertomalla (3.14) vasemmalta $(\mathbf{C} - \mathbf{D})'$:lla saadaan

$$(\mathbf{C} - \mathbf{D})'\mathbf{A}'\mathbf{A}(\mathbf{C} - \mathbf{D}) = \mathbf{0}\,. \tag{3.15}$$

Käyttämällä tulon transponointisääntöä $[\mathbf{A}(\mathbf{C} - \mathbf{D})]' = (\mathbf{C} - \mathbf{D})'\mathbf{A}'$ voimme kirjoittaa (3.15):n yhtäpitävästi

$$[\mathbf{A}(\mathbf{C} - \mathbf{D})]'\mathbf{A}(\mathbf{C} - \mathbf{D}) = \mathbf{0}\,, \tag{3.16}$$

mistä *SS1*:n perusteella seuraa, että $\mathbf{A}(\mathbf{C} - \mathbf{D}) = \mathbf{0}$ eli $\mathbf{AC} = \mathbf{AD}$.

Ominaisuuksia *SS4* ja *SS5* käsiteltiin jo luvussa 2.3 (s. 86) matriisin sarakkeiden lineaarisen riippumattomuuden yhteydessä, jolloin myös käsiteltiin matriisin nolla-avaruutta $\mathscr{N}(\mathbf{A})$. Nyt *SS4* merkitsee itse asiassa luvun 2.3 tulosta (2.58) (s. 89):

$$\mathscr{N}(\mathbf{A}) = \mathscr{N}(\mathbf{A}'\mathbf{A}). \tag{3.17}$$

Sääntö *SS8*. Luvussa 2.3 (s. 90) havaitsimme, että jos $\mathbf{B}$:n sarakkeet ovat vapaat, niin silloin myös matriisin $\mathbf{B}'\mathbf{B}$ sarakkeet ovat vapaat. Jos neliömatriisin $\mathbf{B}'\mathbf{B}$ sarakkeet ovat vapaat, niin sillä on olemassa käänteismatriisi $(\mathbf{B}'\mathbf{B})^{-1}$. Vastaavasti jos $\mathbf{A}$:n vaakarivit ovat vapaat, niin silloin on $(\mathbf{A}\mathbf{A}')^{-1}$ olemassa. Kertomalla yhtälön

$$\mathbf{L}\mathbf{A} = \mathbf{M}\mathbf{A} \tag{3.18}$$

oikealta matriisilla $\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ saamme $\mathbf{A}$:n supistetuksi, joten *SS8* on todellakin voimassa. Tuloksen *SS9* todistus on aivan vastaava.

Astesupistussäännöt *SS10* ja *SS11*: ks. Puntanen, Styan & Isotalo (2011, Ch. 6).

Esimerkki 3.2. (a) Milloin yhtälö

$$(\mathbf{A}+\mathbf{B})^2 = \mathbf{A}^2 + 2\mathbf{A}\mathbf{B} + \mathbf{B}^2 \tag{3.19}$$

on voimassa? Koska

$$(\mathbf{A}+\mathbf{B})^2 = (\mathbf{A}+\mathbf{B})(\mathbf{A}+\mathbf{B}) = \mathbf{A}^2 + \mathbf{A}\mathbf{B} + \mathbf{B}\mathbf{A} + \mathbf{B}^2, \tag{3.20}$$

havaitsemme, että (3.19) toteutuu jos ja vain jos $\mathbf{A}\mathbf{B} + \mathbf{B}\mathbf{A} = 2\mathbf{A}\mathbf{B}$, mikä puolestaan toteutuu jos ja vain jos $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$.

(b) Onkohan olemassa sellaista epäsingulaarista matriisia $\mathbf{A}$, joka toteuttaa yhtälön

$$\mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{A} = \mathbf{A}\,. \tag{3.21}$$

Kerrotaan (3.21) oikealta $\mathbf{A}^{-1}$:llä, jolloin saadaan yhtälö

$$\mathbf{A}\mathbf{B}\mathbf{A}^{-1} - \mathbf{B} = \mathbf{I}\,. \tag{3.22}$$

Ottamalla jälki (trace) (3.22):n kummaltakin puolelta ja käyttämällä jäljen kommutatiivisuutta päädytään ristiriitaan, joten vastaus (b)-kohdan kysymykseen on *ei*.

(c) Onko olemassa sellaisia matriiseja $\mathbf{P}$ ja $\mathbf{Q}$ jotka toteuttavat yhtälön

$$\mathbf{P}\mathbf{Q} - \mathbf{Q}\mathbf{P} = \mathbf{I}? \tag{3.23}$$

Tuloksen $\operatorname{tr}(\mathbf{P}\mathbf{Q}) = \operatorname{tr}(\mathbf{Q}\mathbf{P})$ avulla on helposti havaittavissa, että vastaus on *ei*.

(d) Olkoon $\mathbf{A}$ neliömatriisi. Osoita että

$$\operatorname{tr}(\mathbf{A}'\mathbf{A}) \geq \operatorname{tr}(\mathbf{A}^2)\,, \tag{3.24}$$

ja että yhtäsuuruus on voimassa jos ja vain jos $\mathbf{A}$ on symmetrinen. Todistuksen aluksi havaitsemme ensinnä että

$$\operatorname{tr}(\mathbf{A}' - \mathbf{A})(\mathbf{A}' - \mathbf{A})' := \operatorname{tr} \mathbf{D}\mathbf{D}' \geq 0 \,. \tag{3.25}$$

Lisäksi

$$\begin{aligned} \operatorname{tr}(\mathbf{A}' - \mathbf{A})(\mathbf{A}' - \mathbf{A})' &= \operatorname{tr}(\mathbf{A}' - \mathbf{A})(\mathbf{A} - \mathbf{A}') \\ &= \operatorname{tr}(\mathbf{A}'\mathbf{A} - \mathbf{A}'\mathbf{A}' - \mathbf{A}\mathbf{A} + \mathbf{A}\mathbf{A}') \\ &= \operatorname{tr}(\mathbf{A}'\mathbf{A}) - \operatorname{tr}[(\mathbf{A}^2)'] - \operatorname{tr}(\mathbf{A}^2) + \operatorname{tr}(\mathbf{A}\mathbf{A}') \\ &= 2\operatorname{tr}(\mathbf{A}'\mathbf{A}) - 2\operatorname{tr}(\mathbf{A}^2) \,, \end{aligned} \tag{3.26}$$

missä on käytetty hyväksi yhtälöä $\operatorname{tr}[(\mathbf{A}^2)'] = \operatorname{tr}(\mathbf{A}^2)$. Väite (3.24) seuraa (3.25):sta ja (3.26):stä. Yhtäsuuruusehto on välitön seuraus *K13*:sta. □

3.3 Ositettujen matriisien kertolasku

Ositettujen matriisien kertolasku – mikäli ositteet ovat sopivan kokoisia – voidaan suorittaa siten, että kuvitellaan ositteiden olevan reaalilukuja. Täten esimerkiksi:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix} = \begin{pmatrix} \mathbf{AE} + \mathbf{BG} & \mathbf{AF} + \mathbf{BH} \\ \mathbf{CE} + \mathbf{DG} & \mathbf{CF} + \mathbf{DH} \end{pmatrix} . \tag{3.27}$$

Itse asiassa olemme jo useasti soveltaneet ositettujen matriisien kertolaskusääntöä. Esimerkiksi kertolasku $\mathbf{Ax}$, kun $\mathbf{A} \in \mathbb{R}^{n\times m}$ ja $\mathbf{x} \in \mathbb{R}^m$, voidaan kirjoittaa muodossa

$$\mathbf{Ax} = (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \cdots + \mathbf{a}_m x_m \,, \tag{3.28}$$

mikä on aivan sama sääntö kuin jos $\mathbf{a}_i$:t olisivatkin reaalilukuja. Jos $\mathbf{A}$ ositetaan vaakariveittäin, saamme

$$\mathbf{Ax} = \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{a}'_{(1)}\mathbf{x} \\ \mathbf{a}'_{(2)}\mathbf{x} \\ \vdots \\ \mathbf{a}'_{(n)}\mathbf{x} \end{pmatrix} . \tag{3.29}$$

Kertolasku $\mathbf{A}_{n\times m}\mathbf{B}_{m\times p}$ voidaan kirjoittaa muodossa

$$\mathbf{AB} = \mathbf{A}(\mathbf{b}_1 : \mathbf{b}_2 : \ldots : \mathbf{b}_p) = (\mathbf{Ab}_1 : \mathbf{Ab}_2 : \ldots : \mathbf{Ab}_p) \,, \tag{3.30}$$

ja

$$\mathbf{AB} = \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix} \mathbf{B} = \begin{pmatrix} \mathbf{a}'_{(1)}\mathbf{B} \\ \mathbf{a}'_{(2)}\mathbf{B} \\ \vdots \\ \mathbf{a}'_{(n)}\mathbf{B} \end{pmatrix}. \tag{3.31}$$

Havaitsemme myös, että

$$\mathbf{AA}' = (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_m \end{pmatrix} = \mathbf{a}_1\mathbf{a}'_1 + \mathbf{a}_2\mathbf{a}'_2 + \cdots + \mathbf{a}_m\mathbf{a}'_m\,. \tag{3.32}$$

Esimerkki 3.3.

(a) $$\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} \mathbf{a}' \\ \mathbf{B}' \end{pmatrix} (\mathbf{a} : \mathbf{B}) = \begin{pmatrix} \mathbf{a}'\mathbf{a} & \mathbf{a}'\mathbf{B} \\ \mathbf{B}'\mathbf{a} & \mathbf{B}'\mathbf{B} \end{pmatrix}$$

$$= \begin{pmatrix} \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} (1, 2, 3) + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} (4, 5, 6)$$

(b) $$\begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{pmatrix} = (-\mathbf{1}_3 : \mathbf{I}_3) \begin{pmatrix} y_0 \\ \mathbf{y} \end{pmatrix} = -y_0\mathbf{1}_3 + \mathbf{y} = \begin{pmatrix} y_1 - y_0 \\ y_2 - y_0 \\ y_3 - y_0 \end{pmatrix}$$

□

Esimerkki 3.4. Kuten jo (1.66):ssa (s. 30) todettiin, on matriisin $\mathbf{A}_{n\times m}$ elementtien neliösumma

$$\|\mathbf{A}\|_F^2 = \operatorname{tr}(\mathbf{A}'\mathbf{A}) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2\,, \tag{3.33}$$

missä $\|\mathbf{A}\|_F$ viittaa $\mathbf{A}$:n euklidisen eli Frobeniuksen normiin. Tämä perustuu siis siihen, että

$$\begin{aligned} \operatorname{tr}(\mathbf{A}'\mathbf{A}) &= \operatorname{tr} \begin{pmatrix} \mathbf{a}'_1\mathbf{a}_1 & \mathbf{a}'_1\mathbf{a}_2 & \ldots & \mathbf{a}'_1\mathbf{a}_m \\ \mathbf{a}'_2\mathbf{a}_1 & \mathbf{a}'_2\mathbf{a}_2 & \ldots & \mathbf{a}'_2\mathbf{a}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}'_m\mathbf{a}_1 & \mathbf{a}'_m\mathbf{a}_2 & \ldots & \mathbf{a}'_m\mathbf{a}_m \end{pmatrix} \\ &= \mathbf{a}'_1\mathbf{a}_1 + \mathbf{a}'_2\mathbf{a}_2 + \cdots + \mathbf{a}'_m\mathbf{a}_m \\ &= \|\mathbf{a}_1\|^2 + \|\mathbf{a}_2\|^2 + \cdots + \|\mathbf{a}_m\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2\,. \end{aligned} \tag{3.34}$$

Olemme myös havainneet, että

$$\|\mathbf{A}\|_F^2 = [\operatorname{vec}(\mathbf{A})]' \operatorname{vec}(\mathbf{A}) = \|\operatorname{vec}(\mathbf{A})\|^2, \quad \text{kun } \operatorname{vec}(\mathbf{A}) = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{pmatrix}. \tag{3.35}$$

Matriisien $\mathbf{A}_{n\times m}$ ja $\mathbf{B}_{n\times m}$ sisätulo määriteltiin sivulla 30 vektorien sisätuloa vastaten seuraavasti:

$$\begin{aligned} \langle \mathbf{A}, \mathbf{B} \rangle &= \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ij} = [\operatorname{vec}(\mathbf{A})]' \operatorname{vec}(\mathbf{B}) \\ &= \mathbf{a}_1'\mathbf{b}_1 + \mathbf{a}_2'\mathbf{b}_2 + \cdots + \mathbf{a}_m'\mathbf{b}_m = \operatorname{tr}(\mathbf{A}'\mathbf{B}). \end{aligned} \tag{3.36}$$

Matriisien $\mathbf{A}$ ja $\mathbf{B}$ etäisyys (neliöitynä) on (perustele viimeinen yhtälö!):

$$\begin{aligned} \mathrm{d}^2(\mathbf{A}, \mathbf{B}) &= \sum_{i=1}^{n} \sum_{j=1}^{m} (a_{ij} - b_{ij})^2 = \|\mathbf{A} - \mathbf{B}\|_F^2 \\ &= \operatorname{tr}[(\mathbf{A} - \mathbf{B})'(\mathbf{A} - \mathbf{B})] = \operatorname{tr}(\mathbf{A}'\mathbf{A}) + \operatorname{tr}(\mathbf{B}'\mathbf{B}) - 2\operatorname{tr}(\mathbf{A}'\mathbf{B}). \end{aligned} \tag{3.37}$$

Mainittakoon vielä seuraava epäyhtälö:

$$[\operatorname{tr}(\mathbf{A}'\mathbf{B})]^2 \leq \operatorname{tr}(\mathbf{A}'\mathbf{A}) \cdot \operatorname{tr}(\mathbf{B}'\mathbf{B}). \tag{3.38}$$

Epäyhtälö (3.38) seuraa Cauchyn–Schwarzin epäyhtälöstä

$$(\mathbf{u}'\mathbf{v})^2 \leq \mathbf{u}'\mathbf{u} \cdot \mathbf{v}'\mathbf{v}, \tag{3.39}$$

kun huomataan, että voimme valita $\mathbf{u} = \operatorname{vec}(\mathbf{A})$, $\mathbf{v} = \operatorname{vec}(\mathbf{B})$. Milloin (3.38):ssä on voimassa yhtäsuuruus? □

3.4 Lineaarinen yhtälöryhmä

Lineaarinen yhtälöryhmä

$$\begin{aligned} 2x_1 + 3x_2 &= 6 \\ 4x_1 + 5x_2 &= 7 \end{aligned} \tag{3.40}$$

voidaan esittää matriisien avulla seuraavasti:

$$\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 6 \\ 7 \end{pmatrix}, \quad \mathbf{A}\mathbf{x} = \mathbf{y}. \tag{3.41}$$

Yleisessä tapauksessa voimme tarkastella m tuntemattoman ja n yhtälön yhtälöryhmää

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= y_2 \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= y_n \end{aligned} \tag{3.42}$$

joka matriisien avulla voidaan esittää seuraavasti:

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} x_1 + \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} x_2 + \cdots + \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix} x_m = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \tag{3.43a}$$

$$\mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \cdots + \mathbf{a}_m x_m = \mathbf{y}, \tag{3.43b}$$

$$(\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \mathbf{y}, \tag{3.43c}$$

$$\mathbf{A}\mathbf{x} = \mathbf{y}. \tag{3.43d}$$

Tilanne (3.43d):ssä on siis se, että matriisi $\mathbf{A}$ $(n \times m)$ ja vektori $\mathbf{y}$ $(n \times 1)$ on annettu ja ongelmana on löytää sellainen vektori $\mathbf{x}$ $(m \times 1)$, että yhtälö $\mathbf{A}\mathbf{x} = \mathbf{y}$ toteutuu. Tunnetusti lineaarisella yhtälöryhmällä on joko

(a) ei yhtään ratkaisua, tai

(b) yksikäsitteinen ratkaisu, tai

(c) äärettömän monta ratkaisua.

Sarakevaruuden avulla voimme luonnehtia ratkaisujen olemassaoloa siten, että

> yhtälöllä $\mathbf{A}\mathbf{x} = \mathbf{y}$ on ratkaisu (ts. yhtälö on *ratkeava*) täsmälleen silloin kun $\mathbf{y}$ kuuluu $\mathbf{A}$:n sarakevaruuteen $\mathscr{C}(\mathbf{A})$.

Jos ratkaisu on olemassa, niin se on yksikäsitteinen vain jos $\mathbf{A}$:n sarakkeet ovat vapaat. Tällöin nimittäin $\mathbf{A}$:n sarakkeet muodostavat $\mathscr{C}(\mathbf{A})$:n kannan ja $\mathbf{x}$:n komponentit muodostavat $\mathbf{y}$:n koordinaatit tämän kannan suhteen. Kunkin vektorin koordinaatit annetun kannan suhteen ovat yksikäsitteiset.

Jos $\mathbf{A}$:n sarakkeet ovat vapaat, niin silloin myös neliömatriisin $\mathbf{A}'\mathbf{A}$ sarakkeet ovat vapaat ja sillä on olemassa käänteismatriisi $(\mathbf{A}'\mathbf{A})^{-1}$. Kertomalla (ratkeava) yhtälö $\mathbf{A}\mathbf{x} = \mathbf{y}$ vasemmalta matriisilla $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ saamme ratkaisuksi $\mathbf{x} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$. Jos $\mathbf{A}$ on neliömatriisi, jonka sarakkeet ovat vapaat, niin tietenkin $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$.

3.4.1 Yleistetty käänteismatriisi

Olkoon $\mathbf{A}$ annettu $n \times m$-matriisi ja $\mathbf{G}$ olkoon jokin $m \times n$-matriisi joka toteuttaa ehdon

$$\mathbf{AGA} = \mathbf{A}\,. \tag{mp1}$$

Tällöin matriisia $\mathbf{G}$ sanotaan $\mathbf{A}$:n *yleistetyksi käänteismatriisiksi* ja siitä voidaan käyttää merkintää $\mathbf{A}^-$. Yleistetty käänteismatriisi ei välttämättä ole yksikäsitteinen – se on aina olemassa, mutta se on yksikäsitteinen vain jos $\mathbf{A}^{-1}$ on olemassa. Kaikkien yleistettyjen käänteismatriisien joukkoa merkitsemme symbolilla $\{\mathbf{A}^-\}$. Jos yhtälö $\mathbf{Ax} = \mathbf{y}$ on ratkeava (eli jos vektori $\mathbf{y}$ kuuluu $\mathbf{A}$:n sarakeavaruuteen), niin yksi sen ratkaisu on

$$\mathbf{Gy}\,. \tag{3.44}$$

Ratkeavan yhtälön $\mathbf{Ax} = \mathbf{y}$ *yleinen ratkaisu* eli kaikkien ratkaisujen joukko saadaan lausekkeesta

$$\mathbf{Gy} + (\mathbf{I}_m - \mathbf{GA})\mathbf{z}\,, \tag{3.45}$$

missä $\mathbf{G}$ on jokin kiinteä $\mathbf{A}$:n yleistetty käänteismatriisi ja $\mathbf{z}$ on mielivaltainen m elementin vektori. Kun siis $\mathbf{z}$:aa varioidaan, saadaan yhtälön $\mathbf{Ax} = \mathbf{y}$ kaikki ratkaisut generoiduksi. Toisaalta on osoitettavissa, että lauseke $\mathbf{Gy}$ generoi kaikki (ratkeavan) yhtälön $\mathbf{Ax} = \mathbf{y}$ ratkaisut, kun $\mathbf{G}$ käy läpi kaikki mahdolliset $\mathbf{A}$:n yleistetyt käänteismatriisit.

Lukija saattaa muistaa lineaarialgebrasta, että ratkeavan yhtälön $\mathbf{Ax} = \mathbf{y}$ yleinen ratkaisu $\mathbf{x}_0$ voidaan myös esittää muodossa

$$\mathbf{x}_0 = \{\text{yhtälön } \mathbf{Ax} = \mathbf{y} \text{ jokin ratkaisu}\} + \{\text{yhtälön } \mathbf{Ax} = \mathbf{0} \text{ yleinen ratkaisu}\}\,. \tag{3.46}$$

Yleistettyyn käänteismatriisiin liittyvät neljä *Mooren–Penrosen ehtoa*:

$$\begin{aligned} &(\text{mp1})\ \mathbf{AA}^+\mathbf{A} = \mathbf{A}\,, && (\text{mp2})\ \mathbf{A}^+\mathbf{AA}^+ = \mathbf{A}^+, \\ &(\text{mp3})\ \mathbf{AA}^+ = (\mathbf{AA}^+)', && (\text{mp4})\ \mathbf{A}^+\mathbf{A} = (\mathbf{A}^+\mathbf{A})'. \end{aligned} \tag{3.47}$$

Ehdon (mp$i$) toteuttavaa matriisia sanotaan joskus $\{i\}$-inverssiksi, jos halutaan täsmentää minkä ehdon ko. matriisi toteuttaa näistä neljästä. Erityisesti sovitaan sanonnoista

$$(\text{mp1})\ \&\ (\text{mp2}) \iff \mathbf{G} \in \{\mathbf{A}^-_{12}\}\text{: refleksiivinen g-inverssi,}$$

$$\text{kaikki 4 ehtoa} \iff \mathbf{G} = \mathbf{A}^+\text{: Mooren–Penrosen inverssi.}$$

Mooren–Penrosen inverssi on yksikäsitteinen ja se on aina olemassa. Jos $\mathbf{A}$ on epäsingulaarinen (neliömatriisi), niin $\mathbf{A}^+ = \mathbf{A}^{-1}$. Harjoitustehtävänä on osoittaa, että jos matriiisin $\mathbf{A}_{n\times m}$ sarakkeet ovat vapaat, niin $\mathbf{A}$:n Mooren–Penrosen inverssi on

$$\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'. \tag{3.48}$$

Lävistäjämatriisin $\mathbf{A}$ Mooren–Penrosen inverssi $\mathbf{A}^+$ saadaan kun nollasta poikkevat lävistäjäelementit korvataan käänteisluvuillaan (ja nollat pysyvät nollina). Esimerkiksi:

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & c \end{pmatrix}^{+} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/a & 0 & 0 \\ 0 & 0 & 1/b & 0 \\ 0 & 0 & 0 & 1/c \end{pmatrix}. \tag{3.49}$$

Yleistettyihin käänteismatriiseihin palataan yksityiskohtaisemmin luvussa 10 (s. 335).

3.4.2 Yleistetty käänteismatriisi & normaaliyhtälö

Tarkastellaan seuraavaa yhtälöryhmää, jossa on y-havaintoja kolmesta eri ryhmästä:

$$\begin{aligned}
\text{ryhmä 1:}\quad & y_{11} = \beta_0 + \beta_1 && + \varepsilon_{11} \\
& y_{12} = \beta_0 + \beta_1 && + \varepsilon_{12} \\
\text{ryhmä 2:}\quad & y_{21} = \beta_0 \qquad + \beta_2 && + \varepsilon_{21} \\
& y_{22} = \beta_0 \qquad + \beta_2 && + \varepsilon_{22} \\
& y_{23} = \beta_0 \qquad + \beta_2 && + \varepsilon_{23} \\
\text{ryhmä 3:}\quad & y_{31} = \beta_0 \qquad\qquad + \beta_3 && + \varepsilon_{31} \\
& y_{32} = \beta_0 \qquad\qquad + \beta_3 && + \varepsilon_{32}
\end{aligned} \tag{3.50}$$

Tällöin on kyseessä lineaarinen malli

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{3.51}$$

jota voidaan käyttää kolmen ryhmän varianssianalyysissä. Mallimatriisi $\mathbf{X}$ on

$$\mathbf{X} = (\mathbf{1}_n : \mathbf{x}_1 : \mathbf{x}_2 : \mathbf{x}_3) = \begin{pmatrix} \mathbf{1}_a & \mathbf{1}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{1}_b & \mathbf{0} & \mathbf{1}_b & \mathbf{0} \\ \mathbf{1}_c & \mathbf{0} & \mathbf{0} & \mathbf{1}_c \end{pmatrix}, \quad a + b + c = n\,, \tag{3.52}$$

ja vektori $\mathbf{y}$ ositettu siten että

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{pmatrix}, \quad \mathbf{y}_1 = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1a} \end{pmatrix}, \quad \mathbf{y}_2 = \begin{pmatrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2b} \end{pmatrix}, \quad \mathbf{y}_3 = \begin{pmatrix} y_{31} \\ y_{32} \\ \vdots \\ y_{3c} \end{pmatrix}, \tag{3.53}$$

missä $\mathbf{y}_i$ sisältää i. ryhmän havainnot. Parametrivektorin $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ pienimmän neliösumman estimaatti $\hat{\boldsymbol{\beta}}$ ratkaistaan *normaaliyhtälöstä*:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}\,. \tag{3.54}$$

Nyt $\mathbf{X}$:n sarakkeet eivät ole vapaat, koska 1. sarake on muiden summa, joten $\mathbf{X}'\mathbf{X}$:n käänteismatriisi ei ole olemassa. Tällöin

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{1}_a' & \mathbf{1}_b' & \mathbf{1}_c' \\ \mathbf{1}_a' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{1}_b' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{1}_c' \end{pmatrix} \begin{pmatrix} \mathbf{1}_a & \mathbf{1}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{1}_b & \mathbf{0} & \mathbf{1}_b & \mathbf{0} \\ \mathbf{1}_c & \mathbf{0} & \mathbf{0} & \mathbf{1}_c \end{pmatrix} = \begin{pmatrix} n & a & b & c \\ a & a & 0 & 0 \\ b & 0 & b & 0 \\ c & 0 & 0 & c \end{pmatrix}. \tag{3.55}$$

Aukikirjoitettuna normaaliyhtälö on:

$$\begin{aligned} n\hat{\beta}_0 + a\hat{\beta}_1 + b\hat{\beta}_2 + c\hat{\beta}_3 &= \mathbf{1}_n'\mathbf{y} \\ a\hat{\beta}_0 + a\hat{\beta}_1 \qquad\qquad\qquad &= \mathbf{1}_a'\mathbf{y}_1 \\ b\hat{\beta}_0 \qquad\quad + b\hat{\beta}_2 \qquad\quad &= \mathbf{1}_b'\mathbf{y}_2 \\ c\hat{\beta}_0 \qquad\qquad\qquad + c\hat{\beta}_3 &= \mathbf{1}_c'\mathbf{y}_3 \end{aligned} \tag{3.56}$$

Normaaliyhtälöllä (3.56) on ratkaisu mutta ratkaisu ei ole yksikäsitteinen: ratkaisuja on ääretön määrä. Searlen (1982, s. 398) mukaan on kyseenalaista kutsua ei-yksikäsitteistä $\hat{\boldsymbol{\beta}}$:a estimaatiksi; Searle puhuu vain ”normaaliyhtälön ratkaisusta” tässä tapauksessa. Yksi ratkaisu (3.56):lle saadaan valitsemalla $\hat{\beta}_0 = 0$, jolloin

$$\hat{\beta}_1 = \bar{y}_1\,, \quad \hat{\beta}_2 = \bar{y}_2\,, \quad \hat{\beta}_3 = \bar{y}_3\,. \tag{3.57}$$

Matriisin $\mathbf{X}'\mathbf{X}$ yksi yleistetty käänteismatriisi on

$$(\mathbf{X}'\mathbf{X})^- = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/a & 0 & 0 \\ 0 & 0 & 1/b & 0 \\ 0 & 0 & 0 & 1/c \end{pmatrix}, \tag{3.58}$$

jonka avulla saadaan (3.57):n mukainen ratkaisu

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y} = (0, \bar{y}_1, \bar{y}_2, \bar{y}_3)'. \tag{3.59}$$

Mitkä ehdot (3.58):n mukainen $(\mathbf{X}'\mathbf{X})^-$ toteuttaa ehdoista (mp$i$)?

3.5 Lävistäjämatriisilla kertominen

Olkoon $\mathbf{D}_m \in \mathbb{R}^{m\times m}$ lävistäjämatriisi ja $\mathbf{A} \in \mathbb{R}^{n\times m}$. Silloin havaitaan, että

$$\begin{aligned}\mathbf{A}\mathbf{D}_m &= (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m)\mathbf{D}_m \\ &= \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nm} \end{pmatrix} \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_m \end{pmatrix} \\ &= \begin{pmatrix} a_{11}d_1 & a_{12}d_2 & \ldots & a_{1m}d_m \\ a_{21}d_1 & a_{22}d_2 & \ldots & a_{2m}d_m \\ \vdots & \vdots & & \vdots \\ a_{n1}d_1 & a_{n2}d_2 & \ldots & a_{nm}d_m \end{pmatrix} = (d_1\mathbf{a}_1 : d_2\mathbf{a}_2 : \ldots : d_m\mathbf{a}_m)\,. \end{aligned} \tag{3.60}$$

Vastaavasti jos $\mathbf{D}_n$ on $n \times n$-lävistäjämatriisi, saadaan

$$\begin{aligned}\mathbf{D}_n\mathbf{A} = \mathbf{D}_n \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix} &= \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nm} \end{pmatrix} \\ &= \begin{pmatrix} d_1a_{11} & d_1a_{12} & \ldots & d_1a_{1m} \\ d_2a_{21} & d_2a_{22} & \ldots & d_2a_{2m} \\ \vdots & \vdots & & \vdots \\ d_na_{n1} & d_na_{n2} & \ldots & d_na_{nm} \end{pmatrix} = \begin{pmatrix} d_1\mathbf{a}'_{(1)} \\ d_2\mathbf{a}'_{(2)} \\ \vdots \\ d_n\mathbf{a}'_{(n)} \end{pmatrix}. \end{aligned} \tag{3.61}$$

Siis kerrottaessa $\mathbf{A}$ oikealta lävistäjämatriisilla $\mathbf{D}_m$ tulevat $\mathbf{A}$:n sarakkeet kerrotuiksi $\mathbf{D}_m$:n vastaavilla elementeillä; vasemmalta kerrottaessa $\mathbf{A}$:n vaakarivit tulevat kerrotuiksi $\mathbf{D}_n$:n vastaavilla elementeillä. Yhteenvetona siis:

$$\mathbf{A}\mathbf{D}_m = (d_1\mathbf{a}_1 : d_2\mathbf{a}_2 : \ldots : d_m\mathbf{a}_m), \quad \mathbf{D}_n\mathbf{A} = \begin{pmatrix} d_1\mathbf{a}'_{(1)} \\ d_2\mathbf{a}'_{(2)} \\ \vdots \\ d_n\mathbf{a}'_{(n)} \end{pmatrix}, \tag{3.62a}$$

$$\mathbf{D}_n\mathbf{A}_{n\times m}\mathbf{D}_m = \begin{pmatrix} d_1a_{11}d_1 & d_1a_{12}d_2 & \ldots & d_1a_{1m}d_m \\ d_2a_{21}d_1 & d_2a_{22}d_2 & \ldots & d_2a_{2m}d_m \\ \vdots & \vdots & & \vdots \\ d_na_{n1}d_1 & d_na_{n2}d_2 & \ldots & d_na_{nm}d_m \end{pmatrix} = \{d_id_ja_{ij}\}\,. \tag{3.62b}$$

Havaitsemme myös, että

$$\mathbf{A}\mathbf{D}_m\mathbf{A}' = d_1\mathbf{a}_1\mathbf{a}_1' + d_2\mathbf{a}_2\mathbf{a}_2' + \cdots + d_m\mathbf{a}_m\mathbf{a}_m' \,, \tag{3.63a}$$

$$\mathbf{A}'\mathbf{D}_n\mathbf{A} = d_1\mathbf{a}_{(1)}\mathbf{a}_{(1)}' + d_2\mathbf{a}_{(2)}\mathbf{a}_{(2)}' + \cdots + d_n\mathbf{a}_{(n)}\mathbf{a}_{(n)}' \,. \tag{3.63b}$$

3.5.1 Sarakkeiden skaalaus & kosinit

Voimme skaalata matriisin $\mathbf{A}_{n\times m}$ sarakkeet 1:n pituisiksi seuraavasti. Ensinnäkin muistamme, että $\mathbf{A}'\mathbf{A}$:n lävistäjällä ovat $\mathbf{A}$:n sarakkeiden pituuksien neliöt:

$$\operatorname{diag}(\mathbf{A}'\mathbf{A}) = \operatorname{diag}(\|\mathbf{a}_1\|^2, \|\mathbf{a}_2\|^2, \dots, \|\mathbf{a}_m\|^2) \,. \tag{3.64}$$

Merkitään

$$[\operatorname{diag}(\mathbf{A}'\mathbf{A})]^{1/2} = \operatorname{diag}(\|\mathbf{a}_1\|, \|\mathbf{a}_2\|, \dots, \|\mathbf{a}_m\|) \,, \tag{3.65}$$

ja (olettaen että $\mathbf{a}_i \neq \mathbf{0}$, $i = 1, \dots, m$)

$$[\operatorname{diag}(\mathbf{A}'\mathbf{A})]^{-1/2} = \begin{pmatrix} 1/\|\mathbf{a}_1\| & 0 & \dots & 0 \\ 0 & 1/\|\mathbf{a}_2\| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\|\mathbf{a}_m\| \end{pmatrix} . \tag{3.66}$$

Tällöin matriisin $\mathbf{A}$ sarakkeiden skaalaus 1:n pituisiksi on esitettävissä matriisitulona

$$\underline{\mathbf{A}} := \mathbf{A}[\operatorname{diag}(\mathbf{A}'\mathbf{A})]^{-1/2} = (\underline{\mathbf{a}}_1 : \underline{\mathbf{a}}_2 : \dots : \underline{\mathbf{a}}_m) \,. \tag{3.67}$$

Nyt

$$\cos(\mathbf{a}_i, \mathbf{a}_j) = \frac{\mathbf{a}_i'\mathbf{a}_j}{\sqrt{\mathbf{a}_i'\mathbf{a}_i \cdot \mathbf{a}_j'\mathbf{a}_j}} = \frac{\mathbf{a}_i'}{\sqrt{\mathbf{a}_i'\mathbf{a}_i}} \cdot \frac{\mathbf{a}_j'}{\sqrt{\mathbf{a}_j'\mathbf{a}_j}} = \underline{\mathbf{a}}_i'\underline{\mathbf{a}}_j = \cos(\underline{\mathbf{a}}_i, \underline{\mathbf{a}}_j) \,, \tag{3.68}$$

joten $\mathbf{A}$:n sarakkeiden välisten kulmien kosinit ovat matriisin $\underline{\mathbf{A}}'\underline{\mathbf{A}}$ elementtejä. Toisin sanoen: vektoreiden $\mathbf{a}_i$ välisten kulmien kosinit on kätevä laskea siten, että

(COS) (a) ensin skaalataan vektorit $\mathbf{a}_i$ ykkösen pituisiksi: tuloksena $\underline{\mathbf{a}}_1$, $\underline{\mathbf{a}}_2$, $\dots$, $\underline{\mathbf{a}}_m$,

(b) sitten lasketaan sisätulot $\underline{\mathbf{a}}_i'\underline{\mathbf{a}}_j$: tuloksena $\cos(\mathbf{a}_i, \mathbf{a}_j)$.

3.5.2 Korrelaatio-, kovarianssi- ja tulosummamatriisi

Esimerkki 3.5. Tarkastellaan korrelaatio-, kovarianssi- ja tulosummamatriisin välistä yhteyttä kolmen muuttujan aineistossa:

$$\mathbf{T} = \widetilde{\mathbf{U}}'\widetilde{\mathbf{U}} = \mathbf{U}'(\mathbf{I}-\mathbf{J})\mathbf{U} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} = \operatorname{ssp}(\mathbf{U}),\ \text{tulosummamatriisi},$$

$$\mathbf{S} = \frac{1}{n-1}\mathbf{T} = \begin{pmatrix} s_1^2 & s_{12} & s_{13} \\ s_{21} & s_2^2 & s_{23} \\ s_{31} & s_{32} & s_3^2 \end{pmatrix} = \operatorname{cov}_{\mathrm{d}}(\mathbf{U}),\ \text{kovarianssimatriisi},$$

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix} = \operatorname{cor}_{\mathrm{d}}(\mathbf{U}),\ \text{korrelaatiomatriisi.}$$

Merkitään edelleen

$$\mathbf{T}_\delta = \operatorname{diag}(\mathbf{T}) = \begin{pmatrix} t_{11} & 0 & 0 \\ 0 & t_{22} & 0 \\ 0 & 0 & t_{33} \end{pmatrix}, \quad \mathbf{S}_\delta = \operatorname{diag}(\mathbf{S}) = \begin{pmatrix} s_1^2 & 0 & 0 \\ 0 & s_2^2 & 0 \\ 0 & 0 & s_3^2 \end{pmatrix}, \tag{3.69}$$

$$\mathbf{S}_\delta^{1/2} = \begin{pmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{pmatrix}. \tag{3.70}$$

Tällöin

$$\begin{aligned} \mathbf{R} &= \mathbf{S}_\delta^{-1/2}\mathbf{S}\mathbf{S}_\delta^{-1/2} = \mathbf{T}_\delta^{-1/2}\mathbf{T}\mathbf{T}_\delta^{-1/2} \\ &= \begin{pmatrix} 1/s_1 & 0 & 0 \\ 0 & 1/s_2 & 0 \\ 0 & 0 & 1/s_3 \end{pmatrix} \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix} \begin{pmatrix} 1/s_1 & 0 & 0 \\ 0 & 1/s_2 & 0 \\ 0 & 0 & 1/s_3 \end{pmatrix} \\ &= [\operatorname{diag}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}})]^{-1/2}\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}[\operatorname{diag}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}})]^{-1/2} = \tilde{\widetilde{\mathbf{U}}}'\tilde{\widetilde{\mathbf{U}}}, \end{aligned} \tag{3.71}$$

ja

$$\mathbf{S} = \mathbf{S}_\delta^{1/2}\mathbf{R}\mathbf{S}_\delta^{1/2} = \begin{pmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{pmatrix} \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix} \begin{pmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{pmatrix}. \tag{3.72}$$

Merkintä $\mathbf{D}^{1/2}$ tarkoittaa matriisin $\mathbf{D}$ neliöjuurta eli matriisia, jolla on ominaisuus $\mathbf{D}^{1/2}\mathbf{D}^{1/2} = \mathbf{D}$ (ks. s. 198). Kun $\mathbf{D}$ on lävistäjämatriisi, on $\mathbf{D}^{1/2} = \operatorname{diag}(\sqrt{d_{11}}, \dots, \sqrt{d_{nn}})$.

Korostettakoon vielä kaavan

$$\begin{aligned}\mathrm{cor_d}(\mathbf{U}) = \mathbf{R} = \tilde{\tilde{\mathbf{U}}}'\tilde{\tilde{\mathbf{U}}} &= \big[\mathrm{diag}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}})\big]^{-1/2}\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}\big[\mathrm{diag}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}})\big]^{-1/2} \\ &= [\mathrm{diag}(\mathbf{U}'\mathbf{C}\mathbf{U})]^{-1/2}\mathbf{U}'\mathbf{C}\mathbf{U}[\mathrm{diag}(\mathbf{U}'\mathbf{C}\mathbf{U})]^{-1/2} \qquad (3.73)\end{aligned}$$

merkitystä; (3.73):ssä siis

- $\widetilde{\mathbf{U}} = \mathbf{C}\mathbf{U}$ = keskistetty $\mathbf{U}$
- $\tilde{\tilde{\mathbf{U}}} = \widetilde{\mathbf{U}}\big[\mathrm{diag}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}})\big]^{-1/2}$ = sarakkeiltaan 1:n pituiseksi skaalattu keskistetty $\mathbf{U}$.

Siis: muuttujavektoreiden $\mathbf{u}_1, \dots, \mathbf{u}_p$ väliset korrelaatiokertoimet voidaan laskea siten, että

(COR) (1) ensin keskistetään vektorit $\mathbf{u}_i$: tuloksena $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_p$,
(2) skaalataan vektorit $\tilde{\mathbf{u}}_i$ 1:n pituisiksi: tuloksena $\tilde{\tilde{\mathbf{u}}}_1, \dots, \tilde{\tilde{\mathbf{u}}}_p$,
(3) lopuksi lasketaan niiden väliset sisätulot $\tilde{\tilde{\mathbf{u}}}_i'\tilde{\tilde{\mathbf{u}}}_j$: tuloksena $\mathrm{cor_d}(\mathbf{u}_i, \mathbf{u}_j)$.

Kaavojen (COS) ja (COR) välinen ero on juuri *keskistämisessä*: kaavassa (COS) *ei* tehdä keskistystä, mikä taas (COR):ssa on ensimmäinen toimenpide.

```
  -  - SURVO MM  Mon Sep 13 17:26:31 2010          C:\SJP\D\   2000  100 0
*
* EXAMPLE A:  Correlation is 1 BUT cosine is 0.
*             IN THE WORLD OF COLLINEARITY ...
* Belsley (1991, p.20): Conditioning Diagnostics. Wiley, New York.
*
*MATRIX A
*///     U     V
* 1      1    -1
* 2      1    -1+eps
* 3      a     a
*
*MAT SAVE A         /   eps=0.01   a=SQRT(2)
*MAT COR=NRM(CENTER(A))'*NRM(CENTER(A))
*MAT LOAD COR,CUR+1  / cor(U,V) = 0.999994
*MATRIX COR
*NRM(CENTER(A))'*NRM(CENTER(A))
*///             U        V
*U        1.000000 0.999994
*V        0.999994 1.000000
*
*MAT COS=NRM(A)'*NRM(A)
*MAT LOAD COS,CUR+1  /  cos(U,V) = 0.002506
*MATRIX COS
*NRM(A)'*NRM(A)
*///             U        V
*U        1.000000 0.002506
*V        0.002506 1.000000
*                                          (MK10-005, MKT-03-COL1)
```

Kuvio 3.1. *Opetus*: korrelaatio voi olla lähellä ykköstä, mutta kosini lähellä nollaa.

```
 -   - SURVO MM  Mon Sep 13 17:28:21 2010            C:\SJP\D\   2000  100 0
*
* EXAMPLE B: Correlation is 0 BUT cosine is 0.999999999
*            IN THE WORLD OF COLLINEARITY ...
*  Belsley (1991, p.20): Conditioning Diagnostics. Wiley, New York.
*     Let  X and Y be centered vectors such that X'Y = 0.
*     Let us define variables U and V so that
*     U=1+eps*X    V=1+eps*Y  where eps is a real number.
*  (a) What is cor(U,V)?
*          This is 0 for all nonzero eps, since cor(X,Y) = 0
*  (b) What about cos(U,V)?
*      NOTE. If eps = 0, then cor(U,V) = 0/0  but cos(U,V) = 1
*
*DATA KOED a,b,a-1,d
d  1    11   11     111.111111     111.111111
*MATRIX  KOEM
*///      X    Y       U              V
a  1      1    1       1.001000       1.001000
*  2     -1    1       0.999000       1.001000
b  3      0   -2       1.000000       0.998000
*
*VAR U,V TO KOED     / Here U=1+eps*X    V=1+eps*Y      eps=0.001
*MAT SAVE KOEM       / Between the lines a and b there is a mtx (TESTM)
*MAT COR=NRM(CENTER(KOEM))'*NRM(CENTER(KOEM))  /  and a datamtx (TESTD)
*MAT LOAD COR,CUR+1
*MATRIX COR
*NRM(CENTER(KOEM))'*NRM(CENTER(KOEM))
*///             X        Y        U        V
*X        1.000000 0.000000 1.000000 0.000000
*Y        0.000000 1.000000 0.000000 1.000000
*U        1.000000 0.000000 1.000000 0.000000
*V        0.000000 1.000000 0.000000 1.000000
*
*MAT COSINIT=NRM(KOEM)'*NRM(KOEM)  / NRM scales the lengths of columns
*MAT LOAD COSINIT,CUR+1                                          (to 1)
*MATRIX COSINIT
*NRM(KOEM)'*NRM(KOEM)
*///             X        Y        U        V
*X        1.000000 0.000000 0.008165 0.000000
*Y        0.000000 1.000000 0.000000 0.014141
*U        0.008165 0.000000 1.000000 0.999867
*V        0.000000 0.014141 0.999867 1.000000
*                                         (MK10-005, MKT-03-COL2)
```

Kuvio 3.2. *Opetus*: korrelaatio voi olla nolla, mutta kosini lähellä ykköstä!

```
  -   - SURVO MM  Fri Sep 24 18:48:55 2010          C:\SJP\D\   2000  100 0
*
*               KOVARIANSSIMATRIISIN LASKEMINEN
*DATA KOE
*  NIMI      x   y
*  Kalle     1   1
*  Ville     4   1
*  Maija     4   4
*
*CORR KOE,CUR+1   /  VARS=x,y
*Means, std.devs and correlations of KOE  N=3
*Variable  Mean        Std.dev.
*x         3.000000    1.732051
*y         2.000000    1.732051
*Correlations:
*             x       y
* x            1.0000  0.5000
* y            0.5000  1.0000
*
*MAT R!=CORR.M       / CORRin yhteydessä syntyy automaattisesti
*MAT F!=MSN.M        / 2 matriisia CORR.M ja MSN.M
*MAT LOAD F          / MSN.M sisältää keskiarvot, hajonnat ja hav lkm:t
*MATRIX F
*///         mean   stddev        N
*x        3.000000 1.732051 3.000000
*y        2.000000 1.732051 3.000000
*
*MAT STDEV!=DV(F(*,2))   /  DV tekee pystyvektorista diag.matriisin
*MAT COV=STDEV*R*STDEV   /  STDEV on hajontojen muod. diag mtx
*MAT LOAD COV,CUR+1      /  COV on siis kovarianssimatriisi!
*MATRIX COV
*STDEV*R*STDEV
*///            x        y
*x        3.000000 1.500000
*y        1.500000 3.000000
*
*   Lasketaan kovmtx suoraan matriisioperaatioin:
*MATRIX U
*///        x   y
*Kalle      1   1
*Ville      4   1
*Maija      4   4
*
*MAT SAVE U
*MAT DIM U /* rowU=3 colU=2
*MAT REM n=rowU
*MAT C!=IDN(n,n)-CON(n,n,1/n) / C on keskistäjämatriisi
*MAT COV1!=U'*C*U/(n-1)        / COV1 = kovmtx = COV
*MAT COV2!=U'*CENTER(U)/(n-1)  / COV2 = kovmtx = COV
*MAT LOAD COV1
*MATRIX COV1
*///            x        y
*x        3.000000 1.500000
*y        1.500000 3.000000
*                                             (MK10-006, MKT-03-KOV)
```

Kuvio 3.3. Kovarianssimatriisin **S** laskeminen.

```
 -   - SURVO MM  Sat Sep 25 10:56:57 2010           C:\SJP\D\   2000  100 0
 *                   BAD PRODUCTS OF GOOD MATRICES
 *      Paul Halmos (1991), Linear and Multilinear Algebra, pp. 1-20.
 *
 *      Is the product of two symmetric matrices always symmetric?
 *      The question is preposterous -- the product of two symmetric
 *   matrices is extremely unlikely to be symmetric. Write down almost any
 *   two symmetric matrices, say for instance ...  ( ... this is all from
 *   the article of Halmos ...)
 *
 *MATRIX A
 *///   c1   c2
 *r1    1    2
 *r2    2    3
 *
 *MATRIX B
 *///   c1   c2
 *r1    4    5
 *r2    5    6
 *
 *MAT C=A*B
 *MATRIX C
 *///            c1       c2
 *r1             14       17
 *r2             23       28
 *    ... and so C is not symmetric. Once you have the courage, it's no
 *    trouble to construct much simpler examples. One easy example is:
 *MATRIX A1
 *///   c1   c2
 *r1    1    0
 *r2    0    0
 *
 *MATRIX B1
 *///   c1   c2
 *r1    0    1
 *r2    1    0
 *
 *MAT C1=A1*B1
 *MATRIX C1
 *///            c1       c2
 *r1              0        1
 *r2              0        0
 *
 *MAT D1=B1*A1
 *MATRIX D1
 *///            c1       c2
 *r1              0        0
 *r2              1        0
 *       Granted that the product of two symmetric matrices can fail to be
 *   symmetric, it makes sense to ask which matrices can be such products
 *  -- which non-symmetric matrices are products of symmetric ones?
 *      The question belongs to a large class of interesting ones that are
 *   often non-trivial, questions that ask
 *               WHICH BAD MATRICES ARE PRODUCTS OF GOOD ONES ?
 *   Could
 *MATRIX E  ///
 *   0   0   0
 *   1   0   0
 *   1   1   0
 *
 *   for instance, possibly be written as a product of two symmetric
 *   matrices?                                  (MK10-007a, MKT-03-HAL1)
```

Kuvio 3.4. Bad products of good matrices, Halmos (1991).

```
 -  - SURVO MM  Sat Sep 25 10:58:14 2010           C:\SJP\D\   2000  100 0
*  And if the answer happens to be yes, if that non-symmetric
*  matrix is not bad enough to be a counterexample, how about
*  something thoroughly non-symmetric, something like
*MATRIX F  ///
*  1   0   0
*  2   3   0
*  4   5   6
*
* ... could F be a product of two symmetric ones?
*Consider symmetric E1 and E2:
*MATRIX E1  ///
*  0   0   0
*  0   0  -1
*  0  -1   0
*
*MATRIX E2  ///
* -1  -1  -1
* -1  -1   0
* -1   0   0
*
*MAT E12=E1*E2
*MATRIX E12
*///            c1        c2        c3      This is E
*r1              0         0         0        0   0   0
*r2              1         0         0        1   0   0
*r3              1         1         0        1   1   0
*
*  THEREFORE:   E1*E2 = E is a product of two symmetric matrices
*  ... and this equation may or may not be considered shocking but
*  in any event it does answer the first of the two questions raised abo
*  -- a mildly bad matrix is the product of two good ones.
*
*     Consider symmetric F1 and F2:
*MATRIX F1  ///
*   1     -1       1/5
*  -1      2     -28/15
*   1/5   -28/15   127/45
*
*MATRIX F2  ///
*   2908    3303   1980
*   3303    3753   2250
*   1980    2250   1350
*
*MAT F12=F1*F2
*MATRIX F12
*///             1        2        3
*  1      1.000000 0.000000 0.000000
*  2      2.000000 3.000000 0.000000
*  3      4.000000 5.000000 6.000000
*  This is just the matrix F (thoroughly nonsymmetric...)
*  The FRIGHTFUL equation F = F1*F2 is likely to be considered
*frightening  by most people, and it answers the second question:
*a thoroughly bad matrix can be a product of two symmetric good ones.
*  The factoring is far from unique; another possibility is:
*MATRIX G1  ///
*  25/22   -25/22     5/22
* -25/22    409/418   15/418
*  5/22     15/418   -25/418
*                                                  (MK10-007a, MKT-03-HAL2)
```

Kuvio 3.5. Halmos, jatkoa ...

```
 -   - SURVO MM  Sat Sep 25 10:59:30 2010           C:\SJP\D\    2000   100 0
*MATRIX G2  ///
*  103/5    25       132/5
*  25       31       30
*  132/5    30       18
*
*MAT G12=G1*G2
*MATRIX G12
*///            1        2        3
*  1        1.00000  0.00000  0.00000
*  2        2.00000  3.00000  0.00000
*  3        4.00000  5.00000  6.00000
*    Thus: F = G1*G2
*  The bottom line is:
*       EVERY MATRIX IS A PRODUCT OF TWO SYMMETRIC MATRICES !!!!
*
*     The proof has two parts, cogitation and calculation.
*  The cogitation needed is simple, but it relies on
*  a deep theorem, the deepest theorem of linear algebra.
*    The calculation is simple, but inspired -- it seems
*  to pull a rabbit out of a hat, and invites the question
*  of how a mere mortal could have thought to look for that rabbit.
*    The calculation involves the consideration of matrices of the form
*MATRIX  K
*/// c1  c2  c3  c4
*r1   0   0   0   a
*r2   1   0   0   b
*r3   0   1   0   c
*r4   0   0   1   d
*
*MATRIX  L
*///  c1  c2  c3  c4
*r1   b   c   d   -1
*r2   c   d   -1   0
*r3   d  -1   0    0
*r4  -1   0   0    0
*
*MAT E=K*L          /   a=2   b=3   c=4  d=5
*MATRIX E
*///            c1       c2       c3       c4
*r1             -2        0        0        0
*r2              0        4        5       -1
*r3              0        5       -1        0
*r4              0       -1        0        0
*
*MAT H=INV(L)       / E, L, H are symmetric
*MAT N=E*H          / N = (K*L)*INV(L) = K
*MATRIX N
*///            r1       r2       r3       r4
*r1        0.00000  0.00000  0.00000  2.00000
*r2        1.00000 -0.00000 -0.00000  3.00000
*r3        0.00000  1.00000 -0.00000  4.00000
*r4        0.00000  0.00000  1.00000  5.00000
*     This N is precisely K.
*  SO: the crucial point is to observe that matrix K can be expressed
*  as a product of two symmetric matrices: K = E*H [= (K*L)*INV(L)].
*  Matrices of the form K are called companion matrices,
*  and they are what the deepest theorem of linear algebra is about ...
*  (cf Halmos,p.3)                                  (MK10-007a, MKT-03-HAL3)
```

Kuvio 3.6. Halmos, jatkoa …

Luku 4

Käänteismatriisi

Sana "neliä" ei suinkaan liity ratsastukseen, vaan luvunlaskuun. Se täyttää numeraaleissa eli lukusanoissa sopivasti sen aukon, joka muuten jäisi tyhjäksi kolmosen ja vitosen väliin.

Hyttine (1998): Sana viikoksi, *Aamulehti*.

4.1 Määritelmä

Yksikkömatriisilla $\mathbf{I}_n$ on tietysti ominaisuus

$$\mathbf{A}_{n\times n}\mathbf{I}_n = \mathbf{I}_n\mathbf{A} = \mathbf{A}\,, \tag{4.1a}$$

joten matriisilla $\mathbf{I}$ on samantyyppinen rooli matriisilaskennassa kuin luvulla 1 reaaliluvuilla laskettaessa. Reaalilukujen tapauksessa voimme olla kiinnostuneita luvusta b, jolla on ominaisuus

$$a \cdot b = b \cdot a = 1\,, \tag{4.1b}$$

ja tuloshan on tietysti a:n käänteisluku $b = 1/a$ edellyttäen, että $a \neq 0$. Täten on luonnollista etsiä matriisia $\mathbf{A}^{-1}$, jolla on ominaisuus

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n = \mathbf{A}^{-1}\mathbf{A}\,. \tag{4.2}$$

Jos meillä on esimerkiksi lineaarinen ratkeava yhtälöryhmä

$$\mathbf{A}\mathbf{x} = \mathbf{y}\,, \tag{4.3}$$

niin kertomalla (4.3) vasemmalta $\mathbf{A}^{-1}$:llä saataisiin ratkaisuksi $\mathbf{A}^{-1}\mathbf{y}$. Todettakoon heti, että tälläista matriisia $\mathbf{A}^{-1}$ ei aina ole tietenkään olemassa – samoin kuin (4.3) ei ole välttämättä ratkeava.

Neliömatriisin $\mathbf{A}_{n\times n}$ käänteismatriisi määritellään matriisiksi $\mathbf{B}_{n\times n}$, jolla on ominaisuus

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n\,, \tag{4.4}$$

ja sitä merkitään symbolilla $\mathbf{A}^{-1}$. Jos $\mathbf{A}$:n käänteismatriisi on olemassa, matriisia $\mathbf{A}$ sanotaan *epäsingulaariseksi* (kääntyväksi); muuten $\mathbf{A}$ on singulaarinen. Käänteismatriisi on yksikäsitteinen ja

$$\mathbf{A}_{n\times n}^{-1} \text{ on olemassa} \iff \mathrm{r}(\mathbf{A}_{n\times n}) = n. \tag{4.5}$$

On myös osoitettavissa että

$$\mathbf{A}_{n\times n}\mathbf{B}_{n\times n} = \mathbf{I}_n \implies \mathbf{BA} = \mathbf{I}_n\,. \tag{4.6}$$

Mainittakoon tässä yhteydessä termit *vasemman-* ja *oikeanpuoleinen* käänteismatriisi:

$$\begin{aligned}
&\mathbf{A}_{n\times m}\mathbf{B}_{m\times n} = \mathbf{I}_n\text{: } \mathbf{B} \text{ on } \mathbf{A}\text{:n oikeanpuoleinen käänteismatriisi: } \mathbf{A}_R\,,\\
&\mathbf{B}_{m\times n}\mathbf{A}_{n\times m} = \mathbf{I}_m\text{: } \mathbf{B} \text{ on } \mathbf{A}\text{:n vasemmanpuoleinen käänteismatriisi: } \mathbf{A}_L\,.
\end{aligned}$$

Jos $\mathbf{A}$:n sarakkeet ovat vapaat, niin $(\mathbf{A}'\mathbf{A})^{-1}$ on olemassa ja tietenkin

$$(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \cdot \mathbf{A} = \mathbf{I}_m\,, \tag{4.7}$$

joten $\mathbf{G} := (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ on $\mathbf{A}$:n vasemmanpuoleinen käänteismatriisi. On huomattava, että $\mathbf{A}_L$ ei ole välttämättä yksikäsitteinen. Nimittäin tietenkin on voimassa seuraava tulos:

$$\mathrm{r}(\mathbf{A}'\mathbf{U}\mathbf{A}) = \mathrm{r}(\mathbf{A}_{n\times m}) = m \implies (\mathbf{A}'\mathbf{U}\mathbf{A})^{-1}\mathbf{A}'\mathbf{U} \cdot \mathbf{A} = \mathbf{I}_m\,, \tag{4.8}$$

joten myös $\mathbf{F} := (\mathbf{A}'\mathbf{U}\mathbf{A})^{-1}\mathbf{A}'\mathbf{U}$ on $\mathbf{A}$:n vasemmanpuoleinen käänteismatriisi. Jos lineaarinen ratkeava yhtälö $\mathbf{Ax} = \mathbf{y}$ (missä $\mathbf{A}$:n sarakkeet ovat vapaat) kerrotaan vasemmalta matriisilla $\mathbf{G} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, niin yhtälön (yksikäsitteiseksi) ratkaisuksi saadaan

$$\mathbf{x}_0 = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} = \mathbf{Gy} = \mathbf{A}^{+}\mathbf{y}\,; \tag{4.9}$$

matriisi $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ on $\mathbf{A}$:n Mooren–Penrosen inverssi $\mathbf{A}^{+}$. Ratkaisu $\mathbf{x}_0$ voidaan tietysti ilmaista myös $\mathbf{F}$:n avulla:

$$\mathbf{x}_0 = (\mathbf{A}'\mathbf{U}\mathbf{A})^{-1}\mathbf{A}'\mathbf{U}\mathbf{y} = \mathbf{Fy}\,. \tag{4.10}$$

Olkoon esimerkiksi

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} f & g \\ h & i \end{pmatrix}. \tag{4.11}$$

Jos $\mathbf{B}$ olisi $\mathbf{A}$:n käänteismatriisi eli $\mathbf{AB} = \mathbf{I}_2$, niin pitäisi olla mm. $f + h = 1$, mutta myös $f + h = 0$. Nämä yhtälöt eivät voi tietenkään yhtaikaa toteutua, joten $\mathbf{A}$:lla ei ole käänteismatriisia. Sen sijaan

$$\mathbf{F} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \implies \mathbf{F}^{-1} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}, \tag{4.12}$$

sillä

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}. \tag{4.13}$$

Matriisin $\mathbf{A}_{n\times n}$ käänteismatriisi on olemassa täsmälleen silloin kun

♠ $\mathbf{A}$:n determinantti on nollasta poikeava,

tai yhtäpitävästi silloin kun

♡ $\mathbf{A}$:n sarakkeet ovat vapaat eli $\mathbf{A}$:n aste on n.

Palautettakoon mieleen (taas kerran), että matriisin $\mathbf{A}_{n\times n}$ sarakkeet ovat vapaat, jos ja vain jos yhtälö $\mathbf{Ax} = \mathbf{0}$ toteutuu vain kun $\mathbf{x} = \mathbf{0}$ eli nolla-avaruuden avulla esitettynä (ks. s. 86)

$$\mathbf{A}\text{:n sarakkeet vapaat} \iff \mathscr{N}(\mathbf{A}) = \{\mathbf{0}\}\,. \tag{4.14}$$

Vastaavasti määriteltiin, että matriisin $\mathbf{A}$ sarakkeet ovat lineaarisesti riippuvat eli sidotut eli

$$\text{on olemassa vektori } \mathbf{x} \neq \mathbf{0} \text{ siten että } \mathbf{Ax} = \mathbf{0}\,. \tag{4.15}$$

4.1.1 Determinantti, kofaktori

Ensinnäkin 1×1-matriisin eli reaaliluvun determinantti on luku itse. Jos $\mathbf{A}$ on 2×2-matriisi,

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \text{ niin } \mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \tag{4.16}$$

missä

$$\det(\mathbf{A}) = |\mathbf{A}| = ad - bc = \mathbf{A}\text{:n determinantti.} \tag{4.17}$$

Yleisessä tapauksessa kun $\mathbf{A} = \{a_{ij}\}$, on

$$\mathbf{A}^{-1} = \{a^{ij}\} = \frac{1}{\det(\mathbf{A})}\{\operatorname{cof}(a_{ij})\}'\,, \text{ ts. } a^{ij} = \frac{1}{\det(\mathbf{A})}\operatorname{cof}(a_{ji})\,, \tag{4.18}$$

missä

$$\operatorname{cof}(a_{ij}) = (-1)^{i+j}\det(\mathbf{A}_{ij}) = \text{elementtiin } a_{ij} \text{ liittyvä } \textit{kofaktori}, \tag{4.19a}$$

$$\mathbf{A}_{ij} = \text{osamatriisi joka saadaan kun } \mathbf{A}\text{:sta poistetaan } i\text{. vaakarivi ja } j\text{. sarake}, \tag{4.19b}$$

$$\det(\mathbf{A}_{ij}) = \text{elementtiin } a_{ij} \text{ liittyvä } \textit{minori}. \tag{4.19c}$$

Matriisin $\mathbf{A}$ determinantti voidaan määritellä lausekkeena

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij}\operatorname{cof}(a_{ij}), \text{ kehitetty } i\text{. vaakarivin mukaan}, \tag{4.20a}$$

$$\det(\mathbf{A}) = \sum_{i=1}^{n} a_{ij}\operatorname{cof}(a_{ij}), \text{ kehitetty } j\text{. sarakkeen mukaan}. \tag{4.20b}$$

Kaavaa (4.20) sanotaan *Laplacen determinanttikehitelmäksi.* Esimerkiksi kun $n = 3$, s saadaan

$$\begin{aligned}\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - d\begin{vmatrix} b & c \\ h & i \end{vmatrix} + g\begin{vmatrix} b & c \\ e & f \end{vmatrix}.\end{aligned} \tag{4.21}$$

Jos $\mathbf{R}$ on 3×3-korrelaatiomatriisi, niin sen determinantti on

$$\begin{aligned}\det(\mathbf{R}) &= \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix} \\ &= \begin{vmatrix} 1 & r_{23} \\ r_{32} & 1 \end{vmatrix} - r_{12}\begin{vmatrix} r_{21} & r_{23} \\ r_{31} & 1 \end{vmatrix} + r_{13}\begin{vmatrix} r_{21} & 1 \\ r_{31} & r_{32} \end{vmatrix} \\ &= 1 - r_{23}^2 - r_{12}(r_{21} - r_{31}r_{23}) + r_{13}(r_{21}r_{32} - r_{31}) \\ &= 1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}\,.\end{aligned} \tag{4.22}$$

Vastaavasti 2×2-korrelaatiomatriisin tapauksessa on voimassa

$$\mathbf{R} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \implies \mathbf{R}^{-1} = \frac{1}{1 - r^2}\begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix}, \ \det(\mathbf{R}) = 1 - r^2. \tag{4.23}$$

Kovarianssimatriisin $\mathbf{S}_{2\times 2}$ käänteismatriisiksi ja determinantiksi saadaan

$$\begin{aligned}
\mathbf{S}^{-1} &= \frac{1}{|\mathbf{S}|}\begin{pmatrix} s_y^2 & -s_{xy} \\ -s_{21} & s_x^2 \end{pmatrix} = \frac{1}{|\mathbf{S}|}\begin{pmatrix} s_y^2 & -s_x s_y r \\ -s_x s_y r & s_x^2 \end{pmatrix} \\
&= \frac{1}{s_x^2 s_y^2(1-r^2)}\begin{pmatrix} s_y^2 & -s_x s_y r \\ -s_x s_y r & s_x^2 \end{pmatrix} \\
&= \frac{1}{1-r^2}\begin{pmatrix} \frac{1}{s_x^2} & \frac{-r}{s_x s_y} \\ \frac{-r}{s_x s_y} & \frac{1}{s_y^2} \end{pmatrix} := \begin{pmatrix} s^{xx} & s^{xy} \\ s^{yx} & s^{yy} \end{pmatrix}, \qquad (4.24a) \\
|\mathbf{S}| &= s_x^2 s_y^2 - s_{xy}^2 = s_x^2 s_y^2 - (s_x s_y r)^2 = s_x^2 s_y^2 (1-r^2) \le s_x^2 s_y^2 \,. \qquad (4.24b)
\end{aligned}$$

Determinantti voidaan määritellä yhtäpitävästi myös seuraavasti:

$$\det(\mathbf{A}) = \sum (-1)^{f(i_1,\dots,i_n)} a_{1i_1} a_{2i_2} \cdots a_{ni_n}, \qquad (4.25)$$

missä summaus tapahtuu lukujen $\{1,\dots,n\}$ permutaatioiden $\{i_1, \dots, i_n\}$ yli ja $f(i_1,\dots,i_n)$ on ns. paikanvaihtojen lukumäärä, joka tarvitaan, jotta $\{i_1,\dots,i_n\}$:stä saadaan $\{1,\dots,n\}$. Emme puutu tähän määritelmään täsmällisemmin, mutta voimme todeta, että

♠ determinantin esitys (4.25) muodostuu tulojen summista, joita on $n!$ kappaletta, ja joissa jokaisessa tulossa on n tekijää siten, että mukana on täsmälleen yksi elementti $\mathbf{A}$:n kultakin vaaka- ja pystyriviltä.

Esimerkki 4.1. Sisältäköön vektori $\mathbf{x} \in \mathbb{R}^n$ havaintoarvot muuttujasta x ja merkitään $\mathbf{X}_{n\times 2} = (\mathbf{1} : \mathbf{x})$. Matriisi $\mathbf{X}$ voidaan pitää tällöin mallimatriisina yhden selittäjän (+ vakiotermin) regressioanalyysissä. Osoitamme, että

$$\boxed{\begin{aligned}
\det(\mathbf{X}'\mathbf{X}) &= n\,\mathrm{SS}_x = n(n-1)s_x^2\,, \qquad (4.26a) \\
(\mathbf{X}'\mathbf{X})^{-1} &= \frac{1}{\mathrm{SS}_x}\begin{pmatrix} \sum x_i^2/n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \qquad (4.26b)
\end{aligned}}$$

missä siis $\mathrm{SS}_x = \sum_{i=1}^n (x_i - \bar{x})^2 = \mathbf{x}'(\mathbf{I}-\mathbf{J})\mathbf{x}$ ja $s_x^2 = \frac{1}{n-1}\mathrm{SS}_x$. Koska

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{1}' \\ \mathbf{x}' \end{pmatrix}(\mathbf{1} : \mathbf{x}) = \begin{pmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{x} \\ \mathbf{x}'\mathbf{1} & \mathbf{x}'\mathbf{x} \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}, \qquad (4.27)$$

on $\mathbf{X}'\mathbf{X}$:n determinantti

$$\begin{aligned}
\det(\mathbf{X}'\mathbf{X}) &= n\mathbf{x}'\mathbf{x} - (\mathbf{x}'\mathbf{1})(\mathbf{1}'\mathbf{x}) = n[\mathbf{x}'\mathbf{x} - (\mathbf{x}'\mathbf{1})\tfrac{1}{n}(\mathbf{1}'\mathbf{x})] \\
&= n\mathbf{x}'(\mathbf{I} - \mathbf{1}\tfrac{1}{n}\mathbf{1}')\mathbf{x} = n\mathbf{x}'(\mathbf{I}-\mathbf{J})\mathbf{x} = n\,\mathrm{SS}_x\,. \qquad (4.28)
\end{aligned}$$

Havaitsemme tutun tuloksen:

$$\det(\mathbf{X}'\mathbf{X}) = 0 \iff \mathbf{x} \in \mathscr{C}(\mathbf{1}) \iff s_x^2 = 0 \iff x_1 = \cdots = x_n\,. \tag{4.29}$$

Käänteismatriisin lauseke on täten

$$\begin{aligned}(\mathbf{X}'\mathbf{X})^{-1} &= \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} = \frac{1}{n\,\mathrm{SS}_x}\begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \\ &= \frac{1}{\mathrm{SS}_x}\begin{pmatrix} \sum x_i^2/n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} := \begin{pmatrix} t^{00} & t^{01} \\ t^{10} & t^{11} \end{pmatrix}. \end{aligned} \tag{4.30}$$

Koska

$$\mathrm{SS}_x = \mathbf{x}'(\mathbf{I} - \mathbf{J})\mathbf{x} = \mathbf{x}'\mathbf{x} - \mathbf{x}'\mathbf{J}\mathbf{x} = \mathbf{x}'\mathbf{x} - n\bar{x}^2, \tag{4.31}$$

on $\mathbf{x}'\mathbf{x} = \mathrm{SS}_x + n\bar{x}^2$, joten t^{00}:lle saadaan esitys

$$t^{00} = \frac{\mathbf{x}'\mathbf{x}}{n\mathbf{x}'(\mathbf{I} - \mathbf{J})\mathbf{x}} = \frac{\mathrm{SS}_x + n\bar{x}^2}{n\,\mathrm{SS}_x} = \frac{1}{n} + \frac{\bar{x}^2}{\mathrm{SS}_x}\,. \tag{4.32}$$

Mainittakoon ohimennen, että tietyin oletuksin regressiokertoimien $\left(\begin{smallmatrix}\hat{\beta}_0 \\ \hat{\beta}_1\end{smallmatrix}\right)$ kovarianssimatriisi on

$$\operatorname{cov}\begin{pmatrix}\hat{\beta}_0 \\ \hat{\beta}_1\end{pmatrix} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma^2}{\mathrm{SS}_x}\begin{pmatrix} \sum x_i^2/n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} = \sigma^2\begin{pmatrix} t^{00} & t^{01} \\ t^{10} & t^{11} \end{pmatrix}, \tag{4.33}$$

missä σ^2 on lineaarisen mallin virhetermin varianssi. Tällöin mm.

$$\operatorname{var}(\hat{\beta}_1) = \frac{\sigma^2}{\mathrm{SS}_x}\,, \qquad \operatorname{cor}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}}{\sqrt{\sum x_i^2/n}}\,. \tag{4.34}$$

□

4.1.2 Multinormaalijakauman tiheysfunktio

Tarkastellaan kaksiulotteista satunnaismuuttujaa $\mathbf{z} = \left(\begin{smallmatrix}x \\ y\end{smallmatrix}\right)$, jonka kovarianssi- ja korrelaatiomatriisi ovat

$$\operatorname{cov}(\mathbf{z}) = \operatorname{cov}\begin{pmatrix}x \\ y\end{pmatrix} = \mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \sigma_x\sigma_y\varrho \\ \sigma_x\sigma_y\varrho & \sigma_y^2 \end{pmatrix}, \tag{4.35a}$$

$$\operatorname{cor}(\mathbf{z}) = \boldsymbol{\rho} = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}, \qquad \varrho = \operatorname{cor}(x, y)\,. \tag{4.35b}$$

Käänteismatriisit ja determinantit ovat

$$\begin{aligned}
\boldsymbol{\Sigma}^{-1} &= \frac{1}{|\boldsymbol{\Sigma}|}\begin{pmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{21} & \sigma_x^2 \end{pmatrix} = \frac{1}{|\boldsymbol{\Sigma}|}\begin{pmatrix} \sigma_y^2 & -\sigma_x\sigma_y\varrho \\ -\sigma_x\sigma_y\varrho & \sigma_x^2 \end{pmatrix} \\
&= \frac{1}{\sigma_x^2\sigma_y^2(1-\varrho^2)}\begin{pmatrix} \sigma_y^2 & -\sigma_x\sigma_y\varrho \\ -\sigma_x\sigma_y\varrho & \sigma_x^2 \end{pmatrix} \\
&= \frac{1}{1-\varrho^2}\begin{pmatrix} \frac{1}{\sigma_x^2} & \frac{-\varrho}{\sigma_x\sigma_y} \\ \frac{-\varrho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{pmatrix} := \begin{pmatrix} \sigma^{xx} & \sigma^{xy} \\ \sigma^{yx} & \sigma^{yy} \end{pmatrix}, \qquad (4.36a)
\end{aligned}$$

$$|\boldsymbol{\Sigma}| = \sigma_x^2\sigma_y^2 - \sigma_{xy}^2 = \sigma_x^2\sigma_y^2 - (\sigma_x\sigma_y\varrho)^2 = \sigma_x^2\sigma_y^2(1-\varrho^2) \leq \sigma_x^2\sigma_y^2\,, \qquad (4.36b)$$

$$\boldsymbol{\rho}^{-1} = \frac{1}{1-\varrho^2}\begin{pmatrix} 1 & -\varrho \\ -\varrho & 1 \end{pmatrix}, \quad |\boldsymbol{\rho}| = 1-\varrho^2 \leq 1\,. \qquad (4.36c)$$

Kannattaa panna merkille, että kovarianssimatriisin determinantti on pienempi tai yhtäsuuri kuin tarkasteltavien muuttujien varianssien tulo ja että yhtäsuuruus pätee vain kun $\varrho = \operatorname{cor}(x, y) = 0$. Vastaava tulos osoittautuu pätevän myös yleisessä tapauksessa:

$$|\operatorname{cov}(\mathbf{z})| = |\boldsymbol{\Sigma}_{p\times p}| \leq \sigma_1^2\sigma_2^2\cdots\sigma_p^2\,, \qquad (4.37)$$

missä yhtäsuuruus pätee jos ja vain jos $\mathbf{z}$:n komponentit ovat korreloimattomia:

$$|\boldsymbol{\Sigma}_{p\times p}| = \sigma_1^2\sigma_2^2\cdots\sigma_p^2 \iff \boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)\,. \qquad (4.38)$$

Vastaavasti korrelaatiomatriisille on voimassa

$$|\boldsymbol{\rho}| \leq 1\,, \qquad |\boldsymbol{\rho}| = 1 \iff \boldsymbol{\rho} = \mathbf{I}_p\,. \qquad (4.39)$$

Olkoon $\mathbf{z}$ satunnaisvektori, joka noudattaa p-ulotteista normaalijakaumaa parametrein $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ eli $\mathbf{z} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ja olkoon $\boldsymbol{\Sigma}$:lla käänteismatriisi eli $\boldsymbol{\Sigma}$ on posiivisesti definiitti. Tällöin $\mathbf{z}$:n tiheysfunktio on

$$n(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{1}{2}p}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}. \qquad (4.40)$$

Kun $p = 2$ ja $\varrho = \operatorname{cor}(x, y)$, niin tiheysfunktion lauseke on

$$\begin{aligned}
n(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\varrho^2}}\cdot \\
&\qquad \exp\Big\{\tfrac{-1}{2}\big[\sigma^{xx}(x-\mu_x)^2 - 2\sigma^{xy}(x-\mu_x)(y-\mu_y) + \sigma^{yy}(y-\mu_y)^2\big]\Big\} \\
&= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\varrho^2}}\cdot \\
&\qquad \exp\Big\{\frac{-1}{2(1-\varrho^2)}\Big[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\varrho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\Big]\Big\}. \qquad (4.41)
\end{aligned}$$

Havaitsemme (4.40):n nojalla, että ne $\mathbf{z}$-vektorit, joilla tiheysfunktio saavuttaa saman arvon, muodostuvat niistä $\mathbf{z}$-arvoista, joilla

$$(\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu}) = c^2 \quad (= \text{jokin tietty vakio}). \tag{4.42}$$

Ehdon (4.42) mukaiset $\mathbf{z}$:n arvot muodostavat tietyn $\boldsymbol{\mu}$-keskisen ellipsoidin ja 2-ulotteisessa tapauksessa siten ellipsin.

4.1.3 Käänteismatriisin ominaisuuksia

Seuraavassa on lueteltu muutamia tärkeitä käänteismatriisin ja determinantin ominaisuuksia:

KM1. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \quad$ ($\mathbf{A}$ ja $\mathbf{B}$ neliömatriiseja)

KM2. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$

KM3. $\mathbf{A}$ symmetrinen $\implies$ $\mathbf{A}^{-1}$ symmetrinen

KM4. $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \mathbf{I}_n \implies \mathbf{A}^{-1} = \mathbf{A}' \quad$ ($\mathbf{A}_{n\times n}$ on ortogonaalinen)

KM5. $[\text{diag}(d_1, d_2, \ldots, d_n)]^{-1} = \text{diag}(1/d_1, 1/d_2, \ldots, 1/d_n) \quad$ (kun $d_i \neq 0$)

KM6. $\begin{pmatrix} \mathbf{I} & \mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I} & -\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$

KM7. $\begin{pmatrix} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{U}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^{-1} \end{pmatrix}$

KM8. $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{C}^{-1} \\ \mathbf{0} & \mathbf{C}^{-1} \end{pmatrix}$

D1. $\det(\mathbf{AB}) = \det(\mathbf{A}) \cdot \det(\mathbf{B}) \quad$ ($\mathbf{A}$ ja $\mathbf{B}$ neliömatriiseja)

D2. $\det(\mathbf{A}') = \det(\mathbf{A})$

D3. $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \mathbf{I}_n \implies \det(\mathbf{A}) = \pm 1$

D4. $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$

D5. $\det(\alpha\mathbf{A}_{n\times n}) = \alpha^n \det(\mathbf{A})$

D6. $\det(\mathbf{D}) = d_{11}d_{22}\cdots d_{nn}$, jos $\mathbf{D}$ on lävistäjä- tai ylä(-ala)kolmiomatriisi

D7. $\det(\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_n) = \det(\mathbf{a}_1 : \mathbf{a}_2 + \alpha\mathbf{a}_1 : \ldots : \mathbf{a}_n) \quad$ (sarakkeen lisääminen (vakiolla kerrottuna) toiseen sarakkeeseen ei muuta determinanttia – vastaava ominaisuus pätee myös vaakariveille)

D8. $\det(\mathbf{a}_1 : \mathbf{a}_2 : \mathbf{a}_3 : \ldots : \mathbf{a}_n) = -\det(\mathbf{a}_2 : \mathbf{a}_1 : \mathbf{a}_3 : \ldots : \mathbf{a}_n)$

D9. $\det\begin{pmatrix} \mathbf{U} & \mathbf{F} \\ \mathbf{0} & \mathbf{V} \end{pmatrix} = \det(\mathbf{U}) \cdot \det(\mathbf{V}) \quad$ ($\mathbf{U}$ ja $\mathbf{V}$ neliömatriiseja)

On vielä erityisesti syytä korostaa (ja toistaa) matriisin $\mathbf{A}_{n\times m}$ ja matriisin $\mathbf{A}'\mathbf{A}$ sarakkeiden vapauden välistä yhteyttä. Nimittäin (2.61):n (s. 89) mukaan on voimassa

$$\mathbf{A}\text{:n sarakkeet vapaat} \iff \mathbf{A}'\mathbf{A}\text{:n sarakkeet vapaat} \iff \det(\mathbf{A}'\mathbf{A}) \neq 0. \tag{4.43}$$

Itse asiassa, ks. (2.62) (s. 90), matriisin $\mathbf{A}$ aste on sama kuin $\mathbf{A}'\mathbf{A}$:n aste. Täten jos $\mathbf{A}$:n sarakkeet ovat vapaat, niin $m \times m$-neliömatriisin $\mathbf{A}'\mathbf{A}$ aste on m, ja käänteismatriisi $(\mathbf{A}'\mathbf{A})^{-1}$ on olemassa. Vastaavasti, jos $\mathbf{A}$:n vaakarivit ovat vapaat, on $(\mathbf{A}\mathbf{A}')^{-1}$ olemassa.

4.2 Esimerkkejä

4.2.1 Tasakorrelaatiomatriisi

Esimerkki 4.2. Esimerkissä 1.5 (s. 41) tarkasteltiin ns. *tasakorrelaatiomatriisia*

$$\boldsymbol{\rho}_{n\times n} = (1-\varrho)\mathbf{I}_n + \varrho\mathbf{1}_n\mathbf{1}_n' = \begin{pmatrix} 1 & \varrho & \ldots & \varrho \\ \varrho & 1 & \ldots & \varrho \\ \vdots & \vdots & \ddots & \vdots \\ \varrho & \varrho & \ldots & 1 \end{pmatrix}. \tag{4.44}$$

Osoita seuraavat tulokset (ominaisarvoihin palataan myöhemmin):

(a) $\det(\boldsymbol{\rho}) = (1-\varrho)^{n-1}[1+(n-1)\varrho], \quad 0 \leq \det(\boldsymbol{\rho}) \leq 1,$

(b) $\boldsymbol{\rho}^{-1} = \dfrac{1}{1-\varrho}\Big(\mathbf{I}_n - \dfrac{\varrho}{1+(n-1)\varrho}\mathbf{1}_n\mathbf{1}_n'\Big),$

(c) $\boldsymbol{\rho}$:n ominaisarvot $\mathrm{ch}_1(\boldsymbol{\rho}) \geq \cdots \geq \mathrm{ch}_n(\boldsymbol{\rho})$ ovat

$$\begin{cases} 1+(n-1)\varrho \quad \text{kertalukuna } 1, \\ 1-\varrho \quad \text{kertalukuna } n-1, \end{cases}$$

(d) $\boldsymbol{\rho}$ on ei-negatiivisesti definiitti [eli $\mathrm{ch}_{\min}(\boldsymbol{\rho}) \geq 0$] jos ja vain jos

$$\frac{-1}{n-1} \leq \varrho \leq 1\,. \tag{4.45}$$

Aina kun tarkastelemme tasakorrelaatiomatriisia – korrelaatiomatriisi on nimittäin aina ei-negatiivisesti definiitti – oletamme, että (4.45) on voimassa.

Yleisemmässä tilanteessa saadaan *intraclass-kovarianssimatriisi*:

$$\mathbf{\Sigma}_{n\times n} = (a-b)\mathbf{I}_n + b\mathbf{1}_n\mathbf{1}_n' = \begin{pmatrix} a & b & \dots & b \\ b & a & \dots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \dots & a \end{pmatrix}. \tag{4.46}$$

(a*) $\det(\mathbf{\Sigma}) = (a-b)^{n-1}[a+(n-1)b], \quad 0 \leq \det(\mathbf{\Sigma}) \leq a^n,$

(b*) $\mathbf{\Sigma}^{-1} = \dfrac{1}{a-b}\Big(\mathbf{I}_n - \dfrac{b}{a+(n-1)b}\mathbf{1}_n\mathbf{1}_n'\Big),$

(c*) $\mathbf{\Sigma}$:n ominaisarvot ovat

$$\begin{cases} a+(n-1)b \quad \text{kertalukuna } 1, \\ a-b \quad \text{kertalukuna } n-1, \end{cases}$$

(d*) $\mathbf{\Sigma}$ on ei-negatiivisesti definiitti [eli $\mathrm{ch}_{\min}(\mathbf{\Sigma}) \geq 0$] jos ja vain jos

$$\frac{-a}{n-1} \leq b \leq a\,. \tag{4.47}$$

□

Esimerkki 4.3. Lasketaan determinantti 3×3-tasakorrelaatiomatriisista

$$\mathbf{R} = \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix}. \tag{4.48}$$

Lopputulos on siis $\det(\mathbf{R}) = (1-r)^2(1+2r)$. Determinantin laskemiseksi tehdään $\mathbf{A}$:lle seuraavat operaatiot:

(1) Lisätään $\mathbf{R}$:n 1. sarakkeeseen 3. sarake -1:llä kerrottuna; matriisikertolaskuna tämä tarkoittaa seuraavaa:

$$\mathbf{R}_1 = \mathbf{R}\mathbf{E}_1 = \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1-r & r & r \\ 0 & 1 & r \\ r-1 & r & 1 \end{pmatrix}. \tag{4.49}$$

(2) Lisätään $\mathbf{R}_1$:n 2. sarakkeeseen 3. sarake -1:llä kerrottuna:

$$\mathbf{R}_2 = \mathbf{R}\mathbf{E}_1\mathbf{E}_2 = \begin{pmatrix} 1-r & r & r \\ 0 & 1 & r \\ r-1 & r & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1-r & 0 & r \\ 0 & 1-r & r \\ r-1 & r-1 & 1 \end{pmatrix} \tag{4.50}$$

Itse asiassa nämä molemmat operaatiot voidaan tehdä "yhdellä" kertolaskulla:

$$\mathbf{R}_2 = \mathbf{RF} = \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1-r & 0 & r \\ 0 & 1-r & r \\ r-1 & r-1 & 1 \end{pmatrix}, \tag{4.51}$$

missä

$$\mathbf{F} = \mathbf{E}_1\mathbf{E}_2 = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0}_2 \\ -\mathbf{1}_2' & 1 \end{pmatrix}. \tag{4.52}$$

(3) Lisätään $\mathbf{R}_2$:n 1. vaakarivi ja 2. vaakarivi viimeiseen vaakariviin. Matriisikertolaskuna tämä tarkoittaa $\mathbf{R}_2$:n kertomista vasemmalta $\mathbf{G}$:llä:

$$\mathbf{G} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0}_2 \\ \mathbf{1}_2' & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \tag{4.53}$$

jolloin

$$\begin{aligned} \mathbf{GR}_2 = \mathbf{GRF} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1-r & 0 & r \\ 0 & 1-r & r \\ r-1 & r-1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1-r & 0 & r \\ 0 & 1-r & r \\ 0 & 0 & 1+2r \end{pmatrix}, \end{aligned} \tag{4.54}$$

eli siis

$$\begin{aligned} \mathbf{GRF} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1-r & 0 & r \\ 0 & 1-r & r \\ 0 & 0 & 1+2r \end{pmatrix}. \end{aligned} \tag{4.55}$$

Matriisi $\mathbf{GRF}$ on siis yläkolmiomatriisi ja sen determinantti on lävistäjäelementtien tulo. Sivumennen todettakoon, että $\mathbf{G} = \mathbf{F}^{-1}$, sillä yleisesti on voimassa

$$\begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{K}_{p\times n} & \mathbf{I}_p \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{K}_{p\times n} & \mathbf{I}_p \end{pmatrix}. \tag{4.56}$$

Lukija voi vielä vahvistaa, että 4×4 -tapauksessa

$$\begin{pmatrix} \mathbf{I}_3 & \mathbf{0}_3 \\ -\mathbf{1}_3' & 1 \end{pmatrix} \mathbf{R}_{4\times 4} \begin{pmatrix} \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{1}_3' & 1 \end{pmatrix} = \begin{pmatrix} 1-r & 0 & 0 & r \\ 0 & 1-r & 0 & r \\ 0 & 0 & 1-r & r \\ 0 & 0 & 0 & 1+3r \end{pmatrix}. \tag{4.57}$$

□

4.2.2 Autoregressiivinen prosessi

Esimerkki 4.4. Esimerkissä 1.8 (s. 45) tarkasteltiin ns. autoregressiivistä prosessia,

$$y_i = \varrho y_{i-1} + u_i\,, \quad i = 1, \dots, n, \quad |\varrho| < 1\,, \tag{4.58}$$

missä u_i:t ($i = \dots, -2, -1, 0, 1, 2, \dots$) ovat riippumattomia satunnaismuuttujia, joilla on kaikilla sama odotusarvo 0 ja varianssi σ_u^2. Merkitsemme prosessia lyhyesti symbolilla AR(1). Tällöin voidaan näyttää, että

$$\begin{aligned}
\operatorname{cov}(\mathbf{y}) = \sigma^2\mathbf{V} &= \frac{\sigma_u^2}{1-\varrho^2}\begin{pmatrix} 1 & \varrho & \varrho^2 & \dots & \varrho^{n-1} \\ \varrho & 1 & \varrho & \dots & \varrho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \varrho^{n-1} & \varrho^{n-2} & \varrho^{n-3} & \dots & 1 \end{pmatrix} \\
&= \frac{\sigma_u^2}{1-\varrho^2}\{\varrho^{|i-j|}\} = \frac{\sigma_u^2}{1-\varrho^2}\mathbf{V}, \qquad (4.59a)
\end{aligned}$$

$$\operatorname{cor}(\mathbf{y}) = \begin{pmatrix} 1 & \varrho & \varrho^2 & \dots & \varrho^{n-1} \\ \varrho & 1 & \varrho & \dots & \varrho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \varrho^{n-1} & \varrho^{n-2} & \varrho^{n-3} & \dots & 1 \end{pmatrix} = \mathbf{V} = \{\varrho^{|i-j|}\}\,. \tag{4.59b}$$

Tällöin siis peräkkäisten y_i-arvojen korrelaatio on

$$\operatorname{cor}(y_i, y_{i+1}) = \varrho\,, \ \operatorname{cor}(y_i, y_{i+2}) = \varrho^2, \text{ jne.} \tag{4.60}$$

Osoitamme [ks. Puntanen, Styan & Isotalo (2011, §9.6)] että

$$\mathbf{V}^{-1} = \frac{1}{1-\varrho^2}\begin{pmatrix} 1 & -\varrho & 0 & \dots & 0 & 0 & 0 \\ -\varrho & 1+\varrho^2 & -\varrho & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\varrho & 1+\varrho^2 & -\varrho \\ 0 & 0 & 0 & \dots & 0 & -\varrho & 1 \end{pmatrix}. \tag{4.61}$$

Tätä varten merkitään

$$\mathbf{e} = \begin{pmatrix} y_1 \\ y_2 - \varrho y_1 \\ y_3 - \varrho y_2 \\ \vdots \\ y_n - \varrho y_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -\varrho & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\varrho & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -\varrho & 1 \end{pmatrix}\mathbf{y} = \mathbf{L}\mathbf{y}\,, \tag{4.62}$$

jolloin on helppo havaita, että

$$\begin{aligned}
\operatorname{cov}(\mathbf{e}) = \sigma^2\begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & (1-\varrho^2)\mathbf{I}_{n-1} \end{pmatrix} &:= \sigma^2\mathbf{D} \\
= \sigma^2(1-\varrho^2)\begin{pmatrix} \frac{1}{1-\varrho^2} & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_{n-1} \end{pmatrix}&. \qquad (4.63)
\end{aligned}$$

Toisaalta satunnaisvektorin $\mathbf{Ly}$ kovarianssimatriisi on [ks. (7.7), s. 214]

$$\operatorname{cov}(\mathbf{e}) = \operatorname{cov}(\mathbf{Ly}) = \sigma^2 \mathbf{LVL}', \tag{4.64}$$

joten

$$\mathbf{LVL}' = \mathbf{D}\,. \tag{4.65}$$

Täten $\mathbf{V} = \mathbf{L}^{-1}\mathbf{D}(\mathbf{L}')^{-1}$, mistä $\mathbf{V}$:n käänteismatriisiksi saadaan

$$\mathbf{V}^{-1} = \mathbf{L}'\mathbf{D}^{-1}\mathbf{L} = (\mathbf{D}^{-1/2}\mathbf{L})'\mathbf{D}^{-1/2}\mathbf{L}\,, \tag{4.66}$$

missä

$$\begin{aligned}
\mathbf{D}^{-1/2}\mathbf{L} &= \frac{1}{\sqrt{1-\varrho^2}} \begin{pmatrix} \sqrt{1-\varrho^2} & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_{n-1} \end{pmatrix} \mathbf{L} \\
&= \frac{1}{\sqrt{1-\varrho^2}} \begin{pmatrix} \sqrt{1-\varrho^2} & 0 & 0 & \dots & 0 & 0 & 0 \\ -\varrho & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\varrho & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -\varrho & 1 \end{pmatrix} \\
&:= \frac{1}{\sqrt{1-\varrho^2}} \mathbf{K}\,.
\end{aligned} \tag{4.67}$$

On helppo päätellä että

$$\begin{aligned}
\mathbf{V}^{-1} &= \frac{1}{1-\varrho^2}\mathbf{K}'\mathbf{K} \\
&= \frac{1}{1-\varrho^2} \begin{pmatrix} 1 & -\varrho & 0 & \dots & 0 & 0 & 0 \\ -\varrho & 1+\varrho^2 & -\varrho & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\varrho & 1+\varrho^2 & -\varrho \\ 0 & 0 & 0 & \dots & 0 & -\varrho & 1 \end{pmatrix}.
\end{aligned} \tag{4.68}$$

Vakuuttaudu siitä, että $\det(\mathbf{V}) = (1-\varrho^2)^{n-1}$. □

Esimerkki 4.5. Muodostettava sellainen $(n-1)\times n$-matriisi $\mathbf{L}$, jolla on ominaisuus

$$\mathbf{Ly} = \mathbf{L}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} y_2 - y_1 \\ y_3 - y_2 \\ y_4 - y_3 \\ \vdots \\ y_n - y_{n-1} \end{pmatrix}. \tag{4.69}$$

Näemme välittömästi, että haettu $\mathbf{L}$ on

$$\mathbf{L}_{(n-1)\times n} = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}. \tag{4.70}$$

Määritetään sitten sellainen symmetrinen $n\times n$-matriisi $\mathbf{A}$, jolla on ominaisuus

$$\mathbf{y}'\mathbf{A}\mathbf{y} = (y_2 - y_1)^2 + (y_3 - y_2)^2 + (y_4 - y_3)^2 + \cdots + (y_n - y_{n-1})^2. \tag{4.71}$$

Koska $\mathbf{y}'\mathbf{A}\mathbf{y} = \|\mathbf{L}\mathbf{y}\|^2 = \mathbf{y}'\mathbf{L}'\mathbf{L}\mathbf{y}$, vastaus on tietenkin

$$\mathbf{A} = \mathbf{L}'\mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}, \tag{4.72}$$

ja siis jos $n = 5$, niin

$$\mathbf{A} = \mathbf{L}'\mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}. \tag{4.73}$$

Kääntyykö $\mathbf{A}$? Mikä on $\mathbf{A}$:n aste eli sen vapaiden sarakkeiden lukumäärä? Koska

$$\mathrm{r}(\mathbf{A}_{n\times n}) = \mathrm{r}(\mathbf{L}'\mathbf{L}) = \mathrm{r}(\mathbf{L}_{(n-1)\times n}) \leq n - 1\,, \tag{4.74}$$

ei $\mathbf{A}$:lla ole käänteismatriisia. Matriisin $\mathbf{L}'$ sarakkeet ovat vapaat sillä

$$\begin{aligned} \mathbf{L}' &= \begin{pmatrix} -1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_{n-1} \end{pmatrix} \\ &= \begin{pmatrix} -\lambda_1 \\ \lambda_1 - \lambda_2 \\ \lambda_2 - \lambda_3 \\ \vdots \\ \lambda_{n-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \iff \lambda_1 = \lambda_2 = \cdots = \lambda_{n-1} = 0\,. \end{aligned} \tag{4.75}$$

Täten $\mathbf{A}$:n aste on $\mathrm{r}(\mathbf{A}) = \mathrm{r}\big(\mathbf{L}_{(n-1)\times n}\big) = n - 1$. $\square$

```
 -  - SURVO MM  Thu Sep 23 15:41:57 2010           C:\SJP\D\   2000  100 0
*  KOKEILUJA INTRACLASS-KORRELAATIOMATRIISILLA
*      - kääntäminen ja ominaisarvot
*
*MAT ONE=CON(n,1)    /  ONE = tolppa-1     n=4  r=0.7
*MAT R!=(1-r)*IDN(n,n)+r*ONE*ONE'
*MAT INVR=INV(R,det)          / *INVR~INV(R) det=0.0837 4*4
*MAT LOAD INVR,##.###,CUR+1    /
*MATRIX INVR
*INV(R)
*///           1      2      3      4
*  1       2.581 -0.753 -0.753 -0.753
*  2      -0.753  2.581 -0.753 -0.753
*  3      -0.753 -0.753  2.581 -0.753
*  4      -0.753 -0.753 -0.753  2.581
*
*Determinantin lauseke:
*  detR=(1+(n-1)*r)*((1-r)^(n-1))
*  detR=0.0837             kun  n=4   r=0.7
*  detR=0.9477                  n=4   r=0.1
*  detR=0                       n=4   r=-1/3
*  detR=-0.04812208             n=4   r=-0.34
*
*MAT SPECTRAL DECOMPOSITION OF R TO T,L
*MAT NAME T AS Ominaisvektorit
*MAT LOAD T,##.###,CUR+1
*MATRIX T
*Ominaisvektorit
*///         ev1    ev2    ev3    ev4
*  1      -0.500  0.087  0.707 -0.492
*  2      -0.500  0.087 -0.707 -0.492
*  3      -0.500  0.609  0.000  0.615
*  4      -0.500 -0.783  0.000  0.369
*
*MAT NAME L AS Ominaisarvot
*MAT LOAD L,CUR+1
*MATRIX L
*Ominaisarvot
*///      eigenval
*ev1      3.100000
*ev2      0.300000
*ev3      0.300000
*ev4      0.300000
*....................
*  det=((1-r)^(n-1))*(1+(n-1)*r)    n=4  r=0.7    det=0.0837
*  ch1=1+(n-1)*r   ch1=3.1     oa1=3.1     oa1=MAT_L(1,1)
*  ch2=1-r         ch2=0.3     oa2=0.3     oa2=MAT_L(2,1)
*  DET=ch1*ch2^3   DET=0.0837    determinantti on ominaisarvojen tulo
*  HUOM. On oltava voimassa: chi >= 0 joten    - 1/(n-1) <= r <= 1
*
*MAT D!=DV(L)   /  DV(L) tekee om.arvoista (vektori L) diagmatriisin
*MAT NJR!=T*D^(0.5)*T'      / NJR = R:n neliöjuuri,
*MAT LOAD NJR,##.###,CUR+1  /   R = T*D*T' = R:n ominaisarvohajotelma
*MATRIX NJR
*///           1      2      3      4
*  1       0.851  0.303  0.303  0.303
*  2       0.303  0.851  0.303  0.303
*  3       0.303  0.303  0.851  0.303
*  4       0.303  0.303  0.303  0.851
*
*MAT KK!=TRACE((R-NJR*NJR)'*(R-NJR*NJR))
*MAT LOAD KK,##.##########,CUR+1   /  RNJ todellakin on R*R
*MATRIX KK
*///              trace
*trace     0.0000000000
*                                              (MK10-008, MKT-04-tas1)
```

```
 -  - SURVO MM  Thu Sep 23 15:54:27 2010          C:\SJP\D\   2000  100 0
*
*       KORRELAATIOMATRIISIN ELEMENTIT r:n potensseja
*         - käänteismatriisi ja ominaisarvot
*
*MAT R=CON(n,n)         /  n=5 r=0.9  CON(a,b) = axb-mtx täynnä ykkösiä
*MAT TRANSFORM R BY r^abs(I#-J#)  /   (1,2)-elementti = r
*MAT NAME R AS R                  /   (1,3)-elementti = r^2  jne
*MAT INVR=INV(R,det)        / *INVR~INV(R) det=0.00130321 5*5
*MAT LOAD INVR,##.###,CUR+1
*MATRIX INVR
*INV(R)
*///          1      2      3      4      5
*  1       5.263 -4.737  0.000 -0.000 -0.000
*  2      -4.737  9.526 -4.737  0.000  0.000
*  3       0.000 -4.737  9.526 -4.737 -0.000
*  4      -0.000  0.000 -4.737  9.526 -4.737
*  5      -0.000 -0.000 -0.000 -4.737  5.263
*
*MAT W!=(1-r^2)*INVR      /     SIIS: inv(R) = W/(1-r^2)
*MAT LOAD W,##.###,CUR+1
*MATRIX W
*///          1      2      3      4      5
*  1       1.000 -0.900  0.000 -0.000 -0.000
*  2      -0.900  1.810 -0.900  0.000  0.000
*  3       0.000 -0.900  1.810 -0.900 -0.000
*  4      -0.000  0.000 -0.900  1.810 -0.900
*  5      -0.000 -0.000 -0.000 -0.900  1.000
*
*MATRIX W1 ///   HUOM: W1 = W
*          1      -r       0       0       0
*         -r       1+r^2  -r       0       0
*          0      -r       1+r^2  -r       0
*          0       0      -r       1+r^2  -r
*          0       0       0      -r       1
*
*Matriisi W voidaan esittää muodossa W = Q’Q, missä
*MATRIX  Q  ///
*  sqrt(1-r^2)  0   0   0   0
*  -r           1   0   0   0
*   0          -r   1   0   0
*   0           0  -r   1   0
*   0           0   0  -r   1
*
*MAT SAVE Q          /      Q’Q = W = (1-r^2)*inv(R)
*MAT A1!=Q’*Q        /      inv(R) = W/(1-r^2) = Q’Q/(1-r^2)
*MAT LOAD A1,##.###,CUR+1
*MATRIX A1
*///          1      2      3      4      5
*  1       1.000 -0.900  0.000  0.000  0.000
*  2      -0.900  1.810 -0.900  0.000  0.000
*  3       0.000 -0.900  1.810 -0.900  0.000
*  4       0.000  0.000 -0.900  1.810 -0.900
*  5       0.000  0.000  0.000 -0.900  1.000
*                                            (MK10-008, MKT-04-ar)
```

4.3 Käänteismatriisi ositetussa muodossa

Monissa tilastollisissa menetelmissä tarvitaan tavan takaa tuloksia, jotka liittyvät matriisin kääntämiseen ositetussa muodossa. Käymme tässä lyhyesti läpi joitakin näistä tuloksista.

Olkoon symmetrinen kääntyvä $\mathbf{A}_{n\times n}$ ositettu siten että

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \tag{4.76}$$

missä $\mathbf{A}_{11}$ on neliömatriisi. Symmetrisyys ei ole oleellista seuraavissa tarkasteluissa, mutta se tietysti yksinkertaistaa joitakin lausekkeita. Hyvin usein sovellamme osittaiskääntötuloksia nimenomaan symmetriseen matriisiin. Merkitään $\mathbf{A}$:n vastaavasti ositettua käänteismatriisia

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{pmatrix}. \tag{4.77}$$

Tällöin $\mathbf{A}$ voidaan esittää kolmen ositetun matriisin tulona

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{4.78a}$$

eli

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22\cdot 1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} := \mathbf{F}'\mathbf{G}\mathbf{F}, \tag{4.78b}$$

missä

$$\mathbf{A}_{22\cdot 1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{A}_{11}\text{:n Schurin komplementti } \mathbf{A}\text{:ssa.} \tag{4.79}$$

Myös seuraava merkintä on kirjallisuudessa yleinen:

$$\mathbf{A}_{22\cdot 1} = \mathbf{A}/\mathbf{A}_{11} = \mathbf{A}_{11}\text{:n Schurin komplementti } \mathbf{A}\text{:ssa.} \tag{4.80}$$

Kommentti 4.1. Schurin komplementilla on tilastotieteessä useita sovelluksia, ks. esim. Ouellette (1981), and Styan (1985). Muista lähteistä mainittakoon Carlson (1986), Cottle (1974), Henderson & Searle (1981a), Zhang (2005), Puntanen & Styan (2005a,b). Termin ”Schurin komplementti” otti käyttöön Haynsworth (1968a,b). Schur (1917) osoitti mm. että jos $\mathbf{A}_{11}$ on kääntyvä neliömatriisi, niin

$$|\mathbf{A}| = |\mathbf{A}_{11}||\mathbf{A}/\mathbf{A}_{11}| = |\mathbf{A}_{11}||\mathbf{A}_{22\cdot 1}| = |\mathbf{A}_{11}||\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}|, \tag{4.81}$$

ja täten determinantti on multiplikatiivinen Schurin komplementin suhteen. Tuloksen (4.81) nojalla Haynsworth (1968a,b) otti käyttöön merkinnän (4.80). Koska lohkolävistämatriisille pätee

$$\begin{pmatrix} \mathbf{F}_{f\times f} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{g\times g} \end{pmatrix} = \begin{pmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_g \end{pmatrix} \begin{pmatrix} \mathbf{I}_f & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{pmatrix} \implies \begin{vmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{vmatrix} = |\mathbf{F}||\mathbf{G}|\,, \tag{4.82}$$

näemme (4.78b):sta että

$$|\mathbf{A}| = |\mathbf{A}_{11}||\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}|\,. \tag{4.83}$$

on voimassa. □

Kaavasta (4.78) voi vakuuttautua suoraviivaisella kertolaskulla. Koska $\mathbf{G}$ on lohkolävistäjämatriisi ja

$$\mathbf{F}^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{4.84}$$

saamme $\mathbf{A}$:n käänteismatriisille esityksen

$$\begin{aligned} \mathbf{A}^{-1} &= \mathbf{F}^{-1}\mathbf{G}^{-1}(\mathbf{F}')^{-1} = \mathbf{F}^{-1}\mathbf{G}^{-1}(\mathbf{F}^{-1})' \\ &= \begin{pmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22\cdot 1}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{pmatrix}. \end{aligned} \tag{4.85}$$

Kaavasta (4.85) seuraa välittömästi että

$$\boxed{\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22\cdot 1}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22\cdot 1}^{-1} \\ -\mathbf{A}_{22\cdot 1}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{A}_{22\cdot 1}^{-1} \end{pmatrix}.} \tag{4.86}$$

Lähtemällä (4.78):n sijasta liikkeelle hajotelmasta

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{pmatrix}, \tag{4.87}$$

saamme tuloksen

$$\boxed{\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11\cdot 2}^{-1} & -\mathbf{A}_{11\cdot 2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11\cdot 2}^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11\cdot 2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{pmatrix},} \tag{4.88}$$

missä

$$\mathbf{A}_{11\cdot 2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} = \mathbf{A}_{22}\text{:n Schurin komplementti } \mathbf{A}\text{:ssa.} \tag{4.89}$$

Käänteismatriisi $\mathbf{A}^{-1}$ on tietenkin yksikäsitteinen, joten (4.86) ja (4.88) antavat saman tuloksen ja täten esimerkiksi

$$\mathbf{A}^{22} = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} = \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11\cdot 2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\,. \tag{4.90}$$

Yhtälön (4.90) perusteella voimme päätellä, että

$$\mathbf{A}_{22} = \mathbf{A}_{22}^{-1} \iff \mathbf{A}_{12} = \mathbf{A}_{21}' = \mathbf{0}\,. \tag{4.91}$$

Kommentti 4.2. Korostettakoon vielä kerran, että matriisin $\mathbf{A}$ symmetrisyys ei ole edellisissä tarkasteluissa mitenkään oleellista. Lähtökohtana oleva hajotelma (4.78a) pätee myös epäsymmetriselle $\mathbf{A}$:lle, mutta (4.78b):n mukainen merkintä $\mathbf{A} = \mathbf{F}'\mathbf{G}\mathbf{F}$ on korvattava esityksellä $\mathbf{A} = \mathbf{E}'\mathbf{G}\mathbf{F}$. □

4.3.1 Summan käänteismatriisi

Matriisien summan kääntämistä tarvitaan aina silloin tällöin. Hyvä yhteenveto aiheesta on artikkelissa Henderson & Searle (1981a). Seuraavassa on kolme erikoistapausta, joissa siis oletetaan, että tietyt käänteismatriisit ovat olemassa.

(a) $(\mathbf{B} - \mathbf{U}\mathbf{D}^{-1}\mathbf{V}')^{-1} = \mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{U}\mathbf{S}^{-1}\mathbf{V}'\mathbf{B}^{-1}$, missä $\mathbf{S} = \mathbf{D} - \mathbf{V}'\mathbf{B}^{-1}\mathbf{U}$.

(b) $(\mathbf{B} + \alpha\mathbf{u}\mathbf{v}')^{-1} = \mathbf{B}^{-1} - \dfrac{\alpha}{1 + \alpha\mathbf{v}'\mathbf{B}^{-1}\mathbf{u}}\mathbf{B}^{-1}\mathbf{u}\mathbf{v}'\mathbf{B}^{-1}$, $\alpha \in \mathbb{R}$.

(c) $(\mathbf{A} + \alpha\mathbf{I})^{-1} = \mathbf{A}^{-1} - \alpha\mathbf{A}^{-1}(\mathbf{I} + \alpha\mathbf{A}^{-1})^{-1}\mathbf{A}^{-1}$, $\alpha \in \mathbb{R}$.

Kaavan (b) ovat ilmeisesti ensimmäisenä esittäneet Sherman & Morrison (1949, 1950) ja kaavan (a) Woodbury (1950).

Esimerkki 4.6. Tarkastellaan esimerkkinä havaintomatriisia $\mathbf{U}_{n\times p}$, joka on ositettu seuraavasti:

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{u}_{(n)}' \end{pmatrix}. \tag{4.92}$$

Tällöin tietenkin

$$\mathbf{U}'\mathbf{U} = (\mathbf{U}_1' : \mathbf{u}_{(n)}) \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{u}_{(n)}' \end{pmatrix} = \mathbf{U}_1'\mathbf{U}_1 + \mathbf{u}_{(n)}\mathbf{u}_{(n)}'\,, \tag{4.93}$$

$$\mathbf{U}_1'\mathbf{U}_1 = \mathbf{U}'\mathbf{U} - \mathbf{u}_{(n)}\mathbf{u}_{(n)}' = \mathbf{u}_{(1)}\mathbf{u}_{(1)}' + \cdots + \mathbf{u}_{(n-1)}\mathbf{u}_{(n-1)}'\,, \tag{4.94}$$

joten kaavan (b) perusteella

$$\begin{aligned}(\mathbf{U}_1'\mathbf{U}_1)^{-1} &= (\mathbf{U}'\mathbf{U} - \mathbf{u}_{(n)}\mathbf{u}_{(n)}')^{-1} \\ &= (\mathbf{U}'\mathbf{U})^{-1} + \frac{(\mathbf{U}'\mathbf{U})^{-1}\mathbf{u}_{(n)}\mathbf{u}_{(n)}'(\mathbf{U}'\mathbf{U})^{-1}}{1 - \mathbf{u}_{(n)}'(\mathbf{U}'\mathbf{U})^{-1}\mathbf{u}_{(n)}}\,. \end{aligned} \tag{4.95}$$

Kaavasta (4.95) ”näkyy” miten viimeisen vaakarivin poisjättö $\mathbf{U}$:sta vaikuttaa $\mathbf{U}'\mathbf{U}$:n käänteismatriisiin. Koska $\mathbf{U}'\mathbf{i}_n = \mathbf{u}_{(n)}$, voimme kirjoittaa (4.95):n myös muodossa

$$\begin{aligned}(\mathbf{U}_1'\mathbf{U}_1)^{-1} &= (\mathbf{U}'\mathbf{U})^{-1} + \frac{(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{i}_n\mathbf{i}_n'\mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}}{1-\mathbf{i}_n'\mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{i}_n} \\ &= (\mathbf{U}'\mathbf{U})^{-1} + \frac{(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{i}_n\mathbf{i}_n'\mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}}{1-p_{nn}}, \qquad (4.96)\end{aligned}$$

missä p_{nn} on ortogonaaliprojektorin $\mathbf{P_U} = \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'$ viimeinen lävistäjäelementti. Mainittakoon vielä, että kertomalla (4.96) vasemmalta $\mathbf{u}_{(n)}'$:lla ja oikealta $\mathbf{u}_{(n)}$:llä saadaan

$$\mathbf{u}_{(n)}'(\mathbf{U}_1'\mathbf{U}_1)^{-1}\mathbf{u}_{(n)} = p_{nn} + \frac{p_{nn}^2}{1-p_{nn}} = \frac{p_{nn}}{1-p_{nn}}. \qquad (4.97)$$

□

4.4 Muuntaminen lohkolävistäjämuotoon

Kertomalla hajotelma

$$\mathbf{A} = \begin{pmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22\cdot 1}\end{pmatrix}\begin{pmatrix}\mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I}\end{pmatrix} = \mathbf{F}'\mathbf{G}\mathbf{F} \qquad (4.98)$$

vasemmalta $(\mathbf{F}^{-1})'$:lla ja oikealta $\mathbf{F}^{-1}$:llä saadaan yhtälö

$$(\mathbf{F}^{-1})'\mathbf{A}\mathbf{F}^{-1} = \mathbf{G} \qquad (4.99)$$

eli

$$\begin{pmatrix}\mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I}\end{pmatrix}\mathbf{A}\begin{pmatrix}\mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I}\end{pmatrix} = \begin{pmatrix}\mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22\cdot 1}\end{pmatrix}. \qquad (4.100)$$

Tämä tarkoittaa, että $\mathbf{A}$ on muunnettu lohkolävistäjämuotoon. Lohkodiagonalisoinnista on erityistä hyötyä mm. silloin, kun haluamme tehdä satunnaisvektorille $\mathbf{z}$ sellaisen lineaarisen muunnoksen $\mathbf{Bz}$, että $\mathbf{Bz}$:n kovarianssimatriisi on lohkolävistäjämatriisi. Jos $\operatorname{cov}(\mathbf{z}) = \boldsymbol{\Sigma}$, niin $\operatorname{cov}(\mathbf{Bz}) = \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'$, ja täten lohkodiagonalisoinnin tekevä $\mathbf{B}$ on

$$\mathbf{B} = \begin{pmatrix}\mathbf{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I}\end{pmatrix}. \qquad (4.101)$$

Jos esimerkiksi $\mathbf{z} = \binom{\mathbf{x}}{y}$ on $(p+1)$:n elementin satunnaisvektori, jonka kovarianssimatriisi on

$$\operatorname{cov}(\mathbf{u}) = \operatorname{cov}\begin{pmatrix}\mathbf{x} \\ y\end{pmatrix} = \begin{pmatrix}\operatorname{cov}(\mathbf{x}) & \operatorname{cov}(\mathbf{x}, y) \\ \operatorname{cov}(y, \mathbf{x}) & \operatorname{var}(y)\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}_{\mathbf{x}y}' & \sigma_y^2\end{pmatrix} = \boldsymbol{\Sigma}, \qquad (4.102)$$

niin

$$\mathbf{B} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0}_p \\ -\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} & 1 \end{pmatrix}, \tag{4.103}$$

ja

$$\operatorname{cov}(\mathbf{Bz}) = \mathbf{B\Sigma B}' = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \mathbf{0}_p \\ \mathbf{0}'_p & \sigma_y^2 - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y} \end{pmatrix}, \tag{4.104}$$

ts.

$$\boxed{\begin{aligned} \operatorname{cov}\left[\begin{pmatrix} \mathbf{I}_p & \mathbf{0}_p \\ -\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} & 1 \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ y \end{pmatrix}\right] &= \operatorname{cov}\begin{pmatrix} \mathbf{x} \\ y - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \mathbf{0}_p \\ \mathbf{0}'_p & \sigma_y^2 - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y} \end{pmatrix}. \qquad (4.105) \end{aligned}}$$

Satunnaisvektori $\mathbf{x}$ ja satunnaismuuttuja $y - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}$ ovat siis korreloimattomia,

$$\operatorname{cov}(\mathbf{x}, y - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}) = \mathbf{0}_p\,, \tag{4.106a}$$

ja

$$\operatorname{var}(y - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}) = \sigma_y^2 - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y} := \sigma_{y\cdot\mathbf{x}}^2\,. \tag{4.106b}$$

On osoitettavissa että

$$\operatorname{cor}^2(y, \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}) = \frac{\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}}{\sigma_y^2} := \varrho_{y\cdot\mathbf{x}}^2\,, \tag{4.107}$$

missä

$$\varrho_{y\cdot\mathbf{x}} = \text{ populaation yhteiskorrelaatiokerroin.} \tag{4.108}$$

Palaamme lohkodiagonalisointiin satunnaisvektorien yhteydessä tarkemmin luvussa 7.4 (s. 223), mutta mainittakoon jo tässä yhteydessä kerroinvektorin $\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}$ keskeiset ominaisuudet:

$$\min_{\mathbf{b}\in\mathbb{R}^p} \operatorname{var}(y - \mathbf{b}'\mathbf{x}) = \operatorname{var}(y - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}) = \sigma_{y\cdot\mathbf{x}}^2\,, \tag{4.109}$$

$$\max_{\mathbf{b}\neq\mathbf{0}} \operatorname{cor}^2(y, \mathbf{b}'\mathbf{x}) = \operatorname{cor}^2(y, \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}) = \varrho_{y\cdot\mathbf{x}}^2\,. \tag{4.110}$$

Ks. myös (7.90) (s. 226) ja (7.94) (s. 226).

4.4.1 Ositetun matriisin determinantti

Koska matriisien $\mathbf{F}$ ja $\mathbf{U}$ determinantit ovat 1:iä kun

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22\cdot 1} \end{pmatrix}\begin{pmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} = \mathbf{F}'\mathbf{GF}\,, \tag{4.111a}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{A}_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix}\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{pmatrix} = \mathbf{U}'\mathbf{VU}\,, \tag{4.111b}$$

saamme $\mathbf{A}$:n determinantille esitykset

$$
\begin{aligned}
|\mathbf{A}| &= |\mathbf{A}_{11}||\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}| && (4.112a)\\
&= |\mathbf{A}_{22}||\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|\,. && (4.112b)
\end{aligned}
$$

Koska $\mathbf{F}$ (ja $\mathbf{U}$) on epäsingulaarinen, voimme välittömästi päätellä, että $\mathbf{A}$:n aste on sama kuin $\mathbf{G}$:n aste (= $\mathbf{V}$:n aste) (epäsingulaarisella matriisilla kertominen ei muuta matriisin astetta – ks. tarkemmin luku 9.3, s. 306) eli

$$
\begin{aligned}
\mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{G}) &= \mathrm{r}\begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{pmatrix} \\
&= \mathrm{r}(\mathbf{A}_{11}) + \mathrm{r}(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}) && (4.113a)\\
&= \mathrm{r}(\mathbf{F}) = \mathrm{r}\begin{pmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix} \\
&= \mathrm{r}(\mathbf{A}_{22}) + \mathrm{r}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})\,. && (4.113b)
\end{aligned}
$$

Huomattakoon, että hajotelma (4.78) (s. 145) ja astesääntö (4.113a) samoin kuin determinantin hajoitelma (4.112a) ovat voimassa kunhan $\mathbf{A}_{11}^{-1}$ on olemassa; $\mathbf{A}$:n käänteismatriisin ei tarvitse olla olemassa. Vastaavasti (4.87), (4.112b) ja (4.113b) pätevät, jos $\mathbf{A}_{22}$ on epäsingulaarinen. Kyseiset determinantti- ja astesäännöt ovat voimassa myös epäsymmetrisen $\mathbf{A}$:n tapauksessa.

Esimerkki 4.7. Erikoistapauksena determinantin kaavoista (4.112) saamme seuraavan hyödyllisen tuloksen:

$$
\det\begin{pmatrix} \mathbf{I}_u & \mathbf{U}_{u\times v} \\ \mathbf{V}_{v\times u} & \mathbf{I}_v \end{pmatrix} = |\mathbf{I}_u|\,|\mathbf{I}_v - \mathbf{VU}| = |\mathbf{I}_v|\,|\mathbf{I}_u - \mathbf{UV}|\,, \tag{4.114}
$$

mistä seuraa että

$$
|\mathbf{I}_v - \mathbf{VU}| = |\mathbf{I}_u - \mathbf{UV}|\,. \tag{4.115}
$$

Erityisesti jos $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, on voimassa yhtälö

$$
|\mathbf{I}_n - \mathbf{ab}'| = |\mathbf{I}_1 - \mathbf{b}'\mathbf{a}| = 1 - \mathbf{b}'\mathbf{a} = 1 - \mathbf{a}'\mathbf{b}\,. \tag{4.116}
$$

Samoin saamme (kun $\mathbf{A}$ ja $\mathbf{D}$ kääntyviä)

$$
\det\begin{pmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{D} \end{pmatrix} = |\mathbf{A}|\,|\mathbf{D} - \mathbf{VA}^{-1}\mathbf{U}| = |\mathbf{D}|\,|\mathbf{A} - \mathbf{UD}^{-1}\mathbf{V}|\,, \tag{4.117}
$$

ja (kun $\alpha \neq 0$ ja $\mathbf{A}$ kääntyvä)

$$
\det\begin{pmatrix} \mathbf{A} & \mathbf{u} \\ \mathbf{v}' & \alpha \end{pmatrix} = |\mathbf{A}|(\alpha - \mathbf{v}'\mathbf{A}^{-1}\mathbf{u}) = \alpha|\mathbf{A} - \tfrac{1}{\alpha}\mathbf{uv}'|\,, \tag{4.118}
$$

$$
\mathrm{r}\begin{pmatrix} \mathbf{A} & \mathbf{u} \\ \mathbf{v}' & \alpha \end{pmatrix} = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\alpha - \mathbf{v}'\mathbf{A}^{-1}\mathbf{u}) = \mathrm{r}(\alpha) + \mathrm{r}(\mathbf{A} - \tfrac{1}{\alpha}\mathbf{uv}')\,, \tag{4.119}
$$

eli

$$|\mathbf{A}|(\alpha - \mathbf{v}'\mathbf{A}^{-1}\mathbf{u}) = \alpha|\mathbf{A} - \tfrac{1}{\alpha}\mathbf{u}\mathbf{v}'|\,, \tag{4.120}$$

$$\mathrm{r}(\mathbf{A}) + \mathrm{r}(\alpha - \mathbf{v}'\mathbf{A}^{-1}\mathbf{u}) = 1 + \mathrm{r}(\mathbf{A} - \tfrac{1}{\alpha}\mathbf{u}\mathbf{v}')\,. \tag{4.121}$$

Yhtälöstä (4.121) voimme johtaa edelleen seuraavan tuloksen:

$$\mathrm{r}(\mathbf{A} - \tfrac{1}{\alpha}\mathbf{u}\mathbf{v}') = \mathrm{r}(\mathbf{A}) - 1\,, \ \text{kun } \alpha = \mathbf{v}'\mathbf{A}^{-1}\mathbf{u} \neq 0\,. \tag{4.122}$$

□

4.5 Positiivisesti definiitin matriisin kääntäminen

Olkoon symmetrinen $\mathbf{A}_{n\times n}$ positiivisesti definiitti. Silloin on olemassa sellainen $\mathbf{L}_{n\times n}$, että $\mathbf{A} = \mathbf{L}'\mathbf{L}$ (ks. yksityiskohtaisemmin luku 6.3, s. 193). Esitetään nyt $\mathbf{L}$ ositetussa muodossa $\mathbf{L} = (\mathbf{L}_1 : \mathbf{L}_2)$, jolloin

$$\mathbf{A} = \mathbf{L}'\mathbf{L} = \begin{pmatrix} \mathbf{L}_1' \\ \mathbf{L}_2' \end{pmatrix} (\mathbf{L}_1 : \mathbf{L}_2) = \begin{pmatrix} \mathbf{L}_1'\mathbf{L}_1 & \mathbf{L}_1'\mathbf{L}_2 \\ \mathbf{L}_2'\mathbf{L}_1 & \mathbf{L}_2'\mathbf{L}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}. \tag{4.123}$$

Tällöin

$$\boxed{\begin{aligned} \mathbf{A}^{-1} &= \begin{pmatrix} (\mathbf{L}_1'\mathbf{Q}_2\mathbf{L}_1)^{-1} & -(\mathbf{L}_1'\mathbf{L}_1)^{-1}\mathbf{L}_1'\mathbf{L}_2(\mathbf{L}_2'\mathbf{Q}_1\mathbf{L}_2)^{-1} \\ -(\mathbf{L}_2'\mathbf{Q}_1\mathbf{L}_2)^{-1}\mathbf{L}_2'\mathbf{L}_1(\mathbf{L}_1'\mathbf{L}_1)^{-1} & (\mathbf{L}_2'\mathbf{Q}_1\mathbf{L}_2)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{pmatrix}, \qquad (4.124) \end{aligned}}$$

sillä mm.

$$\begin{aligned} \mathbf{A}^{22} &= (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} = [\mathbf{L}_2'\mathbf{L}_2 - \mathbf{L}_2'\mathbf{L}_1(\mathbf{L}_1'\mathbf{L}_1)^{-1}\mathbf{L}_1'\mathbf{L}_2]^{-1} \\ &= (\mathbf{L}_2'[\mathbf{I} - \mathbf{L}_1(\mathbf{L}_1'\mathbf{L}_1)^{-1}\mathbf{L}_1']\mathbf{L}_2)^{-1} = [\mathbf{L}_2'(\mathbf{I} - \mathbf{P}_{\mathbf{L}_1})\mathbf{L}_2]^{-1} \\ &= (\mathbf{L}_2'\mathbf{Q}_1\mathbf{L}_2)^{-1}, \end{aligned} \tag{4.125}$$

missä

$$\mathbf{P}_{\mathbf{L}_1} = \mathbf{P}_1 = \mathbf{L}_1(\mathbf{L}_1'\mathbf{L}_1)^{-1}\mathbf{L}_1'\,, \quad \mathbf{Q}_1 = \mathbf{I} - \mathbf{P}_1\,, \tag{4.126a}$$

$$\mathbf{L}_2'\mathbf{Q}_1\mathbf{L}_2 = \mathbf{L}_2'(\mathbf{I} - \mathbf{P}_1)\mathbf{L}_2 = \mathbf{A}_{22\cdot 1}\,. \tag{4.126b}$$

Matriisin $\mathbf{A}$ determinantille saadaan lausekkeet

$$|\mathbf{A}| = |\mathbf{L}_1'\mathbf{L}_1||\mathbf{L}_2'\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{L}_2| \tag{4.127a}$$

$$= |\mathbf{L}_2'\mathbf{L}_2||\mathbf{L}_1'\,(\mathbf{I} - \mathbf{P}_2)\,\mathbf{L}_1|\,. \tag{4.127b}$$

Matriisin $\mathbf{A}$ aste on

$$\begin{aligned}\mathrm{r}(\mathbf{A}) &= \mathrm{r}(\mathbf{L}_1'\mathbf{L}_1) + \mathrm{r}[\mathbf{L}_2'(\mathbf{I}-\mathbf{P}_1)\mathbf{L}_2] \\ &= \mathrm{r}(\mathbf{L}_1) + \mathrm{r}[(\mathbf{I}-\mathbf{P}_1)\mathbf{L}_2] \\ &= \mathrm{r}(\mathbf{L}_2) + \mathrm{r}[(\mathbf{I}-\mathbf{P}_2)\mathbf{L}_1]. \end{aligned} \tag{4.128}$$

On osoitettavissa, että astekaavat [samoin kuin determinanttikaavat (4.112)] ovat voimassa vaikka $\mathbf{A}$ olisikin singulaarinen; edellytyksenä on, että $\mathbf{A}$ on ei-negatiivisesti definiitti eli $\mathbf{A}$ on esitettävissä muodossa $\mathbf{A} = \mathbf{L}'\mathbf{L}$.

Esimerkki 4.8. Olkoon $\mathbf{z} = \binom{\mathbf{x}}{y}$ $(p+1)$:n elementin satunnaisvektori, jonka kovarianssimatriisi on

$$\mathrm{cov}(\mathbf{z}) = \mathrm{cov}\begin{pmatrix}\mathbf{x} \\ y\end{pmatrix} = \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}_{\mathbf{x}y}' & \sigma_y^2\end{pmatrix} = \mathbf{\Sigma}\,. \tag{4.129}$$

Tällöin $\mathbf{\Sigma}$:n käänteismatriisin viimeinen lävistäjäelementti σ^{yy} on

$$\sigma^{yy} = \frac{1}{\sigma_y^2 - \boldsymbol{\sigma}_{\mathbf{x}y}'\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}} = \frac{1}{\sigma_{y\cdot\mathbf{x}}^2}\,. \tag{4.130}$$

Determinantin lausekkeeksi saadaan

$$|\mathbf{\Sigma}| = |\mathbf{\Sigma}_{\mathbf{xx}}|(\sigma_y^2 - \boldsymbol{\sigma}_{\mathbf{x}y}'\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y})\,, \tag{4.131}$$

mistä $\sigma_{y\cdot\mathbf{x}}^2$:lle saadaan esitykset

$$\begin{aligned}\sigma_{y\cdot\mathbf{x}}^2 &= \sigma_y^2 - \boldsymbol{\sigma}_{\mathbf{x}y}'\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y} = \sigma_y^2\Big(1 - \frac{\boldsymbol{\sigma}_{\mathbf{x}y}'\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}}{\sigma_y^2}\Big) \\ &= \sigma_y^2(1-\varrho_{y\cdot\mathbf{x}}^2) = \frac{|\mathbf{\Sigma}|}{|\mathbf{\Sigma}_{\mathbf{xx}}|}\,. \end{aligned} \tag{4.132}$$

□

Esimerkki 4.9. Olkoon $\mathbf{L}_{n\times n}$ ortogonaalinen matriisi, joka on ositettu siten että

$$\mathbf{L}_{n\times n} = (\mathbf{X}_{n\times p} : \mathbf{Z}_{n\times q})\,, \quad p+q=n\,. \tag{4.133}$$

Tällöin $\mathbf{L}'\mathbf{L} = \mathbf{L}\mathbf{L}' = \mathbf{I}_n,\ \mathbf{L}' = \mathbf{L}^{-1}$ ja

$$\mathbf{L}'\mathbf{L} = \begin{pmatrix}\mathbf{X}' \\ \mathbf{Z}'\end{pmatrix}(\mathbf{X}:\mathbf{Z}) = \begin{pmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}\end{pmatrix} = \begin{pmatrix}\mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q\end{pmatrix} = \mathbf{I}_n\,. \tag{4.134}$$

Olkoon $\mathbf{V}$ jokin symmetrinen positiivisesti definiitti matriisi eli on olemassa epäsingulaarinen $\mathbf{F}$ siten että $\mathbf{V} = \mathbf{F}\mathbf{F}'$. Tällöin on voimassa seuraava tulos:

Olkoon $\mathbf{L} = (\mathbf{X} : \mathbf{Z})$ *ortogonaalinen* $n \times n$*-matriisi ja* $\mathbf{V}_{n\times n}$ *olkoon positiivisesti definiitti. Tällöin*

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}\mathbf{X}\,. \tag{4.135}$$

On montakin tapaa todistaa (4.135), ja yksi tapa on käyttää ositetun matriisin kääntämissääntöä. Merkitään

$$\mathbf{A} = \mathbf{L}'\mathbf{V}\mathbf{L} = \begin{pmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{pmatrix} \mathbf{V}(\mathbf{X} : \mathbf{Z}) = \begin{pmatrix} \mathbf{X}'\mathbf{V}\mathbf{X} & \mathbf{X}'\mathbf{V}\mathbf{Z} \\ \mathbf{Z}'\mathbf{V}\mathbf{X} & \mathbf{Z}'\mathbf{V}\mathbf{Z} \end{pmatrix}. \tag{4.136}$$

Ositetun matriisin kääntötulosten perusteella $\mathbf{A}^{-1}$:n vasen ylänurkka $\mathbf{A}^{11}$ on

$$\mathbf{A}^{11} = [\mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}\mathbf{X}]^{-1}. \tag{4.137}$$

Toisaalta $\mathbf{L}$:n ortogonaalisuudesta seuraa, että

$$\begin{aligned}\mathbf{A}^{-1} &= (\mathbf{L}'\mathbf{V}\mathbf{L})^{-1} = \mathbf{L}^{-1}\mathbf{V}^{-1}(\mathbf{L}')^{-1} = \mathbf{L}'\mathbf{V}^{-1}\mathbf{L} \\ &= \begin{pmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{pmatrix} \mathbf{V}^{-1}(\mathbf{X} : \mathbf{Z}) = \begin{pmatrix} \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{V}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{V}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z} \end{pmatrix}. \end{aligned} \tag{4.138}$$

Väite seuraa välittömästi yhtälöistä (4.137) ja (4.121).

Miten käy, jos $\mathbf{L} = (\mathbf{X} : \mathbf{Z})$ ei olekaan ortogonaalinen, mutta edelleen $\mathbf{L}$ on kääntyvä ja $\mathbf{X}'\mathbf{Z} = \mathbf{0}$? Tällöin on helppo havaita, että

$$\mathbf{L}^{-1} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \end{pmatrix}, \tag{4.139}$$

sillä

$$\begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \end{pmatrix} (\mathbf{X} : \mathbf{Z}) = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix}. \tag{4.140}$$

Täten (4.121):ta vastaten saamme

$$\begin{aligned}\mathbf{A}^{-1} &= \mathbf{L}^{-1}\mathbf{V}^{-1}(\mathbf{L}')^{-1} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \end{pmatrix} \mathbf{V}^{-1}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} : \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}] \\ &= \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & \cdot \\ \cdot & \cdot \end{pmatrix}. \end{aligned} \tag{4.141}$$

Yhtälöiden (4.137) ja (4.141) nojalla on voimassa

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = [\mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}\mathbf{X}]^{-1}, \tag{4.142}$$

eli

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}\mathbf{X}\,, \tag{4.143}$$

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{4.144}$$

Ks. esim. Rao (1967, s. 358). □

4.5.1 Matriisin X′X käänteismatriisi regressioanalyysissä

Tarkastellaan regressioanalyysissä esiintyvää $n \times (k+1)$-mallimatriisia

$$\mathbf{X} = (\mathbf{1} : \mathbf{X}_0) = (\mathbf{1} : \mathbf{x}_1 : \dots : \mathbf{x}_k) = \begin{pmatrix} \mathbf{x}'_{(1)} \\ \mathbf{x}'_{(2)} \\ \vdots \\ \mathbf{x}'_{(n)} \end{pmatrix} \in \mathbb{R}^{n\times(k+1)}. \tag{4.145}$$

Oletamme siis, että mallissa on mukana vakiotermi. Matriisi $\mathbf{X}_0$ on varsinaisten selittäjien muodostama havaintomatriisi. Olkoon $\mathbf{X}$:n sarakkeet vapaat. Osoitamme, että

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} n & \mathbf{1}'\mathbf{X}_0 \\ \mathbf{X}_0'\mathbf{1} & \mathbf{X}_0'\mathbf{X}_0 \end{pmatrix}^{-1} = \begin{pmatrix} \cdot & \cdot \\ \cdot & \mathbf{T}_{\mathbf{xx}}^{-1} \end{pmatrix}, \tag{4.146}$$

missä $\mathbf{T}_{\mathbf{xx}}$ on selittäjien tulosummamatriisi:

$$\mathbf{T}_{\mathbf{xx}} = \mathbf{X}_0'(\mathbf{I} - \mathbf{J})\mathbf{X}_0 = \mathbf{X}_0'\mathbf{C}\mathbf{X}_0\,. \tag{4.147}$$

Käytämme merkintää

$$\begin{pmatrix} n & \mathbf{1}'\mathbf{X}_0 \\ \mathbf{X}_0'\mathbf{1} & \mathbf{X}_0'\mathbf{X}_0 \end{pmatrix}^{-1} = \begin{pmatrix} t^{00} & t^{01} & \dots & t^{0k} \\ t^{10} & t^{11} & \dots & t^{1k} \\ \vdots & \vdots & \ddots & \vdots \\ t^{k0} & t^{k1} & \dots & t^{kk} \end{pmatrix} := \begin{pmatrix} \cdot & \cdot \\ \cdot & \mathbf{T}^{22} \end{pmatrix}. \tag{4.148}$$

Suoraan sijoittamalla saadaan

$$\mathbf{T}^{22} = (\mathbf{X}_0'\mathbf{X}_0 - \mathbf{X}_0'\mathbf{1}\tfrac{1}{n}\mathbf{1}'\mathbf{X}_0)^{-1} = [\mathbf{X}_0'(\mathbf{I} - \mathbf{J})\mathbf{X}_0]^{-1} = \mathbf{T}_{\mathbf{xx}}^{-1}\,. \tag{4.149}$$

Matriisin $\mathbf{X}'\mathbf{X}$ determinantiksi saadaan

$$\begin{aligned}\det(\mathbf{X}'\mathbf{X}) = \det\begin{pmatrix} n & \mathbf{1}'\mathbf{X}_0 \\ \mathbf{X}_0'\mathbf{1} & \mathbf{X}_0'\mathbf{X}_0 \end{pmatrix} &= n|\mathbf{X}_0'\mathbf{X}_0 - \mathbf{X}_0'\mathbf{1}\tfrac{1}{n}\mathbf{1}'\mathbf{X}_0| \\ &= |\mathbf{X}_0'\mathbf{X}_0|\,[n - \mathbf{1}'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{1}]\,,\end{aligned} \tag{4.150}$$

ts.

$$|\mathbf{X}'\mathbf{X}| = n|\mathbf{X}_0'(\mathbf{I} - \mathbf{J})\mathbf{X}_0| = n|\mathbf{T}_{\mathbf{xx}}| = |\mathbf{X}_0'\mathbf{X}_0|\,(n - \mathbf{1}'\mathbf{P}_{\mathbf{X}_0}\mathbf{1})\,. \tag{4.151}$$

Jos $\mathbf{X}$ on ositettu siten, että

$$\mathbf{X} = (\mathbf{X}_1 : \mathbf{x}_k), \quad \mathbf{X}_1 = (\mathbf{1} : \mathbf{x}_1 : \ldots : \mathbf{x}_{k-1}), \tag{4.152}$$

niin

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{x}_k \\ \mathbf{x}_k'\mathbf{X}_1 & \mathbf{x}_k'\mathbf{x}_k \end{pmatrix}^{-1} = \begin{pmatrix} \cdot & \cdot \\ \cdot & [\mathbf{x}_k'(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})\mathbf{x}_k]^{-1} \end{pmatrix}, \tag{4.153}$$

eli $(\mathbf{X}'\mathbf{X})^{-1}$:n viimeinen lävistäjäelementti on

$$t^{kk} = \frac{1}{\mathbf{x}_k'(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})\mathbf{x}_k} = \frac{1}{\|(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})\mathbf{x}_k\|^2}. \tag{4.154}$$

Tietenkin on voimassa

$$\begin{aligned} \|(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})\mathbf{x}_k\| = 0 &\iff (\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})\mathbf{x}_k = \mathbf{0} \\ &\iff \mathbf{P}_{\mathbf{X}_1}\mathbf{x}_k = \mathbf{x}_k \iff \mathbf{x}_k \in \mathscr{C}(\mathbf{X}_1). \end{aligned} \tag{4.155}$$

Samoin voimme päätellä, että jos $\mathbf{x}_k$ on ”melkein” $\mathbf{X}_1$:n sarakkeiden lineaarikombinaatio, niin t^{kk} on ”hyvin” suuri:

$$\mathbf{x}_k \text{ ”}{\in}\text{” } \mathscr{C}(\mathbf{X}_1) \implies t^{kk} \approx \infty. \tag{4.156}$$

Edellä määritettiin $(\mathbf{X}'\mathbf{X})^{-1}$:n viimeinen lävistäjäelementti t^{kk}. Selittäjää $\mathbf{x}_i$ vastaava lävistäjäelementti t^{ii} matriisissa $(\mathbf{X}'\mathbf{X})^{-1}$ on edellisten tarkastelujen perusteella tietenkin

$$t^{ii} = \frac{1}{\mathbf{x}_i'(\mathbf{I} - \mathbf{P}_{\mathbf{X}_{(-i)}})\mathbf{x}_i}, \tag{4.157}$$

missä $\mathbf{X}_{(-i)} = (\mathbf{1} : \mathbf{x}_1 : \ldots : \mathbf{x}_{i-1} : \mathbf{x}_{i+1} : \ldots : \mathbf{x}_k)$.

Esimerkki 4.10. Lineaarisessa mallissa

$$\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n\} = \{\mathbf{y}, (\mathbf{1} : \mathbf{X}_0)\boldsymbol{\beta}, \sigma^2\mathbf{I}_n\} \tag{4.158}$$

parametrivektorin $\boldsymbol{\beta}$ pienimmän neliösumman estimaatti on

$$\begin{aligned} \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} &= \begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}}_{\mathbf{x}} \end{pmatrix} \in \mathbb{R}^{k+1} \\ &= \text{regressiokertoimien PNS-estimaatit, OLSE}(\boldsymbol{\beta}). \end{aligned} \tag{4.159}$$

Osoittautuu, ks. (8.107) (s. 271), että

$$\hat{\beta}_0 = \bar{y} - \hat{\boldsymbol{\beta}}'_{\mathbf{x}}\bar{\mathbf{x}} = \bar{y} - (\hat{\beta}_1\bar{x}_1 + \cdots + \hat{\beta}_k\bar{x}_k)\,, \tag{4.160a}$$

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathbf{x}} &= (\hat{\beta}_1, \dots, \hat{\beta}_k)' = (\mathbf{X}'_0\mathbf{C}\mathbf{X}_0)^{-1}\mathbf{X}'_0\mathbf{C}\mathbf{y} \\ &= (\widetilde{\mathbf{X}}'_0\widetilde{\mathbf{X}}_0)^{-1}\widetilde{\mathbf{X}}'_0\tilde{\mathbf{y}} = \mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y} = \mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{s}_{\mathbf{x}y} \in \mathbb{R}^k.\end{aligned} \tag{4.160b}$$

Lisäksi $\mathbf{X}_0\hat{\boldsymbol{\beta}}_{\mathbf{x}}$ on se $\mathbf{X}_0$:n sarakkeiden lineaarikombinaatio, jolla on maksimaalinen korrelaatio $\mathbf{y}$:n kanssa eli

$$\max_{\mathbf{b}} \operatorname{cor}_{\mathrm{d}}^2(\mathbf{y}, \mathbf{X}_0\mathbf{b}) = \operatorname{cor}_{\mathrm{d}}^2(\mathbf{y}, \mathbf{X}_0\hat{\boldsymbol{\beta}}_{\mathbf{x}}) = \frac{\mathbf{t}'_{\mathbf{x}y}\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}}{t_{yy}} = R^2_{y\cdot\mathbf{x}}\,, \tag{4.161}$$

missä $R^2_{y\cdot\mathbf{x}}$ on regressiomallin selitysaste; ks. (8.223) (s. 288), luku 8.8 (s. 287). On helppo vakuuttautua, että

$$\operatorname{cor}_{\mathrm{d}}^2(\mathbf{y}, \mathbf{X}_0\hat{\boldsymbol{\beta}}_{\mathbf{x}}) = \operatorname{cor}_{\mathrm{d}}^2(\mathbf{y}, \mathbf{X}\hat{\boldsymbol{\beta}})\,. \tag{4.162}$$

□

Harjoitustehtäviä

4.1. Muodosta sellaiset 2×2-matriisit $\mathbf{A}$ ja $\mathbf{B}$, joiden kaikki elementit ovat $\neq 0$ mutta $\mathbf{AB} = \mathbf{0}$.

4.2. Muodosta jokin epäsymmetrinen ortogonaalinen 2×2-matriisi $\mathbf{D}$. ($\mathbf{D}_{n\times n}$ on ortogonaalinen jos $\mathbf{D}'\mathbf{D} = \mathbf{I}_n$.)

4.3. Muodosta jokin epäsymmetrinen ortogonaalinen 3×3-matriisi $\mathbf{E}$.

4.4. Muodosta jokin epäsymmetrinen idempotentti ($\mathbf{P}^2 = \mathbf{P}$) 3×3-matriisi $\mathbf{P}$.

4.5. Jos neliömatriisi $\mathbf{A}_{n\times n}$ on esitettävissä tulona $\mathbf{A} = \mathbf{LL}'$, missä $\mathbf{L}$ on jokin $n \times q$-matriisi, sanomme että $\mathbf{A}$ on ei-negatiivisesti definiitti matriisi. Osoita, että jos $a_{nn} = 0$, niin $\mathbf{A}$:n viimeinen sarake on välttämättä nollavektori (ja samoin viimeinen vaakarivi).

4.6. Olkoon $\mathbf{P}_{n\times n}$ symmetrinen ja idempotentti ts. $\mathbf{P}$ on ortogonaaliprojektori. Osoitettava:

(a) $\mathbf{P}$ ja $\mathbf{I} - \mathbf{P}$ ovat ei-negatiivisesti definiittejä,

(b) $0 \leq p_{ii} \leq 1$, $i = 1, \dots, n$.

(c) $-1/2 \leq p_{ij} \leq 1/2$, $i \neq j$. Ohje: Merkitään $\mathbf{P} = (\mathbf{p}_1 : \ldots : \mathbf{p}_n) = \mathbf{PP} = \mathbf{P}'$. Tällöin

$$\begin{aligned} p_{11} &= \mathbf{p}_1'\mathbf{p}_1 = p_{11}^2 + p_{21}^2 + p_{31}^2 + \cdots + p_{n1}^2, \\ \implies p_{21}^2 &= p_{11} - p_{11}^2 - (p_{31}^2 + \cdots + p_{n1}^2) \leq p_{11} - p_{11}^2 \leq 1/4 \text{ (miksi?)} \end{aligned}$$

4.7. Muodosta jokin symmetrinen 2×2-matriisi $\mathbf{F}$, joka on ei-negatiivisesti definiitti; symmetrinen 2×2-matriisi $\mathbf{G}$, joka ei ole ei-negatiivisesti definiitti.

4.8. Määritä matriisit $\mathbf{y}_{n\times 1}$, $\mathbf{A}_{n\times 3}$, $\mathbf{b}_{3\times 1}$ ja $\mathbf{c}_{n\times 1}$ siten että voit esittää seuraavan yhtälöryhmän muodossa $\mathbf{y} = \mathbf{Ab} + \mathbf{c}$.

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_n + \delta + \varepsilon_n. \end{aligned}$$

4.9. Olkoon $\mathbf{P} = \{p_{ij}\}$ siirtymätodennäköisyysmatriisi, jolloin siis $\mathbf{P}$:n jokaisen vaakarivin elementtien summa on 1.

(a) Osoita että $\mathbf{P}^2$ on myös siirtymätodennäköisyysmatriisi.

(b) Osoita, että vektori $\mathbf{1}$ kuuluu $\mathbf{P}$:n sarakeavaruuteen $\mathscr{C}(\mathbf{P})$ eli on olemassa vektori $\mathbf{a}$ siten, että $\mathbf{Pa} = \mathbf{1}$.

4.10. Olkoon

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Määritä jokin vektori $\mathbf{u} \in \mathbb{R}^7$, joka kuuluu sarakeavaruuteen $\mathscr{C}(\mathbf{A})$, sekä vektori $\mathbf{v} \in \mathbb{R}^7$, joka ei kuulu sarakeavaruuteen $\mathscr{C}(\mathbf{A})$.

4.11. Tarkastellaan muuttujavektorin $\mathbf{y}$ versioita $\tilde{\mathbf{y}}$, $\tilde{\tilde{\mathbf{y}}}$ ja $\overset{*}{\mathbf{y}}$, jotka on määritelty seuraavasti:

$$\tilde{y}_i = y_i - \bar{y}, \quad \tilde{\tilde{y}}_i = \frac{y_i - \bar{y}}{\sqrt{\mathrm{SS}_y}}, \quad \overset{*}{y}_i = \frac{y_i - \bar{y}}{s_y},$$

missä $s_y = y$:n keskihajonta: $s_y^2 = \frac{1}{n-1}\,\mathrm{SS}_y$, $\mathrm{SS}_y = \sum_{i=1}^n (y_i - \bar{y})^2$. Määritä $n \times n$-matriisit $\mathbf{F}$ ja $\mathbf{G}$ siten että (a) $\tilde{\mathbf{y}} = \mathbf{Fy}$, (b) $\tilde{\tilde{\mathbf{y}}} = \mathbf{G}\tilde{\mathbf{y}}$. Määritä:

(c) vektorin $\tilde{\tilde{\mathbf{y}}}$ pituuden neliö $\|\tilde{\tilde{\mathbf{y}}}\|^2$,

(d) muuttujan $\tilde{\tilde{\mathbf{y}}}$ otosvarianssi,

(e) vektorin $\overset{*}{\mathbf{y}}$ pituuden neliö $\|\overset{*}{\mathbf{y}}\|^2$,

(f) muuttujan $\overset{*}{\mathbf{y}}$ otosvarianssi.

4.12. Olkoon $\mathbf{H} = \mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, missä matriisin $\mathbf{X}_{n\times p}$ sarakkeet ovat vapaat. Osoita, että $\operatorname{tr}(\mathbf{H}) = p$.

4.13. Osoitettava:

$$\begin{pmatrix} \mathbf{I}_m & \mathbf{E}_{m\times n} \\ \mathbf{0}_{n\times m} & \mathbf{I}_n \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I} & -\mathbf{E} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}.$$

4.14. Olkoon $\mathbf{A} \in \mathbb{R}^{n\times p}$ ja $\mathbf{B} \in \mathbb{R}^{n\times p}$. Osoita että

(a) $\mathscr{C}\begin{pmatrix} \mathbf{A} \\ \mathbf{0}_{n\times p} \end{pmatrix} \cap \mathscr{C}\begin{pmatrix} \mathbf{0}_{n\times p} \\ \mathbf{B} \end{pmatrix} = \{\mathbf{0}\}\,,$ (b) $\mathscr{C}\begin{pmatrix} \mathbf{A}\mathbf{A}' \\ \mathbf{B}' \end{pmatrix} \cap \mathscr{C}\begin{pmatrix} \mathbf{B} \\ \mathbf{0}_{p\times p} \end{pmatrix} = \{\mathbf{0}\}\,,$

(c) $\operatorname{r}\begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} = \operatorname{r}(\mathbf{A}) + \operatorname{r}(\mathbf{B})\,.$

4.15. Määritä 6 havainnon muuttujavektorit $\mathbf{x}$ ja $\mathbf{y}$, jotka ovat ortogonaalisia mutta eivät korreloimattomia.

4.16. Määritä neliömatriisi $\mathbf{A}$ siten että $\bar{y}^2 = \mathbf{y}'\mathbf{A}\mathbf{y}$, missä $\bar{y}$ = muuttujavektorin $\mathbf{y}$ elementtien (n kpl) keskiarvo.

4.17. Tarkastellaan $n \times 4$-matriisia $\mathbf{X} = (\mathbf{1} : \mathbf{x} : \mathbf{y} : \mathbf{z})$. Osoita, että jos $\mathbf{X}$:n sarakkeet ovat keskenään ortogonaaliset, niin $\operatorname{cor}_{\mathrm{d}}(\mathbf{x}, \mathbf{y}) = 0$.

4.18. Olkoon $\mathbf{A}$ neliömatriisi. Osoita että $\operatorname{tr}(\mathbf{A}'\mathbf{A}) \geq \frac{1}{4}\operatorname{tr}\big[(\mathbf{A} + \mathbf{A}')^2\big]$. Milloin yhtäsuuruus toteutuu?

4.19. Olkoot $n \times n$-matriisit $\mathbf{A}$ ja $\mathbf{B}$ symmetrisiä. Osoita, että

(a) $\dfrac{1}{2}[\operatorname{tr}(\mathbf{A}^2) + \operatorname{tr}(\mathbf{B}^2)] \geq \operatorname{tr}(\mathbf{A}\mathbf{B})\,,$

(b) $\operatorname{tr}\big[(\mathbf{A}\mathbf{B})^2\big] + 2\operatorname{tr}(\mathbf{A}^2\mathbf{B}^2) \geq 0\,.$

4.20. Määritä $\mathbf{H} = \mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, kun

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_c \end{pmatrix}, \quad a + b + c = n\,.$$

4.21. Olkoon $\mathbf{y}' = (3,3,6)$, $\mathbf{1}' = (1,1,1)$, $\mathbf{J} = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$, $\mathbf{C} = \mathbf{I} - \mathbf{J}$ ja

$$\mathbf{A} = \begin{pmatrix} 3 & 3 \\ 6 & 3 \\ 6 & 6 \end{pmatrix}.$$

Laske käsin:

(a) $\mathbf{Jy}$, (b) $\mathbf{JA}$, (c) $\mathbf{Cy}$, (d) $\mathbf{CA}$, (e) $\mathbf{y}'\mathbf{Cy}/(n-1)$.

4.22. Osoita: $\mathbf{A}$ on idempotentti jos ja vain jos $\mathbf{I} - \mathbf{A}$ on idempotentti.

4.23. Olkoon neliömatriisilla $\mathbf{A}$ ominaisuus $\mathbf{A}^2(\mathbf{I} - \mathbf{A}) = \mathbf{0}$. Seuraako tästä, että $\mathbf{A}$ on idempotentti?

4.24. Todista: Olkoon $\mathbf{y}$ jokin annettu $\mathbb{R}^n$:n vektori, $\mathbf{y} \neq \mathbf{0}$. Tällöin

$$\mathbf{y}'\mathbf{Qx} = 0 \ \forall\, \mathbf{Q}_{n\times p} \implies \mathbf{x} = \mathbf{0}\,.$$

Vihje: Olkoon $\mathbf{y}$:n elementti $y_k \neq 0$. Valitaan $\mathbf{Q} = \mathbf{i}_k\mathbf{i}_i'$, jolloin ...

4.25. Olkoon

$$\mathbf{X}_{n\times 2} = \begin{pmatrix} \mathbf{1}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_b \end{pmatrix}, \quad n = a + b \ (0 < a < n)\,.$$

Määritä a:n arvo siten, että

(a) $\operatorname{tr}(\mathbf{X}'\mathbf{X})^{-1}$ on mahdollisimman (1) suuri, (2) pieni,

(b) $\det(\mathbf{X}'\mathbf{X})^{-1}$ on mahdollisimman (1) suuri, (2) pieni.

4.26. Olkoon $\mathbf{R}_{3\times 3} = \operatorname{cor_d}(\mathbf{x}_1 : \mathbf{x}_2 : \mathbf{y})$ tasakorrelaatiomatriisi:

$$\mathbf{R} = (1-r)\mathbf{I}_3 + r\mathbf{1}_3\mathbf{1}_3' = \begin{pmatrix} \mathbf{R_{xx}} & \mathbf{r}_{\mathbf{x}y} \\ \mathbf{r}_{\mathbf{x}y} & 1 \end{pmatrix} = \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix}, \ \mathbf{R_{xx}} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}.$$

Määritä $\mathbf{R}^{-1}$. Osoita lisäksi, että

$$\mathbf{r}_{\mathbf{x}y}'\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{\mathbf{x}y} = \frac{2r^2}{1+r} \quad (= \text{selitysaste kun } y\text{:tä selitetään } x_1\text{:llä ja } x_2\text{:lla})$$

4.27. Määritä vektori $\mathbf{a}$ siten että $\boldsymbol{\sigma}_{\mathbf{x}y} = \boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{a}$, kun

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^k \\ \rho & 1 & \rho & \dots & \rho^{k-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^k & \rho^{k-1} & \rho^{k-2} & \dots & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}_{\mathbf{x}y}' & \sigma_y^2 \end{pmatrix} = \{\rho^{|i-j|}\}\,.$$

4.28. Matriisien $\mathbf{A}$ ja $\mathbf{B}$ euklidinen etäisyys (neliöitynä) on

$$\|\mathbf{A}-\mathbf{B}\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}(a_{ij}-b_{ij})^2 = \mathrm{tr}[(\mathbf{A}-\mathbf{B})'(\mathbf{A}-\mathbf{B})]\,.$$

Osoita: $\|\mathbf{A}-\mathbf{B}\|_F^2 = \mathrm{tr}(\mathbf{A}'\mathbf{A}) + \mathrm{tr}(\mathbf{B}'\mathbf{B}) - 2\,\mathrm{tr}(\mathbf{A}'\mathbf{B})$.

4.29. Cauchyn–Schwarzin epäyhtälössä $(\mathbf{u}'\mathbf{v})^2 \le \mathbf{u}'\mathbf{u}\cdot\mathbf{v}'\mathbf{v}$ vallitsee yhtäsuuruus täsmälleen silloin kun vektorit $\mathbf{u}$ ja $\mathbf{v}$ ovat lineaarisesti riippuvia. Päättele tämän perusteella milloin epäyhtälössä

$$[\mathrm{tr}(\mathbf{A}'\mathbf{B})]^2 \le \mathrm{tr}(\mathbf{A}'\mathbf{A})\cdot\mathrm{tr}(\mathbf{B}'\mathbf{B})$$

on voimassa yhtäsuuruus.

4.30. Olkoon $a+b+c=n$ ja

$$\mathbf{A} = \begin{pmatrix} n & a & b & c\\ a & a & 0 & 0\\ b & 0 & b & 0\\ c & 0 & 0 & c \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 1/a & 0 & 0\\ 0 & 0 & 1/b & 0\\ 0 & 0 & 0 & 1/c \end{pmatrix}.$$

Mitkä seuraavista ehdoista toteutuvat?

(1) $\mathbf{AGA}=\mathbf{A}$, (2) $\mathbf{GAG}=\mathbf{G}$, (3) $(\mathbf{AG})'=\mathbf{AG}$, (4) $(\mathbf{GA})'=\mathbf{GA}$.

4.31. Tarkastellaan ns. *latinalaisia neliömatriiseja*

$$\mathbf{L}_1 = \begin{pmatrix} 1 & 2 & 3 & 4\\ 2 & 1 & 4 & 3\\ 3 & 4 & 1 & 2\\ 4 & 3 & 2 & 1 \end{pmatrix}, \qquad \mathbf{L}_2 = \begin{pmatrix} 1 & 2 & 3 & 4\\ 3 & 1 & 4 & 2\\ 2 & 4 & 1 & 3\\ 4 & 3 & 2 & 1 \end{pmatrix}.$$

Osoita, että $\mathbf{L}_1$ ja $\mathbf{L}_2$ ovat singulaarisia ja määritä niiden asteet. Muodosta jotkut $\mathbf{L}_1^{\perp}$ ja $\mathbf{L}_2^{\perp}$.

Latinalaiseksi neliöksi sanotaan $k\times k$-neliömatriisia, jonka jokaisella rivillä ja sarakkeella esiintyy k lukua $a_1,\dots,a_k$ siten, että kukin luku esiintyy vain yhden kerran jokaisella rivillä ja sarakkeella. Sanomme, että 4×4-latinalainen neliö $\mathbf{L}$ on perusmuodossa, jos $\mathbf{L}$:n ensimmäinen vaakarivi on $1,2,3,4$. Osoita, että kaikkiaan on olemassa 24 perusmuotoista 4×4-latinalaista neliötä.

Latinalaisen neliön sanotaan olevan *criss-cross*-tyyppiä, mikäli sen diagonaalielementit ovat yhtäsuuria ja sen backwards-diagonaalielementit ovat yhtäsuuria. Vakuuttaudu, että $\mathbf{L}_1$ ja $\mathbf{L}_2$ ovat ainoat perusmuotoiset 4×4 criss-cross-latinalaiset neliöt.

Luku 5

Ominaisarvoista

Tekijä ei vieläkään kirjaimellisesti mainitse Rylleä hevoseksi. Torjumme kuitenkin liian mielikuvituksellisena vääristelevän olettamuksen, että häntä olisi korpitiellä vetänyt kesy strutsi.

Mika Waltari (1949): *Neljä päivänlaskua.*

5.1 Määritelmä

Tässä luvussa tarkastelemme joitakin reaalisten matriisien ominaisarvoihin ja ominaisvektoreihin liittyviä ominaisuuksia.

Luku λ (reaaliluku tai kompleksiluku) on neliömatriisin $\mathbf{A}_{n\times n}$ *ominaisarvo*, jos on olemassa sellainen vektori $\mathbf{t} \neq \mathbf{0}$, että

$$\mathbf{A}\mathbf{t} = \lambda\mathbf{t} \tag{5.1}$$

eli

$$(\mathbf{A} - \lambda\mathbf{I}_n)\mathbf{t} = \mathbf{0}. \tag{5.2}$$

Tällöin $\mathbf{t}$ on ominaisarvoa λ vastaava ominaisvektori: $\lambda \in \operatorname{ch}(\mathbf{A})$, kun $\operatorname{ch}(\mathbf{A})$ merkitsee $\mathbf{A}$:n kaikkien ominaisarvojen joukkoa. Pari $(\lambda, \mathbf{t})$ on yksi $\mathbf{A}$:n ominaispari. On selvää, että (5.1) toteutuu triviaalisti, kun $\mathbf{t} = \mathbf{0}$, mutta nollavektori ei siis kelpaa ominaisvektoriksi.

Ominaisarvot (ja vastaavat ominaisvektorit) eivät aina ole reaalisia. Jos kuitenkin $\mathbf{A}$ on symmetrinen, niin $\mathbf{A}$:n ominaisarvot ovat reaalisia ja vastaavat ominaisvektorit voidaan valita reaalisiksi; ks. esim. Rao (1973, s. 38), Searle (1982, ss. 290–291).

Yhtälöstä (5.2) huomaamme, että ominaisarvoa λ vastaava ominaisvektoriksi käy mikä hyvänsä vektori $\mathbf{t}$, joka toteuttaa ehdon

$$\mathbf{t} \in \mathscr{N}(\mathbf{A} - \lambda\mathbf{I}_n)\,, \quad \mathbf{t} \neq \mathbf{0}\,. \tag{5.3}$$

Aliavaruutta $\mathscr{N}(\mathbf{A}-\lambda\mathbf{I}_n)$ sanotaan $\mathbf{A}$:n ominaisarvoa λ vastaavaksi ominaisavaruudeksi ja

$$\dim \mathscr{N}(\mathbf{A}-\lambda\mathbf{I}_n) = \lambda\text{:n geometrinen kertaluku.} \tag{5.4}$$

Yhtälöllä (5.2) on nollasta poikkeva ratkaisu $\mathbf{t}$ täsmälleen silloin kun matriisin $\mathbf{A}-\lambda\mathbf{I}_n$ sarakkeet ovat sidotut (miksi?), mikä tapahtuu täsmälleen silloin kun ns. *ominaisyhtälö*

$$\det(\mathbf{A}-\lambda\mathbf{I}_n) = 0 \tag{5.5}$$

toteutuu. Funktiota

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A}-\lambda\mathbf{I}_n) \tag{5.6}$$

sanotaan $\mathbf{A}$:n *ominaispolynomiksi.* Matriisin $\mathbf{A}$ ominaisarvot ovat siis ominaisyhtälön juuret. Ominaispolynomi on n. asteen polynomi, jolla on n (mahdollisesti kompleksista) juurta, jotka tietenkään kaikki eivät ole välttämättä erisuuria. Esimerkiksi jos $\mathbf{A} = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, niin

$$p_{\mathbf{A}}(\lambda) = \det \begin{pmatrix} 0-\lambda & 1 \\ -1 & 0-\lambda \end{pmatrix} = \lambda^2 + 1\,, \tag{5.7}$$

joten $\mathbf{A}$:lla ei ole yhtään reaalista ominaisarvoa.

Kannattaa huomata, että ominaisarvot ovat ”yksikäsitteisiä” lukuja; ominaisarvoa λ vastaava ominaisvektori $\mathbf{t}$ saadaan ratkaistua yhtälöstä $\mathbf{At} = \lambda\mathbf{t}$ eli ratkaisuksi käy mikä hyvänsä aliavaruuden $\mathscr{N}(\mathbf{A}-\lambda\mathbf{I}_n)$ vektori (nollavektorista poikkeava).

Tarkastellaan sitten hiukan matriisin $\mathbf{A}_{n\times n}$ ominaispolynomia, joka voidaan siis kirjoittaa muodossa

$$\begin{aligned} p_{\mathbf{A}}(\lambda) = \det(\mathbf{A}-\lambda\mathbf{I}_n) &= \det \begin{pmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{pmatrix} \\ &= (-\lambda)^n + c_{n-1}(-\lambda)^{n-1} + \dots + c_1(-\lambda) + c_0\,, \end{aligned} \tag{5.8}$$

missä c_i:t ovat sopivia kertoimia. Toisaalta koska λ_i:t ovat ominaisyhtälön juuria, ominaispolynomi voidaan kirjoittaa muodossa

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A}-\lambda\mathbf{I}_n) = (\lambda_1-\lambda)\cdots(\lambda_n-\lambda)\,, \tag{5.9}$$

mistä nähdään, että $p_{\mathbf{A}}(0) = \det(\mathbf{A}) = \lambda_1\cdots\lambda_n$. Polynomien (5.8) ja (5.9) kertoimien on tietenkin oltava samoja. Esityksestä (5.9) on kohtalaisen helppo päätellä, että $(-\lambda)^{n-1}$:n kerroin on $\lambda_1+\dots+\lambda_n$. On hiukan työläämpi näyttää, että (5.8):ssa $(-\lambda)^{n-1}$:n kerroin on $\operatorname{tr}(\mathbf{A})$, ks. esim. Schott (2005, s. 91), Searle (1982, s. 278); ks myös (4.25) (s. 133). Täten aina on voimassa

$$\det(\mathbf{A}) = \lambda_1 \cdots \lambda_n\,, \quad \mathrm{tr}(\mathbf{A}) = \lambda_1 + \cdots + \lambda_n\,. \tag{5.10}$$

Olkoot $\mathbf{A}$:n ominaisarvot reaalisia. Tällöin usein asetamme ominaisarvot suuruusjärjestykseen ja merkitsemme

- $\mathrm{ch}_i(\mathbf{A}) = \lambda_i, \quad i = 1, 2, \dots, n,$
- $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n,$
- $\mathrm{ch}(\mathbf{A}) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$, kaikkien ominaisarvojen joukko monikerrat mukaanlukien,
- $\mathrm{nzch}(\mathbf{A}) = \{\, \lambda_i : \lambda_i \neq 0 \,\}$, 0:sta poikkeavien ominaisarvojen joukko.

Jos vektori $\mathbf{t}$ toteuttaa (5.1):n, niin myös $\alpha\mathbf{t}$ toteuttaa (5.1):n, koska (5.1) voidaan kertoa puolittain reaaliluvulla α ($\neq 0$). Toisin sanoen: ominaisvektorin pituus ja ”etumerkki” eivät ole yksikäsitteisiä. Usein ominaisvektorit normeerataan pituudeltaan ykkösiksi.

Välittömästi havaitaan, että yksikkömatriisin $\mathbf{I}_n$ ominaisarvot ovat ykkösiä, sillä tällöin ominaisyhtälö saa muodon

$$\det(\mathbf{I}_n - \lambda\mathbf{I}_n) = (1 - \lambda)^n = 0\,. \tag{5.11}$$

Täten 1 on n-kertainen ominaisarvo ja $\mathrm{ch}(\mathbf{I}_n) = \{1, \dots, 1\}$. Ominaisyhtälö on vastaavasti $\mathbf{I}_n\mathbf{t} = \mathbf{t}$ eli mikä hyvänsä $\mathbb{R}^n$:n vektori $\mathbf{t} \neq \mathbf{0}$ voidaan valita ominaisvektoriksi. Tässä tilanteessa ominaisvektoreissa on ei-yksikäsitteisyyttä myös pituuden ja etumerkin lisäksi; näin käy aina kun kyseessä on useampikertainen ominaisarvo.

Esimerkki 5.1. Tarkastellaan korrelaatiomatriisia

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}, \quad \varrho^2 \leq 1\,. \tag{5.12}$$

Tällöin $\boldsymbol{\rho}$:n ominaisarvot λ_1 ja λ_2 ovat yhtälön $|\boldsymbol{\rho} - \lambda\mathbf{I}_2| = 0$, eli yhtälön

$$\det \begin{pmatrix} 1-\lambda & \varrho \\ \varrho & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - \varrho^2 = 0 \tag{5.13}$$

juuret; ts. λ:n on toteutettava yhtälö

$$(1-\lambda)^2 = \varrho^2. \tag{5.14}$$

Yhtälön (5.14) ratkaisut ovat (suuruusjärjestyksessä, kun $\varrho > 0$)

$$\mathrm{ch}_1(\boldsymbol{\rho}) = \lambda_1 = 1 + \varrho\,, \quad \mathrm{ch}_2(\boldsymbol{\rho}) = \lambda_2 = 1 - \varrho\,. \tag{5.15}$$

Havaitsemme että

$$\lambda_1 + \lambda_2 = (1+\varrho) + (1-\varrho) = 2 = \operatorname{tr}(\boldsymbol{\rho}), \tag{5.16}$$

$$\lambda_1 \lambda_2 = (1+\varrho)(1-\varrho) = 1-\varrho^2 = \det(\boldsymbol{\rho}). \tag{5.17}$$

Vastaavat ominaisvektorit saadaan yhtälöistä $(\boldsymbol{\rho} - \lambda_i \mathbf{I}_2)\mathbf{t}_i = \mathbf{0}$, eli

$\lambda_1 = 1+\varrho$:

$$\begin{pmatrix} 1-(1+\varrho) & \varrho \\ \varrho & 1-(1+\varrho) \end{pmatrix} \mathbf{t}_1 = \begin{pmatrix} -\varrho & \varrho \\ \varrho & -\varrho \end{pmatrix} \begin{pmatrix} t_{11} \\ t_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{5.18}$$

$\lambda_2 = 1-\varrho$:

$$\begin{pmatrix} 1-(1-\varrho) & \varrho \\ \varrho & 1-(1-\varrho) \end{pmatrix} \mathbf{t}_2 = \begin{pmatrix} \varrho & \varrho \\ \varrho & \varrho \end{pmatrix} \begin{pmatrix} t_{12} \\ t_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{5.19}$$

Täten ominaisvektorit ovat

$$\mathbf{t}_1 = \begin{pmatrix} a \\ a \end{pmatrix}, \quad \mathbf{t}_2 = \begin{pmatrix} b \\ -b \end{pmatrix}, \quad a \neq 0,\ b \neq 0, \tag{5.20}$$

ja erityisesti ortonormaalit ominaisvektorit ovat

$$\mathbf{t}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{t}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{5.21}$$

□

Esimerkki 5.2. Matriisin $\mathbf{A} = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ ominaisarvot saadaan yhtälöstä

$$\det(\mathbf{A} - \lambda \mathbf{I}_2) = \det \begin{pmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - 1 = 0, \tag{5.22}$$

eli $\operatorname{ch}(\mathbf{A}) = \{2, 0\}$ ja vastaavat ominaisvektorit ovat samat kuin (5.20):ssä. Jos $\mathbf{A} = \mathbf{1}_3 \mathbf{1}_3'$ niin ominaisarvot saadaan yhtälöstä

$$\det(\mathbf{A} - \lambda \mathbf{I}_3) = \det \begin{pmatrix} 1-\lambda & 1 & 1 \\ 1 & 1-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{pmatrix} = 0. \tag{5.23}$$

Determinantin $\det(\mathbf{A} - \lambda \mathbf{I}_3)$ laskemisessa on hivenen vaivaa, ks. Esimerkki 4.3

(s. 138), ja helpommalla päästäänkin jos huomataan että

$$\mathbf{A}\mathbf{t}_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \mathbf{t}_1 = 3 \cdot \mathbf{t}_1 \,, \quad \text{kun } \mathbf{t}_1 = \mathbf{1}_3 \,, \tag{5.24}$$

$$\mathbf{A}\mathbf{t}_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = 0 \cdot \mathbf{t}_2 \,, \tag{5.25}$$

$$\mathbf{A}\mathbf{t}_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = 0 \cdot \mathbf{t}_3 \,. \tag{5.26}$$

Ominaisarvoa 0 vastaaviksi ominaisvektoreiksi $\mathbf{t}_2$ ja $\mathbf{t}_3$ voidaan itse asiassa valita mitkä hyvänsä aliavaruuden $\mathscr{N}(\mathbf{A} - 0 \cdot \mathbf{I}_3) = \mathscr{N}(\mathbf{A})$ vektorit. Kannattaa panna merkille, että $\mathbf{t}_2$ ja $\mathbf{t}_3$ voidaan valita keskenään ortogonaalisiksi. □

Jos tunnetaan $\mathbf{A}$:n ominaisarvot, niin mitä voidaan sanoa matriisin $\mathbf{A} + d\mathbf{I}$, $d \in \mathbb{R}$, ominaisarvoista? Helposti havaitaan, että

$$\begin{aligned} \mathbf{A}\mathbf{t} = \lambda\mathbf{t} &\implies (\mathbf{A} + d\mathbf{I})\mathbf{t} = \lambda\mathbf{t} + d\mathbf{t} = (\lambda + d)\mathbf{t} \\ &\implies \lambda + d \in \mathrm{ch}(\mathbf{A} + d\mathbf{I}) \,. \end{aligned} \tag{5.27}$$

Samoin (jos $\mathbf{A}$ kääntyy):

$$\mathbf{A}\mathbf{t} = \lambda\mathbf{t} \implies \mathbf{t} = \lambda\mathbf{A}^{-1}\mathbf{t} \implies \mathbf{A}^{-1}\mathbf{t} = \frac{1}{\lambda}\mathbf{t} \implies \frac{1}{\lambda} \in \mathrm{ch}(\mathbf{A}^{-1}) \,, \tag{5.28}$$

ts.

$$(\lambda, \mathbf{t}) \text{ on } \mathbf{A}\text{:n ominaispari} \implies (1/\lambda, \mathbf{t}) \text{ on } \mathbf{A}^{-1}\text{:n ominaispari.} \tag{5.29}$$

Jos $(\mathbf{A} + d\mathbf{I})^{-1}$ on olemassa, niin

$$\lambda = \mathbf{A}\text{:n ominaisarvo} \implies \frac{1}{\lambda + d} = (\mathbf{A} + d\mathbf{I})^{-1}\text{:n ominaisarvo.} \tag{5.30}$$

Symmetrisen $\mathbf{A}$:n erisuuria ominaisarvoja vastaavat ominaisvektorit ovat ortogonaalisia. Tämä perustellaan seuraavasti. Olkoot $(\lambda, \mathbf{t})$ ja $(\mu, \mathbf{u})$ matriisin $\mathbf{A}$ ominaispareja:

$$\mathbf{A}\mathbf{t} = \lambda\mathbf{t} \,, \quad \mathbf{A}\mathbf{u} = \mu\mathbf{u} \,. \tag{5.31}$$

Kerrotaan ensimmäinen em. yhtälöistä vasemmalta $\mathbf{u}'$:lla ja toinen $\mathbf{t}'$:lla, jolloin $\mathbf{u}'\mathbf{A}\mathbf{t} = \lambda\mathbf{u}'\mathbf{t}$ ja $\mathbf{t}'\mathbf{A}\mathbf{u} = \mu\mathbf{t}'\mathbf{u}$. Koska $\mathbf{u}'\mathbf{A}\mathbf{t} = \mathbf{t}'\mathbf{A}\mathbf{u}$, on oltava

$$\lambda\mathbf{u}'\mathbf{t} = \mu\mathbf{t}'\mathbf{u} \;\; \text{eli} \;\; (\lambda - \mu)\mathbf{t}'\mathbf{u} = 0 \,, \tag{5.32}$$

mistä λ:n ja μ:n erisuuruuden vuoksi seuraa, että $\mathbf{t}'\mathbf{u} = 0$.

On osoitettavissa, että

♠ jos λ on symmetrisen matriisin $\mathbf{A}$ p-kertainen ominaisarvo eli ominaisyhtälön $\det(\mathbf{A} - \lambda \mathbf{I}_n) = 0$ p-kertainen juuri, on aina mahdollista löytää p ortogonaalista λ:aa vastaavaa ominaisvektoria.

Jos λ on ominaisyhtälön (5.5) p-kertainen juuri, sanotaan lukua p ominaisarvon λ *algebralliseksi kertaluvuksi*. Ominaisarvon λ geometrinen kertaluku määriteltiin aliavaruuden $\mathscr{N}(\mathbf{A} - \lambda \mathbf{I}_n)$ dimensiona; ks. (5.4) (s. 162). On osoitettavissa, että matriisilla $\mathbf{A}_{n\times n}$ on olemassa n lineaarisesti riippumatonta ominaisvektoria täsmälleen silloin kun jokaisen ominaisarvon algebrallinen kertaluku on sama kuin geometrinen kertaluku. Näin käy mm. silloin kun $\mathbf{A}$ on symmetrinen. Jos $\mathbf{A}$ ei ole symmetrinen, asiat mutkistuvat.

5.2 Ominaisarvohajotelma

Olkoon $\mathbf{A}_{n\times n}$ symmetrinen matriisi, ja olkoon $\mathbf{T}_{n\times n}$ matriisi, jonka sarakkeina ovat $\mathbf{A}$:n 1:n pituisiksi normeeratut ominaisvektorit: $\mathbf{T} = (\mathbf{t}_1 : \mathbf{t}_2 : \ldots : \mathbf{t}_n)$. Oletetaan lisäksi, että ominaisvektorit on valittu siten, että ne ovat keskenään ortonormaaleja. (Kuten edellä todettiin, ”valinta” tulee kyseeseen vain kun jonkin ominaisarvon kertaluku on suurempi kuin yksi.) Tällöin matriisi $\mathbf{T}$ on ortogonaalinen eli se toteuttaa ehdon

$$\mathbf{T}'\mathbf{T} = \mathbf{T}'\mathbf{T} = \mathbf{I}_n \,, \quad \mathbf{T}' = \mathbf{T}^{-1}. \tag{5.33}$$

Merkitään ominaisarvojen muodostamaa lävistäjämatriisia $\mathbf{\Lambda}$:lla:

$$\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)\,. \tag{5.34}$$

Tällöin on tietenkin voimassa yhtälö

$$(\mathbf{A}\mathbf{t}_1 : \mathbf{A}\mathbf{t}_2 : \ldots : \mathbf{A}\mathbf{t}_n) = (\lambda_1\mathbf{t}_1 : \lambda_2\mathbf{t}_2 : \ldots : \lambda_n\mathbf{t}_n)\,, \tag{5.35}$$

eli

$$\mathbf{A}(\mathbf{t}_1 : \mathbf{t}_2 : \ldots : \mathbf{t}_n) = (\mathbf{t}_1 : \mathbf{t}_2 : \ldots : \mathbf{t}_n)\begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}, \tag{5.36a}$$

$$\mathbf{A}\mathbf{T} = \mathbf{T}\mathbf{\Lambda}\,. \tag{5.36b}$$

Kertomalla (5.36b) oikealta $\mathbf{T}'$:lla saadaan yhtälö

$$\boxed{\mathbf{A} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}'.} \tag{OAH}$$

Hajotelmaa (OAH) sanotaan $\mathbf{A}$:n *ominaisarvohajotelmaksi* (tai spektraaliesitykseksi) ja se osoittautuu erittäin käyttökelpoiseksi monissa tarkasteluissa. Jos $\mathbf{A}$:n kaikki ominaisarvot ovat erisuuret, on ominaisarvohajotelma tietenkin yksikäsitteinen – $\mathbf{T}$:n sarakkeiden etumerkkejä ja järjestystä lukuunottamatta. Usein on tapana järjestää ominaisarvot suuruusjärjestykseen siten että

$$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = \lambda_{\min}\,. \tag{5.37}$$

Kertomalla (5.36b) vasemmalta $\mathbf{T}'$:lla saadaan tietenkin

$$\boxed{\mathbf{T}'\mathbf{A}\mathbf{T} = \mathbf{\Lambda}\,.} \tag{5.38}$$

Havaitsemme myös, että

$$\begin{aligned}
\mathbf{A} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}' = (\mathbf{t}_1 : \mathbf{t}_2 : \ldots : \mathbf{t}_n)\mathbf{\Lambda}\mathbf{T}' &= (\lambda_1\mathbf{t}_1 : \lambda_2\mathbf{t}_2 : \ldots : \lambda_n\mathbf{t}_n)\begin{pmatrix}\mathbf{t}_1'\\ \mathbf{t}_2'\\ \vdots\\ \mathbf{t}_n'\end{pmatrix}\\
&= \lambda_1\mathbf{t}_1\mathbf{t}_1' + \lambda_2\mathbf{t}_2\mathbf{t}_2' + \cdots \lambda_k\mathbf{t}_k\mathbf{t}_k' + \lambda_{k+1}\mathbf{t}_{k+1}\mathbf{t}_{k+1}' + \cdots + \lambda_n\mathbf{t}_n\mathbf{t}_n'\\
&= \mathbf{T}_{(k)}'\mathbf{\Lambda}_{(k)}\mathbf{T}_{(k)} + \mathbf{T}_{[n-k]}'\mathbf{\Lambda}_{[n-k]}\mathbf{T}_{[n-k]}\,,
\end{aligned} \tag{5.39}$$

missä

$$\mathbf{\Lambda}_{(k)} = \operatorname{diag}(\lambda_1,\ldots,\lambda_k)\,,\quad \mathbf{\Lambda}_{[n-k]} = \operatorname{diag}(\lambda_{k+1},\ldots,\lambda_n)\,, \tag{5.40a}$$

$$\mathbf{T} = (\mathbf{T}_{(k)} : \mathbf{T}_{[n-k]}),\quad \mathbf{T}_{(k)} \in \mathbb{R}^{n\times k},\quad \mathbf{T}_{[n-k]} \in \mathbb{R}^{n\times(n-k)}\,. \tag{5.40b}$$

Triviaalisti on voimassa

$$\mathbf{A} - \lambda_1\mathbf{t}_1\mathbf{t}_1' = \mathbf{T}_{[n-1]}'\mathbf{\Lambda}_{[n-1]}\mathbf{T}_{[n-1]}\,, \tag{5.41a}$$

$$\mathbf{A} - \mathbf{T}_{(k)}'\mathbf{\Lambda}_{(k)}\mathbf{T}_{(k)} = \mathbf{T}_{[n-k]}'\mathbf{\Lambda}_{[n-k]}\mathbf{T}_{[n-k]}\,. \tag{5.41b}$$

Esimerkki 5.3. Olkoon symmetrisellä $\mathbf{A}_{n\times n}$:lla ominaisarvohajotelma $\mathbf{A} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}'$. Tällöin matriisin jäljen kommutatiivisuuden nojalla saamme

$$\operatorname{tr}(\mathbf{A}) = \operatorname{tr}(\mathbf{T}\cdot\mathbf{\Lambda}\mathbf{T}') = \operatorname{tr}(\mathbf{\Lambda}\mathbf{T}'\cdot\mathbf{T}) = \operatorname{tr}(\mathbf{\Lambda})\,. \tag{5.42}$$

Koska tulon determinantti on tulon tekijöiden determinanttien tulo (kunhan rivi- ja sarakeluvut ovat sopivat), on voimassa

$$|\mathbf{A}| = |\mathbf{T}\mathbf{\Lambda}\mathbf{T}'| = |\mathbf{T}|\,|\mathbf{\Lambda}|\,|\mathbf{T}'| = |\mathbf{T}'|\,|\mathbf{T}|\,|\mathbf{\Lambda}| = |\mathbf{T}'\mathbf{T}|\,|\mathbf{\Lambda}| = |\mathbf{\Lambda}|\,. \tag{5.43}$$

Näin on näytetty että symmetriselle $\mathbf{A}$:lle pätee

$$\operatorname{tr}(\mathbf{A}) = \lambda_1 + \lambda_2 + \cdots + \lambda_n\,,\quad |\mathbf{A}| = \lambda_1\lambda_2\cdots\lambda_n\,. \tag{5.44}$$

Tulos (5.44) pätee (5.10):n (s. 163) mukaan itse asiassa myös mille hyvänsä matriisille, ei ainoastaan symmetrisille. □

5.2.1 Neliömuodon x′Ax maksimiarvo

Olkoon $\mathbf{A}_{n\times n}$ symmetrinen ja tarkastellaan neliömuotoa $f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$. Mikä on neliömuodon $\mathbf{x}'\mathbf{A}\mathbf{x}$ suurin arvo, kun asetetaan rajoitus $\mathbf{x}'\mathbf{x} = 1$? Tehtävä ratkeaa helposti ominaisarvohajoitelman $\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}'$ perusteella;

$$\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)\,, \quad \lambda_1 \geq \dots \geq \lambda_n\,, \quad \mathbf{T}'\mathbf{T} = \mathbf{T}\mathbf{T}' = \mathbf{I}_n\,. \tag{5.45}$$

Nimittäin

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{T}\boldsymbol{\Lambda}\mathbf{T}'\mathbf{x} := \mathbf{z}'\boldsymbol{\Lambda}\mathbf{z} \qquad [\mathbf{z} = \mathbf{T}'\mathbf{x}] \\ &= \lambda_1 z_1^2 + \lambda_2 z_2^2 + \dots + \lambda_n z_n^2 \\ &\leq \lambda_1(z_1^2 + z_2^2 + \dots + z_n^2) = \lambda_1 \mathbf{z}'\mathbf{z} = \lambda_1\,, \end{aligned} \tag{5.46}$$

sillä $\mathbf{z}'\mathbf{z} = \mathbf{x}'\mathbf{T}\mathbf{T}'\mathbf{x} = \mathbf{x}'\mathbf{x} = 1$. Yläraja edellä saavutetaan kun valitaan $\mathbf{x} = \mathbf{t}_1$, sillä yhtälön $\mathbf{A}\mathbf{t}_1 = \lambda_1\mathbf{t}_1$ perusteella

$$\mathbf{t}_1'\mathbf{A}\mathbf{t}_1 = \mathbf{t}_1'(\lambda_1\mathbf{t}_1) = \lambda_1\mathbf{t}_1'\mathbf{t}_1 = \lambda_1\,. \tag{5.47}$$

Siis:

$$\operatorname{ch}_{\max}(\mathbf{A}) = \operatorname{ch}_1(\mathbf{A}) = \lambda_1 = \max_{\mathbf{x}'\mathbf{x}=1} \mathbf{x}'\mathbf{A}\mathbf{x}\,. \tag{5.48}$$

Koska osamäärä (ns. *Rayleigh'n osamäärä*)

$$\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \tag{5.49}$$

on riippumaton $\mathbf{x}$:n (euklidisesta) pituudesta (miksi?), on voimassa:

$$\operatorname{ch}_{\max}(\mathbf{A}) = \operatorname{ch}_1(\mathbf{A}) = \lambda_1 = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \max_{\mathbf{x}'\mathbf{x}=1} \mathbf{x}'\mathbf{A}\mathbf{x}\,. \tag{5.50}$$

Vastaavasti on osoitettavissa että

$$\operatorname{ch}_{\min}(\mathbf{A}) = \operatorname{ch}_n(\mathbf{A}) = \lambda_n = \min_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \min_{\mathbf{x}'\mathbf{x}=1} \mathbf{x}'\mathbf{A}\mathbf{x}\,. \tag{5.51}$$

Merkitään sitten

$$\mathbf{T}_{(k)} = (\mathbf{t}_1 : \dots : \mathbf{t}_k) \quad \mathbf{T}\text{:n } k \text{ ensimmäistä saraketta}\,, \tag{5.52a}$$

$$\mathbf{T}_{[k]} = (\mathbf{t}_{n-k+1} : \dots : \mathbf{t}_n) \quad \mathbf{T}\text{:n } k \text{ viimeistä saraketta}\,. \tag{5.52b}$$

Toiseksi suurin ominaisarvo λ_2 saadaan seuraavasti:

$$\operatorname{ch}_2(\mathbf{A}) = \lambda_2 = \max_{\mathbf{x}\in\mathcal{L}} \mathbf{x}'\mathbf{A}\mathbf{x}\,, \tag{5.53}$$

missä maksimointi tapahtuu seuraavassa joukossa:

$$\begin{aligned}\mathcal{L} &= \{\, \mathbf{x} \in \mathbb{R}^n : \mathbf{x}'\mathbf{x} = 1,\ \mathbf{x}'\mathbf{t}_1 = 0 \,\} \\ &= \{\, \mathbf{x} \in \mathbb{R}^n : \mathbf{x}'\mathbf{x} = 1,\ \mathbf{x} \in \mathscr{C}(\mathbf{t}_2 : \ldots : \mathbf{t}_n) \,\} \\ &= \{\, \mathbf{x} \in \mathbb{R}^n : \mathbf{x}'\mathbf{x} = 1,\ \mathbf{x} \in \mathscr{C}(\mathbf{T}_{[n-1]}) \,\}\,. \end{aligned} \tag{5.54}$$

Yllä on käytetty hyväksi hajotelmaa

$$\mathbb{R}^n = \mathscr{C}(\mathbf{t}_1) \boxplus \mathscr{C}(\mathbf{t}_1)^\perp = \mathscr{C}(\mathbf{t}_1) \boxplus \mathscr{C}(\mathbf{T}_{[n-1]})\,, \tag{5.55}$$

ja siten

$$\mathbf{x}'\mathbf{t}_1 = 0 \iff \mathbf{x} \in \mathscr{C}(\mathbf{T}_{[n-1]})\,. \tag{5.56}$$

Ominaisarvo λ_i voidaan luonnehtia vastaavalla tavalla.

Varmistetaan vielä, että (5.53):n mukainen maksimointi antaa tulokseksi λ_2:n. Olkoon $\mathbf{x} \in \mathscr{C}(\mathbf{T}_{[n-1]})$, $\mathbf{x}'\mathbf{x} = 1$, jolloin on olemassa $\mathbf{a} \in \mathbb{R}^{n-1}$ siten, että $\mathbf{x} = \mathbf{T}_{[n-1]}\mathbf{a}$. Tällöin

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{a}'\mathbf{T}'_{[n-1]}\mathbf{A}\mathbf{T}_{[n-1]}\mathbf{a} = \mathbf{a}'\boldsymbol{\Lambda}_{[n-1]}\mathbf{a}\,. \tag{5.57}$$

Lausekkeen (5.57) maksimi ehdolla $\mathbf{a}'\mathbf{a} = 1$ (mikä on yhtäpitävää ehdon $\mathbf{x}'\mathbf{x} = 1$ kanssa) on tietenkin λ_2.

Seuraavassa on vielä yhteenveto joistakin ominaisarvojen ominaisuuksista. Olkoon $\mathbf{A}_{n\times n}$ symmetrinen. Tällöin:

(a) $\mathrm{ch}_1(\mathbf{A}) = \lambda_1 = \max\limits_{\mathbf{x}\neq\mathbf{0}} \dfrac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \max\limits_{\mathbf{x}'\mathbf{x}=1} \mathbf{x}'\mathbf{A}\mathbf{x}$,

(b) $\mathrm{ch}_n(\mathbf{A}) = \lambda_n = \min\limits_{\mathbf{x}\neq\mathbf{0}} \dfrac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \min\limits_{\mathbf{x}'\mathbf{x}=1} \mathbf{x}'\mathbf{A}\mathbf{x}$,

(c) $\mathrm{ch}_n(\mathbf{A}) = \lambda_n \leq \dfrac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \leq \lambda_1 = \mathrm{ch}_1(\mathbf{A})$,

(d) $\mathrm{ch}_2(\mathbf{A}) = \lambda_2 = \max\limits_{\mathbf{x}'\mathbf{x}=1,\, \mathbf{t}_1'\mathbf{x}=0} \mathbf{x}'\mathbf{A}\mathbf{x}$,

(e) $\mathrm{ch}_{k+1}(\mathbf{A}) = \lambda_{k+1} = \max\limits_{\mathbf{x}'\mathbf{x}=1,\, \mathbf{T}'_{(k)}\mathbf{x}=\mathbf{0}} \mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_{k+1},\ k = 1, \ldots, n-1$.

5.2.2 Antiominaisarvot

Olkoon $\mathbf{A}_{n\times n}$ symmetrinen positiivisesti definiitti matriisi (ks. luku 6.3, s. 193), jolloin sen kaikki ominaisarvot ovat nollaa suurempia. Huomaamme, että ominaisarvoa $\lambda \in \mathbb{R}$ vastaava ominaisvektori $\mathbf{t}$ on sellainen vektori, että kun se kerrotaan matriisilla $\mathbf{A}$, niin tulokseksi tulee ”vain” $\mathbf{t}$:n monikerta $\lambda\mathbf{t}$. Toisin sanoen, vektorin $\mathbf{t}$ kertominen $\mathbf{A}$:lla antaa tulokseksi $\mathbf{t}$:n monikerran, joka siis voi olla eripituinen kuin $\mathbf{t}$. Tämä merkitsee, että vektorien $\mathbf{t}$ ja $\mathbf{At}$ välisen kulman kosini on 1:

$$\cos(\mathbf{t}, \mathbf{At}) = \frac{\mathbf{t}'\mathbf{At}}{\sqrt{\mathbf{t}'\mathbf{t} \cdot \mathbf{t}'\mathbf{A}'\mathbf{At}}} = \frac{\mathbf{t}'\mathbf{At}}{\sqrt{\mathbf{t}'\mathbf{t} \cdot \mathbf{t}'\mathbf{A}^2\mathbf{t}}} = 1\,. \tag{5.58}$$

Etsimme siis sellaista vektoria $\mathbf{x}$, jolla on mahdollimman pieni kulma vektorin $\mathbf{Ax}$ kanssa (0 °) ts $\cos(\mathbf{x}, \mathbf{Ax})$ olisi mahdollisimman suuri (1). Harjoitustehtävänä on näyttää että (5.58) on yhtäpitävä seuraavan kanssa:

$$\mathbf{t}'\mathbf{At} - (\mathbf{t}'\mathbf{A}^{-1}\mathbf{t})^{-1} = 0, \quad \text{missä } \mathbf{t}'\mathbf{t} = 1. \tag{5.59}$$

Voimme muuttaa tehtävää siten, että haemme sellaista vektoria $\mathbf{x}$, jolla on mahdollimman suuri kulma vektorin $\mathbf{Ax}$ kanssa ts. $\cos(\mathbf{x}, \mathbf{Ax})$ olisi mahdollisimman pieni. On osoitettavissa, että

$$\mu_1 := \frac{2\sqrt{\lambda_1\lambda_n}}{\lambda_1 + \lambda_n} \leq \frac{\mathbf{x}'\mathbf{Ax}}{\sqrt{\mathbf{x}'\mathbf{x} \cdot \mathbf{x}'\mathbf{A}^2\mathbf{x}}} \leq 1\,, \tag{5.60}$$

missä $(\lambda_i, \mathbf{t}_i)$ on matriisin $\mathbf{A}$ i. ominaispari. Lausekkeen $\cos(\mathbf{x}, \mathbf{Ax})$ minimiarvoa μ_1 sanotaan ensimmäiseksi antiominaisarvoksi ja siihen johtavaa vektoria (vektoriparia) vastaavaksi antiominaisvektoriksi:

$$\frac{1}{\sqrt{\lambda_1}}\mathbf{t}_1 \pm \frac{1}{\sqrt{\lambda_n}}\mathbf{t}_n\,, \tag{5.61}$$

jotka voidaan normeerata ykkösen pituisiksi:

$$\frac{\sqrt{\lambda_n}\mathbf{t}_1 \pm \sqrt{\lambda_1}\mathbf{t}_n}{\sqrt{\lambda_1 + \lambda_n}}\,. \tag{5.62}$$

Kirjallisuutta antiominaisarvoista: Gustafson (2012), Rao (2007), and Puntanen, Styan & Isotalo (2011, s. 237).

5.2.3 Kovarianssimatriisin $\Sigma_{2\times2}$ ominaisarvot

Tarkastellaan yksinkertaista esimerkkiä, jossa kovarianssimatriisina on

$$\operatorname{cov}(\mathbf{z}) = \operatorname{cov}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} = \mathbf{\Sigma}\,. \tag{5.63}$$

Karakteristinen polynomi on

$$p_{\mathbf{\Sigma}}(\lambda) = \lambda^2 - (\sigma_x^2 + \sigma_y^2)\lambda + (\sigma_x^2\sigma_y^2 - \sigma_{xy}^2) = \lambda^2 - \operatorname{tr}(\mathbf{\Sigma})\lambda + \det(\mathbf{\Sigma})\,, \tag{5.64}$$

ja $\mathbf{\Sigma}$:n ominaisarvot ovat

$$\lambda_1, \lambda_2 = \frac{1}{2}\left[\sigma_x^2 + \sigma_y^2 \pm \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2}\,\right]. \tag{5.65}$$

Päivänselvästi

$$\operatorname{tr}(\mathbf{\Sigma}) = \sigma_x^2 + \sigma_y^2 = \lambda_1 + \lambda_2\,, \tag{5.66a}$$

$$\det(\mathbf{\Sigma}) = \sigma_x^2\sigma_y^2\left(1 - \frac{\sigma_{xy}^2}{\sigma_x^2\sigma_y^2}\right) = \sigma_x^2\sigma_y^2(1-\varrho_{xy}^2) = \lambda_1\lambda_2\,. \tag{5.66b}$$

Ominaisarvoa λ_1 vastaava ominaisvektori $\mathbf{t}_1 = \binom{t_{11}}{t_{21}}$ on nollasta eroava ratkaisu yhtälölle $(\mathbf{\Sigma} - \lambda_1\mathbf{I}_2)\mathbf{t}_1 = \mathbf{0}$, ts.,

$$(\sigma_x^2 - \lambda_1)t_{11} + \sigma_{xy}t_{21} = 0\,, \tag{5.67a}$$

$$\sigma_{xy}t_{11} + (\sigma_y^2 - \lambda_1)t_{21} = 0\,. \tag{5.67b}$$

Olkoon $\sigma_x^2 > \sigma_y^2$. Tällöin on vaivatonta päätellä, että t_{21} voidaan asettaa 0:ksi jos ja vain jos $\sigma_{xy} = 0$, jolloin t_{11}:ksi voidaan valita mikä hyvänsä nollasta pokkeava reaaliluku. Kannattaa panna merkille, että suhde t_{21}/t_{11} määrittää ominaisvektorin $\mathbf{t}_1$.

Erityisesti jos $\sigma_x^2 = \sigma_y^2 = \sigma^2$ ja $\sigma_{xy} = \sigma^2\varrho_{xy} > 0$, niin saamme

$$\lambda_1 = \sigma^2 + \sigma_{xy} = \sigma^2(1+\varrho_{xy})\,, \quad \lambda_2 = \sigma^2 - \sigma_{xy} = \sigma^2(1-\varrho_{xy})\,, \tag{5.68}$$

ja normeeratut ominaisvektorit ovat

$$\mathbf{t}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}, \quad \mathbf{t}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix} = \begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}, \tag{5.69}$$

missä $\theta = 45°$. Mikäli $\sigma_{xy} < 0$, niin $\lambda_1 = \sigma^2 - \sigma_{xy}$ ja $\mathbf{t}_1 = \frac{1}{\sqrt{2}}(-1,1)'$. Jos $\sigma_{xy} = 0$, niin $\lambda_1 = \lambda_2 = \sigma^2$ ja $\mathbf{t}_1$:ksi voidaan valita mikä hyvänsä $\mathbb{R}^2$:n nollasta poikkeava vektori.

Tarkastellaan sitten satunnaisvektoria $\mathbf{z} \sim \mathrm{N}_2(\boldsymbol{\mu}, \mathbf{\Sigma})$, missä $\mathbf{\Sigma}$ on positiivisesti definitti. Silloin $\mathbf{z}$:n tiheysfunktio on

$$n(\mathbf{z}; \boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{1/2}|\mathbf{\Sigma}|^{1/2}}\mathrm{e}^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}. \tag{5.70}$$

On ilmeistä, että $\mathbf{z}$:n arvojen joukko, joilla tiheysfunktion (5.70) arvo on sama ovat ellipsejä; pistejoukkoja, jotka toteuttavat yhtälön

$$\mathcal{A} = \{\, \mathbf{z} \in \mathbb{R}^2 : (\mathbf{z}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu}) = c^2 \,\}\,, \tag{5.71}$$

missä c on jokin annettu nollasta poikkeava reaaliluku. Miten ellipsin akselit sijatsevat? Pääakseli (1. pääakseli) on $\boldsymbol{\mu}$:n kautta kulkeva pisin jana, jonka päätepisteet ovat ellipsin kehällä. Toisin sanoen, haluamme määrittää pisteen $\mathbf{z}_1$ ellipsin kehältä siten, että sen euklidinen etäisyys $\boldsymbol{\mu}$:stä on mahdollisimman suuri:

$$\max\|\mathbf{z}-\boldsymbol{\mu}\|^2 \quad \text{ehdolla } \mathbf{z} \in \mathcal{A}\,. \tag{5.72}$$

Kun merkitään $\mathbf{u} = \mathbf{z} - \boldsymbol{\mu}$, niin em. tehtävä saa asun

$$\max \mathbf{u}'\mathbf{u} \quad \text{ehdolla } \mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u} = c^2. \tag{5.73}$$

Ei ole hankala päätellä että (5.73):n ratkaisu on

$$\mathbf{u}_1 = \mathbf{z}_1 - \boldsymbol{\mu} = \pm c\sqrt{\lambda_1}\mathbf{t}_1\,, \quad \text{ja } \mathbf{u}_1'\mathbf{u}_1 = c^2\lambda_1. \tag{5.74}$$

Ellipsin 2. pääakseli saadaan määrittämällä $\mathcal{A}$:n kehältä vektori, jonka etäisyys $\boldsymbol{\mu}$:stä on mahdollisimman pieni. Yhtäpitävä määritelmä on se, että haetaan vektoria $\mathbf{z}_2 \in \mathcal{A}$, joka maksimoi lausekkeen $\|\mathbf{z} - \boldsymbol{\mu}\|$ ehdoilla $\mathbf{z} \in \mathcal{A}$, $\mathbf{z}'\mathbf{z}_1 = 0$.

Yksi tapa vakuuttautua siitä, että (5.74) on (5.73):n ratkaisu, on huomata, että (5.73) voidaan esittää muodossa

$$\max \mathbf{v}'\boldsymbol{\Sigma}\mathbf{v} \quad \text{ehdolla } \mathbf{v}'\mathbf{v} = c^2, \tag{5.75}$$

missä $\mathbf{v} = \boldsymbol{\Sigma}^{-1/2}\mathbf{u}$ ja täten $\mathbf{u} = \boldsymbol{\Sigma}^{1/2}\mathbf{v}$. Yksityiskohdat jätetään harjoitustehtäväksi.

Kommentti 5.1. Todettakoon vielä yhteenvetona, että ellipsin $\mathcal{A}$ keskipiste on $\boldsymbol{\mu}$ ja akselit määräytyvät lausekkeista $\mathbf{u}_i = \pm c\sqrt{\lambda_i}\mathbf{t}_i$, missä $\lambda_i = \mathrm{ch}_i(\boldsymbol{\Sigma})$ ja $\mathbf{t}_i$ on vastaava ominaisvektori. Jos $\sigma_x^2 = \sigma_y^2 = \sigma^2$ ja $\sigma_{xy} = \sigma^2\varrho_{xy} > 0$, niin 1. pääakseli $\mathbf{u}_1$ kulkee 45°:n kulmassa pisteen $\boldsymbol{\mu} = \left(\begin{smallmatrix}\mu_x\\ \mu_y\end{smallmatrix}\right)$ kautta. Akseleiden (itse asiassa puoliakseleiden) pituudet ovat

$$c\sqrt{\sigma^2 + \sigma_{xy}} = c\sigma\sqrt{1 + \varrho_{xy}} \quad \text{ja} \quad c\sqrt{\sigma^2 - \sigma_{xy}} = c\sigma\sqrt{1 - \varrho_{xy}}\,. \tag{5.76}$$

5.2.4 Tulon AB nollasta poikkeavat ominaisarvot

Todistamme seuraavan hyödyllisen tuloksen:

Olkoon $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{B} \in \mathbb{R}^{m\times n}$. Tällöin matriiseilla $\mathbf{AB} \in \mathbb{R}^{n\times n}$ ja $\mathbf{BA} \in \mathbb{R}^{m\times m}$ on samat nollasta poikkeavat ominaisarvot eli

$$\mathrm{nzch}(\mathbf{AB}) = \mathrm{nzch}(\mathbf{BA})\,. \tag{5.77}$$

Tuloksen (5.77) osoittamiseksi palautamme aluksi mieleen vakiolla kertomisen vaikutuksen determinanttiin. Nimittäin koska

$$|\alpha\mathbf{C}_{n\times n}| = \alpha^n|\mathbf{C}_{n\times n}|\,,\ \alpha \in \mathbb{R}\,, \tag{5.78}$$

on voimassa

$$|\lambda\mathbf{I}_n - \mathbf{C}_{n\times n}| = |(-1)(\mathbf{C}_{n\times n} - \lambda\mathbf{I}_n)| = (-1)^n|\mathbf{C} - \lambda\mathbf{I}_n|\,, \tag{5.79}$$

joten yhtälöillä $|\mathbf{C}-\lambda\mathbf{I}|=0$ ja $|\lambda\mathbf{I}-\mathbf{C}|=0$ on samat juuret. Täten $\mathbf{AB}$:n ominaisarvot saadaan yhtälöstä

$$|\lambda\mathbf{I}_n-\mathbf{AB}|=0\,. \tag{5.80}$$

Käyttäen esimerkin 4.7 (s. 150) tulosta $|\mathbf{I}_v-\mathbf{VU}|=|\mathbf{I}_u-\mathbf{UV}|$ ja ominaisuutta (5.78):ta saamme

$$\begin{aligned}|\lambda\mathbf{I}_n-\mathbf{AB}| &= |\lambda(\mathbf{I}_n-\tfrac{1}{\lambda}\mathbf{AB})|=\lambda^n|\mathbf{I}_n-\tfrac{1}{\lambda}\mathbf{AB}|\\ &= \lambda^n|\mathbf{I}_m-\tfrac{1}{\lambda}\mathbf{BA}|=\lambda^n|\tfrac{1}{\lambda}(\lambda\mathbf{I}_m-\mathbf{BA})|\\ &= \lambda^n\lambda^{-m}|\lambda\mathbf{I}_m-\mathbf{BA}|=\lambda^{n-m}|\lambda\mathbf{I}_m-\mathbf{BA}|\,,\end{aligned} \tag{5.81}$$

eli

$$|\lambda\mathbf{I}_n-\mathbf{AB}|=\lambda^{n-m}|\lambda\mathbf{I}_m-\mathbf{BA}|\,. \tag{5.82}$$

Yhtälöstä (5.82) voimme päätellä, että yhtälöillä

$$|\lambda\mathbf{I}_n-\mathbf{AB}|=0 \quad \text{ja} \quad |\lambda\mathbf{I}_m-\mathbf{BA}|=0 \tag{5.83}$$

on samat nollasta poikkeavat juuret, ja täten (5.77) on todistettu.

Toinen (yksinkertaisempi) tapa todistaa (5.77) on seuraava; ks. Seber (1984, s. 518). Olkoon λ nollasta poikkeava $\mathbf{AB}$:n ominaisarvo. Tällöin on olemassa sellainen $\mathbf{u}\neq\mathbf{0}$ että $\mathbf{ABu}=\lambda\mathbf{u}$, mistä seuraa, että $\mathbf{BABu}=\lambda\mathbf{Bu}$, missä $\mathbf{v}:=\mathbf{Bu}\neq\mathbf{0}$ (sillä $\mathbf{ABu}\neq\mathbf{0}$). Täten $(\mathbf{BA})\mathbf{v}=\lambda\mathbf{v}$, ja λ on $\mathbf{BA}$:n ominaisarvo. Aivan vastaavasti voidaan näyttää, että jos λ on $\mathbf{BA}$:n nollasta poikkeava ominaisarvo, niin λ on myös $\mathbf{AB}$:n ominaisarvo.

Esimerkkinä voimme tarkastella matriisin $\mathbf{aa}'$ ($\mathbf{a}\in\mathbb{R}^n$) ominaisarvoja. Tällöin (5.77):n mukaan

$$\mathrm{nzch}(\mathbf{aa}')=\mathrm{nzch}(\mathbf{a}'\mathbf{a})=\{\mathbf{a}'\mathbf{a}\}, \tag{5.84}$$

eli $\mathbf{aa}'$:n ainoa nollasta poikkeava ominaisarvo on $\mathbf{a}'\mathbf{a}$; luku 0 on $\mathbf{aa}'$:n $(n-1)$-kertainen ominaisarvo.

5.2.5 Idempotentin matriisin ominaisarvot

Ortogonaaliprojektorin $\mathbf{P}_\mathbf{A}=\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, $\mathbf{A}\in\mathbb{R}^{n\times m}$, nollasta poikkeavat ominaisarvot ovat kaikki ykkösiä sillä

$$\mathrm{nzch}[\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']=\mathrm{nzch}[\mathbf{A}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}]=\mathrm{nzch}(\mathbf{I}_m)\,. \tag{5.85}$$

Itse asiassa on helppo osoittaa seuraava tulos:

Olkoon $\mathbf{P}$ *symmetrinen. Tällöin*

$$\mathbf{P}^2=\mathbf{P} \iff \mathrm{ch}(\mathbf{P})=\{0,\dots,0,1,\dots,1\}\,. \tag{5.86}$$

Olkoon symmetrisellä $\mathbf{P}$:llä ominaisarvohajoitelma $\mathbf{P} = \mathbf{T\Lambda T'}$. Tällöin

$$\begin{aligned}\mathbf{P}^2 = \mathbf{P} &\iff (\mathbf{T\Lambda T'})(\mathbf{T\Lambda T'}) = \mathbf{T\Lambda T'} \\ &\iff \mathbf{T\Lambda}^2\mathbf{T'} = \mathbf{T\Lambda T'} \iff \mathbf{\Lambda}^2 = \mathbf{\Lambda} \\ &\iff \lambda_i^2 = \lambda_i \iff \lambda_i = 0 \text{ tai } \lambda_i = 1\,, \end{aligned} \tag{5.87}$$

ja näin on (5.86) todistettu.

Tarkastellaan sitten idempotenttia mutta mahdollisesti epäsymmetristä matriisia $\mathbf{P}$. Olkoon λ matriisin $\mathbf{P}$ ominaisarvo ja $\mathbf{u}$ vastaava ominaisvektori. Tällöin

$$\begin{aligned}\mathbf{Pu} = \lambda\mathbf{u} &\implies \mathbf{P}^2\mathbf{u} = \mathbf{Pu} = \lambda\mathbf{Pu} = \lambda^2\mathbf{u} \\ &\implies \mathbf{Pu} = \lambda\mathbf{u} = \lambda^2\mathbf{u} \implies \lambda = 0 \text{ tai } \lambda = 1\,, \end{aligned} \tag{5.88}$$

eli idempotentin matriisin ominaisarvot ovat aina nollia ja/tai ykkösiä:

$$\boxed{\mathbf{P}^2 = \mathbf{P} \implies \operatorname{ch}(\mathbf{P}) = \{0, \dots, 0, 1, \dots, 1\}\,.} \tag{5.89}$$

On korostettava, että epäsymmetrisen $\mathbf{P}$:n tapauksessa implikaatio (5.89) pätee vain " $\implies$ "-suunnassa. Esimerkiksi

$$\operatorname{ch}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \{0, 1\}, \text{ mutta } \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}. \tag{5.90}$$

Matriisia $\mathbf{P}$ sanotaan *tripotentiksi*, jos $\mathbf{P}^3 = \mathbf{P}$. On helppo osoittaa seuraava tripotentin matriisin ominaisuus:

$$\mathbf{P}^3 = \mathbf{P} \implies \operatorname{ch}(\mathbf{P}) = \{-1, \dots, -1, 0, \dots, 0, 1, \dots, 1\}\,. \tag{5.91}$$

Matriisi $\mathbf{P}$ on *nilpotentti*, jos on olemassa sellainen positiivinen kokonaisluku q, että $\mathbf{P}^q = \mathbf{0}$.

5.3 Singulaariarvohajotelma

Entuudestaan tiedämme, että jos neliömatriisi $\mathbf{A}_{n\times n}$ on symmetrinen, on sillä ominaisarvohajotelma

$$\mathbf{A} = \mathbf{T\Lambda T'}, \tag{OAH}$$

missä $\mathbf{T}$ on ortogonaalinen matriisi sarakkeina $\mathbf{A}$:n 1:n pituisiksi normeeratut ominaisvektorit ja $\mathbf{\Lambda}$ on $\mathbf{A}$:n ominaisarvojen muodostama lävistäjämatriisi:

$$\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)\,, \quad \lambda_1 \geq \dots \geq \lambda_n\,, \quad \mathbf{T'T} = \mathbf{TT'} = \mathbf{I}_n\,. \tag{5.92}$$

Entä jos $\mathbf{A}$ ei ole symmetrinen tai jos $\mathbf{A}$ ei ole neliömatriisi? Mikä olisi tässä tapauksessa (OAH):ta vastaava hajotelma ts. hajotelma joka "toimisi (OAH):n tyyliin"? Vastaukseksi osoittautuu *singulaariarvohajotelma*.

Olkoon matriisin $\mathbf{A}_{n\times m}$ $(m \leq n)$ aste r. Tällöin $\mathbf{A}$:lla on singulaariarvohajotelma

$$\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{\Delta}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}' = \mathbf{U}\mathbf{\Delta}\mathbf{V}', \tag{SAH}$$

missä $\mathbf{U}_{n\times n}$ ja $\mathbf{V}_{m\times m}$ ovat ortogonaalisia ja

$$\mathbf{U}_{n\times n} = (\mathbf{U}_1 : \mathbf{U}_0)\,, \quad \mathbf{U}_1 \in \mathbb{R}^{n\times r}\,, \quad r = \mathrm{r}(\mathbf{A})\,, \tag{5.93a}$$

$$\mathbf{V}_{m\times m} = (\mathbf{V}_1 : \mathbf{V}_0)\,, \quad \mathbf{V}_1 \in \mathbb{R}^{m\times r}\,, \quad \mathbf{\Delta} \in \mathbb{R}^{n\times m}. \tag{5.93b}$$

Matriisi

$$\mathbf{\Delta}_1 = \mathrm{diag}(\delta_1, \delta_2, \ldots, \delta_r) \tag{5.94}$$

on $\mathbf{A}$:n (nollasta poikeavien) singulaariarvojen (jotka osoittautuvat $\mathbf{A}'\mathbf{A}$:n nollasta poikkeavien ominaisarvojen neliöjuuriksi) muodostama lävistäjämatriisi:

$$\delta_i = \mathrm{sg}_i(\mathbf{A}) = +\sqrt{\mathrm{ch}_i(\mathbf{A}'\mathbf{A})} > 0\,, \quad i = 1, 2, \ldots, r\,. \tag{5.95}$$

Hajotelma (SAH) voidaan esittää myös muodossa

$$\mathbf{A} = \mathbf{U}\mathbf{\Delta}\mathbf{V}' = (\mathbf{U}_1 : \mathbf{U}_0)\begin{pmatrix} \mathbf{\Delta}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{V}_1' \\ \mathbf{V}_0' \end{pmatrix} = \mathbf{U}_1\mathbf{\Delta}_1\mathbf{V}_1'\,, \tag{5.96}$$

ja edelleen

$$\begin{aligned}
\mathbf{A} &= (\mathbf{U}_1 : \mathbf{U}_0)\begin{pmatrix} \mathbf{\Delta}_1 & \mathbf{0}_{r\times(m-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(m-r)} \end{pmatrix}\begin{pmatrix} \mathbf{V}_1' \\ \mathbf{V}'0 \end{pmatrix} \\
&= \mathbf{U}_*\mathbf{\Delta}_*\mathbf{V}' \\
&= \delta_1\mathbf{u}_1\mathbf{v}_1' + \cdots + \delta_r\mathbf{u}_r\mathbf{v}_r' \\
&= \mathbf{U}\begin{pmatrix} \mathbf{\Delta}_{*(m\times m)} \\ \mathbf{0}_{(n-m)\times m} \end{pmatrix}\mathbf{V}',
\end{aligned} \tag{5.97}$$

missä

$$
\begin{aligned}
\mathbf{\Delta}_1 &= \operatorname{diag}(\delta_1, \dots, \delta_r)\,, \quad \delta_1 \geq \cdots \geq \delta_r > 0\,, \quad \mathbf{\Delta} \in \mathbb{R}^{n\times m}, \\
\mathbf{\Delta} &= \begin{pmatrix} \mathbf{\Delta}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{\Delta}_* \\ \mathbf{0} \end{pmatrix} \in \mathbb{R}^{n\times m}, \quad \mathbf{\Delta}_1 \in \mathbb{R}^{r\times r}, \quad \mathbf{\Delta}_* \in \mathbb{R}^{m\times m}, \\
\delta_{r+1} &= \delta_{r+2} = \cdots = \delta_m = 0\,, \\
\delta_i &= \operatorname{sg}_i(\mathbf{A}) = \sqrt{\operatorname{ch}_i(\mathbf{A}'\mathbf{A})} \\
&= \mathbf{A}\text{:n } i.\text{ singulaariarvo}\,, \quad i = 1, \dots, m\,, \\
\mathbf{\Delta}_* &= \operatorname{diag}(\delta_1, \dots, \delta_r, \delta_{r+1}, \dots, \delta_m) = \mathbf{\Delta}\text{:n } m \text{ ensimmäistä vaakariviä}\,, \\
\mathbf{U}_{n\times n} &= (\mathbf{U}_1 : \mathbf{U}_0)\,, \quad \mathbf{U}_1 \in \mathbb{R}^{n\times r}, \quad \mathbf{U}'\mathbf{U} = \mathbf{U}\mathbf{U}' = \mathbf{I}_n\,, \\
\mathbf{V}_{m\times m} &= (\mathbf{V}_1 : \mathbf{V}_0)\,, \quad \mathbf{V}_1 \in \mathbb{R}^{m\times r}, \quad \mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}_m\,, \\
\mathbf{U}_* &= (\mathbf{u}_1 : \ldots : \mathbf{u}_m) = \mathbf{U}\text{:n } m \text{ ensimmäistä saraketta}\,, \quad \mathbf{U}_* \in \mathbb{R}^{n\times m}, \\
\mathbf{V}'\mathbf{A}'\mathbf{A}\mathbf{V} &= \mathbf{\Delta}'\mathbf{\Delta} = \mathbf{\Delta}_*^2 = \begin{pmatrix} \mathbf{\Delta}_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times m}, \\
\mathbf{U}'\mathbf{A}\mathbf{A}'\mathbf{U} &= \mathbf{\Delta}\mathbf{\Delta}' = \mathbf{\Delta}_\#^2 = \begin{pmatrix} \mathbf{\Delta}_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{n\times n}, \\
\mathbf{u}_i &= \mathbf{A}\text{:n } i.\text{ vasemmanpuoleinen singulaarivektori} \\
&= \mathbf{A}\mathbf{A}'\text{:n } i.\text{ ominaisvektori}, \\
\mathbf{v}_i &= \mathbf{A}\text{:n } i.\text{ oikeanpuoleinen singulaarivektori} \\
&= \mathbf{A}'\mathbf{A}\text{:n } i.\text{ ominaisvektori}\,.
\end{aligned}
$$

Koska $\mathbf{\Delta}_*$ muodostuu $\mathbf{\Delta}$:n m:stä ensimmäisestä vaakarivistä, eli

$$\mathbf{\Delta}_{n\times m} = \begin{pmatrix} \mathbf{\Delta}_{*(m\times m)} \\ \mathbf{0}_{(n-m)\times m} \end{pmatrix}, \tag{5.98}$$

niin voimme esittää singulaariarvohajotelman myös muodossa

$$\mathbf{A} = \mathbf{U}_*\mathbf{\Delta}_*\mathbf{V}', \tag{5.99}$$

missä $\mathbf{U}_* = (\mathbf{u}_1 : \ldots : \mathbf{u}_m)$ eli se muodostuu $\mathbf{U}$:n m:stä ensimmäisestä sarakkeesta.

Merkintä $\operatorname{sg}(\mathbf{A})$ tarkoittaa matriisin $\mathbf{A}_{n\times m}$ ($m \leq n$) kaikkien singulaariarvojen joukkoa:

$$\operatorname{sg}(\mathbf{A}) = \{\operatorname{sg}_1(\mathbf{A}), \dots, \operatorname{sg}_r(\mathbf{A}), \dots, \operatorname{sg}_m(\mathbf{A})\} \tag{5.100}$$

ja

$$\operatorname{nzsg}(\mathbf{A}) = \{\operatorname{sg}_1(\mathbf{A}), \dots, \operatorname{sg}_r(\mathbf{A})\} \tag{5.101}$$

tarkoittaa nollasta poikeavien singulaariarvojen joukkoa; $r = \operatorname{rank}(\mathbf{A})$.

Koska

$$\mathbf{A} = (\mathbf{U}_1 : \mathbf{U}_0) \begin{pmatrix} \mathbf{\Delta}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{V}_1' \\ \mathbf{V}_0' \end{pmatrix} = \mathbf{U}_1 \mathbf{\Delta}_1 \mathbf{V}_1' \,, \tag{5.102}$$

niin myös

$$\mathbf{A}' = \mathbf{V}_1 \mathbf{\Delta}_1 \mathbf{U}_1' \,. \tag{5.103}$$

Yhtälöistä (5.102) ja (5.103) seuraa välittömästi että

$$\mathscr{C}(\mathbf{A}) = \mathscr{C}(\mathbf{U}_1) \text{ eli } \mathbf{U}_1 \text{ muodostaa } \mathscr{C}(\mathbf{A})\text{:n kannan,} \tag{5.104a}$$

$$\mathscr{C}(\mathbf{A}') = \mathscr{C}(\mathbf{V}_1) \text{ eli } \mathbf{V}_1 \text{ muodostaa } \mathscr{C}(\mathbf{A}')\text{:n kannan.} \tag{5.104b}$$

Täten (SAH):n avulla saadaan oheistuotteena annetun matriisin ortonormaali kanta.

Jos $\mathbf{A}_{n\times n}$ on symmetrinen, niin $\mathbf{A}$:lla on ominaisarvohajotelma

$$\mathbf{A} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}' = (\mathbf{T}_1 : \mathbf{T}_0) \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{T}_1' \\ \mathbf{T}_0' \end{pmatrix} = \mathbf{T}_1 \mathbf{\Lambda}_1 \mathbf{T}_1' \,, \tag{OAH}$$

missä $\mathbf{\Lambda}_1 = \operatorname{diag}(\lambda_1, \dots, \lambda_r)$, $\operatorname{ch}_i(\mathbf{A}) = \lambda_i \neq 0$, $i = 1, \dots, r$, $r = \operatorname{r}(\mathbf{A})$ ja $\mathbf{T}_1$ muodostuu näitä nollasta eroavia ominaisarvoja vastaavista ortonormaaleista ominaisvektoreista. Jos kaikki ominaisarvot λ_i ovat vähintään nollia eli $\mathbf{A}$ on ei-negatiivisesti definiitti, on $\mathbf{A}$:n ominaisarvohajotelma sama kuin sen singulaariarvohajotelma. Mikäli $\mathbf{A}$:lla on negatiivisia ominaisarvoja, saadaan OAH:sta SAH, kun negatiiviset λ_i:t korvataan itseisarvoillaan ja vastaavat sarakkeet $\mathbf{T}_1$:ssä OAH:n oikealla puolella kerrotaan -1:llä.

Seuraavaksi esitämme SAH:n todistuksen samaan tapaan kuin Searle (1982, p. 316); muina lähteinä mainittakoon Stewart (1998, p. 62), Golub & Van Loan (1996, §2.5.3), ja Horn & Olkin (1996, Th. 2.1). SAH:n historiaa ovat käsitelleet mm. Horn & Johnson (1990, §3.0) ja Stewart (1993); ks. myös Horn & Olkin (1996, §2).

SVD:n todistus. Tarkastellaan symmetrisen ei-negatiivisesti definiitin $m \times m$-matriisin $\mathbf{A}'\mathbf{A}$ ominaisarvohajotelmaa

$$\mathbf{A}'\mathbf{A} = \mathbf{V}\mathbf{\Delta}_*^2\mathbf{V}' = (\mathbf{V}_1 : \mathbf{V}_0)\mathbf{\Delta}_*^2(\mathbf{V}_1 : \mathbf{V}_0)', \tag{5.105}$$

missä $\mathbf{\Delta}_*^2$ on $m \times m$-diagonaalimatriisi, jonka diagonaalielementit ovat $\mathbf{A}'\mathbf{A}$:n ominaisarvot:

$$\mathbf{\Delta}_*^2 = \begin{pmatrix} \mathbf{\Delta}_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \mathbf{\Delta}_1^2 \in \mathbb{R}^{r\times r}, \quad r = \operatorname{rank}(\mathbf{A})\,, \tag{5.106a}$$

$$\mathbf{\Delta}_1^2 = \operatorname{diag}(\delta_1^2, \dots, \delta_r^2)\,, \quad \delta_1^2 \geq \dots \geq \delta_r^2 > 0 = \delta_{r+1}^2 = \dots = \delta_m^2\,, \tag{5.106b}$$

$$\delta_i^2 = \operatorname{ch}_i(\mathbf{A}'\mathbf{A})\,, \quad i = 1, \dots, m\,, \tag{5.106c}$$

$$\mathbf{V} = (\mathbf{V}_1 : \mathbf{V}_0)\,, \quad \mathbf{V}_1 = (\mathbf{v}_1 : \dots : \mathbf{v}_r) \in \mathbb{R}^{m\times r}. \tag{5.106d}$$

Huomattakoon, että matriiseilla $\mathbf{A}'\mathbf{A}$ ja $\mathbf{A}\mathbf{A}'$ on samat nollasta poikkeavat ominaisarvot:

$$\{\delta_1^2, \dots, \delta_r^2\} = \mathrm{nzch}(\mathbf{A}'\mathbf{A}) = \mathrm{nzch}(\mathbf{A}\mathbf{A}') \,. \tag{5.107}$$

Matriisin $\mathbf{V}_1 \in \mathbb{R}^{m\times r}$ sarakkeet ovat $\mathbf{A}'\mathbf{A}$:n nollasta poikkeavia ominaisarvoja δ_i^2 vastaavat ortonormaalit ominaisvektorit, ja matriisin $\mathbf{V}_0 \in \mathbb{R}^{m\times(m-r)}$ sarakkeet ovat $\mathbf{A}'\mathbf{A}$:n ortonormaalit ominaisvektorit vastaten ominaisarvoa 0. Lisäksi on voimassa $\mathbf{V}\mathbf{V}' = \mathbf{V}'\mathbf{V} = \mathbf{I}_m$. Nyt (5.105) voidaan esittää muodossa

$$\mathbf{A}'\mathbf{A} = \mathbf{V}\mathbf{\Delta}_*^2\mathbf{V}' = (\mathbf{V}_1 : \mathbf{V}_0)\begin{pmatrix}\mathbf{\Delta}_1^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{V}_1'\\ \mathbf{V}_0'\end{pmatrix} = \mathbf{V}_1\mathbf{\Delta}_1^2\mathbf{V}_1' \,, \tag{5.108}$$

joten

$$\mathbf{\Delta}_1^2 = \mathbf{V}_1'\mathbf{A}'\mathbf{A}\mathbf{V}_1 \,. \tag{5.109}$$

Määritellään $n \times r$-matriisi $\mathbf{U}_1$ seuraavasti:

$$\mathbf{U}_1 = \mathbf{A}\mathbf{V}_1\mathbf{\Delta}_1^{-1} \in \mathbb{R}^{n\times r}. \tag{5.110}$$

Silloin $\mathbf{U}_1$:n sarakkeet ovat ortonormaalit, koska (5.109):n nojalla on voimassa

$$\mathbf{U}_1'\mathbf{U}_1 = \mathbf{\Delta}_1^{-1}\mathbf{V}_1'\mathbf{A}'\mathbf{A}\mathbf{V}_1\mathbf{\Delta}_1^{-1} = \mathbf{\Delta}_1^{-1}\mathbf{\Delta}_1^2\mathbf{\Delta}_1^{-1} = \mathbf{I}_r \,. \tag{5.111}$$

On helppo päätellä, että on mahdollista muodostaa $n \times (n-r)$-matriisi $\mathbf{U}_0$ siten että

$$\mathbf{U} = (\mathbf{U}_1 : \mathbf{U}_0) \in \mathbb{R}^{n\times n} \tag{5.112}$$

on ortogonaalinen. Kun (5.110) kerrotaan oikealta matriisilla $\mathbf{\Delta}_1\mathbf{V}_1'$, saadaan

$$\mathbf{U}_1\mathbf{\Delta}_1\mathbf{V}_1' = \mathbf{A}\mathbf{V}_1\mathbf{V}_1' = \mathbf{A} \,, \tag{5.113}$$

missä on käytetty yhtälöä $\mathbf{V}_1\mathbf{V}_1' = \mathbf{P}_{\mathbf{A}'}$. Näin olemme lopulta päätyneet singulaariarvohajotelmaan

$$\mathbf{A} = (\mathbf{U}_1 : \mathbf{U}_0)\begin{pmatrix}\mathbf{\Delta}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{V}_1'\\ \mathbf{V}_0'\end{pmatrix} = \mathbf{U}\mathbf{\Delta}\mathbf{V}'. \tag{5.114}$$

□

5.3.1 Lausekkeen $\mathbf{x}'\mathbf{A}\mathbf{y}$ maksimiarvo

Olkoon $\mathbf{A}$ $n \times m$-matriisi, jonka aste on $r > 0$, ja olkoon $\mathbf{A}$:lla singulaariarvohajotelma

$$\mathbf{A} = \mathbf{U}\mathbf{\Delta}\mathbf{V}' = \mathbf{U}_1\mathbf{\Delta}_1\mathbf{V}_1' = \delta_1\mathbf{u}_1\mathbf{v}_1' + \dots + \delta_r\mathbf{u}_r\mathbf{v}_r' \,. \tag{5.115}$$

Osoitamme että

$$\max_{\mathbf{x}'\mathbf{x}=\mathbf{y}'\mathbf{y}=1} \mathbf{x}'\mathbf{A}\mathbf{y} = \delta_1 \,, \tag{5.116}$$

ja että maksimi saavutetaan kun

$$\mathbf{x} = \mathbf{u}_1 \,, \qquad \mathbf{y} = \mathbf{v}_1 \,. \tag{5.117}$$

Tarkastellaan vektoreita $\mathbf{x} \in \mathbb{R}^n$ ja $\mathbf{y} \in \mathbb{R}^n$ jotka toteuttavat ehdon $\mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{y} = 1$. Cauchyn–Schwarzin epäyhtälön nojalla (ks. luku 6.7.1, s. 208) saamme

$$\begin{aligned}
(\mathbf{x}'\mathbf{A}\mathbf{y})^2 &= (\mathbf{x}'\mathbf{U}\boldsymbol{\Delta}\mathbf{V}'\mathbf{y})^2 = (\mathbf{x}'\mathbf{U} \cdot \boldsymbol{\Delta}\mathbf{V}'\mathbf{y})^2 \\
&\le \mathbf{x}'\mathbf{U}\mathbf{U}'\mathbf{x} \cdot \mathbf{y}'\mathbf{V}\boldsymbol{\Delta}'\boldsymbol{\Delta}\mathbf{V}'\mathbf{y} = \mathbf{y}'\mathbf{V}\boldsymbol{\Delta}_*^2\mathbf{V}'\mathbf{y} := \mathbf{t}'\boldsymbol{\Delta}_*^2\mathbf{t} \\
&= \delta_1^2 t_1^2 + \delta_2^2 t_2^2 + \dots + \delta_r^2 t_r^2 \\
&\le \delta_1^2 (t_1^2 + t_2^2 + \dots + t_r^2) \\
&\le \delta_1^2 (t_1^2 + t_2^2 + \dots + t_r^2 + \dots + t_m^2) \\
&= \delta_1^2 \mathbf{y}'\mathbf{V}\mathbf{V}'\mathbf{y} \\
&= \delta_1^2 \,,
\end{aligned} \tag{5.118}$$

missä on merkitty $\mathbf{t} = \mathbf{V}'\mathbf{y} \in \mathbb{R}^m$. On vaivaton päätellä, että maksimi saavutetaan kun $\mathbf{x}$ ja $\mathbf{y}$ valitaan kuten (5.117):ssa.

Huomattakoon, että tietysti on voimassa

$$\operatorname{sg}_1(\mathbf{A}) = \delta_1 = \max_{\mathbf{x}\neq\mathbf{0},\,\mathbf{y}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{A}\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x} \cdot \mathbf{y}'\mathbf{y}}} \,. \tag{5.119}$$

Toiseksi suurimmalla singulaariarvolla δ_2 on seuraava ominaisuus:

$$\operatorname{sg}_2(\mathbf{A}) = \delta_2 = \max \mathbf{x}'\mathbf{A}\mathbf{y} \,, \tag{5.120}$$

missä maksimointi tapahtuu joukossa

$$\{\, \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m : \mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{y} = 1,\ \mathbf{x}'\mathbf{u}_1 = \mathbf{y}'\mathbf{v}_1 = 0 \,\} \,. \tag{5.121}$$

Matriisin $\mathbf{A}$ i:nneksi suurin singulaariarvo δ_i voidaan määritellä seuraavasti:

$$\max_{\substack{\mathbf{x}\neq\mathbf{0},\,\mathbf{y}\neq\mathbf{0} \\ \mathbf{U}_{(k)}\mathbf{x}=\mathbf{0},\mathbf{V}_{(k)}\mathbf{y}=\mathbf{0}}} \frac{\mathbf{x}'\mathbf{A}\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x} \cdot \mathbf{y}'\mathbf{y}}} = \delta_{k+1} \,, \quad k = 1, \dots, r-1 \,, \tag{5.122}$$

missä $\mathbf{U}_{(k)} = (\mathbf{u}_1 : \ldots : \mathbf{u}_k)$ ja $\mathbf{V}_{(k)} = (\mathbf{v}_1 : \ldots : \mathbf{v}_k)$; maksimi saavutetaan kun $\mathbf{x} = \mathbf{u}_{k+1}$ ja $\mathbf{y} = \mathbf{v}_{k+1}$.

Harjoitustehtäviä

5.1. Olkoon $\mathbf{X} \in \mathbb{R}^{n\times p}$, $\mathbf{V} \in \mathrm{NND}_n$, $\mathbf{H} = \mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ ja $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$. Merkitään

$$\tau = \frac{1}{2}\|\mathbf{H}\mathbf{V} - \mathbf{V}\mathbf{H}\|_F^2 \,.$$

Osoita, että $\tau = \operatorname{tr}(\mathbf{H}\mathbf{V}^2) - \operatorname{tr}(\mathbf{H}\mathbf{V})^2 = \|\mathbf{H}\mathbf{V}\mathbf{M}\|_F^2$.

5.2. Olkoon $\mathbf{V}_{n\times n}$ positiivisesti definiitti matriisi, jolla on ominaisarvohajotelma $\mathbf{V} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}'$, missä lävistäjämatriisi $\mathbf{\Lambda}$:n elementit ovat $\mathbf{V}$:n ominaisarvot $\lambda_1 \geq \ldots \geq \lambda_n > 0$. Olkoon

$$\mathbf{X} = (\mathbf{t}_1 + \mathbf{t}_n : \mathbf{t}_2 + \mathbf{t}_{n-1}) \cdot \frac{1}{\sqrt{2}}\,.$$

Osoita, että

(a) $\mathbf{X}'\mathbf{X} = \mathbf{I}_2\,,$

(b) $\mathbf{X}'\mathbf{V}\mathbf{X} = \dfrac{1}{2}\begin{pmatrix} \lambda_1 + \lambda_n & 0 \\ 0 & \lambda_2 + \lambda_{n-1} \end{pmatrix}$

[Huom: $\det(\mathbf{X}'\mathbf{V}\mathbf{X}) = \frac{1}{2}(\lambda_1 + \lambda_n)(\lambda_2 + \lambda_{n-1})$],

(c) $\mathbf{X}'\mathbf{V}^2\mathbf{X} = \dfrac{1}{2}\begin{pmatrix} \lambda_1^2 + \lambda_n^2 & 0 \\ 0 & \lambda_2^2 + \lambda_{n-1}^2 \end{pmatrix},$

(d) $\tau = \dfrac{1}{4}(\lambda_1 - \lambda_n)^2(\lambda_2 - \lambda_{n-1})^2,$

(e) $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \frac{1}{2}\begin{pmatrix} \frac{\lambda_1+\lambda_n}{\lambda_1\lambda_n} & 0 \\ 0 & \frac{\lambda_2+\lambda_{n-1}}{\lambda_2\lambda_{n-1}} \end{pmatrix}.$

Ohje: Käytä hyväksi matriisin $\mathbf{X}$ esitystä

$$\begin{aligned} \mathbf{X} &= (\mathbf{t}_1 + \mathbf{t}_n : \mathbf{t}_2 + \mathbf{t}_{n-1}) \cdot \frac{1}{\sqrt{2}} = \mathbf{T}(\mathbf{i}_1 + \mathbf{i}_n : \mathbf{i}_2 + \mathbf{i}_{n-1}) \cdot \frac{1}{\sqrt{2}} \\ &= \mathbf{T}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ 1 & 0 \end{pmatrix}\frac{1}{\sqrt{2}} := \mathbf{T}\mathbf{U}. \end{aligned}$$

Luku 6

Joitakin erityisiä matriiseja

Se on semmosta meitin tämmösten miesten
... Minun niin kun sinunkin ...

Väinö Linna/Reeti Leppänen:
Täällä pohjantähden alla.

6.1 Yksikkövektori $\mathbf{i}_j$

Jo johdantoluvussa ositimme yksikkömatriisin $\mathbf{I}_n$ seuraavasti:

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = (\mathbf{i}_1 : \mathbf{i}_2 : \dots : \mathbf{i}_n)\,, \tag{6.1}$$

missä $\mathbf{i}_j$ on $\mathbf{I}_n$:n j. sarake. Vektorin $\mathbf{i}_j$ avulla saadaan kätevästi poimittua tiettyjä sarakkeita, vaakarivejä ja yksittäisiä elementtejä matriisista $\mathbf{A}$. Esimerkiksi matriisin $\mathbf{A}_{n\times n}$ i. sarake ja j. vaakarivi voidaan esittää siten että $\mathbf{A}\mathbf{i}_i = \mathbf{a}_i$, $\mathbf{i}_j'\mathbf{A} = \mathbf{a}_{(j)}'$. Matriisin $\mathbf{A}$ elementti a_{ij} on tällöin $a_{ij} = \mathbf{i}_i'\mathbf{A}\mathbf{i}_j$.

Esimerkki 6.1. Olkoon $\mathbf{M}_{n\times n} = \{m_{ij}\}$ jokin ortogonaaliprojektori eli idempotentti symmetrinen ($\mathbf{M}^2 = \mathbf{M} = \mathbf{M}'$) matriisi. Osoitetaan, että vektorin $\mathbf{y}$ ortogonaaliprojektion $\mathbf{e} = \mathbf{M}\mathbf{y}$ i. elementin neliö voidaan esittää muodossa

$$e_i^2 = m_{ii}\,\mathbf{y}'\mathbf{P}_{\mathbf{M}\mathbf{i}_i}\mathbf{y}\,, \tag{6.2}$$

missä $\mathbf{P}_{\mathbf{M}\mathbf{i}_i}$ on ortogonaaliprojektori suoralle $\mathscr{C}(\mathbf{M}\mathbf{i}_i)$. Vektori $\mathbf{e}$ voidaan tietyssä tilanteessa tulkita regressioanalyysin *residuaalivektoriksi*. On omat matemaattiset etunsa havaita, että e_i^2 voidaan esittää muodossa $e_i^2 = \text{vakio}\cdot\mathbf{y}'\mathbf{Q}\mathbf{y}$, missä $\mathbf{Q}$ on sopiva ortogonaaliprojektori.

Olkoon merkintöjen yksinkertaistamiseksi $\mathbf{u} = \mathbf{i}_i$. Tällöin $e_i = \mathbf{u}'\mathbf{M}\mathbf{y}$, joten

$$\begin{aligned} e_i^2 &= (\mathbf{u}'\mathbf{M}\mathbf{y})(\mathbf{u}'\mathbf{M}\mathbf{y}) = (\mathbf{y}'\mathbf{M}\mathbf{u})(\mathbf{u}'\mathbf{M}\mathbf{y}) = \mathbf{y}'(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M})\mathbf{y} \\ &= \mathbf{u}'\mathbf{M}\mathbf{u} \cdot \mathbf{y}' \frac{\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}}{\mathbf{u}'\mathbf{M}\mathbf{u}} \mathbf{y} \\ &= m_{ii}\, \mathbf{y}'\mathbf{P}_{\mathbf{M}\mathbf{u}}\mathbf{y}\,, \end{aligned} \tag{6.3}$$

missä

$$\mathbf{P}_{\mathbf{M}\mathbf{u}} = \mathbf{M}\mathbf{u}(\mathbf{u}'\mathbf{M}' \cdot \mathbf{M}\mathbf{u})^{-1}\mathbf{u}'\mathbf{M}' = \mathbf{M}\mathbf{u}(\mathbf{u}'\mathbf{M}\mathbf{u})^{-1}\mathbf{u}'\mathbf{M}\,. \tag{6.4}$$

Voimme soveltaa lauseketta (6.2) laskeaksemme keskistetyn $\mathbf{y}$:n eli $\tilde{\mathbf{y}}$:n i. elementin neliön $(y_i - \bar{y})^2 = \tilde{y}_i^2$. Tällöin $\mathbf{M} = \mathbf{C} = \mathbf{I} - \mathbf{J}$, joten $\tilde{y}_i = \mathbf{u}'\mathbf{C}\mathbf{y}$, ja

$$\tilde{y}_i^2 = c_{ii}\mathbf{y}'\mathbf{P}_{\mathbf{C}\mathbf{u}}\mathbf{y} = (1 - \tfrac{1}{n})\mathbf{y}'\mathbf{P}_{\mathbf{C}\mathbf{u}}\mathbf{y}\,. \tag{6.5}$$

Yhtälöstä (6.5) seuraa, että

$$\frac{(y_i - \bar{y})^2}{1 - \frac{1}{n}} = \frac{(y_i - \bar{y})^2}{\frac{n-1}{n}} = \mathbf{y}'\mathbf{P}_{\mathbf{C}\mathbf{u}}\mathbf{y}\,. \tag{6.6}$$

□

Esimerkki 6.2. Olkoon havaintomatriisi $\mathbf{U}_{n\times p}$ ositettu seuraavasti:

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{u}'_{(n)} \end{pmatrix}. \tag{6.7}$$

Tällöin tietenkin

$$\mathbf{U}'\mathbf{U} = \mathbf{u}_{(1)}\mathbf{u}'_{(1)} + \mathbf{u}_{(2)}\mathbf{u}'_{(2)} + \cdots + \mathbf{u}_{(n)}\mathbf{u}'_{(n)}\,. \tag{6.8}$$

Olkoon tehtävänä määrittää matriisi $\mathbf{W}$, jolla on ominaisuus

$$\mathbf{U}'\mathbf{W}\mathbf{U} = \mathbf{u}_{(1)}\mathbf{u}'_{(1)} + \mathbf{u}_{(2)}\mathbf{u}'_{(2)} + \cdots + \mathbf{u}_{(n-1)}\mathbf{u}'_{(n-1)}\,, \tag{6.9}$$

eli $\mathbf{W}$:n avulla saadaan viimeinen havaintovektori pudotetuksi pois $\mathbf{U}'\mathbf{U}$:sta. Merkitään

$$\mathbf{P} = \mathbf{i}_n\mathbf{i}'_n = \begin{pmatrix} \mathbf{0}_{(n-1)\times(n-1)} & \mathbf{0} \\ \mathbf{0}' & 1 \end{pmatrix} = \mathbf{P}' = \mathbf{P}^2, \tag{6.10a}$$

$$\mathbf{Q} = \mathbf{I}_n - \mathbf{P} = \begin{pmatrix} \mathbf{I}_{n-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix} = \mathbf{Q}' = \mathbf{Q}^2. \tag{6.10b}$$

Tällöin

$$\begin{aligned} \mathbf{P}\mathbf{U} &= \begin{pmatrix} \mathbf{0}_{(n-1)\times(n-1)} & \mathbf{0} \\ \mathbf{0}' & 1 \end{pmatrix} \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{u}'_{(n)} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{(n-1)\times n} \\ \mathbf{u}'_{(n)} \end{pmatrix}, \\ \mathbf{Q}\mathbf{U} &= \begin{pmatrix} \mathbf{I}_{n-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix} \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{u}'_{(n)} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{0}' \end{pmatrix}, \end{aligned} \tag{6.11}$$

joten $\mathbf{W}$ on $\mathbf{Q}$:

$$\begin{aligned}(\mathbf{QU})'\mathbf{QU} &= \mathbf{U}'\mathbf{Q}'\mathbf{QU} = \mathbf{U}'\mathbf{QU} \\ &= \mathbf{U}'(\mathbf{I}_n - \mathbf{i}_n\mathbf{i}_n')\mathbf{U} = \begin{pmatrix}\mathbf{U}_1 \\ \mathbf{0}'\end{pmatrix}' \begin{pmatrix}\mathbf{U}_1 \\ \mathbf{0}'\end{pmatrix} = \mathbf{U}_1'\mathbf{U}_1 \\ &= \mathbf{u}_{(1)}\mathbf{u}_{(1)}' + \mathbf{u}_{(2)}\mathbf{u}_{(2)}' + \cdots + \mathbf{u}_{(n-1)}\mathbf{u}_{(n-1)}' \,. \end{aligned} \tag{6.12}$$

□

Esimerkki 6.3. Tarkastellaan $n \times m$-matriiseja $\mathbf{A}$ ja $\mathbf{B}$. Olkoon tehtävänä määrittää matriisit $\mathbf{F}$ ja $\mathbf{G}$, joilla on ominaisuus:

$$(\mathbf{A} : \mathbf{B})\mathbf{F} = \mathbf{A} + \mathbf{B} = \mathbf{G}\begin{pmatrix}\mathbf{A} \\ \mathbf{B}\end{pmatrix}. \tag{6.13}$$

On helppo nähdä, että vastaus on seuraava:

$$(\mathbf{A} : \mathbf{B})\begin{pmatrix}\mathbf{I}_m \\ \mathbf{I}_m\end{pmatrix} = \mathbf{A} + \mathbf{B} = (\mathbf{I}_n : \mathbf{I}_n)\begin{pmatrix}\mathbf{A} \\ \mathbf{B}\end{pmatrix}. \tag{6.14}$$

Havaitsemme myös, että

$$\begin{pmatrix}\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}\end{pmatrix}\begin{pmatrix}\mathbf{I}_m \\ \mathbf{I}_m\end{pmatrix} = \begin{pmatrix}\mathbf{A} \\ \mathbf{B}\end{pmatrix}, \tag{6.15a}$$

$$(\mathbf{I}_n : \mathbf{I}_n)\begin{pmatrix}\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}\end{pmatrix} = (\mathbf{A} : \mathbf{B})\,. \tag{6.15b}$$

□

6.2 Ortogonaalinen matriisi

Kuten tiedämme, kertolasku $\mathbf{A}'\mathbf{A}$ voidaan tehdä seuraavasti, kun $\mathbf{A}_{n\times m}$ on ositettu sarakkeittain, $\mathbf{A} = (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m)$:

$$\mathbf{A}'\mathbf{A} = \begin{pmatrix}\mathbf{a}_1'\mathbf{a}_1 & \mathbf{a}_1'\mathbf{a}_2 & \ldots & \mathbf{a}_1'\mathbf{a}_m \\ \mathbf{a}_2'\mathbf{a}_1 & \mathbf{a}_2'\mathbf{a}_2 & \ldots & \mathbf{a}_2'\mathbf{a}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_m'\mathbf{a}_1 & \mathbf{a}_m'\mathbf{a}_2 & \ldots & \mathbf{a}_m'\mathbf{a}_m\end{pmatrix} = \{\mathbf{a}_i'\mathbf{a}_j\} = \{\langle\mathbf{a}_i, \mathbf{a}_j\rangle\}\,. \tag{6.16}$$

Siten matriisin $\mathbf{A}'\mathbf{A}$ elementit ovat $\mathbf{A}$:n sarakkeiden keskinäisiä sisätuloja. Jos $\mathbf{A}'\mathbf{A} = \mathbf{I}_m$, niin silloinhan

$$\mathbf{a}_i'\mathbf{a}_j = \langle\mathbf{a}_i, \mathbf{a}_j\rangle = \begin{cases}0, & \text{kun } i \neq j, \\ 1, & \text{kun } i = j.\end{cases} \tag{6.17}$$

Matriisin $\mathbf{A}$ sarakkeet ovat tällöin ortogonaalisia toisiaan kohtaan ja lisäksi 1:n pituisia eli $\mathbf{A}$:n sarakkeet ovat *ortonormaaleja*. Jos $\mathbf{A}$ on neliömatriisi, jolla on ominaisuus $\mathbf{A}'_{n\times n}\mathbf{A}_{n\times n} = \mathbf{I}_n$, niin $\mathbf{A}$:ta sanotaan *ortogonaaliseksi*. Tällöin myös $\mathbf{A}\mathbf{A}' = \mathbf{I}_n$ ja $\mathbf{A}'$ on siten $\mathbf{A}$:n käänteismatriisi:

$$\mathbf{A}' = \mathbf{A}^{-1}. \tag{6.18}$$

Esimerkiksi seuraavat matriisit ovat ortogonaalisia:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1/a & 1/b & 1/c \\ 1/a & -1/b & 1/c \\ 1/a & 0 & -2/c \end{pmatrix}, \tag{6.19}$$

missä $a = 3$, $b = 2$, $c = 6$. Sen sijaan matriisi

$$\mathbf{X} = (\mathbf{x}_1 : \mathbf{x}_2 : \mathbf{x}_3) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{1}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_2 \end{pmatrix} \tag{6.20}$$

ei ole ortogonaalinen – sen sarakkeet kyllä ovat keskenään ortogonaalisia, mutta eivät ykkösen pituisia. Saamme tietysti skaalatuksi $\mathbf{X}$:n sarakkeet pituudeltaan ykkösiksi, mutta tämäkään matriisi ei ole ortogonaalinen, koska se ei ole neliömatriisi.

Ortogonaalisilla matriiseilla on monia mielenkiintoisia ominaisuuksia. Esimerkiksi

♠ vektorien pituudet ja niiden väliset kulmat eivät muutu ortogonaalisessa muunnoksessa,

ts. jos $\mathbf{Q}$ on ortogonaalinen, niin

$$\|\mathbf{Qx}\|^2 = \mathbf{x}'\mathbf{Q}'\mathbf{Qx} = \mathbf{x}'\mathbf{Ix} = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2, \tag{6.21a}$$

$$\langle \mathbf{Qx}, \mathbf{Qy} \rangle = \mathbf{x}'\mathbf{Q}'\mathbf{Qy} = \mathbf{x}'\mathbf{Iy} = \mathbf{x}'\mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle\,. \tag{6.21b}$$

Itse asiassa on todistettavissa seuraava tärkeä tulos:

$$\boxed{\mathbf{UU}' = \mathbf{VV}' \iff \textit{on olemassa ortogonaalinen } \mathbf{Q}\textit{:}\ \mathbf{U} = \mathbf{VQ}\,. \qquad (6.22)}$$

Tulos (6.22) on suunnassa ”$\Leftarrow$” triviaali, mutta toinen osa todistuksesta onkin jo hankalampi; ks. Puntanen, Styan & Isotalo (2011, §19.2).

Mainittakoon myös seuraava hyödyllinen tulos:

$$\mathbf{Q} \text{ on ortogonaalinen} \implies \operatorname{tr}(\mathbf{A}) = \operatorname{tr}(\mathbf{QAQ}')\,. \tag{6.23}$$

Tulos (6.23) perustuu matriisin jäljen kommutatiivisuuteen:

$$\operatorname{tr}(\mathbf{Q} \cdot \mathbf{AQ}') = \operatorname{tr}(\mathbf{AQ}' \cdot \mathbf{Q}) = \operatorname{tr}(\mathbf{AI}_n) = \operatorname{tr}(\mathbf{A})\,. \tag{6.24}$$

Koska matriisitulon determinantti on tulon tekijöiden determinanttien tulo (kunhan rivi- ja sarakeluvut ovat sopivat), on ortogonaalisella $\mathbf{Q}$:lla ominaisuus $|\mathbf{Q}'\mathbf{Q}| = |\mathbf{I}| = 1$. Koska $|\mathbf{Q}| = |\mathbf{Q}'|$, on $|\mathbf{Q}'\mathbf{Q}| = |\mathbf{Q}|^2 = 1$, joten

$$|\mathbf{Q}| = \pm 1\,, \tag{6.25a}$$

$$|\mathbf{A}| = |\mathbf{QAQ}'|\,. \tag{6.25b}$$

Jos vektorille $\mathbf{x} \in \mathbb{R}^2$ halutaan tehdä sellainen ortogonaalinen muunnos (rotaatio) $\mathbf{t} = \mathbf{Qx}$, että uusi vektori $\mathbf{t}$ on kulman α verran myötäpäivään $\mathbf{x}$:stä, on sopiva muunnosmatriisi

$$\mathbf{Q} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} := \mathbf{Q}(\alpha)\,. \tag{6.26}$$

Tällöin $\mathbf{Q}$ on ortogonaalinen, sillä kaavan $\sin^2\alpha + \cos^2\alpha = 1$ perusteella

$$\mathbf{Q}'\mathbf{Q} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6.27}$$

Vastaavasti osoittautuu, että matriisi

$$\mathbf{P} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} := \mathbf{P}(\alpha) \tag{6.28}$$

rotatoi vektorin $\mathbf{x}$ kulman α verran vastapäivään.

Kommentti 6.1. *Peilaus.* Merkitään

$$\mathbf{G}(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}. \tag{6.29}$$

Tällöin $\mathbf{G}(\alpha)\mathbf{u}_{(i)}$ *peilaa* havainnon $\mathbf{u}_{(i)}$ suoran $y = \tan\left(\frac{\alpha}{2}\right) x$ suhteen. Kokeile tietokoneella! Itse asiassa on osoitettavissa, ks. Horn & Johnson (1990, s. 68), että mikä hyvänsä 2×2 ortogonaalinen matriisi voidaan esittää muodossa $\mathbf{Q}(\alpha)$ tai $\mathbf{G}(\alpha)$ jollakin α:n arvolla; ks harjoitustehtävä 6.1 (s. 212). □

Koska

$$\cos 30^\circ = \frac{\sqrt{3}}{2}, \quad \cos 45^\circ = \frac{1}{\sqrt{2}}, \quad \cos 60^\circ = \frac{1}{2}, \quad \cos 90^\circ = 0\,, \tag{6.30}$$

on

$$\mathbf{Q}(30^\circ) = \frac{1}{2}\begin{pmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{pmatrix}, \qquad \mathbf{Q}(45^\circ) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \tag{6.31a}$$

$$\mathbf{Q}(60^\circ) = \frac{1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix}, \qquad \mathbf{Q}(90^\circ) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{6.31b}$$

Kaavan (6.26) mukainen rotaatio voidaan tulkita myös siten että koordinaattiakseleita kierretään α astetta vastapäivään. Matriisin $\mathbf{Q}$ transpoosi aiheuttaa koordinaattiakseleiden kierron myötäpäivään.

Esimerkki 6.4. Kuviot 6.1a ja 6.1b havainnollistavat seuraavaa tilannetta, jossa $\mathbf{x}$ ja $\mathbf{y}$ ovat vektoreita, joille tehdään ortogonaalinen muunnos (45° myötäpäivään).

$$\mathbf{Q} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \|\mathbf{x}\|^2 = 10, \quad \|\mathbf{y}\|^2 = 8,$$

$$\langle \mathbf{x}, \mathbf{y} \rangle = 8, \quad \cos(\mathbf{x}, \mathbf{y}) = \frac{8}{\sqrt{80}} = \frac{2}{\sqrt{5}},$$

$$\mathbf{Qx} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 4 \\ -2 \end{pmatrix} := \mathbf{u}, \quad \|\mathbf{u}\|^2 = (16+4)/2 = 10,$$

$$\mathbf{Qy} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 4 \\ 0 \end{pmatrix} := \mathbf{v}, \quad \|\mathbf{v}\|^2 = (16+0)/2 = 8,$$

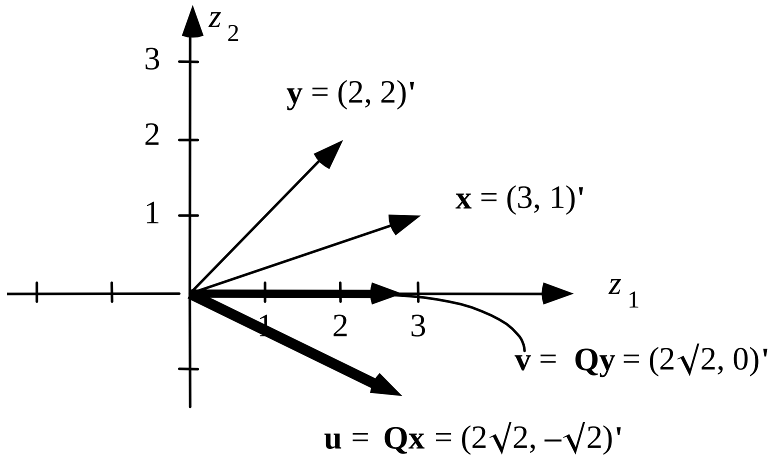
$$\langle \mathbf{Qx}, \mathbf{Qy} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{2}(16+0) = 8.$$

□

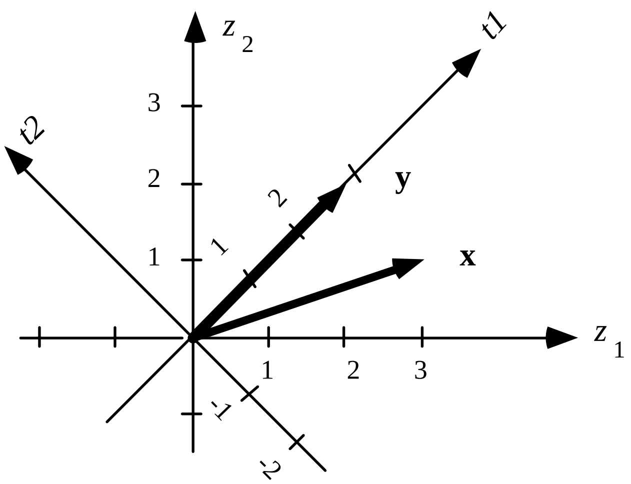
6.2.1 Havaintovektorien ortogonaalinen muunnos

Luvussa 2.2 tarkastelimme (tilastollisten) muuttujien lineaarikombinaatioita. Olkoon muuttujien x ja y havaintomatriisi

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix}. \tag{6.32}$$



Kuvio 6.1a. Esimerkin 6.4 vektoreille **x** ja **y** tehty ortogonaalinen muunnos (45° myötäpäivään).



Kuvio 6.1b. Esimerkin 6.4 tilanteessa tehty koordinaatistolle ortogonaalinen muunnos (45° vastapäivään). Uudessa koordinaatistossa **x**:n ja **y**:n koordinaatit ovat **u** ja **v**.

Tehdään nyt jokaiselle havaintovektorille $\mathbf{u}_{(i)}$ ortogonaalinen muunnos eli kerrotaan jokainen havaintovektori ortogonaalisella matriisilla $\mathbf{A}$,

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}' \\ \mathbf{b}' \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}. \tag{6.33}$$

Merkitään uusia havaintovektoreita alleviivatuin symbolein:

$$\left(\underline{\mathbf{u}}_{(1)} : \underline{\mathbf{u}}_{(2)} : \ldots : \underline{\mathbf{u}}_{(n)}\right) := \left(\mathbf{A}\mathbf{u}_{(1)} : \mathbf{A}\mathbf{u}_{(2)} : \ldots : \mathbf{A}\mathbf{u}_{(n)}\right), \tag{6.34}$$

eli $\underline{\mathbf{U}}' = \mathbf{A}\mathbf{U}'$, joten

$$\underline{\mathbf{U}} = \mathbf{U}\mathbf{A}'. \tag{6.35}$$

Sarakkeittain (eli muuttujittain) ositettuna on

$$\underline{\mathbf{U}} = (\underline{\mathbf{x}} : \underline{\mathbf{y}}) = \mathbf{U}(\mathbf{a} : \mathbf{b}) = (\mathbf{U}\mathbf{a} : \mathbf{U}\mathbf{b}). \tag{6.36}$$

Uudet muuttujavektorit $\underline{\mathbf{x}} = \mathbf{U}\mathbf{a}$ ja $\underline{\mathbf{y}} = \mathbf{U}\mathbf{b}$ ovat siis alkuperäisten muuttujien lineaarikombinaatioita. Päädymme täten luvun 2.2 (s. 85) tulokseen, jonka mukaan havaintovektoreille tehtävä lineaarinen muunnos (matriisilla $\mathbf{A}$ kertominen) tarkoittaa alkuperäisten muuttujien korvaamista niiden lineaarikombinaatioilla.

Jos havaintovektoreille tehtävä muunnos on ortogonaalinen, tiedämme että

- havaintovektoreiden pituudet eivät muutu
- havaintovektoreiden väliset kulmat eivät muutu.

Tämä perustuu (kuten jo tiedämme) seuraavaan päättelyyn:

$$\begin{aligned}
& \mathbf{A}'\mathbf{A} = \mathbf{I}_2 \; (= \mathbf{A}\mathbf{A}') \\
\implies & \underline{\mathbf{u}}'_{(i)}\underline{\mathbf{u}}_{(j)} = \mathbf{u}'_{(i)}\mathbf{A}'\mathbf{A}\mathbf{u}_{(j)} = \mathbf{u}'_{(i)}\mathbf{u}_{(j)}, \quad \underline{\mathbf{u}}'_{(i)}\underline{\mathbf{u}}_{(i)} = \mathbf{u}'_{(i)}\mathbf{A}'\mathbf{A}\mathbf{u}_{(i)} = \mathbf{u}'_{(i)}\mathbf{u}_{(i)}, \\
\implies & \cos\left(\underline{\mathbf{u}}_{(i)}, \underline{\mathbf{u}}_{(j)}\right) = \frac{\underline{\mathbf{u}}'_{(i)}\underline{\mathbf{u}}_{(j)}}{\sqrt{\underline{\mathbf{u}}'_{(i)}\underline{\mathbf{u}}_{(i)} \cdot \underline{\mathbf{u}}'_{(j)}\underline{\mathbf{u}}_{(j)}}} = \frac{\mathbf{u}'_{(i)}\mathbf{u}_{(j)}}{\sqrt{\mathbf{u}'_{(i)}\mathbf{u}_{(i)} \cdot \mathbf{u}'_{(j)}\mathbf{u}_{(j)}}} \\
& = \cos\left(\mathbf{u}_{(i)}, \mathbf{u}_{(j)}\right).
\end{aligned}$$

Tapanahan on useimmiten esittää havaintovektorit pisteinä xy-tasossa eli pisteparvena. Tällöin ortogonaalisen muunnoksen ominaisuus on se, että pisteparven muoto sinänsä säilyy – sitä vain ”kierretään” tai ”peilataan”. Pisteparven kiertämisen sijasta voimme myös ajatella pisteparven pysyvän ennallaan mutta kierto koskeekin koordinaattiakseleita.

On huomattava, että uusien muuttujavektoreiden pituudet ja keskinäiset kulmat voivat muuttua, jos havaintovektoreille tehdään ortogonaalinen muunnos. Tällöinhän

$$\|\underline{\mathbf{x}}\|^2 = \|\mathbf{U}\mathbf{a}\|^2 = \mathbf{a}'\mathbf{U}'\mathbf{U}\mathbf{a}, \quad \underline{\mathbf{x}}'\underline{\mathbf{y}} = \mathbf{a}'\mathbf{U}'\mathbf{U}\mathbf{b}, \tag{6.37}$$

$$\cos(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = \frac{\underline{\mathbf{x}}'\underline{\mathbf{y}}}{\sqrt{\underline{\mathbf{x}}'\underline{\mathbf{x}} \cdot \underline{\mathbf{y}}'\underline{\mathbf{y}}}} = \frac{\mathbf{a}'\mathbf{U}'\mathbf{U}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{U}'\mathbf{U}\mathbf{b} \cdot \mathbf{b}'\mathbf{U}'\mathbf{U}\mathbf{b}}}. \tag{6.38}$$

Jos $\mathbf{U}$ on keskistetty, niin myös $\underline{\mathbf{U}}$ on keskistetty ja

$$\cos(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = \mathrm{cor_d}(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = \frac{\mathbf{a}'\mathbf{S}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{b} \cdot \mathbf{b}'\mathbf{S}\mathbf{b}}}\,, \tag{6.39}$$

missä $\mathbf{S} = \frac{1}{n-1}\mathbf{U}'\mathbf{U} = \mathrm{cov_d}(\mathbf{U}) =$ kovarianssimatriisi. Täten on selvää, että havaintovektoreille tehtävä ortogonaalinen muunnos voi muuttaa muuttujien välistä korrelaatiota.

Jos muuttujavektoreille $\mathbf{x}$ ja $\mathbf{y}$ tehdään ortogonaalinen muunnos, niin se tarkoittaa $\mathbf{x}$:n ja $\mathbf{y}$:n kertomista $n \times n$-matriisilla $\mathbf{B}$, $\mathbf{B}'\mathbf{B} = \mathbf{I}_n$. Tällöin tietenkin

$$\|\mathbf{Bx}\| = \|\mathbf{x}\|,\ \|\mathbf{By}\| = \|\mathbf{y}\| \text{ ja } \cos(\mathbf{Bx}, \mathbf{By}) = \cos(\mathbf{x}, \mathbf{y})\,. \tag{6.40}$$

Voimme ajatella, että x ja y ovat satunnaismuuttujia, joiden muodostama satunnaisvektori on

$$\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})\,, \tag{6.41}$$

ja että $\mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \dots : \mathbf{u}_{(n)})$ on satunnaisotos (kaksiulotteisesta) populaatiosta, jonka odotusarvo on $\boldsymbol{\mu}$ ja kovarianssimatriisi on $\boldsymbol{\Sigma}$ eli jokaisella $\mathbf{u}_{(i)}$:llä on sama jakauma kuin $\mathbf{u}$:lla ja $\mathbf{u}_{(i)}$:t ovat riippumattomia. Satunnaisvektorille $\mathbf{u}$ tehtävä lineaarinen muunnos tarkoittaa $\mathbf{u}$:n kertomista $\mathbf{A}$:lla, jolloin uusi satunnaisvektori $\underline{\mathbf{u}}$ on

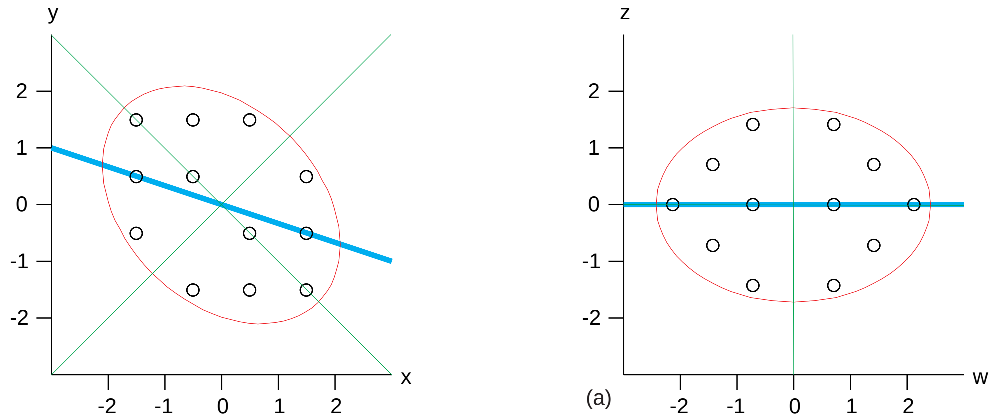
$$\underline{\mathbf{u}} = \begin{pmatrix} \underline{x} \\ \underline{y} \end{pmatrix} = \mathbf{Au}\,. \tag{6.42}$$

Pyrimme säilyttämään termin "muuttujavektori" tarkoittamaan havaintomatriisin pystyriviä eli yhden tietyn muuttujan x havaittujen arvojen (n kpl) muodostamaa pystyvektoria $\mathbf{x}$. On syytä korostaa näin määritellyn vektorin ja (6.42):n mukaisen satunnaisvektorin $\mathbf{u}$ eroa. Vektorin $\mathbf{u}$ elementit ovat kahden eri satunnaismuuttujan x ja y arvoja. Alan kirjallisuudessa näkee $\mathbf{u}$:stakin käytettävän nimitystä "muuttujavektori" – asiayhteydestä käy ilmi mistä on kysymys.

Esimerkki 6.5. Kuviossa (toimituskentässä) 2.8 (s. 87) tarkastelimme Survon avulla lineaarista muunnosta, jossa keskistettyä havaintoaineistoa rotatoitiin seuraavasti:

$$\begin{pmatrix} w \\ z \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} \mathbf{f}' \\ \mathbf{g}' \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} := \mathbf{A}' \begin{pmatrix} x \\ y \end{pmatrix} \implies \underline{\mathbf{U}} = \mathbf{UA}\,. \tag{6.43}$$

Oheisista kuvioista 6.2 näemme kuinka pisteparvi on kiertynyt 45° vastapäivään. Uudet muuttujat w ja z ovat korreloimattomia. Muuttujavektoreiden (arvot keskistettyjä) avulla esitettynä tilanne on se, että $\mathrm{cor_d}(\mathbf{x}, \mathbf{y}) = \cos(\mathbf{x}, \mathbf{y}) = -1/3$, mutta $\mathrm{cor_d}(\mathbf{w}, \mathbf{z}) = \cos(\mathbf{w}, \mathbf{z}) = 0$, sillä $\mathbf{w}'\mathbf{z} = \mathbf{f}'\mathbf{U}'\mathbf{U}\mathbf{g} = 0$. □



Kuvio 6.2. Esimerkin 6.5 muunnoksen vaikutus muuttujavektoreihin.

Esimerkki 6.6. Kuvioissa 6.3 ja 6.4 generoidaan ensin 100 havaintoa 2-ulotteisesta satunnaismuuttujasta $\binom{x_1}{y_1}$, joka noudattaa jakaumaa $\mathrm{N}_2(\mathbf{0}, \mathbf{I}_2)$. Sen jälkeen tehdään muunnos (*kutistus* x-akselilla ja *venytys* y-akselilla)

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} := \mathbf{D}\mathbf{u} \implies \underline{\mathbf{U}} = \mathbf{U}\mathbf{D}' = \mathbf{U}\mathbf{D}.$$

Tämän jälkeen tehdään pisteparven kierto 45° astetta vastapäivään:

$$\begin{pmatrix} x_3 \\ y_3 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} := \mathbf{B} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \implies \underline{\underline{\mathbf{U}}} = \underline{\mathbf{U}}\mathbf{B}',$$

eli satunnaisvektorille $\mathbf{u}$ tehdään muunnos $\underline{\underline{\mathbf{u}}} = \mathbf{B}\mathbf{D}\mathbf{u}$. Pisteparvi siirretään vielä toiseen kohtaan lisäämällä jokaiseen havaintovektoriin vektori $(1, -1)'$. □

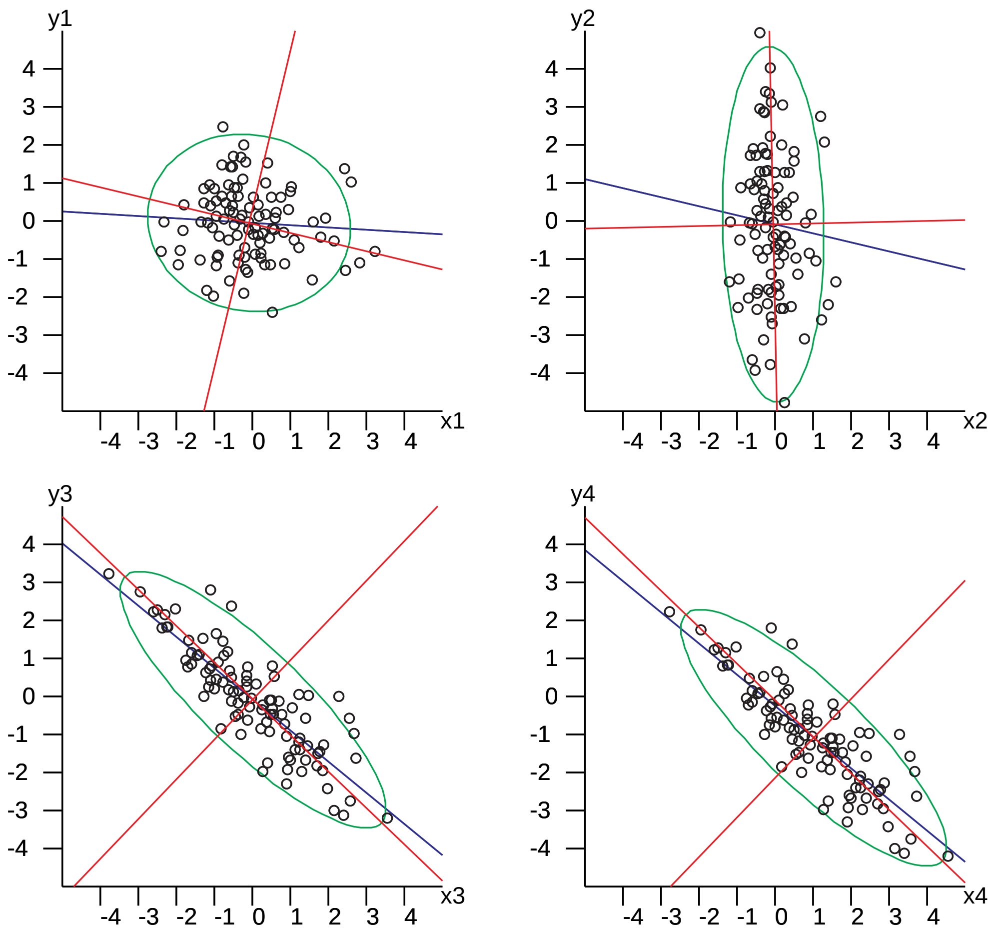
Korostettakoon vielä yhteenvetona, että xy-pisteparven

- kutistus tai venytys eli muunnos $\underline{\mathbf{U}} = \mathbf{U} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = (a\mathbf{x} : b\mathbf{y})$,
- siirto eli muunnos $\underline{\mathbf{U}} = \mathbf{U} + \mathbf{1}_n(c, d) = (\mathbf{x} + c\mathbf{1}_n : \mathbf{y} + d\mathbf{1}_n)$,

eivät vaikuta x:n ja y:n väliseen korrelaatioon, mutta

- pisteparven kierto

saattaa vaikuttaa korrelaatioon.



Kuvio 6.3. Esimerkkiin 6.6 liittyvä kuvio: kutistus, venytys, kierto, siirto.

```
   -   - SURVO MM  Wed Sep 29 19:10:12 2010            C:\SJP\D\   2000  100 0
*
*MAT U1=ZER(100,2)
*MAT #TRANSFORM U1 BY PROBIT(RND(0)) / U1 (=U) on täynnä N(0,1)-tavaraa
*MAT U1(0,1)="X1"       / sarakkeille nimet
*MAT U1(0,2)="Y1"
*MATRIX D ///
*  0.5   0
*  0     2
*
*MAT SAVE D
*MATRIX A  ///
*  q      -q
*  q       q
*
*MAT SAVE A    /  q=1/SQRT(2)  q=0.70710678118655
*MAT LOAD A,CUR+1
*MATRIX A
*///            1        2
*  1       0.70711 -0.70711
*  2       0.70711  0.70711
*
*Toinen mahdollisuus määritellä rotaatiomatriisi on seuraava:
*pi=3.1415926535893        alfa=pi/4  alfa=0.78539816339733
*Näin voidaan määritellä kiertokulma alfa (pi/4 = 45 astetta).
*MATRIX  ROT  ///
*cos(alfa)   -sin(alfa)
*sin(alfa)    cos(alfa)
*
*MAT SAVE ROT       /  Tämä = A
*MAT U2=U1*D        /  U1:n kutistus (x) ja venytys (y)
*MAT U2(0,1)="X2"   /  U2:n sarakkeille nimet X2 ja Y2
*MAT U2(0,2)="Y2"
*MAT U3=U2*A'       /  U3 on U2 rotatoituna 45 astetta vastapäivään
*MAT U3(0,1)="X3"   /  U3:n sarakkeille nimet X3 ja Y3
*MAT U3(0,2)="Y3"
*MATRIX K_ARVO ///
*  1
* -1
*
*MAT SAVE K_ARVO  / keskiarvo-vektori joka lisätään jokaiseen havaintoon
*MAT DIM U1 /* rowU1=100 colU1=2
*MAT ONE=CON(rowU1,1)  / tolppa-1, 100 elementtiä
*MAT U4=U3+ONE*K_ARVO' / U4 = U3 johon lisätty keskiarvot
*MAT U4(0,1)="X4"      / U4:n sarakkeille nimet X4 ja Y4
*MAT U4(0,2)="Y4"
*CORR U2.MAT,CUR+1  /   VARS=X2,Y2
*Means, std.devs and correlations of U2.MAT  N=100
*Variable  Mean        Std.dev.
*X2       -0.034560    0.544386
*Y2       -0.073516    1.908275
*Correlations:
*                  X2      Y2
* X2           1.0000 -0.0676
* Y2          -0.0676  1.0000
*
*CORR U3.MAT,CUR+1  /   VARS=X3,Y3
*Means, std.devs and correlations of U3.MAT  N=100
*Variable  Mean        Std.dev.
*X3        0.027545    1.427977
*Y3       -0.076421    1.377952
*Correlations:
*                  X3      Y3
* X3           1.0000 -0.8500
* Y3          -0.8500  1.0000
*                                     (MK10-009, MK-05-FIG-NJ1)
```

Kuvio 6.4. Esimerkkiin 6.6 liittyvä toimituskenttä.

6.3 Ei-negatiivisesti definiitti matriisi

Jos neliömatriisi $\mathbf{A}_{n\times n}$ on esitettävissä tulona

$$\mathbf{A} = \mathbf{L}\mathbf{L}', \tag{6.44}$$

missä $\mathbf{L}$ on jokin matriisi, jossa on n vaakariviä, sanomme että $\mathbf{A}$ on *ei-negatiivisesti definiitti* matriisi. Voisimme päästä samaan määritelmään myös seuraavasti neliömuodon $\mathbf{x}'\mathbf{A}\mathbf{x}$ avulla: Symmetrinen matriisi $\mathbf{A}$ on ei-negatiivisesti definiitti, jos

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \text{ kaikilla vektoreilla } \mathbf{x} \in \mathbb{R}^n, \tag{6.45}$$

eli 2×2-tapauksessa

$$\mathbf{x}'\mathbf{A}\mathbf{x} = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2 \geq 0 \text{ kaikilla } x_1, x_2 \in \mathbb{R}\,. \tag{6.46}$$

Nämä määritelmät ovat yhtäpitäviä, mutta (6.45):ssä on erikseen mainittava $\mathbf{A}$:n symmetrisyys. Kun tarkastelemme matriisien definiittisyyttä, oletamme tarkasteltavat matriisit symmetrisiksi yleensä ilman eri mainintaa.

Esimerkki 6.7. Mitkä seuraavista matriiseista ovat ei-negatiivisesti definiittejä?

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 1.00001 \\ 1.00001 & 1 \end{pmatrix},$$

$$\mathbf{D} = \begin{pmatrix} 1 & 0.999999 \\ 0.999999 & 1 \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix},$$

$$\mathbf{F} = \begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 1 & 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1 & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} & 1 \end{pmatrix}.$$

□

Mikäli (6.44):ssa $\mathbf{A}$:n aste – mikä on sama kuin $\mathbf{L}$:n aste – on n (eli $\mathbf{A}$ on epäsingulaarinen) on $\mathbf{A}$ *positiivisesti definiitti*. Neliömuoto-ehtona esitettynä tämä merkitsee, että

$$\mathbf{x}'\mathbf{A}\mathbf{x} > 0 \text{ kaikilla vektoreilla } \mathbf{x} \neq \mathbf{0}\,. \tag{6.47}$$

Jos $\mathbf{A}_{n\times n}$ on ei-negatiivisesti definiitti, mutta sen aste on pienempi kuin n (jolloin $\mathbf{A}$ on singulaarinen), on olemassa sellainen $\mathbf{x} \neq \mathbf{0}$, jolla on ominaisuus $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$:

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \quad \forall\, \mathbf{x} \in \mathbb{R}^n, \quad \exists\, \mathbf{x} \neq \mathbf{0} : \mathbf{x}'\mathbf{A}\mathbf{x} = 0\,. \tag{6.48}$$

Havaitsemme, että keskistäjämatriisi $\mathbf{C} = \mathbf{I} - \mathbf{J}$ toteuttaa ehdon

$$\mathbf{x}'\mathbf{C}\mathbf{x} = (\mathbf{C}\mathbf{x})'\mathbf{C}\mathbf{x} \geq 0\,, \tag{6.49}$$

joten $\mathbf{C}$ on ei-negatiivisesti definiitti. Koska vektorilla $\mathbf{1}$ on ominaisuus $\mathbf{1}'\mathbf{C}\mathbf{1} = \mathbf{1}'(\mathbf{I} - \mathbf{J})\mathbf{1} = 0$, ei $\mathbf{C}$ ole positiivisesti definiitti. Vastaavalla tavalla voimme päätellä, että mikä hyvänsä ortogonaaliprojektori (symmetrinen idempotetti) $\mathbf{P}$ on ei-negatiivisesti definiitti.

Jos $\mathbf{A}$ on ei-negatiivisesti definiitti, voimme merkitä lyhyesti

$$\mathbf{A} \geq_{\mathsf{L}} \mathbf{0}\,, \quad \text{ts.} \quad \mathbf{A} \in \mathrm{NND}_n\,, \tag{6.50a}$$

missä NND_n = kaikkien ei-negatiivisesti definiittien matriisien joukko. Jos $\mathbf{A}$ on positiivisesti definiitti,

$$\mathbf{A} >_{\mathsf{L}} \mathbf{0}\,, \quad \text{ts.} \quad \mathbf{A} \in \mathrm{PD}_n\,, \tag{6.50b}$$

missä PD_n = kaikkien positiivisten definiittien matriisien joukko. Näiden merkintätapojen ei pidä antaa johtaa siihen mielikuvaan, että *jokainen* $\mathbf{A}$:n elementti olisi ≥ 0. Sen sijaan valitsemalla ehdossa (6.45) $\mathbf{x} = \mathbf{i}_i$ voimme heti päätellä, että

$$\mathbf{A} \text{ on ei-negatiivisesti definiitti} \implies \mathbf{A}\text{:n lävistäjäelementit } a_{ii} \geq 0\,, \tag{6.51}$$

ja $a_{ii} > 0, i = 1, \dots, n$, jos $\mathbf{A}$ on positiivisesti definiitti.

Jos merkitään

$$\mathbf{A} = \{a_{ij}\} = \mathbf{Q}'\mathbf{Q} = \{\mathbf{q}_i'\mathbf{q}_j\} = (\mathbf{Q}'\mathbf{q}_1 : \dots : \mathbf{Q}'\mathbf{q}_n)\,, \tag{6.52}$$

missä $\mathbf{Q} = (\mathbf{q}_1 : \dots : \mathbf{q}_n)$, on Cauchyn–Schwarzin epäyhtälön (ks. luku 6.7, s. 207) nojalla voimassa

$$(\mathbf{q}_i'\mathbf{q}_j)^2 \leq \mathbf{q}_i'\mathbf{q}_i \cdot \mathbf{q}_j'\mathbf{q}_j\,, \tag{6.53}$$

eli ei-negatiivisesti definiitin matriisin elementtien välillä on seuraava epäyhtälö:

$$a_{ij}^2 \leq a_{ii}a_{jj}\,, \quad i, j = 1, \dots, n\,. \tag{6.54}$$

Erityisesti havaitsemme, että jos ei-negatiivisesti definiitin matriisin lävistäjäelementit ovat keskenään yhtäsuuria, niin

$$a_{11} = a_{22} = \dots = a_{nn} \implies a_{ij}^2 \leq a_{ii}^2\,, \quad j = 1, \dots, n\,. \tag{6.55}$$

Samoin jos ei-negatiivisesti definiitin matriisin i. lävistäjäelementti on 0, niin kaikki i. sarakkeen (ja i. rivin) elementit ovat nollia:

$$a_{ii} = 0 \implies \mathbf{q}_i'\mathbf{q}_i = 0 \implies \mathbf{q}_i = \mathbf{0} \implies \mathbf{a}_i = \mathbf{Q}'\mathbf{q}_i = \mathbf{0}\,, \tag{6.56}$$

missä siis $\mathbf{A} = (\mathbf{a}_1 : \ldots : \mathbf{a}_n)$.

Merkintä $\mathbf{A} \leq_{\mathsf{L}} \mathbf{B}$ tarkoittaa, että

$$\text{"}\mathbf{A} \text{ on } \mathbf{B}\text{:tä pienempi Löwnerin mielessä".} \tag{6.57}$$

Kyseessä on ns. *Löwnerin järjestys*:

$$\boxed{\mathbf{A} \leq_{\mathsf{L}} \mathbf{B} \iff \mathbf{B} - \mathbf{A} \geq_{\mathsf{L}} \mathbf{0} \iff \text{ on olemassa } \mathbf{F}\colon \mathbf{B} - \mathbf{A} = \mathbf{F}\mathbf{F}'. \quad (6.58)}$$

Määritelmällä (6.58) on keskeinen merkitys esim. lineaaristen mallien estimoinnin yhteydessä. Kannattaa panna merkille, että (6.58):ssa ei $\mathbf{A}$:n ja $\mathbf{B}$:n tarvitse välttämättä olla symmetrisiä, erotus $\mathbf{B}-\mathbf{A}$ kuitenkin on symmetrinen. Matriisiepäyhtälöstä $\mathbf{A} \leq_{\mathsf{L}} \mathbf{B}$ seuraa monia hyödyllisiä tuloksia:

$$\begin{aligned} &\mathbf{A} \leq_{\mathsf{L}} \mathbf{B} \implies \\ &a_{ii} \leq b_{ii}\,, \ \operatorname{tr}(\mathbf{A}) \leq \operatorname{tr}(\mathbf{B})\,, \ \det(\mathbf{A}) \leq \det(\mathbf{B})\,, \ \operatorname{ch}_i(\mathbf{A}) \leq \operatorname{ch}_i(\mathbf{B})\,. \end{aligned} \tag{6.59}$$

Seuraavat Löwnerin järjestyksen ominaisuudet on helppo osoittaa:

LÖ1. $\mathbf{A}, \mathbf{B} \geq_{\mathsf{L}} \mathbf{0} \implies \mathbf{A} + \mathbf{B} \geq_{\mathsf{L}} \mathbf{0}$,

LÖ2. $\mathbf{A} \geq_{\mathsf{L}} \mathbf{0} \iff \mathbf{U}'\mathbf{A}\mathbf{U} \geq_{\mathsf{L}} \mathbf{0} \ \forall \ \mathbf{U}$,

LÖ3. $\mathbf{A} \geq_{\mathsf{L}} \mathbf{B}$ ja $\mathbf{B} \geq_{\mathsf{L}} \mathbf{A} \implies \mathbf{A} = \mathbf{B}$ (antisymmetrisyys),

LÖ4. $\mathbf{A} \geq_{\mathsf{L}} \mathbf{A}$ (refleksisyys),

LÖ5. $\mathbf{A} \geq_{\mathsf{L}} \mathbf{B}$ ja $\mathbf{B} \geq_{\mathsf{L}} \mathbf{C} \implies \mathbf{A} \geq_{\mathsf{L}} \mathbf{C}$ (transitiivisuus).

Ominaisuudet *LÖ3–LÖ5* merkitsevät, että Löwnerin järjestys on ns. *osittaisjärjestys*; perusteellinen lähde tähän aiheeseen on Marshall, Olkin & Arnold (2011). Järjestys on saanut nimensä Karl Löwnerin artikkelista vuodelta (1934).

On huomattava, että aina ei kumpikaan järjestyksistä $\mathbf{A} \leq_{\mathsf{L}} \mathbf{B}$, $\mathbf{B} \leq_{\mathsf{L}} \mathbf{A}$ ole välttämättä voimassa. Jos esimerkiksi

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{6.60}$$

niin $\mathbf{A} \geq_{\mathsf{L}} \mathbf{0}$ ja $\mathbf{B} \geq_{\mathsf{L}} \mathbf{0}$, mutta kumpikaan järjestyksistä $\mathbf{A} \leq_{\mathsf{L}} \mathbf{B}$, $\mathbf{B} \leq_{\mathsf{L}} \mathbf{A}$ ei ole voimassa.

Matriisin definiittisyys liittyy myös erityisesti matriisin ominaisarvoihin sekä moniin muihin mielenkiintoisiin tarkasteluihin, joihin palataan myöhemmin. Monet tilastotieteessä keskeiset matriisit – kuten korrelaatio- ja kovarianssimatriisi – ovat ei-negatiivisesti definiittejä. Kovarianssi- ja korrelaatiomatriisi voidaan esittää muodossa

$$\mathbf{S} = \tfrac{1}{n-1}\,\mathbf{U}'\mathbf{C}\mathbf{U} = \tfrac{1}{n-1}\,(\mathbf{C}\mathbf{U})'\mathbf{C}\mathbf{U}, \quad \mathbf{R} = \tilde{\tilde{\mathbf{U}}}'\tilde{\tilde{\mathbf{U}}}, \tag{6.61}$$

missä $\mathbf{C}$ on keskistäjämatriisi ja $\tilde{\tilde{\mathbf{U}}} = \tilde{\mathbf{U}}\big[\mathrm{diag}(\tilde{\mathbf{U}}'\tilde{\mathbf{U}})\big]^{-1/2}$ sarakkeiltaan 1:n pituiseksi skaalattu $\tilde{\mathbf{U}}$ ($\tilde{\mathbf{U}}$ on keskistetty havaintomatriisi $\mathbf{U}$). Kaavoista (6.61) nähdään, että

> *otoskovarianssi- ja otoskorrelaatiomatriisi ovat aina ei-negatiivisesti definiittejä. Sama tulos pätee myös satunnaisvektoreiden kovarianssi- ja korrelaatiomatriiseille.*

Olkoon $\mathbf{A}$ symmetrinen. On osoitettavissa, että seuraavat väitteet ovat yhtäpitäviä:

PD1. $\mathbf{A}$ on positiivisesti definiitti,

PD2. $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ kaikilla $\mathbf{x} \neq \mathbf{0}$,

PD3. $\mathrm{ch}_i(\mathbf{A}) > 0, \ i = 1, \dots, n$,

PD4. on olemassa $\mathbf{F}\colon \mathbf{A} = \mathbf{F}\mathbf{F}'$, $\mathbf{F} \in \mathbb{R}^{n\times n}$, $\mathrm{r}(\mathbf{F}) = n$,

PD5. kaikki johtavat pääminorit > 0,

PD6. $\mathrm{tr}(\mathbf{AUU}') \leq \mathrm{tr}(\mathbf{BUU}')$ kaikilla $\mathbf{U} \neq \mathbf{0}$.

Samoin seuraavat väitteet ovat yhtäpitäviä:

NND1. $\mathbf{A}$ ei-negatiivisesti definiitti,

NND2. $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ kaikilla $\mathbf{x}$,

NND3. $\mathrm{ch}_i(\mathbf{A}) \geq 0, \ i = 1, \dots, n$,

NND4. on olemassa $\mathbf{F}\colon \mathbf{A} = \mathbf{F}\mathbf{F}'$,

NND5. kaikki pääminorit ≥ 0,

NND6. $\mathrm{tr}(\mathbf{AUU}') \leq \mathrm{tr}(\mathbf{BUU}') \ \forall\, \mathbf{U}$.

Harjoitustehtäväksi jätetään seuraavan tuloksen todistus:

$$\mathbf{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathrm{NND}_2 \iff \mathrm{tr}(\mathbf{A}) \geq 0 \ \& \ \det(\mathbf{A}) \geq 0\,. \tag{6.62}$$

Termi *pääminori* tarkoittaa $\mathbf{A}$:n sellaisen osamatriisin determinanttia, josta on poistettu samat pystyrivit ja vaakarivit. Jos $\mathbf{A}_{kk}$ on $\mathbf{A}$:n sellainen alimatriisi, joka muodostuu $\mathbf{A}$:n vasemmasta ylänurkasta $(k \times k)$, niin *johtava pääminori* D_k on

$$D_k = \det(\mathbf{A}_{kk})\,. \tag{6.63}$$

Pääminoreita on kaikkiaan 2^n kpl (miksi?) mutta johtavia pääminoreita vain n kpl. Ehto (PD5) on tällöin näillä merkinnöillä:

$$\mathbf{A} \text{ on positiivisesti definiitti} \iff D_k > 0,\ k = 1, \ldots, n\,. \tag{6.64}$$

On huomattava, että ei-negatiivisesti definiittisyyttä *ei* voida luonnehtia johtavien pääminoreiden avulla ehdolla

$$D_k \geq 0, \quad k = 1, \ldots, n\,. \tag{6.65}$$

Jos nimittäin $\mathbf{A}$ on ei-negatiivisesti definiitti, niin silloin kaikki $\mathbf{A}$:n pääminorit ovat ≥ 0, mistä seuraa, että erityisesti johtavat pääminorit ovat ≥ 0 eli ehto (6.65) toteutuu. Ehto (6.65) on siis välttämätön mutta ei rittävä ehto $\mathbf{A}$:n ei-negatiivisesti definiittisyydelle.

Symmetrisen matriisin $\mathbf{A}$ definiittisyys voidaan kätevästi luonnehtia $\mathbf{A}$:n ositteiden avulla seuraavasti. Olkoon symmetrinen $\mathbf{A}$ ositettu seuraavasti:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \tag{6.66}$$

missä $\mathbf{A}_{11}$ on neliömatriisi. Silloin seuraavat kolme väitettä ovat yhtäpitäviä:

(a) $\mathbf{A} \geq_{\mathsf{L}} \mathbf{0}$,

(b_1) $\mathbf{A}_{11} \geq_{\mathsf{L}} \mathbf{0}$, ($\mathrm{b}_2$) $\mathscr{C}(\mathbf{A}_{12}) \subset \mathscr{C}(\mathbf{A}_{11})$, ($\mathrm{b}_3$) $\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-}\mathbf{A}_{12} \geq_{\mathsf{L}} \mathbf{0}$,

(c_1) $\mathbf{A}_{22} \geq_{\mathsf{L}} \mathbf{0}$, ($\mathrm{c}_2$) $\mathscr{C}(\mathbf{A}_{21}) \subset \mathscr{C}(\mathbf{A}_{22})$, ($\mathrm{c}_3$) $\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-}\mathbf{A}_{21} \geq_{\mathsf{L}} \mathbf{0}$.

Em. tulosta sanotaan usein Albertin lauseeksi, ks. Albert (1969) ja (1972, Th. 9.1.6) sekä Puntanen, Styan & Isotalo (2011, Th. 14). Jos $\mathbf{A}_{11}$ ja $\mathbf{A}_{22}$ ovat positiivisesti definiittejä, on välittömästi voimassa $\mathscr{C}(\mathbf{A}_{12}) \subset \mathscr{C}(\mathbf{A}_{11})$ ja $\mathscr{C}(\mathbf{A}_{21}) \subset \mathscr{C}(\mathbf{A}_{22})$. Täten seuraavat kolme väitettä ovat yhtäpitäviä:

(a) $\mathbf{A} >_{\mathsf{L}} \mathbf{0}$,

(b_1) $\mathbf{A}_{11} >_{\mathsf{L}} \mathbf{0}$, ($\mathrm{b}_3$) $\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} >_{\mathsf{L}} \mathbf{0}$,

(c_1) $\mathbf{A}_{22} >_{\mathsf{L}} \mathbf{0}$, ($\mathrm{c}_3$) $\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} >_{\mathsf{L}} \mathbf{0}$.

Valitsemalla $\mathbf{A}_{12} = \mathbf{I}$ ja merkitsemällä $\mathbf{A}_{11} = \mathbf{U}$, $\mathbf{A}_{22}^{-1} = \mathbf{V}$, saamme

$$\mathbf{U} \geq_{\mathsf{L}} \mathbf{V} \iff \mathbf{V}^{-1} \geq_{\mathsf{L}} \mathbf{U}^{-1} \quad \text{ja} \quad \mathbf{U} >_{\mathsf{L}} \mathbf{V} \iff \mathbf{V}^{-1} >_{\mathsf{L}} \mathbf{U}^{-1}. \tag{6.67}$$

Lainaamme tähän Pukelsheimin (1993, s. 13) näppärän todistuksen tulokselle (6.67). Olkoot $\mathbf{A}$ ja $\mathbf{B}$ positiivisesti definiittejä ja $\mathbf{A} \leq_{\mathsf{L}} \mathbf{B}$ eli $\mathbf{B} - \mathbf{A} \geq_{\mathsf{L}} \mathbf{0}$. Tällöin (miksi?)

$$\mathbf{0} \leq_{\mathsf{L}} (\mathbf{B} - \mathbf{A})\mathbf{A}^{-1}(\mathbf{B} - \mathbf{A}) + (\mathbf{B} - \mathbf{A}) = \mathbf{B}\mathbf{A}^{-1}\mathbf{B} - \mathbf{B}. \tag{6.68}$$

Väite seuraa, kun kerrotaan (6.68) vasemmalta ja oikealta matriisilla $\mathbf{B}^{-1}$.

6.4 Matriisin neliöjuuri

Aikaisemmin on jo tarkasteltu lävistäjämatriisin (jonka kaikki elementit > 0) neliöjuurta:

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}, \quad \mathbf{D}^{1/2} = \begin{pmatrix} \sqrt{d_1} & 0 & \dots & 0 \\ 0 & \sqrt{d_2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sqrt{d_n} \end{pmatrix}. \tag{6.69}$$

Yleisessä tapauksessa neliömatriisin $\mathbf{A}$ neliöjuurella $\mathbf{A}^{1/2}$ tarkoitetaan sellaista matriisia, jolla on ominaisuus

$$\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}\,. \tag{6.70}$$

Tällöin on selvää, että $\mathbf{A}$:n neliöjuureksi ei välttämättä kelpaa sellainen matriisi, jonka elementteinä ovat $\mathbf{A}$:n elementtien neliöjuuret. Jos esim. $\mathbf{P}$ on idempotentti matriisi, niin silloin tietenkin $\mathbf{P}$ itse on neliöjuuri $\mathbf{P}$:stä.

Olkoon symmetrisellä positiivisesti definiitillä $n \times n$-matriisilla $\mathbf{A}$ ominaisarvohajotelma

$$\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}', \tag{OAH}$$

missä $\mathbf{T}$ on ortogonaalinen matriisi ja $\boldsymbol{\Lambda}$ on lävistäjämatriisi:

$$\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)\,, \;\; \lambda_1 \geq \dots \geq \lambda_n > 0\,, \;\; \mathbf{T}'\mathbf{T} = \mathbf{T}\mathbf{T}' = \mathbf{I}_n\,. \tag{6.71}$$

Lävistäjämatriisin $\boldsymbol{\Lambda}$ elementit λ_i ovat $\mathbf{A}$:n ominaisarvoja ja $\mathbf{T}$:n sarakkeet $\mathbf{t}_i$ ovat vastaavia ominaisvektoreita. Kertomalla (OAH):n vasemmalta $\mathbf{T}'$:lla ja oikealta $\mathbf{T}$:llä saamme hajotelman

$$\mathbf{T}'\mathbf{A}\mathbf{T} = \boldsymbol{\Lambda}\,. \tag{6.72}$$

Tällöin on tietenkin voimassa

$$\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}' = \mathbf{T}\boldsymbol{\Lambda}^{1/2}\mathbf{T}' \cdot \mathbf{T}\boldsymbol{\Lambda}^{1/2}\mathbf{T}' = \mathbf{U} \cdot \mathbf{U}\,, \tag{6.73}$$

joten matriisi $\mathbf{U} = \mathbf{T}\boldsymbol{\Lambda}^{1/2}\mathbf{T}'$ on $\mathbf{A}$:n symmetrinen positiivisesti definiitti neliöjuuri:

$$\boxed{\mathbf{A}^{1/2} = \mathbf{T}\boldsymbol{\Lambda}^{1/2}\mathbf{T}'.} \tag{6.74}$$

Kaikilla matriiseilla ei tietenkään ole hajotelmaa (OAH). Symmetrisellä $\mathbf{A}$:lla hajotelma on sinänsä aina olemassa, mutta osa ominaisarvoista voi olla negatiivisia. Jos $\mathbf{A}$ on positiivisesti definiitti, niin silloin sillä on hajotelma (OAH), missä kaikki λ_i:t ovat nollaa suurempia, ja $\mathbf{A}$:n neliöjuuri voidaan muodostaa (6.74):n mukaan. Jos $\mathbf{A}$ on singulaarinen ei-negatiivisesti definiitti matriisi, niin jotkut λ_i:t ovat nollia. Olkoon $\mathbf{A}$:n aste $\mathrm{r}(\mathbf{A}) = r$, jolloin

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0, \; \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0.$$

Voimme kirjoittaa (OAH):n tällöin seuraavasti:

$$\mathbf{A} = \mathbf{T\Lambda T}' = (\mathbf{T}_1 : \mathbf{T}_0) \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{T}_1' \\ \mathbf{T}_0' \end{pmatrix} = \mathbf{T}_1\mathbf{\Lambda}_1\mathbf{T}_1', \tag{6.75}$$

missä $\mathbf{\Lambda}_1 = \mathrm{diag}(\lambda_1, \dots, \lambda_r)$, ja $\mathbf{T}_1$ sisältää näitä (nollasta poikkeavia) ominaisarvoja vastaavat ominaisvektorit ja $\mathbf{T}_0$ muodostuu ominaisarvoa 0 vastaavista ortonormaaleista ominaisvektoreista. Matriisin $\mathbf{A}$ neliöjuuri on tällöin tietenkin

$$\mathbf{A}^{1/2} = \mathbf{T}_1\mathbf{\Lambda}_1^{1/2}\mathbf{T}_1' = \mathbf{T\Lambda}^{1/2}\mathbf{T}'. \tag{6.76}$$

Ei-negatiivisesti definiitin matriisin ei-negatiivisesti definiitti neliöjuuri osoittautuu yksikäsitteiseksi; ks. esim. Bapat (2012, s. 24).

Erityisesti huomattakoon, että $\mathbf{T}$:n ortogonaalisuuden vuoksi

$$\mathbf{A}^k = (\mathbf{T\Lambda T}')^k = \mathbf{T\Lambda}^k\mathbf{T}', \; k = 1, 2, \dots, \tag{6.77}$$

ja jos $\mathbf{A}$ on epäsingulaarinen, niin

$$\mathbf{A}^{-1} = (\mathbf{T\Lambda T}')^{-1} = \mathbf{T\Lambda}^{-1}\mathbf{T}'. \tag{6.78}$$

Jos $\mathbf{A}$ on positiivisesti definiitti, niin $\mathbf{A}^{1/2}$:n käänteismatriisi on

$$\mathbf{A}^{-1/2} = (\mathbf{T\Lambda}^{1/2}\mathbf{T}')^{-1} = \mathbf{T\Lambda}^{-1/2}\mathbf{T}', \tag{6.79}$$

joten

$$\mathbf{A}^{-1/2}\mathbf{A} = \mathbf{T\Lambda}^{-1/2}\mathbf{T}' \cdot \mathbf{T\Lambda T}' = \mathbf{T\Lambda}^{1/2}\mathbf{T}' = \mathbf{A}^{1/2}, \tag{6.80}$$

$$\mathbf{A}^{-1/2}\mathbf{A}\mathbf{A}^{-1/2} = \mathbf{T\Lambda}^{-1/2}\mathbf{T}' \cdot \mathbf{T\Lambda T}' \cdot \mathbf{T\Lambda}^{-1/2}\mathbf{T}' = \mathbf{I}. \tag{6.81}$$

On myös hyödyllistä havaita, että hajotelmasta $\mathbf{A} = \mathbf{T\Lambda T}'$ seuraa $\mathbf{A}$:lle esitys

$$\mathbf{A} = \mathbf{T\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{T}' = \mathbf{T\Lambda}^{1/2}(\mathbf{T\Lambda}^{1/2})' := \mathbf{WW}'. \tag{6.82}$$

Jos merkitsemme

$$\mathbf{F} = \mathbf{W}^{-1} = (\mathbf{T\Lambda}^{1/2})^{-1} = \mathbf{\Lambda}^{-1/2}\mathbf{T}', \tag{6.83}$$

saamme hajotelman

$$\boxed{\mathbf{FAF}' = \mathbf{\Lambda}^{-1/2}\mathbf{T}'\mathbf{AT\Lambda}^{-1/2} = \mathbf{\Lambda}^{-1/2}\mathbf{T}' \cdot \mathbf{T\Lambda T}' \cdot \mathbf{T\Lambda}^{-1/2} = \mathbf{I}_n. \quad (6.84)}$$

Kaavan (6.81) perusteella on luonnollista määritellä $\mathbf{A}^0$ seuraavasti:

$$\mathbf{A}^0 = \mathbf{T}\mathbf{\Lambda}^0\mathbf{T}' = (\mathbf{T}_1 : \mathbf{T}_0)\begin{pmatrix}\mathbf{\Lambda}_1^0 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{T}_1'\\ \mathbf{T}_0'\end{pmatrix} = \mathbf{T}_1\mathbf{T}_1', \tag{6.85}$$

missä $\mathbf{\Lambda}_1^0 = \mathbf{I}_r$. On helppo näyttää, että

$$\mathbf{A}^0 = \mathbf{T}_1\mathbf{T}_1' = \mathbf{P}_{\mathbf{A}}. \tag{6.86}$$

6.5 Epäyhtälöitä

6.5.1 Muutamia ”traceen” liittyviä epäyhtälöitä

Olkoot $\mathbf{A}$ ja $\mathbf{B}$ symmetrisiä ei-negatiivisesti definiittejä $n \times n$-matriiseja. Olkoon $\mathbf{A}$:lla ominaisarvohajotelma $\mathbf{A} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}'$, missä $\mathbf{T}$ on ortogonaalinen matriisi ja $\mathbf{\Lambda}$ on lävistäjämatriisi:

$$\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)\,, \quad \lambda_1 \geq \cdots \geq \lambda_n \geq 0\,, \quad \mathbf{T}'\mathbf{T} = \mathbf{I}_n\,. \tag{6.87}$$

Osoita että

$$\boxed{\operatorname{tr}(\mathbf{AB}) \leq \lambda_1 \operatorname{tr}(\mathbf{B}) = \lambda_{\max}(\mathbf{A}) \operatorname{tr}(\mathbf{B})\,. \tag{6.88}}$$

Ratkaisu: Koska $\mathbf{A} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}'$, on myös $\mathbf{T}'\mathbf{A}\mathbf{T} = \mathbf{\Lambda}$, joten

$$\begin{aligned}
\operatorname{tr}(\mathbf{AB}) &= \operatorname{tr}(\mathbf{T}\mathbf{\Lambda}\mathbf{T}' \cdot \mathbf{B}) = \operatorname{tr}(\mathbf{\Lambda} \cdot \mathbf{T}'\mathbf{B}\mathbf{T}) \qquad [\operatorname{tr}(\mathbf{FG}) = \operatorname{tr}(\mathbf{GF})]\\
&:= \operatorname{tr}(\mathbf{\Lambda} \cdot \mathbf{U}) \qquad [\text{kun merk. } \mathbf{U} = \mathbf{T}'\mathbf{B}\mathbf{T}]\\
&= \lambda_1 u_{11} + \lambda_2 u_{22} + \cdots + \lambda_n u_{nn}\\
&\leq \lambda_1 (u_{11} + u_{22} + \cdots + u_{nn}) \qquad [\mathbf{U} \geq_{\mathsf{L}} \mathbf{0} \text{ joten } u_{ii} \geq 0 \ \ \forall i]\\
&= \lambda_1 \operatorname{tr}(\mathbf{U}) = \lambda_1 \operatorname{tr}(\mathbf{T}'\mathbf{B}\mathbf{T})\\
&= \lambda_1 \operatorname{tr}(\mathbf{T}\mathbf{T}'\mathbf{B}) = \lambda_1 \operatorname{tr}(\mathbf{B})\,.
\end{aligned} \tag{6.89}$$

Samoin on voimassa epäyhtälö

$$\boxed{\operatorname{tr}(\mathbf{AB}) \leq \operatorname{tr}(\mathbf{A}) \operatorname{tr}(\mathbf{B})\,, \tag{6.90}}$$

sillä (6.88):n todistuksen perusteella

$$\begin{aligned}
\operatorname{tr}(\mathbf{AB}) &= \lambda_1 u_{11} + \lambda_2 u_{22} + \cdots + \lambda_n u_{nn}\\
&\leq (\lambda_1 + \lambda_2 + \cdots + \lambda_n)(u_{11} + u_{22} + \cdots + u_{nn})\\
&= \operatorname{tr}(\mathbf{\Lambda}) \operatorname{tr}(\mathbf{U}) = \operatorname{tr}(\mathbf{T}'\mathbf{A}\mathbf{T}) \operatorname{tr}(\mathbf{T}'\mathbf{B}\mathbf{T})\\
&= \operatorname{tr}(\mathbf{A}) \operatorname{tr}(\mathbf{B})\,.
\end{aligned} \tag{6.91}$$

Todistamme seuraavan tuloksen:

Olkoon $\mathbf{V}$ *symmetrinen ei-negatiivisesti definiitti* $p \times p$*-matriisi. Silloin*

$$\mathbf{1}'\mathbf{V}\mathbf{1} \geq \frac{p}{p-1}[\mathbf{1}'\mathbf{V}\mathbf{1} - \operatorname{tr}(\mathbf{V})], \tag{6.92}$$

ts. (*olettaen että* $\mathbf{1}'\mathbf{V}\mathbf{1} \neq 0$),

$$1 \geq \frac{p}{p-1}\left(1 - \frac{\operatorname{tr}(\mathbf{V})}{\mathbf{1}'\mathbf{V}\mathbf{1}}\right). \tag{6.93}$$

Yhtäsuuruus (6.92)*:ssä saavutetaan jos ja vain jos* $\mathbf{V} = \delta^2\mathbf{1}\mathbf{1}'$ *jollakin* $\delta \in \mathbb{R}$.

Kuten Vehkalahti, Puntanen & Tarkkonen (2009, Th. 1), kirjoitetaan (6.92) muodossa $(p-1)\mathbf{1}'\mathbf{V}\mathbf{1} \geq p\mathbf{1}'\mathbf{V}\mathbf{1} - p\operatorname{tr}(\mathbf{V})$ ts.

$$\frac{\mathbf{1}'\mathbf{V}\mathbf{1}}{\mathbf{1}'\mathbf{1}} \leq \operatorname{tr}(\mathbf{V}) = \lambda_1 + \lambda_2 + \cdots + \lambda_p\,, \tag{6.94}$$

missä $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ ovat $\mathbf{V}$:n ominaisarvot; $\lambda_i = \operatorname{ch}_i(\mathbf{V})$. Koska

$$\max_{\mathbf{z} \neq \mathbf{0}} \frac{\mathbf{z}'\mathbf{V}\mathbf{z}}{\mathbf{z}'\mathbf{z}} = \lambda_1 = \operatorname{ch}_1(\mathbf{V})\,,$$

epäyhtälö (6.94) on voimassa. Yhtäsuuruus (6.94):ssä merkitsee, että

$$\lambda_1 \geq \frac{\mathbf{1}'\mathbf{V}\mathbf{1}}{\mathbf{1}'\mathbf{1}} = \lambda_1 + \lambda_2 + \cdots + \lambda_p\,,$$

mikä on voimassa jos ja vain jos $\lambda_2 = \cdots = \lambda_p = 0$ ja $(\lambda_1, \mathbf{1})$ on $\mathbf{V}$:n ominaispari eli $\mathbf{V}$ on muotoa $\mathbf{V} = \lambda_1\mathbf{1}\mathbf{1}'$.

Lähteenä tulokseen (6.92) mainittakoon myös Vehkalahti (2000, Lemma 4.1). Epäyhtälön(6.93) oikeanpuoleinen termi eli

$$\frac{p}{p-1}\left(1 - \frac{\operatorname{tr}(\mathbf{V})}{\mathbf{1}'\mathbf{V}\mathbf{1}}\right) := \alpha(\mathbf{1})\,, \tag{6.95}$$

voidaan tulkita Cronbachin *alphaksi*, ks. Cronbach (1951).

Osoitetaan seuraavaksi tulos:

Maksimi: $\operatorname{tr}(\mathbf{G}'\mathbf{A}\mathbf{G})$ ehdolla $\mathbf{G}'\mathbf{G} = \mathbf{I}_k$. *Olkoon $\mathbf{A}_{n\times n}$ symmetrinen matriisi, jonka ominaisarvohajotelma on*

$$\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}', \quad \boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n), \quad \operatorname{r}(\mathbf{A}) = r. \tag{6.96}$$

missä $\mathbf{T} = (\mathbf{t}_1 : \ldots : \mathbf{t}_n)$, $\mathbf{T}'\mathbf{T} = \mathbf{I}_n$. Olkoon k jokin annettu kokonaisluku, $k \leq n$. Silloin

$$\max_{\mathbf{G}'\mathbf{G}=\mathbf{I}_k} \operatorname{tr}(\mathbf{G}'\mathbf{A}\mathbf{G}) = \lambda_1 + \cdots + \lambda_k, \tag{6.97}$$

missä yläraja saavutetaan kun

$$\mathbf{G} = (\mathbf{t}_1 : \ldots : \mathbf{t}_k) := \mathbf{T}_{(k)}. \tag{6.98}$$

Tuloksen (6.97) osoittamiseksi kirjoitetaan ensin

$$\operatorname{tr}(\mathbf{G}'\mathbf{A}\mathbf{G}) = \operatorname{tr}(\mathbf{G}'\mathbf{T}\boldsymbol{\Lambda}\mathbf{T}'\mathbf{G}) := \operatorname{tr}(\mathbf{Y}'\boldsymbol{\Lambda}\mathbf{Y}) = \operatorname{tr}(\mathbf{Y}\mathbf{Y}'\boldsymbol{\Lambda}), \tag{6.99}$$

missä $\mathbf{Y} = \mathbf{T}'\mathbf{G}$. Nyt $\mathbf{Y}'\mathbf{Y} = \mathbf{G}'\mathbf{T}\mathbf{T}'\mathbf{G} = \mathbf{I}_k$, ja matriisi $\mathbf{Y}\mathbf{Y}' := \mathbf{P}_{n\times n}$ on ortogonaaliprojektori, jonka lävistäjäelementeillä on ominaisuus

$$0 \leq p_{ii} \leq 1, \quad i = 1, \dots, n, \quad p_{11} + \cdots + p_{nn} = k = \operatorname{r}(\mathbf{P}). \tag{6.100}$$

Täten

$$\operatorname{tr}(\mathbf{G}'\mathbf{A}\mathbf{G}) = \operatorname{tr}(\mathbf{P}\boldsymbol{\Lambda}) = p_{11}\lambda_1 + \cdots + p_{kk}\lambda_k + \cdots + p_{nn}\lambda_n, \tag{6.101}$$

joten (6.100):n perusteella

$$\operatorname{tr}(\mathbf{G}'\mathbf{A}\mathbf{G}) \leq \lambda_1 + \cdots + \lambda_k. \tag{6.102}$$

Yläraja (6.102):ssa saavutetaan tietenkin kun (6.98) on voimassa.

Lähteinä tulokseen (6.97) mainittakoon Rao (1973, p. 63) ja Harville (1997, Th. 21.12.5). Tulos (6.97) voidaan todistaa myös *Poincarén lauseen* nojalla; ks. esim. Rao (1973, p. 64), Rao (1980, Th. 2.1), Scott & Styan (1985, p. 213), Abadir & Magnus (2005, p. 347), ja Mäkeläinen (1970a, Th. 4.1, Cor. 4.2.2).

Poincarén lause: *Olkoon $\mathbf{A}_{n\times n}$ symmetrinen ja $\mathbf{G}_{n\times k}$ toteuttaa ehdon $\mathbf{G}'\mathbf{G} = \mathbf{I}_k$, $k \leq n$. Silloin*

$$\operatorname{ch}_{n-k+i}(\mathbf{A}) \leq \operatorname{ch}_i(\mathbf{G}'\mathbf{A}\mathbf{G}) \leq \operatorname{ch}_i(\mathbf{A}), \quad i = 1, \dots, k. \tag{6.103}$$

Yläraja (6.103):ssa saavutetaan kun

$$\mathbf{G} = (\mathbf{t}_1 : \ldots : \mathbf{t}_k) = \mathbf{T}_{(k)}, \tag{6.104}$$

ja alaraja kun

$$\mathbf{G} = (\mathbf{t}_{n-k+1} : \ldots : \mathbf{t}_n) := \mathbf{T}_{[k]}, \tag{6.105}$$

missä $\mathbf{T}$:n sarakkeet ovat $\mathbf{A}$:n ortonormaalit ominaisvektorit.

6.5.2 Approksimointi alempiasteisella matriisilla

Seuraavan tuloksen ovat esittäneet mm. Puntanen, Styan & Isotalo (2011, Prop. 19.3):

Olkoon $\mathrm{r}(\mathbf{A}_{n\times}) = r$ *ja* $\mathrm{r}(\mathbf{B}_{n\times m}) = k, k < r$. *Silloin*

$$\begin{aligned}(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})' &\geq_{\mathsf{L}} (\mathbf{A}-\mathbf{A}\mathbf{P}_{\mathbf{B}'})(\mathbf{A}-\mathbf{A}\mathbf{P}_{\mathbf{B}'})' \\ &= \mathbf{A}(\mathbf{I}_m - \mathbf{P}_{\mathbf{B}'})\mathbf{A}', \end{aligned} \tag{6.106}$$

ja täten

$$\|\mathbf{A}-\mathbf{B}\|_F^2 \geq \|\mathbf{A}-\mathbf{A}\mathbf{P}_{\mathbf{B}'}\|_F^2 \,. \tag{6.107}$$

Jos matriisilla $\mathbf{B}$ *on täysiastehajotelma* $\mathbf{B} = \mathbf{F}\mathbf{G}'$, *missä* $\mathbf{G}'\mathbf{G} = \mathbf{I}_k$, *niin* $\mathbf{P}_{\mathbf{B}'} = \mathbf{P}_{\mathbf{G}}$, *ja*

$$(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})' \leq \mathbf{A}(\mathbf{I}_m - \mathbf{G}\mathbf{G}')\mathbf{A}'. \tag{6.108}$$

Eckart & Young (1936) laativat tärkeän artikkelin matriisin $\mathbf{A}$ approksimoinnissa alempiasteisen matriisin $\mathbf{B}$ avulla ks. myös Householder & Young (1938). Tätä tulosta sanotaan usein *Eckartin ja Youngin lauseeksi.*

Eckart & Young: *Olkoon matriisin* $\mathbf{A}_{n\times m}, \mathrm{r}(\mathbf{A}) = r$, *singulaariarvohajotelma*

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}' = \mathbf{U}_1\boldsymbol{\Delta}_1\mathbf{V}_1' = \delta_1\mathbf{u}_1\mathbf{v}_1' + \cdots + \delta_r\mathbf{u}_r\mathbf{v}_r' \,. \tag{6.109}$$

Olkoon $\mathbf{B}$ $n\times m$*-matriisi, jonka aste on* k $(< r)$. *Then*

$$\min_{\mathbf{B}} \|\mathbf{A}-\mathbf{B}\|_F^2 = \delta_{k+1}^2 + \cdots + \delta_r^2 \,, \tag{6.110}$$

ja minimi saavutetaan kun

$$\mathbf{B} = \dot{\mathbf{B}}_k = \delta_1\mathbf{u}_1\mathbf{v}_1' + \cdots + \delta_k\mathbf{u}_k\mathbf{v}_k' \,. \tag{6.111}$$

Todistus: ks. myös Puntanen, Styan & Isotalo (2011, Prop. 19.4).

Stewart (1993, s. 563) korostaa, että itse asiassa Eckart & Young (1936) löysivät uudelleen tuloksen, jona aiemmin oli esittänyt yleisemmässä muodossa Schmidt (1907); ks. myös Hubert, Meulman & Heiser (2000, s. 71). Nimityksen ”Eckart–Young theorem” sijasta Stewart (1993) käyttää nimitystä ”Schmidt’s approximation theorem” ja samoin tekevät Ben-Israel & Greville (2003, s. 213), jotka käyttävät myös termiä *paras k-asteinen approksimaatio*.

Kirjoitetaan vielä Eckartin & Youngin tulos symmetrisen $\mathbf{A}$:n tapauksessa.

Eckart & Young: symmetrinen $\mathbf{A}$. *Olkoon $\mathbf{A}_{n\times n}$ symmetrinen matriisi, jonka aste on r ja jonka ominaisarvohajotelma on*

$$\mathbf{A} = \mathbf{T}\mathbf{\Delta}\mathbf{T}' = \lambda_1 \mathbf{t}_1\mathbf{t}_1' + \cdots + \lambda_r \mathbf{t}_r\mathbf{t}_r' \,. \tag{6.112}$$

Silloin

$$\min_{\mathrm{r}(\mathbf{B})=k} \|\mathbf{A}-\mathbf{B}\|_F^2 = \min_{\mathrm{r}(\mathbf{B})=k} \mathrm{tr}(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})' = \lambda_{k+1}^2 + \cdots + \lambda_r^2 \,, \tag{6.113}$$

ja minimi saavutetaan kun

$$\mathbf{B} = \mathbf{T}_{(k)}\mathbf{\Lambda}_{(k)}\mathbf{T}_{(k)}' = \lambda_1 \mathbf{t}_1\mathbf{t}_1' + \cdots + \lambda_k \mathbf{t}_k\mathbf{t}_k' \,. \tag{6.114}$$

6.5.3 Kaksi matriisiepäyhtälöä

Olkoon $\mathbf{X}$ $n\times p$-matriisi, jonka sarakkeet ovat vapaat ja $\mathbf{V}$ olkoon symmetrinen positiivisesti definiitti $n \times n$-matriisi. Todista seuraavat matriisiepäyhtälöt:

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}' \leq_{\mathsf{L}} \mathbf{V}, \tag{6.115a}$$

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \leq_{\mathsf{L}} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{6.115b}$$

Ratkaisu: Olkoon $\mathbf{V}^{1/2}$ matriisin $\mathbf{V}$ positiivisesti definiitti neliöjuuri. Tällöin (6.115a) seuraa siitä, että

$$\begin{aligned}
\mathbf{G} &:= \mathbf{V} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}' \\
&= \mathbf{V}^{1/2}\left[\mathbf{I} - \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{V}^{-1/2}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1/2}\right]\mathbf{V}^{1/2} \\
&= \mathbf{V}^{1/2}\left(\mathbf{I} - \mathbf{P}_{\mathbf{V}^{-1/2}\mathbf{X}}\right)\mathbf{V}^{1/2} \\
&:= \mathbf{F}'\mathbf{F} \geq_{\mathsf{L}} \mathbf{0}\,,
\end{aligned} \tag{6.116}$$

missä $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ on ortogonaaliprojektori $\mathscr{C}(\mathbf{A})$:lle ja

$$\mathbf{F} = (\mathbf{I} - \mathbf{P}_{\mathbf{V}^{-1/2}\mathbf{X}})\mathbf{V}^{1/2}. \tag{6.117}$$

Epäyhtälö (6.115b) saadaan kertomalla (6.116) vasemmalta $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$:lla ja oikealta $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$:lla.

Yhtäsuuruus (6.115b):ssä toteutuu täsmälleen silloin kun

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{F}'\mathbf{F}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}\,, \tag{6.118}$$

mikä on yhtäpitävää sen kanssa että

$$\mathbf{F}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \left(\mathbf{I} - \mathbf{P}_{\mathbf{V}^{-1/2}\mathbf{X}}\right)\mathbf{V}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}\,, \tag{6.119}$$

ts. $\mathbf{V}^{1/2}\mathbf{X} = \mathbf{P}_{\mathbf{V}^{-1/2}\mathbf{X}}\mathbf{V}^{1/2}\mathbf{X}$, eli

$$\begin{aligned}\mathbf{V}^{1/2}\mathbf{X} &= \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{V}^{1/2}\mathbf{X} \\ &= \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}. \qquad (6.120)\end{aligned}$$

Kertomalla (6.120) vasemmalta matriisilla $\mathbf{V}^{-1/2}$ ja oikealta $(\mathbf{X}'\mathbf{X})^{-1}$:llä (ja transponoimalla puolittain) saadaan tulos

$$\boxed{\begin{aligned}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \qquad (6.121)\\ &\iff \\ (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \qquad (6.122)\end{aligned}}$$

Olisimme voineet todistaa (6.115a):n ja (6.115b):n käyttämällä esimerkin 4.9 tulosta (4.144) (s. 154), jonka mukaan

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - \mathbf{D}, \qquad (4.144)$$

missä

$$\mathbf{D} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

ja $(\mathbf{X} : \mathbf{Z})$ on epäsingulaarinen $n \times n$-matriisi ja $\mathbf{X}'\mathbf{Z} = \mathbf{0}$. On selvää, että $\mathbf{D}$ on ei-negatiivisesti definiitti ja lisäksi $\mathbf{D} = \mathbf{0}$ täsmälleen silloin kun

$$\mathbf{X}'\mathbf{V}\mathbf{Z} = \mathbf{0}. \qquad (6.123)$$

Täten kaikki kolme ehtoa (6.121), (6.122) ja (6.123) ovat yhtäpitäviä.

Saamme tulkinnan (6.115b):lle lineaaristen mallien yhteydestä. Olkoon $\mathbf{y}$ satunnaisvektori, jonka odotusarvo on $\mathbf{X}\boldsymbol{\beta}$, ja kovarianssimatriisi $\mathbf{V}$, missä $\mathrm{r}(\mathbf{X}_{n\times p}) = p$ ja $\mathbf{V}$ on positiivisesti definiitti. Merkitään

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathrm{OLSE}(\boldsymbol{\beta}), \qquad (6.124a)$$

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathrm{BLUE}(\boldsymbol{\beta}). \qquad (6.124b)$$

On helppo nähdä että $\hat{\boldsymbol{\beta}}$ ja $\tilde{\boldsymbol{\beta}}$ ovat $\boldsymbol{\beta}$:n harhattomia estimaattoreita eli $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ ja $\mathrm{E}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. Lisäksi

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \quad \mathrm{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \qquad (6.125)$$

Täten (6.115b) merkitsee, että

$$\mathrm{cov}(\tilde{\boldsymbol{\beta}}) \leq_{\mathsf{L}} \mathrm{cov}(\hat{\boldsymbol{\beta}}), \qquad (6.126)$$

josta seuraa mm., että

$$\mathrm{var}(\tilde{\beta}_i) \leq \mathrm{var}(\hat{\beta}_i), \quad i = 1, \dots, p. \qquad (6.127)$$

Estimaattorin (6.124b) lausekkeen (paras lineaarinen harhaton estimaattori) esitti ensimmäisenä ilmeisesti Aitken (1935, 1939). Yhtälön (6.123):n mukaan saamme seuraavan ehdon OLSE($\boldsymbol{\beta}$):n ja BLUE($\boldsymbol{\beta}$):n yhtäsuuruudelle:

$$\mathrm{OLSE}(\boldsymbol{\beta}) = \mathrm{BLUE}(\boldsymbol{\beta}) \iff \mathbf{X}'\mathbf{V}\mathbf{Z} = \mathbf{0}. \tag{6.128}$$

Ks. esim. Puntanen & Styan (1989).

6.6 BLUE:n määrittely

Määritellään nyt vielä BLUE, best linear unbiased estimator, täsmällisesti, kun kyseessä on lineaarinen malli

$$\mathscr{M} = \{\mathbf{y}, \mathrm{E}(\mathbf{y}), \mathrm{cov}(\mathbf{y})\} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}\,, \tag{6.129}$$

eli

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \quad \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}\,, \quad \mathrm{cov}(\mathbf{y}) = \mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}, \tag{6.130}$$

missä $\mathbf{X} \in \mathbb{R}^{n\times p}$, $\boldsymbol{\beta} \in R^p$. Sitä varten on ensin määriteltävä mitä lineaarisessa mallissa tarkoitetaan *estimoituvalla* parametrifunktiolla. Olkoon $\mathbf{K} \in \mathbb{R}^{q\times p}$. Silloin parametrifunktio $\mathbf{K}\boldsymbol{\beta}$ on *estimoituva*, jos on olemassa $\mathbf{F}$ siten että $\mathbf{F}\mathbf{y}$ on $\mathbf{K}\boldsymbol{\beta}$:n harhaton estimaattori.

On helppo osoittaa, että $\mathbf{K}\boldsymbol{\beta}$ on estimoituva jos ja vain jos $\mathscr{C}(\mathbf{K}') \subset \mathscr{C}(\mathbf{X}')$, ja että $\mathbf{X}\boldsymbol{\beta}$ on aina estimoituva.

(a) Määritelmä 1: Olkoon $\mathbf{k}'\boldsymbol{\beta}$ estimoituva. Silloin $\mathbf{g}'\mathbf{y}$ on BLUE($\mathbf{k}'\boldsymbol{\beta}$) mallissa $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ jos $\mathbf{g}'\mathbf{y}$ on $\mathbf{k}'\boldsymbol{\beta}$:n harhaton estimaattori ja sillä on minimivarianssi kaikkien lineaaristen harhattomien estimaattorien joukossa:

$$\mathrm{E}(\mathbf{g}'\mathbf{y}) = \mathbf{k}'\boldsymbol{\beta} \ \text{ ja } \ \mathrm{var}(\mathbf{g}'\mathbf{y}) \le \mathrm{var}(\mathbf{f}'\mathbf{y}) \text{ kaikilla } \mathbf{f}:\ \mathrm{E}(\mathbf{f}'\mathbf{y}) = \mathbf{k}'\boldsymbol{\beta}.$$

(b) Määritelmä 2: Olkoon $\mathbf{K}\boldsymbol{\beta}$ estimoituva. Silloin $\mathbf{G}\mathbf{y}$ on BLUE($\mathbf{K}\boldsymbol{\beta}$) mallissa $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ jos

$$\mathrm{E}(\mathbf{G}\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \ \text{ ja } \ \mathrm{cov}(\mathbf{G}\mathbf{y}) \le_{\mathsf{L}} \mathrm{cov}(\mathbf{F}\mathbf{y}) \text{ kaikilla } \mathbf{F}:\ \mathrm{E}(\mathbf{F}\mathbf{y}) = \mathbf{K}\boldsymbol{\beta}.$$

Seuraavaa tulosta kutsutaan usein nimellä *Fundamental* BLUE *Equation*:

$$\mathbf{G}\mathbf{y} = \mathrm{BLUE}(\mathbf{K}\boldsymbol{\beta}) \text{ mallissa } \mathscr{M} \iff \mathbf{G}(\mathbf{X} : \mathbf{V}\mathbf{M}) = (\mathbf{K} : \mathbf{0})\,, \tag{6.131}$$

Lineaaristen mallien keskeistä kalustoa on myös *Gaussin–Markovin lause*:

$$\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) \text{ under } \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\}\,. \tag{6.132}$$

Todistetaan sitten (6.132). Tarkastellaan mallia $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\}$, jossa oletamme että $\sigma^2 = 1$. Silloin

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^{+}\mathbf{y}\,, \quad \operatorname{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}. \tag{6.133}$$

Olkoon $\mathbf{By}$ jokin toinen $\boldsymbol{\beta}$:n lineaarinen harhaton estimaattori, ts. $\mathbf{B}$ toteuttaa yhtälön $\mathbf{BX} = \mathbf{I}_p$. Tällöin

$$\begin{aligned}
\operatorname{cov}(\mathbf{By} - \hat{\boldsymbol{\beta}}) &= \operatorname{cov}(\mathbf{By}) + \operatorname{cov}(\hat{\boldsymbol{\beta}}) - \operatorname{cov}(\mathbf{By}, \hat{\boldsymbol{\beta}}) - \operatorname{cov}(\hat{\boldsymbol{\beta}}, \mathbf{By}) \\
&= \mathbf{BB}' + (\mathbf{X}'\mathbf{X})^{-1} - \mathbf{B}(\mathbf{X}^{+})' - \mathbf{X}^{+}\mathbf{B}' \\
&= \mathbf{BB}' + (\mathbf{X}'\mathbf{X})^{-1} - \mathbf{BX}(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}' \\
&= \mathbf{BB}' - (\mathbf{X}'\mathbf{X})^{-1},
\end{aligned} \tag{6.134}$$

mistä seuraa että

$$\mathbf{BB}' - (\mathbf{X}'\mathbf{X})^{-1} = \operatorname{cov}(\mathbf{By}) - \operatorname{cov}(\hat{\boldsymbol{\beta}}) = \operatorname{cov}(\mathbf{By} - \hat{\boldsymbol{\beta}}) \geq_{\mathsf{L}} \mathbf{0}\,, \tag{6.135}$$

missä viimeinen epäyhtälö seuraa siitä, että jokainen kovarianssimatriisi on ei-negatiivisesti definiitti. Nyt (6.135):sta seuraa että

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) \leq_{\mathsf{L}} \operatorname{cov}(\mathbf{By}) \quad \text{kaikilla } \mathbf{B}\text{: } \mathbf{BX} = \mathbf{I}_p\,. \tag{6.136}$$

6.7 Cauchyn–Schwarzin epäyhtälö

Cauchyn–Schwarzin epäyhtälö,

$$\boxed{(\mathbf{x}'\mathbf{y})^2 \leq \mathbf{x}'\mathbf{x} \cdot \mathbf{y}'\mathbf{y}\,, \tag{CS}}$$

eli

$$\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \|\mathbf{x}\|^2 \cdot \|\mathbf{y}\|^2,$$

on monissa tarkasteluissa erittäin hyödyllinen. On monta tapaa todistaa (CS), ja erityisesti aiemmin johdettu projektorin lauseke antaa yhden kätevän tavan. Olkoon $\mathbf{Q}_{\mathbf{x}} = \mathbf{I} - \mathbf{P}_{\mathbf{x}}$. Koska

$$\|\mathbf{Q}_{\mathbf{x}}\mathbf{y}\|^2 = \mathbf{y}'\mathbf{Q}_{\mathbf{x}}\mathbf{y} \geq 0\,, \tag{6.137}$$

on tietenkin voimassa (olettaen että $\mathbf{x} \neq \mathbf{0}$)

$$\mathbf{y}'\mathbf{Q}_{\mathbf{x}}\mathbf{y} = \mathbf{y}'[\mathbf{I} - \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}']\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} \geq 0 \tag{6.138}$$

eli toisin kirjoitettuna

$$\mathbf{y}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} \leq \mathbf{y}'\mathbf{y}, \tag{6.139a}$$

$$\frac{\mathbf{y}'\mathbf{x} \cdot \mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} \leq \mathbf{y}'\mathbf{y}, \tag{6.139b}$$

mistä (CS) seuraakin. Jos $\mathbf{x} = \mathbf{0}$ tai $\mathbf{y} = \mathbf{0}$, niin (CS) on ilman mutta voimassa. Havaitsemme (6.137):sta, että yhtäsuuruus (CS):ssä on voimassa jos ja vain jos

$$\mathbf{Q_x y} = (\mathbf{I} - \mathbf{P_x})\mathbf{y} = \mathbf{0}\,, \tag{6.140}$$

mikä toteutuu täsmälleen silloin kun $\mathbf{y} \in \mathscr{C}(\mathbf{x})$ ts. on olemassa sellainen reaaliluku λ, että

$$\mathbf{y} = \lambda\mathbf{x}\,. \tag{6.141}$$

Yhtäsuuruus (CS):ssä pätee siis täsmälleen silloin kun joukko $\{\mathbf{x}, \mathbf{y}\}$ on sidottu.

Voimme soveltaa Cauchyn–Schwarzin epäyhtälöä seuraavassa arvioinnissa:

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|^2 &= (\mathbf{x}+\mathbf{y})'(\mathbf{x}+\mathbf{y}) \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\mathbf{x}'\mathbf{y} \\ &\leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2|\mathbf{x}'\mathbf{y}| \\ &\leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2,\end{aligned} \tag{6.142}$$

joten olemme päätyneet *kolmioepäyhtälöön*

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|\,, \tag{6.143}$$

missä yhtäsuuruus on voimassa täsmälleen silloin kun $\mathbf{x}$ ja $\mathbf{y}$ ovat kohtisuorassa: $\mathbf{x}'\mathbf{y} = 0$.

6.7.1 Cauchyn–Schwarzin epäyhtälön versioita

Cauchyn–Schwarzin epäyhtälöstä saadaan eri versioita valitsemalla vektorit $\mathbf{x}$ ja $\mathbf{y}$ sopivasti. Esimerkiksi:

- $\mathbf{x} = \mathbf{Lu}$, $\mathbf{y} = \mathbf{Lv}$:

$$\boxed{(\mathbf{u}'\mathbf{L}'\mathbf{Lv})^2 \leq (\mathbf{u}'\mathbf{L}'\mathbf{Lu})(\mathbf{v}'\mathbf{L}'\mathbf{Lv})\,, \qquad \text{(CS1)}}$$

  yhtäsuuruus voimassa joss $\mathbf{Lu} = \lambda\mathbf{Lv}$ jollakin λ:n arvolla.

- $\mathbf{x} = \mathbf{A}^{1/2}\mathbf{u}$, $\mathbf{y} = \mathbf{A}^{1/2}\mathbf{v}$, $\mathbf{A}$ on ei-negatiivisesti definiitti:

$$\boxed{(\mathbf{u}'\mathbf{Av})^2 \leq (\mathbf{u}'\mathbf{Au})(\mathbf{v}'\mathbf{Av})\,, \qquad \text{(CS2)}}$$

  yhtäsuuruus voimassa joss $\mathbf{Au} = \lambda\mathbf{Av}$ jollakin λ:n arvolla.

- $\mathbf{x} = \mathbf{A}^{1/2}\mathbf{u}$, $\mathbf{y} = \mathbf{A}^{-1/2}\mathbf{v}$, $\mathbf{A}$ on positiivisesti definiitti:

$$(\mathbf{u}'\mathbf{A}^{1/2}\mathbf{A}^{-1/2}\mathbf{v})^2 \leq (\mathbf{u}'\mathbf{A}^{1/2}\mathbf{A}^{1/2}\mathbf{u})(\mathbf{v}'\mathbf{A}^{-1/2}\mathbf{A}^{-1/2}\mathbf{v}) \tag{CS3}$$

  eli

$$\boxed{(\mathbf{u}'\mathbf{v})^2 \leq (\mathbf{u}'\mathbf{A}\mathbf{u})(\mathbf{v}'\mathbf{A}^{-1}\mathbf{v})\,, \qquad \text{(CS3)}}$$

  yhtäsuuruus voimassa joss $\mathbf{A}\mathbf{u} = \lambda\mathbf{v}$ jollakin λ:n arvolla.

- $\mathbf{x} = \mathbf{A}^{1/2}\mathbf{t}$, $\mathbf{y} = \mathbf{A}^{-1/2}\mathbf{t}$, $\mathbf{A}$ on positiivisesti definiitti, $\mathbf{t}'\mathbf{t} = 1$:

$$\boxed{(\mathbf{t}'\mathbf{A}\mathbf{t})^{-1} \leq \mathbf{t}'\mathbf{A}^{-1}\mathbf{t}\,. \qquad \text{(CS4)}}$$

6.8 Permutaatiomatriisi

Matriisin $\mathbf{A}_{4\times 4}$ sarakkeitten järjestystä voidaan vaihtaa vektoreiden $\mathbf{i}_j$ avulla esim. seuraavasti:

$$\begin{aligned}\mathbf{A}\mathbf{E}_{23} &= (\mathbf{a}_1 : \mathbf{a}_2 : \mathbf{a}_3 : \mathbf{a}_4)(\mathbf{i}_1 : \mathbf{i}_3 : \mathbf{i}_2 : \mathbf{i}_4) \\ &= (\mathbf{A}\mathbf{i}_1 : \mathbf{A}\mathbf{i}_3 : \mathbf{A}\mathbf{i}_2 : \mathbf{A}\mathbf{i}_4) \\ &= (\mathbf{a}_1 : \mathbf{a}_3 : \mathbf{a}_2 : \mathbf{a}_4),\end{aligned} \tag{6.144}$$

missä siis

$$\mathbf{E}_{23} = (\mathbf{i}_1 : \mathbf{i}_3 : \mathbf{i}_2 : \mathbf{i}_4) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{6.145}$$

Matriisi $\mathbf{E}_{23}$, joka saadaan vaihtamalla $\mathbf{I}$:n 2. ja 3. sarake keskenään, on ns. *alkeispermutaatiomatriisi*. Kertomalla $\mathbf{A}\mathbf{E}_{23}$ edelleen oikealta matriisilla

$$\mathbf{E}_{14} = (\mathbf{i}_4 : \mathbf{i}_2 : \mathbf{i}_3 : \mathbf{i}_1) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{6.146}$$

saadaan tulokseksi

$$\mathbf{A}\mathbf{E}_{23}\mathbf{E}_{14} = (\mathbf{a}_4 : \mathbf{a}_3 : \mathbf{a}_2 : \mathbf{a}_1). \tag{6.147}$$

Matriisi

$$\mathbf{Q} = \mathbf{E}_{23}\mathbf{E}_{14} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} = (\mathbf{i}_4 : \mathbf{i}_3 : \mathbf{i}_2 : \mathbf{i}_1) \tag{6.148}$$

on esimerkki ns. *permutaatiomatriisista*. Yleisesti voimme määritellä että permutaatiomatriisi $\mathbf{Q}$ on yksikkömatriisin $\mathbf{I}$ sellainen versio, jossa sarakkeiden järjestystä on vaihdettu; alkeispermutaatiomatriisissa $\mathbf{E}_{ij}$ on vaihdettu vain kahden sarakkeen paikkaa. Matriisi $\mathbf{Q}$ voidaan aina esittää alkeispermutaatiomatriisien $\mathbf{E}_{ij}$ tulona. Alkeispermutaatiomatriisit $\mathbf{E}_{ij}$ ovat aina symmetrisiä mutta permutaatiomatriisi $\mathbf{Q}$ ei välttämättä ole symmetrinen. Sen sijaan sekä $\mathbf{Q}$ että $\mathbf{E}_{ij}$ ovat ortogonaalisia:

$$\mathbf{Q}'\mathbf{Q} = \mathbf{I}, \quad \mathbf{E}'_{ij}\mathbf{E}_{ij} = \mathbf{E}_{ij}\mathbf{E}_{ij} = \mathbf{I}. \tag{6.149}$$

6.9 Kroneckerin tulo

Matriisien $\mathbf{A}_{n\times m}$ and $\mathbf{B}_{p\times q}$ Kroneckerin tulo määritellään seuraavasti:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \dots & a_{nm}\mathbf{B} \end{pmatrix} \in \mathbb{R}^{np\times mq}. \tag{6.150}$$

Jos esimerkiksi

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}, \tag{6.151}$$

niin

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1\begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} & 2\begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} & 3\begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} \\ 4\begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} & 5\begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} & 6\begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} \end{pmatrix}$$

ja

$$\mathbf{B} \otimes \mathbf{A} = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} \otimes \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 0\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} & 1\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \\ 2\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} & 3\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \end{pmatrix}.$$

Tässä tilanteessa siis $\mathbf{A}\otimes\mathbf{B} \neq \mathbf{B}\otimes\mathbf{A}$. Itse asiassa voimme havaita, että matriiseissa $\mathbf{A}\otimes\mathbf{B}$ ja $\mathbf{B}\otimes\mathbf{A}$ on samat elementit ja rivejä ja sarakkeita permutoimalla ne saadaan samoiksi.

Tulos

$$\mathrm{vec}(\mathbf{A}_{n\times p}\mathbf{B}_{p\times q}) = (\mathbf{I}_q \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B}), \tag{6.152}$$

missä $\mathrm{vec}(\mathbf{C}_{n\times m}) = \mathrm{vec}(\mathbf{c}_1 : \mathbf{c}_2 : \ldots : \mathbf{c}_m) = \begin{pmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_m \end{pmatrix} \in \mathbb{R}^{nm}$, voidaan päätellä seuraavasti:

$$\mathrm{vec}(\mathbf{A}_{n\times p}\mathbf{B}_{p\times q}) = \mathrm{vec}[\mathbf{A}(\mathbf{b}_1 : \mathbf{b}_2 : \ldots \mathbf{b}_q)] = \begin{pmatrix} \mathbf{A}\mathbf{b}_1 \\ \vdots \\ \mathbf{A}\mathbf{b}_q \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{A} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_q \end{pmatrix} = (\mathbf{I}_q \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})\,. \tag{6.153}$$

Seuraavassa on luetteloitu joitakin Kroneckerin tulon ominaisuuksia:

KR1. $\mathbf{a}' \otimes \mathbf{b} = \mathbf{b}\mathbf{a}' = \mathbf{b} \otimes \mathbf{a}'$,

KR2. $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$,

KR3. $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$, jos $\mathbf{A}$ ja $\mathbf{B}$ ovat kääntyviä,

KR4. $(\mathbf{A} \otimes \mathbf{B})^{-} = \mathbf{A}^{-} \otimes \mathbf{B}^{-}$,

KR5. $\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}$, yleensä,

KR6. $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D}$,

KR7. $\mathrm{vec}(\mathbf{A}_{n\times p}\mathbf{B}_{p\times q}) = (\mathbf{B}' \otimes \mathbf{I}_n)\,\mathrm{vec}(\mathbf{A}) = (\mathbf{I}_q \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})$,

KR8. $\mathrm{vec}(\mathbf{a}\mathbf{b}') = \mathbf{b} \otimes \mathbf{a}$,

KR9. $\mathrm{vec}(\mathbf{A}\mathbf{B}\mathbf{C}) = (\mathbf{C}' \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})$,

KR10. $\mathbf{P}_{\mathbf{A}\otimes\mathbf{B}} = \mathbf{P}_{\mathbf{A}} \otimes \mathbf{P}_{\mathbf{B}}$.

Lähteitä Kroneckeriin tuloon: Henderson & Searle (1979, 1981b), Harville (1997, Ch. 16), Abadir & Magnus (2005, ss. 273–284), Seber (2008, §11.1).

Esimerkki 6.8. Olkoon $\mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)}) = (\mathbf{u}_1 : \mathbf{u}_2 : \ldots : \mathbf{u}_p)'$ satunnaisotos p-ulotteisesta jakaumasta, jonka odotusarvo on $\boldsymbol{\mu}$ ja kovarianssimatriisi on $\boldsymbol{\Sigma}$. Vakuuttaudu (ks. myös s. 53), että

(a) $\mathrm{cov}(\mathbf{u}_i, \mathbf{u}_j) = \sigma_{ij}\mathbf{I}_n$,

(b) $\mathrm{cov}(\mathbf{u}_{(i)}, \mathbf{u}_{(j)}) = \mathbf{0}_{p\times p},\ i \neq j$,

(c) $\operatorname{cov}(\mathbf{u}_{(i)}) = \mathbf{\Sigma}$,

(d) $\operatorname{cov}(\mathbf{u}_i) = \sigma_i^2 \mathbf{I}_n$,

(e) $\operatorname{cov}[\operatorname{vec}(\mathbf{u}_i : \mathbf{u}_j)] = \operatorname{cov}\begin{pmatrix}\mathbf{u}_i \\ \mathbf{u}_j\end{pmatrix} = \begin{pmatrix}\sigma_i^2\mathbf{I}_n & \sigma_{ij}\mathbf{I}_n \\ \sigma_{ji}\mathbf{I}_n & \sigma_j^2\mathbf{I}_n\end{pmatrix} = \begin{pmatrix}\sigma_i^2 & \sigma_{ij} \\ \sigma_{ji} & \sigma_j^2\end{pmatrix} \otimes \mathbf{I}_n$,

(f) $\operatorname{cov}\big[\operatorname{vec}\big(\mathbf{u}_{(i)} : \mathbf{u}_{(j)}\big)\big] = \operatorname{cov}\begin{pmatrix}\mathbf{u}_{(i)} \\ \mathbf{u}_{(j)}\end{pmatrix} = \begin{pmatrix}\mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}\end{pmatrix} = \begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix} \otimes \mathbf{\Sigma}$,

(g) $\operatorname{cov}[\operatorname{vec}(\mathbf{U})] = \begin{pmatrix}\sigma_1^2\mathbf{I}_n & \sigma_{12}\mathbf{I}_n & \dots & \sigma_{1p}\mathbf{I}_n \\ \sigma_{21}\mathbf{I}_n & \sigma_2^2\mathbf{I}_n & \dots & \sigma_{2p}\mathbf{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1}\mathbf{I}_n & \sigma_{p2}\mathbf{I}_n & \dots & \sigma_p^2\mathbf{I}_n\end{pmatrix} = \mathbf{\Sigma} \otimes \mathbf{I}_n \in \mathbb{R}^{pn \times pn}$.

□

Harjoitustehtäviä

6.1 (Ortogonaalinen rotaatio). Merkitään

$$\mathbf{A}_\theta = \begin{pmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix}, \qquad \mathbf{B}_\theta = \begin{pmatrix}\cos\theta & \sin\theta \\ \sin\theta & -\cos\theta\end{pmatrix}.$$

Osoita että mielivaltainen 2×2 ortogonaalinen matriisi $\mathbf{Q}$ on esitettävisää muodossa $\mathbf{A}_\theta$ tai $\mathbf{B}_\theta$ jollakin θ:n arvolla. Muunnos $\mathbf{A}_\theta\mathbf{u}_{(i)}$ rotatoi havaintoa $\mathbf{u}_{(i)}$ kulman θ verran vastapäivään ja $\mathbf{B}_\theta\mathbf{u}_{(i)}$ peilaa the havainnon $\mathbf{u}_{(i)}$ suoran $y = \tan\left(\frac{\theta}{2}\right)x$ suhteen. Ks. myös kuvio 8.12 (s. 283).

Horn & Johnson (1990, p. 68).

Luku 7

Satunnaisvektoreista

Vuorotellen he huomauttivat kukin jotakin runonsanankäänteistä, sen rakenteesta, sen persialaisista runoista saamista vaikutteista ja niin edespäin, kunnes heidän taidokkaasti peräkkäin sovittamansa sanat hitaasti kietoivat mitäänsanomattoman runon kuin jalokivikukkiin.

Mika Waltari (1949): *Mikael Hakim.*

7.1 Satunnaisvektorin kovarianssimatriisi

Tarkastellaan p-ulotteista satunnaisvektoria $\mathbf{z}$, jonka odotusarvovektori ja kovarianssimatriisi ovat

$$\mathrm{E}(\mathbf{z}) = \begin{pmatrix} \mathrm{E}(z_1) \\ \mathrm{E}(z_2) \\ \vdots \\ \mathrm{E}(z_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \boldsymbol{\mu}, \ \operatorname{cov}(\mathbf{z}) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{pmatrix} = \boldsymbol{\Sigma}\,. \tag{7.1}$$

Voimme merkitä lyhyesti $\mathbf{z} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Tällöin siis:

$$\operatorname{var}(z_i) = \mathrm{E}(z_i - \mu_i)^2 = \mathrm{E}(z_i^2) - \mu_i^2 = \sigma_i^2, \ \mu_i = \mathrm{E}(z_i)\,, \tag{7.2a}$$

$$\operatorname{cov}(z_i, z_j) = \mathrm{E}(z_i - \mu_i)(z_j - \mu_j) = \mathrm{E}(z_i z_j) - \mu_i \mu_j = \sigma_{ij} = \varrho_{ij}\sigma_i\sigma_j\,, \tag{7.2b}$$

$$\operatorname{cor}(z_i, z_j) = \frac{\operatorname{cov}(z_i, z_j)}{\sqrt{\operatorname{var}(z_i)\operatorname{var}(z_j)}} = \frac{\sigma_{ij}}{\sigma_i\sigma_j} = \varrho_{ij}\,, \tag{7.2c}$$

missä $i, j = 1, \ldots, p$ ja $\mathrm{E}(y)$ tarkoittaa satunnaismuuttujan y odotusarvoa. Jos muodostamme $p \times p$-matriisin

$$
\begin{aligned}
(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})' &= \begin{pmatrix} z_1-\mu_1 \\ \vdots \\ z_p-\mu_p \end{pmatrix} (z_1-\mu_1, \ldots, z_p-\mu_p) \\
&= \begin{pmatrix} (z_1-\mu_1)^2 & (z_1-\mu_1)(z_2-\mu_2) & \ldots & (z_1-\mu_1)(z_p-\mu_p) \\ (z_2-\mu_2)(z_1-\mu_1) & (z_2-\mu_2)^2 & \ldots & (z_2-\mu_1)(z_p-\mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (z_p-\mu_p)(z_1-\mu_1) & (z_p-\mu_p)(z_2-\mu_2) & \ldots & (z_p-\mu_p)^2 \end{pmatrix},
\end{aligned} \tag{7.3}
$$

niin odotusarvo tästä satunnaismatriisista on tietenkin $\mathbf{z}$:n kovarianssimatriisi (ks. myös s. 40):

$$
\boxed{\operatorname{cov}(\mathbf{z}) = \mathrm{E}(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})'.} \tag{7.4}
$$

Koska

$$
\begin{aligned}
\mathrm{E}(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})' &= \mathrm{E}(\mathbf{z}\mathbf{z}' - \mathbf{z}\boldsymbol{\mu}' - \boldsymbol{\mu}\mathbf{z}' - \boldsymbol{\mu}\boldsymbol{\mu}') \\
&= \mathrm{E}(\mathbf{z}\mathbf{z}') - \mathrm{E}(\mathbf{z})\boldsymbol{\mu}' - \boldsymbol{\mu}\,\mathrm{E}(\mathbf{z}') + \boldsymbol{\mu}\boldsymbol{\mu}' \\
&= \mathrm{E}(\mathbf{z}\mathbf{z}') - \boldsymbol{\mu}\boldsymbol{\mu}' - \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}',
\end{aligned} \tag{7.5}
$$

saamme kovarianssimatriisille $\boldsymbol{\Sigma}$ lausekkeen

$$
\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{z}) = \mathrm{E}(\mathbf{z}\mathbf{z}') - \boldsymbol{\mu}\boldsymbol{\mu}'. \tag{7.6}
$$

On vaivatonta osoittaa seuraava erittäin tärkeä tulos:

$$
\boxed{\mathrm{E}(\mathbf{z}) = \boldsymbol{\mu},\ \operatorname{cov}(\mathbf{z}) = \boldsymbol{\Sigma} \implies \mathrm{E}(\mathbf{A}\mathbf{z}) = \mathbf{A}\boldsymbol{\mu},\ \operatorname{cov}(\mathbf{A}\mathbf{z}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'.} \tag{7.7}
$$

Nimittäin ensinnäkin $\mathbf{Az}$:n odotusarvo on $\mathrm{E}(\mathbf{Az}) = \mathbf{A}\,\mathrm{E}(\mathbf{z}) = \mathbf{A}\boldsymbol{\mu}$, joten $\mathbf{Az}$:n kovarianssimatriisiksi saadaan todellakin lauseke

$$
\begin{aligned}
\operatorname{cov}(\mathbf{Az}) &= \mathrm{E}[(\mathbf{Az} - \mathbf{A}\boldsymbol{\mu})(\mathbf{Az} - \mathbf{A}\boldsymbol{\mu})'] \\
&= \mathrm{E}[\mathbf{A}(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})'\mathbf{A}'] \\
&= \mathbf{A}\,\mathrm{E}\big[(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})'\big]\,\mathbf{A}' \\
&= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'.
\end{aligned} \tag{7.8}
$$

7.1.1 Lineaarikombinaation varianssi

Erikoistapauksena (7.7):sta saamme $\mathbf{z}$:n elementtien lineaarikombinaation $\mathbf{a}'\mathbf{z}$ varianssiksi

$$\mathrm{var}(\mathbf{a}'\mathbf{z}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}\,. \tag{7.9}$$

Jos esimerkiksi $p = 2$, niin

$$\begin{aligned} \mathrm{var}(\mathbf{1}_2'\mathbf{z}) = \mathrm{var}(z_1 + z_2) &= \mathbf{1}_2'\boldsymbol{\Sigma}\mathbf{1}_2 \\ &= \mathrm{var}(z_1) + \mathrm{var}(z_2) + 2 \cdot \mathrm{cov}(z_1, z_2) \\ &= \sigma_1^2 + \sigma_2^2 + 2\sigma_{12}. \end{aligned} \tag{7.10}$$

Täten:

♠ satunnaismuuttujien z_1 ja z_2 summan varianssi on varianssien summa täsmälleen silloin kun z_1 ja z_2 ovat korreloimattomia. (7.11)

Kommentti 7.1. Koska varianssi $\mathrm{var}(\mathbf{a}'\mathbf{z}) = \mathrm{E}(\mathbf{a}'\mathbf{z} - \mathbf{a}'\boldsymbol{\mu})^2$ on aina suurempi tai yhtä suuri kuin 0, on

$$\mathrm{var}(\mathbf{a}'\mathbf{z}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \geq 0 \;\forall\, \mathbf{a} \in \mathbb{R}^p. \tag{7.12}$$

Epäyhtälö (7.12) osoittaa, että

♠ satunnaisvektorin $\mathbf{z}$ kovarianssimatriisi $\boldsymbol{\Sigma}$
on aina ei-negatiivisesti definiitti. (7.13)

Voimme päätellä edelleen, että kovarianssimatriisi $\boldsymbol{\Sigma}$ on singulaarinen, jos on olemassa vektori $\mathbf{a} \neq \mathbf{0}$ siten että $\mathrm{var}(\mathbf{a}'\mathbf{z}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = 0$, ts.

$$\mathrm{cov}(\mathbf{z}) = \boldsymbol{\Sigma} \text{ on singulaarinen } \iff \exists\, \mathbf{a} \neq \mathbf{0} \text{ s.e. } \mathbf{a}'\mathbf{z} = \text{vakio tn:llä 1}. \tag{7.14}$$

□

7.1.2 Neliömuodon odotusarvo

Olkoon $\mathbf{y}$ p-ulotteinen satunnaisvektori, jonka odotusarvo on $\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu}$, ja kovarianssimatriisi $\mathrm{cov}(\mathbf{y}) = \boldsymbol{\Sigma}$. Olkoon $\mathbf{A}$ symmetrinen $p \times p$-matriisi. Tällöin neliömuodon $\mathbf{y}'\mathbf{A}\mathbf{y}$ odotusarvo on

$$\boxed{\mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}\,.} \tag{7.15}$$

Tulos (7.15) päätellään seuraavasti:

$$\begin{aligned}
\mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) &= \mathrm{E}[\mathrm{tr}(\mathbf{y}'\mathbf{A}\mathbf{y})] = \mathrm{E}[\mathrm{tr}(\mathbf{A}\mathbf{y}\mathbf{y}')] \\
&= \mathrm{tr}[\mathrm{E}(\mathbf{A}\mathbf{y}\mathbf{y}')] = \mathrm{tr}[\mathbf{A}\,\mathrm{E}(\mathbf{y}\mathbf{y}')] \\
&= \mathrm{tr}[\mathbf{A}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}')] \qquad [\boldsymbol{\Sigma} = \mathrm{E}(\mathbf{y}\mathbf{y}') - \boldsymbol{\mu}\boldsymbol{\mu}'] \\
&= \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \mathrm{tr}(\mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}') \\
&= \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}\,. \qquad (7.16)
\end{aligned}$$

Esimerkki 7.1. (a) Olkoon $\mathbf{y}$ satunnaisvektori, jonka odotusarvo on $\mathrm{E}(\mathbf{y}) = \mu\mathbf{1}_n$, ja kovarianssimatriisi $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{I}_n$. Tilanne voidaan tulkita siten, että meillä on n havainnon yksinkertainen satunnaisotos $y_1, y_2, \dots, y_n$ populaatiosta, jonka odotusarvo on μ ja varianssi σ^2. Tällöin y:n neliösumman

$$t_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \mathbf{y}'(\mathbf{I}_n - \mathbf{J})\mathbf{y} = \mathbf{y}'\mathbf{C}\mathbf{y} \qquad (7.17)$$

odotusarvo on

$$\begin{aligned}
\mathrm{E}(\mathbf{y}'\mathbf{C}\mathbf{y}) &= \mathrm{tr}(\mathbf{C}\sigma^2\mathbf{I}) + \mu\mathbf{1}'\mathbf{C}\mathbf{1}\mu \\
&= \sigma^2\,\mathrm{tr}(\mathbf{I} - \mathbf{J}) \qquad [\mathbf{C}\mathbf{1} = \mathbf{0}] \\
&= \sigma^2[\mathrm{tr}(\mathbf{I}) - \mathrm{tr}(\mathbf{J})] \\
&= \sigma^2(n-1)\,. \qquad (7.18)
\end{aligned}$$

Näin olemme päätyneet tunnettuun tulokseen, jonka mukaan yksinkertaisen (palauttaen tehdyn) satunnaisotoksen tapauksessa otosvarianssi s_y^2 on perusjoukon varianssin σ^2 harhaton estimaattori:

$$\mathrm{E}(\tfrac{1}{n-1}\mathbf{y}'\mathbf{C}\mathbf{y}) = \mathrm{E}(s_y^2) = \sigma^2. \qquad (7.19)$$

(b) Oletetaan että $\mathrm{E}(\mathbf{y}) = \mu\mathbf{1}_n$ mutta kovarianssimatriisi $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{V}$, missä $\mathbf{V}$ on tasakorrelaatiomatriisi:

$$\sigma^2\mathbf{V} = \sigma^2\big[(1-\varrho)\mathbf{I}_n + \varrho\mathbf{1}_n\mathbf{1}_n'\big] = \sigma^2\begin{pmatrix} 1 & \varrho & \dots & \varrho \\ \varrho & 1 & \dots & \varrho \\ \vdots & \vdots & \ddots & \vdots \\ \varrho & \varrho & \dots & 1 \end{pmatrix} \in \mathrm{NND}_n\,. \qquad (7.20)$$

Menetellen kuten (a)-kohdassa saamme

$$\begin{aligned}
\mathrm{E}(\mathbf{y}'\mathbf{C}\mathbf{y}) &= \mathrm{tr}(\mathbf{C}\mathbf{V}\sigma^2) + \mu\mathbf{1}'\mathbf{C}\mathbf{1}\mu = \sigma^2\,\mathrm{tr}(\mathbf{C}\mathbf{V}) \\
&= \sigma^2\,\mathrm{tr}\big[\mathbf{C}((1-\varrho)\mathbf{I} + \varrho\mathbf{1}\mathbf{1}')\big] \\
&= \sigma^2(1-\varrho)\,\mathrm{tr}(\mathbf{C}) \\
&= \sigma^2(1-\varrho)(n-1)\,. \qquad (7.21)
\end{aligned}$$

Täten s_y^2 ei olekaan (jos $\varrho \neq 0$) harhaton estimaattori σ^2:lle, sillä

$$\mathrm{E}(\tfrac{1}{n-1}\mathbf{y}'\mathbf{C}\mathbf{y}) = \mathrm{E}(s_y^2) = \sigma^2(1-\varrho)\,. \tag{7.22}$$

Jos yksinkertainen satunnaisotos valitaan palauttamatta, saadaan $\mathbf{y}$:n korrelaatiomatriisiksi tasakorrelaatiomatriisi; ks. (b)-kohta esimerkissä 1.5 (s. 41).

□

Esimerkki 7.2. Tarkastellaan neliömuotoa

$$\mathbf{y}'\mathbf{A}\mathbf{y} = (y_2-y_1)^2 + (y_3-y_2)^2 + (y_4-y_3)^2 + \cdots + (y_n-y_{n-1})^2, \tag{7.23}$$

missä esimerkin 4.5 (s. 141) mukaan

$$\mathbf{A} = \mathbf{L}'\mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}. \tag{7.24}$$

Määritettävä $\mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y})$, kun $\mathrm{E}(\mathbf{y}) = \mathbf{1}\mu$ ja

(a) $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{I}$. Tällöin

$$\begin{aligned} \mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) &= \mathrm{tr}(\mathbf{A}\sigma^2\mathbf{I}) + \mu\mathbf{1}'\mathbf{A}\mathbf{1}\mu \\ &= \sigma^2\,\mathrm{tr}(\mathbf{A}) \qquad [\mathbf{A}\mathbf{1} = \mathbf{0}] \\ &= \sigma^2 2(n-1). \end{aligned} \tag{7.25}$$

(b) $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{W}$, missä

$$\mathbf{W} = \begin{pmatrix} 1 & \varrho & \varrho^2 & \dots & \varrho^{n-1} \\ \varrho & 1 & \varrho & \dots & \varrho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \varrho^{n-1} & \varrho^{n-2} & \varrho^{n-3} & \dots & 1 \end{pmatrix}. \tag{7.26}$$

Tällöin

$$\begin{aligned} \mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) &= \mathrm{tr}(\mathbf{A}\mathbf{W}\sigma^2) + \mu\mathbf{1}'\mathbf{A}\mathbf{1}\mu = \sigma^2\,\mathrm{tr}(\mathbf{A}\mathbf{W}) \\ &= \sigma^2[(1-\varrho) + 2(1-\varrho) + \cdots + 2(1-\varrho) + (1-\varrho)] \\ &= \sigma^2 2(n-1)(1-\varrho). \end{aligned} \tag{7.27}$$

□

7.2 Satunnaisvektorien x ja y keskinäinen kovarianssimatriisi

Jos $\mathbf{x}$ on p-ulotteinen ja $\mathbf{y}$ on q-ulotteinen satunnaisvektori, joiden odotusarvot ovat $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$ ja $\mathrm{E}(\mathbf{y}) = \boldsymbol{\nu}$, niin $\mathbf{x}$:n ja $\mathbf{y}$:n *keskinäinen* kovarianssimatriisi on matriisimerkinnöin

$$\operatorname{cov}(\mathbf{x}, \mathbf{y}) = \mathrm{E}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\nu})' = \{\operatorname{cov}(x_i, y_j)\} \in \mathbb{R}^{p \times q}\,. \tag{7.28}$$

Tällöin tietenkin $\operatorname{cov}(\mathbf{x}, \mathbf{x}) = \operatorname{cov}(\mathbf{x})$.

Tulosta (7.5) (s. 214) vastaten on helppo osoittaa, että

$$\operatorname{cov}(\mathbf{x}, \mathbf{y}) = \mathrm{E}(\mathbf{x}\mathbf{y}') - \boldsymbol{\mu}\boldsymbol{\nu}'. \tag{7.29}$$

Varoitus: Tulon $\mathbf{x}\mathbf{y}'$ odotusarvo on vain poikkeustapauksessa $\mathbf{x}$:n ja $\mathbf{y}$:n odotusarvojen tulo: ehto on se että satunnaisvektorit $\mathbf{x}$ ja $\mathbf{y}$ ovat tilastollisesti riippumattomia.

Havaitsemme edelleen, että

$$\begin{aligned} \operatorname{cov}(\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}) &= \mathrm{E}[(\mathbf{A}\mathbf{x} - \mathbf{A}\boldsymbol{\mu})(\mathbf{B}\mathbf{y} - \mathbf{B}\boldsymbol{\nu})'] \\ &= \mathrm{E}[\mathbf{A}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\nu})'\mathbf{B}'] = \mathbf{A}\,\mathrm{E}\big[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\nu})'\big]\mathbf{B}' \\ &= \mathbf{A}\operatorname{cov}(\mathbf{x}, \mathbf{y})\mathbf{B}', \end{aligned} \tag{7.30}$$

joten

$$\boxed{\operatorname{cov}(\mathbf{x}, \mathbf{y}) = \boldsymbol{\Sigma}_{\mathbf{xy}} \implies \operatorname{cov}(\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}) = \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{xy}}\mathbf{B}'. \qquad (7.31)}$$

Jo luvussa 1.3 (s. 41) korostettiin merkintöjen

$$\operatorname{cov}(\mathbf{x}, \mathbf{y}) = \boldsymbol{\Sigma}_{\mathbf{xy}} \in \mathbb{R}^{p \times q} \quad \text{ja} \quad \operatorname{cov}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \boldsymbol{\Sigma} \in \mathbb{R}^{(p+q)(p+q)} \tag{7.32}$$

välistä eroa. Nyt siis $\operatorname{cov}(\mathbf{x}, \mathbf{y}) = \boldsymbol{\Sigma}_{\mathbf{xy}}$ tarkoittaa satunnaisvektorien $\mathbf{x}$ (p elementtiä) ja $\mathbf{y}$ (q elementtiä) keskinäistä kovarianssimatriisia (cross-covariance matrix), mikä on $p \times q$-matriisi. Sen sijaan $\boldsymbol{\Sigma}$ on $(p+q)$-ulotteisen satunnaisvektorin kovarianssimatriisi, joka voidaan esittää ositetussa muodossa

$$\operatorname{cov}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \operatorname{cov}(\mathbf{x}) & \operatorname{cov}(\mathbf{x}, \mathbf{y}) \\ \operatorname{cov}(\mathbf{y}, \mathbf{x}) & \operatorname{cov}(\mathbf{y}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}} \\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \boldsymbol{\Sigma}_{\mathbf{yy}} \end{pmatrix} = \boldsymbol{\Sigma}\,. \tag{7.33}$$

Erityisesti jos $\mathbf{z} = \left(\begin{smallmatrix} \mathbf{x} \\ y \end{smallmatrix}\right)$ on $(p+1)$:n elementin satunnaisvektori, niin

$$\operatorname{cov}(\mathbf{z}) = \operatorname{cov}\begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} = \begin{pmatrix} \operatorname{cov}(\mathbf{x}) & \operatorname{cov}(\mathbf{x}, y) \\ \operatorname{cov}(y, \mathbf{x}) & \operatorname{var}(y) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}'_{\mathbf{x}y} & \sigma_y^2 \end{pmatrix}, \tag{7.34}$$

missä

$$\boldsymbol{\sigma}_{\mathbf{x}y} = \operatorname{cov}(\mathbf{x}, y) = \begin{pmatrix} \operatorname{cov}(x_1, y) \\ \operatorname{cov}(x_2, y) \\ \vdots \\ \operatorname{cov}(x_p, y) \end{pmatrix} = \begin{pmatrix} \sigma_{1y} \\ \sigma_{2y} \\ \vdots \\ \sigma_{py} \end{pmatrix}. \tag{7.35}$$

7.2.1 Satunnaisvektorien summan kovarianssimatriisi

Olkoot $\mathbf{x}$ ja $\mathbf{y}$ kumpikin p elementin satunnaisvektoreita. Tällöin niiden summan kovarianssimatriisi saadaan seuraavasti:

$$\begin{aligned} \operatorname{cov}(\mathbf{x}+\mathbf{y}) &= \mathrm{E}[\mathbf{x}+\mathbf{y}-(\boldsymbol{\mu}+\boldsymbol{\nu})][\mathbf{x}+\mathbf{y}-(\boldsymbol{\mu}+\boldsymbol{\nu})]' \\ &= \mathrm{E}[(\mathbf{x}-\boldsymbol{\mu})+(\mathbf{y}-\boldsymbol{\nu})][(\mathbf{x}-\boldsymbol{\mu})+(\mathbf{y}-\boldsymbol{\nu})]' \\ &= \mathrm{E}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})' + \mathrm{E}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\nu})' \\ &\quad + \mathrm{E}(\mathbf{y}-\boldsymbol{\nu})(\mathbf{x}-\boldsymbol{\mu})' + \mathrm{E}(\mathbf{y}-\boldsymbol{\nu})(\mathbf{y}-\boldsymbol{\nu})', \end{aligned} \tag{7.36}$$

joten

$$\begin{aligned} \operatorname{cov}(\mathbf{x}+\mathbf{y}) &= \operatorname{cov}(\mathbf{x}) + \operatorname{cov}(\mathbf{y}) + \operatorname{cov}(\mathbf{x},\mathbf{y}) + \operatorname{cov}(\mathbf{y},\mathbf{x}) \\ &= \boldsymbol{\Sigma}_{\mathbf{xx}} + \boldsymbol{\Sigma}_{\mathbf{yy}} + \boldsymbol{\Sigma}_{\mathbf{xy}} + \boldsymbol{\Sigma}_{\mathbf{xy}}'. \end{aligned} \tag{7.37}$$

Sanomme, että satunnaisvektorit $\mathbf{x}$ ja $\mathbf{y}$ ovat korreloimattomia, jos jokainen $\mathbf{x}$:n elementti on korreloimaton jokaisen $\mathbf{y}$:n elementin kanssa eli jos

$$\operatorname{cov}(\mathbf{x},\mathbf{y}) = \boldsymbol{\Sigma}_{\mathbf{xy}} = \mathbf{0}\,. \tag{7.38}$$

Täten (7.36):n perusteella saamme (7.11):tä vastaten:

> Jos satunnaisvektorit $\mathbf{x}$ ja $\mathbf{y}$ ovat korreloimattomia, niin summan $\mathbf{x}+\mathbf{y}$ kovarianssimatriisi on kovarianssimatriisien summa $\operatorname{cov}(\mathbf{x}) + \operatorname{cov}(\mathbf{y})$. (7.39)

On mielenkiintoista havaita, että (7.36):n nojalla riittävä ja välttämätön ehto sille että

$$\operatorname{cov}(\mathbf{x}+\mathbf{y}) = \operatorname{cov}(\mathbf{x}) + \operatorname{cov}(\mathbf{y})\,, \tag{7.40}$$

on seuraava:

$$\operatorname{cov}(\mathbf{x},\mathbf{y}) = -\operatorname{cov}(\mathbf{y},\mathbf{x}), \ \text{ ts. } \boldsymbol{\Sigma}_{\mathbf{xy}} = -\boldsymbol{\Sigma}_{\mathbf{xy}}'\,. \tag{7.41}$$

Ehto (7.41) merkitsee, että $\boldsymbol{\Sigma}_{\mathbf{xy}}$ on vinosymmetrinen.

Tuloksen (7.36) yleistyksenä on tietenkin voimassa

$$\operatorname{cov}(\mathbf{Ax}+\mathbf{By}) = \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{A}' + \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{yy}}\mathbf{B}' + \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{xy}}\mathbf{B}' + \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{yx}}\mathbf{A}'. \tag{7.42}$$

Edelleen saamme

$$\begin{aligned} \operatorname{cov}(\mathbf{x}+\mathbf{y},\mathbf{z}) &= \mathrm{E}[\mathbf{x}+\mathbf{y}-(\boldsymbol{\mu}+\boldsymbol{\nu})](\mathbf{z}-\boldsymbol{\delta})' \qquad [\mathrm{E}(\mathbf{z})=\boldsymbol{\delta}] \\ &= \mathrm{E}[(\mathbf{x}-\boldsymbol{\mu})+(\mathbf{y}-\boldsymbol{\nu})](\mathbf{z}-\boldsymbol{\delta})' \\ &= \mathrm{E}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\delta})' + \mathrm{E}(\mathbf{y}-\boldsymbol{\nu})(\mathbf{z}-\boldsymbol{\delta})' \\ &= \operatorname{cov}(\mathbf{x},\mathbf{z}) + \operatorname{cov}(\mathbf{y},\mathbf{z}) \\ &= \boldsymbol{\Sigma}_{\mathbf{xz}} + \boldsymbol{\Sigma}_{\mathbf{yz}}\,, \end{aligned} \tag{7.43}$$

sekä

$$\operatorname{cov}(\mathbf{Ax}+\mathbf{By},\mathbf{Cz}) = \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{xz}}\mathbf{C}' + \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{yz}}\mathbf{C}'. \tag{7.44}$$

7.3 Yksinkertainen lineaarinen malli

Olkoon $\mathbf{y}$ satunnaisvektori (n elementtiä), jonka odotusarvo on $\mathrm{E}(\mathbf{y}) = \mathbf{1}\mu$, ja kovarianssimatriisi $\operatorname{cov}(\mathbf{y}) = \mathbf{V}$. Tällöin meillä on tarkasteltavana hyvin yksinkertainen lineaarinen malli

$$\mathscr{M}_0\colon\ y_i = \mu + \varepsilon_i\,, \quad i = 1,2,\dots,n\,, \tag{7.45a}$$

eli matriisimerkinnöin

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \ \text{ts. } \mathbf{y} = \mu\mathbf{1} + \boldsymbol{\varepsilon}\,, \tag{7.45b}$$

missä $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ ja

$$\operatorname{cov}(\mathbf{y}) = \mathrm{E}(\mathbf{y}-\mu\mathbf{1})(\mathbf{y}-\mu\mathbf{1})' = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \operatorname{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}. \tag{7.46}$$

Lyhyesti mallia voidaan merkitä kolmikolla

$$\mathscr{M}_0 = \{\mathbf{y}, \mathbf{1}\mu, \mathbf{V}\} = \{\mathbf{y}, \mathrm{E}(\mathbf{y}), \operatorname{cov}(\mathbf{y})\}\,. \tag{7.47}$$

Mallissa $\mathscr{M}_0$ ei ole mukana yhtään ”varsinaista” selittäjämuuttujaa. Ainoa selittäjä on $\mathbf{1}$ eli ainoa selittäjä on ”vakio”. Merkitään

$$\hat{\mu} = \frac{1}{n}\mathbf{1}'\mathbf{y} := \mathbf{a}'\mathbf{y} = \bar{y}, \quad \tilde{\mu} = \frac{1}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}}\mathbf{1}'\mathbf{V}^{-1}\mathbf{y} := \mathbf{b}'\mathbf{y}\,. \tag{7.48}$$

Tällöin $\hat{\mu}$ ja $\tilde{\mu}$ ovat parametrin μ estimaattoreita. [Osoittautuu, että $\hat{\mu} = \mathrm{OLSE}(\mu) = \mu$:n pienimmän neliösumman estimaattori ja $\tilde{\mu} = \mathrm{BLUE}(\mu) = \mu$:n paras lineaarinen harhaton estimaattori.] Havaitsemme, että molemmat ovat harhattomia:

(1) $\mathrm{E}(\hat{\mu}) = \mu$, sillä

$$\mathrm{E}(\hat{\mu}) = \mathrm{E}(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\,\mathrm{E}(\mathbf{y}) = \frac{1}{n}\mathbf{1}'\mathbf{1}\mu = \frac{1}{n}n\mu = \mu\,. \tag{7.49}$$

(2) $\mathrm{E}(\tilde{\mu}) = \mu$, sillä

$$\mathrm{E}(\tilde{\mu}) = \mathrm{E}(\mathbf{b}'\mathbf{y}) = \mathbf{b}'\,\mathrm{E}(\mathbf{y}) = \frac{1}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}}\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}\mu = \mu\,. \tag{7.50}$$

Estimaattoreiden varianssit ovat:

(3) $\mathrm{var}(\hat{\mu}) = \frac{1}{n^2}\mathbf{1}'\mathbf{V}\mathbf{1}$, sillä

$$\mathrm{var}(\hat{\mu}) = \mathrm{var}(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\mathbf{V}\mathbf{a} = \frac{1}{n}\mathbf{1}'\mathbf{V}\frac{1}{n}\mathbf{1} = \frac{1}{n^2}\mathbf{1}'\mathbf{V}\mathbf{1}\,. \tag{7.51}$$

(4) $\mathrm{var}(\tilde{\mu}) = \frac{1}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}}$, sillä

$$\begin{aligned}\mathrm{var}(\tilde{\mu}) &= \mathrm{var}(\mathbf{b}'\mathbf{y}) = \mathbf{b}'\mathbf{V}\mathbf{b}\\ &= \frac{1}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}}\mathbf{1}'\mathbf{V}^{-1}\cdot\mathbf{V}\cdot\mathbf{V}^{-1}\mathbf{1}\frac{1}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}} = \frac{1}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}}.\end{aligned} \tag{7.52}$$

Osoitamme vielä, että varianssien välillä on epäyhtälö

$$\mathrm{var}(\tilde{\mu}) = \frac{1}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}} \le \frac{1}{n^2}\mathbf{1}'\mathbf{V}\mathbf{1} = \mathrm{var}(\hat{\mu})\,. \tag{7.53}$$

Nimittäin Cauchyn–Schwarzin epäyhtälön version (CS3) (s. 209) mukaan on voimassa

$$(\mathbf{u}'\mathbf{v})^2 \le \mathbf{u}'\mathbf{A}\mathbf{u}\cdot\mathbf{v}'\mathbf{A}^{-1}\mathbf{v}\,, \tag{CS3}$$

eli jos $\mathbf{u} = \mathbf{v} = \mathbf{1}$, ja $\mathbf{A} = \mathbf{V}$, niin $(\mathbf{1}'\mathbf{1})^2 \le \mathbf{1}'\mathbf{V}\mathbf{1}\cdot\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}$, ts.

$$n^2 \le \mathbf{1}'\mathbf{V}\mathbf{1}\cdot\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}\,. \tag{7.54}$$

Epäyhtälöt (7.53) ja (7.54) ovat tietenkin identtiset.

Mainittakoon, että yhtäsuuruus epäyhtälössä (7.53) pätee [ks. yhtäsuuruusehto (CS3):lle, s. 209] täsmälleen silloin kun on olemassa sellainen λ, että

$$\mathbf{V}\mathbf{1} = \lambda\mathbf{1}\,. \tag{7.55}$$

Yhtälö (7.55) merkitsee, että $\mathbf{1}$ on $\mathbf{V}$:n ominaisvektori; samoin se merkitsee, että $\mathbf{V}$:n rivisummat ovat identtiset. Näin voimme päätellä että mallin $\{\mathbf{y}, \mathbf{1}\mu, \mathbf{V}\}$ tilanteessa

$$\mathrm{OLSE}(\mu) = \mathrm{BLUE}(\mu) \iff \mathbf{V}\mathbf{1} = \lambda\mathbf{1} \text{ jollakin } \lambda\text{:n arvolla.} \tag{7.56}$$

Palataan vielä epäyhtälöön (7.53). Kerrotaan se puolittain n:llä, jolloin saadaan

$$\frac{n}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}} \leq \frac{1}{n}\mathbf{1}'\mathbf{V}\mathbf{1}\,. \tag{7.57}$$

Jos merkitään $\mathbf{t} = \frac{1}{\sqrt{n}}\mathbf{1}$, (7.57) voidaan esittää muodossa

$$(\mathbf{t}'\mathbf{V}^{-1}\mathbf{t})^{-1} \leq \mathbf{t}'\mathbf{V}\mathbf{t}\,, \tag{7.58}$$

missä siis $\mathbf{t}'\mathbf{t} = 1$. Ks. myös ominaisvektorin luonnehdinta (5.59) (s. 170).

Esimerkki 7.3. Olkoon $\mathbf{y}$ satunnaisvektori, jonka odotusarvo on $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ ja kovarianssimatriisi $\mathrm{cov}(\mathbf{y}) = \mathbf{V}$ eli kyseessä on lineaarinen malli

$$\mathscr{M} = \{\mathbf{y}, \mathrm{E}(\mathbf{y}), \mathrm{cov}(\mathbf{y})\} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}\,, \tag{7.59}$$

eli $\mathbf{y}$:n uskotaan syntyvän seuraavasti:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \quad \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}\,, \quad \mathrm{cov}(\mathbf{y}) = \mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}. \tag{7.60}$$

Merkitään

$$\begin{aligned} \mathbf{H} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', & \mathbf{G} &= \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}, \\ \hat{\boldsymbol{\mu}} &= \hat{\mathbf{y}} = \mathbf{H}\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}}, & \tilde{\boldsymbol{\mu}} &= \mathbf{G}\mathbf{y} = \mathbf{X}\tilde{\boldsymbol{\beta}}. \end{aligned}$$

Osoita

(a) $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad \mathrm{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$,

(b) $\mathrm{E}(\hat{\boldsymbol{\mu}}) = \mathbf{X}\boldsymbol{\beta}, \quad \mathrm{cov}(\hat{\boldsymbol{\mu}}) = \mathbf{H}\mathbf{V}\mathbf{H}$,

(c) $\mathrm{E}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad \mathrm{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$,

(d) $\mathrm{E}(\tilde{\boldsymbol{\mu}}) = \mathbf{X}\boldsymbol{\beta}, \quad \mathrm{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'$,

(e) $\mathrm{cov}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) = \mathrm{cov}(\hat{\boldsymbol{\beta}}) - \mathrm{cov}(\tilde{\boldsymbol{\beta}})$,

(f) $\mathrm{cov}(\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}) = \mathrm{cov}(\hat{\boldsymbol{\mu}}) - \mathrm{cov}(\tilde{\boldsymbol{\mu}})$,

(g) $\mathrm{cov}\begin{pmatrix}\hat{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\beta}}\end{pmatrix} = \begin{pmatrix}\mathrm{cov}(\hat{\boldsymbol{\beta}}) & \mathrm{cov}(\tilde{\boldsymbol{\beta}}) \\ \mathrm{cov}(\tilde{\boldsymbol{\beta}}) & \mathrm{cov}(\tilde{\boldsymbol{\beta}})\end{pmatrix}$.

□

7.4 Muunnos korreloimattomiksi osavektoreiksi

Olkoon $p+q$ elementin satunnaisvektorin $\mathbf{z} = \binom{\mathbf{x}}{\mathbf{y}}$ positiivisesti definiitti kovarianssimatriisi

$$\operatorname{cov}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \begin{pmatrix}\operatorname{cov}(\mathbf{x}) & \operatorname{cov}(\mathbf{x},\mathbf{y})\\ \operatorname{cov}(\mathbf{y},\mathbf{x}) & \operatorname{cov}(\mathbf{y})\end{pmatrix} = \begin{pmatrix}\mathbf{\Sigma_{xx}} & \mathbf{\Sigma_{xy}}\\ \mathbf{\Sigma_{yx}} & \mathbf{\Sigma_{yy}}\end{pmatrix} = \mathbf{\Sigma}\,. \tag{7.61}$$

Hajotelman (4.99) (s. 148) perusteella matriisi $\mathbf{\Sigma}$ voidaan muuntaa lohkolävistäjämuotoon seuraavasti:

$$\begin{pmatrix}\mathbf{I}_p & \mathbf{0}\\ -\mathbf{\Sigma_{yx}}\mathbf{\Sigma_{xx}^{-1}} & \mathbf{I}_q\end{pmatrix}\mathbf{\Sigma}\begin{pmatrix}\mathbf{I}_p & -\mathbf{\Sigma_{xx}^{-1}}\mathbf{\Sigma_{xy}}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix} = \begin{pmatrix}\mathbf{\Sigma_{xx}} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma_{yy\cdot x}}\end{pmatrix}, \tag{7.62}$$

eli

$$\mathbf{B\Sigma B'} = \mathbf{\Omega}\,. \tag{7.63}$$

Tehdään nyt $\mathbf{z}$:lle muunnos $\mathbf{u} = \mathbf{Bz}$:

$$\mathbf{u} = \begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix} = \mathbf{Bz} = \begin{pmatrix}\mathbf{I}_p & \mathbf{0}\\ -\mathbf{\Sigma_{yx}}\mathbf{\Sigma_{xx}^{-1}} & \mathbf{I}_q\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}, \tag{7.64}$$

jolloin

$$\mathbf{u}_1 = \mathbf{x}\,, \tag{7.65a}$$

$$\mathbf{u}_2 = \mathbf{y} - \mathbf{\Sigma_{yx}}\mathbf{\Sigma_{xx}^{-1}}\mathbf{x}\,. \tag{7.65b}$$

Koska

$$\operatorname{cov}\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix} = \operatorname{cov}(\mathbf{By}) = \mathbf{B\Sigma B'} = \begin{pmatrix}\mathbf{\Sigma_{xx}} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma_{yy\cdot x}}\end{pmatrix}, \tag{7.66}$$

ovat $\mathbf{u}_1$ ja $\mathbf{u}_2$ korreloimattomia,

$$\operatorname{cov}(\mathbf{u}_1,\mathbf{u}_2) = \operatorname{cor}(\mathbf{u}_1,\mathbf{u}_2) = \mathbf{0}\,, \tag{7.67}$$

ja niiden kovarianssimatriisit ovat

$$\operatorname{cov}(\mathbf{u}_1) = \mathbf{\Sigma_{xx}}\,, \tag{7.68a}$$

$$\operatorname{cov}(\mathbf{u}_2) = \mathbf{\Sigma_{yy\cdot x}} = \mathbf{\Sigma_{yy}} - \mathbf{\Sigma_{yx}}\mathbf{\Sigma_{xx}^{-1}}\mathbf{\Sigma_{xy}}\,. \tag{7.68b}$$

Matriisi $\mathbf{\Sigma_{yy\cdot x}}$ on $\mathbf{\Sigma_{xx}}$:n *Schurin komplementti* $\mathbf{\Sigma}$:ssa.

Voimme myös lähestyä edellä tarkasteltua kysymystä siten, että määritämme sellaisen matriisin $\mathbf{L} \in \mathbb{R}^{q\times p}$, jolla on ominaisuus

$$\operatorname{cov}(\mathbf{x},\mathbf{y}-\mathbf{Lx}) = \mathbf{0}\,. \tag{7.69}$$

Tällöin

$$\operatorname{cov}(\mathbf{x},\mathbf{y}-\mathbf{Lx}) = \operatorname{cov}(\mathbf{x},\mathbf{y}) - \operatorname{cov}(\mathbf{x},\mathbf{Lx}) = \mathbf{\Sigma_{xy}} - \mathbf{\Sigma_{xx}}\mathbf{L}' = \mathbf{0} \tag{7.70}$$

eli $\mathbf{L\Sigma_{xx}} = \mathbf{\Sigma_{yx}}$, mistä $\mathbf{L}$:n lausekkeeksi saadaan $\mathbf{L} = \mathbf{\Sigma_{yx}}\mathbf{\Sigma_{xx}^{-1}}$.

Esimerkki 7.4. Tarkastellaan kaksiulotteista satunnaismuuttujaa $\mathbf{z}$, jonka kovarianssimatriisi on

$$\operatorname{cov}(\mathbf{z}) = \operatorname{cov}\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \sigma_x\sigma_y\varrho_{xy} \\ \sigma_x\sigma_y\varrho_{xy} & \sigma_y^2 \end{pmatrix}. \tag{7.71}$$

Tällöin muunnosta

$$\mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ -\mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^{-1} & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \tag{7.72}$$

vastaten saamme

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\sigma_{yx}\sigma_x^{-2} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{7.73}$$

ts. uusiksi korreloimattomiksi muuttujiksi saadaan

$$u_1 = x\,, \tag{7.74a}$$

$$u_2 = y - \frac{\sigma_{yx}}{\sigma_x^2}\,x = y - \varrho_{xy}\,\frac{\sigma_y}{\sigma_x}\,x\,. \tag{7.74b}$$

Lisäksi on

$$\begin{aligned} \operatorname{var}(u_2) = \sigma_{y\cdot x}^2 &= \sigma_y^2 - \sigma_{yx}\sigma_x^{-2}\sigma_{xy} \\ &= \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2} = \sigma_y^2\Big(1 - \frac{\sigma_{xy}^2}{\sigma_x^2\sigma_y^2}\Big) \\ &= \sigma_y^2(1 - \varrho_{xy}^2) \le \sigma_y^2 = \operatorname{var}(y)\,. \end{aligned} \tag{7.75}$$

Erityisesti jos

$$\operatorname{cov}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}, \tag{7.76}$$

niin

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\varrho & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{7.77}$$

ts.

$$u_1 = x\,, \tag{7.78a}$$

$$u_2 = y - \varrho x\,. \tag{7.78b}$$

□

7.5 Lausekkeen $\operatorname{cov}(\mathbf{y}-\mathbf{Fx})$ minimointi

Olkoon tehtävänä määrittää sellainen matriisi $\mathbf{F}_{q\times p}$, jolla on ominaisuus

$$\operatorname{cov}(\mathbf{y}-\mathbf{Fx}) \text{ on mahdollisimman pieni Löwnerin mielessä,} \tag{7.79}$$

eli jos $\mathbf{F}_*$ antaa ko. minimin, niin

$$\operatorname{cov}(\mathbf{y}-\mathbf{F}_*\mathbf{x}) \leq_{\mathsf{L}} \operatorname{cov}(\mathbf{y}-\mathbf{Fx}) \quad \text{kaikilla } \mathbf{F}\in\mathbb{R}^{q\times p}. \tag{7.80}$$

Oletamme, että $p+q$ elementin satunnaisvektorin $\mathbf{z}=\binom{\mathbf{x}}{\mathbf{y}}$ positiivisesti definiitti kovarianssimatriisi $\mathbf{\Sigma}$ on ositettu kuten (7.61):ssa (s. 223). Osoitamme, että tällöin

$$\boxed{\operatorname{cov}(\mathbf{y}-\mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}) \leq_{\mathsf{L}} \operatorname{cov}(\mathbf{y}-\mathbf{Fx}) \quad \textit{kaikilla } \mathbf{F}\in\mathbb{R}^{q\times p}.} \tag{7.81}$$

Merkitään

$$\mathbf{F}_* = \mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}. \tag{7.82}$$

On selvää, että jos (7.81) on voimassa, niin satunnaisvektori $\mathbf{F}_*\mathbf{x}$ on jossakin mielessä "tilastollisesti lähellä" satunnaisvektoria $\mathbf{y}$, ja siten $\mathbf{F}_*\mathbf{x}$:llä voisi olla käyttöä $\mathbf{y}$:tä arvioitaessa.

Havaitsemme että

$$\begin{aligned}
\operatorname{cov}(\mathbf{y}-\mathbf{Fx}) &= \operatorname{cov}\big[(\mathbf{y}-\mathbf{F}_*\mathbf{x})+(\mathbf{F}_*\mathbf{x}-\mathbf{Fx})\big] \\
&= \operatorname{cov}(\mathbf{u}_1+\mathbf{u}_2) \qquad [\mathbf{u}_1=\mathbf{y}-\mathbf{F}_*\mathbf{x},\ \mathbf{u}_2=(\mathbf{F}_*-\mathbf{F})\mathbf{x}] \\
&= \operatorname{cov}(\mathbf{u}_1)+\operatorname{cov}(\mathbf{u}_2)+\operatorname{cov}(\mathbf{u}_1,\mathbf{u}_2)+\operatorname{cov}(\mathbf{u}_2,\mathbf{u}_1) \\
&= \operatorname{cov}(\mathbf{u}_1)+\operatorname{cov}(\mathbf{u}_2) \\
&= \operatorname{cov}(\mathbf{y}-\mathbf{F}_*\mathbf{x})+\operatorname{cov}\big[(\mathbf{F}_*-\mathbf{F})\mathbf{x}\big]\,,
\end{aligned} \tag{7.83}$$

sillä $\mathbf{u}_1$:n ja $\mathbf{x}$:n korreloimattomuudesta seuraa, että myös $\mathbf{u}_1$:n ja $\mathbf{u}_2$ ovat korreloimattomia:

$$\operatorname{cov}(\mathbf{u}_1,\mathbf{u}_2)=\operatorname{cov}[\mathbf{u}_1,(\mathbf{F}_*-\mathbf{F})\mathbf{x}]=\operatorname{cov}(\mathbf{u}_1,\mathbf{x})(\mathbf{F}_*-\mathbf{F})'=\mathbf{0}\,. \tag{7.84}$$

Täten saamme (7.83):n perusteella

$$\operatorname{cov}(\mathbf{y}-\mathbf{Fx})-\operatorname{cov}(\mathbf{y}-\mathbf{F}_*\mathbf{x})=\operatorname{cov}\big[(\mathbf{F}_*-\mathbf{F})\mathbf{u}_1\big]\,. \tag{7.85}$$

Koska $\operatorname{cov}\big[(\mathbf{F}_*-\mathbf{F})\mathbf{u}_1\big]$ on ei-negatiivisesti definiitti, on erotus $\operatorname{cov}(\mathbf{y}-\mathbf{Fx})-\operatorname{cov}(\mathbf{y}-\mathbf{F}_*\mathbf{x})$ triviaalisti myös ei-negatiivisesti definiitti eli

$$\operatorname{cov}(\mathbf{y}-\mathbf{Fx})-\operatorname{cov}(\mathbf{y}-\mathbf{F}_*\mathbf{x}) \geq_{\mathsf{L}} \mathbf{0} \tag{7.86}$$

ts.

$$\operatorname{cov}(\mathbf{y}-\mathbf{Fx}) \geq_{\mathsf{L}} \operatorname{cov}(\mathbf{y}-\mathbf{F}_*\mathbf{x}) \quad \text{kaikilla } \mathbf{F}\in\mathbb{R}^{q\times p}. \tag{7.87}$$

Yhtäsuuruus (7.87):ssa on voimassa vain jos $\mathbf{F} = \mathbf{F}_*$.

Voimme soveltaa em. tulosta tilanteeseen, jossa $\mathbf{z} = \binom{\mathbf{x}}{y}$ on $(p+1)$:n elementin satunnaisvektori:

$$\operatorname{cov}(\mathbf{z}) = \operatorname{cov}\begin{pmatrix}\mathbf{x}\\ y\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y}\\ \boldsymbol{\sigma}'_{\mathbf{x}y} & \sigma_y^2\end{pmatrix}, \quad \boldsymbol{\sigma}_{\mathbf{x}y} = \operatorname{cov}(\mathbf{x}, y) = \begin{pmatrix}\sigma_{1y}\\ \sigma_{2y}\\ \vdots\\ \sigma_{py}\end{pmatrix}. \tag{7.88}$$

Mikä on se $\mathbf{x}$:n lineaarikombinaatio $\mathbf{f}'_*\mathbf{x}$, jolla on ominaisuus

$$\operatorname{var}(y - \mathbf{f}'_*\mathbf{x}) = \min_{\mathbf{f}} \operatorname{var}(y - \mathbf{f}'\mathbf{x})? \tag{7.89}$$

Tuloksen (7.81) mukaan

$$\operatorname{var}(y - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}) \le \operatorname{var}(y - \mathbf{f}'\mathbf{x}) \quad \text{kaikilla } \mathbf{f} \in \mathbb{R}^p, \tag{7.90}$$

joten haettu kerroinvektori $\mathbf{f}_*$ on

$$\mathbf{f}_* = \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}\,, \tag{7.91}$$

ja erotuksen $y - \mathbf{f}'_*\mathbf{x}$ varianssi on

$$\operatorname{var}(y - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}) = \sigma_y^2 - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y} := \sigma_{y\cdot\mathbf{x}}^2. \tag{7.92}$$

Havaitsemme, että $\sigma_{y\cdot\mathbf{x}}^2$ voidaan esittää seuraavasti:

$$\begin{aligned}\sigma_{y\cdot\mathbf{x}}^2 &= \sigma_y^2 - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}\\ &= \sigma_y^2\Big(1 - \frac{\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}}{\sigma_y^2}\Big) = \sigma_y^2(1 - \varrho_{y\cdot\mathbf{x}}^2)\,,\end{aligned} \tag{7.93}$$

missä

$$\begin{aligned}\varrho_{y\cdot\mathbf{x}}^2 = \operatorname{cor}^2(y, \mathbf{f}'_*\mathbf{x}) &= \frac{\operatorname{cov}^2(y, \mathbf{f}'_*\mathbf{x})}{\operatorname{var}(y)\operatorname{var}(\mathbf{f}'_*\mathbf{x})} = \frac{(\boldsymbol{\sigma}'_{\mathbf{x}y}\mathbf{f}_*)^2}{\sigma_y^2\cdot\mathbf{f}'_*\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{f}_*}\\ &= \frac{(\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y})^2}{\sigma_y^2\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}} = \frac{\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}}{\sigma_y^2}\,.\end{aligned} \tag{7.94}$$

Toisaalta Cauchyn–Schwarzin epäyhtälön $(\mathbf{u}'\mathbf{v})^2 \le \mathbf{u}'\mathbf{A}\mathbf{u}\cdot\mathbf{v}'\mathbf{A}^{-1}\mathbf{v}$ mukaan:

$$\begin{aligned}\operatorname{cor}^2(y, \mathbf{f}'\mathbf{x}) &= \frac{(\boldsymbol{\sigma}'_{\mathbf{x}y}\mathbf{f})^2}{\sigma_y^2\cdot\mathbf{f}'\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{f}} \le \frac{\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}\cdot\mathbf{f}'\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{f}}{\sigma_y^2\cdot\mathbf{f}'\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{f}}\\ &= \frac{\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}}{\sigma_y^2} = \varrho_{y\cdot\mathbf{x}}^2 = \operatorname{cor}^2(y, \mathbf{f}'_*\mathbf{x})\,.\end{aligned} \tag{7.95}$$

Täten voimme päätellä seuraavan tuloksen:

Tehtävät

$$\min_{\mathbf{f}} \operatorname{var}(y - \mathbf{f}'\mathbf{x}) \quad ja \quad \max_{\mathbf{f}} \operatorname{cor}^2(y, \mathbf{f}'\mathbf{x}) \tag{7.96}$$

johtavat (oleellisesti) samaan tulokseen: $\mathbf{f}_* = \mathbf{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}$.

Otosvarianssia ja -korrelaatiota koskevat vastaavat tulokset on esitetty luvussa 8.8 (s. 287). Maksimaalista korrelaatiota $\varrho_{y\cdot\mathbf{x}}$ sanotaan *populaation yhteiskorrelaatiokertoimeksi*.

Huomattakoon, että jos

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}'_{\mathbf{x}y} & \sigma_y^2 \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{\Sigma}^{\mathbf{xx}} & \boldsymbol{\sigma}^{12} \\ (\boldsymbol{\sigma}^{12})' & \sigma^{yy} \end{pmatrix}, \tag{7.97}$$

niin $\mathbf{\Sigma}$:n käänteismatriisin viimeinen lävistäjäelementti σ^{yy} on

$$\sigma^{yy} = \frac{1}{\sigma_y^2 - \boldsymbol{\sigma}'_{\mathbf{x}y}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}} = \frac{1}{\sigma_{y\cdot\mathbf{x}}^2}. \tag{7.98}$$

7.6 Paras lineaarinen prediktori, BLP

Olkoon $f(\mathbf{x})$ jokin p-ulotteisen satunnaisvektorin $\mathbf{x}$ reaalilukuarvoinen funktio (esim. $f(\mathbf{x}) = \mathbf{a}'\mathbf{x}$, $\mathbf{a} \in \mathbb{R}^p$) ja olkoon y jokin kiinteä vakio (reaaliluku) tai yksidimensioinen satunnaismuuttuja. Tällöin $f(\mathbf{x})$:n *keskineliövirhe* y:n suhteen määritellään lausekkeena

$$\operatorname{MSE}[f(\mathbf{x}); y] = \operatorname{E}[y - f(\mathbf{x})]^2. \tag{7.99}$$

Vastaavasti jos $f(\mathbf{x})$ on q-ulotteinen satunnaisvektori (esim. $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$, $\mathbf{A} \in \mathbb{R}^{q\times p}$) ja $\mathbf{y}$ on vektori (joko satunnaisvektori tai kiinteä $\mathbb{R}^q$:n vektori) niin $f(\mathbf{x})$:n keskineliövirhematriisi $\mathbf{y}$:n suhteen on

$$\operatorname{MSEM}[f(\mathbf{x}); \mathbf{y}] = \operatorname{E}[\mathbf{y} - f(\mathbf{x})][\mathbf{y} - f(\mathbf{x})]'. \tag{7.100}$$

Tarkastellaan $(p+1)$-ulotteista satunnaisvektoria $\mathbf{z}$:

$$\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ y \end{pmatrix}, \quad \operatorname{E}(\mathbf{z}) = \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \mu_y \end{pmatrix}, \quad \operatorname{cov}(\mathbf{z}) = \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}'_{\mathbf{x}y} & \sigma_y^2 \end{pmatrix}. \tag{7.101}$$

Ajatellaan, että meidän on mahdollista havainnoida satunnaisvektorin $\mathbf{x}$ arvoja mutta ei y:n arvoja ja tavoitteena on muodostaa jokin "hyvä" $\mathbf{x}$:n funktio $f(\mathbf{x})$, jonka avulla saataisiin "hyvin" arvioitua y:n arvo, kun $\mathbf{x}$ on saatu havainnoiduksi. Tällöin satunnaismuuttujaa $f(\mathbf{x})$ sanotaan y:n *prediktoriksi* ja sen realisoitunutta arvoa voidaan sanoa ennusteeksi. Erotusta $y - f(\mathbf{x})$ sanotaan ennustevirheeksi. Tuntuu luonnolliselta valita funktio f siten, että keskineliövirhe $\mathrm{MSE}[f(\mathbf{x}); y] = \mathrm{E}[y - f(\mathbf{x})]^2$ olisi mahdollisimman pieni. Tällaista prediktoria sanotaan parhaaksi prediktoriksi ja siitä käytetään merkintää $\mathrm{BP}(y; \mathbf{x})$. Täten $\mathrm{BP}(y; \mathbf{x})$:llä on ominaisuus

$$\min_{f(\mathbf{x})} \mathrm{E}[y - f(\mathbf{x})]^2 = \mathrm{E}[y - \mathrm{BP}(y; \mathbf{x})]^2. \tag{7.102}$$

On osoitettavissa, että

$$\spadesuit \text{ ehdollinen odotusarvo } \mathrm{E}(y \mid \mathbf{x}) \text{ on } y\text{:n paras prediktori}, \tag{7.103}$$

ts.

$$\mathrm{E}(y \mid \mathbf{x}) = \mathrm{BP}(y; \mathbf{x}) \,. \tag{7.104}$$

Tällöin on pidettävä $\mathrm{E}(y \mid \mathbf{x})$:tä satunnaismuuttujana, ei reaalilukuna. Ehdollinen odotusarvo $\mathrm{E}(y \mid \mathbf{x} = \underline{\mathbf{x}})$ on $\underline{\mathbf{x}}$:n funktio, ja voimme käyttää siitä merkintää $\mathrm{E}(y \mid \mathbf{x} = \underline{\mathbf{x}}) := m(\underline{\mathbf{x}})$. Kun $\underline{\mathbf{x}}$ korvataan $\mathbf{x}$:llä, $m(\underline{\mathbf{x}})$:stä saadaan satunnaismuuttuja $m(\mathbf{x})$, jota merkitään $\mathrm{E}(y \mid \mathbf{x})$. Tuloksen (7.104) todistus löytyy mm. seuraavista lähteistä: Christensen (2011, Theorem 6.3.4), Rao (1973, p. 264), ja Searle, Casella & McCulloch (1992, §7.2). Ks. myös harjoitustehtävä 7.1 (s. 250).

Tarkastellaan sitten *lineaarisia* (epähomogeenisia) prediktoreita:

$$\{\mathrm{LP}(y; \mathbf{x})\} = \{\, f(\mathbf{x}) : f(\mathbf{x}) = \mathbf{a}'\mathbf{x} + b,\ \mathbf{a} \in \mathbb{R}^p,\ b \in \mathbb{R} \,\} \,. \tag{7.105}$$

Silloin $\mathbf{x}$:ään perustuva y:n *paras lineaarinen prediktori* (BLP) on

$$\mathrm{BLP}(y; \mathbf{x}) = \mathbf{a}_*'\mathbf{x} + b_* \,, \tag{7.106}$$

jos se minimoi keskineliövirheen:

$$\begin{aligned} \min_{\mathbf{a}, b} \mathrm{MSE}(\mathbf{a}'\mathbf{x} + b; y) &= \min_{\mathbf{a}, b} \mathrm{E}[y - (\mathbf{a}'\mathbf{x} + b)]^2 \\ &= \mathrm{E}[y - (\mathbf{a}_*'\mathbf{x} + b_*)]^2. \end{aligned} \tag{7.107}$$

Havaitsemme heti, että keskineliövirhe voidaan esittää muodossa

$$\begin{aligned} \mathrm{E}[y - (\mathbf{a}'\mathbf{x} + b)]^2 &= \mathrm{E}[y - \mathbf{a}'\mathbf{x} - (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}}) + (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}} - b)]^2 \\ &= \mathrm{E}[y - \mathbf{a}'\mathbf{x} - (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}})]^2 + \mathrm{E}(\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}} - b)^2 \\ &\qquad + 2\,\mathrm{E}\big[\big(y - \mathbf{a}'\mathbf{x} - (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}})\big)(\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}} - b)\big] \\ &= \mathrm{var}(y - \mathbf{a}'\mathbf{x}) + (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}} - b)^2 \\ &= \text{“varianssi”} + \text{“harha}^2\text{”} \,. \end{aligned} \tag{7.108}$$

Termi "harha^2" saadaan nollaksi kun b:ksi valitaan

$$b_* = \mu_y - \mathbf{a}'\boldsymbol{\mu}_\mathbf{x}\,. \tag{7.109}$$

Olkoon (yksinkertaisuuden vuoksi) $\boldsymbol{\Sigma}$ positiivisesti definiitti. Tällöin lausekkeen (7.108) ensimmäinen termi $\operatorname{var}(y - \mathbf{a}'\mathbf{x})$ on (7.90):n (s. 226) nojalla pienimmillään kun $\mathbf{a}$:ksi valitaan

$$\mathbf{a}_* = \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}\,, \tag{7.110}$$

ja täten $\mathrm{BLP}(y;\mathbf{x})$ on

$$\begin{aligned}\mathrm{BLP}(y;\mathbf{x}) &= (\mu_y - \boldsymbol{\sigma}_{\mathbf{x}y}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_\mathbf{x}) + \boldsymbol{\sigma}_{\mathbf{x}y}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x} \\ &= \mu_y + \boldsymbol{\sigma}_{\mathbf{x}y}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})\,. \end{aligned} \tag{7.111}$$

Ennustevirhe on

$$e_{y\cdot\mathbf{x}} = y - \mathrm{BLP}(y;\mathbf{x}) = y - \mu_y - \boldsymbol{\sigma}_{\mathbf{x}y}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})\,, \tag{7.112}$$

ja ennustevirheen $e_{y\cdot\mathbf{x}}$ varianssiksi saadaan

$$\begin{aligned}\operatorname{var}(e_{y\cdot\mathbf{x}}) &= \sigma_y^2 - \boldsymbol{\sigma}_{\mathbf{x}y}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y} \\ &= \sigma_y^2\Big(1 - \frac{\boldsymbol{\sigma}_{\mathbf{x}y}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}}{\sigma_y^2}\Big) = \sigma_y^2(1 - \varrho_{y\cdot\mathbf{x}}^2)\,, \end{aligned} \tag{7.113}$$

missä $\varrho_{y\cdot\mathbf{x}}$ on populaation yhteiskorrelaatiokerroin:

$$\varrho_{y\cdot\mathbf{x}} = \operatorname{cor}[y, \mathrm{BLP}(y;\mathbf{x})] = \operatorname{cor}(y, \boldsymbol{\sigma}_{\mathbf{x}y}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x})\,. \tag{7.114}$$

Tarkastelleen yleisempää tilannetta, jossa

$$\mathbf{z} = \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_\mathbf{x}\\ \boldsymbol{\mu}_\mathbf{y}\end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}}\\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \boldsymbol{\Sigma}_{\mathbf{yy}}\end{pmatrix}. \tag{7.115}$$

Tällöin lineaarinen prediktori $\mathbf{A}_*\mathbf{x} + \mathbf{b}_*$ on $\mathbf{y}$:n paras lineaarinen prediktori, BLP, jos Löwnerin järjestys

$$\mathrm{MSEM}(\mathbf{A}_*\mathbf{x} + \mathbf{b}_*;\mathbf{y}) \leq_\mathsf{L} \mathrm{MSEM}(\mathbf{A}\mathbf{x} + \mathbf{b};\mathbf{y}) \tag{7.116}$$

on voimassa kaikilla lineaarisilla prediktoreilla $\mathbf{A}\mathbf{x} + \mathbf{b}$; edellä

$$\begin{aligned}\mathrm{MSEM}(\mathbf{A}\mathbf{x} + \mathbf{b};\mathbf{y}) &= \mathrm{E}[\mathbf{y} - (\mathbf{A}\mathbf{x} + \mathbf{b})][\mathbf{y} - (\mathbf{A}\mathbf{x} + \mathbf{b})]' \\ &= \text{ keskineliövirhematriisi.} \end{aligned} \tag{7.117}$$

Havaitsemme, että

$$\begin{aligned}\mathrm{E}[\mathbf{y}-&(\mathbf{A}\mathbf{x}+\mathbf{b})][\mathbf{y}-(\mathbf{A}\mathbf{x}+\mathbf{b})]' \\ &= \mathrm{E}[(\mathbf{y}-\mathbf{A}\mathbf{x})-\mathbf{b}][(\mathbf{y}-\mathbf{A}\mathbf{x})-\mathbf{b}]' \\ &= \mathrm{E}([(\mathbf{y}-\mathbf{A}\mathbf{x})-(\boldsymbol{\mu}_{\mathbf{y}}-\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}})]+[(\boldsymbol{\mu}_{\mathbf{y}}-\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}})-\mathbf{b}])(\cdots)' \\ &= \operatorname{cov}(\mathbf{y}-\mathbf{A}\mathbf{x})+(\boldsymbol{\mu}_{\mathbf{y}}-\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}}-\mathbf{b})(\boldsymbol{\mu}_{\mathbf{y}}-\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}}-\mathbf{b})'. \end{aligned} \tag{7.118}$$

On suoraviivaista päätellä (7.81):n (s. 225) nojalla, että $\mathbf{A}_* = \mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}$ ja $\mathbf{b}_* = \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{A}_*\boldsymbol{\mu}_{\mathbf{x}}$, joten

$$\begin{aligned}\mathrm{BLP}(\mathbf{y};\mathbf{x}) &= \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_{\mathbf{x}} + \mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x} \\ &= \boldsymbol{\mu}_{\mathbf{y}} + \mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}) . \end{aligned} \tag{7.119}$$

Ennustevirheeksi saadaan

$$\mathbf{e}_{\mathbf{y}\cdot\mathbf{x}} = \mathbf{y} - \mathrm{BLP}(\mathbf{y};\mathbf{x}) = \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}) , \tag{7.120}$$

ja

$$\operatorname{cov}(\mathbf{e}_{\mathbf{y}\cdot\mathbf{x}}) = \mathbf{\Sigma}_{\mathbf{yy}} - \mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{\Sigma}_{\mathbf{xy}} = \mathbf{\Sigma}_{\mathbf{yy}\cdot\mathbf{x}} . \tag{7.121}$$

Satunnaisvektoria $\mathrm{BLP}(\mathbf{y};\mathbf{x})$ sanotaan $\mathbf{y}$:n *regressioksi* $\mathbf{x}$*:n suhteen* ja vektoria $\mathbf{e}_{\mathbf{y}\cdot\mathbf{x}}$ vastaavaksi $\mathbf{y}$:n *residuaaliksi*:

$$\mathrm{BLP}(\mathbf{y};\mathbf{x}) = \mathbf{y}\text{:n regressio } \mathbf{x}\text{:n suhteen} , \tag{7.122a}$$

$$\mathbf{e}_{\mathbf{y}\cdot\mathbf{x}} = \mathbf{y}\text{:n residuaali } \mathbf{x}\text{:n eliminoinnin jälkeen} . \tag{7.122b}$$

Matriisi $\operatorname{cov}(\mathbf{e}_{\mathbf{y}\cdot\mathbf{x}}) = \mathbf{\Sigma}_{\mathbf{yy}\cdot\mathbf{x}}$ on $\mathbf{y}$:n osittaiskovarianssimatriisi $\mathbf{x}$:n eliminoinnin jälkeen; tätä ”eliminointi”-sanontaa ei parane ottaa kirjaimellisesti, sillä kyseessä ei välttämättä ole mikään varsinainen eliminointi. Matriisin

$$[\operatorname{diag}(\mathbf{\Sigma}_{\mathbf{yy}\cdot\mathbf{x}})]^{-1/2}\mathbf{\Sigma}_{\mathbf{yy}\cdot\mathbf{x}}[\operatorname{diag}(\mathbf{\Sigma}_{\mathbf{yy}\cdot\mathbf{x}})]^{-1/2} \tag{7.123}$$

elementit ovat osittaiskorrelaatiokertoimia:

$$\varrho_{ij\cdot\mathbf{x}} = \frac{\sigma_{ij\cdot\mathbf{x}}}{\sqrt{\sigma_{ii\cdot\mathbf{x}}\sigma_{jj\cdot\mathbf{x}}}} = \operatorname{cor}(e_{y_i\cdot\mathbf{x}}, e_{y_j\cdot\mathbf{x}}) , \tag{7.124}$$

missä

$$e_{y_i\cdot\mathbf{x}} = y_i - \mu_{y_i} - \boldsymbol{\sigma}'_{\mathbf{x}y_i}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}) , \quad \boldsymbol{\sigma}_{\mathbf{x}y_i} = \operatorname{cov}(\mathbf{x}, y_i) . \tag{7.125}$$

7.7 Muunnosmatriisina $\mathbf{\Sigma}^{-1/2}$

Ongelma NJ1. Olkoon n elementin satunnaisvektorin $\mathbf{y}$ positiivisesti definiitti kovarianssimatriisi $\operatorname{cov}(\mathbf{y}) = \mathbf{\Sigma}$. Millä muunnoksella

$$\mathbf{z} = \mathbf{B}\mathbf{y} \tag{7.126}$$

on sellainen ominaisuus, että $\mathbf{z}$:n komponentit ovat korreloimattomia ja jokaisella on sama varianssi 1 eli $\operatorname{cov}(\mathbf{z}) = \mathbf{I}_n$?

Yksi vastaus tähän kysymykseen on $\boldsymbol{\Sigma}$:n positiivisesti definiitin neliöjuuren käänteismatriisi eli kun

$$\boldsymbol{\Sigma} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}' = \mathbf{T}\boldsymbol{\Lambda}^{1/2}\mathbf{T}' \cdot \mathbf{T}\boldsymbol{\Lambda}^{1/2}\mathbf{T}' = \boldsymbol{\Sigma}^{1/2} \cdot \boldsymbol{\Sigma}^{1/2}, \tag{7.127}$$

niin

$$\mathbf{B} = \boldsymbol{\Sigma}^{-1/2} = \mathbf{T}\boldsymbol{\Lambda}^{-1/2}\mathbf{T}'. \tag{7.128}$$

Tällöin nimittäin

$$\begin{aligned} \operatorname{cov}(\mathbf{z}) = \operatorname{cov}(\mathbf{B}\mathbf{y}) = \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}' &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} \\ &= \mathbf{T}\boldsymbol{\Lambda}^{-1/2}\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T}\boldsymbol{\Lambda}^{-1/2}\mathbf{T}' \\ &= \mathbf{T}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1/2}\mathbf{T}' = \mathbf{T}\mathbf{T}' = \mathbf{I}_n\,. \end{aligned} \tag{7.129}$$

Huomaamme myös, että jos

$$\mathbf{F} = \boldsymbol{\Lambda}^{-1/2}\mathbf{T}', \tag{7.130}$$

niin satunnaisvektorin $\mathbf{w} = \mathbf{F}\mathbf{y}$ kovarianssimatriisi on

$$\begin{aligned} \operatorname{cov}(\mathbf{w}) = \operatorname{cov}(\mathbf{F}\mathbf{y}) = \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}' &= \boldsymbol{\Lambda}^{-1/2}\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T}\boldsymbol{\Lambda}^{-1/2} \\ &= \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1/2} = \mathbf{I}_n\,. \end{aligned} \tag{7.131}$$

Matriisi $\mathbf{F} = \boldsymbol{\Lambda}^{-1/2}\mathbf{T}'$ ei siis tietenkään ole välttämättä symmetrinen.

Voimme asettaa kysymyksen myös toisin päin:

Ongelma NJ2. Olkoon $\operatorname{cov}(\mathbf{z}) = \mathbf{I}_n$ ja olkoon $\boldsymbol{\Sigma}$ jokin annettu positiivisesti definiitti $n \times n$ matriisi. Mikä on se matriisi $\mathbf{G}$, jolla on ominaisuus

$$\operatorname{cov}(\mathbf{G}\mathbf{z}) = \mathbf{G}\mathbf{G}' = \boldsymbol{\Sigma}? \tag{7.132}$$

Huomaamme välittömästi, että sopiviksi $\mathbf{G}$-matriiseiksi kelpaavat

$$\mathbf{G}_1 = \mathbf{T}\boldsymbol{\Lambda}^{1/2}\mathbf{T}' = \boldsymbol{\Sigma}^{1/2}, \tag{7.133a}$$

$$\mathbf{G}_2 = \mathbf{T}\boldsymbol{\Lambda}^{1/2}. \tag{7.133b}$$

Esimerkki 7.5. Palautetaan mieleen sivulla 56 tarkasteltu esimerkki. Jos $\mathbf{x} \sim \mathrm{N}_2(\mathbf{0}, \mathbf{I}_2)$, ja merkitään

$$\begin{aligned} v_1 &= \sigma_1 x_1\,, \\ v_2 &= \sigma_2\big(\varrho x_1 + \sqrt{1-\varrho^2}x_2\big)\,, \end{aligned}$$

ts.

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2\varrho & \sigma_2\sqrt{1-\varrho^2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} := \mathbf{B}\mathbf{x}\,, \tag{7.134}$$

niin

$$\operatorname{cov}(\mathbf{v}) = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho \\ \sigma_1\sigma_2\varrho & \sigma_2^2 \end{pmatrix} := \mathbf{\Sigma}\,. \tag{7.135}$$

Tulos (7.135) seuraa siitä että

$$\mathbf{B}\mathbf{B}' = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2\varrho & \sigma_2\sqrt{1-\varrho^2} \end{pmatrix} \begin{pmatrix} \sigma_1 & \sigma_2\varrho \\ 0 & \sigma_2\sqrt{1-\varrho^2} \end{pmatrix} = \mathbf{\Sigma}\,. \tag{7.136}$$

□

Esimerkki 7.6. Olkoon $\mathbf{y}$ kahden elementin satunnaisvektori, jonka kovarianssimatriisi on $\operatorname{cov}(\mathbf{y}) = \mathbf{\Sigma}$ ja olkoon $\mathbf{\Sigma}$:n ominaisarvohajotelma $\mathbf{\Sigma} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}'$, jolloin $\mathbf{T}'\mathbf{\Sigma}\mathbf{T} = \mathbf{\Lambda}$, ts.

$$\begin{pmatrix} \mathbf{t}_1' \\ \mathbf{t}_2' \end{pmatrix} \mathbf{\Sigma}(\mathbf{t}_1 : \mathbf{t}_2) = \begin{pmatrix} \mathbf{t}_1'\mathbf{\Sigma}\mathbf{t}_1 & \mathbf{t}_1'\mathbf{\Sigma}\mathbf{t}_2 \\ \mathbf{t}_2'\mathbf{\Sigma}\mathbf{t}_1 & \mathbf{t}_2'\mathbf{\Sigma}\mathbf{t}_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}. \tag{7.137}$$

Jos nyt tehdään $\mathbf{y}$:lle muunnos (ortogonaalinen rotaatio) $\mathbf{z} = \mathbf{T}'\mathbf{y}$, niin

$$\operatorname{cov}(\mathbf{z}) = \operatorname{cov}(\mathbf{T}'\mathbf{y}) = \mathbf{T}'\mathbf{\Sigma}\mathbf{T} = \mathbf{\Lambda}\,, \tag{7.138}$$

eli

$$\operatorname{var}(z_i) = \operatorname{var}(\mathbf{t}_i'\mathbf{y}) = \mathbf{t}_i'\mathbf{\Sigma}\mathbf{t}_i = \lambda_i\,, \quad i = 1,2, \tag{7.139a}$$

$$\operatorname{cov}(z_1, z_2) = \mathbf{t}_1'\mathbf{\Sigma}\mathbf{t}_2 = 0\,. \tag{7.139b}$$

Toisin sanoen: uudet muuttujat $z_1 = \mathbf{t}_1'\mathbf{y}$ ja $z_2 = \mathbf{t}_2'\mathbf{y}$ ovat korreloimattomia ja niiden varianssit ovat $\mathbf{\Sigma}$:n ominaisarvot λ_1 ja λ_2. Itse asiassa osoittautuu, että $z_1 = \mathbf{t}_1'\mathbf{y}$ on se $\mathbf{y}$:n komponenttien lineaarikombinaatio $\mathbf{a}'\mathbf{y}$, jolla on suurin varianssi ehdolla $\mathbf{a}'\mathbf{a} = 1$:

$$\operatorname{ch}_1(\mathbf{\Sigma}_{n\times n}) = \lambda_1 = \max_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\mathbf{\Sigma}\mathbf{a} = \max_{\mathbf{a}\neq\mathbf{0}} \frac{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}{\mathbf{a}'\mathbf{a}}, \tag{7.140a}$$

$$\operatorname{ch}_n(\mathbf{\Sigma}_{n\times n}) = \lambda_n = \min_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\mathbf{\Sigma}\mathbf{a} = \min_{\mathbf{a}\neq\mathbf{0}} \frac{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}{\mathbf{a}'\mathbf{a}}. \tag{7.140b}$$

Kuvioiden 7.1 ja 7.2 toimituskentissä on tehty joitakin tähän esimerkkiin liittyviä laskelmia. □

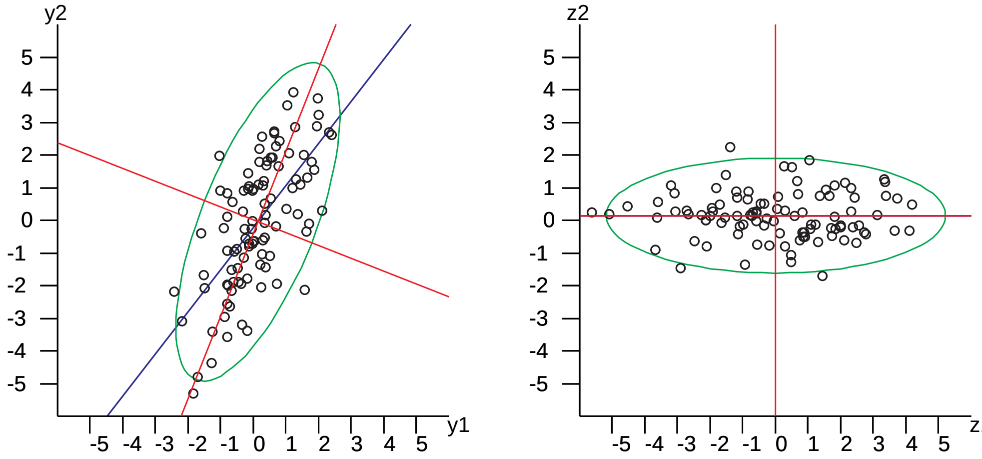
7.8 Faktorianalyysin malli

Faktorianalyysissä tarkastellaan satunnaismuuttujia $x_1, x_2, \dots, x_p$, joiden välillä uskotaan olevan "jotakin yhteistä": uskomme, että x-muuttujien keskinäisen korrelaation aiheuttajana on joitakin yhteisiä tekijöitä, joita faktorianalyysisissä kutsutaan *faktoreiksi* ja merkitään $f_1, f_2, \dots, f_r$. Uskomme, että muuttujan x_i arvo muodostuu seuraavan yhtälön mukaisesti:

$$x_i = \mu_i + a_{i1}f_1 + a_{i2}f_2 + \cdots + a_{ir}f_r + u_i\,, \quad i = 1,\dots,p\,, \tag{7.141a}$$

```
 -   - SURVO MM  Fri Oct 08 22:22:04 2010          C:\SJP\D\   2000  100 0
*
*Oheisesta pisteparvesta korrelaatio- ja kovarianssimatriisi ovat:
*MATRIX R
*///            Y1       Y2
*Y1       1.000000 0.665325
*Y2       0.665325 1.000000
*
*MATRIX S
*///            Y1       Y2
*Y1       1.054409 1.361496
*Y2       1.361496 3.971511
*
*
*  Muod. sellainen Y1:n ja Y2:n lin.komb. Z1, jolla max-varianssi
*                  Y1:n ja Y2:n lin.komb. Z2, jolla max-var ehdolla cor(Z
*  Sitä varten lasketaan S:n ominaisvektorit matriisiin T:
*MAT SPECTRAL DECOMPOSITION OF S TO T,L
*MAT LOAD T
*MATRIX T
*S(S)
*///            ev1      ev2
*Y1          0.36674  0.93033
*Y2          0.93033 -0.36674
*
*MAT LOAD L     /       tässä ovat S:n ominaisarvot
*MATRIX L
*L(S)
*///      eigenval
*ev1      4.508215
*ev2      0.517705
*
*  Suurinta ominaisarvoa lambda1=MAT_L(1,1) lambda1=4.508215487105   vast
*  ominaisvektori on t1 = T:n ensimmäinen sarake. lambda2=MAT_L(2,1) lamb
*  Tällöin uudet muuttujat z1 ja z2 saadaan seuraavasti:
*    z1 = t1'*y =  MAT_T(1,1)*y1+MAT_T(2,1)*y2
*    z2 = t2'*y =  MAT_T(1,2)*y1+MAT_T(2,2)*y2
*  eli z = T'*y mistä seuraa että z-muuttujien
*  havaintomatriisi U2 on: U2 = U1*T
*MAT U2=U1*T         / *U2~U1*S(S) 100*2
*MAT U2(0,1)="Z1"     / U2:n sarakkeille nimet Z1 ja Z2
*MAT U2(0,2)="Z2"
*CORR U2.MAT,CUR+1    /   VARS=Z1,Z2
*Means, std.devs and correlations of U2.MAT  N=100
*Variable  Mean        Std.dev.
*Z1        0.026301    2.123256
*Z2        0.157209    0.719517
*Correlations:
*                   Z1      Z2
* Z1            1.0000 -0.0000
* Z2           -0.0000  1.0000
*
*  var(z1) = MAT_MSN.M(1,2)^2=4.5082154871049  = lambda1=4.508215487105
*  var(z2) = MAT_MSN.M(2,2)^2=0.5177051530283  = lambda2=0.5177051530283
*         Uusien muuttujien varianssit ovat S:n om.arvoja
*
```

Kuvio 7.1. Esimerkkiin 7.6 liittyvä toimituskenttä. Määritetään uudet korreloimattomat muuttujat $z_1 = \mathbf{a}'\mathbf{y}$ ja $z_2 = \mathbf{b}'\mathbf{y}$ siten, että z_1:n varianssi on mahdollisimman suuri ehdolla $\mathbf{a}'\mathbf{a} = 1$.



Kuvio 7.2. Esimerkkiin 7.6 liittyviä kuvioita. Määritetään uudet korreloimattomat muuttujat $z_1 = \mathbf{a}'\mathbf{y}$ ja $z_2 = \mathbf{b}'\mathbf{y}$ siten, että z_1:n varianssi on mahdollisimman suuri ehdolla $\mathbf{a}'\mathbf{a} = 1$.

eli

$$x_i = \mu_i + \mathbf{a}'_{(i)}\mathbf{f} + u_i\,, \quad i = 1, \ldots, p\,, \tag{7.141b}$$

missä $\mathbf{a}'_{(i)} = (a_{i1}, a_{i2}, \ldots, a_{ir})$ ja $\mathbf{f}' = (f_1, f_2, \ldots, f_r)$. Matriisimuodossa voimme ilmaista (7.141):n seuraavasti:

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{f} + \mathbf{u}\,, \tag{7.142}$$

missä

$$\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}\,, \quad \mathrm{E}(\mathbf{f}) = \mathbf{0}\,, \quad \mathrm{E}(\mathbf{u}) = \mathbf{0}\,, \tag{7.143}$$

ja $p \times r$-matriisi $\mathbf{A}$ on faktorimatriisi:

$$\mathbf{A}_{p\times r} = \begin{pmatrix} \mathbf{a}'_{(1)} \\ \mathbf{a}'_{(2)} \\ \vdots \\ \mathbf{a}'_{(p)} \end{pmatrix} = (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_r)\,. \tag{7.144}$$

Satunnaisvektorin $\mathbf{f}$ elementit f_1, f_2, $\ldots$, f_r ovat (yhteis)faktoreita ja satunnaisvektorin $\mathbf{u}$ elementit u_1, u_2, $\ldots$, u_p ovat ominaisfaktoreita. Yleensä yhteisfaktoreita on huomattavasti vähemmän kuin ominaisfaktoreita eli $r < p$.

Muuttujien x_i ja f_i luonteissa on voimallinen ero: x_i:n arvot ovat havaittavissa mutta faktoreiden arvoja *ei* voida suoraan havaita tai mitata. Faktorit ovat hypoteettisia, teoreettisia muuttujia, eräänlaisia piilomuuttujia, joita ei konkreettisesti ole olemassa (kuten esim. ns. ”älykkyyden dimensiot”). Soveltajan tarkoituksena on faktorimatriisin elementtien (faktorilatausten) avulla tulkita mitä nämä faktorit ovat.

Malliin liittyvistä oletuksista mainittakoon vielä seuraavat:

(1) $\mathbf{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jolloin siis $\operatorname{cov}(\mathbf{x}) = \boldsymbol{\Sigma}$,

(2) $\mathbf{f} \sim \mathrm{N}_r(\mathbf{0}, \boldsymbol{\Phi})$,

(3) $\mathbf{u} \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Psi})$, missä $\operatorname{cov}(\mathbf{u}) = \boldsymbol{\Psi} = \operatorname{diag}(\psi_1^2, \dots, \psi_p^2)$,

(4) $\operatorname{cov}(\mathbf{f}, \mathbf{u}) = \mathrm{E}(\mathbf{f}\mathbf{u}') = \mathbf{0}$ eli $\mathbf{f}$ ja $\mathbf{u}$ ovat korreloimattomia,

(5) $\operatorname{rank}(\mathbf{A}) = r$.

Oletuksista seuraa (7.39):n mukaisesti ($\mathbf{Af}$:n ja $\mathbf{u}$:n korreloimattomuuden nojalla):

$$\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{x}) = \operatorname{cov}(\mathbf{Af} + \mathbf{u}) = \operatorname{cov}(\mathbf{Af}) + \operatorname{cov}(\mathbf{u}) = \mathbf{A}\boldsymbol{\Phi}\mathbf{A}' + \boldsymbol{\Psi}. \tag{7.145}$$

Yhtälöä (7.145) sanotaan *faktorianalyysin perusyhtälöksi.*

Mikäli faktorit oletetaan korreloimattomiksi ja variansseiltaan ykkösiksi eli $\operatorname{cov}(\mathbf{f}) = \mathbf{I}_r = \operatorname{cor}(\mathbf{f})$, niin faktorianalyysin perusyhtälö on

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}' + \boldsymbol{\Psi}. \tag{7.146}$$

Tällöin $\mathbf{A}$:n elementit ovat faktorien ja muuttujien välisiä kovariansseja, sillä

$$\begin{aligned} \operatorname{cov}(\mathbf{x}, \mathbf{f}) &= \mathrm{E}[(\mathbf{x} - \boldsymbol{\mu})\mathbf{f}'] \\ &= \mathrm{E}[(\mathbf{Af} + \mathbf{u})\mathbf{f}'] = \mathrm{E}(\mathbf{Aff}') + \mathrm{E}(\mathbf{uf}') = \mathbf{A}\,\mathrm{E}(\mathbf{ff}') = \mathbf{A}\,. \end{aligned} \tag{7.147}$$

Erityisesti jos lisäksi $\boldsymbol{\Sigma} = \operatorname{cor}(\mathbf{x})$, niin

$$\mathbf{A} = \operatorname{cor}(\mathbf{x}, \mathbf{f}) = \{\operatorname{cor}(x_i, f_j)\}. \tag{7.148}$$

Olkoon $\boldsymbol{\Sigma}$:n estimaatiksi otoksesta (standardoiduista muuttujista) saatu korrelaatiomatriisi $\mathbf{R}$. Miten estimoida faktorimatriisi $\mathbf{A}$? Yksi lähtökohta on arvioida ensin jotenkin $\boldsymbol{\Psi}$ ja sitten määrittää $\mathbf{A}$ siten, että $\|(\mathbf{R} - \boldsymbol{\Psi}) - \mathbf{AA}'\|$ minimoituu. Ratkaisuksi $\mathbf{A}$ saadaan $(\mathbf{R} - \boldsymbol{\Psi})$:n r ensimmäistä ominaisvektoria.

Emme puutu faktorianalyysiin sen kummemmin – lukijalle on jo varmaan selvinnyt matriisilaskennan tärkeä rooli tässä yhteydessä.

7.9 Jakaumista

Luvussa 1 sivulla 54 jo tarkastelimme satunnaisvektoria $\mathbf{z}$, joka noudattaa p-ulotteista normaalijakaumaa parametrein $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ eli $\mathbf{z} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Jos $\boldsymbol{\Sigma}$ on positiivisesti definiitti, $\mathbf{z}$:n tiheysfunktio on

$$n(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{1}{2}p} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}. \tag{7.149}$$

Kun $n = 2$, on tiheysfunktion lauseke

$$n(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho_{12}^2}} \cdot \exp\biggl[\frac{-1}{2(1-\varrho_{12}^2)}\biggl(\frac{(z_1-\mu_1)^2}{\sigma_1^2} - 2\varrho_{12}\frac{(z_1-\mu_1)(z_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(z_2-\mu_2)^2}{\sigma_2^2}\biggr)\biggr]. \tag{7.150}$$

Jos $\mathbf{z}$:n komponentit z_1 ja z_2 ovat korreloimattomia eli $\varrho_{12} = 0$, tiheysfunktion lauseke voidaan kirjoittaa kahden yksiulotteisen normaalijakauman tiheysfunktion tulona,

$$n(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\Sigma}) = n(z_1;\mu_1,\sigma_1^2)\, n(z_2;\mu_2,\sigma_2^2)\,, \tag{7.151}$$

mikä merkitsee, että z_1 ja z_2 ovat riippumattomia satunnaismuuttujia. Multinormaalijakauman tapauksessa korreloimattomuus ja riippumattomuus ovat siis yhtäpitäviä ominaisuuksia – tunnetustihan ei muuten näin välttämättä ole.

Multinormaalijakauman ominaisuuksiin emme kovin paljoa tässä puutu; lähteinä mainittakoon esim. Rao (1973, pp. 525–528), ja Seber (2008, §20.5). Seuraavassa kuitenkin muutama tärkeä tulos:

NJ1. $\mathbf{z} \sim \mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma}) \iff \mathbf{t}'\mathbf{z} \sim \mathrm{N}_1(\mathbf{t}'\boldsymbol{\mu},\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})\ \forall\,\mathbf{t} \in \mathbb{R}^p$.

NJ2. $\mathbf{z} \sim \mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma}) \implies \mathbf{A}\mathbf{z} \sim \mathrm{N}_p(\mathbf{A}\boldsymbol{\mu},\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

NJ3. $\mathbf{z} \sim \mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma}) \implies$ $\mathbf{z}$:n elementtien reunajakaumat ovat $\mathrm{N}_1(\mu_i,\sigma_i^2)$.

NJ4. $\mathbf{z} \sim \mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\mathbf{y} \sim \mathrm{N}_p(\boldsymbol{\nu},\boldsymbol{\Gamma})$ ja $\mathbf{z}$ ja $\mathbf{y}$ riippumattomia

$$\implies a\mathbf{z} + b\mathbf{y} \sim \mathrm{N}_p(a\boldsymbol{\mu} + b\boldsymbol{\nu}, a^2\boldsymbol{\Sigma} + b^2\boldsymbol{\Gamma}).$$

NJ5. Olkoot $\mathbf{z}$ ja $\mathbf{y}$ riippumattomia p-ulotteisia satunnaismuuttujia. Tällöin

$$\mathbf{z} + \mathbf{y} \sim \mathrm{N}_p \implies \mathbf{z} \sim \mathrm{N}_p \text{ ja } \mathbf{y} \sim \mathrm{N}_p\,.$$

NJ6. Olkoon $\mathbf{z}$ ositettu seuraavasti:

$$\mathbf{z} = \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}, \quad \mathrm{E}(\mathbf{z}) = \begin{pmatrix}\boldsymbol{\mu}_\mathbf{x}\\ \boldsymbol{\mu}_\mathbf{y}\end{pmatrix}, \quad \mathrm{cov}(\mathbf{z}) = \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}}\\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \boldsymbol{\Sigma}_{\mathbf{yy}}\end{pmatrix}.$$

Tällöin $\mathbf{y}$:n ehdollinen jakauma, kun $\mathbf{x}$:llä on kiinteä arvo $\mathbf{x} = \underline{\mathbf{x}}$, on normaalijakauma parametrein

$$\mathrm{E}(\mathbf{y} \mid \underline{\mathbf{x}}) = \boldsymbol{\mu}_\mathbf{y} + \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\underline{\mathbf{x}} - \boldsymbol{\mu}_\mathbf{x})\,,$$
$$\mathrm{cov}(\mathbf{y} \mid \underline{\mathbf{x}}) = \boldsymbol{\Sigma}_{\mathbf{yy}\cdot\mathbf{x}} = \boldsymbol{\Sigma}_{\mathbf{yy}} - \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xy}}\,.$$

NJ7. Erityisesti jos

$$\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} \sim \mathrm{N}_{p+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \mu_y \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}'_{\mathbf{x}y} & \sigma_y^2 \end{pmatrix},$$

niin

$$\begin{aligned} \mathrm{E}(y \mid \underline{\mathbf{x}}) &= \mu_y + \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\underline{\mathbf{x}} - \boldsymbol{\mu}_{\mathbf{x}}), \\ \mathrm{var}(y \mid \underline{\mathbf{x}}) &= \sigma_{y\cdot\mathbf{x}}^2 = \sigma_y^2 - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y} = \sigma_y^2\Big(1 - \frac{\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}}{\sigma_y^2}\Big) \\ &= \frac{1}{\sigma^{yy}} = \sigma_y^2(1 - \varrho_{y\cdot\mathbf{x}}^2) \le \sigma_y^2 = \mathrm{var}(y), \end{aligned}$$

missä σ^{yy} on $\boldsymbol{\Sigma}$:n käänteismatriisin viimeinen lävistäjäelementti.

NJ8. Kahden muuttujan tapauksessa saadaan

$$\begin{aligned} \mathrm{E}(y \mid \underline{x}) &= \mu_y + \varrho_{xy}\frac{\sigma_y}{\sigma_x}(\underline{x} - \mu_x) = \mu_y + \frac{\sigma_{xy}}{\sigma_x^2}\,(\underline{x} - \mu_x) \\ &= \Big(\mu_y - \frac{\sigma_{xy}}{\sigma_x^2}\,\mu_x\Big) + \frac{\sigma_{xy}}{\sigma_x^2}\,\underline{x}, \\ \mathrm{var}(y \mid \underline{x}) &= \sigma_{y\cdot x}^2 = \sigma_y^2(1 - \varrho_{xy}^2) \le \sigma_y^2 = \mathrm{var}(y). \end{aligned}$$

Kannattaa erityisesti panna merkille, että multinormaalijakauman tapauksessa ehdollinen varianssi on riippumaton annetusta $\underline{x}$:n (tai $\underline{\mathbf{x}}$:n) arvosta ja kahden muuttujan tilanteessa ehdollinen odotusarvo muodostaa suoran, jonka kulmakerroin on $\frac{\sigma_{xy}}{\sigma_x^2}$ ja joka kulkee pisteen (μ_x, μ_y) kautta. Tässä on mielenkiintoinen yhteys otoksesta laskettuun regressiosuoraan, sillä sen kulmakerroin on $\frac{s_{xy}}{s_x^2}$ ja se kulkee pisteen $(\bar{x}, \bar{y})$ kautta.

7.9.1 Tasa-arvokäyrät

Olkoon $a > 0$ jokin annettu reaaliluku. Tarkastellaan niiden pisteiden (vektoreiden) $\mathbf{z} \in \mathbb{R}^2$ joukkoa, jotka toteuttavat ehdon

$$n(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\mathrm{e}^{(\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})} = a^2. \tag{7.152}$$

Ehto (7.152) tarkoittaa siis sellaisten pisteiden $\mathbf{z}$ joukkoa, joissa normaalijakauman tiheysfunktiolla on vakioarvo a^2. On tietenkin selvää (ota logaritmit puolittain), että tällaisten pisteiden joukko on

$$\mathcal{A} = \{\, \mathbf{z} \in \mathbb{R}^2 : (\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) = c^2 \,\}, \tag{7.153}$$

missä $c^2 = -2\log(2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}a^2)$. Joukon $\mathcal{A}$ pisteet muodostavat *tasa-arvokäyrän* (*contour*), mikä tässä tapauksessa on $\boldsymbol{\mu}$-keskinen ellipsi.

Sivulla 172 tarkasteltiin pääakseleiden suuntia ja pituuksia. Ensimmäinen pääakseli on pisin $\boldsymbol{\mu}$:n kautta kulkeva jana, jonka päätepisteet ovat ellipsin $\mathcal{A}$ kehällä eli etsimme pistettä (vektoria) $\mathbf{z}_1 \in \mathcal{A}$, jolla on maksimaalinen etäisyys $\boldsymbol{\mu}$:stä. Vastaavasti toinen pääakseli määräytyy sen pisteen mukaan, jolla on lyhin etäisyys $\boldsymbol{\mu}$:stä.

Ellipsin $\mathcal{A}$ pääakselien suunnat ja pituudet määräytyvät $\boldsymbol{\Sigma}^{-1}$:n ominaisvektorien ja ominaisarvojen perusteella; $\boldsymbol{\Sigma}^{-1}$:n ominaisarvojen ja -vektoreiden sijasta voimme tietysti tarkastella $\boldsymbol{\Sigma}$:n ominaisarvoja ja -vektoreita; matriiseilla $\boldsymbol{\Sigma}$ ja $\boldsymbol{\Sigma}^{-1}$ on samat ominaisvektorit ja ominaisarvot ovat toistensa käänteislukuja.

Jos $\mathrm{ch}_1(\boldsymbol{\Sigma}) = \lambda_1$ ja $\mathrm{ch}_2(\boldsymbol{\Sigma}) = \lambda_2$ ja $\lambda_1 \geq \lambda_2 > 0$, ja vastaavat ortonormaalit ominaisvektorit ovat $\mathbf{t}_1$ ja $\mathbf{t}_2$, niin $\mathcal{A}$:n mukaisella ellipsillä on seuraavat ominaisuudet:

E1. Ellipsin keskipiste on $\boldsymbol{\mu}$.

E2. Pääakselien pituudet ovat $2c\sqrt{\lambda_i}$.

E3. Jos $\lambda_1 > \lambda_2$, niin i. pääakseli kulkee $\boldsymbol{\mu}$-keskisesti siten että sillä on sama suunta kuin (origosta lähtevällä) vektorilla $\mathbf{t}_i$. Jos $\lambda_1 = \lambda_2$, niin ominaisvektorit eivät ole yksikäsitteisesti määriteltyjä (mikä hyvänsä ortonormaali vektoripari $\{\mathbf{t}_1, \mathbf{t}_2\}$ kelpaa) ja siten pääakselien suunnatkaan eivät ole yksikäsitteiset (kyseessä ympyrä).

E4. Ellipsin pinta-ala on verrannollinen lukuun $c^2 \det(\boldsymbol{\Sigma}) = c^2\lambda_1\lambda_2$.

Useampiulotteisessa tapauksessa tasa-arvokäyristä muodostuu ellipsoidin pinta.

7.9.2 Mahalanobisin etäisyys

Olkoon $\mathbf{z}$ p-ulotteinen satunnaisvektori ja $\mathrm{E}(\mathbf{z}) = \boldsymbol{\mu}$, $\mathrm{cov}(\mathbf{z}) = \boldsymbol{\Sigma}$. Tällöin lauseketta

$$(\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) = \mathrm{MHLN}^2(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{7.154}$$

sanotaan satunnaisvektorin $\mathbf{z}$ *Mahalanobisin etäisyydeksi* odotusarvovektorista $\boldsymbol{\mu}$ (neliöitynä). Korostettakoon, että (7.154):ssa emme aina välttämättä pidä $\mathbf{z}$:aa satunnaisvektorina, vaan tulkitsemme (7.154):n yksinkertaisesti tiettynä $\mathbb{R}^p$:n vektorien $\mathbf{z}$ ja $\boldsymbol{\mu}$ sekä matriisin $\boldsymbol{\Sigma} \in \mathrm{PD}_p$ funktiona.

Yksiulotteisessa tapauksessa ($p = 1$) saamme

$$\mathrm{MHLN}(z, \mu, \sigma^2) = \frac{z - \mu}{\sigma}\,, \tag{7.155}$$

eli $\mathrm{MHLN}(z, \mu, \sigma^2)$ kertoo kuinka monta hajonnan mittaa z poikkeaa μ:sta. Jos esimerkiksi

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}, \tag{7.156}$$

niin

$$\begin{aligned}\mathrm{MHLN}^2(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= (\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu}) = (\mathbf{z}-\boldsymbol{\mu})' \begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix} (\mathbf{z}-\boldsymbol{\mu}) \\ &= \frac{(z_1-\mu_1)^2}{\sigma_1^2} + \frac{(z_2-\mu_2)^2}{\sigma_2^2} . \end{aligned} \tag{7.157}$$

Täten $\mathrm{MHLN}^2(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ on tässä tapauksessa sitä suurempi mitä useamman hajonnan mitan z_1 poikkeaa μ_1:stä ja z_2 poikkeaa μ_2:sta ja siten Mahalanobisin etäisyys on luonteva mitta tilastolliselle etäisyydelle. Jos $\boldsymbol{\Sigma} = \left(\begin{smallmatrix} 1 & \varrho \\ \varrho & 1 \end{smallmatrix}\right)$, niin

$$\begin{aligned}\mathrm{MHLN}^2(\mathbf{z}, \mathbf{0}, \boldsymbol{\Sigma}) = \mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z} &= \frac{1}{1-\varrho^2}(z_1^2 + z_2^2 - 2\varrho z_1 z_2) \\ &= \frac{1}{1-\varrho^2}(\|\mathbf{z}\|^2 - 2\varrho z_1 z_2) . \end{aligned} \tag{7.158}$$

Olkoon $\varrho > 0$. Tällöin esimerkiksi vektoreilla $\mathbf{a} = (1,1)'$ ja $\mathbf{b} = (-1,1)'$ on sama euklidinen pituus $\sqrt{2}$, mutta positiivisesta ϱ:sta johtuen niillä on erisuuret Mahalanobisin etäisyydet origosta; ks. Esimerkki 1.3 (s. 27).

Positiivisesti definiitin $\boldsymbol{\Sigma}$:n avulla voimme määritellä normin siten että

$$\|\mathbf{a}\|^2_{\boldsymbol{\Sigma}^{-1}} = \mathbf{a}'\boldsymbol{\Sigma}^{-1}\mathbf{a} , \tag{7.159}$$

jolloin Mahalanobisin etäisyys (neliöitynä) on

$$\|\mathbf{z}-\boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}^{-1}} = (\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu}) . \tag{7.160}$$

Samalla tasa-arvokäyrällä (tai -pinnalla) olevat $\mathbf{z}$:n arvot ovat tällöin siis normin (7.159) mielessä (eli Mahalanobisin etäisyyden mielessä) yhtä kaukana $\boldsymbol{\mu}$:stä.

Muistutettakoon lukijaa siitä, että jo sivulla 27 määriteltiin havaintovektorin $\mathbf{u}_{(i)}$ Mahalanobisin etäisyys (neliöitynä) keskiarvovektorista $\bar{\mathbf{u}}$ lausekkeena

$$\mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S}) = (\mathbf{u}_{(i)} - \bar{\mathbf{u}})'\mathbf{S}^{-1}(\mathbf{u}_{(i)} - \bar{\mathbf{u}}) = \|\mathbf{u}_{(i)} - \bar{\mathbf{u}}\|^2_{\mathbf{S}^{-1}} , \tag{7.161}$$

missä $\mathbf{U}' = (\mathbf{u}_{(1)} : \ldots : \mathbf{u}_{(n)})$ on havaintomatriisin transpoosi, ja

$$\bar{\mathbf{u}} = \frac{1}{n}\mathbf{U}'\mathbf{1}_n = \frac{1}{n}(\mathbf{u}_{(1)} + \cdots + \mathbf{u}_{(n)}) , \quad \mathbf{S} = \frac{1}{n-1}\mathbf{U}'(\mathbf{I}-\mathbf{J})\mathbf{U} . \tag{7.162}$$

Havaintojen $\mathbf{u}_{(i)}$ ja $\mathbf{u}_{(j)}$ Mahalanobisin etäisyys (neliöitynä) on

$$\mathrm{MHLN}^2(\mathbf{u}_{(i)}, \mathbf{u}_{(j)}, \mathbf{S}) = (\mathbf{u}_{(i)} - \mathbf{u}_{(j)})'\mathbf{S}^{-1}(\mathbf{u}_{(i)} - \mathbf{u}_{(j)}) . \tag{7.163}$$

Esimerkki 7.7. Jos $\operatorname{cov}(\mathbf{z}) = \left(\begin{smallmatrix} 1 & \varrho \\ \varrho & 1 \end{smallmatrix}\right) = \mathbf{\Sigma}$, niin $\mathbf{\Sigma}$:n ominaisarvot ovat $\lambda_1 = 1 + \varrho$, $\lambda_2 = 1 - \varrho$, ja vastaavat ortonormaalit ominaisvektorit $\mathbf{t}_1 = \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$, $\mathbf{t}_2 = \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)$ Jos nyt $\mathbf{z} \sim \mathrm{N}_2(\boldsymbol{\mu}, \mathbf{\Sigma})$, niin tasa-arvokäyrällä eli ellipsillä

$$\mathcal{B} = \{\, \mathbf{z} \in \mathbb{R}^2 : (\mathbf{z} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) = c^2 \,\} \tag{7.164}$$

on ominaisuudet:

E1. Ellipsin keskipiste on $\boldsymbol{\mu}$.

E2. Pääakselien pituudet ovat $2c\sqrt{1+\varrho}$ ja $2c\sqrt{1-\varrho}$.

E3. Jos $\varrho > 0$, niin $\operatorname{ch}_{\max}(\mathbf{\Sigma}) = \lambda_1 = 1 + \varrho$. Tällöin 1. pääakseli kulkee $\boldsymbol{\mu}$:n kautta siten, että sillä on sama suunta kuin vektorilla $\mathbf{t}_1$ eli kulmakerroin on 1 (2. pääakselin kulmakerroin on -1).

Jos $\varrho < 0$, niin $\operatorname{ch}_{\max}(\mathbf{\Sigma}) = \lambda_2 = 1 - \varrho$ ja 1. pääakseli kulkee vektorin $\mathbf{t}_2$ suuntaisesti eli kulmakerroin on -1 (2. pääakselin kulmakerroin on 1). Jos $\varrho = 0$, niin kyseessä on ympyrä koska varianssit yhtä suuria.

E4. Ellipsin pinta-ala on verrannollinen lukuun $c^2(1+\varrho)(1-\varrho) = c^2(1-\varrho^2)$ eli mitä suurempi on ϱ^2 sitä pienempi on ellipsin pinta-ala.

Kuviossa 7.3 on piirretty ellipsit $(\mathbf{z} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) = c^2$, kun $c = 1, 2, 3$, ja $\mathbf{\Sigma}$:ssa $\varrho = 0.6$ ja 0.9. Toimituskentässä on käytetty hyväksi tietoa

$$(\mathbf{z} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) \sim \chi^2(2)\,, \text{ jos } \mathbf{z} \sim \mathrm{N}_2(\boldsymbol{\mu}, \mathbf{\Sigma})\,. \tag{7.165}$$

Ellipsi on piirrettävissä Survon `BINORM`-määritteen avulla, jossa parametreina ovat muuttujien odotusarvot, hajonnat ja korrelaatiokerroin:

$$\texttt{BINORM=mean1,mean2,s1,s2,kor}. \tag{7.166}$$

`CONTOUR`-määritteessä annetaan parametreiksi χ^2-jakauman mukaiset konfidenssitasot $1 - \alpha$ (arvolla 0 saadaan pääakselit). Esimerkiksi voimme kirjoittaa `CONTOUR=0.95`, jolloin $\alpha = 0.05$. Survo piirtää tällöin ellipsin

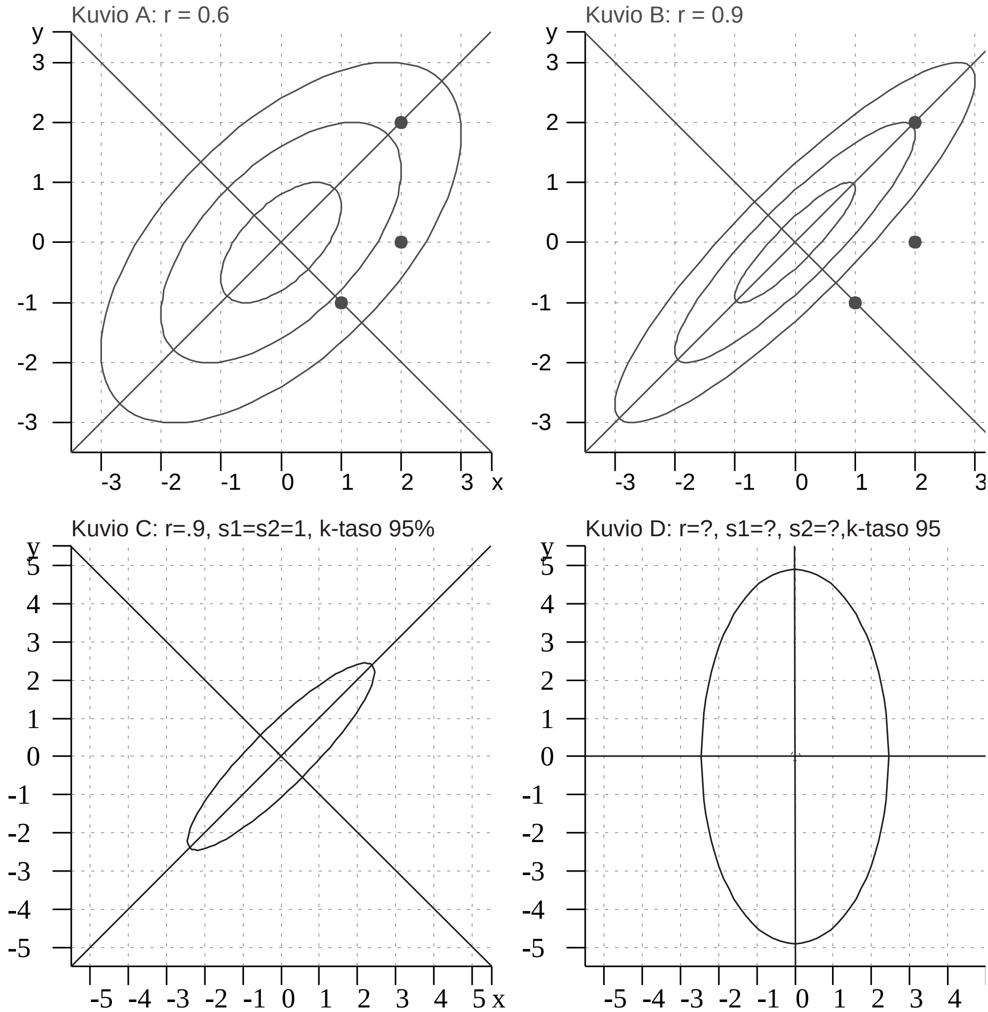
$$(\mathbf{z} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) = \chi^2_{\alpha,2}\,, \tag{7.167}$$

missä $\chi^2_{\alpha,2}$ on $\chi^2(2)$-jakauman kriittinen arvo riskitasolla α eli

$$\mathrm{P}\big[(\mathbf{z} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) > \chi^2_{\alpha,2}\big] = \alpha\,, \tag{7.168a}$$

$$\mathrm{P}\big[(\mathbf{z} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) \le \chi^2_{\alpha,2}\big] = 1 - \alpha\,. \tag{7.168b}$$

Koska nyt $c^2 = \chi^2_{\alpha,2}$, voimme määrittää α:n $\chi^2(2)$-jakauman kertymäfunktion avulla. Esimerkiksi kun $c = 2$, meillä on α:n ratkaisemiseksi yhtälö $4 = \chi^2_{\alpha,2}$: $\chi^2(2)$-jakauman kertymäfunktion arvo pisteessä 4 on $1 - \alpha$. Jos Survossa määritellään `a=Chi2.F(df,4)`, niin `a`:n arvoksi saadaan χ^2(`df`)-jakauman kertymäfunktion arvo pisteessä 4 eli `a` on juuri etsimämme $1 - \alpha$. $\square$



Kuvio 7.3. Esimerkkiin 7.7 liittyviä kuvioita. Kuvioihin C ja D on piirretty 95 %:n konfidenssiellipsit tietyillä $\mathbf{\Sigma}$:n arvoilla.

7.9.3 χ^2- ja F-jakauma

Emme puutu jakaumiin sen tarkemmin – kopioimme vain muutaman määritelmän Kaavakokoelmasta. Lähteistä mainittakoon Seber (2008, §20.5).

1. Keskinen χ^2-jakauma: $\mathbf{z} \sim \mathrm{N}_p(\mathbf{0}, \mathbf{I}_p)$: $\mathbf{z}'\mathbf{z} = \chi^2_p \sim \chi^2(p)$.
2. Epäkeskinen χ^2-jakauma: $\mathbf{z} \sim \mathrm{N}_p(\boldsymbol{\mu}, \mathbf{I}_p)$: $\mathbf{z}'\mathbf{z} = \chi^2_{p,\delta} \sim \chi^2(p, \delta)$, $\delta = \boldsymbol{\mu}'\boldsymbol{\mu}$.
3. Epäkeskinen F-jakauma: $F = \frac{\chi^2_{p,\delta}/p}{\chi^2_q/q} \sim \mathrm{F}(p, q, \delta)$, missä $\chi^2_{p,\delta}$ and χ^2_q ovat riippumattomia.
4. t-jakauma: $\mathrm{t}^2(p) = \mathrm{F}(1, p)$.
5. Olkoon $\mathbf{z} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, missä $\boldsymbol{\Sigma}$ on positiivisesti definiitti, ja olkoot $\mathbf{A}$ ja $\mathbf{B}$ symmetrisiä ja $\mathbf{b} \in \mathbb{R}^p$. Silloin (i)–(v) ovat voimassa:

   (i) $\mathbf{z}'\mathbf{A}\mathbf{z} \sim \chi^2(r, \delta) \iff \mathbf{A}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{A}$, jolloin $r = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) = \mathrm{r}(\mathbf{A}\boldsymbol{\Sigma})$, $\delta = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$,

   (ii) $\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z} = \mathbf{z}'[\mathrm{cov}(\mathbf{z})]^{-1}\mathbf{z} \sim \chi^2(r, \delta)$, missä $r = p$, $\delta = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$,

   (iii) $(\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) \sim \chi^2(p)$,

   (iv) $\mathbf{z}'\mathbf{A}\mathbf{z}$ ja $\mathbf{z}'\mathbf{B}\mathbf{z}$ ovat riippumattomia $\iff \mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$,

   (v) $\mathbf{z}'\mathbf{A}\mathbf{z}$ ja $\mathbf{b}'\mathbf{z}$ riippumattomia $\iff \mathbf{A}\boldsymbol{\Sigma}\mathbf{b} = \mathbf{0}$.

   Olkoon $\mathbf{z} \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, missä $\boldsymbol{\Sigma}$ on mahdollisesti singulaarinen. Tällöin

   (vi) $\mathbf{z}'\mathbf{A}\mathbf{z} \sim \chi^2(r) \iff \boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}$, jolloin $r = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) = \mathrm{r}(\mathbf{A}\boldsymbol{\Sigma})$,

   (vii) $\mathbf{z}'\mathbf{A}\mathbf{z}$ ja $\mathbf{x}'\mathbf{B}\mathbf{x}$ ovat riippumattomia $\iff \boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma} = \mathbf{0}$,

   (viii) $\mathbf{z}'\boldsymbol{\Sigma}^{-}\mathbf{z} = \mathbf{z}'[\mathrm{cov}(\mathbf{z})]^{-}\mathbf{z} \sim \chi^2(r)$ missä $\mathrm{r}(\boldsymbol{\Sigma}) = r$ ja $\boldsymbol{\Sigma}^{-}$ on mielivaltainen $\boldsymbol{\Sigma}$:n yleistetty käänteismatriisi.

Edellä satunnaismuuttujien riippumattomuus tarkoittaa tietenkin niiden tilastollista riippumattomuutta: satunnaisvektorit $\mathbf{x}$ ja $\mathbf{y}$ ovat tilastollisesti riippumattomia jos ja vain jos satunnaisvektorin $\binom{\mathbf{x}}{\mathbf{y}}$ tiheysfunktio (tai yhtäpitävästi kertymäfunktio) on $\mathbf{x}$:n ja $\mathbf{y}$:n tiheysfunktioiden (vastaavasti kertymäfunktioiden) tulo. Intuitiivisesti tämä tarkoittaa, että tieto $\mathbf{x}$:n arvosta ei mitenkään anna informaatiota $\mathbf{y}$:n arvosta. Esimerkiksi jos x ja y ovat diskreettejä satunnaismuuttujia, joiden arvot ovat $x_1, \dots, x_r$ ja $y_1, \dots, y_c$, niin x ja y ovat tilastollisesti riippumattomia jos ja vain jos

$$\mathrm{P}(x = x_i,\, y = y_j) = \mathrm{P}(x = x_i)\,\mathrm{P}(y = y_j)\,, \; i = 1, \dots, r,\; j = 1, \dots, c. \quad (7.169)$$

Vielä pari sanaa Wishart-jakaumasta ja Hotellingin T^2:sta.

1. Wishart-jakauma. Olkoon $\mathbf{U}' = (\mathbf{u}_{(1)} : \ldots : \mathbf{u}_{(n)})$ satunnaisotos jakaumasta $\mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, ts. $\mathbf{u}_{(i)}$:t ovat riippumattomia ja kukin $\mathbf{u}_{(i)} \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. Silloin satunnaismatriisi $\mathbf{W} = \mathbf{U}'\mathbf{U} = \sum_{i=1}^n \mathbf{u}_{(i)}\mathbf{u}'_{(i)}$ noudattaa Wishart-jakaumaa parametrein n ja $\boldsymbol{\Sigma}$: $\mathbf{W} \sim \mathrm{W}_p(n, \boldsymbol{\Sigma})$.
2. Olkoon $\mathbf{U}' = (\mathbf{u}_{(1)} : \ldots : \mathbf{u}_{(n)})$ satunnaisotos $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:sta. Silloin $\bar{\mathbf{u}} = \frac{1}{n}\mathbf{U}'\mathbf{1}_n$ ja $\mathbf{T} = \mathbf{U}'(\mathbf{I} - \mathbf{J})\mathbf{U}$ ovat riippumattomia ja $\mathbf{T} \sim \mathrm{W}_p(n-1, \boldsymbol{\Sigma})$.
3. Hotellingin T^2-jakauma. Oletetaan että $\mathbf{W} \sim \mathrm{W}_p(m, \boldsymbol{\Sigma})$, $\mathbf{v} \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, $\mathbf{v}$ ja $\mathbf{W}$ ovat riippumattomia, ja että $\mathbf{W}$ on positiivisesti definiitti. Hotellingin T^2-jakaumaksi sanotaan satunnaismuuttujan

$$T^2 = m \cdot \mathbf{v}'\mathbf{W}^{-1}\mathbf{v} = \mathbf{v}'\big(\tfrac{1}{m}\mathbf{W}\big)^{-1}\mathbf{v} \tag{7.170}$$

$$= (\text{normal r.v.})' \left(\frac{\text{Wishart}}{\text{df}}\right)^{-1} (\text{normal r.v.}) \tag{7.171}$$

jakaumaa: $T^2 \sim \mathrm{T}^2(p, m)$.

Esimerkki 7.8. Olkoon $\mathbf{y}$ normaalisti jakautunut satunnaisvektori, jonka odotusarvo on $\mathrm{E}(\mathbf{y}) = \mathbf{1}_n\mu$, ja kovarianssimatriisi $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{I}_n$, eli $\mathbf{y}$:n elementit y_i ovat samoin jakautuneita ja korreloimattomia. Toisin sanoen: meillä on n havainnon satunnaisotos $y_1, y_2, \ldots, y_n$ normaalijakaumasta, jonka odotusarvo on μ ja varianssi σ^2. Tällöin y:n neliösumma

$$t_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y} = \mathbf{y}'\mathbf{C}\mathbf{y}\,, \tag{7.172}$$

ja keskiarvo $\bar{y} = \frac{1}{n}\mathbf{1}'\mathbf{y} := \mathbf{b}'\mathbf{y}$ ovat riippumattomia, sillä $(\mathbf{I}-\mathbf{J})\mathbf{1}\frac{1}{n} = \mathbf{0}$. Samoin neliömuodot $t_{yy} = \mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y}$ ja $n\bar{y}^2 = \mathbf{y}'\mathbf{J}\mathbf{y}$ ovat riippumattomia ja

$$\mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y}/\sigma^2 \sim \chi^2(n-1)\,, \quad \mathbf{y}'\mathbf{J}\mathbf{y}/\sigma^2 \sim \chi^2(1, \delta)\,, \tag{7.173}$$

missä $\delta = \mu\mathbf{1}'\mathbf{J}\mathbf{1}\mu/\sigma^2 = n\mu^2/\sigma^2$. Täten

$$F = \frac{\mathbf{y}'\mathbf{J}\mathbf{y}/1}{\mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y}/(n-1)} = \frac{n\bar{y}^2}{s_y^2} = \frac{\bar{y}^2}{s_y^2/n} \sim \mathrm{F}(1, n-1, \delta)\,, \tag{7.174}$$

missä epäkeskisyysparametri $\delta = n\mu^2/\sigma^2 = 0 \iff \mu = 0$. Näin saamme tutun t-testisuureen hypoteesin $\mu = 0$ testaamiseksi:

$$t = \frac{\bar{y}}{s_y/\sqrt{n}} \sim \mathrm{t}(1, n-1)\,. \tag{7.175}$$

Tarkastellaan vielä edellistä tilannetta sillä muutoksella, että kovarianssimatriisi $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{V} = \boldsymbol{\Sigma}$, missä $\mathbf{V}$ on tasakorrelaatiomatriisi:

$$\mathbf{V} = \big[(1-\varrho)\mathbf{I} + \varrho\mathbf{1}\mathbf{1}'\big] = \begin{pmatrix} 1 & \varrho & \ldots & \varrho \\ \varrho & 1 & \ldots & \varrho \\ \vdots & \vdots & \ddots & \vdots \\ \varrho & \varrho & \ldots & 1 \end{pmatrix}. \tag{7.176}$$

On mielenkiintoista havaita, että nytkin (vaikka $\mathbf{V} \neq \mathbf{I}_n$) riippumattomuusehto $\mathbf{C\Sigma b} = \mathbf{0}$ on voimassa:

$$\begin{aligned} \mathbf{A\Sigma b} &= \mathbf{C}\sigma^2[(1-\varrho)\mathbf{I} + \varrho\mathbf{11}']\mathbf{1}\tfrac{1}{n} \\ &= [(1-\varrho)\mathbf{C1} + \varrho\mathbf{C1}n]\sigma^2\tfrac{1}{n} = \mathbf{0}\,. \end{aligned} \tag{7.177}$$

□

7.9.4 Havainnon vierauden testaus

Esimerkki 7.9. Olkoon taustatilanne sama kuin esimerkissä 7.8, mutta oletetaan, että odotusarvovektori on

$$\mathrm{E}(\mathbf{y}) = \begin{pmatrix} \mu \\ \vdots \\ \mu \\ \mu+\delta \end{pmatrix} = \mu\mathbf{1} + \delta\mathbf{u} := \boldsymbol{\nu}\,, \tag{7.178}$$

missä $\mathbf{u} = \mathbf{i}_i = (0, 0, \dots, 0, 1)'$. Tarkastellaan i. y-arvon y_i poikkeamaa keskiarvosta. (Merkinnällisen mukavuuden takia i. havainto on viimeinen havainto.) Kuinka suuri poikkeama on "hälyttävä"? Onko kaikilla y-arvoilla sama odotusarvo μ vai onko $\delta \neq 0$, jolloin y_i:llä on poikkeava odotusarvo?

Esimerkin 6.1 (s. 181) perusteella on voimassa

$$(y_i - \bar{y})^2 = c_{ii}\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y} = (1 - \tfrac{1}{n})\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y}, \tag{7.179}$$

missä $\mathbf{C} = \mathbf{I} - \mathbf{J}$. Tällöin

$$\frac{(y_i - \bar{y})^2}{1 - \frac{1}{n}} = \mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y}\,. \tag{7.180}$$

Tarkastellaan suhdetta

$$\frac{\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y}/1}{\mathbf{y}'\mathbf{C}\mathbf{y}/(n-1)} = \frac{(y_i - \bar{y})^2}{(1 - \frac{1}{n})s_y^2} := F^{\#}, \tag{7.181}$$

missä $s_y^2 = \mathrm{var_d}(\mathbf{y})$. Osamäärän $F^{\#}$ osoittaja ja nimittäjä (jaettuna σ^2:lla) noudattavat χ^2-jakaumaa:

$$\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y}/\sigma^2 \sim \chi^2(1, \theta), \quad \theta = \boldsymbol{\nu}'\mathbf{P}_{\mathbf{Cu}}\boldsymbol{\nu}/\sigma^2 = \delta^2/\sigma^2, \tag{7.182a}$$

$$\mathbf{y}'\mathbf{C}\mathbf{y}/\sigma^2 \sim \chi^2(n-1)\,, \tag{7.182b}$$

mutta neliömuodot $\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y}$ ja $\mathbf{y}'\mathbf{C}\mathbf{y}$ eivät ole riippumattomia, sillä

$$\mathbf{CP}_{\mathbf{Cu}} = \mathbf{C}\cdot\mathbf{Cu}(\mathbf{u}'\mathbf{Cu})^{-1}\mathbf{u}'\mathbf{C} = \mathbf{P}_{\mathbf{Cu}} \neq \mathbf{0}\,. \tag{7.183}$$

Täten $F^{\#}$ ei noudata F-jakaumaa eikä sitä voi käyttää testisuureena.

Merkitään

$$\mathbf{Q} = \begin{pmatrix} \mathbf{C}_{n-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix} = \mathbf{Q}' = \mathbf{Q}^2, \tag{7.184}$$

missä $\mathbf{C}_{n-1} = \mathbf{I}_{n-1} - \frac{1}{n-1}\mathbf{1}_{n-1}\mathbf{1}'_{n-1}$. Koska

$$\mathbf{Cu} = \begin{pmatrix} -1/n \\ \vdots \\ -1/n \\ 1-1/n \end{pmatrix} = \tfrac{1}{n}\begin{pmatrix} -\mathbf{1}_{n-1} \\ n-1 \end{pmatrix}, \tag{7.185}$$

on helppo nähdä että $\mathbf{QCu} = \mathbf{0}$, joten myös $\mathbf{QP}_{\mathbf{Cu}} = \mathbf{0}$. Täten neliömuodot $\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y}$ ja $\mathbf{y}'\mathbf{Q}\mathbf{y}$ ovat riippumattomia, $\mathrm{r}(\mathbf{Q}) = n-2$, ja

$$\frac{\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y}}{\mathbf{y}'\mathbf{Q}\mathbf{y}/(n-2)} = F \sim \mathrm{F}(1, n-2, \theta)\,. \tag{7.186}$$

Nyt tietenkin

$$\mathbf{y}'\mathbf{Q}\mathbf{y} = \mathbf{y}_1'(\mathbf{I}_{n-1} - \mathbf{J}_{n-1})\mathbf{y}_1\,, \tag{7.187}$$

kun $\mathbf{y}$ on ositettu siten että $\mathbf{y} = \binom{\mathbf{y}_1}{y_i}$.Tällöin

$$\mathbf{y}'\mathbf{Q}\mathbf{y} = \sum_{k=1}^{n-1}(y_k - \bar{y}_{(i)})^2, \qquad \bar{y}_{(i)} = \tfrac{1}{n-1}\sum_{k=1}^{n-1} y_k\,, \tag{7.188}$$

$$\tfrac{1}{n-2}\,\mathbf{y}'\mathbf{Q}\mathbf{y} = s^2_{(i)} = \text{otosvarianssi ilman } i.\ \text{havaintoa.} \tag{7.189}$$

Hypoteesin $\delta = 0$ testaamiseksi saamme täten F-testisuureen:

$$\frac{\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y}}{\mathbf{y}'\mathbf{Q}\mathbf{y}/(n-2)} = \frac{(y_i - \bar{y})^2}{(1-\frac{1}{n})s^2_{(i)}} \sim \mathrm{F}(1, n-2)\,, \tag{7.190}$$

ts. jos $\delta = 0$, niin

$$\frac{y_i - \bar{y}}{s_{(i)}\sqrt{1-\frac{1}{n}}} = t \sim \mathrm{t}(n-2)\,. \tag{7.191}$$

Suuri t-arvo kertoo, että i. havainto on vakava kandidaatti poikkeavaksi (vieraaksi) havainnoksi (*outlieriksi*). □

7.9.5 How deviant can you be?

Esimerkki 7.10. Edellisessä esimerkissä johdimme testisuureen, joka kertoo erotuksen $(y_i - \bar{y})^2$ ”kohtuuttomasta suuruudesta” ts. tilastollisesta merkitsevyydestä. Toinen luonteva kysymyksen asettelu on seuraava:

$$\text{Kuinka monta hajonnan mittaa voi } y_i \text{ poiketa } \bar{y}\text{:stä?} \tag{7.192}$$

Nyt oletamme, että $\mathbf{y}$ on $\mathbb{R}^n$:n muuttujavektori eli se sisältää muuttujan y havaitut arvot. Osoitamme, että seuraava epäyhtälö, ns. Samuelsonin epäyhtälö, on vastaus (7.192):een:

$$y_i - \bar{y} \leq \frac{n-1}{\sqrt{n}} s_y \,. \tag{7.193}$$

Esimerkin 6.1 (s. 181) perusteella $(y_i - \bar{y})^2$ voidaan kirjoittaa muodossa

$$(y_i - \bar{y})^2 = c_{ii}\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y} = \frac{n-1}{n}\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y} \,. \tag{7.194}$$

Cauchyn–Schwarzin epäyhtälön mukaan $(\mathbf{u}'\mathbf{v})^2 \leq \mathbf{u}'\mathbf{u}\cdot\mathbf{v}'\mathbf{v}$, missä yhtäsuuruus toteutuu täsmälleen silloin jos on olemassa sellainen reaaliluku λ että $\mathbf{u} = \lambda\mathbf{v}$. Merkitään $\mathbf{u} = \mathbf{C}\mathbf{y}$, $\mathbf{v} = \mathbf{P}_{\mathbf{Cu}}\mathbf{y}$. Koska $\mathbf{C}\mathbf{P}_{\mathbf{Cu}} = \mathbf{P}_{\mathbf{Cu}}$, niin

$$\mathbf{u}'\mathbf{v} = \mathbf{y}'\mathbf{C}\mathbf{P}_{\mathbf{Cu}}\mathbf{y} = \mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y} \,, \quad \mathbf{u}'\mathbf{u} = \mathbf{y}'\mathbf{C}\mathbf{y} \,, \quad \mathbf{v}'\mathbf{v} = \mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y} \,. \tag{7.195}$$

Täten $(\mathbf{u}'\mathbf{v})^2 \leq \mathbf{u}'\mathbf{u}\cdot\mathbf{v}'\mathbf{v}$ voidaan kirjoittaa muodossa

$$(\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y})^2 \leq \mathbf{y}'\mathbf{C}\mathbf{y}\cdot\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y} \,, \tag{7.196}$$

eli

$$\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y} \leq \mathbf{y}'\mathbf{C}\mathbf{y} = (n-1)s_y^2 \,. \tag{7.197}$$

Sijoittamalla (7.197):n (7.194):een saamme epäyhtälön

$$(y_i - \bar{y})^2 = \frac{n-1}{n}\mathbf{y}'\mathbf{P}_{\mathbf{Cu}}\mathbf{y} \leq \frac{(n-1)^2}{n} s_y^2. \tag{7.198}$$

Näin olemme päätyneet epäyhtälöön (7.193). Yhtäsuuruus (7.193):ssa on voimassa (minkähän takia?) kun $y_1 = y_2 = \cdots = y_{n-1}$ ja y_n (eli y_i) on mielivaltainen.

Jos joku suomalainen innostuksissaan väittää, että hän on y-arvonsa suhteen 10 000 hajontaa koko kansan keskiarvon yläpuolella, voimme heti rauhoittaa häntä. Nimittäin jos $n = 5\,000\,000$, niin (7.193):n mukaan yksittäinen havainto voi olla enintään 2236 hajonnan mittaa keskiarvon yläpuolella.

Lähteinä Samuelsonin epäyhtälöön mainittakoon Olkin (1992), Thompson (1935), Samuelson (1968), Wolkowicz & Styan (1979). □

7.10 Otos multinormaalijakaumasta

Erityisesti monimuuttujamenetelmissä tarkastellaan satunnaisotoksia useampiulotteisesta jakaumasta (ks. luku 1.3, s. 35):

> ”Olkoon $\mathbf{u}_{(1)}, \mathbf{u}_{(2)}, \ldots, \mathbf{u}_{(n)}$ satunnaisotos populaatiosta, joka noudattaa multinormaalijakaumaa $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.” (7.199)

Koko satunnaisotos voidaan tällöin esittää matriisina

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix} = (\mathbf{u}_1 : \mathbf{u}_2 : \ldots : \mathbf{u}_p)\,. \tag{7.200}$$

Kuten luvussa 1.3 todettiin, näin johdettu matriisi $\mathbf{U}$ on siis itse asiassa ”tavallisen” empiirisen havaintomatriisin teoreettinen vastine; kun otos poimitaan, satunnaismatriisille realisoituu jokin arvo. Sanonnan (7.199) voimme esittää täten yhtäpitävästi:

> ”Olkoon $\mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)})$ satunnaisotos populaatiosta, joka noudattaa multinormaalijakaumaa $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.” (7.201)

Korostettakoon vielä, että (7.199) tarkoittaa että

- $\mathbf{u}_{(1)}, \mathbf{u}_{(2)}, \ldots, \mathbf{u}_{(n)}$ ovat riippumattomia p-ulotteisia satunnaisvektoreita, joilla jokaisella on sama jakauma, tässä tapauksessa $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p$ ovat n-ulotteisia satunnaisvektoreita ja

$$\mathbf{u}_i \sim \mathrm{N}_n(\mu_i \mathbf{1}, \sigma_{ii}\mathbf{I})\,, \quad \mathrm{cov}(\mathbf{u}_i, \mathbf{u}_j) = \sigma_{ij}\mathbf{I}_n\,. \tag{7.202}$$

Todistamme nyt, että $\bar{\mathbf{u}}$ on $\boldsymbol{\mu}$:n ja otoskovarianssimatriisi $\mathbf{S}$ on $\boldsymbol{\Sigma}$:n harhaton estimaattori, kun $\mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)})$ on satunnaisotos $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:sta. Itse asiassa normaalijakaumaoletusta emme harhattomuuden näyttämiseen tarvitse – riittää kun kyseessä on satunnaisotos jakaumasta, jonka odotusarvo on $\boldsymbol{\mu}$ ja kovarianssimatriisi on $\boldsymbol{\Sigma}$. Merkitään (kuten ennenkin)

$$\bar{\mathbf{u}} = \tfrac{1}{n}(\mathbf{u}_{(1)} + \mathbf{u}_{(2)} + \cdots + \mathbf{u}_{(n)}) = \tfrac{1}{n}\mathbf{U}'\mathbf{1} = \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \vdots \\ \bar{u}_p \end{pmatrix}, \tag{7.203a}$$

$$\boldsymbol{\Sigma} = \tfrac{1}{n-1}\mathbf{T} = \tfrac{1}{n-1}\mathbf{U}'(\mathbf{I} - \mathbf{J})\mathbf{U} = \tfrac{1}{n-1}\Big(\sum_{i=1}^{n} \mathbf{u}_{(i)}\mathbf{u}'_{(i)} - n\bar{\mathbf{u}}\bar{\mathbf{u}}'\Big)\,. \tag{7.203b}$$

Tällöin

> *Jos* $\mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)})$ *on satunnaisotos* $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*:sta, niin*
>
> $$\mathrm{E}(\bar{\mathbf{u}}) = \boldsymbol{\mu}\,, \quad \mathrm{E}(\mathbf{S}) = \boldsymbol{\Sigma}\,. \tag{7.204a}$$

Havaitsemme ensinnäkin että

$$\begin{aligned}\mathrm{E}(\bar{\mathbf{u}}) &= \mathrm{E}[\tfrac{1}{n}(\mathbf{u}_{(1)} + \mathbf{u}_{(2)} + \cdots + \mathbf{u}_{(n)})] \\ &= \tfrac{1}{n}[\mathrm{E}(\mathbf{u}_{(1)}) + \mathrm{E}(\mathbf{u}_{(2)}) + \cdots + \mathrm{E}(\mathbf{u}_{(n)})] \\ &= \tfrac{1}{n}(\boldsymbol{\mu} + \boldsymbol{\mu} + \cdots + \boldsymbol{\mu}) = \boldsymbol{\mu}\,. \end{aligned} \tag{7.205}$$

Satunnaisvektorien $\mathbf{u}_{(i)}$ riippumattomuudesta seuraa että

$$\begin{aligned}\mathrm{cov}(\bar{\mathbf{u}}) &= \mathrm{cov}[\tfrac{1}{n}(\mathbf{u}_{(1)} + \mathbf{u}_{(2)} + \cdots + \mathbf{u}_{(n)})] \\ &= \tfrac{1}{n^2}[\mathrm{cov}(\mathbf{u}_{(1)}) + \mathrm{cov}(\mathbf{u}_{(2)}) + \cdots + \mathrm{cov}(\mathbf{u}_{(n)})] \\ &= \tfrac{1}{n^2}(\boldsymbol{\Sigma} + \boldsymbol{\Sigma} + \cdots + \boldsymbol{\Sigma}) = \tfrac{1}{n}\boldsymbol{\Sigma}\,. \end{aligned} \tag{7.206}$$

Toisin sanoen (ks. *NJ4*):

$$\begin{aligned} &\mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)}) \text{ satunnaisotos } \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})\text{:sta} \\ \implies\ &\bar{\mathbf{u}} \sim \mathrm{N}_p(\boldsymbol{\mu}, \tfrac{1}{n}\boldsymbol{\Sigma})\,. \end{aligned} \tag{7.207}$$

Tulosummamatriisin $\mathbf{T}$ odotusarvo on

$$\mathrm{E}(\mathbf{T}) = \mathrm{E}(\sum_{i=1}^{n} \mathbf{u}_{(i)}\mathbf{u}'_{(i)} - n\bar{\mathbf{u}}\bar{\mathbf{u}}') = \sum_{i=1}^{n} \mathrm{E}(\mathbf{u}_{(i)}\mathbf{u}'_{(i)}) - n\,\mathrm{E}(\bar{\mathbf{u}}\bar{\mathbf{u}}')\,. \tag{7.208}$$

Tällöin, koska $\boldsymbol{\Sigma} = \mathrm{E}(\mathbf{u}_{(i)}\mathbf{u}'_{(i)}) - \boldsymbol{\mu}\boldsymbol{\mu}'$,

$$\sum_{i=1}^{n} \mathrm{E}(\mathbf{u}_{(i)}\mathbf{u}'_{(i)}) = \sum_{i=1}^{n}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}') = n\boldsymbol{\Sigma} + n\boldsymbol{\mu}\boldsymbol{\mu}'. \tag{7.209}$$

Koska $\mathrm{cov}\,\bar{\mathbf{u}} = \frac{1}{n}\boldsymbol{\Sigma} = \mathrm{E}(\bar{\mathbf{u}}\bar{\mathbf{u}}') - \boldsymbol{\mu}\boldsymbol{\mu}'$, on $n\,\mathrm{E}(\bar{\mathbf{u}}\bar{\mathbf{u}}') = \boldsymbol{\Sigma} + n\boldsymbol{\mu}\boldsymbol{\mu}'$, ja täten

$$\mathrm{E}(\mathbf{T}) = n\boldsymbol{\Sigma} + n\boldsymbol{\mu}\boldsymbol{\mu}' - (\boldsymbol{\Sigma} + n\boldsymbol{\mu}\boldsymbol{\mu}') = (n-1)\boldsymbol{\Sigma}\,. \tag{7.210}$$

Yhtälö (7.210) osoittaa otoskovarianssimatriisin $\mathbf{S} = \frac{1}{n-1}\mathbf{T}$ harhattomuuden $\boldsymbol{\Sigma}$:n estimaattorina.

7.10.1 Keskiarvovektorin ja tulosummamatriisin riippumattomuus

Seuraava tulos jätetään todistamatta:

Olkoon $\mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)})$ *satunnaisotos* $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*:sta. Tällöin* $\bar{\mathbf{u}}$ *ja* $\mathbf{T} = \mathbf{U}'(\mathbf{I} - \mathbf{J})\mathbf{U}$ *ovat riipumattomia, ja tulosummamatriisi* $\mathbf{T}$ *on jakautunut kuten*

$$\sum_{i=1}^{n-1} \mathbf{y}_{(i)}\mathbf{y}'_{(i)}\,, \tag{7.211}$$

missä $\mathbf{y}_{(i)}$*:t ovat riippumattomia ja kukin* $\sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$.

Wishart-jakauman määritelmästä (s. 243) huomaamme välittömästi, että $\mathbf{T}$ noudattaa keskistä Wishart-jakaumaa parametrein $n-1$ ja $\mathbf{\Sigma}$:

$$\mathbf{T} = \mathbf{y}_{(1)}\mathbf{y}'_{(1)} + \cdots + \mathbf{y}_{(n-1)}\mathbf{y}'_{(n-1)} \sim \mathrm{W}(n-1, \mathbf{\Sigma})\,. \tag{7.212}$$

7.10.2 Hotellingin T^2

Tarkastellaan (taas kerran) tilannetta, jossa $\mathbf{U}' = \big(\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \ldots : \mathbf{u}_{(n)}\big)$ on satunnaisotos $\mathrm{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$:sta. Tällöin (7.207):n mukaan $\bar{\mathbf{u}} \sim \mathrm{N}_p(\boldsymbol{\mu}, \frac{1}{n}\mathbf{\Sigma})$. Täten saamme $\bar{\mathbf{u}}$:n ja $\boldsymbol{\mu}_0$:n Mahalanobisin etäisyyden neliön jakauman seuraavasti:

$$\begin{aligned}\mathrm{MHLN}^2(\bar{\mathbf{u}}, \boldsymbol{\mu}_0, \tfrac{1}{n}\mathbf{\Sigma}) &= (\bar{\mathbf{u}} - \boldsymbol{\mu}_0)' \left(\tfrac{1}{n}\mathbf{\Sigma}\right)^{-1} (\bar{\mathbf{u}} - \boldsymbol{\mu}_0) \\ &= n(\bar{\mathbf{u}} - \boldsymbol{\mu}_0)'\mathbf{\Sigma}^{-1}(\bar{\mathbf{u}} - \boldsymbol{\mu}_0) \sim \chi^2(p, \theta)\,,\end{aligned} \tag{7.213}$$

missä epäkeskisyysparametri θ on nolla jos ja vain jos $\boldsymbol{\mu} = \boldsymbol{\mu}_0$. Täten Mahalanobisin etäisyyttä voidaan käyttää testisuureena hypoteesille $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, mikäli $\mathbf{\Sigma}$ on tunnettu. Jos $\mathbf{\Sigma}$ korvataan harhattomalla estimaatillaan eli otoskovarianssimatriisilla $\mathbf{S}$, saadaan lauseke

$$\boxed{T^2 = n(\bar{\mathbf{u}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{u}} - \boldsymbol{\mu}_0)\,,} \tag{7.214}$$

jota sanotaan *Hotellingin* T^2-testisuureeksi. On osoitettavissa, että

$$\frac{n-p}{(n-1)p}\,T^2 \sim \mathrm{F}(p, n-p, \theta)\,, \tag{7.215}$$

missä $\theta = n(\boldsymbol{\mu} - \boldsymbol{\mu}_0)'\mathbf{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)$. Voimme myös käyttää merkintää

$$T^2 \sim \frac{(n-1)p}{n-p}\,\mathrm{F}(p, n-p, \theta)\,. \tag{7.216}$$

Täten hypoteesi $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ hylätään riskitasolla α, jos

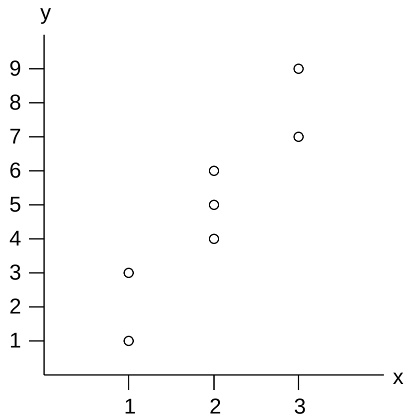
$$n(\bar{\mathbf{u}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{u}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{n-p}\,F_{\alpha;p,n-p}\,. \tag{7.217}$$

Luku $F_{\alpha;p,n-p}$ on F-jakauman kriittinen arvo riskitasolla α.

Jos $p = 1$ ja otos on $u_1, u_2, \ldots, u_n$ saamme T^2:n arvoksi

$$T^2 = n(\bar{u} - \mu_0)(s_u^2)^{-1}(\bar{u} - \mu_0) = \frac{(\bar{u} - \mu_0)^2}{s_u^2/n} = t^2 \sim \mathrm{F}(1, n-1, \theta)\,. \tag{7.218}$$

Täten T^2 on tietenkin tavallisen t-testin yleistys useampiulotteiseen tapaukseen.



Kuvio 7.4. (HT 7.1) Satunnaismuuttujien x ja y yhteisjakauma, jokainen pistepari yhtä todennäköinen.

Harjoitustehtäviä

7.1. Oletetaan, että kuvion 7.4 pisteparvi esittää kaksiulotteista diskreettiä todennäköisyysjakaumaa, jossa kunkin pisteparin todennäköisyys on $1/7 = 1/n$. Olkoon $\mathrm{E}(y \mid x)$ satunnaismuuttuja, jonka arvot ovat ehdollisia odotusarvoja $\mathrm{E}(y \mid x = 1)$, $\mathrm{E}(y \mid x = 2)$, $\mathrm{E}(y \mid x = 3)$ todennäköisyyksin $\mathrm{P}(x = 1)$, $\mathrm{P}(x = 2)$, $\mathrm{P}(x = 3)$. Määritellään satunnaismuuttuja $\mathrm{var}(y \mid x)$ vastaavalla tavalla. Osoita, että hajotelma

$$\mathrm{var}(y) = \mathrm{var}[\mathrm{E}(y \mid x)] + \mathrm{E}[\mathrm{var}(y \mid x)] := \mathrm{var}[m(x)] + \mathrm{E}[v(x)]$$

(kerrottuna luvulla n) voidaan esittää muodossa

$$\sum_{i=1}^{g}\sum_{i=1}^{n_i}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{g} n_i(\bar{y}_i - \bar{y})^2 + \sum_{i=1}^{g}\sum_{i=1}^{n_i}(y_{ij} - \bar{y}_i)^2,$$

$$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}\,, \quad \mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y} = \mathbf{y}'(\mathbf{H} - \mathbf{J})\mathbf{y} + \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}\,,$$

missä $n_i = i$. ”ryhmän” koko, g = ryhmien lukumäärä, SST, SSR ja SSE ovat varianssianalyysiin liittyviä nelösummia ja

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{1}_{n_g} \end{pmatrix}, \quad \mathbf{H} = \mathbf{P}_{\mathbf{X}} = \begin{pmatrix} \mathbf{J}_{n_1} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{J}_{n_g} \end{pmatrix}.$$

7.2. Osoita, että jos $p = 2$, niin Hotellingin T^2 voidaan esittää muodossa

$$T^2 = \frac{1}{1 - r^2}(t_1 + t_2 + 2rt_1t_2)\,,$$

missä $t_i = (\bar{u}_i - \mu_{0i})/(s_i/\sqrt{n})$, $i = 1, 2$, eli t_i:t ovat tavallisia yksiulotteisia hypoteesien $\mu_i = \mu_{0i}$ t-testisuureita; ks. Mustonen (1995, s. 45).

7.3. Olkoon $\mathbf{S}$ positiivisesti definiitti kovarianssimatriisi, jonka ominaisarvohajotelma on $\mathbf{S} = \mathbf{T\Lambda T}'$. Ota selvää seuraavasta päättelystä, joka osoittaa, että $\mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S})$ on tiettyjen keskistettyjen, skaalattujen (varianssit 1) ja korreloimattomien muuttujien neliösumma.

$$\begin{aligned}
\mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S}) &= (\mathbf{u}_{(i)} - \bar{\mathbf{u}})'\mathbf{S}^{-1}(\mathbf{u}_{(i)} - \bar{\mathbf{u}}) \\
&= (\mathbf{u}_{(i)} - \bar{\mathbf{u}})'\mathbf{T\Lambda}^{-1}\mathbf{T}'(\mathbf{u}_{(i)} - \bar{\mathbf{u}}) \\
&= \mathbf{w}'_{(i)}\mathbf{\Lambda}^{-1}\mathbf{w}_{(i)} \\
&= \frac{w_{i1}^2}{\mathrm{var_s}(w_1)} + \frac{w_{i2}^2}{\mathrm{var_s}(w_2)} \\
&= \mathrm{MHLN}^2(\mathbf{w}_{(i)}, \mathbf{0}, \mathbf{\Lambda}) \\
&= \mathrm{MHLN}^2\big(\mathbf{w}_{(i)}, \bar{\mathbf{w}}, \mathrm{cov_d}(\mathbf{W})\big)\,.
\end{aligned}$$

7.4. Olkoot $\mathbf{A}$ ja $\mathbf{B}$ $n \times n$-matriiseja ja $\mathbf{y}$ on satunnaisvektori, jonka odotusarvo on $\mathbf{X}\boldsymbol{\beta} \in \mathbb{R}^n$ ja kovarianssimatriisi on $\mathbf{V}$ eli kyseessä on lineaarinen malli $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. Tällöin sanotaan, että yhtälö $\mathbf{Ay} = \mathbf{By}$ toteutuu todennäköisyydellä 1 jos

$$\mathbf{A}(\mathbf{X} : \mathbf{V}) = \mathbf{B}(\mathbf{X} : \mathbf{V})\,.$$

Osoita, että ylläoleva ehto totetuu täsmälleen silloin kun jokin seuraavista yhtäpitävistä ehdoista totetuu:

(1) $\mathbf{AX} = \mathbf{BX}$ ja $\mathbf{AV} = \mathbf{BV}$,

(2) $\mathbf{AX} = \mathbf{BX}$ ja $\mathbf{AVM} = \mathbf{BVM}$, missä $\mathbf{M} = \mathbf{I} - \mathbf{P_X}$,

(3) $\mathbf{AX} = \mathbf{BX}$ ja $\mathrm{cov}(\mathbf{Ay} - \mathbf{By}) = \mathbf{0}$,

(4) $\mathbf{AX} = \mathbf{BX}$, $\mathrm{cov}(\mathbf{Ay}) = \mathrm{cov}(\mathbf{By})$, ja $2\,\mathrm{cov}(\mathbf{Ay}) = \mathrm{cov}(\mathbf{Ay}, \mathbf{By}) + \mathrm{cov}(\mathbf{By}, \mathbf{Ay})$.

Todistuksessa tarvitaan tulosta (9.149) (s. 327):

$$\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{VM}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{VM}).$$

Groß & Trenkler (1998a, Th. 1).

7.5. Olkoon satunnaisvektorin $\mathbf{u} = (x, y, z)'$ kovarianssimatriisi

$$\mathrm{cov}(\mathbf{u}) = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & -2 \end{pmatrix}.$$

Määritä $\mathrm{cor}(\mathbf{u})$, $\mathrm{var}(x - y)$ ja $\mathrm{cov}(x, y - z)$.

Luku 8

Korrelaatio & regressio

Minulla on tunne, että pieni matka tekisi minulle hyvää, ja haluaisin haistella vieraan kaupungin ilmaa ja katsella toisenlaisia taloja kuin joka päivä näen ympärilläni. Myös haluaisin nuuskia vierasta teatteria ja kenties suudella jokakuta kaunista näyttelijätärtä, jota ennen en ole tavannut, koska kaikki uusi viehättää mieltäni ja rakastan suuresti kauneutta, kuten hyvin tiedät.

Mika Waltari (1949): *Neljä päivänlaskua.*

8.1 Johdanto

Korrelaatiomatriisia on käsitelty jo useaan otteeseen edellisissä luvuissa, mutta keskeisenä käsitteenä se ansaitsee vielä oman lukunsa – oli miten oli, joka tapauksessa tässä luvussa siis on paljon vanhan kertausta. Palautamme aluksi mieleen ortogonaaliprojektorin ominaisuuksia ja sitten käsittelemme erityisesti ortogonaaliprojektoreita $\mathbf{J}$ ja $\mathbf{C} = \mathbf{I} - \mathbf{J}$.

8.1.1 Ortogonaaliprojektori A:n sarakeavaruuteen

Aloitamme matriisin $\mathbf{A}_{n\times m}$ sarakeavaruuden *ortogonaalikomplementin* määritelmällä:

$$\begin{aligned}\mathscr{C}(\mathbf{A})^{\perp} &= \text{sarakeavaruuden } \mathscr{C}(\mathbf{A}_{n\times m}) \text{ ortogonaalikomplementti}\\ &= \text{niiden } \mathbb{R}^n\text{:n vektoreiden joukko jotka ovat kohtisuoras-}\\ &\quad\ \text{sa jokaisen } \mathscr{C}(\mathbf{A})\text{:n vektorin kanssa.}\end{aligned}$$

Käsite ”kohtisuoruus” riippuu annetusta sisätulosta: vektorit ovat kohtisuorassa, jos niiden sisätulo on 0. Ellei toisin todeta, tarkoitamme sisätulolla tavallista euklidista sisätuloa:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = \langle \mathbf{y}, \mathbf{x} \rangle\,. \tag{8.1}$$

Jos $\mathbf{V}$ on symmetrinen positiivisesti definiitti matriisi, niin sisätulon yleisempi lauseke on

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{V}} = \mathbf{x}'\mathbf{V}\mathbf{y}\,. \tag{8.2}$$

Tällöin $\mathbf{V}$:tä sanotaan sisätulomatriisiksi. tavanomaisen sisätulon tapauksessa sisätulomatriisi on $\mathbf{I}$.

Vektoriavaruuksien terminologiaa käyttäen vektoriavaruus $\mathcal{T}$ on aliavaruuksien $\mathcal{U}$ ja $\mathcal{V}$ suora summa, jos $\mathcal{T}$ on $\mathcal{U}$:n ja $\mathcal{V}$:n summa ja $\mathcal{U} \cap \mathcal{V} = \{\mathbf{0}\}$:

$$\mathcal{T} = \mathcal{U} \oplus \mathcal{V} \iff \mathcal{T} = \mathcal{U} + \mathcal{V} \text{ ja } \mathcal{U} \cap \mathcal{V} = \{\mathbf{0}\}\,. \tag{8.3}$$

Jos $\mathcal{U}$ ja $\mathcal{V}$ ovat kohtisuorassa keskenään, merkitsemme $\mathcal{T} = \mathcal{U} \boxplus \mathcal{V}$.

On helppo vakuuttautua, että $\mathbb{R}^n$ on esitettävissä suorana summana

$$\mathbb{R}^n = \mathscr{C}(\mathbf{A}) \boxplus \mathscr{C}(\mathbf{A})^{\perp}. \tag{8.4}$$

Tällöin jokainen $\mathbf{y} \in \mathbb{R}^n$ voidaan esittää yksikäsitteisesti muodossa

$$\mathbf{y} = \hat{\mathbf{y}} + \dot{\mathbf{y}}, \quad \hat{\mathbf{y}} \in \mathscr{C}(\mathbf{A}), \quad \dot{\mathbf{y}} \in \mathscr{C}(\mathbf{A})^{\perp}. \tag{8.5}$$

Vektori $\hat{\mathbf{y}}$ on $\mathbf{y}$:n projektio $\mathscr{C}(\mathbf{A})$:lle ja $\dot{\mathbf{y}}$ on $\mathbf{y}$:n projektio $\mathscr{C}(\mathbf{A})^{\perp}$:lle. Toisin sanoen on olemassa vektorit $\mathbf{b}$ ja $\mathbf{c}$ (eivät välttämättä yksikäsitteisiä) siten että

$$\mathbf{y} = \mathbf{A}\mathbf{b} + \mathbf{A}^{\perp}\mathbf{c} = \hat{\mathbf{y}} + \dot{\mathbf{y}}\,, \tag{8.6}$$

missä $\mathbf{A}^{\perp}$ on matriisi, jolla on ominaisuus $\mathscr{C}(\mathbf{A}^{\perp}) = \mathscr{C}(\mathbf{A})^{\perp}$. Vektorit $\mathbf{A}\mathbf{b}$ ja $\mathbf{A}^{\perp}\mathbf{c}$ ovat yksikäsitteisiä (mutta siis $\mathbf{b}$ ja $\mathbf{c}$ välttämättä eivät); kunhan $\mathbf{y}$ on annettu. Jos $\mathbf{P}$ on sellainen matriisi, että kertolasku $\mathbf{P}\mathbf{y}$ antaa kaikilla vektoreilla $\mathbf{y} \in \mathbb{R}^n$ tuloksen

$$\mathbf{P}\mathbf{y} = \hat{\mathbf{y}}\,, \tag{8.7}$$

jolloin tietenkin

$$(\mathbf{I} - \mathbf{P})\mathbf{y} = \dot{\mathbf{y}} \in \mathscr{C}(\mathbf{A})^{\perp}, \tag{8.8}$$

niin $\mathbf{P}$ on *ortogonaaliprojektori* $\mathbf{A}$:n sarakeavaruudelle $\mathscr{C}(\mathbf{A})$, ja sitä merkitään symbolilla $\mathbf{P}_{\mathbf{A}}$. Ortogonaaliprojektorin $\mathbf{P}_{\mathbf{A}}$ ekplisiittiseksi lausekkeeksi osoittautuu

$$\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}', \tag{8.9}$$

missä $(\mathbf{A}'\mathbf{A})^{-}$ voi olla mikä hyvänsä $\mathbf{A}'\mathbf{A}$:n yleistetty käänteismatriisi. Mikäli $\mathbf{A}$:n sarakkeet ovat vapaat, niin

$$\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'. \tag{8.10}$$

Palaamme lausekkeen (8.9) johtamiseen luvussa 9.7 (s. 318). Voimme tässä yhteydessä kuitenkin varmistaa, että (8.10):n mukainen $\mathbf{P}_{\mathbf{A}}$ toimii halutulla tavalla. Jos nimittäin (8.6) kerrotaan vasemmalta matriisilla $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, niin saadaan

$$\mathbf{P}_{\mathbf{A}}\mathbf{y} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A}\mathbf{b} = \mathbf{A}\mathbf{b} = \hat{\mathbf{y}}\,. \tag{8.11}$$

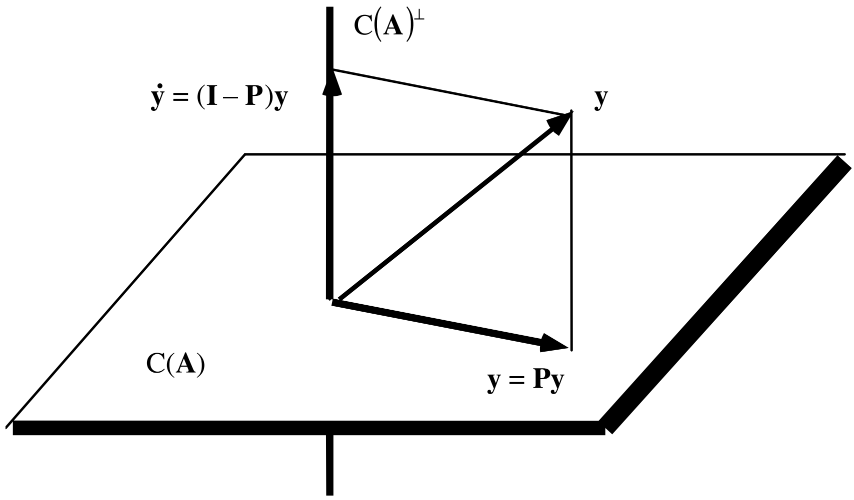
Vektorin $\mathbf{y}$ projektio $\hat{\mathbf{y}}$ samoin kuin projektori $\mathbf{P_A}$ ovat aina yksikäsitteisiä – projektori riippuu vain sarakeavaruudesta, johon projisoidaan:

$$\mathbf{P_A} = \mathbf{P_B} \iff \mathscr{C}(\mathbf{A}) = \mathscr{C}(\mathbf{B})\,. \tag{8.12}$$

Vektorin $\mathbf{y}$ projektio $\hat{\mathbf{y}} = \mathbf{Ab}$ on aina yksikäsitteinen, mutta kerroinvektori $\mathbf{b}$ on yksikäsitteinen vain jos $\mathbf{A}$:n sarakkeet ovat vapaat.

On helppo päätellä, että kun $\mathbf{P_A}$ on ortogonaaliprojektori $\mathscr{C}(\mathbf{A})$:lle, niin $\mathbf{I} - \mathbf{P_A}$ on ortogonaaliprojektori $\mathscr{C}(\mathbf{A})^\perp$:lle:

$$\mathbf{I} - \mathbf{P_A} = \text{ ortogonaaliprojektori } \mathscr{C}(\mathbf{A})^\perp\text{:lle.} \tag{8.13}$$



Kuvio 8.1. Vektorin $\mathbf{y}$ ortogonaaliprojektio $\mathscr{C}(\mathbf{A})$:lle on $\hat{\mathbf{y}}$; vektori $\dot{\mathbf{y}}$ on $\mathbf{y}$:n ortogonaaliprojektio $\mathscr{C}(\mathbf{A})^\perp$:lle. [Hattu puuttuu]

Ortogonaaliprojektiolla on luonnollisesti minimointiominaisuus

$$\|\mathbf{y} - \mathbf{P_A y}\|^2 \leq \|\mathbf{y} - \mathbf{Ab}\|^2 \ \forall\ \mathbf{b} \in \mathbb{R}^m. \tag{8.14}$$

Tulos (8.14) todistetaan seuraavasti:

$$\begin{aligned}
\|\mathbf{y} - \mathbf{Ab}\|^2 &= \|(\mathbf{y} - \mathbf{P_A y}) + (\mathbf{P_A y} - \mathbf{Ab})\|^2 \\
&= \|(\mathbf{I} - \mathbf{P_A})\mathbf{y} + (\mathbf{P_A y} - \mathbf{Ab})\|^2 \\
&= \|(\mathbf{I} - \mathbf{P_A})\mathbf{y}\|^2 + \|(\mathbf{P_A y} - \mathbf{Ab})\|^2 + 2\mathbf{y}'(\mathbf{I} - \mathbf{P_A})(\mathbf{P_A y} - \mathbf{Ab}) \\
&= \|(\mathbf{I} - \mathbf{P_A})\mathbf{y}\|^2 + \|(\mathbf{P_A y} - \mathbf{Ab})\|^2 \\
&\geq \|(\mathbf{I} - \mathbf{P_A})\mathbf{y}\|^2 \ \forall\ \mathbf{b} \in \mathbb{R}^m,
\end{aligned} \tag{8.15}$$

missä viimeinen yhtäsuuruus perustuu yhtälöön

$$(\mathbf{I} - \mathbf{P_A})(\mathbf{P_A y} - \mathbf{Ab}) = \mathbf{0}\,. \tag{8.16}$$

Jos kerromme yhtälön (8.6) vasemmalta matriisilla $\mathbf{A}'$, saamme ns. *normaaliyhtälön*

$$\mathbf{A}'\mathbf{A}\mathbf{b} = \mathbf{A}'\mathbf{y}\,. \tag{8.17}$$

Olkoon $\mathbf{b}_*$ mielivaltainen normaaliyhtälön ratkaisu. Silloin $\mathbf{A}'(\mathbf{y} - \mathbf{A}\mathbf{b}_*) = \mathbf{0}$, joten $\mathbf{y} - \mathbf{A}\mathbf{b}_*$ on kohtisuorassa jokaisen $\mathscr{C}(\mathbf{A})$:n vektorin kanssa. Koska

$$\mathbf{y} = \mathbf{A}\mathbf{b}_* + (\mathbf{y} - \mathbf{A}\mathbf{b}_*), \quad \text{missä } \mathbf{A}\mathbf{b}_* \in \mathscr{C}(\mathbf{A}),\ \mathbf{y} - \mathbf{A}\mathbf{b}_* \in \mathscr{C}(\mathbf{A})^\perp, \tag{8.18}$$

niin vektori $\mathbf{A}\mathbf{b}_*$ on $\mathbf{y}$:n projektio $\mathscr{C}(\mathbf{A})$:lle eli $\mathbf{A}\mathbf{b}_* = \mathbf{P}_\mathbf{A}\mathbf{y}$, ja

$$\min_{\mathbf{b}} \|\mathbf{y} - \mathbf{A}\mathbf{b}\|^2 = \|\mathbf{y} - \mathbf{A}\mathbf{b}_*\|^2 = \mathbf{y}'(\mathbf{I} - \mathbf{P}_\mathbf{A})\mathbf{y}\,. \tag{8.19}$$

Osoitetaan vielä, että jos $\mathbf{b}_*$ on mielivaltainen normaaliyhtälön (8.17) ratkaisu, niin se minimoi lausekkeen $\|\mathbf{y} - \mathbf{A}\mathbf{b}\|^2$. Tällöin, aivan samoin kuin (8.15):ssa,

$$\begin{aligned}
\|\mathbf{y} - \mathbf{A}\mathbf{b}\|^2 &= \|\mathbf{y} - \mathbf{A}\mathbf{b}_* + (\mathbf{A}\mathbf{b}_* - \mathbf{A}\mathbf{b})\|^2 \\
&= \|\mathbf{y} - \mathbf{A}\mathbf{b}_*\|^2 + \|\mathbf{A}\mathbf{b}_* - \mathbf{A}\mathbf{b}\|^2 + 2(\mathbf{A}\mathbf{b}_* - \mathbf{A}\mathbf{b})'(\mathbf{y} - \mathbf{A}\mathbf{b}_*) \\
&= \|\mathbf{y} - \mathbf{A}\mathbf{b}_*\|^2 + \|\mathbf{A}(\mathbf{b}_* - \mathbf{b})\|^2 + 2\mathbf{z}'\underbrace{\mathbf{A}'(\mathbf{y} - \mathbf{A}\mathbf{b}_*)}_{=\mathbf{0}} \\
&= \|\mathbf{y} - \mathbf{A}\mathbf{b}_*\|^2 + \|\mathbf{A}(\mathbf{b}_* - \mathbf{b})\|^2,
\end{aligned} \tag{8.20}$$

missä $\mathbf{z} = \mathbf{b}_* - \mathbf{b}$. Täten on varmistettu seuraava tulos:

Jos $\mathbf{b}_*$ on normaaliyhtälön $\mathbf{A}'\mathbf{A}\mathbf{b} = \mathbf{A}'\mathbf{y}$ ratkaisu, niin

$$\min_{\mathbf{b}} \|\mathbf{y} - \mathbf{A}\mathbf{b}\|^2 = \|\mathbf{y} - \mathbf{A}\mathbf{b}_*\|^2. \tag{8.21}$$

Aiemmin olemme jo määritelleet ortogonaaliprojektorin seuraavasti:

Neliömatriisi $\mathbf{P}$ on ortogonaaliprojektori, jos se toteuttaa ehdon

$$\mathbf{P} = \mathbf{P}^2 = \mathbf{P}'. \tag{8.22}$$

Tällöin $\mathbf{P}$ on ortogonaaliprojektori $\mathscr{C}(\mathbf{P})$:lle.

Jos lisäksi $\mathbf{P}$:n sarakeavaruus on sama kuin $\mathbf{A}$:n sarakeavaruus eli $\mathscr{C}(\mathbf{P}) = \mathscr{C}(\mathbf{A})$, niin $\mathbf{P}$ on ortogonaaliprojektori nimenomaan $\mathscr{C}(\mathbf{A})$:lle ja siitä käytetään merkintää $\mathbf{P}_\mathbf{A}$. Toistettakoon taas, että (8.22) takaa, että $\mathbf{P}$ on ortogonaaliprojektori kun sisätulomatriisi on $\mathbf{I}$ eli kun kyseessä on tavanomainen sisätulo.

8.1.2 Ortogonaaliprojektorit J ja I – J

Kuten olemme aiemmin todenneet, tärkeimmät matriisit havaintoaineiston keskistämisen yhteydessä ovat $\mathbf{J}$ ja $\mathbf{I} - \mathbf{J}$:

$$\mathbf{J} = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = \frac{1}{n}\mathbf{1}\mathbf{1}' = \mathbf{P_1} = \frac{1}{n}\begin{pmatrix} \mathbf{1}' \\ \mathbf{1}' \\ \vdots \\ \mathbf{1}' \end{pmatrix}. \tag{8.23}$$

Matriisi $\mathbf{J}$ on symmetrinen $n \times n$-matriisi ja se on idempotentti eli $\mathbf{J}^2 = \mathbf{J}$, ja

$$\operatorname{tr}(\mathbf{J}) = 1\,, \quad \operatorname{r}(\mathbf{J}) = 1\,. \tag{8.24}$$

Itse asiassa on osoitettavissa seuraava hyödyllinen ortogonaaliprojektorin ominaisuus:

$$\operatorname{r}(\mathbf{P_A}) = \operatorname{tr}(\mathbf{P_A}) = \operatorname{r}(\mathbf{A})\,. \tag{8.25}$$

Täten muuttujan y keskiarvovektori $\bar{\bar{\mathbf{y}}}$, joka siis on $\mathbf{y}$:n ortogonaaliprojektio suoralle $\mathscr{C}(\mathbf{1})$, voidaan esittää matriisitulona

$$\bar{\bar{\mathbf{y}}} = \mathbf{J}\mathbf{y} = \begin{pmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix}. \tag{8.26}$$

Keskistetty $\mathbf{y}$ eli $\tilde{\mathbf{y}}$ on

$$\tilde{\mathbf{y}} = \mathbf{y} - \bar{\bar{\mathbf{y}}} = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} = \mathbf{y} - \mathbf{J}\mathbf{y} = (\mathbf{I} - \mathbf{J})\mathbf{y} = \mathbf{C}\mathbf{y}\,. \tag{8.27}$$

Matriisi $\mathbf{I} - \mathbf{J}$ on keskistäjämatriisi ja sitä merkitsemme $\mathbf{C}$:llä.

Kuten tiedämme, keskistäjämatriisi $\mathbf{C} = \mathbf{I} - \mathbf{J}$ on myös ortogonaaliprojektori; se on symmetrinen ja idempotentti. Lisäksi keskistäjämatriisilla on ominaisuus

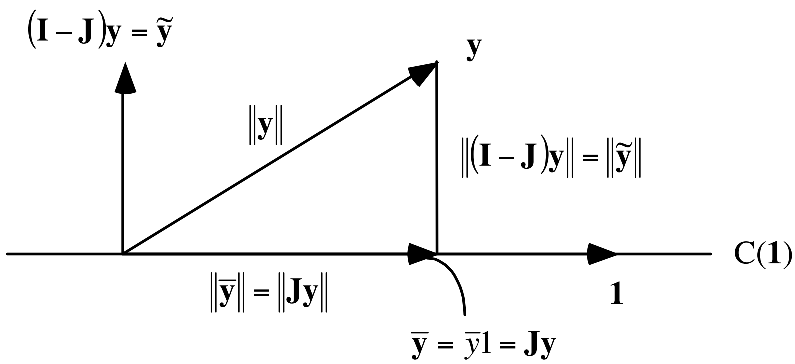
$$\mathbf{C}\mathbf{J} = (\mathbf{I} - \mathbf{J})\mathbf{J} = \mathbf{J} - \mathbf{J} = \mathbf{0}\,, \tag{8.28}$$

joten keskiarvovektori $\bar{\bar{\mathbf{y}}} = \mathbf{J}\mathbf{y}$ ja keskistetty vektori $\tilde{\mathbf{y}} = (\mathbf{I} - \mathbf{J})\mathbf{y}$ ovat keskenään ortogonaalisia (kuten kuviosta 8.2 ”näkyy”):

$$\mathbf{y}'\mathbf{J}(\mathbf{I} - \mathbf{J})\mathbf{y} = 0\,. \tag{8.29}$$

Ominaisuudesta (8.29) voimme päätellä, että

$$\mathscr{C}(\mathbf{J}) = \mathscr{C}(\mathbf{1}) \text{ ja } \mathscr{C}(\mathbf{I} - \mathbf{J}) \text{ ovat ortogonaalisia aliavaruuksia.} \tag{8.30}$$



Kuvio 8.2. Vektorin $\mathbf{y}$ projisointi suoralle $\mathscr{C}(\mathbf{1})$ projektorin $\mathbf{J}$ avulla. (Kuviossa pitäisi $\bar{\mathbf{y}}$ korvata $\bar{\bar{\mathbf{y}}}$:llä.)

Minne $\mathbf{I} - \mathbf{J}$ sitten projisoi? Se on ortogonaaliprojektori omalle sarakeavaruudelleen $\mathscr{C}(\mathbf{I}-\mathbf{J})$, joka on $\mathscr{C}(\mathbf{1})$:n sarakeavaruuden ortogonaalikomplementti:

$$\begin{aligned} \mathscr{C}(\mathbf{1})^{\perp} &= \text{niiden vektoreiden joukko jotka ovat kohtisuorassa jokaisen} \\ &\quad\ \mathscr{C}(\mathbf{1})\text{:n vektorin kanssa eli kohtisuorassa } \mathbf{1}\text{:n kanssa} \\ &= \{\, \mathbf{u} \in \mathbb{R}^n : \mathbf{u}'\mathbf{1} = 0 \,\}. \end{aligned}$$

Täten siis

$$\mathscr{C}(\mathbf{I} - \mathbf{J}) = \mathscr{C}(\mathbf{1})^{\perp} = \mathscr{C}(\mathbf{J})^{\perp}. \tag{8.31}$$

On osoitettavissa, että

$$\mathrm{r}(\mathbf{I} - \mathbf{J}) = \mathrm{tr}(\mathbf{I} - \mathbf{J}) = n - 1 = \dim \mathscr{C}(\mathbf{1})^{\perp}. \tag{8.32}$$

Triviaalisti on aina voimassa $\mathbf{y} = \bar{\bar{\mathbf{y}}} + (\mathbf{y} - \bar{\bar{\mathbf{y}}})$, mutta nyt on lisäksi voimassa [koska $\bar{\bar{\mathbf{y}}}'(\mathbf{y} - \bar{\bar{\mathbf{y}}}) = 0$]

$$\|\mathbf{y}\|^2 = \|\bar{\bar{\mathbf{y}}}\|^2 + \|\mathbf{y} - \bar{\bar{\mathbf{y}}}\|^2, \tag{8.33a}$$

ts.

$$\|\mathbf{y}\|^2 = \|\mathbf{J}\mathbf{y}\|^2 + \|(\mathbf{I} - \mathbf{J})\mathbf{y}\|^2, \tag{8.33b}$$

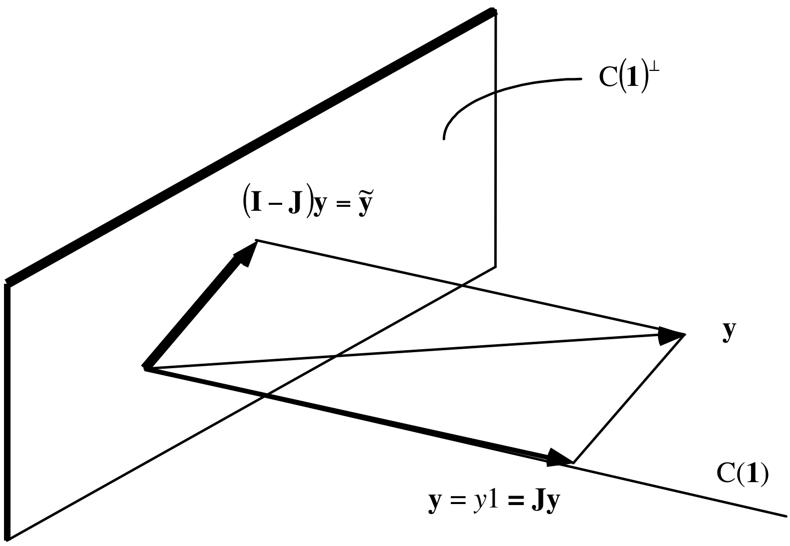
$$\mathbf{y}'\mathbf{y} = \mathbf{y}'\mathbf{J}\mathbf{y} + \mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y}, \tag{8.33c}$$

$$\sum_{i=1}^{n} y_i^2 = n\bar{y}^2 + \sum_{i=1}^{n} (y_i - \bar{y})^2. \tag{8.33d}$$

Kuvio 8.3 valaisee tilannetta.

Palautamme mieleen, että $\mathbf{y}$:n otosvarianssi voidaan kirjoittaa muodossa

$$\mathrm{var}_{\mathrm{d}}(\mathbf{y}) = \tfrac{1}{n-1}\|\tilde{\mathbf{y}}\|^2 = \tfrac{1}{n-1}\tilde{\mathbf{y}}'\tilde{\mathbf{y}} = \tfrac{1}{n-1}\mathbf{y}'\mathbf{C}\mathbf{y}\,. \tag{8.34}$$



Kuvio 8.3. Vektorin $\mathbf{y}$ projisointi $\mathscr{C}(\mathbf{1})$:lle ja $\mathscr{C}(\mathbf{1})^{\perp}$:lle. (Yläviiva puuttuu)

Täten $\mathbf{y}$:n otosvarianssi kertoo siitä kuinka kaukana muuttujavektori $\mathbf{y}$ on vakiovektorista $\mathbf{1}$. Muuttujien $\mathbf{x}$ ja $\mathbf{y}$ otoskovarianssi on

$$\operatorname{cov}_{\mathrm{d}}(\mathbf{x}, \mathbf{y}) = \tfrac{1}{n-1}\mathbf{x}'\mathbf{C}\mathbf{y}\,, \tag{8.35}$$

joten (otos)korrelaatiokerroin voidaan esittää seuraavasti:

$$\begin{aligned} r_{xy} &= \operatorname{cor}_{\mathrm{s}}(x, y) = \operatorname{cor}_{\mathrm{d}}(\mathbf{x}, \mathbf{y}) = \cos(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \operatorname{cor}_{\mathrm{d}}(\mathbf{C}\mathbf{x}, \mathbf{C}\mathbf{y}) \\ &= \frac{t_{xy}}{\sqrt{t_{xx}t_{yy}}} = \frac{\mathrm{SP}_{xy}}{\sqrt{\mathrm{SS}_x\,\mathrm{SS}_y}} = \frac{s_{xy}}{s_x s_y} \\ &= \frac{\mathbf{x}'\mathbf{C}\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{C}\mathbf{x}\cdot\mathbf{y}'\mathbf{C}\mathbf{y}}} = \frac{\tilde{\mathbf{x}}'\tilde{\mathbf{y}}}{\sqrt{\tilde{\mathbf{x}}'\tilde{\mathbf{x}}\cdot\tilde{\mathbf{y}}'\tilde{\mathbf{y}}}}\,. \end{aligned} \tag{8.36}$$

On syytä korostaa, että

> korrelaatiokerroin on nimenomaan keskistettyjen muuttujavektoreiden välisen kulman kosini.

Jos $\mathbf{x}$:n tai $\mathbf{y}$:n hajonta on nolla, niin silloin $\mathbf{x} \in \mathscr{C}(\mathbf{1})$ tai $\mathbf{y} \in \mathscr{C}(\mathbf{1})$, jolloin myös $\mathbf{x}$:n ja $\mathbf{y}$:n välinen kovarianssi on nolla, sillä $\mathbf{x}'(\mathbf{I}-\mathbf{J})\mathbf{y} = 0$. Tässä tilanteessa korrelaatio on $0/0$ eli sitä ei ole määritelty.

Kuten olemme aiemmin todenneet, on mahdollista, että alkuperäisten muuttujavektorien kosini on 0 mutta keskistettyjen kosini (eli korrelaatio) on 1. Jos esim.

$$\mathbf{x} = (1, 1, \sqrt{2})' \text{ ja } \mathbf{y} = (-1, -1, \sqrt{2})', \tag{8.37}$$

niin $\cos(\mathbf{x}, \mathbf{y}) = 0$, mutta $\text{cor}_\text{d}(\mathbf{x}, \mathbf{y}) = 1$. Jos korrelaatio on 1 (itseisarvoltaan), niin silloin kaikki pisteparit (x_i, y_i) ovat samalla suoralla (jnka ei ole kummankaan koordinaattiakselin suuntainen). Tämä merkitsee, että on olemassa vakiot a ja b $(b \neq 0)$ siten että

$$y_i = a + bx_i\,, \quad i = 1, \dots, n\,, \tag{8.38}$$

eli vektorimerkinnöin

$$\mathbf{y} = a\mathbf{1} + b\mathbf{x}\,. \tag{8.39}$$

Yhtälö (8.39) merkitsee tietysti sitä, että

$$\text{vektorijoukko } \{\mathbf{x}, \mathbf{y}, \mathbf{1}\} \text{ on sidottu,} \tag{8.40}$$

eli vektorit $\mathbf{x}$, $\mathbf{y}$ ja $\mathbf{1}$ kuuluvat samaan tasoon. Kirjataan tämä tulos vielä täsmällisesti:

Olkoon $\text{var}_\text{d}(\mathbf{x}) \neq 0$ *ja* $\text{var}_\text{d}(\mathbf{y}) \neq 0$. *Tällöin*

$$\text{cor}_\text{d}{}^2(\mathbf{x}, \mathbf{y}) = 1 \iff \mathbf{y} \in \mathscr{C}(\mathbf{1} : \mathbf{x}) \text{ eli } \text{r}(\mathbf{1} : \mathbf{x} : \mathbf{y}) = 2\,. \tag{8.41}$$

Toisaalta on mahdollista, että muuttujien korrelaatio on 0, mutta kosini on hyvin lähellä 1:tä (ks. kuvio 8.5). Esimerkiksi jos

$$(\mathbf{x} : \mathbf{y}) = \begin{pmatrix} 1.001 & 1.001 \\ 0.999 & 1.001 \\ 1.000 & 0.998 \end{pmatrix}, \quad (\tilde{\mathbf{x}} : \tilde{\mathbf{y}}) = \begin{pmatrix} 0.001 & 0.001 \\ -0.001 & 0.001 \\ 0.000 & -0.002 \end{pmatrix}, \tag{8.42a}$$

niin

$$\cos(\mathbf{x}, \mathbf{y}) = 0.999999, \qquad \text{cor}_\text{d}(\mathbf{x}, \mathbf{y}) = \cos(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = 0\,. \tag{8.42b}$$

Jos $\cos(\mathbf{x}, \mathbf{y})$ on tasan 1, niin silloinhan korrelaatio on välttämättä myös 1.

On helppo vakuuttautua, että jos $\mathbf{x}$ on annettu muuttujavektori $[\text{var}(\mathbf{x}) \neq 0]$, niin kaikkien niiden $\mathbf{y}$-vektorien joukko, joilla on 0-korrelaatio $\mathbf{x}$:n kanssa on

$$\{\, \mathbf{y} : \text{cor}_\text{d}(\mathbf{x}, \mathbf{y}) = 0 \,\} = \mathscr{C}(\mathbf{C}\mathbf{x})^\perp; \tag{8.43}$$

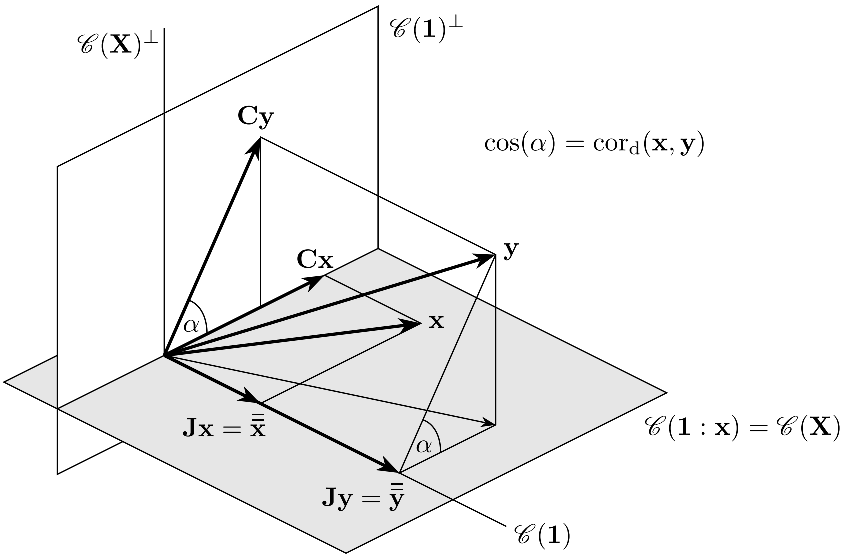
joukosta $\mathscr{C}(\mathbf{C}\mathbf{x})^\perp$ on kuitenkin poistettava $\mathbf{0}$-vektori, sillä $\text{cor}_\text{d}(\mathbf{x}, \mathbf{0}) = 0/0$.

8.2 Muuttujien standardointi

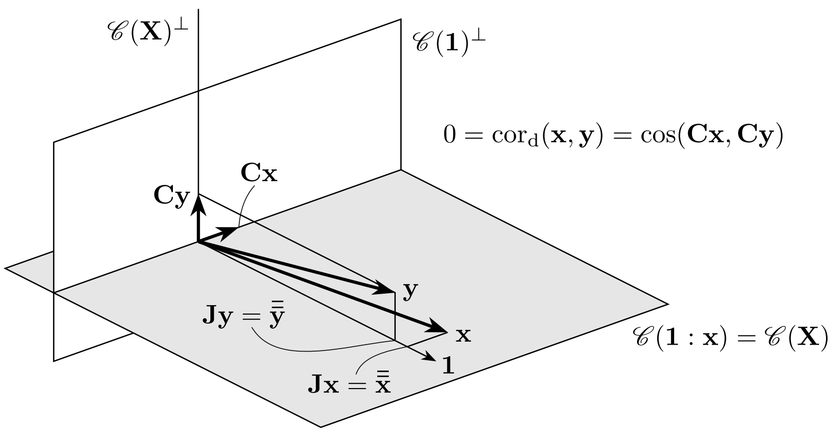
Tarkastellaan keskistetyn muuttujavektorin $\tilde{\mathbf{y}}$ skaalattuja versioita $\tilde{\tilde{\mathbf{y}}}$ ja $\overset{*}{\mathbf{y}}$, jotka on määritelty seuraavasti:

$$\tilde{\tilde{y}}_i = \frac{y_i - \bar{y}}{t_{yy}} = \frac{1}{\sqrt{t_{yy}}} y_i - \frac{\bar{y}}{\sqrt{t_{yy}}} := \alpha y_i - \beta\,, \tag{8.44a}$$

$$\overset{*}{y}_i = \frac{y_i - \bar{y}}{s_y} = \frac{1}{s_y} y_i - \frac{\bar{y}}{s_y} := \gamma y_i - \delta\,. \tag{8.44b}$$



Kuvio 8.4. Muuttujien $\mathbf{x}$ ja $\mathbf{y}$ välinen korrelaatio geometrisesti.



Kuvio 8.5. Vektoreiden $\mathbf{x}$ ja $\mathbf{y}$ välinen kosini on suuri, mutta vastaavat muuttujat ovat korreloimattomia. Vektorit $\mathbf{x}$ ja $\mathbf{y}$ ovat ”hyvin lähellä” vektoria $\mathbf{1}$.

Muuttujien $\tilde{\tilde{y}}$ ja $\overset{*}{y}$ keskiarvot ovat tietenkin nollia. Lisäksi vanhastaan tiedämme, että muunnoksilla (8.44a) ja (8.44b) on seuraavat vaikutukset variansseihin:

$$\operatorname{var}_{\mathrm{s}}(\tilde{\tilde{y}}) = \alpha^2 \operatorname{var}_{\mathrm{s}}(y) = \frac{1}{t_{yy}}\left(\frac{1}{n-1}t_{yy}\right) = \frac{1}{n-1}, \tag{8.45a}$$

$$\operatorname{var}_{\mathrm{s}}(\overset{*}{y}) = \gamma^2 \operatorname{var}_{\mathrm{s}}(y) = \frac{1}{s_y^2}\, s_y^2 = 1\,. \tag{8.45b}$$

Käsittelemme nyt matriisimerkinnöin joitakin näiden muuttujien ominaisuuksia. Vektori $\tilde{\tilde{\mathbf{y}}}$ esitettävissä muodossa $\tilde{\tilde{\mathbf{y}}} = \alpha\mathbf{Cy}$. Täten muuttujavektorin $\tilde{\tilde{\mathbf{y}}}$ pituuden neliö on

$$\|\tilde{\tilde{\mathbf{y}}}\|^2 = \tilde{\tilde{\mathbf{y}}}'\tilde{\tilde{\mathbf{y}}} = \alpha^2\mathbf{yCy} = \frac{1}{t_{yy}}\mathbf{y}'\mathbf{Cy} = \frac{1}{t_{yy}}t_{yy} = 1\,, \tag{8.46}$$

ja varianssi on

$$\operatorname{var}_{\mathrm{d}}(\tilde{\tilde{\mathbf{y}}}) = \frac{1}{n-1}\tilde{\tilde{\mathbf{y}}}'\mathbf{C}\tilde{\tilde{\mathbf{y}}} = \frac{1}{n-1}\tilde{\tilde{\mathbf{y}}}'\tilde{\tilde{\mathbf{y}}} = \frac{1}{n-1}\,, \tag{8.47}$$

koska $\mathbf{C}\tilde{\tilde{\mathbf{y}}} = \tilde{\tilde{\mathbf{y}}}$. Muuttujaa $\overset{*}{y}$ vastaavan muuttujavektorin $\overset{*}{\mathbf{y}} = \beta\mathbf{Cy}$ pituuden neliö on

$$\|\overset{*}{\mathbf{y}}\|^2 = \overset{*}{\mathbf{y}}'\overset{*}{\mathbf{y}} = \beta^2\mathbf{y}'\mathbf{Cy} = \frac{1}{s_y^2}\mathbf{y}'\mathbf{Cy} = \frac{1}{s_y^2}(n-1)s_y^2 = n-1\,, \tag{8.48}$$

ja varianssi on tietenkin

$$\operatorname{var}(\overset{*}{\mathbf{y}}) = \frac{1}{n-1}\overset{*}{\mathbf{y}}'\mathbf{C}\overset{*}{\mathbf{y}} = \frac{1}{n-1}\overset{*}{\mathbf{y}}'\overset{*}{\mathbf{y}} = \frac{1}{n-1}(n-1) = 1\,. \tag{8.49}$$

On selvää, että joissakin yhteyksissä (varsinkin jos halutaan saattaa muuttujat samanarvoiseen asemaan hajontojensa suhteen) on kätevää tarkastella alkuperäisten muuttujien sijasta muuttujia, jotka ovat tyyppiä (8.44a) tai (8.44b). Voimme kutsua näitä muuttujia *standardoiduiksi* muuttujiksi. Tällöin on kuitenkin huomattava, että standardoinnilla voidaan tarkoittaa keskistetyn muuttujavektorin skaalaamista

- joko pituudeltaan ykköseksi (jolloin varianssi on $\frac{1}{n-1}$)
- tai hajonnaltaan ykköseksi (jolloin pituuden neliö on $n-1$).

Muuttujavektorin $\tilde{\tilde{\mathbf{y}}}$ keskiarvo on siis 0 ja pituus on 1. Olkoon $\tilde{\tilde{\mathbf{x}}}$ vastaavalla tavalla skaalattu muuttujavektori $\tilde{\tilde{\mathbf{x}}} = \frac{1}{\sqrt{t_{xx}}}\mathbf{Cx}$. Tällöin $\mathbf{x}$:n ja $\mathbf{y}$:n välinen korrelaatio on yksinkertaisesti $\tilde{\tilde{\mathbf{x}}}$:n ja $\tilde{\tilde{\mathbf{y}}}$:n sisätulo

$$r_{xy} = \tilde{\tilde{\mathbf{x}}}'\tilde{\tilde{\mathbf{y}}}\,. \tag{8.50}$$

Muuttujien $\tilde{\tilde{\mathbf{x}}}$:n ja $\tilde{\tilde{\mathbf{y}}}$:n välinen korrelaatio on tietenkin sama kuin $\mathbf{x}$:n ja $\mathbf{y}$:n välinen korrelaatio. Muuttujien $\tilde{\tilde{\mathbf{x}}}$:n ja $\tilde{\tilde{\mathbf{y}}}$:n välinen kovarianssi on tällöin

$$\operatorname{cov_d}(\tilde{\tilde{\mathbf{x}}}, \tilde{\tilde{\mathbf{y}}}) = \tfrac{1}{n-1}\, \tilde{\tilde{\mathbf{x}}}'\tilde{\tilde{\mathbf{y}}} = \tfrac{1}{n-1} r_{xy}\,. \tag{8.51}$$

Hajonnoiltaan 1:ksi standardoitujen muuttujien $\overset{*}{\mathbf{x}}$ ja $\overset{*}{\mathbf{y}}$ välinen korrelaatiokerroin on (koska $\overset{*}{\mathbf{x}}'\overset{*}{\mathbf{x}} = \overset{*}{\mathbf{y}}'\overset{*}{\mathbf{y}} = n-1$)

$$\operatorname{cor_d}(\overset{*}{\mathbf{x}}, \overset{*}{\mathbf{y}}) = \frac{\overset{*}{\mathbf{x}}'\mathbf{C}\overset{*}{\mathbf{y}}}{\sqrt{\overset{*}{\mathbf{x}}'\mathbf{C}\overset{*}{\mathbf{x}} \cdot \overset{*}{\mathbf{y}}'\mathbf{C}\overset{*}{\mathbf{y}}}} = \frac{\overset{*}{\mathbf{x}}'\overset{*}{\mathbf{y}}}{\sqrt{\overset{*}{\mathbf{x}}'\overset{*}{\mathbf{x}} \cdot \overset{*}{\mathbf{y}}'\overset{*}{\mathbf{y}}}} = \tfrac{1}{n-1}\, \overset{*}{\mathbf{x}}'\overset{*}{\mathbf{y}} = r_{xy}\,, \tag{8.52}$$

ja niiden välinen kovarianssi on

$$\operatorname{cov_d}(\overset{*}{\mathbf{x}}, \overset{*}{\mathbf{y}}) = \tfrac{1}{n-1}\overset{*}{\mathbf{x}}'\overset{*}{\mathbf{y}} = r_{xy}\,. \tag{8.53}$$

8.3 Osittaiskorrelaatio

Olkoot $\mathbf{x}$, $\mathbf{y}$ ja $\mathbf{z}$ keskistettyjä ja ykkösen pituisia muuttujavektoreita, jolloin

$$\mathbf{x}'\mathbf{y} = r_{xy},\;\; \mathbf{x}'\mathbf{z} = r_{xz},\;\; \mathbf{y}'\mathbf{z} = r_{yz}, \tag{8.54a}$$
$$\mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{y} = \mathbf{z}'\mathbf{z} = 1, \tag{8.54b}$$
$$\mathbf{I} - \mathbf{P_z} = \mathbf{I} - \mathbf{z}\mathbf{z}' := \mathbf{M_z}. \tag{8.54c}$$

Merkitään

$$\begin{aligned} \mathbf{e_{x\cdot z}} &= \operatorname{res}(\mathbf{x};\mathbf{z}) \\ &= \mathbf{x} - \mathbf{P_z}\mathbf{x} = (\mathbf{I} - \mathbf{P_z})\mathbf{x} = \mathbf{M_z}\mathbf{x} = (\mathbf{I} - \mathbf{z}\mathbf{z}')\mathbf{x} = \mathbf{x} - r_{xz}\mathbf{z}, && (8.55\text{a}) \\ \mathbf{e_{y\cdot z}} &= \operatorname{res}(\mathbf{y};\mathbf{z}) \\ &= \mathbf{y} - \mathbf{P_z}\mathbf{y} = (\mathbf{I} - \mathbf{P_z})\mathbf{y} = \mathbf{M_z}\mathbf{y} = (\mathbf{I} - \mathbf{z}\mathbf{z}')\mathbf{y} = \mathbf{y} - r_{yz}\mathbf{z}. && (8.55\text{b}) \end{aligned}$$

Tällöin regressioanalyysin termein

$$\begin{aligned} \mathbf{e_{x\cdot z}} &= \text{residuaali kun } \mathbf{x}\text{:ää selitetään } \mathbf{z}\text{:lla}, \\ \mathbf{e_{y\cdot z}} &= \text{residuaali kun } \mathbf{y}\text{:tä selitetään } \mathbf{z}\text{:lla}. \end{aligned}$$

Näiden residuaalien korrelaatiokerroin on muuttujien x ja y välinen osittaiskorrelaatiokerroin kun z on vakioitu:

$$\boxed{\operatorname{cor_d}(\mathbf{e_{x\cdot z}}, \mathbf{e_{y\cdot z}}) = \operatorname{pcor_d}(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = r_{xy\cdot z}. \qquad (8.56)}$$

Osoitamme nyt (tuloshan on lukijalle jo tuttu), että näin määritelty osittaiskorrelaatiokerroin on

$$r_{xy\cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}. \tag{8.57}$$

Koska $\mathbf{x}$, $\mathbf{y}$ ja $\mathbf{z}$ ovat keskistettyjä, ovat residuaalit $\mathbf{e}_{\mathbf{x}\cdot\mathbf{z}}$ ja $\mathbf{e}_{\mathbf{y}\cdot\mathbf{z}}$ myös keskistettyjä eli esim. $\mathbf{C}\mathbf{e}_{\mathbf{x}\cdot\mathbf{z}} = \mathbf{e}_{\mathbf{x}\cdot\mathbf{z}}$. Täten

$$\begin{aligned}\operatorname{cor}_{\mathrm{d}}(\mathbf{e}_{\mathbf{x}\cdot\mathbf{z}}, \mathbf{e}_{\mathbf{y}\cdot\mathbf{z}}) &= \cos(\mathbf{e}_{\mathbf{x}\cdot\mathbf{z}}, \mathbf{e}_{\mathbf{y}\cdot\mathbf{z}}) \\ &= \frac{\mathbf{e}'_{\mathbf{x}\cdot\mathbf{z}}\mathbf{e}_{\mathbf{y}\cdot\mathbf{z}}}{\sqrt{\mathbf{e}'_{\mathbf{x}\cdot\mathbf{z}}\mathbf{e}_{\mathbf{x}\cdot\mathbf{z}} \cdot \mathbf{e}'_{\mathbf{y}\cdot\mathbf{z}}\mathbf{e}_{\mathbf{y}\cdot\mathbf{z}}}} = \frac{a}{\sqrt{b \cdot c}},\end{aligned} \tag{8.58}$$

missä

$$\begin{aligned} a &= [(\mathbf{I} - \mathbf{P_z})\mathbf{x}]'(\mathbf{I} - \mathbf{P_z})\mathbf{y} = \mathbf{x}'(\mathbf{I} - \mathbf{P_z})\mathbf{y} \\ &= \mathbf{x}'(\mathbf{I} - \mathbf{zz}')\mathbf{y} = \mathbf{x}'\mathbf{y} - \mathbf{x}'\mathbf{z} \cdot \mathbf{z}'\mathbf{y} = r_{xy} - r_{xz}r_{yz}, && \text{(8.59a)} \\ b &= \mathbf{x}'(\mathbf{I} - \mathbf{zz}')\mathbf{x} = \mathbf{x}'\mathbf{x} - \mathbf{x}'\mathbf{z} \cdot \mathbf{z}'\mathbf{x} = 1 - r_{xz}^2, && \text{(8.59b)} \\ c &= \mathbf{y}'(\mathbf{I} - \mathbf{zz}')\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{z} \cdot \mathbf{z}'\mathbf{y} = 1 - r_{yz}^2. && \text{(8.59c)} \end{aligned}$$

Osittaiskorrelaatiokerroin on siis tiettyjen residuaalien välinen tavallinen korrelaatiokerroin. Osittaiskorrelaatiokerroin

> $r_{xy\cdot z}$ kertoo x:n ja y:n välisen lineaarisen riippuvuuden voimakkuudesta kun z:n vaikutus on eliminoitu.

On helppo huomata, että osittaiskorrelaatiokertoimen ja tavallisen korrelaatiokertoimen välillä on selvä vastaavuus: r_{xy} kertoo x:n ja y:n välisen lineaarisen riippuvuuden voimakkuudesta, kun "vakiomuuttuja $\mathbf{1}$ on eliminoitu" – vektorilla $\mathbf{1}$ on sama rooli r_{xy}:tä laskettaessa kuin $\mathbf{z}$:lla on $r_{xy\cdot z}$:aan nähden. Tämä ilmenee hyvin myös tilannetta havainnollistavista kuvioista. Voimme tulkita keskistetyn arvon tiettynä residuaalina:

$$\begin{aligned}\tilde{\mathbf{y}} &= \operatorname{res}(\mathbf{y}; \mathbf{1}) \\ &= \mathbf{y} - \mathbf{P_1}\mathbf{y} = \mathbf{y} - \mathbf{J}\mathbf{y} = \text{residuaali kun } \mathbf{y}\text{:tä selitetään } \mathbf{1}\text{:llä.}\end{aligned} \tag{8.60}$$

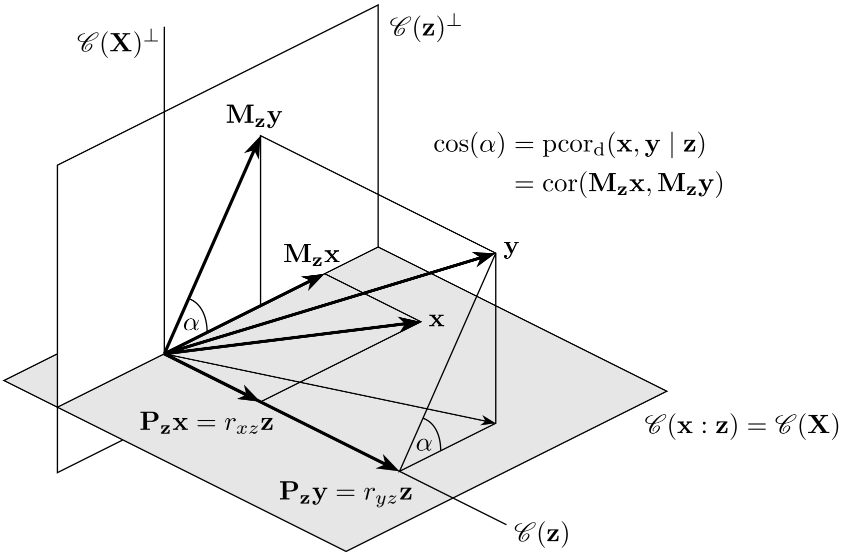
8.4 Korrelaatiomatriisi ja joitakin hajotelmia

Kuten muistamme, korrelaatiokertoimet havaintomatriisista $\mathbf{U}$ saadaan matriisioperaatioin seuraavasti:

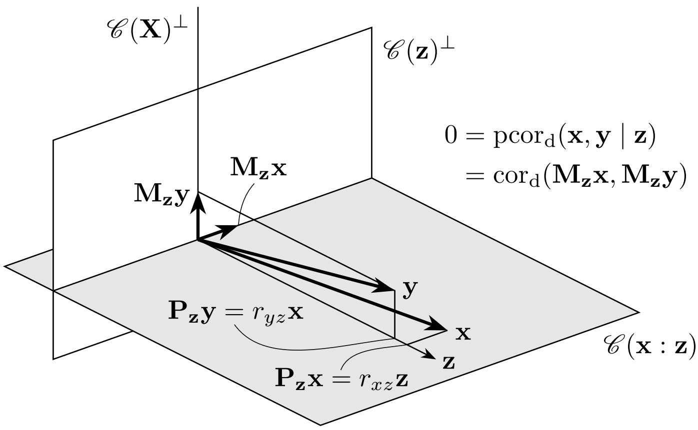
$$\begin{aligned}\mathbf{R} = \operatorname{cor}_{\mathrm{d}}(\mathbf{U}) &= [\operatorname{diag}(\mathbf{U}'\mathbf{C}\mathbf{U})]^{-1/2}\mathbf{U}'\mathbf{C}\mathbf{U}[\operatorname{diag}(\mathbf{U}'\mathbf{C}\mathbf{U})]^{-1/2} \\ &= [\operatorname{diag}(\mathbf{T})]^{-1/2}\mathbf{T}[\operatorname{diag}(\mathbf{T})]^{-1/2} \\ &= [\operatorname{diag}(\mathbf{S})]^{-1/2}\mathbf{S}[\operatorname{diag}(\mathbf{S})]^{-1/2}.\end{aligned} \tag{8.61}$$

Tarkastellaan sitten keskistetyn $\mathbf{U}$:n skaalattuja versioita:

$$\tilde{\tilde{\mathbf{U}}} = \widetilde{\mathbf{U}}[\operatorname{diag}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}})]^{-1/2} = \widetilde{\mathbf{U}}[\operatorname{diag}(\mathbf{T})]^{-1/2}, \tag{8.62}$$



Kuvio 8.6. Osittaiskorrelaatio $r_{xy\cdot z}$ geometrisesti.



Kuvio 8.7. Osittaiskorrelaatio $r_{xy\cdot z}$ geometrisesti. Huomaa myös, että jos $\mathbf{y}$:tä selitetään $\mathbf{x}$:llä ja $\mathbf{y}$:llä, niin $\mathbf{x}$:n regressiokerroin on 0.

missä sarakkeiden pituudet ykkösiä ja hajonnat $\frac{1}{n-1}$;

$$\overset{*}{\mathbf{U}} = \widetilde{\mathbf{U}}[\mathrm{diag}(\mathbf{S})]^{-1/2} = \sqrt{n-1}\,\tilde{\tilde{\mathbf{U}}}, \tag{8.63}$$

missä sarakkeiden pituudet ovat $\sqrt{n-1}$ ja hajonnat ykkösiä. Tällöin

$$\mathbf{R} = \tilde{\tilde{\mathbf{U}}}'\tilde{\tilde{\mathbf{U}}} = \frac{1}{n-1}\overset{*}{\mathbf{U}}'\overset{*}{\mathbf{U}}, \tag{8.64}$$

ja

$$\mathrm{cov_d}(\tilde{\tilde{\mathbf{U}}}) = \frac{1}{n-1}\tilde{\tilde{\mathbf{U}}}'\tilde{\tilde{\mathbf{U}}} = \frac{1}{n-1}\mathbf{R}, \tag{8.65}$$

$$\mathrm{cov_d}(\overset{*}{\mathbf{U}}) = \frac{1}{n-1}\overset{*}{\mathbf{U}}'\overset{*}{\mathbf{U}} = \frac{1}{n-1}(n-1)\tilde{\tilde{\mathbf{U}}}'\tilde{\tilde{\mathbf{U}}} = \mathbf{R}. \tag{8.66}$$

Korostettakoon vielä kerran, että kumpikin skaalaustapa (8.62) ja (8.63) johtaa siihen, että kaikilla uusilla muuttujilla on keskenään samat varianssit: tapauksessa (8.62) varianssit ovat $\frac{1}{n-1}$ ja tapauksessa (8.63) varianssit ovat ykkösiä.

8.4.1 Ositettu havaintomatriisi

Olkoon $(k+1)$:n muuttujan havaintomatriisi $\mathbf{U}$ ositettu seuraavasti:

$$\mathbf{U} = (\mathbf{x}_1 : \ldots : \mathbf{x}_k : \mathbf{y}) = (\mathbf{X}_0 : \mathbf{y}). \tag{8.67}$$

Osoitetaan aluksi, että havaintomatriisin $\mathbf{U} = (\mathbf{X}_0 : \mathbf{y})$ tulosummamatriisi on

$$\mathbf{T} = \mathbf{U}'\mathbf{C}\mathbf{U} = \widetilde{\mathbf{U}}'\widetilde{\mathbf{U}} = \begin{pmatrix} \mathbf{X}_0'\mathbf{C}\mathbf{X}_0 & \mathbf{X}_0'\mathbf{C}\mathbf{y} \\ \mathbf{y}'\mathbf{C}\mathbf{X}_0 & \mathbf{y}'\mathbf{C}\mathbf{y} \end{pmatrix} = \begin{pmatrix} \widetilde{\mathbf{X}}_0'\widetilde{\mathbf{X}}_0 & \widetilde{\mathbf{X}}_0'\tilde{\mathbf{y}} \\ \tilde{\mathbf{y}}'\widetilde{\mathbf{X}}_0 & \tilde{\mathbf{y}}'\tilde{\mathbf{y}} \end{pmatrix}. \tag{8.68}$$

Koska

$$\widetilde{\mathbf{U}} = \mathbf{C}\mathbf{U} = (\mathbf{C}\mathbf{X}_0 : \mathbf{C}\mathbf{y}) = (\widetilde{\mathbf{X}}_0 : \mathbf{y} - \bar{\bar{\mathbf{y}}}) = (\widetilde{\mathbf{X}}_0 : \tilde{\mathbf{y}}), \tag{8.69}$$

on todellakin

$$\begin{aligned} \mathbf{T} = \widetilde{\mathbf{U}}'\widetilde{\mathbf{U}} &= (\mathbf{C}\mathbf{X}_0 : \mathbf{C}\mathbf{y})'(\mathbf{C}\mathbf{X}_0 : \mathbf{C}\mathbf{y}) \\ &= \begin{pmatrix} \mathbf{X}_0'\mathbf{C} \\ \mathbf{y}'\mathbf{C} \end{pmatrix}(\mathbf{C}\mathbf{X}_0 : \mathbf{C}\mathbf{y}) \\ &= \begin{pmatrix} \mathbf{X}_0'\mathbf{C}\mathbf{X}_0 & \mathbf{X}_0'\mathbf{C}\mathbf{y} \\ \mathbf{y}'\mathbf{C}\mathbf{X}_0 & \mathbf{y}'\mathbf{C}\mathbf{y} \end{pmatrix} = \begin{pmatrix} \widetilde{\mathbf{X}}_0'\widetilde{\mathbf{X}}_0 & \widetilde{\mathbf{X}}_0'\tilde{\mathbf{y}} \\ \tilde{\mathbf{y}}'\widetilde{\mathbf{X}}_0 & \tilde{\mathbf{y}}'\tilde{\mathbf{y}} \end{pmatrix}. \end{aligned} \tag{8.70}$$

Voimme käyttää merkintää

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{\mathbf{xx}} & \mathbf{t}_{\mathbf{x}y} \\ \mathbf{t}_{\mathbf{x}y}' & t_{yy} \end{pmatrix}. \tag{8.71}$$

Matriisin $\mathbf{X}_0$ sisältämien muuttujien ja $\mathbf{y}$:n välinen tulosummavektori on siis

$$\mathbf{t}_{\mathbf{x}y} = \mathbf{X}_0'\mathbf{C}\mathbf{y} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_k' \end{pmatrix} \mathbf{C}\mathbf{y} = \begin{pmatrix} \mathbf{x}_1'\mathbf{C}\mathbf{y} \\ \mathbf{x}_2'\mathbf{C}\mathbf{y} \\ \vdots \\ \mathbf{x}_k'\mathbf{C}\mathbf{y} \end{pmatrix} = \begin{pmatrix} t_{1y} \\ t_{2y} \\ \vdots \\ t_{ky} \end{pmatrix}, \tag{8.72}$$

ja vastaava kovarianssivektori on

$$\mathbf{s}_{\mathbf{x}y} = \frac{1}{n-1}\mathbf{t}_{\mathbf{x}y} = \begin{pmatrix} s_{1y} \\ s_{2y} \\ \vdots \\ s_{ky} \end{pmatrix} = \operatorname{cov_d}(\mathbf{X}_0, \mathbf{y}). \tag{8.73}$$

Jos korrelaatiomatriisi $\mathbf{R}$ on ositettu vastaavalla tavalla, niin on osoitettavissa, että

$$\mathbf{r}_{\mathbf{x}y}'\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{\mathbf{x}y} = \frac{\mathbf{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}}{t_{yy}} = \frac{\mathbf{s}_{\mathbf{x}y}'\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{s}_{\mathbf{x}y}}{s_y^2} = R_{y;\mathbf{x}}^2\,, \tag{8.74a}$$

$$1 - \mathbf{r}_{\mathbf{x}y}'\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{\mathbf{x}y} = \frac{t_{yy} - \mathbf{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}}{t_{yy}}. \tag{8.74b}$$

Lauseke $R_{y;\mathbf{x}}^2$ on yhteiskorrelaatiokertoimen neliö (selitysaste), kun y:tä selitetään muttujilla $x_1, \dots, x_k$ sekä vakiolla.

8.4.2 Lineaarikombinaation kovarianssimatriisi

Olkoon muuttuja y muuttujien u_1, u_2, …, u_p lineaarikombinaatio:

$$y = a_1u_1 + a_2u_2 + \dots + a_pu_p. \tag{8.75}$$

Kun $\mathbf{y}$ on y:n havaitut arvot sisältävä muuttujavektori ja $\mathbf{U} = (\mathbf{u}_1 : \mathbf{u}_2 : \dots : \mathbf{u}_p)$ on u-muuttujien havaintomatriisi, saamme $\mathbf{y}$:lle lausekkeen

$$\mathbf{y} = \mathbf{U}\mathbf{a} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \dots + a_p\mathbf{u}_p. \tag{8.76}$$

Tällöin $\mathbf{y}$:n (keskiarvosta laskettujen poikkeamien) neliösumma on

$$t_{yy} = \mathbf{y}'\mathbf{C}\mathbf{y} = \mathbf{a}'\mathbf{U}'\mathbf{C}\mathbf{U}\mathbf{a} = \mathbf{a}'\mathbf{T}\mathbf{a}, \tag{8.77a}$$

joten $\mathbf{y}$:n otosvarianssi on

$$\operatorname{var_s}(y) = \tfrac{1}{n-1}\mathbf{y}'\mathbf{C}\mathbf{y} = \tfrac{1}{n-1}\mathbf{a}'\mathbf{U}'\mathbf{C}\mathbf{U}\mathbf{a} = \mathbf{a}'\mathbf{S}\mathbf{a}\,, \tag{8.77b}$$

missä $\mathbf{S} = \operatorname{cov_d}(\mathbf{U})$. Erityisesti kahden muuttujan tapauksessa saamme

$$\operatorname{var_d}(\mathbf{U}_{n\times 2}\mathbf{1}_2) = \mathbf{1}'\mathbf{S}\mathbf{1} = s_1^2 + s_2^2 + 2s_{12} = s_1^2 + s_2^2 + 2r_{12}s_1s_2, \tag{8.78}$$

joten voimme välittömästi päätellä vanhan tuloksen:

$$\begin{aligned}&\text{Summan } \mathbf{u}_1+\mathbf{u}_2 \text{ otosvarianssi on otosvarianssien } (\neq 0) \text{ summa}\\ &\iff \text{ yhteenlaskettavat ovat korreloimattomia.}\end{aligned} \tag{8.79}$$

Jos $s_1^2 = 0$ tai $s_2^2 = 0$, niin korrelaatiokerroin ($\approx 0/0$) ei ole määritelty – silti varianssien summa on tietysti summan varianssi.

Tulos (8.79) ei päde sellaisenaan useamman muuttujan tapauksessa (pätee vain "$\Leftarrow$"-suunnassa). Esimerkiksi jos

$$\operatorname{cov_d}(\mathbf{u}_1 : \mathbf{u}_2 : \mathbf{u}_3) = \mathbf{S} = \begin{pmatrix} s_1^2 & 0 & -r \\ 0 & s_2^2 & r \\ -r & r & s_3^2 \end{pmatrix}, \tag{8.80a}$$

niin

$$\operatorname{var}(\mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3) = \mathbf{1}'\mathbf{S}\mathbf{1} = s_1^2 + s_2^2 + s_3^2, \tag{8.80b}$$

mutta silti yhteenlaskettavat eivät ole korreloimattomia.

Jos $\mathbf{u}$ on p elementin satunnaisvektori ja y on sen elementtien lineaarikombinaatio $y = \mathbf{a}'\mathbf{u}$, niin saamme luvusta 7 tutun lausekkeen

$$\operatorname{var}(y) = \operatorname{var}(\mathbf{a}'\mathbf{u}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}, \tag{8.81}$$

missä $\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{u})$. Lauseke (8.81) on siis teoreettinen satunnaismuuttujan varianssi ja (8.77b) on otoksesta laskettu vastaava otosvarianssi.

Olkoon $\mathbf{x}$ jokin toinen $\mathbf{u}_i$-muuttujien lineaarikombinaatio: $\mathbf{x} = \mathbf{U}\mathbf{b}$. Tällöin

$$t_{xy} = \mathbf{x}'\mathbf{C}\mathbf{y} = \mathbf{a}'\mathbf{U}'\mathbf{C}\mathbf{U}\mathbf{b} = \mathbf{a}'\mathbf{T}\mathbf{b}, \tag{8.82}$$

joten

$$\operatorname{cov_d}(\mathbf{x}, \mathbf{y}) = \tfrac{1}{n-1}\mathbf{x}'\mathbf{C}\mathbf{y} = \tfrac{1}{n-1}\mathbf{a}'\mathbf{U}'\mathbf{C}\mathbf{U}\mathbf{b} = \mathbf{a}'\mathbf{S}\mathbf{b}, \tag{8.83}$$

ja

$$\operatorname{cor_d}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{a}'\mathbf{S}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a} \cdot \mathbf{b}'\mathbf{S}\mathbf{b}}} = \frac{\mathbf{a}'\mathbf{T}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{T}\mathbf{a} \cdot \mathbf{b}'\mathbf{T}\mathbf{b}}}. \tag{8.84}$$

Matriisimuodossa voimme kirjoittaa

$$\begin{aligned}\operatorname{cov_d}(\mathbf{x} : \mathbf{y}) &= \operatorname{cov_d}(\mathbf{U}\mathbf{a} : \mathbf{U}\mathbf{b}) = \operatorname{cov_d}\big(\mathbf{U}(\mathbf{a} : \mathbf{b})\big)\\ &= \operatorname{cov_d}(\mathbf{U}\mathbf{A}) = (\mathbf{a} : \mathbf{b})' \operatorname{cov_d}(\mathbf{U})(\mathbf{a} : \mathbf{b})\\ &= \begin{pmatrix} \mathbf{a}' \\ \mathbf{b}' \end{pmatrix} \mathbf{S}(\mathbf{a} : \mathbf{b}) = \begin{pmatrix} \mathbf{a}'\mathbf{S}\mathbf{a} & \mathbf{a}'\mathbf{S}\mathbf{b} \\ \mathbf{b}'\mathbf{S}\mathbf{a} & \mathbf{b}'\mathbf{S}\mathbf{b} \end{pmatrix},\end{aligned} \tag{8.85a}$$

eli

$$\operatorname{cov_d}(\mathbf{U}\mathbf{A}) = \mathbf{A}' \operatorname{cov_d}(\mathbf{U})\mathbf{A} = \mathbf{A}'\mathbf{S}\mathbf{A}. \tag{8.85b}$$

Jos nyt $\mathbf{u}$ on p elementin satunnaisvektori ja y ja x ovat sen elementtien lineaarikombinaatioita:

$$x = \mathbf{a}'\mathbf{u}, \quad y = \mathbf{b}'\mathbf{u}, \tag{8.86a}$$

niin

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mathbf{a}'\mathbf{u} \\ \mathbf{b}'\mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{a}' \\ \mathbf{b}' \end{pmatrix} \mathbf{u} = \mathbf{A}'\mathbf{u}, \tag{8.86b}$$

$$\operatorname{cov} \begin{pmatrix} x \\ y \end{pmatrix} = \operatorname{cov}(\mathbf{A}'\mathbf{u}) = \mathbf{A}'\mathbf{\Sigma}\mathbf{A}. \tag{8.86c}$$

Toisin sanoen:

- $\mathbf{A}'\mathbf{u}$ on satunnaisvektori elementteinään $\mathbf{u}$:n elementtien lineaarikombinaatiot
- satunnaisvektorin $\mathbf{A}'\mathbf{u}$ kovarianssimatriisi on $\mathbf{A}'\mathbf{\Sigma}\mathbf{A} = \operatorname{cov}(\mathbf{A}'\mathbf{u})$
- $\mathbf{U}' = (\mathbf{u}_{(1)} : \mathbf{u}_{(2)} : \dots : \mathbf{u}_{(n)})$ on $\mathbf{u}$:n havaittujen arvojen muodostaman havaintomatriisin transpoosi
- $\mathbf{AU}' = (\mathbf{Au}_{(1)} : \mathbf{Au}_{(2)} : \dots : \mathbf{Au}_{(n)})$ on $\mathbf{Au}$:n havaittujen arvojen muodostaman havaintomatriisin transpoosi
- $\mathbf{UA}'$ on $\mathbf{Au}$:n havaittujen arvojen muodostama havaintomatriisi
- $\mathbf{UA}'$:sta laskettu otoskovarianssimatriisi on $\mathbf{A}'\mathbf{SA} = \operatorname{cov}_{\mathrm{d}}(\mathbf{UA})$.

8.4.3 Korrelaatiomatriisin käänteismatriisi

Korrelaatiomatriisin käänteismatriisilla on yllättävän monia tilastotieteellisesti mielenkiintoisia ominaisuuksia. Tarkastelemme tässä joitakin sellaisia hyvin lyhyesti.

Merkitään

$$\mathbf{R}^{-1} = \begin{pmatrix} \mathbf{R_{xx}} & \mathbf{r_{x}}_y \\ \mathbf{r}'_{\mathbf{x}y} & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{R^{xx}} & \mathbf{r}^{\mathbf{x}y} \\ (\mathbf{r}^{\mathbf{x}y})' & r^{yy} \end{pmatrix} = \{\underline{r}^{ij}\}, \ \mathbf{R}_{\mathbf{xx}}^{-1} = \{r^{ij}\}, \tag{8.87a}$$

$$\mathbf{T}^{-1} = \begin{pmatrix} \mathbf{T_{xx}} & \mathbf{t}_{\mathbf{x}y} \\ \mathbf{t}'_{\mathbf{x}y} & t_{yy} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{T^{xx}} & \mathbf{t}^{\mathbf{x}y} \\ (\mathbf{t}^{\mathbf{x}y})' & t^{yy} \end{pmatrix} = \{\underline{t}^{ij}\}, \ \mathbf{T}_{\mathbf{xx}}^{-1} = \{t^{ij}\}. \tag{8.87b}$$

On erityisesti korostettava alleviivattujen ja alleviivaamattomien r^{ij}:tten eroa:

$$\mathbf{R}_{\mathbf{xx}}^{-1} = \mathbf{R^{xx}} \iff \mathbf{r}_{\mathbf{x}y} = \mathbf{0}. \tag{8.88}$$

Vastaavasti:

$$\mathbf{T}_{\mathbf{xx}}^{-1} = \mathbf{T^{xx}} \qquad \iff \mathbf{t}_{\mathbf{x}y} = \mathbf{0}. \tag{8.89}$$

Osittaiskääntötulosten perusteella $\mathbf{R}^{-1}$:n viimeinen lävistäjäelementti r^{yy} on

$$r^{yy} = \frac{1}{r_{yy\cdot 12\ldots k}} = \frac{1}{1 - \mathbf{r}'_{\mathbf{x}y}\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{\mathbf{x}y}}. \tag{8.90}$$

missä

$$r_{yy\cdot 12\ldots k} = 1 - \mathbf{r}'_{\mathbf{x}y}\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{\mathbf{x}y} = \mathbf{R}_{\mathbf{xx}}\text{:n Schurin komplementti } \mathbf{R}\text{:ssä.} \tag{8.91}$$

On osoitettavissa, että

$$\begin{aligned}\mathbf{r}'_{\mathbf{x}y}\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{\mathbf{x}y} = R^2_{y\cdot\mathbf{x}} = {}& \text{selitysaste lineaarisessa mallissa,} \\ & \text{kun } \mathbf{y}\text{:tä selitetään } \mathbf{x}\text{-muuttujilla (+ vakiolla).}\end{aligned} \tag{8.92}$$

Täten $\mathbf{R}^{-1}$:n viimeisellä lävistäjäelementillä r^{yy} on mielenkiintoinen tulkinta (mikähän on luonnollisesti muiden lävistäjäelementtien tulkinta?):

$$r^{yy} = \frac{1}{1 - \mathbf{r}'_{\mathbf{x}y}\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{\mathbf{x}y}} = \frac{1}{1 - R^2_{y\cdot\mathbf{x}}}. \tag{8.93}$$

Jos siis $R^2_{y\cdot\mathbf{x}}$ $(0 \leq R^2_{y\cdot\mathbf{x}} \leq 1)$ on äärimmäisen lähellä lukua 1, niin r^{yy} (≥ 1) on astronomisen suuri. Mainittakoon, että selitysasteelle on voimassa esitys

$$R^2_{y\cdot\mathbf{x}} = \mathbf{r}'_{\mathbf{x}y}\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{\mathbf{x}y} = \frac{\mathbf{t}'_{\mathbf{x}y}\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}}{t_{yy}} = \frac{\mathbf{s}'_{\mathbf{x}y}\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{s}_{\mathbf{x}y}}{s^2_y} = \frac{\mathrm{SSR}}{\mathrm{SST}}, \tag{8.94}$$

missä $\mathrm{SSR} = \mathbf{t}'_{\mathbf{x}y}\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}$ on ns. regressioneliösumma ja $\mathrm{SST} = t_{yy}$ on kokonaisneliösumma, ks. (8.100) (s. 270).

Matriisin $\mathbf{T}^{-1}$ viimeinen lävistäjäelementti t^{yy} on vastaavasti

$$t^{yy} = \frac{1}{t_{yy\cdot\mathbf{x}}} = \frac{1}{t_{yy} - \mathbf{t}'_{\mathbf{x}y}\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}}, \tag{8.95a}$$

missä

$$t_{yy\cdot\mathbf{x}} = t_{yy} - \mathbf{t}'_{\mathbf{x}y}\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y} = \mathbf{T}_{\mathbf{xx}}\text{:n Schurin komplementti } \mathbf{T}\text{:ssä.} \tag{8.95b}$$

Edellä tarkasteltujen tunnuslukujen teoreettiset vastineet ovat luonnollisesti

$$\sigma^2_{y\cdot\mathbf{x}} = \sigma^2_y - \boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y} = \sigma^2_y(1 - \varrho^2_{y\cdot\mathbf{x}}) = \frac{1}{\sigma^{yy}}, \tag{8.96a}$$

$$\varrho^2_{y\cdot\mathbf{x}} = \frac{\boldsymbol{\sigma}'_{\mathbf{x}y}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\sigma}_{\mathbf{x}y}}{\sigma^2_y}, \tag{8.96b}$$

missä

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}'_{\mathbf{x}y} & \sigma^2_y \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}^{\mathbf{xx}} & \boldsymbol{\sigma}^{\mathbf{x}y} \\ (\boldsymbol{\sigma}^{\mathbf{x}y})' & \sigma^{yy} \end{pmatrix}. \tag{8.96c}$$

8.4.4 Projektorin H hajotelma

Luvun 9.13 (s. 330) mukaan ortogonaaliprojektori $\mathbf{P}_{(\mathbf{1}:\mathbf{X}_0)}$ voidaan esittää seuraavasti:

$$\begin{aligned}\mathbf{H} = \mathbf{P}_{(\mathbf{1}:\mathbf{X}_0)} &= \mathbf{P}_{\mathbf{1}} + \mathbf{P}_{(\mathbf{I}-\mathbf{P}_{\mathbf{1}})\mathbf{X}_0} = \mathbf{J} + \mathbf{P}_{\mathbf{C}\mathbf{X}_0} \\ &= \mathbf{J} + \mathbf{C}\mathbf{X}_0(\mathbf{X}_0'\mathbf{C}\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{C}\,. \end{aligned} \tag{8.97}$$

Täten

$$\mathbf{M} = \mathbf{I} - \mathbf{P}_{(\mathbf{1}:\mathbf{X}_0)} = \mathbf{I} - \mathbf{J} - \mathbf{P}_{\mathbf{C}\mathbf{X}_0} = \mathbf{C} - \mathbf{P}_{\mathbf{C}\mathbf{X}_0}\,, \tag{8.98}$$

ja

$$\begin{aligned}\mathrm{SSE} &= \mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y} = \mathbf{y}'(\mathbf{I}-\mathbf{P}_{(\mathbf{1}:\mathbf{X}_0)})\mathbf{y} \\ &= \mathbf{y}'(\mathbf{C}-\mathbf{P}_{\mathbf{C}\mathbf{X}_0})\mathbf{y} \\ &= \mathbf{y}'[\mathbf{C}-\mathbf{C}\mathbf{X}_0(\mathbf{X}_0'\mathbf{C}\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{C}]\mathbf{y} \\ &= \mathbf{y}'\mathbf{C}\mathbf{y} - \mathbf{y}'\mathbf{C}\mathbf{X}_0(\mathbf{X}_0'\mathbf{C}\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{C}\mathbf{y} \\ &= t_{yy} - \mathbf{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{t}_{\mathbf{x}y} \\ &= \mathrm{SST} - \mathrm{SSR} \\ &= t_{yy}\Big(1 - \frac{\mathbf{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{t}_{\mathbf{x}y}}{t_{yy}}\Big) \\ &= t_{yy}(1 - R_{y\cdot\mathbf{x}}^2) \\ &= t_{yy}\Big(1 - \frac{\mathrm{SSR}}{\mathrm{SST}}\Big) \\ &= t_{yy\cdot\mathbf{x}}\,, \end{aligned} \tag{8.99}$$

missä

$$\mathrm{SST} = t_{yy} = \mathbf{y}'(\mathbf{I}-\mathbf{J})\mathbf{y} = \text{ kokonaisneliösumma,} \tag{8.100a}$$
$$\mathrm{SSR} = \mathbf{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{t}_{\mathbf{x}y} = \mathbf{y}'(\mathbf{H}-\mathbf{J})\mathbf{y} = \text{ regressioneliösumma,} \tag{8.100b}$$
$$\mathrm{SSE} = \mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y} = \text{ jäännösneliösumma,} \tag{8.100c}$$
$$R_{y\cdot\mathbf{x}}^2 = \mathrm{SSR}\,/\,\mathrm{SST} = \text{ selitysaste.} \tag{8.100d}$$

Näin on päädytty regressioanalyysin neliösummahajotelmaan:

$$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}\,. \tag{8.101}$$

Vakuuttaudu, että (vakiotermisessä mallissa)

$$\mathrm{SST} = \|(\mathbf{I}-\mathbf{J})\mathbf{y}\|^2, \quad \mathrm{SSR} = \|(\mathbf{H}-\mathbf{J})\mathbf{y}\|^2, \quad \mathrm{SSR} = \|(\mathbf{I}-\mathbf{H})\mathbf{y}\|^2. \tag{8.102}$$

Ks. myös HT 8.6 (s. 299) ja kuvio 8.15 (s. 297).

Hajotelman (8.97) perusteella voidaan kätevästi johtaa lausekkeet regressiokertoimille $\hat{\beta}_0$ ja $\hat{\boldsymbol{\beta}}_{\mathbf{x}}$, kun

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}}_{\mathbf{x}} \end{pmatrix}, \quad \mathbf{Hy} = \mathbf{X}\hat{\boldsymbol{\beta}}\,. \tag{8.103}$$

Tällöin saamme seuraavat esitykset $\mathbf{Hy}$:lle:

$$\mathbf{Hy} = (\mathbf{1} : \mathbf{X}_0) \begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}}_{\mathbf{x}} \end{pmatrix} = \hat{\beta}_0\mathbf{1} + \mathbf{X}_0\hat{\boldsymbol{\beta}}_{\mathbf{x}}\,, \tag{8.104a}$$

$$\mathbf{Hy} = \mathbf{Jy} + \mathbf{CX}_0(\mathbf{X}_0'\mathbf{CX}_0)^{-1}\mathbf{X}_0'\mathbf{Cy}, \tag{8.104b}$$

joten

$$\hat{\beta}_0\mathbf{1} + \mathbf{X}_0\hat{\boldsymbol{\beta}}_{\mathbf{x}} = \mathbf{Jy} + \mathbf{CX}_0(\mathbf{X}_0'\mathbf{CX}_0)^{-1}\mathbf{X}_0'\mathbf{Cy}. \tag{8.105}$$

Kertomalla (8.105) vasemmalta keskistäjämatrisilla $\mathbf{C}$ saadaan

$$\mathbf{CX}_0\hat{\boldsymbol{\beta}}_{\mathbf{x}} = \mathbf{CX}_0(\mathbf{X}_0'\mathbf{CX}_0)^{-1}\mathbf{X}_0'\mathbf{Cy}. \tag{8.106}$$

Kun (8.106) kerrotaan vasemmalta matriisilla $\mathbf{X}_0'$, saadaan (aiemmin jo esitetty) tulos:

$$\hat{\boldsymbol{\beta}}_{\mathbf{x}} = (\mathbf{X}_0'\mathbf{CX}_0)^{-1}\mathbf{X}_0'\mathbf{Cy} = \mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}\,. \tag{8.107}$$

Kertomalla (8.105) vasemmalta vaakavektorilla $\mathbf{1}'$ aadaan

$$n\hat{\beta}_0 + \mathbf{1}'\mathbf{X}_0\hat{\boldsymbol{\beta}}_{\mathbf{x}} = \mathbf{1}'\mathbf{y}, \tag{8.108}$$

joten

$$\hat{\beta}_0 = \bar{y} - \bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}_{\mathbf{x}}\,, \quad \text{missä } \bar{\mathbf{x}}' = (\bar{x}_1, \dots, \bar{x}_k). \tag{8.109}$$

Erityisesti yhden selittäjän (+ vakion) tapauksessa saadaan

$$\mathbf{H} = \mathbf{P}_{(\mathbf{1}:\mathbf{x})} = \mathbf{J} + \mathbf{P}_{\mathbf{Cx}} = \mathbf{J} + \mathbf{Cx}(\mathbf{x}'\mathbf{Cx})^{-1}\mathbf{x}'\mathbf{C}\,, \tag{8.110}$$

ja

$$\begin{aligned} \mathrm{SSE} &= \mathbf{y}'\mathbf{My} \\ &= \mathbf{y}'\mathbf{Cy} - \mathbf{y}'\mathbf{Cx}(\mathbf{x}'\mathbf{Cx})^{-1}\mathbf{x}'\mathbf{Cy} \\ &= t_{yy} - t_{yx}t_{xx}^{-1}t_{xy} \\ &= t_{yy}\Big(1 - \frac{t_{xy}^2}{t_{xx}t_{yy}}\Big) \\ &= t_{yy}(1 - r_{xy}^2)\,. \end{aligned} \tag{8.111}$$

Voimme päätellä (8.111):sta, että aina on voimassa

$$\mathrm{SSE} = t_{yy}(1 - r_{xy}^2) \le t_{yy}\,. \tag{8.112}$$

Mainittakoon vielä, että (8.110):n nojalla

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{t_{xx}^2} = \frac{1}{n} + \frac{1}{n-1}\frac{(x_i - \bar{x})^2}{s_x^2}\,, \quad i = 1, \dots, n\,, \tag{8.113}$$

missä s_x^2 on x:n otosvarianssi.

8.4.5 Korrelaatiomatriisin determinantti

Ositetun matriisin determinantin lauseke johtaa välittömästi tulokseen

$$\begin{vmatrix} \mathbf{R_{xx}} & \mathbf{r_{x}}_y \\ \mathbf{r}'_{\mathbf{x}y} & 1 \end{vmatrix} = |\mathbf{R_{xx}}|(1 - \mathbf{r}'_{\mathbf{x}y}\mathbf{R}^{-1}_{\mathbf{xx}}\mathbf{r}_{\mathbf{x}y}) = |\mathbf{R_{xx}}|(1 - R^2_{y\cdot\mathbf{x}}). \tag{8.114}$$

Jos $\mathbf{R_{xx}}$ on edelleen ositettu vastaavalla tavalla, niin

$$\mathbf{R_{xx}} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{r}_{12} \\ \mathbf{r}'_{12} & 1 \end{pmatrix} \implies$$
$$|\mathbf{R_{xx}}| = |\mathbf{R}_{11}|(1 - \mathbf{r}'_{12}\mathbf{R}^{-1}_{11}\mathbf{r}_{12}) = |\mathbf{R}_{11}|(1 - R^2_k), \tag{8.115}$$

missä

$$R^2_k = R^2 \text{ kun } x_k\text{:ta selitetään muilla } x\text{-muuttujilla.} \tag{8.116}$$

Tulosummamatriisille ja kovarianssimatriisille saamme vastaavasti

$$|\mathbf{T}| = |\mathbf{T_{xx}}|(t_{yy} - \mathbf{t}'_{\mathbf{x}y}\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{t}_{\mathbf{x}y}) = |\mathbf{T_{xx}}|\,\mathrm{SSE}, \tag{8.117}$$
$$|\mathbf{S}| = |\mathbf{S_{xx}}|(s^2_y - \mathbf{s}'_{\mathbf{x}y}\mathbf{S}^{-1}_{\mathbf{xx}}\mathbf{s}_{\mathbf{x}y}). \tag{8.118}$$

Osoitamme nyt seuraavan tuloksen:

Tulosummamatriisin $\mathbf{T}$ *determinantille on voimassa epäyhtälö*

$$0 \leq |\mathbf{T}| \leq t_{11}t_{22}\cdots t_{kk}t_{yy}\,, \tag{8.119}$$

missä yläraja saavutetaan täsmälleen silloin kun $\mathbf{T}$ *on lävistäjämatriisi ja alaraja, kun matriisin*

$$\mathbf{U} = (\mathbf{x}_1 : \ldots : \mathbf{x}_k : \mathbf{y}) = (\mathbf{X}_0 : \mathbf{y}). \tag{8.120}$$

sarakkeet ovat lineaarisesti riippuvia tai $\mathbf{1}_n \in \mathscr{C}(\mathbf{U})$.

Jos $\mathbf{T}$ on singulaarinen, niin $|\mathbf{T}| = 0$, joten (8.119) toteutuu. Olkoon $\mathbf{T}$ positiivisesti definiitti. Tällöin

$$\begin{aligned} |\mathbf{T}| &= |\mathbf{T_{xx}}|(t_{yy} - \mathbf{t}'_{\mathbf{x}y}\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{t}_{\mathbf{x}y}) \\ &= |\mathbf{T_{xx}}|t_{yy}\Big(1 - \frac{\mathbf{t}'_{\mathbf{x}y}\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{t}_{\mathbf{x}y}}{t_{yy}}\Big) \\ &= |\mathbf{T_{xx}}|t_{yy}\Big(1 - \frac{\mathrm{SSR}}{\mathrm{SST}}\Big) \\ &= |\mathbf{T_{xx}}|\,t_{yy}(1 - R^2_{y\cdot\mathbf{x}}) \leq |\mathbf{T_{xx}}|\,t_{yy}. \end{aligned} \tag{8.121}$$

Toistamalla tätä päättelyä päädymme lopulta (8.119):n oikeanpuoleiseen epäyhtälöön. Huomattakoon, että $\mathbf{T}$:n positiivisesti definiittisyydestä seuraa, että kaikki $\mathbf{T}$:n lävistäjäelementit ja kaikki johtavat pääminorit ($\mathbf{T}$:n vasemmasta nurkasta muodostettujen alimatriisien determinantit) ovat nollaa suurempia. Alarajan saavuttamiseen palataan sivulla 273.

Epäyhtälöä (8.119) vastaten on kovarianssimatriisille tietysti voimassa

$$|\mathbf{S}| \leq s_1^2 s_2^2 \cdots s_k^2 s_y^2 \,, \tag{8.122}$$

sekä $|\mathbf{R}| \leq 1$, ja $|\mathbf{R}| = 1 \iff \mathbf{R} = \mathbf{I}_{k+1}$. Huomattakoon vielä, että

$$|\mathbf{T}| = t_{11} t_{22} \cdots t_{kk} t_{yy} |\mathbf{R}|, \quad |\mathbf{S}| = s_1^2 s_2^2 \cdots s_k^2 s_y^2 |\mathbf{R}|. \tag{8.123}$$

Vastaavat epäyhtälöt pätevät tietenkin satunnaisvektorien korrelaatio- ja kovarianssimatriiseille ja ylipäänsä ei-negatiivisesti definiiteille matriiseille.

Selvitetään seuraavaksi milloin muuttujien $x_1, \dots, x_k$ korrelaatiomatriisin $\mathbf{R}_{\mathbf{xx}}$ determinantti on nolla. Jos muuttujien varianssit ovat nollasta poikkeavia, on selvää, että $|\mathbf{R}_{\mathbf{xx}}| \neq 0 \iff |\mathbf{T}_{\mathbf{xx}}| \neq 0 \iff |\mathbf{S}_{\mathbf{xx}}| \neq 0$. Merkitään

$$\mathbf{X} = (\mathbf{1} : \mathbf{X}_0) = (\mathbf{1} : \mathbf{x}_1 : \ldots : \mathbf{x}_k). \tag{8.124}$$

Tällöin

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{1}' \\ \mathbf{X}_0' \end{pmatrix} (\mathbf{1} : \mathbf{X}_0) = \begin{pmatrix} n & \mathbf{1}'\mathbf{X}_0 \\ \mathbf{X}_0'\mathbf{1} & \mathbf{X}_0'\mathbf{X}_0 \end{pmatrix}, \tag{8.125}$$

joten

$$\begin{aligned} |\mathbf{X}'\mathbf{X}| &= n|\mathbf{X}_0'\mathbf{X}_0 - \mathbf{X}_0'\mathbf{1}\tfrac{1}{n}\mathbf{1}'\mathbf{X}_0| \\ &= |\mathbf{X}_0'\mathbf{X}_0|[n - \mathbf{1}'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-}\mathbf{X}_0'\mathbf{1}], \end{aligned} \tag{8.126}$$

ts.

$$|\mathbf{X}'\mathbf{X}| = n|\mathbf{X}_0'(\mathbf{I} - \mathbf{J})\mathbf{X}_0| = n|\mathbf{T}_{\mathbf{xx}}| = |\mathbf{X}_0'\mathbf{X}_0| \left(n - \mathbf{1}'\mathbf{P}_{\mathbf{X}_0}\mathbf{1}\right). \tag{8.127}$$

Näin voimme päätellä seuraavan tuloksen; ks. myös (9.163)–(9.166) (s. 329):

Oletetaan että muuttujien $x_1, \dots, x_k$ varianssit Seuraavat väitteet ovat yhtäpitäviä:

(i) $\det(\mathbf{R}_{\mathbf{xx}}) \neq 0$, *(olettaen että muuttujien varianssit eivät ole nollia)*

(ii) $\det(\mathbf{S}_{\mathbf{xx}}) \neq 0$,

(iii) $\mathrm{r}(\mathbf{X}) = \mathrm{r}(\mathbf{1} : \mathbf{x}_1 : \ldots : \mathbf{x}_k) = k + 1$,

(iv) $\mathrm{r}(\mathbf{X}_0) = \mathrm{r}(\mathbf{x}_1 : \ldots : \mathbf{x}_k) = k$ *ja* $\mathbf{1} \notin \mathscr{C}(\mathbf{X}_0)$,

(v) $\det(\mathbf{X}'\mathbf{X}) \neq 0$.

Olkoon $p \geq 3$. On hyödyllistä huomata, että jos tällöin yhdenkin muuttujaparin korrelaatio on itseisarvoltaan 1, niin silloin korrelaatiomatriisi $\mathbf{R}$ on singulaarinen. Nimittäin $\mathrm{cor_d}(\mathbf{x}_i, \mathbf{x}_j) = 1$ merkitsee (8.41):n mukaan sitä että vektorit $\{\mathbf{1}, \mathbf{x}_i, \mathbf{x}_j\}$ ovat sidotut ja täten $\mathrm{r}(\mathbf{X}) = \mathrm{r}(\mathbf{1} : \mathbf{x}_1 : \ldots : \mathbf{x}_k) < k+1$. Toisaalta korrelaatiomatriisin singulaarisuus *ei* implikoi, että jokin korrelaatio olisi välttämättä (itseisarvoltaan) 1. Yhteenvetona:

$$\text{jokin } \mathrm{cor_d}^2(\mathbf{x}_i, \mathbf{x}_j) = 1 \implies \det(\mathbf{R}) = 0, \tag{8.128a}$$

$$\det(\mathbf{R}) = 0 \not\Rightarrow \mathrm{cor_d}^2(\mathbf{x}_i, \mathbf{x}_j) = 1 \text{ jollakin muuttujaparilla.} \tag{8.128b}$$

Esimerkki 8.1. Olkoot $\mathbf{x}$, $\mathbf{y}$ ja $\mathbf{z}$ keskistettyjä ja 1:n pituisia muuttujavektoreita, joiden korrelaatiomatriisi on

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & -1/\sqrt{2} \\ 0 & 1 & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} & 1 \end{pmatrix}. \tag{8.129}$$

Tässä tapauksessa $\mathbf{R}$:n determinantti on 0, mutta silti yksikään lävistäjän ulkopuolinen elementti ei ole 1. Osoitamme seuraavaksi, että vektorit $\mathbf{x}$, $\mathbf{y}$ ja $\mathbf{z}$ toteuttavat yhtälön

$$\mathbf{x} - \mathbf{y} + \sqrt{2}\mathbf{z} = \mathbf{0}. \tag{8.130}$$

Koska $\mathbf{x}$, $\mathbf{y}$ ja $\mathbf{z}$ ovat keskistettyjä ja 1:n pituisia, on $\mathbf{R} = \mathbf{U}'\mathbf{U}$, kun $\mathbf{U} = (\mathbf{x} : \mathbf{y} : \mathbf{z})$. Merkitään $\mathbf{t} = (1, 1, \sqrt{2})'$, jolloin

$$\mathbf{Rt} = \mathbf{UU}'\mathbf{t} = \begin{pmatrix} 1 & 0 & -1/\sqrt{2} \\ 0 & 1 & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ \sqrt{2} \end{pmatrix} = \mathbf{0}. \tag{8.131}$$

Tällöin $\mathbf{U}'\mathbf{Ut} = \mathbf{0} \Rightarrow \mathbf{t}'\mathbf{U}'\mathbf{Ut} = 0 \Rightarrow (\mathbf{Ut})'\mathbf{Ut} = 0 \Rightarrow \mathbf{Ut} = \mathbf{0}$ ts.

$$(\mathbf{x} : \mathbf{y} : \mathbf{z}) \begin{pmatrix} 1 \\ -1 \\ \sqrt{2} \end{pmatrix} = \mathbf{0} \text{ eli } \mathbf{x} - \mathbf{y} + \sqrt{2}\mathbf{z} = \mathbf{0}. \tag{8.132}$$

8.4.6 Yleistetty varianssi

Kovarianssimatriisin $\mathbf{S}$ determinanttia sanotaan $\mathbf{U}$:n muuttujien *yleistetyksi varianssiksi*. Jos kyseessä on kaksi muuttujaa $\mathbf{x}$ ja $\mathbf{y}$ (joiden varianssit $\neq 0$), niin

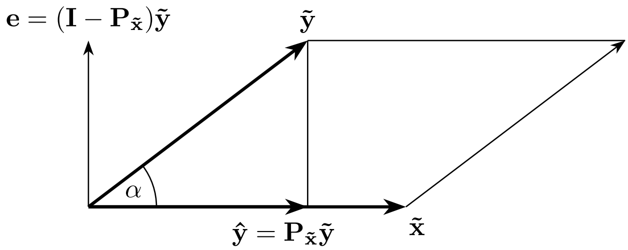
$$\begin{aligned} \det(\mathbf{S}) &= \det \begin{pmatrix} s_x^2 & s_{xy} \\ s_{yx} & s_y^2 \end{pmatrix} = s_x^2 s_y^2 - s_{xy}^2 \\ &= s_x^2 s_y^2 \Big(1 - \frac{s_{xy}^2}{s_x^2 s_y^2}\Big) = s_x^2 s_y^2 (1 - r_{xy}^2). \end{aligned} \tag{8.133}$$

Täten

$$0 \leq \det(\mathbf{S}) \leq s_x^2 s_y^2. \tag{8.134}$$

Determinantin $\det(\mathbf{S})$ suurin arvo saavutetaan silloin kun $\mathbf{x}$ ja $\mathbf{y}$ ovat korreloimattomia ja pienin arvo kun $r_{xy}^2 = 1$. Tarkastellaan nyt $\mathbf{x}$:n ja $\mathbf{y}$:n keskistettyjä versioita $\tilde{\mathbf{x}}$ ja $\tilde{\mathbf{y}}$:

$$\widetilde{\mathbf{U}} = \mathbf{CU} = (\mathbf{Cx} : \mathbf{Cy}) = (\tilde{\mathbf{x}} : \tilde{\mathbf{y}}).$$



Kuvio 8.8. Keskistettyjen muuttujavektoreiden virittämä puolisuunnikas.

Keskistettyjen muuttujavektoreiden $\tilde{\mathbf{x}}$ ja $\tilde{\mathbf{y}}$ muodostaman puolisuunnikkaan pinta-alan neliö on

$$(\text{ala})^2 = \|\tilde{\mathbf{x}}\|^2 \|\mathbf{e}\|^2. \tag{8.135}$$

Koska $\|\tilde{\mathbf{x}}\|^2 = t_{xx}$ ja

$$\begin{aligned} \|\mathbf{e}\|^2 &= \tilde{\mathbf{y}}' \left(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{x}}}\right) \tilde{\mathbf{y}} = \tilde{\mathbf{y}}'[\mathbf{I} - \tilde{\mathbf{x}}(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}]\tilde{\mathbf{y}} \\ &= t_{yy} - \frac{t_{xy}^2}{t_{xx}} = t_{yy}(1 - r_{xy}^2), \end{aligned} \tag{8.136}$$

on

$$(\text{ala})^2 = t_{xx}t_{yy}(1 - r_{xy}^2) = (n-1)^2 s_x^2 s_y^2 (1 - r_{xy}^2) = (n-1)^2 \det(\mathbf{S}). \tag{8.137}$$

Täten yleistetty varianssi on verrannollinen puolisuunnikkaan pinta-alan neliöön:

$$\det(\mathbf{S}) = \frac{1}{(n-1)^2}(\text{ala})^2. \tag{8.138}$$

Kolmiulotteisessa tapauksessa yleistetty varianssi on verrannollinen keskistettyjen muuttujavektoreiden $\tilde{\mathbf{x}}$, $\tilde{\mathbf{y}}$ ja $\tilde{\mathbf{z}}$ virittämän ”laatikon” tilavuuteen: laatikon pohjasärmät muodostuvat vektoreista $\tilde{\mathbf{x}}$ ja $\tilde{\mathbf{y}}$ ja sen kolmannen särmän muodostaa $\tilde{\mathbf{z}}$. Yleisesti on voimassa seuraava tulos [ks. esim. Johnson & Wichern (2007, s. 100), Anderson (2003, s. 259):

$$\det(\mathbf{S}_{p\times p}) = \frac{1}{(n-1)^p}(\text{volume})^2. \tag{8.139}$$

Kaavassa (8.139) ”volume” tarkoittaa keskistettyjen muuttujavektoreiden virittämän useampiulotteisen ”laatikon” tilavuutta.

Voimme tulkita yleistetyn varianssin myös havaintoavaruudessa (edellähän tarkasteltiin muuttuja-avaruutta). Nimittäin on osoitettavissa, että jos $\mathcal{A}$ on ellipsoidi

$$\mathcal{A} = \{\, \mathbf{z} \in \mathbb{R}^p : (\mathbf{z} - \bar{\mathbf{u}})'\mathbf{S}^{-1}(\mathbf{z} - \bar{\mathbf{u}}) = d^2 \,\}, \tag{8.140}$$

niin

$$\mathrm{Vol}(\mathcal{A}) = k_p|\mathbf{S}|^{\frac{1}{2}}d^p, \tag{8.141}$$

missä k_p on p:stä riippuva vakio ja ”Vol” tarkoittaa ellipsin (ellipsoidin) pinta-alaa (tilavuutta). Toisin sanoen:

$$[\mathrm{Vol}(\mathcal{A})]^2 = \text{vakio} \cdot \text{yleistetty varianssi}. \tag{8.142}$$

Yleistetty varianssi on siis tietynlainen muuttujajoukon kokonaisvaihtelun mitta – se on yksittäinen luku joka kuvaa useiden muuttujien vaihtelua. Sillä on tietysti esim. sellainen heikkous, että sen arvo on nolla jos jonkin muuttujaparin korrelaatio on ± 1 (olivatpa muut korrelaatiot ja varianssit mitä hyvänsä).

Toinen varsin luonteva aineiston kokonaisvaihtelua mittaava tunnusluku on varianssien summa eli trace($\mathbf{S}$), josta käytetään usein nimitystä *kokonaisvarianssi*:

$$\text{kokonaisvarianssi} = \mathrm{tr}(\mathbf{S}). \tag{8.143}$$

Tämä puolestaan ei ota mitenkään huomioon muuttujien välisiä korrelaatioita. Mustonen (1997) on ehdottanut vastaavaksi mittaluvuksi lauseketta

$$\mathrm{Mvar}(\mathbf{S}) = \max \sum_{i=1}^{p} s_{ii\cdot 12\ldots i-1}, \tag{8.144}$$

missä $s_{ii\cdot 12\ldots i-1}$ on i. muuttujan jäännösvarianssi kun aikaisemmat muuttujat on eliminoitu ja maksimia etsitään kaikkien permutaatioiden yli. Kahden muuttujan tapauksessa saamme (kun $s_1 \geq s_2$)

$$\mathrm{Mvar}(\mathbf{S}) = s_1^2 + (1 - r_{12}^2)s_2^2. \tag{8.145}$$

Mainittakoon vielä, että jos $\mathbf{S}_{p\times p}$:llä on ominaisarvohajotelma

$$\mathbf{S} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}', \tag{OAH}$$

missä $\mathbf{Q}$ on ortogonaalinen matriisi ja $\boldsymbol{\Lambda}$ on $\mathbf{S}$:n ominaisarvojen muodostama lävistäjämatriisi:

$$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p), \qquad \lambda_1 \geq \ldots \geq \lambda_p \geq 0, \qquad \mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}_p, \tag{8.146}$$

niin

$$\mathrm{tr}(\mathbf{S}) = \mathrm{tr}(\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}') = \mathrm{tr}(\mathbf{Q}'\mathbf{Q}\boldsymbol{\Lambda}) = \mathrm{tr}(\boldsymbol{\Lambda}) = \lambda_1 + \lambda_2 + \cdots + \lambda_p, \tag{8.147a}$$

$$|\mathbf{S}| = |\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}'| = |\mathbf{Q}||\boldsymbol{\Lambda}||\mathbf{Q}'| = |\boldsymbol{\Lambda}| = \lambda_1\lambda_2\cdots\lambda_p. \tag{8.147b}$$

8.5 Lineaarikombinaation maksimivarianssi

Olkoon tehtävänä määrittää se muuttujavektoreiden $\mathbf{u}_1$, $\mathbf{u}_2$, ..., $\mathbf{u}_p$ lineaarikombinaatio

$$\mathbf{y} = \mathbf{Ua} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_p\mathbf{u}_p, \tag{8.148}$$

jolla on mahdollisimman suuri varianssi. Meidän on siis maksimoitava lauseke

$$\mathrm{var_d}(\mathbf{y}) = \mathbf{a}'\mathbf{Sa}, \tag{8.149}$$

missä $\mathbf{S} = \mathrm{cov_d}(\mathbf{U})$. Tehtävä ei tietenkään ole mielekäs, ellei vektorille $\mathbf{a}$ aseteta jotakin rajoitusta; muutenhan $\mathrm{var_d}(\mathbf{y})$ saadaan miten suureksi hyvänsä. Luonteva rajoitus on ehto $\mathbf{a}'\mathbf{a} = 1$. Ratkaisemme tehtävän käyttämällä $\mathbf{S}$:n on ominaisarvohajotelmaa $\mathbf{S} = \mathbf{Q\Lambda Q}'$, missä $\mathbf{Q}$ on $\mathbf{S}$:n ominaisvektoreiden muodostama ortogonaalinen matriisi ja $\mathbf{\Lambda}$ on $\mathbf{S}$:n ominaisarvojen muodostama lävistäjämatriisi. Jos merkitään $\mathbf{b} = \mathbf{Q}'\mathbf{a}$, niin $\mathbf{b}'\mathbf{b} = \mathbf{a}'\mathbf{QQ}'\mathbf{a} = \mathbf{a}'\mathbf{a} = 1$, ja

$$\begin{aligned}\mathrm{var_d}(\mathbf{y}) &= \mathbf{a}'\mathbf{Sa} = \mathbf{a}'\mathbf{Q\Lambda Q}'\mathbf{a} = \mathbf{b}'\mathbf{\Lambda b}\\ &= b_1^2\lambda_1 + b_2^2\lambda_2 + \cdots + b_p^2\lambda_p\\ &\le \lambda_1(b_1^2 + b_2^2 + \cdots + b_p^2) = \lambda_1\mathbf{b}'\mathbf{b} = \lambda_1. \end{aligned}\tag{8.150}$$

Näin olemme osoittaneet, että kysytty varianssin maksimiarvo on kovarianssimatriisin suurin ominaisarvo:

$$\begin{aligned}\mathrm{ch}_1(\mathbf{S}) = \lambda_1 &= \max_{\mathbf{a}\neq\mathbf{0}} \frac{\mathbf{a}'\mathbf{Sa}}{\mathbf{a}'\mathbf{a}} = \max_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\mathbf{Sa}\\ &= \max_{\mathbf{a}'\mathbf{a}=1} \mathrm{var_d}(\mathbf{Ua})\\ &= \max_{\mathbf{a}'\mathbf{a}=1} \mathrm{var_d}(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_p\mathbf{u}_p). \end{aligned}\tag{8.151}$$

[Ks. myös (5.46), s. 168.] Kyseinen maksimiarvo saavutetaan, kun valitaan $\mathbf{a}$:ksi

$$\mathbf{a}_* = \mathbf{q}_1 = \text{ominaisarvoa } \lambda_1 \text{ vastaava ominaisvektori}, \tag{8.152}$$

sillä (yhtälön $\mathbf{Sq}_1 = \lambda_1\mathbf{q}_1$ perusteella)

$$\mathbf{q}_1'\mathbf{Sq}_1 = \mathbf{q}_1'(\lambda_1\mathbf{q}_1) = \lambda_1\mathbf{q}_1'\mathbf{q}_1 = \lambda_1. \tag{8.153}$$

Täten olemme todistaneet seuraavan tuloksen:

Muuttujavektoreiden $\mathbf{u}_1$, $\mathbf{u}_2$, ..., $\mathbf{u}_p$ lineaarikombinaation

$$\mathbf{y} = \mathbf{Ua} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_p\mathbf{u}_p, \tag{8.154}$$

varianssin maksimi ehdolla $\mathbf{a}'\mathbf{a} = 1$ on $\mathbf{U}$:n kovarianssimatriisin $\mathbf{S}$ suurin ominaisarvo $\mathrm{ch}_1(\mathbf{S}) = \lambda_1$. Tämä arvo saavutetaan kun $\mathbf{a} = \mathbf{q}_1 = \lambda_1$:tä vastaava ominaisvektori.

Maksimivarianssin antavaa lineaarikombinaatiota sanotaan *ensimmäiseksi pääkomponentiksi*. Jos merkitään

$$\mathbf{q}_1 = \begin{pmatrix} q_{11} \\ q_{21} \\ \vdots \\ q_{p1} \end{pmatrix}, \tag{8.155}$$

niin 1. pääkomponentti on siis

$$\mathbf{y}_1 = \mathbf{U}\mathbf{q}_1 = q_{11}\mathbf{u}_1 + q_{21}\mathbf{u}_2 + \cdots + q_{p1}\mathbf{u}_p \in \mathbb{R}^n. \tag{8.156}$$

Jos merkitsemme "symbolisesti" $\mathbf{u}_i$-muuttujien muodostamaa vektoria siten että $\mathbf{u} = (u_1, \dots, u_p)'$, niin 1. pääkomponenttia voidaan merkitä seuraavasti:

$$y_1 = \mathbf{q}_1'\mathbf{u} = q_{11}u_1 + q_{21}u_2 + \cdots + q_{p1}u_p. \tag{8.157}$$

Erityisesti jos pidämme vektoria $\mathbf{u}$ satunnaisvektorina, on 1. pääkomponentti satunnaismuuttuja $y_1 = \mathbf{q}_1'\mathbf{u}$ ja sen varianssi on

$$\operatorname{var}(y_1) = \operatorname{var}(\mathbf{q}_1'\mathbf{u}) = \mathbf{q}_1' \operatorname{cov}(\mathbf{u})\mathbf{q}_1 = \mathbf{q}_1'\mathbf{\Sigma}\mathbf{q}_1 = \operatorname{ch}_1(\mathbf{\Sigma}). \tag{8.158}$$

Toinen pääkomponentti määritetään siten että se on se muuttujavektori $\mathbf{Ub}$, jolla on suurin varianssi ehdoilla

$$\mathbf{b}'\mathbf{b} = 1\,, \quad \operatorname{cor}_{\mathrm{d}}(\mathbf{Ub}, \mathbf{y}_1) = 0\,. \tag{8.159}$$

Tällaiseksi kerroinvektoriksi osoittautuu $\mathbf{S}$:n toiseksi suurinta ominaisarvoa $\operatorname{ch}_2(\mathbf{S}) = \lambda_2$ vastaava ominaisvektori $\mathbf{q}_2$. Näemme välittömästi, että $\mathbf{U}\mathbf{q}_1$ ja $\mathbf{U}\mathbf{q}_2$ ovat korreloimattomia, sillä $\mathbf{Q}$:n sarakkeiden ortogonaalisuuden perusteella

$$\operatorname{cov}_{\mathrm{d}}(\mathbf{U}\mathbf{q}_1, \mathbf{U}\mathbf{q}_2) = \mathbf{q}_1'\mathbf{S}\mathbf{q}_2 = \mathbf{q}_1'(\lambda_2\mathbf{q}_2) = \lambda_1\mathbf{q}_1'\mathbf{q}_2 = 0. \tag{8.160}$$

Vastaavalla tavalla määritellään i. pääkomponentti $\operatorname{ch}_i(\mathbf{S})$:ää vastaavan ominaisvektorin $\mathbf{q}_i$ avulla.

Viimeisellä eli p. pääkomponentilla on seuraava minimiominaisuus

$$\begin{aligned} \operatorname{ch}_p(\mathbf{S}) = \lambda_p = \min_{\mathbf{a}\neq\mathbf{0}} \frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{\mathbf{a}'\mathbf{a}} &= \min_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\mathbf{S}\mathbf{a} \\ &= \min_{\mathbf{a}'\mathbf{a}=1} \operatorname{var}_{\mathrm{d}}(\mathbf{Ua}) \\ &= \min_{\mathbf{a}'\mathbf{a}=1} \operatorname{var}_{\mathrm{d}}(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_p\mathbf{u}_p). \end{aligned} \tag{8.161}$$

Ominaisarvoista voimme vielä palauttaa mileen seuraavan tuloksen: Jos $\mathbf{A}_{p\times p}$ on symmetrinen, niin

$$\operatorname{ch}_p(\mathbf{A}) = \operatorname{ch}_{\min}(\mathbf{A}) \leq \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \leq \operatorname{ch}_{\max}(\mathbf{A}) = \operatorname{ch}_1(\mathbf{A}) \ \forall \mathbf{x} \in \mathbb{R}^p. \tag{8.162}$$

Esimerkki 8.2. Tarkastellaan muuttujavektoreita $\mathbf{u}_1$ ja $\mathbf{u}_2$, joiden kovarianssimatriisi (ja samalla korrelaatiomatriisi) on

$$\mathrm{cov_d}(\mathbf{u}_1 : \mathbf{u}_2) = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} = \mathbf{S}. \tag{8.163}$$

Matriisin $\mathbf{S}$ ominaisarvot ovat

$$\mathrm{ch}_1(\mathbf{S}) = \lambda_1 = 1 + r, \ \mathrm{ch}_2(\mathbf{S}) = \lambda_2 = 1 - r, \tag{8.164}$$

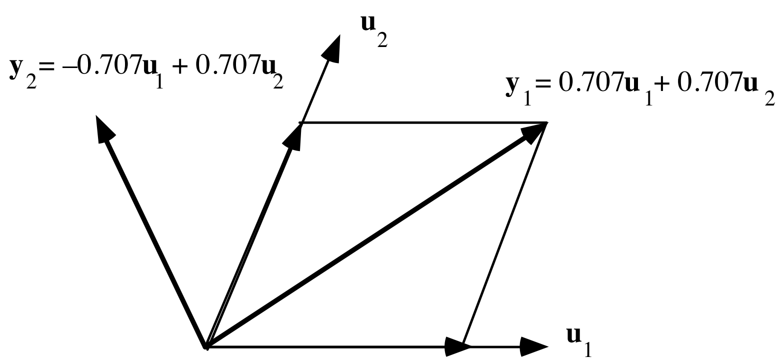
ja vastaavat ortonormaalit ominaisvektorit

$$\mathbf{q}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.707 \\ 0.707 \end{pmatrix}, \ \mathbf{q}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.707 \\ 0.707 \end{pmatrix}. \tag{8.165}$$

Pääkomponentit ovat (jos $r > 0$)

$$\mathbf{y}_1 = 0.707\mathbf{u}_1 + 0.707\mathbf{u}_2, \ \mathrm{var_d}(\mathbf{y}_1) = 1 + r, \tag{8.166a}$$

$$\mathbf{y}_2 = -0.707\mathbf{u}_1 + 0.707\mathbf{u}_2, \ \mathrm{var_d}(\mathbf{y}_2) = 1 - r. \tag{8.166b}$$



Kuvio 8.9. Pääkomponentit muuttuja-avaruudessa: $\mathbf{u}_1$ ja $\mathbf{u}_2$ ovat keskistettyjä ja yhtä pitkiä (ja täten niiden varianssitkin ovat yhtä suuret) muuttujavektoreita, joiden korrelaatio on 0.5.

Voimme olettaa, että $\mathbf{u}_1$ ja $\mathbf{u}_2$ ovat keskistettyjä – keskistämisellä ei tietenkään ole vaikutusta kovarianssimatriisiin $\mathbf{S}$. Keskistetyn muuttujavektorin varianssi on tietenkin suoraan verrannollinen muuttujavektorin pituuden neliöön. Vektoreiden $\mathbf{u}_1$ ja $\mathbf{u}_2$ lineaarikombinaatiot ovat myös keskistettyjä ja täten 1. pääkomponentin määrittäminen (varianssin maksimointi) tarkoittaa sellaisen lineaarikombinaation $a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ etsimistä, jolla on *maksimipituus* (ehdolla $a_1^2 + a_2^2 = 1$). Kuvio 8.9 havainnollistaa tilannetta. Siinä kovarianssimatriisi on (sama kuin korrelaatiomatriisi)

$$\mathrm{cov_d}(\mathbf{u}_1 : \mathbf{u}_2) = \mathbf{S} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \tag{8.167}$$

ja kun $r > 0$,

$$\mathrm{ch}_1(\mathbf{S}) = \lambda_1 = 1.5 = \max_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\mathbf{S}\mathbf{a}, \tag{8.168a}$$

$$\mathrm{ch}_2(\mathbf{S}) = \lambda_2 = 0.5 = \min_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\mathbf{S}\mathbf{a}. \tag{8.168b}$$

Jos kovarianssimatriisi onkin

$$\mathrm{cov}_{\mathrm{d}}(\mathbf{v}_1 : \mathbf{v}_2) = \mathbf{V} = \begin{pmatrix} 1 & 0.5 \cdot 3 \\ 0.5 \cdot 3 & 3 \cdot 3 \end{pmatrix} = \begin{pmatrix} 1 & 1.5 \\ 1.5 & 9 \end{pmatrix}, \tag{8.169}$$

niin $\mathbf{v}_1$:n ja $\mathbf{v}_2$:n välinen korrelaatio on edelleen sama kuin $\mathbf{u}_1$:n ja $\mathbf{u}_2$:n korrelaatio. Kuitenkin pääkomponentit ovat aivan erilaiset:

$$\mathbf{z}_1 = 0.178\mathbf{v}_1 + 0.984\mathbf{v}_2, \;\; \mathrm{var}_{\mathrm{d}}(\mathbf{z}_1) = 9.272, \tag{8.170a}$$

$$\mathbf{z}_2 = 0.984\mathbf{v}_1 - 0.178\mathbf{v}_2, \;\; \mathrm{var}_{\mathrm{d}}(\mathbf{z}_2) = 0.728. \tag{8.170b}$$

Kovarianssimatriisin $\mathbf{V}$ tapauksessa 1. pääkomponenttia laskettaessa painotetaan muuttujaa $\mathbf{v}_2$ huomattavasti enemmän kuin $\mathbf{v}_1$:tä: syynä on $\mathbf{v}_2$:n suurempi varianssi. Kuvion 8.9 tarkastelussa tämä merkitsisi, että pitkää muuttujavektoria kannattaa painottaa enemmän, jotta saadaan 1. pääkomponentin pituus suureksi.

Jos kovarianssimatriisi on $\mathbf{I}_2$, niin $\lambda_1 = \lambda_2 = 1$, ja lineaarikombinaation $a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ pituus on aina 1, kun $\mathbf{a}'\mathbf{a} = 1$. tämä tarkoittaa, että ominaisvektorit eivät ole yksikäsitteisiä; mikä hyvänsä ortonormaali vektoripari kelpaa ominaisvektoreiksi tässä tapauksessa. □

8.6 Ortogonaaliset etäisyydet muuttuja-avaruudessa

Voimme lähestyä pääkomponenttianalyysiä myös seuraavasti. Oletetaan, että meillä on p keskistetyn muuttujan havaintomatriisi

$$\widetilde{\mathbf{U}} = (\tilde{\mathbf{u}}_1 : \tilde{\mathbf{u}}_2 : \ldots : \tilde{\mathbf{u}}_p). \tag{8.171}$$

Olkoon $\mathbf{y}$ jokin $\widetilde{\mathbf{U}}$:n sarakkeiden lineaarikombinaatio eli

$$\mathbf{y} = \widetilde{\mathbf{U}}\mathbf{a}, \;\; \mathbf{a}'\mathbf{a} = 1, \;\; \mathbf{a} \in \mathbb{R}^p. \tag{8.172}$$

Tällöin muuttujavektorin $\tilde{\mathbf{u}}_i$ ortogonaaliprojektio suoralle $\mathscr{C}(\mathbf{y})$ on tietenkin $\mathbf{P}_{\mathbf{y}}\tilde{\mathbf{u}}_i$. Vastaavasti vektorin $\tilde{\mathbf{u}}_i$ etäisyys suorasta $\mathscr{C}(\mathbf{y})$ (neliöön korotettuna) on

$$\|\tilde{\mathbf{u}}_i - \mathbf{P}_{\mathbf{y}}\tilde{\mathbf{u}}_i\|^2 = \|(\mathbf{I}_n - \mathbf{P}_{\mathbf{y}})\tilde{\mathbf{u}}_i\|^2 := \|\mathbf{f}_i\|^2 = \mathbf{f}_i'\mathbf{f}_i, \tag{8.173}$$

missä $\mathbf{f}_i$ on siis tietty residuaalivektori. Olkoon $\mathbf{F}$ näiden residuaalivektoreiden muodostama matriisi:

$$\begin{aligned}\mathbf{F} = (\mathbf{f}_1 : \mathbf{f}_2 : \ldots : \mathbf{f}_p) &= ((\mathbf{I}_n - \mathbf{P}_\mathbf{y})\tilde{\mathbf{u}}_1 : (\mathbf{I}_n - \mathbf{P}_\mathbf{y})\tilde{\mathbf{u}}_2 : \ldots : (\mathbf{I}_n - \mathbf{P}_\mathbf{y})\tilde{\mathbf{u}}_p) \\ &= (\mathbf{I}_n - \mathbf{P}_\mathbf{y})(\tilde{\mathbf{u}}_1 : \tilde{\mathbf{u}}_2 : \ldots : \tilde{\mathbf{u}}_p) \\ &= (\mathbf{I}_n - \mathbf{P}_\mathbf{y})\widetilde{\mathbf{U}}. \end{aligned} \tag{8.174}$$

Merkitään vektoreiden $\tilde{\mathbf{u}}_i$ etäisyyksien [suorasta $\mathscr{C}(\mathbf{y})$] neliöiden summaa $f(\mathbf{y})$:llä:

$$f(\mathbf{y}) = f(\widetilde{\mathbf{U}}\mathbf{a}) = \|\mathbf{f}_1\|^2 + \|\mathbf{f}_2\|^2 + \cdots + \|\mathbf{f}_p\|^2 = \operatorname{tr}(\mathbf{F}'\mathbf{F}) = \|\mathbf{F}\|_F^2. \tag{8.175}$$

Matriisi $\widetilde{\mathbf{U}}$ on siis annettu. Asetamme nyt itsellemme seuraavan ongelman:

> Miten on $\mathbf{y} = \widetilde{\mathbf{U}}\mathbf{a}$ määritettävä, jotta lauseke $f(\mathbf{y})$ olisi pienimmillään? (8.176)

Tämähän on luonteva kysymys: jos $\mathbf{y}$ minimoi $f(\mathbf{y})$:n, niin tällöin suora $\mathscr{C}(\mathbf{y})$ kulkee tietyssä mielessä optimaalisesti muuttujavektorikimpun ”ruuhkan” kautta (jos sitä ruuhkaa nyt sitten onkaan).

Vektorin $\mathbf{y}$ pituus on epäoleellinen tekijä, joten voimme olettaa että

$$\mathbf{y}'\mathbf{y} = 1, \text{ ja siten } \mathbf{P}_\mathbf{y} = \mathbf{y}\mathbf{y}'. \tag{8.177}$$

Havaitsemme vaivatta että $f(\mathbf{y})$ voidaan ilmaista seuravasti:

$$\begin{aligned} f(\mathbf{y}) = \operatorname{tr}(\mathbf{F}'\mathbf{F}) &= \operatorname{tr}[((\mathbf{I}_n - \mathbf{P}_\mathbf{y})\widetilde{\mathbf{U}})'(\mathbf{I}_n - \mathbf{P}_\mathbf{y})\widetilde{\mathbf{U}}] \\ &= \operatorname{tr}[\widetilde{\mathbf{U}}'(\mathbf{I}_n - \mathbf{P}_\mathbf{y})\widetilde{\mathbf{U}}] = \operatorname{tr}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}) - \operatorname{tr}(\widetilde{\mathbf{U}}'\mathbf{P}_\mathbf{y}\widetilde{\mathbf{U}}) \\ &:= c - \operatorname{tr}(\widetilde{\mathbf{U}}'\mathbf{y} \cdot \mathbf{y}'\widetilde{\mathbf{U}}) = c - \operatorname{tr}(\mathbf{y}'\widetilde{\mathbf{U}} \cdot \widetilde{\mathbf{U}}'\mathbf{y}) \\ &= c - \mathbf{y}'\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}'\mathbf{y}. \end{aligned} \tag{8.178}$$

Täten $f(\mathbf{y})$:n minimointi yhtäpitävää lausekkeen

$$g(\mathbf{y}) = \mathbf{y}'\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}'\mathbf{y} = \mathbf{a}'\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}\mathbf{a} = \mathbf{a}'(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}})^2\mathbf{a} = \mathbf{a}'\mathbf{T}^2\mathbf{a} := h(\mathbf{a}) \tag{8.179}$$

maksimoinnin kanssa; $\mathbf{T}$ on $\mathbf{U}$:n tulosummamatriiisi. Näin olemme päätyneet aivan identtiseen tehtävään kuin lineaarikombinaation $\mathbf{Ua}$ varianssin maksimoinnissa – erona on se, että (8.179):ssä esiintyy $\mathbf{T}$:n neliö (siinä voisi yhtä hyvin tietysti olla $\mathbf{T}$:n paikalla $\mathbf{S}$), kun taas $\operatorname{var}_\mathrm{d}(\mathbf{Ua}) = \frac{1}{n-1}\mathbf{a}'\mathbf{T}\mathbf{a}$. Koska

$$\text{matriiseilla } \mathbf{T}^2,\ \mathbf{T},\ \mathbf{S}^2 \text{ ja } \mathbf{S} \text{ on samat ominaisvektorit,} \tag{8.180}$$

(tästä on hyvin helppo vakuuttautua) saavutetaan (8.179):n maksimi silloin kun kun $\mathbf{a} = \mathbf{q}_1 = \mathbf{S}$:n suurinta ominaisarvoa λ_1 vastaava ominaisvektori. Täten vastaus kysymykseen (8.176) on

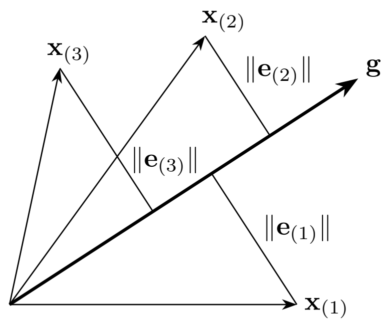
$$f(\mathbf{y}) \text{ on pienimmillään kun } \mathbf{y} = \widetilde{\mathbf{U}}\mathbf{q}_1. \tag{8.181}$$

Mainittakoon että

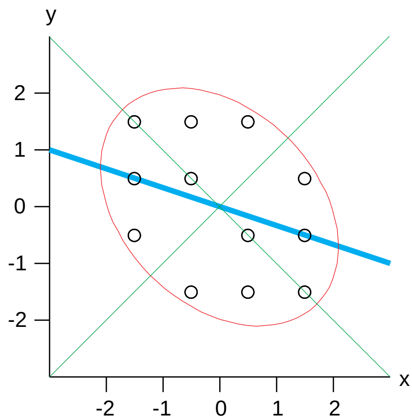
$$\operatorname{ch}_i(\mathbf{T}^2) = [\operatorname{ch}_i(\mathbf{T})]^2,\ \operatorname{ch}_i(\mathbf{S}) = \frac{1}{n-1}\operatorname{ch}_i(\mathbf{T}),\ i = 1, \ldots, p. \tag{8.182}$$

8.7 Ortogonaaliset etäisyydet havaintoavaruudessa

Edellä on tarkasteltu pääkomponentteja lähinnä muuttuja-avaruuden näkökulmasta. Mitä tapahtuu havaintoavaruudessa?



Kuvio 8.10. Suoran $\mathscr{C}(\mathbf{g})$ määrittäminen siten että $\mathbf{x}_{(i)}$:den etäisyyksien neliöiden summa $\|\mathbf{e}_{(1)}\|^2 + \cdots + \|\mathbf{e}_{(n)}\|^2$ minimoituu. (Huom: Kuviossa $\mathbf{x}_{(i)}$ on i. havaintovektori; tekstissä $\mathbf{u}_{(i)}$.)

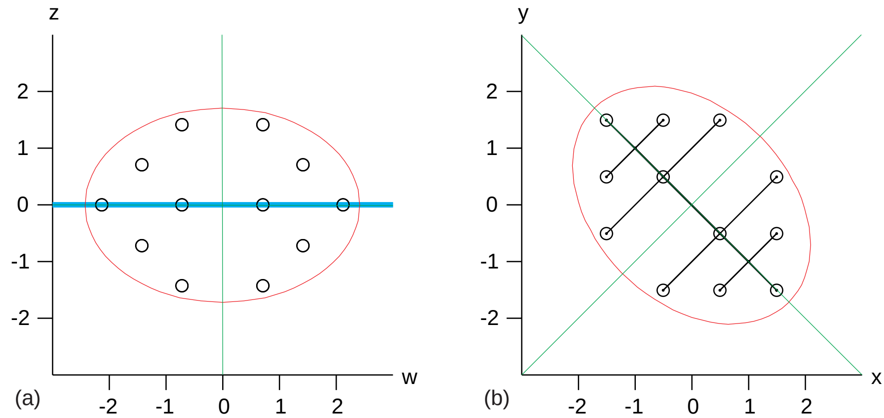


Kuvio 8.11. Pistepari 12:sta havainnosta. Ks. myös kuvio 6.2, s. 190. Merkitään keskistettyä havaintomatriisia $\widetilde{\mathbf{U}}$:lla.

Tarkastellaan p muuttujan keskistettyä havaintomatriisia

$$\widetilde{\mathbf{U}} = (\tilde{\mathbf{u}}_1 : \tilde{\mathbf{u}}_2 : \ldots : \tilde{\mathbf{u}}_p) = \begin{pmatrix} \tilde{\mathbf{u}}'_{(1)} \\ \tilde{\mathbf{u}}'_{(2)} \\ \vdots \\ \tilde{\mathbf{u}}'_{(n)} \end{pmatrix}, \quad \widetilde{\mathbf{U}}' = (\tilde{\mathbf{u}}_{(1)} : \tilde{\mathbf{u}}_{(2)} : \ldots : \tilde{\mathbf{u}}_{(n)}). \tag{8.183}$$

Olkoon $\mathbf{g}$ jokin $\mathbb{R}^p$:n vektori, jonka pituus on 1; $\mathbf{g}'\mathbf{g} = 1$. Tällöin havaintovektorin $\tilde{\mathbf{u}}_{(i)}$ ortogonaaliprojektio suoralle $\mathscr{C}(\mathbf{g})$ on $\mathbf{P}_{\mathbf{g}}\tilde{\mathbf{u}}_{(i)} := \mathbf{p}_{(n)}$, ja kaikkien



Kuvio 8.12. (Jatkoa kuvioon 8.11.) (a) Keskistetty $\widetilde{\mathbf{U}}$ on kerrottu matriisilla $\mathbf{T}$: $\widetilde{\mathbf{U}}\mathbf{T} = (\mathbf{w} : \mathbf{z})$. Mikä on $\mathbf{T} = (\mathbf{t}_1 : \mathbf{t}_2)$? (b) Havaintojen ortogonaaliset etäisyydet suorasta $\mathscr{C}(\mathbf{t}_1)$ merkitty näkyviin.

näiden projektioiden muodostama matriisi on

$$\begin{aligned}(\mathbf{p}_{(1)} : \mathbf{p}_{(2)} : \dots : \mathbf{p}_{(n)}) &= (\mathbf{P_g}\tilde{\mathbf{u}}_{(1)} : \mathbf{P_g}\tilde{\mathbf{u}}_{(2)} : \dots : \mathbf{P_g}\tilde{\mathbf{u}}_{(n)}) \\ &= \mathbf{P_g}\widetilde{\mathbf{U}}' = \mathbf{gg}'\widetilde{\mathbf{U}}' \in \mathbb{R}^{p\times n}. \end{aligned} \tag{8.184}$$

Havaintovektorin $\tilde{\mathbf{u}}_{(i)}$ etäisyys suorasta $\mathscr{C}(\mathbf{g})$ (neliöön korotettuna) on

$$\|\tilde{\mathbf{u}}_{(i)} - \mathbf{P_g}\tilde{\mathbf{u}}_{(i)}\|^2 = \|(\mathbf{I}_p - \mathbf{P_g})\tilde{\mathbf{u}}_{(i)}\|^2 := \|\mathbf{e}_{(i)}\|^2 = \mathbf{e}'_{(i)}\mathbf{e}_{(i)}. \tag{8.185}$$

Merkitään residuaalien $\mathbf{e}_{(i)}$ muodostamaa matriisia $\mathbf{E}'$:llä:

$$\begin{aligned}\mathbf{E}' &= (\mathbf{e}_{(1)} : \mathbf{e}_{(2)} : \dots : \mathbf{e}_{(n)}) \\ &= ((\mathbf{I}_p - \mathbf{P_g})\tilde{\mathbf{u}}_{(1)} : (\mathbf{I}_p - \mathbf{P_g})\tilde{\mathbf{u}}_{(2)} : \dots : (\mathbf{I}_p - \mathbf{P_g})\tilde{\mathbf{u}}_{(n)}) \\ &= (\mathbf{I}_p - \mathbf{P_g})\widetilde{\mathbf{U}}' = (\mathbf{I}_p - \mathbf{gg}')\widetilde{\mathbf{U}}' \in \mathbb{R}^{p\times n}. \end{aligned} \tag{8.186}$$

Matriisin $\mathbf{E}'$ sarakkeet ovat siis havaintovektoreiden ja niiden projektioiden erotuksia $\mathbf{e}_{(i)}$. Olkoon vektorien $\mathbf{e}_{(i)}$ pituuksien neliösumma

$$k(\mathbf{g}) = \|\mathbf{e}_{(1)}\|^2 + \|\mathbf{e}_{(2)}\|^2 + \dots + \|\mathbf{e}_{(n)}\|^2 = \operatorname{tr}(\mathbf{E}'\mathbf{E}) = \|\mathbf{E}\|_F^2. \tag{8.187}$$

Matriisi $\widetilde{\mathbf{U}}$ on siis annettu ja $\|\mathbf{E}\|_F^2$ on siten $\mathbf{g}$:n funktio. Kuten muuttuja-avaruudessakin, voimme kysyä:

$$\text{Mikä on se } \mathbf{g}\text{:n arvo } \mathbf{g}_*, \text{ joka minimoi lausekkeen } k(\mathbf{g})\ ? \tag{8.188}$$

Jos $\mathbf{g}_*$ minimoi $k(\mathbf{g})$:n, niin suora $\mathscr{C}(\mathbf{g}_*)$ kulkee tietyssä mielessä optimaalisesti havaintovektorien kautta: havaintovektorien (-pisteiden) *ortogonaalisten* etäisyyksien neliösumma minimoituu. Pienimmän neliösumman menetelmässähän minimoidaan y-akselin suuntaisia etäisyyksiä. Funktion $k(\mathbf{g})$ lauseke on

$$\begin{aligned} k(\mathbf{g}) &= \operatorname{tr}(\mathbf{E}\mathbf{E}') = \operatorname{tr}[\widetilde{\mathbf{U}}(\mathbf{I}_p - \mathbf{P}_{\mathbf{g}})\widetilde{\mathbf{U}}'] \\ &= \operatorname{tr}(\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}') - \operatorname{tr}(\widetilde{\mathbf{U}}\mathbf{P}_{\mathbf{g}}\widetilde{\mathbf{U}}') \\ &:= c - \operatorname{tr}(\widetilde{\mathbf{U}}\mathbf{g} \cdot \mathbf{g}'\widetilde{\mathbf{U}}') = c - \mathbf{g}'\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}\mathbf{g} \\ &= c - \mathbf{g}'\mathbf{T}\mathbf{g}. \end{aligned} \tag{8.189}$$

Näin olemme päätyneet aivan samaan tulokseen kuin muuttuja-avaruudessa tehdyissä tarkasteluissa eli $k(\mathbf{g})$ minimoituu kun $\mathbf{g}$ on $\mathbf{g}_* = \mathbf{q}_1$, missä $\mathbf{q}_1$ on tulosummamatriisin $\mathbf{T}$ suurinta ominaisarvoa vastaava ominaisvektori, joka on siis sama kuin kovarianssimatriisin $\mathbf{S}$ vastaava ominaisvektori. Kun kaikki havaintovektorit projisoidaan suoralle $\mathscr{C}(\mathbf{q}_1)$, saadaan (8.184):n mukaan projektioiden muodostamaksi matriisiksi

$$\begin{aligned} (\mathbf{p}_{(1)} : \mathbf{p}_{(2)} : \ldots : \mathbf{p}_{(n)}) &= (\mathbf{P}_{\mathbf{q}_1}\tilde{\mathbf{u}}_{(1)} : \mathbf{P}_{\mathbf{q}_1}\tilde{\mathbf{u}}_{(2)} : \ldots : \mathbf{P}_{\mathbf{q}_1}\tilde{\mathbf{u}}_{(n)}) \\ &= \mathbf{P}_{\mathbf{q}_1}\widetilde{\mathbf{U}}' = \mathbf{q}_1\mathbf{q}_1'\widetilde{\mathbf{U}}' \\ &= \mathbf{q}_1(\mathbf{q}_1'\tilde{\mathbf{u}}_{(1)}, \mathbf{q}_1'\tilde{\mathbf{u}}_{(2)}, \ldots, \mathbf{q}_1'\tilde{\mathbf{u}}_{(n)}) \\ &:= \mathbf{q}_1\mathbf{z}_1' \in \mathbb{R}^{p\times n}, \end{aligned} \tag{8.190}$$

missä

$$\mathbf{z}_1 = \widetilde{\mathbf{U}}\mathbf{q}_1 = \begin{pmatrix} \mathbf{q}_1'\tilde{\mathbf{u}}_{(1)} \\ \mathbf{q}_1'\tilde{\mathbf{u}}_{(2)} \\ \vdots \\ \mathbf{q}_1'\tilde{\mathbf{u}}_{(n)} \end{pmatrix} \in \mathbb{R}^n. \tag{8.191}$$

Vektori $\mathbf{z}_1$ sisältää nyt 1. pääkomponentin (otoksesta lasketut) arvot. Tällöin i. havainnon (esim. henkilön) saama arvo ensimmäisellä pääkomponentilla on $\mathbf{z}_1$:n i. elementti

$$z_{i1} = \mathbf{q}_1'\tilde{\mathbf{u}}_{(i)} = \tilde{\mathbf{u}}_{(i)}'\mathbf{q}_1\,. \tag{8.192}$$

Havaintovektorin $\tilde{\mathbf{u}}_{(i)}$ ortogonaaliprojektio suoralle $\mathscr{C}(\mathbf{q}_1)$ on

$$\mathbf{p}_{(i)} = \mathbf{P}_{\mathbf{q}_1}\tilde{\mathbf{u}}_{(i)} = \mathbf{q}_1\mathbf{q}_1'\tilde{\mathbf{u}}_{(i)} = \mathbf{q}_1 z_{i1}\,, \tag{8.193}$$

ja täten projektion $\mathbf{p}_{(i)}$ pituus on juuri $|z_{i1}|$ eli 1. pääkomponentin itseisarvo i. havainnon tapauksessa.

Edellä olevia tarkasteluja voidaan yleistää siten että $\mathscr{C}(\mathbf{g})$:n sijaan tarkastellaankin k-ulotteista aliavaruutta $\mathscr{C}(\mathbf{G})$, missä $\mathbf{G}$ on $p \times k$-matriisi jonka sarakkeet ovat ortonormaaleja eli $\mathbf{G}'\mathbf{G} = \mathbf{I}_k$, $k \leq p$. Kun nyt projisoidaan keskistetyt havaintovektorit $\mathscr{C}(\mathbf{G})$:lle, saadaan residuaalien muodostamaksi matriisiksi

$$\mathbf{E}' = (\mathbf{I}_p - \mathbf{P}_{\mathbf{G}})\widetilde{\mathbf{U}}', \tag{8.194}$$

ja

$$\|\mathbf{E}\|_F^2 = \|\mathbf{E}'\|_F^2 = \|\widetilde{\mathbf{U}}' - \mathbf{P}_{\mathbf{G}}\widetilde{\mathbf{U}}'\|_F^2\,. \tag{8.195}$$

Tehtävänä on nyt minimoida

$$\begin{aligned}\operatorname{tr}(\mathbf{E}\mathbf{E}') &= \operatorname{tr}(\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}') - \operatorname{tr}(\widetilde{\mathbf{U}}\mathbf{P}_{\mathbf{G}}\widetilde{\mathbf{U}}') \\ &= \operatorname{tr}(\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}') - \operatorname{tr}(\widetilde{\mathbf{U}}\mathbf{G}\mathbf{G}'\widetilde{\mathbf{U}}') \\ &= \operatorname{tr}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}) - \operatorname{tr}(\mathbf{G}'\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}\mathbf{G}) \end{aligned} \tag{8.196}$$

eli on ratkaistava $\mathbf{G}$:n suhteen

$$\max \operatorname{tr}(\mathbf{G}'\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}\mathbf{G}) \quad \text{ehdolla } \mathbf{G}'\mathbf{G} = \mathbf{I}_k\,. \tag{8.197}$$

Tuloksen (6.97) (s. 202) nojalla matriisi $\dot{\mathbf{G}}$, joka toteuttaa (8.197):n, muodostuu matriisin $\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}$ k:sta ensimmäisestä ortonormaalista ominaisvektorista:

$$\dot{\mathbf{G}} = \mathbf{Q}_{(k)} = (\mathbf{q}_1 : \ldots : \mathbf{q}_k)\,, \tag{8.198}$$

jolloin

$$\mathbf{P}_{\dot{\mathbf{G}}} = \mathbf{Q}_{(k)}\mathbf{Q}_{(k)}' = \mathbf{q}_1\mathbf{q}_1' + \cdots + \mathbf{q}_k\mathbf{q}_k'\,, \tag{8.199}$$

ja

$$\begin{aligned}\min_{\mathbf{G}'\mathbf{G}=\mathbf{I}_k} \|\mathbf{E}\|_F^2 &= \|\widetilde{\mathbf{U}} - \widetilde{\mathbf{U}}\dot{\mathbf{G}}\dot{\mathbf{G}}'\|_F^2 \\ &= \|\widetilde{\mathbf{U}} - \widetilde{\mathbf{U}}\mathbf{Q}_{(k)}\mathbf{Q}_{(k)}'\|_F^2 \\ &= \operatorname{ch}_{k+1}(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}) + \cdots + \operatorname{ch}_p(\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}) \\ &= \operatorname{sg}_{k+1}^2(\widetilde{\mathbf{U}}) + \cdots + \operatorname{sg}_p^2(\widetilde{\mathbf{U}}) \\ &:= \delta_{k+1}^2 + \cdots + \delta_p^2\,. \end{aligned} \tag{8.200}$$

Projisoidut havaintovektorit $\mathbf{p}_{(i)}$ ovat matriisin

$$\mathbf{Q}_{(k)}\mathbf{Q}_{(k)}'\widetilde{\mathbf{U}}' = \mathbf{Q}_{(k)}\mathbf{Q}_{(k)}'(\tilde{\mathbf{u}}_{(1)} : \tilde{\mathbf{u}}_{(2)} : \ldots : \tilde{\mathbf{u}}_{(n)}) \tag{8.201}$$

sarakkeita, ts,

$$\begin{aligned}(\mathbf{p}_{(1)} : \ldots : \mathbf{p}_{(n)}) &= (\mathbf{q}_1 : \ldots : \mathbf{q}_k) \begin{pmatrix} \mathbf{q}_1' \\ \vdots \\ \mathbf{q}_k' \end{pmatrix} \widetilde{\mathbf{U}}' \\ &= \mathbf{q}_1\mathbf{q}_1'\widetilde{\mathbf{U}}' + \cdots + \mathbf{q}_k\mathbf{q}_k'\widetilde{\mathbf{U}}' \\ &= \mathbf{q}_1\mathbf{z}_1' + \cdots + \mathbf{q}_k\mathbf{z}_k'\,, \end{aligned} \tag{8.202}$$

missä

$$\mathbf{z}_i = \widetilde{\mathbf{U}}\mathbf{q}_i \text{ sisältää } i. \text{ pääkomponentin (otoksesta lasketut) arvot.} \tag{8.203}$$

Olkoon $k = p$ eli muodostetaan p pääkomponenttia (mikä onkin niiden maksimimäärä). Täten i. havainnon kaikkien pääkomponenttien (p kappaletta) arvojen muodostama vektori on $\underline{\mathbf{u}}_{(i)}$:

$$\underline{\mathbf{u}}_{(i)} = \begin{pmatrix} \mathbf{q}_1'\tilde{\mathbf{u}}_{(i)} \\ \mathbf{q}_2'\tilde{\mathbf{u}}_{(i)} \\ \vdots \\ \mathbf{q}_p'\tilde{\mathbf{u}}_{(i)} \end{pmatrix} = \mathbf{Q}'\tilde{\mathbf{u}}_{(i)} \Rightarrow \underline{\mathbf{U}}' = \mathbf{Q}'\widetilde{\mathbf{U}}' \Rightarrow \underline{\mathbf{U}} = \widetilde{\mathbf{U}}\mathbf{Q}\,, \tag{8.204}$$

missä $\mathbf{Q} = (\mathbf{q}_1 : \ldots : \mathbf{q}_p)$ on ortogonaalinen $p \times p$-matriisi. Toisin sanoen: pääkomponenttien arvot i. havainnolle saadaan tekemällä kyseiselle havainnolle ortogonaalinen rotaatio.

Kommentti 8.1. Lukijan kannattaa tutustua keskistetyn havaintomatriisin $\widetilde{\mathbf{U}}$ singulaariarvohajotelman ja pääkomponenttianalyysin väliseen yhteyteen. Samoin matriisin $\widetilde{\mathbf{U}}$ approksimointi alempiasteisen matriisin avulla on mielenkiintoinen ja perehtymisen arvoinen juttu. Käy nimittäin niin. että Eckart–Young-tulos (s. 203) kertoo, että

$$\min_{\mathrm{r}(\mathbf{B})=1} \|\widetilde{\mathbf{X}} - \mathbf{B}\|_F^2 = \delta_2^2 + \cdots + \delta_p^2\,, \tag{8.205}$$

missä merkinnällisistä syistä $\widetilde{\mathbf{X}}$ viittaa keskistettyyn havaintomatriisiin (josta edellä käytettiin merkintää $\widetilde{\mathbf{U}}$). Olkoon matriisin $\widetilde{\mathbf{X}}$ singulaariarvohajotelma

$$\widetilde{\mathbf{X}} = \mathbf{U}\mathbf{\Delta}\mathbf{Q}'. \tag{8.206}$$

missä ”perinteinen” $\mathbf{V}$ on korvattu $\mathbf{Q}$:lla (matriisi $\mathbf{U}$ *ei* viittaa havaintomatriisiin):

$$\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} = \mathbf{Q}\mathbf{\Delta}'\mathbf{\Delta}\mathbf{Q}' := \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'. \tag{8.207}$$

Siis: $\mathbf{Q}$ sisältää matriisin $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$ ortonormaalit ominaisvektorit ja diagonaalimatriisi $\mathbf{\Lambda}$ sisältää sen ominaisarvot (matriisin $\widetilde{\mathbf{X}}$ neliöidyt singulaariarvot).

Minimi (8.205):ssa saavutetaan, kun valitaan

$$\mathbf{B} = \dot{\mathbf{B}}_1 = \delta_1\mathbf{u}_1\mathbf{q}_1'\,, \tag{8.208}$$

missä $\mathbf{q}_1$ on matriisin $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$ ensimmäinen ominaisvektori (tai yhtäpitävästi kovarianssimatriisin $\mathbf{S}$ ensimmäinen ominaisvektori), ts. matriisin $\widetilde{\mathbf{X}}$ ensimmäinen oikeanpuoleinen singulaarivektori:

$$\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\mathbf{q}_1 = \delta_1^2\mathbf{q}_1\,, \quad \widetilde{\mathbf{X}}\mathbf{q}_1 = \delta_1\mathbf{u}_1\,. \tag{8.209}$$

Täten

$$\dot{\mathbf{B}}_1 = \delta_1\mathbf{u}_1\mathbf{q}_1' = \widetilde{\mathbf{X}}\mathbf{q}_1\mathbf{q}_1' = \mathbf{s}_1\mathbf{q}_1'\,, \quad \text{missä } \mathbf{s}_1 = \widetilde{\mathbf{X}}\mathbf{q}_1\,, \tag{8.210}$$

mikä on täsmälleen sama tulos, joka saatiin kun minimoitiin ortogonaalia etäisyyksiä. Siis: matriisin $\widetilde{\mathbf{X}}$ paras 1-asteinen approksimaatio on $\dot{\mathbf{B}}_1 = \mathbf{s}_1\mathbf{q}_1'$. Matriisin $\dot{\mathbf{B}}_1 = \mathbf{s}_1\mathbf{q}_1'$ rivit ovat $\mathbf{q}_1'$:n monikertoja; matriisin $\dot{\mathbf{B}}_1' = \mathbf{q}_1\mathbf{s}_1' = \mathbf{q}_1\mathbf{q}_1'\widetilde{\mathbf{X}}'$ sarakkeet ovat havaintovektorien projektioita aliavaruudelle $\mathscr{C}(\mathbf{q}_1)$.

Ks. esim. Puntanen, Styan & Isotalo (2011, §19.4). □

8.8 Lineaarikombinaation maksimikorrelaatio

Olkoon $(k+1)$:n muuttujan havaintomatriisi $\mathbf{U}$ ositettu seuraavasti:

$$\mathbf{U} = (\mathbf{x}_1 : \ldots : \mathbf{x}_k : \mathbf{y}) = (\mathbf{X}_0 : \mathbf{y})\,, \tag{8.211}$$

ja olkoon tulosummamatriisi

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{\mathbf{xx}} & \mathbf{t}_{\mathbf{x}y} \\ \mathbf{t}_{\mathbf{x}y}' & t_{yy} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_0'\mathbf{C}\mathbf{X}_0 & \mathbf{X}_0'\mathbf{C}\mathbf{y} \\ \mathbf{y}'\mathbf{C}\mathbf{X}_0 & \mathbf{y}'\mathbf{C}\mathbf{y} \end{pmatrix} = \begin{pmatrix} \widetilde{\mathbf{X}}_0'\widetilde{\mathbf{X}}_0 & \widetilde{\mathbf{X}}_0'\tilde{\mathbf{y}} \\ \tilde{\mathbf{y}}'\widetilde{\mathbf{X}}_0 & \tilde{\mathbf{y}}'\tilde{\mathbf{y}} \end{pmatrix}. \tag{8.212}$$

Olkoon tehtävänä määrittää se muuttujavektoreiden $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_k$ lineaarikombinaatio

$$\mathbf{z} = \mathbf{X}_0\mathbf{b} = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \cdots + b_k\mathbf{x}_k\,, \tag{8.213}$$

jolla on mahdollisimman suuri korrelaatio (neliöitynä) muuttujan $\mathbf{y}$ kanssa.

Muuttujavektorien $\mathbf{y}$ ja $\mathbf{X}_0\mathbf{b}$ tulosumma on

$$\mathrm{ssp}(\mathbf{y}, \mathbf{X}_0\mathbf{b}) = \mathbf{y}'\mathbf{C}\mathbf{X}_0\mathbf{b} = \mathbf{t}_{\mathbf{x}y}'\mathbf{b}\,, \tag{8.214}$$

ja neliösummat

$$\mathrm{ssp}(\mathbf{y}) = \mathbf{y}'\mathbf{C}\mathbf{y} = t_{yy}, \quad \mathrm{ssp}(\mathbf{X}_0\mathbf{b}) = \mathbf{b}'\mathbf{X}_0'\mathbf{C}\mathbf{X}_0\mathbf{b} = \mathbf{b}'\mathbf{T}_{\mathbf{xx}}\mathbf{b}\,. \tag{8.215}$$

Muuttujien $\mathbf{y}$ ja $\mathbf{X}_0\mathbf{b}$ välisen korrelaatiokertoimen neliö on täten

$$\mathrm{cor}_{\mathrm{d}}^2(\mathbf{y}, \mathbf{X}_0\mathbf{b}) = \frac{(\mathbf{t}_{\mathbf{x}y}'\mathbf{b})^2}{t_{yy}\,\mathbf{b}'\mathbf{T}_{\mathbf{xx}}\mathbf{b}} := f(\mathbf{b}). \tag{8.216}$$

Lausekkeen $f(\mathbf{b})$ maksimoinnissa voimme kätevästi soveltaa Cauchyn–Schwarzin epäyhtälön versiota (oletus: $\mathbf{A} \in \mathrm{PD}_n$):

$$(\mathbf{u}'\mathbf{v})^2 \leq \mathbf{u}'\mathbf{A}\mathbf{u} \cdot \mathbf{v}'\mathbf{A}^{-1}\mathbf{v}\,, \tag{8.217}$$

missä yhtäsuuruus toteutuu täsmälleen silloin kun $\mathbf{A}\mathbf{u} = \lambda\mathbf{v}$ jollakin λ:n arvolla. Valitaan nyt

$$\mathbf{u} = \mathbf{b}\,, \;\; \mathbf{v} = \mathbf{t}_{\mathbf{x}y}\,, \;\; \mathbf{A} = \mathbf{T}_{\mathbf{xx}}\,. \tag{8.218}$$

Tällöin

$$(\mathbf{t}_{\mathbf{x}y}'\mathbf{b})^2 \leq \mathbf{b}'\mathbf{T}_{\mathbf{xx}}\mathbf{b} \cdot \mathbf{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}, \;\; \forall\, \mathbf{b} \in \mathbb{R}^k, \tag{8.219}$$

joten

$$\begin{aligned}\operatorname{cor}_{\mathrm{d}}^2(\mathbf{y}, \mathbf{X}_0\mathbf{b}) &= \frac{(\mathbf{t}_{\mathbf{x}y}'\mathbf{b})^2}{t_{yy}\,\mathbf{b}'\mathbf{T}_{\mathbf{xx}}\mathbf{b}} \\ &\leq \frac{\mathbf{b}'\mathbf{T}_{\mathbf{xx}}\mathbf{b}\cdot\mathbf{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}}{t_{yy}\,\mathbf{b}'\mathbf{T}_{\mathbf{xx}}\mathbf{b}} \\ &= \frac{\mathbf{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}}{t_{yy}}. \end{aligned} \tag{8.220}$$

Yhtäsuuruus (8.220):ssä toteutuu täsmälleen silloin kun on olemassa $\lambda \in \mathbb{R}$ siten että

$$\mathbf{T}_{\mathbf{xx}}\mathbf{b} = \lambda\mathbf{t}_{\mathbf{x}y} \tag{8.221}$$

eli kun $\mathbf{b}$ on vektorin

$$\hat{\boldsymbol{\beta}}_{\mathbf{x}} = \mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y} = \mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{s}_{\mathbf{x}y} \tag{8.222}$$

monikerta. Täten

$$\max_{\mathbf{b}} \operatorname{cor}_{\mathrm{d}}^2(\mathbf{y}, \mathbf{X}_0\mathbf{b}) = \operatorname{cor}_{\mathrm{d}}^2(\mathbf{y}, \mathbf{X}_0\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}) = \frac{\mathbf{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}}{t_{yy}}. \tag{8.223}$$

Olemme jo aiemmin maininneet että

$$\frac{\mathbf{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}}{t_{yy}} = \frac{\mathbf{s}_{\mathbf{x}y}'\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{s}_{\mathbf{x}y}}{s_y^2} = \mathbf{r}_{\mathbf{x}y}'\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{\mathbf{x}y} = R_{y\cdot\mathbf{x}}^2 = R^2, \tag{8.224}$$

missä R^2 on selitysaste regressiomallissa kun $\mathbf{y}$:tä selitetään $\mathbf{x}$-muuttujilla (+ vakiolla).

Tulokseen (8.223) voidaan päätyä myös ortogonaaliprojektion kautta. Koska korrelaatiokerroin on sama kuin keskistettyjen muuttujavektoreiden välisen kulman kosini, on tehtävänä maksimoida lauseke

$$f(\mathbf{b}) = \operatorname{cor}_{\mathrm{d}}^2(\mathbf{y}, \mathbf{X}_0\mathbf{b}) = \cos^2(\tilde{\mathbf{y}}, \widetilde{\mathbf{X}}_0\mathbf{b})\,. \tag{8.225}$$

Jos vektori $\tilde{\mathbf{y}}$ on annettu, niin mikä on se vektori $\widetilde{\mathbf{X}}_0\hat{\mathbf{b}}$, jonka kulma $\tilde{\mathbf{y}}$:n kanssa on pienin (eli kosini suurin)? Tämä vektori on luonnollisesti [miksi– ks. Puntanen, Styan & Isotalo (2011, Prop. 2.3)] $\tilde{\mathbf{y}}$:n projektio sarakeavaruudelle $\mathscr{C}(\widetilde{\mathbf{X}}_0)$:

$$\widetilde{\mathbf{X}}_0\hat{\mathbf{b}} = \mathbf{P}_{\widetilde{\mathbf{X}}_0}\tilde{\mathbf{y}} = \widetilde{\mathbf{X}}_0(\widetilde{\mathbf{X}}_0'\widetilde{\mathbf{X}}_0)^{-1}\widetilde{\mathbf{X}}_0'\tilde{\mathbf{y}} = \widetilde{\mathbf{X}}_0\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}\,. \tag{8.226}$$

Täten $\hat{\mathbf{b}} = \hat{\boldsymbol{\beta}}_{\mathbf{x}} = \mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}$.

Entäpä jos tehtävänä on määrittää se $\mathbf{a}$:n arvo $\hat{\mathbf{a}}$ jolla on ominaisuus

$$\min_{\mathbf{a}} \operatorname{var}_{\mathrm{d}}(\mathbf{y} - \mathbf{X}_0\mathbf{a}) = \operatorname{var}_{\mathrm{d}}(\mathbf{y} - \mathbf{X}_0\hat{\mathbf{a}})\,? \tag{8.227}$$

Koska

$$\operatorname{var}_{\mathrm{d}}(\mathbf{y} - \mathbf{X}_0\mathbf{a}) = \frac{1}{n-1}\,\|\mathbf{C}\mathbf{y} - \mathbf{C}\mathbf{X}_0\hat{\mathbf{a}}\|^2\,, \tag{8.228}$$

on tehtävänä löytää se $\widetilde{\mathbf{X}}_0$:n sarakkeiden lineaarikombinaatio $\widetilde{\mathbf{X}}_0\hat{\mathbf{a}}$, jolla on ominaisuus

$$\min_{\mathbf{a}} \|\tilde{\mathbf{y}} - \widetilde{\mathbf{X}}_0\mathbf{a}\|^2 = \|\tilde{\mathbf{y}} - \widetilde{\mathbf{X}}_0\hat{\mathbf{a}}\|^2. \tag{8.229}$$

Ratkaisu tähän tehtävään on

$$\widetilde{\mathbf{X}}_0\hat{\mathbf{a}} = \mathbf{P}_{\widetilde{\mathbf{X}}_0}\tilde{\mathbf{y}} = \widetilde{\mathbf{X}}_0\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y} \tag{8.230}$$

eli $\hat{\mathbf{a}} = \hat{\boldsymbol{\beta}}_{\mathbf{x}} = \mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}$. Kannattaa panna merkille, että vastaavanlaiset tulokset saavutetaan myös satunnaismuutujien (populaation) tapauksessa; ks. (7.96) (s. 227).

Kirjattakoon vielä seuraava yhteenveto:

> *Vektorin* $\mathbf{b}$ *ratkaiseminen tehtävistä*
>
> $$\min_{\mathbf{b}} \|\tilde{\mathbf{y}} - \widetilde{\mathbf{X}}_0\mathbf{b}\|^2 \quad \textit{ja} \quad \max_{\mathbf{b}} \cos^2(\tilde{\mathbf{y}}, \widetilde{\mathbf{X}}_0\mathbf{b}) \tag{8.231}$$
>
> *ts.*
>
> $$\min_{\mathbf{b}} \operatorname{var}_{\mathrm{d}}(\mathbf{y} - \mathbf{X}_0\mathbf{b}) \quad \textit{ja} \quad \max_{\mathbf{b}} \operatorname{cor}_{\mathrm{d}}^2(\mathbf{y}, \mathbf{X}_0\mathbf{b}) \tag{8.232}$$
>
> *johtaa (oleellisesti) samaan tulokseen:* $\hat{\boldsymbol{\beta}}_{\mathbf{x}} = \mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y} = \mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{s}_{\mathbf{x}y}$.

8.9 Osittaiskorrelaatiomatriisi

Tarkastellaan sitten osittaiskorrelaatiokertoimien laskemista matriisioperaaatioin. Ositetaan havaintomatriisi $\mathbf{U}$ seuraavasti:

$$\mathbf{U} = (\mathbf{X} : \mathbf{Y}), \quad \mathbf{X} = (\mathbf{x}_1 : \ldots : \mathbf{x}_p), \quad \mathbf{Y} = (\mathbf{y}_1 : \ldots : \mathbf{y}_q). \tag{8.233}$$

Nyt $\mathbf{y}_1, \ldots, \mathbf{y}_q$ ovat ne muuttujat, joiden keskinäiset osittaiskorrelaatiokertoimet haluamme laskea, kun vakioidut muuttujat ovat $\mathbf{X}$:n sisältämät muuttujat $\mathbf{x}_1, \ldots, \mathbf{x}_p$. Tällöin muuttujan $\mathbf{y}_i$ residuaali, kun siitä eliminoidaan muuttujien $\mathbf{x}_1, \ldots, \mathbf{x}_p$ vaikutus, on määritelty seuraavasti:

$$\mathbf{e}_{\mathbf{y}_i\cdot\mathbf{X}} = (\mathbf{I}_n - \mathbf{P}_{(\mathbf{1}:\mathbf{X})})\mathbf{y}_i = \operatorname{res}(\mathbf{y}_i; \mathbf{X}), \quad i = 1, \ldots, m. \tag{8.234}$$

Määritelmässä (8.234) on huomattava vakiovektorin $\mathbf{1} \in \mathbb{R}^n$ rooli – sekin on mukana eliminoitavien muuttujien joukossa. Merkitään residuaalien muodostamaa havaintomatriisia $\mathbf{E}_{\mathbf{Y}\cdot\mathbf{X}}$:lla, jolloin

$$\begin{aligned}
\mathbf{E}_{\mathbf{Y}\cdot\mathbf{X}} := \operatorname{res}(\mathbf{Y}; \mathbf{X}) &= (\mathbf{e}_{\mathbf{y}_1\cdot\mathbf{X}} : \ldots : \mathbf{e}_{\mathbf{y}_q\cdot\mathbf{X}}) \\
&= [(\mathbf{I}_n - \mathbf{P}_{(\mathbf{1}:\mathbf{X})})\mathbf{y}_1 : \ldots : (\mathbf{I}_n - \mathbf{P}_{(\mathbf{1}:\mathbf{X})})\mathbf{y}_q] \\
&= (\mathbf{I}_n - \mathbf{P}_{(\mathbf{1}:\mathbf{X})})(\mathbf{y}_1 : \ldots : \mathbf{y}_q) \\
&= (\mathbf{I}_n - \mathbf{P}_{(\mathbf{1}:\mathbf{X})})\mathbf{Y}.
\end{aligned} \tag{8.235}$$

Matriisin $\mathbf{E}_{\mathbf{Y}\cdot\mathbf{X}} \in \mathbb{R}^{n\times q}$ sarakkeet ovat siis $\mathbf{y}_i$:tten residuaaleja. Näiden residuaalien tavalliset korrelaatiokertoimet ovat osittaiskorrelaatiokertoimia. Esimerkiksi $\mathbf{y}_1$:n ja $\mathbf{y}_2$:n välinen osittaiskorrelaatiokerroin (kun muuttujien $\mathbf{x}_1$, ..., $\mathbf{x}_p$ vaikutus on eliminoitu) on

$$r_{12\cdot m+1,\dots,k} = r_{12\cdot\mathbf{X}} = \operatorname{pcor_d}(\mathbf{y}_1, \mathbf{y}_2 \mid \mathbf{X}) = \operatorname{cor_d}(\mathbf{e}_{\mathbf{y}_1\cdot\mathbf{X}}, \mathbf{y}_{\mathbf{y}_2\cdot\mathbf{X}}). \tag{8.236}$$

Täten $\mathbf{X}$:n sisältämien muuttujien osittaiskorrelaatiomatriisi on korrelaatiomatriisi $\mathbf{E}_{\mathbf{Y}\cdot\mathbf{X}}$:sta eli

$$\operatorname{pcor_d}(\mathbf{Y} \mid \mathbf{X}) = \operatorname{cor_d}(\mathbf{E}_{\mathbf{Y}\cdot\mathbf{X}}) = \operatorname{cor_d}[\operatorname{res}(\mathbf{Y}; \mathbf{X})]. \tag{8.237}$$

Määritetään seuraavaksi osittaiskorrelaatiomatriisin eksplisiittinen lauseke. On osoitettavissa (ks. luku 9, s. 330), että projektorilla $\mathbf{P}_{(\mathbf{1}:\mathbf{X})}$ on hajotelma

$$\mathbf{P}_{(\mathbf{1}:\mathbf{X})} = \mathbf{J} + \mathbf{P}_{\mathbf{CX}}, \tag{8.238}$$

jolloin

$$\mathbf{I}_n - \mathbf{P}_{(\mathbf{1}:\mathbf{X})} = \mathbf{I}_n - \mathbf{J} - \mathbf{P}_{\mathbf{CX}} = \mathbf{C} - \mathbf{P}_{\mathbf{CX}}, \tag{8.239}$$

missä

$$\mathbf{P}_{\mathbf{CX}} = \mathbf{CX}(\mathbf{X}'\mathbf{CX})^{-1}\mathbf{X}'\mathbf{C}. \tag{8.240}$$

Täten

$$\mathbf{E}_{\mathbf{Y}\cdot\mathbf{X}} = (\mathbf{I}_n - \mathbf{P}_{(\mathbf{1}:\mathbf{X})})\mathbf{Y} = (\mathbf{C} - \mathbf{P}_{\mathbf{CX}})\mathbf{Y}. \tag{8.241}$$

Havaitsemme heti, että residuaalien muodostama matriisi $\mathbf{E}_{\mathbf{Y}\cdot\mathbf{X}}$ on keskistetty, sillä $\mathbf{CE}_{\mathbf{Y}\cdot\mathbf{X}} = \mathbf{E}_{\mathbf{Y}\cdot\mathbf{X}}$. Matriisista $\mathbf{E}_{\mathbf{Y}\cdot\mathbf{X}}$ laskettu tulosummamatriisi on tällöin

$$\begin{aligned}\mathbf{T}_{\mathbf{yy}\cdot\mathbf{x}} &= \mathbf{E}'_{\mathbf{Y}\cdot\mathbf{X}}\mathbf{E}_{\mathbf{Y}\cdot\mathbf{X}} \\ &= \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_{(\mathbf{1}:\mathbf{X})})\mathbf{Y} = \mathbf{Y}'(\mathbf{C} - \mathbf{P}_{\mathbf{CX}})\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{CY} - \mathbf{Y}'\mathbf{CX}(\mathbf{X}'\mathbf{CX})^{-1}\mathbf{X}'\mathbf{CY}.\end{aligned} \tag{8.242}$$

Matriisi $\mathbf{T}_{\mathbf{yy}\cdot\mathbf{x}}$ on siis residuaalimuuttujien neliösummien ja tulosummien matriisi eli residuaalien tulosummamatriisi. Voimme käyttää siitä nimitystä

$$\mathbf{T}_{\mathbf{yy}\cdot\mathbf{x}} = \text{osittaistulosummamatriisi}. \tag{8.243}$$

Osittaiskorrelaatiomatriisi on täten

$$\boxed{\operatorname{pcor_d}(\mathbf{Y} \mid \mathbf{X}) = [\operatorname{diag}(\mathbf{T}_{\mathbf{yy}\cdot\mathbf{x}})]^{-1/2}\mathbf{T}_{\mathbf{yy}\cdot\mathbf{x}}[\operatorname{diag}(\mathbf{T}_{\mathbf{yy}\cdot\mathbf{x}})]^{-1/2}. \qquad (8.244)}$$

Olkoon $\mathbf{T}$ kaikkien $\mathbf{x}_i$- ja $\mathbf{y}_i$-muuttujien tulosummamatriisi:

$$\begin{aligned}\mathbf{T} = \mathbf{U}'\mathbf{CU} &= \begin{pmatrix}\mathbf{X}'\mathbf{CX} & \mathbf{X}'\mathbf{CY} \\ \mathbf{Y}'\mathbf{CX} & \mathbf{Y}'\mathbf{CY}\end{pmatrix} \\ &= \begin{pmatrix}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} & \widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}} \\ \widetilde{\mathbf{Y}}'\widetilde{\mathbf{X}} & \widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}\end{pmatrix} = \begin{pmatrix}\mathbf{T}_{\mathbf{xx}} & \mathbf{T}_{\mathbf{xy}} \\ \mathbf{T}_{\mathbf{yx}} & \mathbf{T}_{\mathbf{yy}}\end{pmatrix}.\end{aligned} \tag{8.245}$$

Tällöin $\mathbf{T_{yy\cdot x}}$ voidaan esittää seuraavasti:

$$\mathbf{T_{yy\cdot x}} = \mathbf{T_{yy}} - \mathbf{T_{yx}}\mathbf{T_{xx}^{-1}}\mathbf{T_{xy}} = \tilde{\mathbf{Y}}'(\mathbf{I}_n - \mathbf{P}_{\tilde{\mathbf{X}}})\tilde{\mathbf{Y}}. \tag{8.246}$$

Jakamalla $\mathbf{T_{yy\cdot x}}$:n sopivalla vakiolla saamme *osittaiskovarianssimatriisin*, jonka teoreettinen vastine on

$$\mathbf{\Sigma_{yy\cdot x}} = \mathbf{\Sigma_{xx}} - \mathbf{\Sigma_{yx}}\mathbf{\Sigma_{xx}^{-1}}\mathbf{\Sigma_{xy}}. \tag{8.247}$$

Matriisi $\mathbf{T_{yy\cdot x}}$ on $\mathbf{T_{xx}}$:n Schurin komplementti $\mathbf{T}$:ssä, joten sen käänteismatriisi on $\mathbf{T}^{-1}$:n oikea alanurkka:

$$\mathbf{T}^{-1} = \begin{pmatrix} \mathbf{T_{xx}} & \mathbf{T_{xy}} \\ \mathbf{T_{yx}} & \mathbf{T_{yy}} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{T^{xx}} & \mathbf{T^{xy}} \\ \mathbf{T^{yx}} & \mathbf{T^{yy}} \end{pmatrix} = \begin{pmatrix} \mathbf{T_{xx\cdot y}^{-1}} & \mathbf{T^{xy}} \\ \mathbf{T^{yx}} & \mathbf{T_{yy\cdot x}^{-1}} \end{pmatrix} = \{t^{ij}\}. \tag{8.248}$$

Voimme ilmaista osittaiskorrelaatiomatriisin myös seuraavasti

$$\mathrm{pcor_d}(\mathbf{Y} \mid \mathbf{X}) = [\mathrm{diag}(\mathbf{R_{yy\cdot x}})]^{-1/2}\mathbf{R_{yy\cdot x}}[\mathrm{diag}(\mathbf{R_{yy\cdot x}})]^{-1/2} \tag{8.249}$$

missä

$$\mathbf{R_{yy\cdot x}} = \mathbf{R_{yy}} - \mathbf{R_{yx}}\mathbf{R_{xx}^{-1}}\mathbf{R_{xy}}, \tag{8.250}$$

$$\mathbf{R}^{-1} = [\mathrm{cor_d}(\mathbf{U})]^{-1} = \begin{pmatrix} \mathbf{R_{xx}} & \mathbf{R_{xy}} \\ \mathbf{R_{yx}} & \mathbf{R_{yy}} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{R_{xx\cdot y}^{-1}} & \mathbf{R^{xy}} \\ \mathbf{R^{yx}} & \mathbf{R_{yy\cdot x}^{-1}} \end{pmatrix}. \tag{8.251}$$

Esimerkki 8.3. Tarkastellaan symmetristä matriisia

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix}, \tag{8.252}$$

Olkoon $\mathbf{R}$ ”ehdokas” korrelaatiomatriisiksi. Tällöin on välttämättä oltava

$$|r_{12}|,\ |r_{13}| \text{ ja } |r_{23}| \leq 1. \tag{8.253}$$

Oletetaan että (8.253) on voimassa. Osoita, että tällöin $\mathbf{R}$ on korrelaatiomatriisi (ts. $\mathbf{R}$ on ei-negatiivisesti definiitti) jos ja vain jos jokin seuraavista yhtäpitävistä ehdoista on voimassa.

(a) $|r_{12\cdot 3}| \leq 1$,

(b) $|r_{13\cdot 2}| \leq 1$,

(c) $|r_{23\cdot 1}| \leq 1$,

(d) $r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1$,

(e) $\det(\mathbf{R}) \geq 0$.

Määritä edellisten ehtojen perusteella r siten että seuraava $\mathbf{R}$ todella on korrelaatiomatriisi (tulos: $|r| = 1/\sqrt{2}$)

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & -r \\ 0 & 1 & r \\ -r & r & 1 \end{pmatrix}.$$

Ks. myös HT 1.2 (s. 62). □

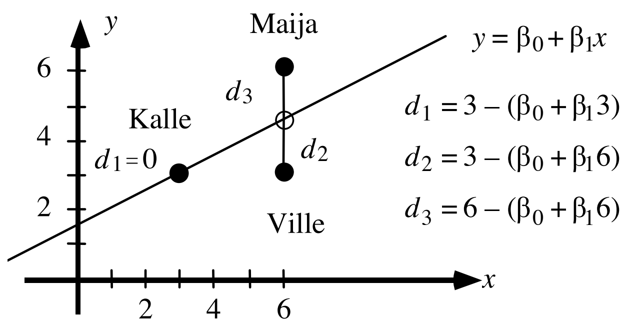
8.10 Lineaarinen malli

Oletetaan, että tehtävänä on sijoittaa kuvioon 8.13 suora $y = \beta_0 + \beta_1 x$ siten, että sijoittamisen kriteeri on seuraava:

> lasketaan havaintojen etäisyydet suorasta – y-akselin suuntaiset etäisyydet – ja minimoidaan näiden etäisyyksien neliöiden summa.

Tämä kriteeri tarkoittaa ns. *pienimmän neliösumman menetelmän* käyttöä.

8.10.1 Pienimmän neliösumman menetelmä



Kuvio 8.13. Minimoidaan havaintojen y-akselin suuntaisten etäisyyksien neliösummaa suorasta $y = \beta_0 + \beta_1 x$.

Olkoon $|d_i|$ = havaintopisteen y-akselin suuntainen etäisyys suorasta $y = \beta_0 + \beta_1 x$:

$$d_i = y_i - (\beta_0 + \beta_1 x_i).$$

On ”selvää”, että d_i:tten neliösumma minimoituu, kun suora kulkee täsmälleen Maijan ja Villen y-arvojen puolivälistä. Näin saadaan kuvion 8.13 mukainen suora, joka on *regressiosuora*. Minimoitavan funktion minimiarvoa sanotaan *jäännösneliösummaksi* (SSE).

Etäisyyksien d_i neliöiden summa on n havainnon tapauksessa

$$\begin{aligned}
\mathrm{SSE}(\mathbf{y}; \beta_0\mathbf{1} + \beta_1\mathbf{x}) &= d_1^2 + d_2^2 + \ldots + d_n^2 \\
&= [y_1 - (\beta_0 + \beta_1 x_1)]^2 + [y_2 - (\beta_0 + \beta_1 x_2)]^2 + \cdots \\
&\quad + [y_n - (\beta_0 + \beta_1 x_n)]^2 \\
&= \|\mathbf{y} - (\beta_0\mathbf{1} + \beta_1\mathbf{x})\|^2 \\
&= \left\| \mathbf{y} - (\mathbf{1} : \mathbf{x}) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \right\|^2 \\
&= \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \qquad (8.254)
\end{aligned}$$

missä

$$\mathbf{X} = (\mathbf{1} : \mathbf{x}) = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \text{ns. } \textit{mallimatriisi},$$

$$\mathbf{X}\boldsymbol{\beta} = (\mathbf{1} : \mathbf{x}) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \beta_0\mathbf{1} + \beta_1\mathbf{x} = \begin{pmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_1 x_2 \\ \vdots \\ \beta_0 + \beta_1 x_n \end{pmatrix}. \qquad (8.255)$$

Meidän on siis määritettävä se sarakeavaruuden $\mathscr{C}(\mathbf{1} : \mathbf{x}) = \mathscr{C}(\mathbf{X})$ vektori eli se mallimatriisin $\mathbf{X}$ sarakkeiden lineaarikombinaatio $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, joka on kaikkein lähinnä annettua vektoria $\mathbf{y}$. Projektiotulosten avulla on pääteltävissä, että minimi saavutetaan, kun $\mathbf{X}\hat{\boldsymbol{\beta}}$ on $\mathbf{y}$:n ortogonaaliprojektio $\mathscr{C}(\mathbf{X})$:lle eli $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P}_{\mathbf{X}}\mathbf{y}$:

$$\begin{aligned}
\min_{\boldsymbol{\beta}} \mathrm{SSE}(\mathbf{y}; \mathbf{X}\boldsymbol{\beta}) &= \min_{\mathbf{t} \in \mathscr{C}(\mathbf{X})} \mathrm{SSE}(\mathbf{y}; \mathbf{t}) \\
&= \min_{\boldsymbol{\beta}} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \\
&= \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\
&= \mathrm{SSE}(\mathbf{y}; \mathbf{X}\hat{\boldsymbol{\beta}}) \\
&= \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \\
&= \|\mathbf{y} - \mathbf{H}\mathbf{y}\|^2 = \mathbf{y}'(\mathbf{I}_n - \mathbf{H})\mathbf{y}, \qquad (8.256)
\end{aligned}$$

missä

$$\mathbf{H} = \mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad \text{ortogonaaliprojektori } \mathscr{C}(\mathbf{X})\text{:lle,} \tag{8.257a}$$

$$\mathbf{H}\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\mathbf{y}} \quad \text{mallin antama } \mathbf{y}\text{-arvo, sovite,} \tag{8.257b}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \quad \text{regressiokertoimien PNS-estimaatit.} \tag{8.257c}$$

Kun matriisimerkinnät puretaan, tulokseksi saadaan lukijan muistamat lausekkeet

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}, \tag{8.258a}$$

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\mathrm{SP}_{xy}}{\mathrm{SS}_x} = r_{xy}\frac{s_y}{s_x}. \tag{8.258b}$$

Regressiosuoran yhtälö on siis

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = (\bar{y} - \hat{\beta}_1\bar{x}) + \hat{\beta}_1 x = \bar{y} + \hat{\beta}_1(x - \bar{x}) \tag{8.259}$$

eli

$$\hat{y} - \bar{y} = \hat{\beta}_1(x - \bar{x}), \tag{8.260}$$

joten regressiosuora kulkee keskiarvopisteen $(\bar{x}, \bar{y})$ kautta.

8.10.2 Lineaarinen malli

Voimme ajatella, että y-havainnot ovat syntyneet seuraavan mekanismin mukaan:

$$\mathscr{M}_{01} : y = \beta_0 + \beta_1 x + \text{satunnaisvirhe } \varepsilon, \tag{8.261a}$$

ts.

$$y_i = \beta_0 + \beta_1 x_i + \text{satunnaisvirhe } \varepsilon_i, \ \ i = 1, 2, \dots, n. \tag{8.261b}$$

Satunnaisvirhe ε on satunnaismuuttuja ja täten myös y on satunnaismuuttuja. Sen sijaan (tilanteesta tosin riippuen) x ei ole satunnaismuuttuja, vaan sellainen muuttuja, jonka arvoja voidaan kontrolloida. Voimme ajatella mallin $\mathscr{M}_{01}$ merkitsevän sitä, että kun x:lle annetaan arvot $x_1, \dots, x_n$, niin havaitsemme y:lle arvot $y_1, \dots, y_n$, joiden uskomme syntyneen (8.261b):n mukaisesti. (Merkinnällisesti emme tee eroa satunnaismuuttujan ja sen havaitun arvon välillä.) Tavallisin oletus satunnaisvirheistä ε_i on, että

$$\varepsilon_i\text{:t ovat toisistaan riippumattomia,} \tag{8.262a}$$

$$\mathrm{E}(\varepsilon_i) = 0 \text{ ja } \mathrm{var}(\varepsilon_i) = \sigma^2, \quad i = 1, 2, \dots, n. \tag{8.262b}$$

Matriisimerkinnöin voimme merkitä mallia lyhyesti

$$\mathscr{M}_{01} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\} = \{\mathbf{y}, (\mathbf{1} : \mathbf{x})\boldsymbol{\beta}, \sigma^2\mathbf{I}\}, \tag{8.263}$$

Tarkastellaan sitten mallia $\mathscr{M}$, jossa on vakiotermin lisäksi k varsinaista selittäjää, eli

$$\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\} = \{\mathbf{y}, (\mathbf{1} : \mathbf{X}_0)\boldsymbol{\beta}, \sigma^2\mathbf{I}\}, \tag{8.264}$$

missä $\mathbf{X} = (\mathbf{1} : \mathbf{X}_0) = (\mathbf{1} : \mathbf{x}_1 : \ldots : \mathbf{x}_k)$. Tällöin

$$\mathbf{y} = \beta_0\mathbf{1} + \beta_1\mathbf{x}_1 + \cdots + \beta_k\mathbf{x}_k + \boldsymbol{\varepsilon}, \quad \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}. \tag{8.265}$$

Malli $\mathscr{M}$ merkitsee että

$$\mathrm{E}(\mathbf{y}) = \beta_0\mathbf{1} + \beta_1\mathbf{x}_1 + \cdots + \beta_k\mathbf{x}_k = \mathbf{X}\boldsymbol{\beta} \in \mathscr{C}(\mathbf{X}), \tag{8.266}$$

eli $\mathbf{y}$:n odotusarvo on jokin $\mathbf{X}$:n sarakkeiden lineaarikombinaatio. Vektori $\boldsymbol{\beta}$ on tuntematon ja se on estimoitava havaitun otoksen perusteella.

8.10.3 PNS-estimointi usean selittäjän tapauksessa

Vektori $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ on $\mathbf{y}$:n odotusarvon $\mathbf{X}\boldsymbol{\beta}$ pienimmän neliösumman estimaatti, jos $\hat{\mathbf{y}}$ minimoi lausekkeen

$$\begin{aligned}\mathrm{SSE}(\mathbf{y}; \mathbf{X}\boldsymbol{\beta}) &= \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta}.\end{aligned} \tag{8.267}$$

Toisin sanoen, vektorin $\hat{\mathbf{y}}$ on toteutettava ehto

$$\min_{\mathbf{t} \in \mathscr{C}(\mathbf{X})} \mathrm{SSE}(\mathbf{y}; \mathbf{t}) = \mathrm{SSE}(\mathbf{y}; \hat{\mathbf{y}}) := \mathrm{SSE}, \tag{8.268}$$

eli

$$\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 \leq \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \ \forall\boldsymbol{\beta}. \tag{8.269}$$

Meidän on siis määritettävä se aliavaruuden $\mathscr{C}(\mathbf{X})$ vektori $\mathbf{X}\hat{\boldsymbol{\beta}}$, joka on kaikkein lähinnä annettua vektoria $\mathbf{y}$. Projektiotulosten avulla on helppo päätellä, että minimi saavutetaan, kun $\hat{\mathbf{y}}$ on $\mathbf{y}$:n projektio aliavaruudelle $\mathscr{C}(\mathbf{X})$:

$$\hat{\mathbf{y}} = \mathbf{P}_{\mathbf{X}}\mathbf{y} = \mathbf{H}\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}}, \tag{8.270}$$

missä

$$\mathbf{H} = \mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \text{ortogonaaliprojektori } \mathscr{C}(\mathbf{X})\text{:lle}, \tag{8.271a}$$

$$\begin{aligned}\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} &= \hat{\beta}_0\mathbf{1} + \hat{\beta}_1\mathbf{x}_1 + \cdots + \hat{\beta}_k\mathbf{x}_k \\ &= \text{mallin antama } \mathbf{y}\text{-arvo, sovite} \\ &= \mathrm{E}(\mathbf{y})\text{:n PNS-estimaatti} \\ &= \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}),\end{aligned} \tag{8.271b}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}}_{\mathbf{x}} \end{pmatrix} \in \mathbb{R}^{k+1} \tag{8.271c}$$

$$= \text{regressiokertoimien PNS-estimaatit, OLSE}(\boldsymbol{\beta}).$$

Olemme nähneet [ks. (8.107), s. 271], että

$$\hat{\beta}_0 = \bar{y} - \hat{\boldsymbol{\beta}}'_{\mathbf{x}}\bar{\mathbf{x}} = \bar{y} - (\hat{\beta}_1\bar{x}_1 + \cdots + \hat{\beta}_k\bar{x}_k)\,, \tag{8.272a}$$

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\mathbf{x}} &= (\hat{\beta}_1, \dots, \hat{\beta}_k)' = (\mathbf{X}_0'\mathbf{C}\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{C}\mathbf{y} \\ &= (\widetilde{\mathbf{X}}_0'\widetilde{\mathbf{X}}_0)^{-1}\widetilde{\mathbf{X}}_0'\tilde{\mathbf{y}} = \mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y} = \mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{s}_{\mathbf{x}y} \in \mathbb{R}^k. \end{aligned} \tag{8.272b}$$

Residuaalivektori on nyt

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{M}\mathbf{y}. \tag{8.273}$$

Jäännösneliösumma SSE on residuaalivektorin normin neliö:

$$\text{SSE} = \|\hat{\boldsymbol{\varepsilon}}\|^2 = \|(\mathbf{I} - \mathbf{H})\mathbf{y}\|^2 = \|\mathbf{M}\mathbf{y}\|^2 = \mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}. \tag{8.274}$$

Komponenttimuodossa voimme kirjoittaa

$$\begin{aligned} \min_{\boldsymbol{\beta}} \text{SSE}(\mathbf{y}; \mathbf{X}\boldsymbol{\beta}) &= \min_{\boldsymbol{\beta}} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})]^2 \\ &= \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik})]^2 \\ &= \sum_{i=1}^n \hat{\varepsilon}_i^2. \end{aligned} \tag{8.275}$$

Voisimme määrittää SSE$(\mathbf{y}; \mathbf{X}\boldsymbol{\beta})$:n minimiarvon myös siten, että merkitsisimme SSE$(\mathbf{y}; \mathbf{X}\boldsymbol{\beta})$:n osittaisderivaatat $\boldsymbol{\beta}$:n suhteen nolliksi, jolloin saisimme ns. normaaliyhtälön

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}. \tag{NY}$$

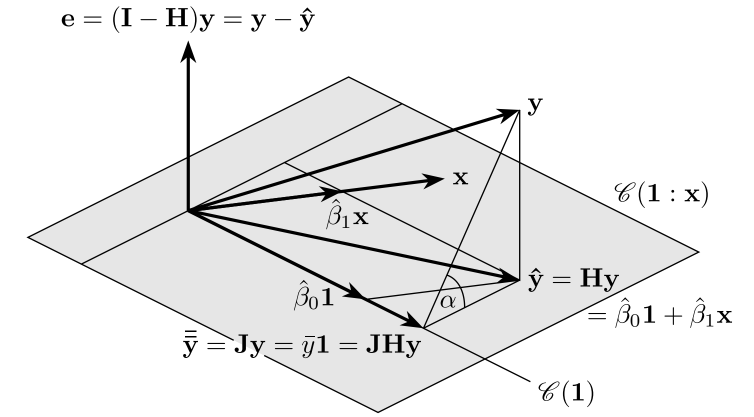
On helppo näyttää, ks. (8.21) (s. 255), että mikä hyvänsä (NY):n ratkaisu minimoi SSE$(\mathbf{y}; \mathbf{X}\boldsymbol{\beta})$:n. Ratkaisu on tietenkin yksikäsitteinen, jos $\mathbf{X}$:n sarakkeet ovat vapaat.

Edellisistä kuvioista ilmenee, että kun $\hat{\mathbf{y}}$ projisoidaan $\mathscr{C}(\mathbf{1})$:lle, saadaan tulokseksi $\mathbf{J}\mathbf{y}$:

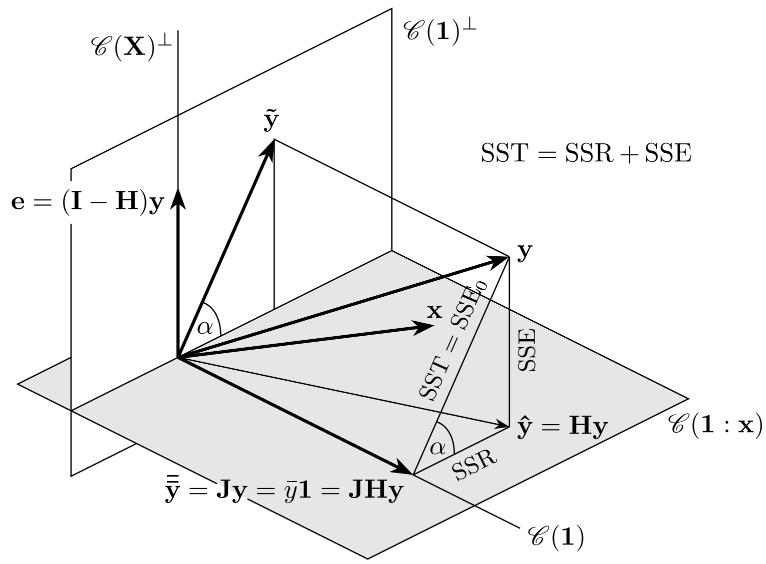
$$\mathbf{J}\hat{\mathbf{y}} = \mathbf{J}\mathbf{H}\mathbf{y} = \mathbf{J}\mathbf{y}. \tag{8.276}$$

Tämä on aivan oleellista, jotta syntyisi suorakulmainen kolmio, johon regressioanalyysin neliösummahajoitelma (ra) perustuu. Tulos (8.276) seuraa yhtälöstä

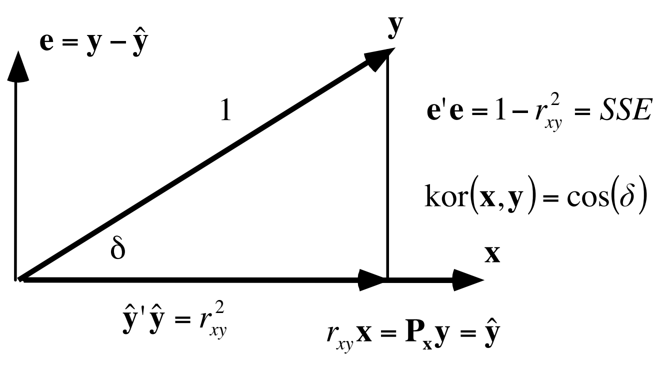
$$\mathbf{J}\mathbf{H} = \mathbf{H}\mathbf{J} = \mathbf{J}, \tag{8.277}$$



Kuvio 8.14. Vektorin $\mathbf{y}$ projisointi sarakeavaruudelle $\mathscr{C}(\mathbf{1} : \mathbf{x})$.



Kuvio 8.15. Hajotelma SST = SSR + SSE geometrisesti.



Kuvio 8.16. Yhden selittäjän regressioanalyysi standardoiduilla muuttujilla.

josta on helppo vakuuttautua. Ensinnäkin

$$\mathbf{HJ} = \mathbf{H1} \cdot \mathbf{1}'/n = \mathbf{1} \cdot \mathbf{1}'/n = \mathbf{J}, \tag{8.278}$$

koska $\mathbf{1}$ kuuluu $\mathbf{X}$:n sarakeavaruuteen. Koska $\mathbf{J}$ on symmetrinen, on myös $\mathbf{HJ}$:n oltava symmetrinen ja täten (8.277) on todistettu. On erityisesti huomattava, että (8.278):ssa oletamme, että kyseessä on malli, jossa vakiotermi on mukana eksplisiittisesti tai ehto $\mathbf{1} \in \mathscr{C}(\mathbf{X})$ toteutuu.

Esimerkki 8.4. *Usean selitettävän muuttujan lineaarinen malli.* Aiemmin on tarkasteltu lineaarista mallia, jossa on vain yksi selitettävä muuttuja eli vastemuuttuja y, jonka havaitut arvot ovat vektorissa $\mathbf{y}$. Olkoon meillä nyt kaksi vastemuuttujaa y_1 and y_2, joiden havaitut arvot ovat matriisissa $\mathbf{Y}$:

$$\mathbf{Y} = (\mathbf{y}_1 : \mathbf{y}_2) = \begin{pmatrix} \mathbf{y}'_{(1)} \\ \vdots \\ \mathbf{y}'_{(n)} \end{pmatrix}. \tag{8.279}$$

Jos nyt uskotaan, että

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i\,, \ \mathrm{E}(\boldsymbol{\varepsilon}_i) = \mathbf{0}\,, \ \mathrm{cov}(\boldsymbol{\varepsilon}_i) = \sigma_{ii}\mathbf{I}_n\,, \ \mathrm{cov}(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2) = \sigma_{12}\mathbf{I}_n\,, \tag{8.280}$$

missä $i = 1, 2$, niin meillä on *usean selitettävän muuttujan lineaarinen malli* (multivariate linear model), joka voidaan esittää muodossa

$$\mathbf{Y}_{n\times 2} = \mathbf{X}_{n\times p}\mathbb{B}_{p\times 2} + \boldsymbol{\varepsilon} n \times 2\,, \tag{8.281}$$

missä

$$\mathbb{B} = (\boldsymbol{\beta}_1 : \boldsymbol{\beta}_2)\,, \quad \boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1 : \boldsymbol{\varepsilon}_2)\,. \tag{8.282}$$

Tässä mallissa (usein) oletetaan, että $\mathbf{Y}$:n vaakarivit (havaintovektorit) ovat korreloimattomia eli $\mathrm{cov}(\mathbf{y}_{(r)}, \mathbf{y}_{(s)}) = \mathbf{0}$, $r \neq s$. Kun käytetään Kroneckerin tuloa ja vec-operaatiota, (8.281) saadaan muotoon

$$\begin{aligned} \mathrm{vec}(\mathbf{Y}) = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} &= \begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{pmatrix} \\ &= (\mathbf{I}_2 \otimes \mathbf{X})\,\mathrm{vec}(\mathbb{B}) + \mathrm{vec}(\boldsymbol{\varepsilon})\,. \end{aligned} \tag{8.283}$$

Luvun 6.9 (s. 210) mukaan saamme

$$\mathrm{cov}[\mathrm{vec}(\mathbf{Y})] = \begin{pmatrix} \sigma_{11}\mathbf{I}_n & \sigma_{12}\mathbf{I}_n \\ \sigma_{21}\mathbf{I}_n & \sigma_{22}\mathbf{I}_n \end{pmatrix} = \boldsymbol{\Sigma} \otimes \mathbf{I}_n\,, \tag{8.284}$$

ja täten usean selitettävän muuttujan malli (8.281) voidaan esittää yhden selitettävän muuttujan mallina

$$\{\mathrm{vec}(\mathbf{Y}),\, (\mathbf{I}_2 \otimes \mathbf{X})\,\mathrm{vec}(\mathbb{B}),\, \boldsymbol{\Sigma} \otimes \mathbf{I}_n\}\,. \tag{8.285}$$

□

Harjoitustehtäviä

8.1. Olkoon

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \quad \hat{\mathbf{y}} = \mathbf{H}\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}}, \quad \hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{M}\mathbf{y}.$$

Osoita, että

(a) $\operatorname{cor}_{\mathrm{d}}(\hat{\mathbf{y}}, \hat{\boldsymbol{\varepsilon}}) = 0$,

(b) $\hat{\boldsymbol{\varepsilon}}'\mathbf{1} = \mathbf{y}'\mathbf{M}\mathbf{1} = 0$.

(c) Mitä voit sanoa $\mathbf{y}$:n ja $\hat{\boldsymbol{\varepsilon}}$:n välisestä korrelaatiosta?

8.2. Määritä sellaiset vektorit $\mathbf{x}$ ja $\mathbf{y}$, että

(a) $\cos(\mathbf{x}, \mathbf{y})$ on suuri mutta $\operatorname{cor}_{\mathrm{d}}(\mathbf{x}, \mathbf{y}) = 0$,

(b) $\cos(\mathbf{x}, \mathbf{y}) = 0$ mutta $\operatorname{cor}_{\mathrm{d}}(\mathbf{x}, \mathbf{y}) = 1$.

8.3. Olkoon $\mathbf{X} = (\mathbf{1} : \mathbf{x}_1 : \mathbf{x}_2)$, $\mathbf{H} = \mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{J} = \mathbf{P}_{\mathbf{1}}$. Osoita että

(a) $\mathbf{H} - \mathbf{J}$ on ortogonaaliprojektori,

(b) $(\mathbf{I} - \mathbf{J})\mathbf{H} = \mathbf{H} - \mathbf{J}$.

8.4. Olkoot $\mathbf{x}$, $\mathbf{y}$ ja $\mathbf{z}$ keskistettyjä ja 1:n pituisia muuttujavektoreita. Piirrä kuvio seuraavista tilanteista (jos mahdollisia):

(a) $r_{xy} = 0$, $r_{xz} = 0$, $r_{yz} = -1$,

(b) $r_{xy} = 1/\sqrt{2}$, $r_{xz} = -1/\sqrt{2}$, $r_{yz} = 1/\sqrt{2}$.

8.5. Olkoon $\mathbf{X} = (\mathbf{1} : \mathbf{X}_0)$, $\mathbf{H} = \mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{J} = \mathbf{P}_{\mathbf{1}}$, $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$. Osoita että tällöin

(a) $\|\mathbf{y} - \bar{\bar{\mathbf{y}}}\|^2 = \|(\mathbf{I} - \mathbf{J})\mathbf{y}\|^2 = \mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y} = \mathrm{SST} =$ kokonaisneliösumma,

(b) $\|\hat{\mathbf{y}} - \bar{\bar{\mathbf{y}}}\|^2 = \|(\mathbf{H} - \mathbf{J})\mathbf{y}\|^2 = \|(\mathbf{I} - \mathbf{J})\mathbf{H}\mathbf{y}\|^2 = \mathbf{y}'(\mathbf{H} - \mathbf{J})\mathbf{y} = \mathrm{SSR}$,

(c) $\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \|(\mathbf{I} - \mathbf{H})\mathbf{y}\|^2 = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'\mathbf{M}\mathbf{y} = \mathrm{SSE}$.

8.6. Todista seuraavat regressioanalyysin tulokset. Oletamme, että $\mathbf{1} \in \mathscr{C}(\mathbf{X})$.

(a) $\mathrm{SST} = \|\mathbf{y} - \bar{\bar{\mathbf{y}}}\|^2 = \|(\mathbf{I} - \mathbf{J})\mathbf{y}\|^2 = \mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y} = \mathbf{y}'\mathbf{y} - n\bar{y}^2 = t_{yy}$.

(b) $$\begin{aligned}\mathrm{SSR} &= \|\hat{\mathbf{y}} - \bar{\bar{\mathbf{y}}}\|^2 = \|(\mathbf{H} - \mathbf{J})\mathbf{y}\|^2 = \|(\mathbf{I} - \mathbf{J})\mathbf{H}\mathbf{y}\|^2 = \mathbf{y}'(\mathbf{H} - \mathbf{J})\mathbf{y} \\ &= \mathbf{y}'\mathbf{P}_{\mathbf{C}\mathbf{X}_0}\mathbf{y} = \mathbf{y}'\mathbf{P}_{\widetilde{\mathbf{X}}_0}\mathbf{y} = \mathbf{t}'_{\mathbf{x}y}\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}.\end{aligned}$$

(c) $$\begin{aligned}\mathrm{SSE} &= \|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \|(\mathbf{I} - \mathbf{H})\mathbf{y}\|^2 = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{y}'(\mathbf{C} - \mathbf{P}_{\mathbf{C}\mathbf{X}_0})\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} = t_{yy} - \mathbf{t}'_{\mathbf{x}y}\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y}.\end{aligned}$$

(d) $\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}$.

(e) (1) $\mathrm{r}(\mathbf{I} - \mathbf{J}) = n - 1, \quad s_y^2 = \mathrm{SST}/(n-1)$,

(2) $\mathrm{r}(\mathbf{H} - \mathbf{J}) = \mathrm{r}(\mathbf{X}) - 1$,

(3) $\mathrm{r}(\mathbf{I} - \mathbf{H}) = n - \mathrm{r}(\mathbf{X}), \quad \mathrm{MSE} = \mathrm{SSE}/[n - \mathrm{r}(\mathbf{X})] = \hat{\sigma}^2$.

(f) $\mathrm{SST} = \sum_{i=1}^n (y_i - \bar{y})^2, \;\; \mathrm{SSR} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \;\; \mathrm{SSE} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$.

8.7. Olkoon $\mathbf{X} = (\mathbf{1} : \mathbf{X}_0)$, $\mathrm{cor_d}(\mathbf{X}_0) = \mathbf{R_{xx}}$, $\mathrm{ssp}(\mathbf{X}_0) = \mathbf{T_{xx}}$. Todista (a)–(g):

(a)
$$\begin{aligned}\mathrm{r}(\mathbf{X}) &= 1 + \mathrm{r}(\mathbf{X}_0) - \dim \mathscr{C}(\mathbf{1}) \cap \mathscr{C}(\mathbf{X}_0) \\ &= 1 + \mathrm{r}\big[(\mathbf{I} - \mathbf{J})\mathbf{X}_0\big] \\ &= 1 + \mathrm{r}(\mathbf{T_{xx}}) = 1 + \mathrm{r}(\mathbf{R_{xx}}).\end{aligned}$$

(b) $|\mathbf{R_{xx}}| \neq 0 \iff \mathrm{r}(\mathbf{X}) = k + 1 \iff \mathrm{r}(\mathbf{X}_0) = k \;\&\; \mathbf{1} \in \mathscr{C}(\mathbf{X}_0) \iff |\mathbf{X}'\mathbf{X}| \neq 0$.

(c) $r_{ij} = 1$ jollakin $i \neq j$ implikoi $|\mathbf{R_{xx}}| = 0$ (mutta ei päinvastoin).

(d) (1) matriisin $\mathbf{X} = (\mathbf{1} : \mathbf{X}_0)$ sarakkeet ovat ortogonaaliset keskenään jos ja vain jos

(2) matriisi $\mathbf{X}_0$ on keskistetty ja $\mathrm{cor_d}(\mathbf{X}_0) = \mathbf{R_{xx}} = \mathbf{I}_k$.

(e) Tarkastellaan seuraavia väitteitä:

(1) matriisin $\mathbf{X}_0$ sarakkeet ovat ortogonaalisia,

(2) $\mathrm{cor_d}(\mathbf{X}_0) = \mathbf{I}_k$.

Vakuuttaudu että (1) ja (2) ovat yhtäpitäviä jos $\mathbf{X}_0$ on keskistetty.

(f) Laadi esimerkit seuraavista tilanteista

(1) $\cos(\mathbf{x}, \mathbf{y})$ on suuri mutta $\mathrm{cor_d}(\mathbf{x}, \mathbf{y}) = 0$,

(2) $\cos(\mathbf{x}, \mathbf{y}) = 0$ mutta $\mathrm{cor_d}(\mathbf{x}, \mathbf{y}) = 1$.

Onko mahdollista että $\cos(\mathbf{x}, \mathbf{y}) = 1$ mutta $\mathrm{cor_d}(\mathbf{x}, \mathbf{y}) = 0$?

(g) $\mathrm{cor_d}(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{y} \in \mathscr{C}(\mathbf{Cx})^\perp = \mathscr{C}(\mathbf{1} : \mathbf{x})^\perp \oplus \mathscr{C}(\mathbf{1})$ [eliminoi tapaukset $\mathbf{y} \in \mathscr{C}(\mathbf{1})$ ja/tai $\mathbf{y} \in \mathscr{C}(\mathbf{1})$]. Piirrä tilanteesta kuvio.

Luku 9

Sarakeavaruus

Tekijä on ilmeisesti täysin suunnittelemattomasti jakanut teoksensa neljäksi eri päivänlaskuksi eli jaksoksi. Tarkoittaako hän, että nämä erilliset luvut olisi luettava päivänlaskun aikaan, vai vihjaako hän kirjoittaneensa luvut eri päivinä, jää hämäräksi. Emme kuitenkaan erehtyne otaksuessamme, että häntä – aivan oikein – on kalvanut jonkinlainen salatajuinen tunne, että hänen edustamansa aikakausi eli lopullisen päivänlaskunsa aikaa ja että hänellä oli paljon laskuja maksamatta.

Mika Waltari (1949): *Neljä päivänlaskua.*

9.1 Sarakeavaruuden määritelmä

Olemme jo aikaisemmissa luvuissa käsitelleet monia sarakeavaruuden ominaisuuksia. Kertaamme tässä niitä ja teemme joitakin erityisesti tilastotieteen kannalta hyödyllisiä lisätarkasteluja.

Matriisin $\mathbf{A}_{n\times m}$ sarakevaruus määriteltiin seuraavasti:

$$\begin{aligned}\mathscr{C}(\mathbf{A}) &= \text{matriisin } \mathbf{A}_{n\times m} = (\mathbf{a}_1 : \ldots : \mathbf{a}_m) \text{ sarakeavaruus} \\ &= \{\, \mathbf{y} \in \mathbb{R}^n : \exists\, \mathbf{x} \in \mathbb{R}^m \text{ s.e. } \mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{a}_1 x_1 + \cdots + \mathbf{a}_m x_m \,\} \\ &= \{\, \mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^m \,\} \subset \mathbb{R}^n. \end{aligned} \tag{9.1}$$

Vektorin $\mathbf{a} \neq \mathbf{0}$ virittämä suora on on yksi esimerkki sarakeavaruudesta:

$$\mathscr{C}(\mathbf{a}) = \{\, \mathbf{y} \in \mathbb{R}^n : \exists\, \lambda \in \mathbb{R} \text{ s.e. } \mathbf{y} = \lambda\mathbf{a} \,\} = \{\lambda\mathbf{a}\}. \tag{9.2}$$

Jos $\mathbf{A}$:ssa on vain kaksi saraketta, niin:

$$\begin{aligned}\mathscr{C}(\mathbf{A}) &= \text{matriisin } \mathbf{A}_{n\times 2} = (\mathbf{a} : \mathbf{b}) \text{ sarakeavaruus} \\ &= \{\, \mathbf{y} \in \mathbb{R}^n : \exists\, \mathbf{x} \in \mathbb{R}^2 \text{ s.e. } \mathbf{y} = \mathbf{A}\mathbf{x} \,\} \\ &= \{\, \mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^2 \,\}. \end{aligned} \tag{9.3}$$

Kuten luvussa 2.2 (s. 76) totesimme, joukko $\mathscr{C}(\mathbf{a}:\mathbf{b})$ muodostaa geometrisesti joko

(i) origon kautta kulkevan tason, jolloin $\dim\mathscr{C}(\mathbf{A}) = 2 = \mathrm{r}(\mathbf{A})$, tai

(ii) origon kautta kulkevan suoran, jolloin $\dim\mathscr{C}(\mathbf{A}) = 1 = \mathrm{r}(\mathbf{A})$, tai

(iii) pelkän origon, jolloin $\dim\mathscr{C}(\mathbf{A}) = 0 = \mathrm{r}(\mathbf{A})$.

Tapauksessa (iii) sekä **a** että **b** ovat nollavektoreita.

On helppo havaita, että sarakeavaruus $\mathscr{C}(\mathbf{A}_{n\times m})$ on $\mathbb{R}^n$:n *aliavaruus*. Vektoriavaruuden $\mathcal{V}$ osajoukko $\mathcal{U}$ on $\mathcal{V}$:n aliavaruus, jos seuraavat ehdot toteutuvat:

AA1. $\mathbf{u},\mathbf{v}\in\mathcal{U} \implies \mathbf{u}+\mathbf{v}\in\mathcal{U}$,

AA2. $\lambda\in\mathbb{R},\ \mathbf{u}\in\mathcal{U} \implies \lambda\mathbf{u}\in\mathcal{U}$,

AA3. $\mathbf{0}\in\mathcal{U}$.

Yleisesti ottaen ei vektoriavaruuden elementtien tietenkään tarvitse olla pystyvektoreita vaikka tässä esityksessä useimmiten niin onkin. Vektoriavaruuden aksioomiin yleisessä tapauksessa emme tässä yhteydessä puutu. Sarakeavaruus $\mathscr{C}(\mathbf{A})$ on todellakin $\mathbb{R}^n$:n aliavaruus, sillä:

AA1. $\mathbf{u},\mathbf{v}\in\mathscr{C}(\mathbf{A}) \implies$ ....

AA2. $\lambda\in R,\ \mathbf{u}\in\mathscr{C}(\mathbf{A}) \implies$ ....

AA3. $\mathbf{0}\in\mathscr{C}(\mathbf{A})$, sillä ....

Aliavaruuksien $\mathcal{U}$ ja $\mathcal{V}$ summalla tarkoitetaan tunnetusti joukkoa

$$\mathcal{U}+\mathcal{V} = \{\,\mathbf{z} : \mathbf{z} = \mathbf{u}+\mathbf{v},\ \ \mathbf{u}\in\mathcal{U},\ \ \mathbf{v}\in\mathcal{V}\,\}. \tag{9.4}$$

Erityisesti huomaamme, että sarakevaruuksien tapauksessa

$$\begin{aligned}\mathscr{C}(\mathbf{A}_{n\times m})+\mathscr{C}(\mathbf{B}_{n\times k}) &= \{\,\mathbf{z}\in\mathbb{R}^n : \mathbf{z} = \mathbf{u}+\mathbf{v},\ \ \mathbf{u}\in\mathscr{C}(\mathbf{A}),\ \ \mathbf{v}\in\mathscr{C}(\mathbf{B})\,\}\\ &= \{\,\mathbf{z}\in\mathbb{R}^n : \mathbf{z} = \mathbf{A}\mathbf{a}+\mathbf{B}\mathbf{b},\ \ \mathbf{a}\in\mathbb{R}^m,\ \ \mathbf{b}\in\mathbb{R}^k\,\}.\end{aligned} \tag{9.5}$$

Toisaalta

$$\begin{aligned}\mathscr{C}(\mathbf{A}_{n\times m}:\mathbf{B}_{n\times k}) &= \left\{\,\mathbf{t}\in\mathbb{R}^n : \mathbf{t} = (\mathbf{A}:\mathbf{B})\binom{\mathbf{u}}{\mathbf{v}},\ \ \mathbf{u}\in\mathbb{R}^m,\ \ \mathbf{v}\in\mathbb{R}^k\,\right\}\\ &= \{\,\mathbf{t}\in\mathbb{R}^n : \mathbf{t} = \mathbf{A}\mathbf{u}+\mathbf{B}\mathbf{v},\ \ \mathbf{u}\in\mathbb{R}^m,\ \ \mathbf{v}\in\mathbb{R}^k\,\}.\end{aligned} \tag{9.6}$$

Täten olemme (täsmällisesti) osoittaneet (aivan ilmeisen) tuloksen:

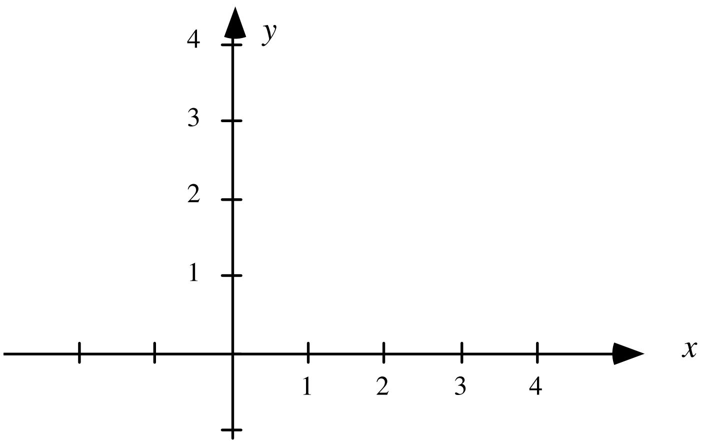
$$\mathscr{C}(\mathbf{A}:\mathbf{B}) = \mathscr{C}(\mathbf{A})+\mathscr{C}(\mathbf{B}), \tag{9.7}$$

jonka yhtenä erikoistapauksena on tietysti

$$\mathscr{C}(\mathbf{a}_1:\mathbf{a}_2:\ldots:\mathbf{a}_m) = \mathscr{C}(\mathbf{a}_1)+\mathscr{C}(\mathbf{a}_2)+\cdots+\mathscr{C}(\mathbf{a}_m). \tag{9.8}$$

Esimerkki 9.1. Piirrä seuraavien matriisien sarakeavaruudet kuvioon 9.1:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} 2 & 0 \\ 4 & 2 \end{pmatrix}, \ \mathbf{C} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix}, \ \mathbf{D} = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 2 & 4 & 6 & 2 \end{pmatrix}.$$



Kuvio 9.1. Esimerkin 9.1 sarakeavaruuksia.

Olemme jo aikaisemmissa luvuissa käsitelleet matriisin $\mathbf{A}_{n\times m}$ sarakeavaruuden *ortogonaalikomplementtia*, joka määritellään seuraavasti:

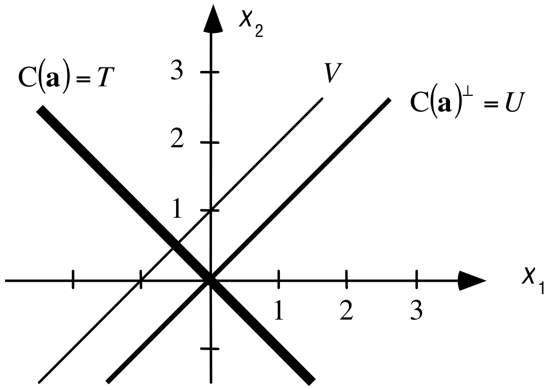
$$\begin{aligned} \mathscr{C}(\mathbf{A})^\perp &= \text{sarakeavaruuden } \mathscr{C}(\mathbf{A}_{n\times m}) \text{ ortogonaalikomplementti} \\ &= \text{niiden } \mathbb{R}^n\text{:n vektoreiden joukko jotka ovat koh-} \\ &\quad\ \text{tisuorassa jokaisen } \mathscr{C}(\mathbf{A})\text{:n vektorin kanssa} \\ &= \{\, \mathbf{u} \in \mathbb{R}^n : \mathbf{u}'\mathbf{v} = 0 \ \forall\ \mathbf{v} \in \mathscr{C}(\mathbf{A}) \,\} \\ &= \{\, \mathbf{u} \in \mathbb{R}^n : \mathbf{u}'\mathbf{A}\mathbf{t} = 0 \ \forall\ \mathbf{t} \in \mathbb{R}^m \,\} \\ &= \{\, \mathbf{u} \in \mathbb{R}^n : \mathbf{A}'\mathbf{u} = \mathbf{0} \,\} = \mathscr{N}(\mathbf{A}'). \end{aligned} \tag{9.9}$$

On paikallaan tähdentää sanan ”kohtisuora” merkitystä: kohtisuoruus eli ortogonaalisuus riippuu annetusta sisätulosta – ellei erikseen mainita, tarkoitamme nimenomaan tavanomaista euklidista sisätuloa.

Esimerkki 9.2. Olkoon $\mathbf{a} = \binom{-1}{1}$ ja tarkastellaan $\mathbb{R}^2$:n osajoukkoja

$$\mathscr{C}(\mathbf{a}), \quad \mathcal{U} = \{\, \mathbf{x} \in \mathbb{R}^2 : \mathbf{a}'\mathbf{x} = 0 \,\}, \quad \mathcal{V} = \{\, \mathbf{x} \in \mathbb{R}^2 : \mathbf{a}'\mathbf{x} = 1 \,\}. \tag{9.10}$$

Havaitsemme, että joukko $\mathcal{U}$ muodostuu suorasta, jonka yhtälö on $-x_1 + x_2 = 0$ eli $x_2 = x_1$. Tämä suora, joka siis kulkee origon kautta ja on $\mathscr{C}(\mathbf{a})$:n ortogonaalikomplementti $\mathscr{C}(\mathbf{a})^\perp$, muodostaa tietenkin aliavaruuden. Sen sijaan joukko $\mathcal{V}$ ei olekaan aliavaruus; se muodostuu suorasta, joka ei kulje origon kautta. Aliavaruuden on aina sisällettävä origo, joten $\mathcal{V}$ ei ole aliavaruus. □



Kuvio 9.2. Esimerkin 9.2 joukkoja: $\mathcal{U}$ ja $\mathcal{T}$ ovat aliavaruuksia mutta $\mathcal{V}$ ei ole.

9.2 Lineaarinen riippuvuus

Palautetaanpa mieleen lineaarisen riippumattomuuden määritelmä:

Matriisin $\mathbf{A}_{n\times m}$ sarakkeet $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^n$ ovat *lineaarisesti riippumattomat* eli *vapaat*, jos

$$\lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2 + \cdots + \lambda_m\mathbf{a}_m = \mathbf{0} \iff \lambda_1 = \lambda_2 = \cdots = \lambda_m = 0, \tag{9.11}$$

eli matriisimerkinnöin:

$$\mathbf{A}\boldsymbol{\lambda} = \mathbf{0} \iff \boldsymbol{\lambda} = \mathbf{0}. \tag{9.12}$$

Nolla-avaruuden avulla esitettynä:

$$\mathbf{A}\text{:n sarakkeet vapaat} \iff \mathscr{N}(\mathbf{A}) = \{\mathbf{0}\}, \tag{9.13}$$

missä

$$\mathscr{N}(\mathbf{A}) = \{\, \mathbf{x} \in \mathbb{R}^m : \mathbf{A}\mathbf{x} = \mathbf{0} \,\} = \text{matriisin } \mathbf{A}_{n\times m} \text{ ydin} \subset \mathbb{R}^m. \tag{9.14}$$

Jos $\mathbf{A}$:n sarakkeet ovat vapaat, sanomme että $\mathbf{A}$:lla on *täysi sarakeaste.*

Vastaavasti määritellään, että vektorit $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^n$ ovat *lineaarisesti riippuvat* eli *sidotut*, jos ne eivät ole vapaat eli

$$\text{on olemassa sellainen vektori } \boldsymbol{\lambda} \neq \mathbf{0} \text{ että } \mathbf{A}\boldsymbol{\lambda} = \mathbf{0}, \tag{9.15}$$

ts.

$$\mathbf{A}\text{:n sarakkeet sidotut} \iff \mathscr{N}(\mathbf{A}) \neq \{\mathbf{0}\}. \tag{9.16}$$

Havaittakoon, että em. käsitteet liittyvät nimenomaan vektori*joukkoihin*. Voimme sanoa yhtäpitävästi:

”joukko $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ on vapaa” tai ”vektorit $\mathbf{a}_1$, $\mathbf{a}_2$, $\dots$, $\mathbf{a}_m$ ovat vapaat”. (9.17)

Esimerkki 9.3. Tarkastellaan esimerkkinä matriisia

$$\mathbf{X} = (\mathbf{1} : \mathbf{x}_1 : \mathbf{x}_2 : \mathbf{x}_3) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{9.18}$$

Koska $\mathbf{X}\boldsymbol{\lambda} = \mathbf{0}$, kun

$$\boldsymbol{\lambda} = \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \tag{9.19}$$

ovat $\mathbf{X}$:n sarakkeet lineaarisesti riippuvia. Nythän $\mathbf{X}$:n ensimmäinen sarake on muiden sarakkeiden summa. Ylipäänsä jos matriisin $\mathbf{A}$ jokin sarake on muiden lineaarikombinaatio, ovat $\mathbf{A}$:n sarakkeet sidotut. Jos tarkastelemme $\mathbf{X}$:n kolmea ensimmäistä saraketta ja merkitsemme $\mathbf{B} = (\mathbf{1} : \mathbf{x}_1 : \mathbf{x}_2)$, niin yhtälöstä $\mathbf{Bt} = \mathbf{0}$, kun $\mathbf{t} = (t_0, t_1, t_2)'$, saamme yhtälöryhmän

$$\begin{aligned} t_0 + t_1 &= 0, \\ t_0 + t_2 &= 0, \\ t_0 &= 0. \end{aligned} \tag{9.20}$$

Tämän yhtälöryhmän ainoa ratkaisu on $\mathbf{t} = \mathbf{0}$, joten joukko $\{\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2\}$ on vapaa. Helposti näemme, että mitkä tahansa kolme $\mathbf{X}$:n saraketta ovat vapaita. Matriisin $\mathbf{X}$ aste – sen vapaiden sarakkeiden (suurin) lukumäärä – on 3. □

Lukijan on hyvä vakuuttautua seuraavista tuloksista:

LR1. Joukko $\{\mathbf{0}\}$ on sidottu.

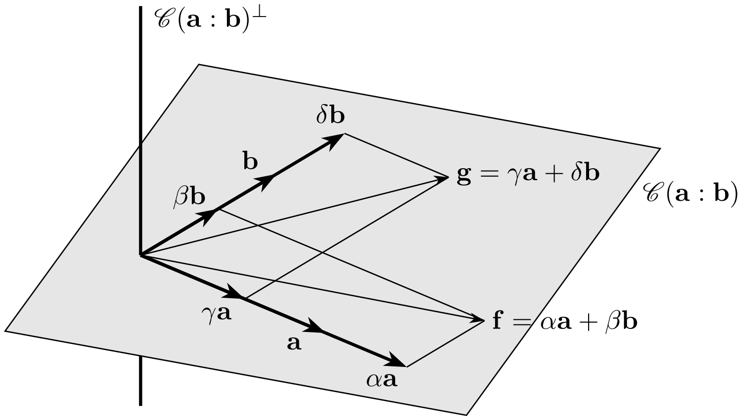
LR2. Vektorijoukko, johon kuuluu nollavektori, on sidottu.

LR3. Vektorijoukko on sidottu $\iff$ jokin vektori on toisten lineaarikombinaatio.

LR4. Nollasta poikkeavat ortogonaaliset vektorit $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ ovat vapaat.

Esimerkki 9.4. Kuvion 9.3 tilanteessa tarkastelemme $\mathbb{R}^3$:n vektoreita $\mathbf{a}$ ja $\mathbf{b}$, jotka eivät ole toistensa monikertoja eikä kumpikaan niistä ole nollavektori, joten $\mathbf{a}$ ja $\mathbf{b}$ ovat lineaarisesti riippumattomia. Kuvioon on merkitty myös $\mathscr{C}(\mathbf{a} : \mathbf{b})$:n ortogonaalikomplementti $\mathscr{C}(\mathbf{a} : \mathbf{b})^{\perp}$, eli niiden vektoreiden joukko jotka ovat kohtisuorassa $\mathscr{C}(\mathbf{a} : \mathbf{b})$:tä vastaan. On ilmeistä, että $\mathbb{R}^3$ voidaan esittää summana

$$\mathbb{R}^3 = \mathscr{C}(\mathbf{a} : \mathbf{b}) + \mathscr{C}(\mathbf{a} : \mathbf{b})^{\perp}. \tag{9.21}$$



Kuvio 9.3. Vektorit $\mathbf{f} = \alpha\mathbf{a} + \beta\mathbf{b}$ ja $\mathbf{g} = \gamma\mathbf{a} + \delta\mathbf{b}$ geometrisesti, $\alpha > 1$, $\beta < 1$, $\gamma < 1$, $\delta > 1$. (Sama kuin kuvio 2.7, s. 80).

9.3 Matriisin aste

Matriisin aste, $\mathrm{r}(\mathbf{A})$, määritellään seuraavasti:

$$\mathrm{r}(\mathbf{A}) = \mathbf{A}\text{:n vapaiden sarakkeiden maksimaalinen lukumäärä.} \tag{9.22}$$

On osoitettavissa, että $\mathrm{r}(\mathbf{A})$ on itse asiassa sama kuin $\mathbf{A}$:n vapaiden vaakarivien lukumäärä eli $\mathrm{r}(\mathbf{A}') = \mathrm{r}(\mathbf{A})$; ks. esim. Bapat (2000, s. 9). Sarakeavaruuden $\mathscr{C}(\mathbf{A})$ dimensio on luonnollisesti

$$\dim \mathscr{C}(\mathbf{A}) = \mathrm{r}(\mathbf{A}). \tag{9.23}$$

Vektoriavaruuden dimensiohan määritellään sen *kantavektorien* lukumääränä: vektorijoukko $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ on vektoriavaruuden $\mathcal{V}$ *kanta*, jos

KV1. $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ on vapaa ja

KV2. $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ virittää $\mathcal{V}$:n eli jokainen $\mathcal{V}$:n vektori voidaan esittää $\mathbf{v}_i$:tten lineaarikombinaationa.

Esimerkki 9.5. Määritä seuraavien matriisien asteet:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} 2 & 0 \\ 4 & 2 \end{pmatrix}, \ \mathbf{C} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix}, \ \mathbf{D} = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 2 & 4 & 6 & 2 \end{pmatrix},$$

$$\mathbf{E} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \ \mathbf{F} = \begin{pmatrix} 1 & 1 \\ 1 & 1.000000001 \end{pmatrix}, \ \mathbf{A} + \mathbf{B}, \ \mathbf{A} - \mathbf{B}, \ \mathbf{E}^2.$$

$\square$

Seuraavat lineaarialgebran tulokset ovat varmaankin lukijalle tuttuja:

KA1. Jokaisella vektoriavaruudella on kanta.

KA2. Jokaisella $\mathcal{V}$:n vektorilla on yksikäsitteinen esitys tietyn kannan suhteen.

KA3. Olkoon $\mathcal{U}$ vektoriavaruuden $\mathcal{V}$ aliavaruus. Tällöin $\dim \mathcal{U} \leq \dim \mathcal{V}$ ja $\dim \mathcal{U} = \dim \mathcal{V}$ vain jos $\mathcal{U} = \mathcal{V}$.

KA4. Jos $\mathcal{U}$ ja $\mathcal{V}$ ovat aliavaruuksia, niin

$$\dim(\mathcal{U} + \mathcal{V}) = \dim \mathcal{U} + \dim \mathcal{V} - \dim \mathcal{U} \cap \mathcal{V}. \tag{9.24}$$

Sarakeavaruuksien tapauksessa *KA3* merkitsee seuraavaa:

$$\mathscr{C}(\mathbf{U}) \subset \mathscr{C}(\mathbf{V}) \implies \mathrm{r}(\mathbf{U}) \leq \mathrm{r}(\mathbf{V}). \tag{9.25}$$

Koska $\mathbf{ABx} := \mathbf{Az} \in \mathscr{C}(\mathbf{A})$, on aina voimassa

$$\mathscr{C}(\mathbf{AB}) \subset \mathscr{C}(\mathbf{A}), \tag{9.26}$$

joten (9.25):n perusteella saamme epäyhtälön

$$\mathrm{r}(\mathbf{AB}) \leq \mathrm{r}(\mathbf{A}). \tag{9.27}$$

Toisaalta koska $\mathscr{C}(\mathbf{B}'\mathbf{A}') \subset \mathscr{C}(\mathbf{B}')$, on myös $\mathrm{r}(\mathbf{B}'\mathbf{A}') \leq \mathrm{r}(\mathbf{B}')$ eli $\mathrm{r}[(\mathbf{AB})'] = \mathrm{r}(\mathbf{AB}) \leq \mathrm{r}(\mathbf{B}') = \mathrm{r}(\mathbf{B})$. Täten

> matriisitulon aste pienenee tai pysyy ennallaan, kun tuloon lisätään tekijöitä:

$$\boxed{\mathrm{r}(\mathbf{AB}) \leq \mathrm{r}(\mathbf{A}), \quad \mathrm{r}(\mathbf{AB}) \leq \mathrm{r}(\mathbf{B}). \tag{9.28}}$$

Seuraavat tulokset osoittautuvat varsin hyödyllisiksi:

A1. $\mathbf{B}$:n vaakarivit vapaat $\implies \mathscr{C}(\mathbf{AB}) = \mathscr{C}(\mathbf{A})$ & $\mathrm{r}(\mathbf{AB}) = \mathrm{r}(\mathbf{A})$,

A2. $\mathbf{C}$:n sarakkeet vapaat $\implies \mathrm{r}(\mathbf{CA}) = \mathrm{r}(\mathbf{A})$,

A3. $\mathbf{C}$:n sarakkeet vapaat $\implies \mathscr{N}(\mathbf{CA}) = \mathscr{N}(\mathbf{A})$,

A4. $\mathscr{N}(\mathbf{A}) \subset \mathscr{N}(\mathbf{BA})$,

A5. $\mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{B}) \iff \exists\, \mathbf{F}\colon \mathbf{A} = \mathbf{BF}$,

A6. $\mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{B}) \iff \mathbf{P_B A} = \mathbf{A}$,

A7. $\mathscr{C}(\mathbf{A}_{n\times m}) = \mathscr{C}(\mathbf{B}_{n\times p}) \iff \exists\, \mathbf{F}\colon \mathbf{A}_{n\times m} = \mathbf{B}_{n\times p}\mathbf{F}_{p\times m}$, missä $\mathscr{C}(\mathbf{B}') \cap \mathscr{C}(\mathbf{F}^{\perp}) = \{\mathbf{0}\}$, *A8.* $\mathscr{C}(\mathbf{A}) = \mathscr{C}(\mathbf{B}) \iff \mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{B})$ & $\mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{B})$,

A9. $\mathscr{C}(\mathbf{F}) = \mathscr{C}(\mathbf{G}) \implies \mathscr{C}(\mathbf{AF}) = \mathscr{C}(\mathbf{AG})$,

A10. $\mathscr{C}(\mathbf{F}) \subset \mathscr{C}(\mathbf{G}) \implies \mathscr{C}(\mathbf{AF}) \subset \mathscr{C}(\mathbf{AG})$,

A11. $\mathrm{r}(\mathbf{A}+\mathbf{B}) \le \mathrm{r}(\mathbf{A} : \mathbf{B}) \le \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B})$,

A12. $\mathrm{r}(\mathbf{A}+\mathbf{B}) \le \mathrm{r}\begin{pmatrix}\mathbf{A}\\ \mathbf{B}\end{pmatrix} \le \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B})$,

A13. $\mathrm{r}(\mathbf{A}+\mathbf{B}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) \Leftrightarrow \dim \mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}) = 0 = \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B}')$.

Tuloksen *A1* todistamiseksi olettakaamme, että $\mathbf{B}$:n vaakarivit ovat vapaat. Tällöin $(\mathbf{BB}')^{-1}$ on olemassa ja

$$\mathrm{r}(\mathbf{A}) \ge \mathrm{r}(\mathbf{AB}) \ge \mathrm{r}[\mathbf{ABB}'(\mathbf{BB}')^{-1}] = \mathrm{r}(\mathbf{A}). \tag{9.29}$$

Koska (9.29):ssa saamme epäyhtälöketjun alkuun ja loppuun $\mathrm{r}(\mathbf{A})$:n, muuttuvat kaikki epäyhtälöt tietenkin yhtälöiksi, joten $\mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{AB})$. Ehdot $\mathscr{C}(\mathbf{AB}) \subset \mathscr{C}(\mathbf{A})$ ja $\mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{AB})$ yhdessä merkitsevät, että $\mathscr{C}(\mathbf{AB}) = \mathscr{C}(\mathbf{A})$.

Epäyhtälöketjun (9.29) mukainen todistustapa on monissa asteeseen liittyvissä tarkasteluissa varsin kätevä. Epämatemaattisesti sanoen

saamme $\mathrm{r}(\mathbf{AB})$:n ”puristetuksi $\mathrm{r}(\mathbf{A})$:n ja $\mathrm{r}(\mathbf{A})$:n välille”.

Todistamme vielä *A11*:n, mutta muut kohdat jätetään harjoitustehtäväksi. Esimerkissä 6.3 havaitsimme, että $n \times m$-matriiseille $\mathbf{A}$ ja $\mathbf{B}$ on voimassa

$$\mathbf{A}+\mathbf{B} = (\mathbf{A} : \mathbf{B})\begin{pmatrix}\mathbf{I}_m\\ \mathbf{I}_m\end{pmatrix}, \quad (\mathbf{A} : \mathbf{B}) = (\mathbf{I}_n : \mathbf{I}_n)\begin{pmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}\end{pmatrix}. \tag{9.30}$$

Koska tulon astesäännön (9.28) perusteella tulon aste on enintään tulon minkä hyvänsä tekijän aste, saamme

$$\mathrm{r}(\mathbf{A}+\mathbf{B}) \le \mathrm{r}(\mathbf{A} : \mathbf{B}), \tag{9.31a}$$

$$\mathrm{r}(\mathbf{A} : \mathbf{B}) \le \mathrm{r}\begin{pmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}\end{pmatrix} = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}), \tag{9.31b}$$

eli todellakin *A11* on voimassa:

$$\boxed{\mathrm{r}(\mathbf{A}+\mathbf{B}) \le \mathrm{r}(\mathbf{A} : \mathbf{B}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}). \qquad (A11)}$$

Lausekkeen (9.31b) viimeinen yhtäsuuruusmerkki ts. yhtälö

$$\mathrm{r}\begin{pmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}\end{pmatrix} = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}), \tag{9.32}$$

saa oikeutuksensa säännöstä

$$\mathrm{r}(\mathbf{A} : \mathbf{B}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) - \dim \mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}), \tag{9.33}$$

sillä

$$\mathscr{C}\begin{pmatrix}\mathbf{A}\\ \mathbf{0}\end{pmatrix} \cap \mathscr{C}\begin{pmatrix}\mathbf{0}\\ \mathbf{B}\end{pmatrix} = \{\mathbf{0}\},\ \mathrm{r}\begin{pmatrix}\mathbf{A}\\ \mathbf{0}\end{pmatrix} = \mathrm{r}(\mathbf{A}),\ \mathrm{r}\begin{pmatrix}\mathbf{0}\\ \mathbf{B}\end{pmatrix} = \mathrm{r}(\mathbf{B}). \tag{9.34}$$

Toisaalta on helppo havaita, että

$$\mathscr{C}(\mathbf{A} + \mathbf{B}) \subset \mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{B}), \tag{9.35}$$

mistä myös voitaisiin päätellä (9.31a):n olevan voimassa; tulos (9.35) perustuu seuraavaan päättelyyn:

$$\mathbf{y} \in \mathscr{C}(\mathbf{A} + \mathbf{B}) \implies \exists\, \mathbf{x} \colon \mathbf{y} = (\mathbf{A} + \mathbf{B})\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{x} \in \mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{B}). \tag{9.36}$$

9.4 Keskistäjämatriisin aste

Tarkastellaan $n \times n$-ortogonaaliprojektorien $\mathbf{J}$ ja $\mathbf{C} = \mathbf{I} - \mathbf{J}$ (= keskistäjämatriisi) asteita ja niiden sarakeavaruuksien kantoja. Koska $\mathscr{C}(\mathbf{J}) = \mathscr{C}(\mathbf{1})$, on selvää, että $\mathbf{J}$:n aste on 1. Keskistäjämatriisi $\mathbf{C}$ on

$$\mathbf{C} = \mathbf{I} - \mathbf{J} = \begin{pmatrix} 1 - 1/n & -1/n & \dots & -1/n \\ -1/n & 1 - 1/n & \dots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \dots & 1 - 1/n \end{pmatrix}. \tag{9.37}$$

Keskistäjämatriisin aste onkin jo mutkikkaampi määrittää. Koska

$$\mathbf{C}\mathbf{1} = (\mathbf{I} - \mathbf{J})\mathbf{1} = \mathbf{0}, \tag{9.38}$$

ovat $\mathbf{C}$:n sarakkeet lineaarisesti riippuvia ja täten $\mathbf{C}$:n aste on enintään $n - 1$. Osoittautuukin, että $\mathbf{C}$:n aste on juuri $n - 1$.

Kun $n = 2$ ja $n = 3$ saamme keskistäjämatriiseiksi

$$\mathbf{C}_2 = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix},\ \mathbf{C}_3 = \begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}. \tag{9.39}$$

On suoraviivaista päätellä, että $\mathrm{r}(\mathbf{C}_2) = 1$ ja $\mathrm{r}(\mathbf{C}_3) = 2$. Aliavaruuden $\mathscr{C}(\mathbf{C}_2)$ kannaksi voimme valita esim. ensimmäisen sarakkeen $\mathbf{C}_2$:sta ja vastaavasti $\mathscr{C}(\mathbf{C}_3)$:n kannaksi kaksi ensimmäistä saraketta $\mathbf{C}_3$:sta.

Luvussa 2.3 (s. 86) osoitimme, että

$$\mathscr{N}(\mathbf{A}') = \mathscr{N}(\mathbf{A}\mathbf{A}'), \tag{9.40}$$

mistä tietysti seuraa, että

$$\mathscr{N}(\mathbf{1}') = \mathscr{N}(\mathbf{1}\mathbf{1}') = \mathscr{N}\left(\tfrac{1}{n}\mathbf{1}\mathbf{1}'\right) = \mathscr{N}(\mathbf{J}). \tag{9.41}$$

Koska

$$\mathbf{u} \in \mathscr{N}(\mathbf{1}') \iff \mathbf{1}'\mathbf{u} = 0 \iff \mathbf{u} \in \mathscr{C}(\mathbf{1})^{\perp}, \tag{9.42}$$

on voimassa yhtälö

$$\mathscr{N}(\mathbf{1}') = \mathscr{N}(\mathbf{J}) = \mathscr{C}(\mathbf{1})^{\perp}. \tag{9.43}$$

Todistamme nyt seuraavassa tuloksen

$$\mathscr{C}(\mathbf{I} - \mathbf{J}) = \mathscr{N}(\mathbf{J}), \tag{9.44a}$$

ts.

$$\mathscr{C}(\mathbf{I} - \mathbf{P_1}) = \mathscr{N}(\mathbf{1}') = \mathscr{C}(\mathbf{1})^{\perp}, \tag{9.44b}$$

joten myös

$$\mathscr{C}(\mathbf{C})^{\perp} = \mathscr{C}(\mathbf{1})\,. \tag{9.44c}$$

Oletetaan ensin, että $\mathbf{u} \in \mathscr{N}(\mathbf{1}') = \mathscr{N}(\mathbf{J})$. Tällöin siis $\mathbf{1}'\mathbf{u} = 0$, mistä välittömästi seuraa, että $\mathbf{J}\mathbf{u} = \mathbf{0}$, joten $(\mathbf{I} - \mathbf{J})\mathbf{u} = \mathbf{u}$, eli $\mathbf{u} \in \mathscr{C}(\mathbf{I} - \mathbf{J})$. Näin on osoitettu että $\mathscr{N}(\mathbf{1}') \subset \mathscr{C}(\mathbf{I} - \mathbf{J})$. Vielä on näytettävä, että $\mathscr{C}(\mathbf{I} - \mathbf{J}) \subset \mathscr{N}(\mathbf{J}) = \mathscr{N}(\mathbf{1}')$.

Oletetaan, että $\mathbf{u} \in \mathscr{C}(\mathbf{I} - \mathbf{J})$. Tällöin on olemassa sellainen vektori $\mathbf{a}$, että

$$\mathbf{u} = (\mathbf{I} - \mathbf{J})\mathbf{a}. \tag{9.45}$$

Kerrotaan (9.45) vasemmalta $\mathbf{J}$:llä, jolloin $\mathbf{J}$:n idempotenttisuuden perusteella

$$\mathbf{J}\mathbf{u} = (\mathbf{J} - \mathbf{J}^2)\mathbf{a} = \mathbf{0}, \tag{9.46}$$

eli $\mathbf{u} \in \mathscr{N}(\mathbf{J})$, joten todellakin $\mathscr{C}(\mathbf{I} - \mathbf{J}) \subset \mathscr{N}(\mathbf{J}) = \mathscr{N}(\mathbf{1}')$. Näin olemme (juurta jaksain) osoittaneet, että (9.44) on voimassa.

On osoitettavissa, että (9.44):n yleistyksenä pätee

$$\mathscr{C}(\mathbf{I} - \mathbf{P_A}) = \mathscr{N}(\mathbf{A}') = \mathscr{C}(\mathbf{A})^{\perp}. \tag{9.47}$$

Keskistäjämatriisin $\mathbf{C}$ sarakeavaruus $\mathscr{C}(\mathbf{C})$ muodostuu siis kaikista sellaisista vektoreista, joiden elementtien summa on nolla:

$$\mathscr{C}(\mathbf{C}) = \{\, \mathbf{u} \in \mathbb{R}^n : \mathbf{1}'\mathbf{u} = 0 \,\} = \mathscr{N}(\mathbf{1}') = \mathscr{C}(\mathbf{1})^{\perp}, \tag{9.48}$$

eli sellaisista vektoreista, jotka ovat kohtisuorassa vektorin $\mathbf{1}$ kanssa. Kuinka monta tällaisen ehdon toteuttavaa lineaarisesti riippumatonta vektoria on olemassa? Vastaus on $\dim \mathscr{C}(\mathbf{C}) = \mathrm{r}(\mathbf{C})$, jonka voi päätellä olevan $n-1$:

$$\mathrm{r}(\mathbf{C}) = n - 1. \tag{9.49}$$

Esimerkiksi seuraavan $n\times(n-1)$-matriisin $\mathbf{L}$ sarakkeet muodostavat $\mathscr{C}(\mathbf{C})$:n kannan (ks. esimerkki 4.5, s. 141):

$$\mathbf{L}_{n\times(n-1)} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & -1 \end{pmatrix}. \tag{9.50}$$

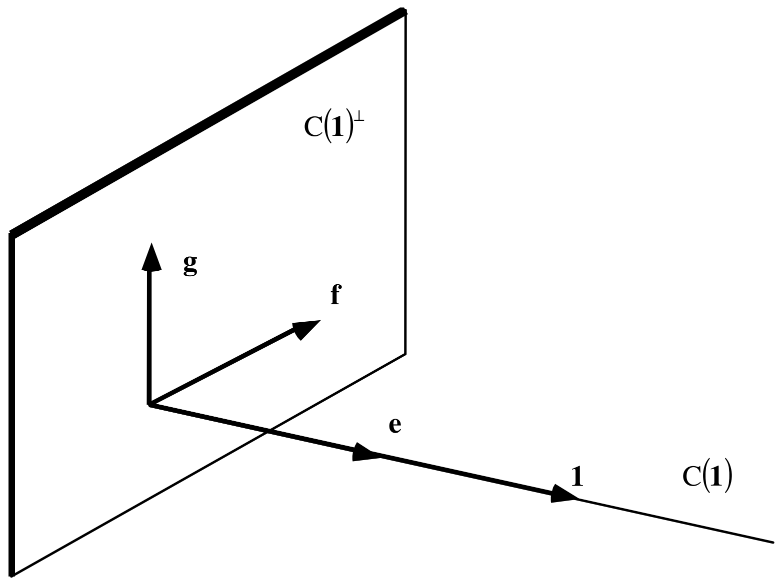
Tarkastellaan vielä tilannetta erityisesti $\mathbb{R}^2$:ssa ja $\mathbb{R}^3$:ssa. Merkitään

$$\mathbf{T} = (\mathbf{t}_1 : \mathbf{t}_2) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{9.51a}$$

$$\mathbf{U} = (\mathbf{u}_1 : \mathbf{u}_2 : \mathbf{u}_3) = \begin{pmatrix} 1/a & 1/b & 1/c \\ 1/a & -1/b & 1/c \\ 1/a & 0 & -2/c \end{pmatrix}, \tag{9.51b}$$

missä $a = \sqrt{3}$, $b = \sqrt{2}$, $c = \sqrt{6}$. Tällöin $\mathbf{T}$:n sarakkeet ovat ortonormaaleja ja samoin $\mathbf{U}$:n. Lisäksi

- kun $n = 2$:
  (1) $\mathbf{t}_1$ muodostaa aliavaruuden $\mathscr{C}(\mathbf{J}) = \mathscr{C}(\mathbf{1})$ ortonormaalin kannan,
  (2) $\mathbf{t}_2$ muodostaa aliavaruuden $\mathscr{C}(\mathbf{C}) = \mathscr{C}(\mathbf{I}-\mathbf{J}) = \mathscr{C}(\mathbf{1})^{\perp}$ ortonormaalin kannan,
  (3) $\mathscr{C}(\mathbf{t}_1 : \mathbf{t}_2) = \mathbb{R}^2$,
- kun $n = 3$:
  (1) $\mathbf{u}_1$ muodostaa aliavaruuden $\mathscr{C}(\mathbf{J}) = \mathscr{C}(\mathbf{1})$ ortonormaalin kannan,
  (2) $\{\mathbf{u}_2, \mathbf{u}_3\}$ muodostaa aliavaruuden $\mathscr{C}(\mathbf{C}) = \mathscr{C}(\mathbf{I}-\mathbf{J}) = \mathscr{C}(\mathbf{1})^{\perp}$ ortonormaalin kannan,
  (3) $\mathscr{C}(\mathbf{u}_1 : \mathbf{u}_2 : \mathbf{u}_3) = \mathbb{R}^3$.



Kuvio 9.4. Vektori $\mathbf{e}$ muodostaaa $\mathscr{C}(\mathbf{1})$:n kannan, vektorit $\mathbf{g}$ ja $\mathbf{f}$ muodostavat $\mathscr{C}(\mathbf{C})$:n eli $\mathscr{C}(\mathbf{1})^{\perp}$:n kannan.

9.5 Ortogonaalikomplementti

Nolla-avaruuden $\mathscr{N}(\mathbf{A}')$ ja ortogonaalikomplementin $\mathscr{C}(\mathbf{A})^{\perp}$ välillä on tärkeä yhteys, ks. (2.64) (s. 90):

$$\boxed{\mathscr{C}(\mathbf{A})^{\perp} = \mathscr{N}(\mathbf{A}'). \qquad (9.52)}$$

Muistisääntönä "nastan" vaikutuksesta $\mathscr{C}(\mathbf{A})$:han voi pitää seuraavaa:

$$\mathscr{C} \text{ muuttuu } \mathscr{N}\text{:ksi ja } \mathbf{A} \text{ muuttuu } \mathbf{A}'\text{:ksi.} \tag{9.53}$$

Erityisesti voimme sopia seuraavasta merkinnästä:

$$\mathbf{A}^{\perp} = \text{matriisi, jonka sarakevaruus on } \mathscr{C}(\mathbf{A})^{\perp} = \mathscr{N}(\mathbf{A}'). \tag{9.54}$$

Koska $\mathbf{A}^{\perp}$ ei välttämättä ole yksikäsitteinen, voisi olla perusteltua käyttää joukkomerkintää

$$\{\mathbf{A}^{\perp}\}, \tag{9.55}$$

eli jos $\mathbf{Z} \in \{\mathbf{A}^{\perp}\}$, niin silloin $\mathbf{Z}$ on matriisi, jolla on ominaisuus $\mathscr{C}(\mathbf{Z}) = \mathscr{N}(\mathbf{A}')$ ts.

$$\mathbf{Z}_{n\times q} \in \{\mathbf{A}^{\perp}_{n\times p}\} \tag{9.56a}$$

$$\iff$$

$$\mathbf{A}'\mathbf{Z} = \mathbf{0} \text{ ja } \mathrm{r}(\mathbf{Z}) = \dim \mathscr{C}(\mathbf{A}^{\perp}) = n - \mathrm{r}(\mathbf{A}). \tag{9.56b}$$

Huom: yhtälöä $\dim \mathscr{C}(\mathbf{A}^{\perp}) = n - \mathrm{r}(\mathbf{A})$ on tarkasteltu sivulla 316.

Olemme jo aikaisemminkin havainneet, että esimerkiksi

$$\mathscr{C}(\mathbf{1})^{\perp} = \mathscr{C}(\mathbf{I} - \mathbf{J}) = \mathscr{C}(\mathbf{C}) = \mathscr{N}(\mathbf{1}'), \tag{9.57}$$

joten $\mathbf{I} - \mathbf{J} \in \{\mathbf{1}^{\perp}\}$. Samoin jo luvussa 2.3 (s. 90) osoitimme, että

$$\mathscr{N}(\mathbf{A}) = \mathscr{N}(\mathbf{A}'\mathbf{A}). \tag{9.58}$$

Ottamalla (9.58):sta "nastat" puolittain ja "pudottamalla nastat alas" saamme tärkeän (jo aikaisemmin monesti käytetyn) tuloksen

$$\mathscr{C}(\mathbf{A}') = \mathscr{C}(\mathbf{A}'\mathbf{A}), \tag{9.59a}$$

ja tietenkin

$$\mathscr{C}(\mathbf{A}) = \mathscr{C}(\mathbf{A}\mathbf{A}'). \tag{9.59b}$$

Matriisin asteelle em. tulokset merkitsevät seuraavaa:

$$\mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{A}\mathbf{A}') = \mathrm{r}(\mathbf{A}'\mathbf{A}). \tag{9.60}$$

Seuraavat tulokset ovat ilmeisiä seurauksia (9.59):stä:

C1. $\mathscr{C}(\mathbf{AB}) = \mathscr{C}(\mathbf{ABB}') = \mathscr{C}(\mathbf{ABB}'\mathbf{A}')$,

C2. $\mathrm{r}(\mathbf{AB}) = \mathrm{r}(\mathbf{ABB}') = \mathrm{r}(\mathbf{ABB}'\mathbf{A}')$,

C3. $\mathscr{C}(\mathbf{AVA}') = \mathscr{C}(\mathbf{AV})$ kun $\mathbf{V}$ on ei-negatiivisesti definiitti,

C4. $\mathrm{r}(\mathbf{AVA}') = \mathrm{r}(\mathbf{AV}) = \mathrm{r}(\mathbf{VA}')$ kun $\mathbf{V}$ on ei-negatiivisesti definiitti,

C5. $\mathscr{C}(\mathbf{AVA}') = \mathscr{C}(\mathbf{A})$ kun $\mathbf{V}$ on positiivisesti definiitti,

C6. $\mathrm{r}(\mathbf{AVA}') = \mathrm{r}(\mathbf{A})$ kun $\mathbf{V}$ on positiivisesti definiitti.

Esimerkiksi *C2* on välitön seuraus seuraavasta:

$$\mathrm{r}(\mathbf{AB}) \geq \mathrm{r}(\mathbf{ABB}') \geq \mathrm{r}(\mathbf{ABB}'\mathbf{A}') = \mathrm{r}[\mathbf{AB}(\mathbf{AB})'] = \mathrm{r}(\mathbf{AB}), \tag{9.61}$$

ja *C4* on voimassa koska (käyttämättä tulosta *C2*)

$$\begin{aligned} \mathrm{r}(\mathbf{AV}) &\geq \mathrm{r}(\mathbf{AVA}') = \mathrm{r}(\mathbf{ALL}'\mathbf{A}') = \mathrm{r}(\mathbf{AL}) \\ &\geq \mathrm{r}(\mathbf{ALL}') = \mathrm{r}(\mathbf{AV}), \end{aligned} \tag{9.62}$$

missä $\mathbf{V} = \mathbf{LL}'$ (tälläinen $\mathbf{L}$ on olemassa koska $\mathbf{V}$ on ei-negatiivisesti definiitti). Muiden kohtien todistukset jätetään harjoitustehtäviksi.

Todistamme nyt seuraavan tuloksen:

C7. $\mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{B}) \iff \mathscr{C}(\mathbf{B})^\perp \subset \mathscr{C}(\mathbf{A})^\perp \qquad [\iff \mathscr{N}(\mathbf{B}') \subset \mathscr{N}(\mathbf{A}')]$.

Tulos *C7* perustuu seuraavaan päättelyyn:

$$\begin{aligned}\mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{B}) &\iff (\mathbf{I}-\mathbf{P_B})\mathbf{A} = \mathbf{0} \iff \mathbf{A}'(\mathbf{I}-\mathbf{P_B}) = \mathbf{0} \\ &\iff \mathscr{C}(\mathbf{I}-\mathbf{P_B}) \subset \mathscr{N}(\mathbf{A}') \iff \mathscr{N}(\mathbf{B}') \subset \mathscr{N}(\mathbf{A}'). \end{aligned} \tag{9.63}$$

Vielä pari tulosta:

C8. $\mathscr{C}(\mathbf{U}+\mathbf{V}) = \mathscr{C}(\mathbf{U} : \mathbf{V})$, kun $\mathbf{U}$ ja $\mathbf{V}$ ei-negatiivisesti definiittejä.

C9. $\mathscr{C}(\mathbf{A} : \mathbf{B})^\perp = [\mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{B})]^\perp = \mathscr{C}(\mathbf{A})^\perp \cap \mathscr{C}(\mathbf{B})^\perp$.

Osoitamme ensin *C8*:n. Jos $\mathbf{U}$ ja $\mathbf{V}$ ovat ei-negatiivisesti definiittejä, niin silloin on olemassa sellaiset $\mathbf{A}$ ja $\mathbf{B}$ että $\mathbf{U} = \mathbf{A}\mathbf{A}'$ ja $\mathbf{V} = \mathbf{B}\mathbf{B}'$ ja täten

$$\mathbf{U}+\mathbf{V} = \mathbf{A}\mathbf{A}' + \mathbf{B}\mathbf{B}' = (\mathbf{A} : \mathbf{B})\begin{pmatrix}\mathbf{A}' \\ \mathbf{B}'\end{pmatrix} := \mathbf{T}\mathbf{T}', \tag{9.64}$$

joten $\mathscr{C}(\mathbf{U}+\mathbf{V}) = \mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}(\mathbf{T})$. Väite seuraa suoraan päättelystä

$$\begin{aligned}\mathscr{C}(\mathbf{U}+\mathbf{V}) &= \mathscr{C}(\mathbf{A} : \mathbf{B}) \\ &= \mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{B}) = \mathscr{C}(\mathbf{A}\mathbf{A}') + \mathscr{C}(\mathbf{B}\mathbf{B}') = \mathscr{C}(\mathbf{U}) + \mathscr{C}(\mathbf{V}). \end{aligned} \tag{9.65}$$

Tuloksen *C9* todistus:

$$\begin{aligned}\mathbf{y} \in \mathscr{C}(\mathbf{A} : \mathbf{B})^\perp = \mathscr{N}\begin{pmatrix}\mathbf{A}' \\ \mathbf{B}'\end{pmatrix} &\iff \begin{pmatrix}\mathbf{A}'\mathbf{y} \\ \mathbf{B}'\mathbf{y}\end{pmatrix} = \begin{pmatrix}\mathbf{0} \\ \mathbf{0}\end{pmatrix} \\ &\iff \mathbf{y} \in \mathscr{N}(\mathbf{A}') \ \&\ \mathbf{y} \in \mathscr{N}(\mathbf{B}') \\ &\iff \mathbf{y} \in \mathscr{N}(\mathbf{A}') \cap \mathscr{N}(\mathbf{B}'). \end{aligned} \tag{9.66}$$

9.6 Suora summa

Olemme jo aikaisemminkin tarkastelleet aliavaruuksien summaa:

$$\mathcal{U} + \mathcal{V} = \{\, \mathbf{u} + \mathbf{v} : \mathbf{u} \in \mathcal{U},\ \mathbf{v} \in \mathcal{V} \,\}. \tag{9.67}$$

Erityisesti olemme huomanneet, että sarakevaruuksien tapauksessa

$$\mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{B}). \tag{9.68}$$

Koska nyt $\mathrm{r}(\mathbf{A} : \mathbf{B}) = \dim[\mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{B})]$, saamme tärkeän tuloksen

$$\boxed{\mathrm{r}(\mathbf{A} : \mathbf{B}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) - \dim \mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}).} \tag{9.69}$$

Täten siis

$$\mathrm{r}(\mathbf{A} : \mathbf{B}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) \iff \mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}) = \{\mathbf{0}\}. \tag{9.70}$$

Koska $\mathrm{r}(\mathbf{Z}) = \mathrm{r}(\mathbf{Z}')$, saamme (9.69):sta välittömästi

$$\begin{aligned} \mathrm{r}\begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} &= \mathrm{r}(\mathbf{A}' : \mathbf{B}') = \mathrm{r}(\mathbf{A}') + \mathrm{r}(\mathbf{B}') - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B}') \\ &= \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B}'). \end{aligned} \tag{9.71}$$

Tulosten *A11* ja *A12* sekä (9.69) perusteella on siis voimassa:

$$\mathrm{r}(\mathbf{A} + \mathbf{B}) \leq \mathrm{r}(\mathbf{A} : \mathbf{B}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) - \dim \mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}), \tag{9.72a}$$
$$\mathrm{r}(\mathbf{A} + \mathbf{B}) \leq \mathrm{r}(\mathbf{A}' : \mathbf{B}') = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B}'). \tag{9.72b}$$

Jos $\mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}) = \{\mathbf{0}\}$, niin sanomme, että $\mathscr{C}(\mathbf{A})$ ja $\mathscr{C}(\mathbf{B})$ ovat *erillisiä*: niiden ainoa yhteinen vektori on nollavektori. [Aliavaruuksien leikkaus ei voi olla tyhjä joukko, koska aliavaruuksiin aina kuuluu nollavektori.] Tällöin sanomme, että $\mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{B})$ on $\mathscr{C}(\mathbf{A})$:n ja $\mathscr{C}(\mathbf{B})$:n suora summa ja merkitsemme

$$\mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{B}) = \mathscr{C}(\mathbf{A}) \oplus \mathscr{C}(\mathbf{B}). \tag{9.73}$$

Erityisesti jos $\mathscr{C}(\mathbf{A})$ ja $\mathscr{C}(\mathbf{B})$ ovat ortogonaalisia keskenään (euklidisen sisätulon suhteen) eli $\mathbf{A}'\mathbf{B} = \mathbf{0}$, voimme merkitä

$$\mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}(\mathbf{A}) \boxplus \mathscr{C}(\mathbf{B}). \tag{9.74}$$

Jos (9.73) on voimassa, sanomme että

$\mathscr{C}(\mathbf{B})$ on $\mathscr{C}(\mathbf{A})$:n *komplementti* $\mathscr{C}(\mathbf{A} : \mathbf{B})$:ssä,

ja mikäli (9.74) toteutuu,

on $\mathscr{C}(\mathbf{B})$ aliavaruuden $\mathscr{C}(\mathbf{A})$ *ortogonaalikomplementti* $\mathscr{C}(\mathbf{A} : \mathbf{B})$:ssä.

Olkoon $\mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{T})$. Tällöin on olemassa sellainen matriisi $\mathbf{B}$, että sarakeavaruudella $\mathscr{C}(\mathbf{U})$ on suorasummahajotelma

$$\mathscr{C}(\mathbf{T}) = \mathscr{C}(\mathbf{A}) \oplus \mathscr{C}(\mathbf{B}), \tag{9.75}$$

ja siten $\mathscr{C}(\mathbf{B})$ on $\mathscr{C}(\mathbf{A})$:n komplementti $\mathscr{C}(\mathbf{T})$:ssä. On syytä korostaa, että ehto (9.75) ei määrää $\mathscr{C}(\mathbf{B})$:tä yksikäsitteisesti [kun siis $\mathscr{C}(\mathbf{A})$ ja $\mathscr{C}(\mathbf{T})$ on annettu], paitsi silloin kun $\mathscr{C}(\mathbf{A}) = \{\mathbf{0}\}$ tai $\mathscr{C}(\mathbf{A}) = \mathscr{C}(\mathbf{T})$; muuten sarakeavaruudella $\mathscr{C}(\mathbf{A})$ on äärettömän monta eri komplementtia.

Koko vektoriavaruuden $\mathbb{R}^n$ suorasummahajotelmalla

$$\mathbb{R}^n = \mathscr{C}(\mathbf{A}) \oplus \mathscr{C}(\mathbf{B}) \tag{9.76}$$

on oma erityinen mielenkiintonsa. Aikaisempien tarkastelujen perusteella on aivan ilmeistä, että aikaisemmin määritelty $\mathscr{C}(\mathbf{A})^{\perp}$ tarkoittaa $\mathscr{C}(\mathbf{A})$:n ortogonaalikomplementtia nimenomaan $\mathbb{R}^n$:ssä:

$$\mathbb{R}^n = \mathscr{C}(\mathbf{A}) \boxplus \mathscr{C}(\mathbf{A})^{\perp} = \mathscr{C}(\mathbf{A}) \boxplus \mathscr{N}(\mathbf{A}'), \tag{9.77a}$$

eli (9.47):n perusteella

$$\mathbb{R}^n = \mathscr{C}(\mathbf{A}) \boxplus \mathscr{C}(\mathbf{I} - \mathbf{P}_{\mathbf{A}}) = \mathscr{C}(\mathbf{P}_{\mathbf{A}}) \boxplus \mathscr{C}(\mathbf{I} - \mathbf{P}_{\mathbf{A}}). \tag{9.77b}$$

Hajotelman (9.77) perusteella saamme seuraavat lausekkeet $\mathbf{A}$:n asteelle:

$$n = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{A}^{\perp}) = \mathrm{r}(\mathbf{A}) + \dim \mathscr{N}(\mathbf{A}') \tag{9.78}$$

eli

$$\begin{aligned} \mathrm{r}(\mathbf{A}_{n\times m}) &= n - \mathrm{r}(\mathbf{A}^{\perp}) = n - \dim \mathscr{N}(\mathbf{A}') \\ &= \mathbf{A}\text{:n vaakarivien lkm} - \dim \mathscr{N}(\mathbf{A}'), \end{aligned} \tag{9.79}$$

ja

$$\mathrm{r}(\mathbf{A}^{\perp}) = \mathbf{A}\text{:n vaakarivien lkm} - \mathrm{r}(\mathbf{A}). \tag{9.80}$$

Toisaalta koska

$$\mathbb{R}^m = \mathscr{C}(\mathbf{A}') \boxplus \mathscr{N}(\mathbf{A}), \tag{9.81}$$

on voimassa myös

$$\begin{aligned} \mathrm{r}(\mathbf{A}_{n\times m}) &= m - \dim \mathscr{N}(\mathbf{A}) \\ &= \mathbf{A}\text{:n sarakkeiden lkm} - \dim \mathscr{N}(\mathbf{A}). \end{aligned} \tag{9.82}$$

Vektoriavaruuksien terminologiaa käyttäen vektoriavaruus $\mathcal{T}$ on aliavaruuksien $\mathcal{U}$ ja $\mathcal{V}$ suora summa, jos $\mathcal{T}$ on $\mathcal{U}$:n ja $\mathcal{V}$:n summa ja $\mathcal{U} \cap \mathcal{V} = \{\mathbf{0}\}$:

$$\mathcal{T} = \mathcal{U} \oplus \mathcal{V} \iff \mathcal{T} = \mathcal{U} + \mathcal{V} \text{ ja } \mathcal{U} \cap \mathcal{V} = \{\mathbf{0}\}. \tag{9.83}$$

Matriisien sarakevaruuksina voimme esittää (9.83):n vielä seuraavasti:

$$\mathscr{C}(\mathbf{T}) = \mathscr{C}(\mathbf{A}) \oplus \mathscr{C}(\mathbf{B}) \iff \mathscr{C}(\mathbf{T}) = \mathscr{C}(\mathbf{A} : \mathbf{B}) \text{ ja } \mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}) = \{\mathbf{0}\}. \tag{9.84}$$

Suora summa voidaan luonnehtia myös yhtäpitävästi siten, että

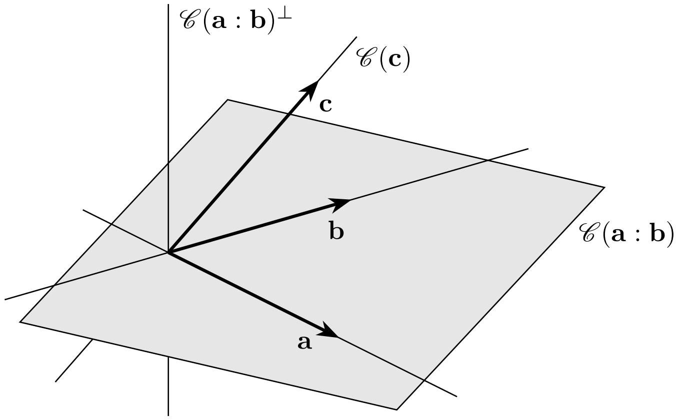
$$\begin{aligned} \mathscr{C}(\mathbf{T}) = \mathscr{C}(\mathbf{A}) \oplus \mathscr{C}(\mathbf{B}) \iff &\text{jokainen } \mathscr{C}(\mathbf{T})\text{:n vektori } \mathbf{t} \text{ voidaan esittää} \\ &\textit{yksikäsitteisesti} \text{ muodossa } \mathbf{t} = \mathbf{a} + \mathbf{b}, \end{aligned} \tag{9.85}$$

missä $\mathbf{a} \in \mathscr{C}(\mathbf{A})$ ja $\mathbf{b} \in \mathscr{C}(\mathbf{B})$, ts.

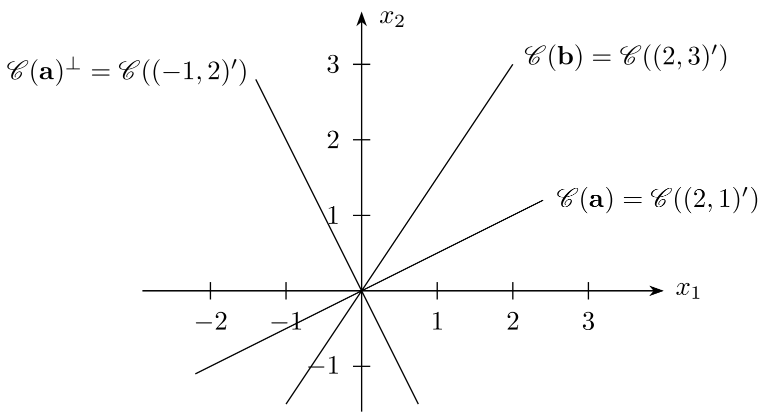
$$\mathbf{t} = \mathbf{Aa} + \mathbf{Bb}, \tag{9.86}$$

missä $\mathbf{Aa}$ ja $\mathbf{Bb}$ ovat yksikäsitteisiä; kertoimet $\mathbf{a}$ ja $\mathbf{b}$ eivät välttämättä yksikäsitteisiä.

On erityisesti huomattava, että vektoriavaruudellinen komplementti ei ole sama kuin joukko-opillinen komplementtikäsite. Tästä voi pyrkiä vakuuttautumaan piirtämällä erilaisia komplementteja kuvioihin 2.6 ja 2.5.



Kuvio 9.5. Kuviossa $\mathbb{R}^3 = \mathscr{C}(\mathbf{a}:\mathbf{b}) \oplus \mathscr{C}(\mathbf{c})$, $\mathbb{R}^3 = \mathscr{C}(\mathbf{a}:\mathbf{b}) \boxplus \mathscr{C}(\mathbf{a}:\mathbf{b})^\perp$. (Sama kuin kuvio 2.5, s. 79)



Kuvio 9.6. Kuvion tilanteessa $\mathbb{R}^2 = \mathscr{C}(\mathbf{a}) \oplus \mathscr{C}(\mathbf{b})$, $\mathbb{R}^2 = \mathscr{C}(\mathbf{a}) \boxplus \mathscr{C}(\mathbf{a})^\perp$. (Sama kuin kuvio 2.6, s. 79

9.7 Ortogonaaliprojektori: sisätulomatriisina I

Aikaisemmin on jo moneen otteeseen tarkasteltu ortogonaaliprojektoria, joka toteuttaa kaksi ehtoa

$$\mathbf{P} \text{ on symmetrinen,} \tag{9.87a}$$

$$\mathbf{P}\text{on idempotentti eli } \mathbf{P}^2 = \mathbf{P}. \tag{9.87b}$$

Tarkastelemme nyt hieman abstraktimmin projektorimatriisin käsitettä ts. sitä miten siihen päädytään.

Olkoot $\mathbf{A}_{n\times m}$ ja $\mathbf{B}_{n\times q}$ sellaisia matriiseja, että vektoriavaruudella $\mathbb{R}^n$ on suorasummahajotelma

$$\mathbb{R}^n = \mathscr{C}(\mathbf{A}) \oplus \mathscr{C}(\mathbf{B}). \tag{9.88}$$

Tällöin siis jokainen $\mathbf{y} \in \mathbb{R}^n$ voidaan esittää yksikäsitteisesti muodossa

$$\mathbf{y} = \mathbf{Aa} + \mathbf{Bb}\,. \tag{9.89}$$

missä siis kertoimet $\mathbf{a}$ ja $\mathbf{b}$ eivät välttämättä ole yksikäsitteisiä – nimenomaan $\mathbf{Aa}$ ja $\mathbf{Bb}$ ovat yksikäsitteisiä. Jos $\mathbf{P}_{n\times n}$ on sellainen matriisi, että

$$\mathbf{Py} = \mathbf{Aa}\,, \tag{9.90}$$

niin $\mathbf{P}$ on *projektori* $\mathbf{A}$:n sarakeavaruudelle $\mathscr{C}(\mathbf{A})$ *suuntaan* $\mathscr{C}(\mathbf{B})$; tässä on "abstrakti" projektorin määritelmä. Vektori $\mathbf{Py} = \mathbf{Aa}$ on tällöin $\mathbf{y}$:n *projektio* $\mathscr{C}(\mathbf{A})$:lle suuntaan $\mathscr{C}(\mathbf{B})$. On syytä huomata, että em. määrittelyssä ei termiä "ortogonaalisuus" käytetty lainkaan.

Kertomalla (9.90):n vasemmalta matriisilla $\mathbf{P}$ saamme yhtälön

$$\mathbf{P}(\mathbf{Py}) = \mathbf{P}(\mathbf{Aa} + \mathbf{0}) = \mathbf{Aa} = \mathbf{Py}. \tag{9.91}$$

Koska (9.91) on voimassa kaikilla $\mathbf{y} \in \mathbb{R}^n$, on $\mathbf{P}$ välttämättä idempotentti ja (9.92a) on voimassa. Samoin on helppo vahvistaa, että $\mathbf{P}$:llä on ominaisuudet (9.92b) ja (9.92c):

$$\mathbf{P}^2 = \mathbf{P}, \quad (\mathbf{I} - \mathbf{P})^2 = \mathbf{I} - \mathbf{P}, \tag{9.92a}$$

$$\mathscr{C}(\mathbf{P}) = \mathscr{C}(\mathbf{A}), \tag{9.92b}$$

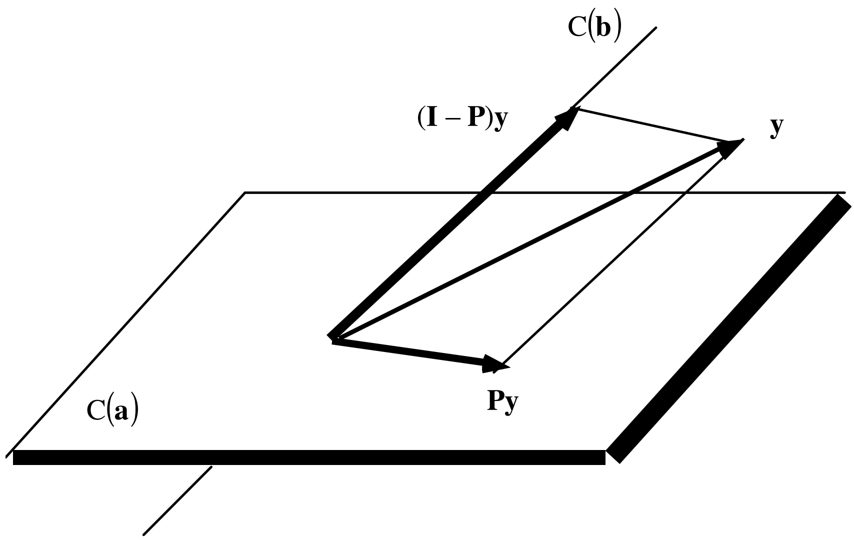
$$\mathscr{N}(\mathbf{P}) = \mathscr{C}(\mathbf{I} - \mathbf{P}). \tag{9.92c}$$

Voimme luonnehtia projektorin myös seuraavasti: Olkoon $\mathbb{R}^n = \mathscr{C}(\mathbf{A}) \oplus \mathscr{C}(\mathbf{B})$. Tällöin $\mathbf{P}$ on projektori sarakeavaruudelle $\mathscr{C}(\mathbf{A})$ suuntaan $\mathscr{C}(\mathbf{B})$ jos ja vain jos

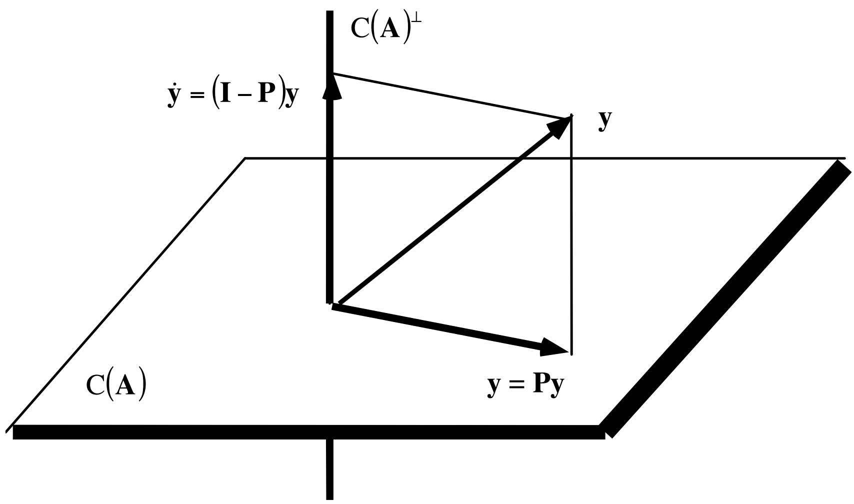
$$\mathbf{P}(\mathbf{Aa} : \mathbf{Bb}) = (\mathbf{Aa} : \mathbf{0}) \ \forall\, \mathbf{a}, \mathbf{b} \tag{9.93a}$$

eli

$$\mathbf{P}(\mathbf{A} : \mathbf{B}) = (\mathbf{A} : \mathbf{0}). \tag{9.93b}$$



Kuvio 9.7. Vektorin $\mathbf{y}$ projektio $\mathscr{C}(\mathbf{A})$:lle suuntaan $\mathscr{C}(\mathbf{b})$. (Kuviossa $\mathbf{a}$ pitää korvata $\mathbf{A}$:lla.)



Kuvio 9.8. Vektorin $\mathbf{y}$ ortogonaaliprojektio $\mathscr{C}(\mathbf{A})$:lle; $\hat{\mathbf{y}} = \mathbf{P}\mathbf{y}$.

Olkoon $\mathbf{P}$ idempotentti matriisi. Osoitetaan seuraavaksi, että tällöin $\mathbb{R}^n$:llä on hajotelma

$$\mathbb{R}^n = \mathscr{C}(\mathbf{P}) \oplus \mathscr{C}(\mathbf{I} - \mathbf{P}). \tag{9.94}$$

Jotta (9.94) olisi voimassa, täytyy seuraavien kahden ehdon toteutua:

$$\mathbb{R}^n = \mathscr{C}(\mathbf{P}) + \mathscr{C}(\mathbf{I} - \mathbf{P}) \tag{9.95a}$$

ja

$$\mathscr{C}(\mathbf{P}) \cap \mathscr{C}(\mathbf{I} - \mathbf{P}) = \{\mathbf{0}\}. \tag{9.95b}$$

Ehto (9.95a) on aivan ilmeinen, sillä jokainen $\mathbf{y} \in \mathbb{R}^n$ voidaan esittää muodossa

$$\mathbf{y} = \mathbf{Py} + (\mathbf{I} - \mathbf{P})\mathbf{y}. \tag{9.96}$$

Ehdon (9.95b) osoittamiseksi valitaan $\mathbf{z} \in \mathscr{C}(\mathbf{P}) \cap \mathscr{C}(\mathbf{I} - \mathbf{P})$, jolloin on olemassa sellaiset $\mathbf{a}$ ja $\mathbf{b}$, että

$$\mathbf{Pa} = (\mathbf{I} - \mathbf{P})\mathbf{b} = \mathbf{z}. \tag{9.97}$$

Kertomalla (9.97) vasemmalta $\mathbf{P}$:llä saadaan

$$\mathbf{PPa} = \mathbf{P}(\mathbf{I} - \mathbf{P})\mathbf{b}, \tag{9.98}$$

mistä $\mathbf{P}$:n idempotenttisuudesta seuraa, että $\mathbf{Pa} = \mathbf{0}$, eli (9.97):n perusteella $\mathbf{z} = \mathbf{0}$. Täten olemme näyttäneet, että (9.95b) todella on voimassa ja näin (9.94) on todistettu.

Huomattakoon, että (9.96) merkitsee sitä, että idempotentti $\mathbf{P}$ on projektori $\mathscr{C}(\mathbf{P})$:lle suuntaan $\mathscr{C}(\mathbf{I} - \mathbf{P})$. Aikaisemmin jo havaittiin, että jos $\mathbf{P}$ on projektori, niin se on idempotentti. Täten voimme päätellä, että $\mathbf{P}$ on projektori *täsmälleen* silloin kun se on idempotentti.

Joskus idempotentista matriisista käytetään myös nimitystä *vino* projektori, jolloin korostetaan sitä, että suunta, jossa projisoidaan, ei olekaan välttämättä kohtisuorassa $\mathscr{C}(\mathbf{A})$:n kanssa. Kuten aiemmin olemme todenneet, saatamme toisinaan käyttää termin ”ortogonaaliprojektori” sijasta lyhyempää termiä ”projektori” – tällöin kuitenkin lukijan oletetaan olevan täsmällisesti selvillä onko kyseessä todellinen ortogonaaliprojektori vai pelkästään idempotentti matriisi.

Jos suunta, jossa projisoidaan, on $\mathscr{C}(\mathbf{B}) = \mathscr{C}(\mathbf{A})^{\perp}$, meillä on hajotelma

$$\mathbb{R}^n = \mathscr{C}(\mathbf{A}) \boxplus \mathscr{C}(\mathbf{A})^{\perp}. \tag{9.99}$$

Tällöin jokainen $\mathbf{y} \in \mathbb{R}^n$ voidaan esittää yksikäsitteisesti muodossa

$$\mathbf{y} = \hat{\mathbf{y}} + \dot{\mathbf{y}}, \quad \hat{\mathbf{y}} \in \mathscr{C}(\mathbf{A}), \quad \dot{\mathbf{y}} \in \mathscr{C}(\mathbf{A})^{\perp}. \tag{9.100}$$

Jos $\mathbf{P}$ on sellainen matriisi, että kertolasku $\mathbf{Py}$ antaa kaikilla vektoreilla $\mathbf{y} \in \mathbb{R}^n$ tuloksen

$$\mathbf{Py} = \hat{\mathbf{y}}, \tag{9.101}$$

niin $\mathbf{P}$ on *ortogonaaliprojektori* $\mathbf{A}$:n sarakeavaruudelle $\mathscr{C}(\mathbf{A})$, ja sitä merkitään symbolilla $\mathbf{P}_{\mathbf{A}}$. Näissä määritelmissä on oleellista, että projisointisuunta on $\mathscr{C}(\mathbf{A})^{\perp}$ eli aliavaruus, joka on kohtisuorassa $\mathscr{C}(\mathbf{A})$:n kanssa tavanomaisen euklidisen sisätulon $\langle \mathbf{t}, \mathbf{u} \rangle = \mathbf{t}'\mathbf{u}$ suhteen Jos sisätuloa vaihdetaan, tilanne muuttuu.

Seuraavassa on lueteltu joitakin yhtäpitäviä ehtoja ortogonaaliprojektorin määrittämiseksi annetun matriisin $\mathbf{A}_{n\times m}$ sarakeavaruudelle; ks. esim. Puntanen, Styan & Isotalo (2011, Th. 2).

(a) $\mathbf{P}(\mathbf{A}\mathbf{b} + \mathbf{A}^{\perp}\mathbf{c}) = \mathbf{A}\mathbf{b}$ *kaikilla* $\mathbf{b}, \mathbf{c}$.

(b) $\mathbf{P}(\mathbf{A} : \mathbf{A}^{\perp}) = (\mathbf{A} : \mathbf{0})$.

(c) $\mathscr{C}(\mathbf{P}) \subset \mathscr{C}(\mathbf{A})$, $\quad \min_{\mathbf{b}} \|\mathbf{y} - \mathbf{A}\mathbf{b}\|^2 = \|\mathbf{y} - \mathbf{P}\mathbf{y}\|^2$ *kaikilla* $\mathbf{y} \in \mathbb{R}^n$.

(d) $\mathscr{C}(\mathbf{P}) = \mathscr{C}(\mathbf{A})$, $\quad \mathbf{P}'\mathbf{P} = \mathbf{P}$.

(e) $\mathscr{C}(\mathbf{P}) = \mathscr{C}(\mathbf{A})$, $\quad \mathbf{P}^2 = \mathbf{P}$, $\quad \mathbf{P}' = \mathbf{P}$.

(f) $\mathscr{C}(\mathbf{P}) = \mathscr{C}(\mathbf{A})$, $\quad \mathbb{R}^n = \mathscr{C}(\mathbf{P}) \boxplus \mathscr{C}(\mathbf{I}_n - \mathbf{P})$.

(g) $\mathbf{P} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}' = \mathbf{A}\mathbf{A}^{+}$.

Matriisi $\mathbf{I}_n - \mathbf{P}_{\mathbf{A}}$ on tietenkin ortogonaaliprojektori $\mathscr{C}(\mathbf{A}^{\perp})$:lle:

$$\mathbf{I}_n - \mathbf{P}_{\mathbf{A}} = \mathbf{P}_{\mathbf{A}^{\perp}}\,. \tag{9.102}$$

Johdetaan sitten $\mathbf{P}$:lle eksplisiittinen esitys (jonka jo tunnemme). Olkoon vektorilla $\mathbf{y} \in \mathbb{R}^n$ hajotelma

$$\mathbf{y} = \mathbf{A}\mathbf{b} + \mathbf{A}^{\perp}\mathbf{c}\,. \tag{9.103}$$

Yhtälön (9.103) kertominen vasemmalta matriisilla $\mathbf{A}'$ johtaa *normaaliyhtälöön*

$$\mathbf{A}'\mathbf{A}\mathbf{b} = \mathbf{A}'\mathbf{y}\,, \tag{9.104}$$

ts.

$$\mathbf{A}'(\mathbf{y} - \mathbf{A}\mathbf{b}) = \mathbf{0}\,. \tag{9.105}$$

Yhtälöstä (9.105) näemme, että jos $\mathbf{b}_*$ on (9.104):n ratkaisu, niin $\mathbf{y} - \mathbf{A}\mathbf{b}_*$ on ortogonaalinen $\mathscr{C}(\mathbf{A})$:n kanssa. Koska

$$\mathbf{y} = \mathbf{A}\mathbf{b}_* + (\mathbf{y} - \mathbf{A}\mathbf{b}_*), \quad \text{missä } \mathbf{A}\mathbf{b}_* \in \mathscr{C}(\mathbf{A}),\ \mathbf{y} - \mathbf{A}\mathbf{b}_* \in \mathscr{C}(\mathbf{A})^{\perp}, \tag{9.106}$$

niin vektori $\mathbf{A}\mathbf{b}_*$ on $\mathbf{y}$:n projektio $\mathscr{C}(\mathbf{A})$:lle. Kannattaa huomata, että (9.104) on tietenkin ratkeava, sillä $\mathbf{A}'\mathbf{y} \in \mathscr{C}(\mathbf{A}'\mathbf{A}) = \mathscr{C}(\mathbf{A}')$. Kaikkien niiden vektoreiden $\mathbf{b}_*$ joukko, jotka toteuttavat (9.104):n, voidaan ilmaista muodossa

$$\mathbf{b}_* = (\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{y} + [\mathbf{I}_m - (\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{A}]\mathbf{t}\,, \tag{9.107}$$

missä $\mathbf{t}$ varioi vapaasti; ks. (10.50) (s. 342). Kun (9.107) kerrotaan vasemmalta matriisilla $\mathbf{A}$ ja käytetään tulosta

$$\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{A} = \mathbf{A}\,, \tag{9.108}$$

saadaan

$$\mathbf{A}\mathbf{b}_{*} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{y}\,, \tag{9.109}$$

ja näin on saatu ortogonaaliprojektorille $\mathbf{P}$ tunnettu esitys. Tulos (9.108) päätellään seuraavasti. Ensinnäkin on ilman muuta voimassa

$$\underline{\mathbf{A}'}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{A} = \underline{\mathbf{A}'}\mathbf{A}\,. \tag{9.110}$$

Koska $\mathrm{r}(\mathbf{A}'\mathbf{A}) = \mathrm{r}(\mathbf{A})$, yhtälöstä (9.110) voidaan astesupistussäännön (RCR) (s. 330) nojalla supistaa alleviivatut matriisit, mistä seuraa (9.108).

Osoitamme vielä, että (d) implikoi (c):n. Nyt

$$\begin{aligned}\|\mathbf{y} - \mathbf{A}\mathbf{b}\|^2 &= \|\mathbf{y} - \mathbf{P}\mathbf{y} + (\mathbf{P}\mathbf{y} - \mathbf{A}\mathbf{b})\|^2 \\ &= \|\mathbf{y} - \mathbf{P}\mathbf{y}\|^2 + \|\mathbf{P}\mathbf{y} - \mathbf{A}\mathbf{b}\|^2 + 2\delta\,,\end{aligned} \tag{9.111}$$

missä

$$\delta = \mathbf{y}'(\mathbf{I}_n - \mathbf{P})'(\mathbf{P}\mathbf{y} - \mathbf{A}\mathbf{b})\,. \tag{9.112}$$

Jos $\delta = 0$, niin tietenkin (c) on voimassa:

$$\|\mathbf{y} - \mathbf{A}\mathbf{b}\|^2 \geq \|\mathbf{y} - \mathbf{P}\mathbf{y}\|^2 \quad \text{kaikilla } \mathbf{b} \in \mathbb{R}^m,\ \mathbf{y} \in \mathbb{R}^n, \tag{9.113}$$

missä yhtäsuuruus pätee vain jos $\mathbf{P}\mathbf{y} = \mathbf{A}\mathbf{b}$. Ehdosta $\mathscr{C}(\mathbf{P}) = \mathscr{C}(\mathbf{A})$ seuraa että on olemassa $\mathbf{t}$ ja $\mathbf{u}$ siten että $\mathbf{P}\mathbf{y} - \mathbf{A}\mathbf{b} = \mathbf{P}\mathbf{y} - \mathbf{P}\mathbf{t} = \mathbf{P}\mathbf{u}$, mikä puolestaan yhdessä oletuksen $\mathbf{P}'\mathbf{P} = \mathbf{P}$ kanssa implikoi, että $\delta = 0$; näin on nähty, että (d):stä seuraa (c). Pantakoon merkille, että

$$\mathbf{P}'\mathbf{P} = \mathbf{P} \iff \mathbf{P}' = \mathbf{P} \text{ and } \mathbf{P} = \mathbf{P}^2, \tag{9.114}$$

joten (d) and (e) ovat yhtäpitäviä.

On selvää (onkohan?), että vektorin $\mathbf{y}$ ortogonaaliprojektio $\hat{\mathbf{y}}$ samoin kuin ortogonaaliprojektori $\mathbf{P}_{\mathbf{A}}$ ovat aina yksikäsitteisiä. Tämä merkitsee sitä, että lauseke $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ on riippumaton $(\mathbf{A}'\mathbf{A})^{-}$:n valinnasta. Ortogonaaliprojektori riippuu vain sarakeavaruudesta, johon projisoidaan:

$$\mathbf{P}_{\mathbf{A}} = \mathbf{P}_{\mathbf{B}} \iff \mathscr{C}(\mathbf{A}) = \mathscr{C}(\mathbf{B}). \tag{9.115}$$

9.8 Ortogonaaliprojektori: sisätulomatriisina V

Olkoon sisätulo $\mathbb{R}^n$:ssä määritelty lausekkeena $\langle \mathbf{t}, \mathbf{u} \rangle_{\mathbf{V}} = \mathbf{t}'\mathbf{V}\mathbf{u}$, missä $\mathbf{V}$ on positiivisesti definiitti $n \times n$-matriisi, ja olkoon $\mathbf{A}^{\perp}_{\mathbf{V}}$ $n \times q$-matriisi, jonka sarakeavaruus on $\mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}} = \mathscr{C}(\mathbf{A})$:n ortogonaalikomplementti, kun sisätulomatriisi on $\mathbf{V}$. Näemme heti, että

$$\begin{aligned} \mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}} &= \{\, \mathbf{y} \in \mathbb{R}^n : \mathbf{z}'\mathbf{A}'\mathbf{V}\mathbf{y} = 0 \text{ kaikilla } \mathbf{z} \in \mathbb{R}^m \,\} \\ &= \{\, \mathbf{y} \in \mathbb{R}^n : \mathbf{A}'\mathbf{V}\mathbf{y} = \mathbf{0} \,\} \\ &= \mathscr{N}(\mathbf{A}'\mathbf{V}) = \mathscr{C}(\mathbf{V}\mathbf{A})^{\perp} = \mathscr{C}(\mathbf{V}^{-1}\mathbf{A}^{\perp}) \,, \end{aligned} \tag{9.116}$$

missä viimeinen yhtälö perustuu siihen, että

$$\mathbf{A}'\mathbf{V} \cdot \mathbf{V}^{-1}\mathbf{A}^{\perp} = \mathbf{0} \implies \mathscr{C}(\mathbf{V}^{-1}\mathbf{A}^{\perp}) \subset \mathscr{N}(\mathbf{A}'\mathbf{V}) \,, \tag{9.117}$$

ja

$$\mathrm{r}(\mathbf{V}^{-1}\mathbf{A}^{\perp}) = \mathrm{r}(\mathbf{A}^{\perp}) = n - \mathrm{r}(\mathbf{A}) \,, \tag{9.118a}$$

$$\dim \mathscr{C}(\mathbf{V}\mathbf{A})^{\perp} = n - \mathrm{r}(\mathbf{V}\mathbf{A}) = n - \mathrm{r}(\mathbf{A}) \,. \tag{9.118b}$$

Matriisin $\mathbf{A}^{\perp}_{\mathbf{V}}$ ominaisuuksia ovat ovat käsitelleet mm. Markiewicz & Puntanen (2015).

Seuraavassa on lueteltu joitakin yhtäpitäviä ehtoja ortogonaaliprojektorin $\mathbf{P}_*$ määrittämiseksi annetun matriisin $\mathbf{A}_{n\times m}$ sarakeavaruudelle sisätulomatriisin ollessa positiivisesti definiitti $\mathbf{V}$, ks. esim. Puntanen, Styan & Isotalo (2011, Prop. 2.6).

(a) $\mathbf{P}_*(\mathbf{A}\mathbf{b} + \mathbf{A}^{\perp}_{\mathbf{V}}\mathbf{c}) = \mathbf{A}\mathbf{b}$ for all $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^q$.

(b) $\mathbf{P}_*(\mathbf{A} : \mathbf{A}^{\perp}_{\mathbf{V}}) = \mathbf{P}_*(\mathbf{A} : \mathbf{V}^{-1}\mathbf{A}^{\perp}) = (\mathbf{A} : \mathbf{0})$.

(c) $\mathscr{C}(\mathbf{P}_*) \subset \mathscr{C}(\mathbf{A})$, $\quad \min_{\mathbf{b}} \|\mathbf{y} - \mathbf{A}\mathbf{b}\|^2_{\mathbf{V}} = \|\mathbf{y} - \mathbf{P}_*\mathbf{y}\|^2_{\mathbf{V}}$ for all $\mathbf{y} \in \mathbb{R}^n$.

(d) $\mathscr{C}(\mathbf{P}_*) = \mathscr{C}(\mathbf{A})$, $\quad \mathbf{P}_*'\mathbf{V}(\mathbf{I}_n - \mathbf{P}_*) = \mathbf{0}$.

(e) $\mathscr{C}(\mathbf{P}_*) = \mathscr{C}(\mathbf{A})$, $\quad \mathbf{P}_*^2 = \mathbf{P}_*$, $\quad (\mathbf{V}\mathbf{P}_*)' = \mathbf{V}\mathbf{P}_*$.

(f) $\mathbf{P}_* = \mathbf{A}(\mathbf{A}'\mathbf{V}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}$, joka on riippumaton $(\mathbf{A}'\mathbf{V}\mathbf{A})^{-}$:n valinnasta.

Matriisi $\mathbf{P}_* = \mathbf{P}_{\mathbf{A};\mathbf{V}}$ on ortogonaaliprojektori sarakeavaruuteen $\mathscr{C}(\mathbf{A})$ kun sisätulo on määritelty $\langle \mathbf{t}, \mathbf{u} \rangle_{\mathbf{V}} = \mathbf{t}'\mathbf{V}\mathbf{u}$.

9.9 Täysiastehajotelma

Olkoon $n \times m$-matriisin $\mathbf{A}$ aste $\mathrm{r}(\mathbf{A}) = r$. Tällöin $\mathbf{A}$:lla on ns. *täysiastehajotelma*

$$\mathbf{A}_{n\times m} = \mathbf{U}_{n\times r}\mathbf{V}'_{r\times m}, \tag{TAH}$$

missä

$$\begin{aligned} \mathbf{U} \text{ on } n \times r\text{-matriisi}, \quad & \mathrm{r}(\mathbf{U}) = r = \mathrm{r}(\mathbf{A}), && (9.119\text{a}) \\ \mathbf{V}' \text{ on } r \times m\text{-matriisi}, \quad & \mathrm{r}(\mathbf{V}) = r. && (9.119\text{b}) \end{aligned}$$

Toisin sanoen: $\mathbf{A}$ esitetään tulona $\mathbf{UV}'$, missä sekä $\mathbf{U}$:n että $\mathbf{V}$:n sarakeasteet ($\mathbf{V}'$:n riviaste) ovat täysiä. Tulos (TAH) todistetaan seuraavasti. Olkoon $\mathbf{U}$ $n \times r$-matriisi, jonka sarakkeet muodostavat $\mathscr{C}(\mathbf{A})$:n kannan. Tällöin jokainen $\mathscr{C}(\mathbf{A})$:n vektori voidaan ilmaista $\mathbf{U}$:n sarakkeiden lineaarikombinaationa. Erityisesti jokainen $\mathbf{A}$:n sarake $\mathbf{a}_i$ voidaan esittää muodossa

$$\mathbf{a}_i = \mathbf{U}_{n\times r}\mathbf{v}_i, \tag{9.120}$$

missä $\mathbf{v}_i$ on jokin $\mathbb{R}^r$:n vektori. Täten on olemassa sellainen $r \times m$-matriisi $\mathbf{V}'$, että

$$(\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_m) = \mathbf{U}(\mathbf{v}_1 : \mathbf{v}_2 : \ldots : \mathbf{v}_m) = \mathbf{UV}'. \tag{9.121}$$

Koska

$$r = \mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{UV}') \le \mathrm{r}(\mathbf{V}') \le r, \tag{9.122}$$

on $r = \mathrm{r}(\mathbf{V})$, joten (TAH) on todistettu.

Täysiastehajotelmalla on monia mielenkiintoisia sovelluksia. Esimerkiksi idempotentilla matriisilla on seuraava ominaisuus.

Olkoon $\mathbf{A}$*:lla täysiastehajotelma* $\mathbf{A} = \mathbf{UV}'$*, missä* $\mathrm{r}(\mathbf{A}) = r > 0$*. Tällöin*

$$\mathbf{A}^2 = \mathbf{A} \iff \mathbf{V}'\mathbf{U} = \mathbf{I}_r. \tag{9.123}$$

Tuloksen (9.123) todistamiseksi oletamme ensin, että $\mathbf{A}$ on idempotentti. Tällöin

$$\mathbf{UV}' = \mathbf{UV}'\mathbf{UV}'. \tag{9.124}$$

Kertomalla (9.124) vasemmalta matriisilla $(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'$ [$\mathbf{U}'\mathbf{U}$:n käänteismatriisi on todella olemassa koska $\mathbf{U}$:n sarakkeet ovat vapaat] saadaan yhtälö

$$\mathbf{V}' = \mathbf{V}'\mathbf{UV}'. \tag{9.125}$$

Kun (9.125) kerrotaan oikealta matriisilla $\mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}$, saadaan väite $\mathbf{I}_r = \mathbf{V}'\mathbf{U}$. Jos toisaalta $\mathbf{V}'\mathbf{U} = \mathbf{I}_r$, niin

$$\mathbf{A}^2 = \mathbf{U}\mathbf{V}'\mathbf{U}\mathbf{V}' = \mathbf{U}\mathbf{I}\mathbf{V}' = \mathbf{U}\mathbf{V}' = \mathbf{A}, \tag{9.126}$$

joten (9.123) on saatu todistetuksi.

Täysiastehajotelman avulla voidaan todistaa myös helposti seuraava tulos:

$$\boxed{\mathbf{A}^2 = \mathbf{A} \implies \mathrm{r}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}). \qquad (9.127\text{a})}$$

Nimittäin (9.123):n nojalla $\mathbf{A}$:n idempotenttisuudesta seuraa sellaisten $\mathbf{U}$:n ja $\mathbf{V}$:n olemassaolo, että

$$\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{U}\mathbf{V}') = \mathrm{tr}(\mathbf{V}'\mathbf{U}) = \mathrm{tr}(\mathbf{I}_r) = r. \tag{9.127b}$$

9.10 Sarakeavaruuden ja asteen ominaisuuksia

Edellä olemme havainneet, että matriisitulon asteelle on voimassa

$$\mathrm{r}(\mathbf{AB}) \leq \mathrm{r}(\mathbf{A}) \text{ ja } \mathrm{r}(\mathbf{AB}) \leq \mathrm{r}(\mathbf{B}). \tag{9.128}$$

Kuinka paljon pienempi tulon $\mathbf{AB}$ aste sitten on kuin $\mathbf{A}$:n aste? Tätä ja muita tämäntapaisia kysymyksiä tarkastelemme tässä luvussa.

Aloitamme seuraavan tuloksen todistuksella:

$$\boxed{\begin{aligned} \mathrm{r}(\mathbf{A}_{n\times a} : \mathbf{B}_{n\times b}) &= \mathrm{r}(\mathbf{A}) + \mathrm{r}[(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{B}] \\ &= \mathrm{r}(\mathbf{A}) + \mathrm{r}\left([\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}']\mathbf{B}\right) \\ &= \mathrm{r}(\mathbf{A}) + \mathrm{r}[(\mathbf{I}_n - \mathbf{A}\mathbf{A}^{+})\mathbf{B}] \\ &= \mathrm{r}(\mathbf{A}) + \mathrm{r}[(\mathbf{I}_n - \mathbf{A}\mathbf{A}^{-})\mathbf{B}]. \qquad (9.129) \end{aligned}}$$

Olkoon $\mathbf{A}^{-} \in \mathbb{R}^{a\times n}$ jokin $\mathbf{A}$:n yleistetty käänteismatriisi. Merkitään

$$\mathbf{L} = \begin{pmatrix} \mathbf{I}_a & -\mathbf{A}^{-}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_b \end{pmatrix}, \tag{9.130}$$

jolloin

$$\begin{aligned} (\mathbf{A} : \mathbf{B})\mathbf{L} &= (\mathbf{A} : \mathbf{B}) \begin{pmatrix} \mathbf{I}_a & -\mathbf{A}^{-}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_b \end{pmatrix} \\ &= (\mathbf{A} : -\mathbf{A}\mathbf{A}^{-}\mathbf{B} + \mathbf{B}) = [\mathbf{A} : (\mathbf{I}_n - \mathbf{A}\mathbf{A}^{-})\mathbf{B}]\,. \end{aligned} \tag{9.131}$$

Matriiisi $\mathbf{L}$ on epäsingulaarinen ja täten

$$\begin{aligned}\mathrm{r}(\mathbf{A}:\mathbf{B}) &= \mathrm{r}[(\mathbf{A}:\mathbf{B})\mathbf{L}] = \mathrm{r}[\mathbf{A}:(\mathbf{I}_n-\mathbf{A}\mathbf{A}^-)\mathbf{B}] \\ &= \mathrm{r}(\mathbf{A}) + \mathrm{r}[(\mathbf{I}_n-\mathbf{A}\mathbf{A}^-)\mathbf{B}] \\ &\quad - \dim \mathscr{C}(\mathbf{A}) \cap \mathscr{C}[(\mathbf{I}_n-\mathbf{A}\mathbf{A}^-)\mathbf{B}]\,. \end{aligned} \tag{9.132}$$

Vielä on osoitettava, että sarakeavaruudet $\mathscr{C}(\mathbf{A})$ ja $\mathscr{C}[(\mathbf{I}_n-\mathbf{A}\mathbf{A}^-)\mathbf{B}]$ ovat erillisiä. Jos $\mathbf{u} \in \mathscr{C}(\mathbf{A}) \cap \mathscr{C}[(\mathbf{I}_n-\mathbf{A}\mathbf{A}^-)\mathbf{B}]$, niin $\mathbf{u}$ voidaan esittää muodossa

$$\mathbf{u} = \mathbf{A}\mathbf{a} = (\mathbf{I}_n-\mathbf{A}\mathbf{A}^-)\mathbf{B}\mathbf{b} \quad \text{joillakin } \mathbf{a} \in \mathbb{R}^a,\ \mathbf{b} \in \mathbb{R}^b. \tag{9.133}$$

Kertomalla (9.133):n matriisilla $\mathbf{A}\mathbf{A}^-$ saadaan

$$\begin{aligned}\mathbf{A}\mathbf{A}^-\mathbf{u} &= \mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{a} = \mathbf{u} \\ &= \mathbf{A}\mathbf{A}^-(\mathbf{I}_n-\mathbf{A}\mathbf{A}^-)\mathbf{B}\mathbf{b} = \mathbf{0}\,,\end{aligned} \tag{9.134}$$

ja siksi $\mathscr{C}(\mathbf{A})$ and $\mathscr{C}[(\mathbf{I}_n-\mathbf{A}\mathbf{A}^-)\mathbf{B}]$ ovat erillisiä. Näin olemme todistaneet kaavan (9.129):n viimeisen yhtälön. Muut (9.129):n esitykset osoitetaan aivan samalla tavalla.

Esimerkki 9.6. Olkoon $p+q$ elementin satunnaisvektorin $\mathbf{z} = \binom{\mathbf{x}}{\mathbf{y}}$ kovarianssimatriiisi

$$\operatorname{cov}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \mathbf{\Sigma} = \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{xx}} & \mathbf{\Sigma}_{\mathbf{xy}} \\ \mathbf{\Sigma}_{\mathbf{yx}} & \mathbf{\Sigma}_{\mathbf{yy}}\end{pmatrix}. \tag{9.135}$$

Ongelma: Osoitettava, että

$$\boxed{\mathscr{C}(\mathbf{\Sigma}_{\mathbf{xy}}) \subset \mathscr{C}(\mathbf{\Sigma}_{\mathbf{xx}}).} \tag{9.136}$$

Ratkaisu: Koska $\mathbf{\Sigma}$ on ei-negatiivisesti definiitti, on olemassa sellainen matriisi $\mathbf{U}_{n\times n}$, että $\mathbf{U}'\mathbf{U} = \mathbf{\Sigma}$. Ositetaan $\mathbf{U}$ siten että

$$\mathbf{\Sigma} = \mathbf{U}'\mathbf{U} = \begin{pmatrix}\mathbf{A}'\\ \mathbf{B}'\end{pmatrix}(\mathbf{A}:\mathbf{B}) = \begin{pmatrix}\mathbf{A}'\mathbf{A} & \mathbf{A}'\mathbf{B} \\ \mathbf{B}'\mathbf{A} & \mathbf{B}'\mathbf{B}\end{pmatrix} = \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{xx}} & \mathbf{\Sigma}_{\mathbf{xy}} \\ \mathbf{\Sigma}_{\mathbf{yx}} & \mathbf{\Sigma}_{\mathbf{yy}}\end{pmatrix}. \tag{9.137}$$

Tuloksen (9.136) päättelemme välittömästi yhtälöistä $\mathbf{\Sigma}_{\mathbf{xy}} = \mathbf{A}'\mathbf{B}$ ja $\mathbf{\Sigma}_{\mathbf{xx}} = \mathbf{A}'\mathbf{A}$. Aivan vastaavasti on tietenkin voimassa

$$\mathscr{C}(\mathbf{\Sigma}_{\mathbf{yx}}) \subset \mathscr{C}(\mathbf{\Sigma}_{\mathbf{yy}}). \tag{9.138}$$

□

Esimerkki 9.7. Olkoon satunnaisvektori $\mathbf{z}$ ja sen kovarianssimatriisi $\mathbf{\Sigma}$ ositettu kuten esimerkissä 9.6. On helppo vakuuttautua, että $\mathbf{\Sigma}$ voidaan muuntaa lohkolävistäjämuotoon seuraavasti:

$$\begin{pmatrix}\mathbf{I}_p & \mathbf{0} \\ -\mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^+ & \mathbf{I}_q\end{pmatrix}\mathbf{\Sigma}\begin{pmatrix}\mathbf{I}_p & -\mathbf{\Sigma}_{\mathbf{xx}}^+\mathbf{\Sigma}_{\mathbf{xy}} \\ \mathbf{0} & \mathbf{I}_q\end{pmatrix} = \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{xx}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\mathbf{yy}\cdot\mathbf{x}}\end{pmatrix}. \tag{9.139}$$

Tehdään nyt $\mathbf{z}$:lle muunnos $\mathbf{t} = \mathbf{B}\mathbf{z}$:

$$\mathbf{t} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \mathbf{B}\mathbf{z} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ -\boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{+} & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}, \tag{9.140}$$

jolloin

$$\mathbf{u} = \mathbf{x}, \quad \mathbf{v} = \mathbf{y} - \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{+}\mathbf{x}. \tag{9.141}$$

Koska

$$\operatorname{cov}\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \operatorname{cov}(\mathbf{B}\mathbf{z}) = \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}' = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{yy}\cdot\mathbf{x}} \end{pmatrix}, \tag{9.142}$$

ovat $\mathbf{u}$ ja $\mathbf{v}$ korreloimattomia, ja niiden kovarianssimatriisit ovat

$$\operatorname{cov}(\mathbf{u}) = \boldsymbol{\Sigma}_{\mathbf{xx}}\,, \tag{9.143}$$

$$\operatorname{cov}(\mathbf{v}) = \boldsymbol{\Sigma}_{\mathbf{yy}\cdot\mathbf{x}} = \boldsymbol{\Sigma}_{\mathbf{yy}} - \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{+}\boldsymbol{\Sigma}_{\mathbf{xy}}\,. \tag{9.144}$$

Itse asiassa edellä olevissa lausekkeissa voidaan $\boldsymbol{\Sigma}_{\mathbf{xx}}^{+}$ korvata millä hyvänsä $\boldsymbol{\Sigma}_{\mathbf{xx}}$:n yleistetyllä käänteismatriisilla. □

Esimerkki 9.8. Olkoon satunnaisvektorin $\mathbf{y}$ kovarianssimatriisi $\boldsymbol{\Sigma}$ ja odotusarvo $\boldsymbol{\mu}$. Todistamme seuraavan tuloksen:

$$\mathbf{y} - \boldsymbol{\mu} \in \mathscr{C}(\boldsymbol{\Sigma}) \text{ todennäköisyydellä } 1. \tag{9.145}$$

Olkoon $\mathbf{Q}_{\boldsymbol{\Sigma}}$ ortogonaaliprojektori $\mathscr{C}(\boldsymbol{\Sigma})^{\perp}$:lle eli $\mathbf{Q}_{\boldsymbol{\Sigma}} = \mathbf{I} - \mathbf{P}_{\boldsymbol{\Sigma}}$. Tällöin

$$\operatorname{cov}(\mathbf{Q}_{\boldsymbol{\Sigma}}\mathbf{y}) = \mathbf{Q}_{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}\mathbf{Q}_{\boldsymbol{\Sigma}} = \mathbf{0}. \tag{9.146}$$

Koska $\mathbf{Q}_{\boldsymbol{\Sigma}}\mathbf{y}$:n kovarianssimatriisi on nollamatriisi, on satunnaisvektorin $\mathbf{Q}_{\boldsymbol{\Sigma}}\mathbf{y}$ arvo jokin vakiovektori todennäköisyydellä 1. Tämän vakiovektorin arvo täytyy tietenkin olla sama kuin $\mathbf{Q}_{\boldsymbol{\Sigma}}\mathbf{y}$:n odotusarvo eli

$$\mathbf{Q}_{\boldsymbol{\Sigma}}\mathbf{y} = \mathrm{E}(\mathbf{Q}_{\boldsymbol{\Sigma}}\mathbf{y}) = \mathbf{Q}_{\boldsymbol{\Sigma}}\,\mathrm{E}(\mathbf{y}) = \mathbf{Q}_{\boldsymbol{\Sigma}}\boldsymbol{\mu}, \tag{9.147}$$

eli $\mathbf{Q}_{\boldsymbol{\Sigma}}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}$, ts.

$$\mathbf{P}_{\boldsymbol{\Sigma}}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{y} - \boldsymbol{\mu}. \tag{9.148}$$

Väite seuraakin välittömästi (9.148):stä. □

Esimerkki 9.9. Todistetaan seuraava tulos:

Olkoon $\mathbf{V}_{n\times n}$ *ei-negatiivisesti definiitti ja olkoon* $\mathbf{M} = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}}$. *Tällöin*

$$\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{M}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{V}\mathbf{M}). \tag{9.149}$$

Ensinnäkin havaitsemme, että

$$\mathscr{C}(\mathbf{X} : \mathbf{VM}) = \mathscr{C}(\mathbf{X}) + \mathscr{C}(\mathbf{VM}) \subset \mathscr{C}(\mathbf{X}) + \mathscr{C}(\mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{V}). \tag{9.150}$$

Matriisin $(\mathbf{X} : \mathbf{VM})$ asteeksi saamme (9.129):n (s. 325) perusteella

$$\mathrm{r}(\mathbf{X} : \mathbf{VM}) = \mathrm{r}(\mathbf{X}) + \mathrm{r}(\mathbf{M} \cdot \mathbf{VM}) = \mathrm{r}(\mathbf{X}) + \mathrm{r}(\mathbf{MV}). \tag{9.151}$$

Toisaalta on voimassa

$$\mathrm{r}(\mathbf{X} : \mathbf{V}) = \mathrm{r}(\mathbf{X}) + \mathrm{r}(\mathbf{MV}), \tag{9.152}$$

joten $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{VM})$. Yhtälön $\mathscr{C}(\mathbf{X} : \mathbf{VM}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{VM})$ osoittamiseksi on vielä näytettävä, että

$$\mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{VM}) = \{\mathbf{0}\}. \tag{9.153}$$

Valitaan vektori $\mathbf{u} \in \mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{VM})$, jolloin on olemassa sellaiset vektorit $\mathbf{a}$ ja $\mathbf{b}$ että

$$\mathbf{u} = \mathbf{Xa} = \mathbf{VMb}. \tag{9.154}$$

Kerrotaan (9.154) vasemmalta matriisilla $\mathbf{M}$, jolloin

$$\mathbf{Mu} = \mathbf{MXa} = \mathbf{MVMb}. \tag{9.155}$$

Koska $\mathbf{MX} = (\mathbf{I} - \mathbf{P_X})\mathbf{X} = \mathbf{0}$, on

$$\mathbf{Mu} = \mathbf{0} = \mathbf{MVMb}. \tag{9.156}$$

Nyt yhtälö $\mathbf{MVMb} = \mathbf{0}$ toteutuu täsmälleen silloin kun

$$\mathbf{VMb} = \mathbf{0}, \tag{9.157}$$

mistä puolestaan seuraa (9.154):n nojalla, että $\mathbf{u} = \mathbf{0}$. Näin on (9.153) todistettu. □

9.11 Matriisitulon aste

Seuraava hyödyllinen tulos ansaitsee erityistä huomiota:

$$\boxed{\begin{aligned} \mathrm{r}(\mathbf{AB}) &= \mathrm{r}(\mathbf{A}) - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B})^{\perp} \\ &= \mathrm{r}(\mathbf{A}) - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{N}(\mathbf{B}') \\ &= \mathrm{r}(\mathbf{A}) - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{I} - \mathbf{P_B}). \end{aligned} \quad (9.158)}$$

Todistaaksemme (9.158):n käytämme ensin tulosta

$$\begin{aligned}\mathrm{r}(\mathbf{B}^{\perp} : \mathbf{A}') &= \mathrm{r}(\mathbf{B}^{\perp}) + \mathrm{r}(\mathbf{A}') - \dim \mathscr{C}(\mathbf{B}^{\perp}) \cap \mathscr{C}(\mathbf{A}') \\ &= \mathrm{r}(\mathbf{B}^{\perp}) + \mathrm{r}(\mathbf{A}) - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B})^{\perp}.\end{aligned} \tag{9.159}$$

Toisaalta (9.129):n (s. 325) nojalla on voimassa

$$\begin{aligned}\mathrm{r}(\mathbf{B}^{\perp} : \mathbf{A}') &= \mathrm{r}(\mathbf{B}^{\perp}) + \mathrm{r}[(\mathbf{I} - \mathbf{P}_{\mathbf{B}^{\perp}})\mathbf{A}'] \\ &= \mathrm{r}(\mathbf{B}^{\perp}) + \mathrm{r}(\mathbf{P}_{\mathbf{B}}\mathbf{A}') = \mathrm{r}(\mathbf{B}^{\perp}) + \mathrm{r}(\mathbf{A}\mathbf{P}_{\mathbf{B}}) \\ &= \mathrm{r}(\mathbf{B}^{\perp}) + \mathrm{r}(\mathbf{A}\mathbf{B}),\end{aligned} \tag{9.160}$$

missä viimeinen yhtälö $\mathrm{r}(\mathbf{A}\mathbf{P}_{\mathbf{B}}) = \mathrm{r}(\mathbf{A}\mathbf{B})$ perustuu seuraavaan:

$$\mathrm{r}(\mathbf{A}\mathbf{B}) \geq \mathrm{r}(\mathbf{A}\mathbf{P}_{\mathbf{B}}) \geq \mathrm{r}(\mathbf{A}\mathbf{P}_{\mathbf{B}}\mathbf{B}) = \mathrm{r}(\mathbf{A}\mathbf{B}). \tag{9.161}$$

Yhdistämällä (9.159) ja (9.160) saadaan (9.158).

Voimme soveltaa tulosta (9.158) keskistetyn $\mathbf{X}_0$:n, kun

$$\mathbf{X} = (\mathbf{1} : \mathbf{X}_0) = (\mathbf{1} : \mathbf{x}_1 : \ldots : \mathbf{x}_k)\,, \tag{9.162}$$

asteen määrittämiseen:

$$\begin{aligned}\mathrm{r}(\widetilde{\mathbf{X}}_0) &= \mathrm{r}(\mathbf{C}\mathbf{X}_0) = \mathrm{r}(\mathbf{X}_0'\mathbf{C}) \\ &= \mathrm{r}(\mathbf{X}_0) - \dim \mathscr{C}(\mathbf{X}_0) \cap \mathscr{C}(\mathbf{C})^{\perp} \\ &= \mathrm{r}(\mathbf{X}_0) - \dim \mathscr{C}(\mathbf{X}_0) \cap \mathscr{C}(\mathbf{1})\,.\end{aligned} \tag{9.163}$$

Täten

$$\mathrm{r}(\widetilde{\mathbf{X}}_0) = \mathrm{r}(\mathbf{X}_0) \quad \Longleftrightarrow \quad \mathbf{1} \notin \mathscr{C}(\mathbf{X}_0)\,. \tag{9.164}$$

Havaitsemme myös, että

$$\mathrm{r}(\mathbf{X}) = \mathrm{r}(\mathbf{X}_0) - \dim \mathscr{C}(\mathbf{X}_0) \cap \mathscr{C}(\mathbf{1}) = \mathrm{r}(\widetilde{\mathbf{X}}_0)\,. \tag{9.165}$$

Täten on voimassa mm.

$$(\mathbf{X}'\mathbf{X})^{-1} \ \exists \quad \Longleftrightarrow \quad (\widetilde{\mathbf{X}}_0'\widetilde{\mathbf{X}}_0)^{-1} \ \exists. \tag{9.166}$$

Tulon astesääntö on liioittelematta erittäin käyttökelpoinen. Alan kirjallisuudessa (9.158) (tai jokin sen versio) esiintyy mm. seuraavissa lähteissä: Marsaglia & Styan (1974a, Cor. 6.2), Rao (1973, p. 28; First Edition 1965, p. 27), Zyskind & Martin (1969, p. 1194), Rao & Bhimasankaram (2000, Th. 3.5.11), Rao & Rao (1998, p. 426).

9.12 Astesupistussääntö (RCR)

Todistetaan sitten seuraava poikkeuksellisen hyödyllinen *astesupistussääntö*, *rank cancellation rule*; ks. Marsaglia & Styan (1974a, s. 271):

$$\mathbf{LAY} = \mathbf{MAY} \ \& \ \mathrm{r}(\mathbf{AY}) = \mathrm{r}(\mathbf{A}) \implies \mathbf{LA} = \mathbf{MA}. \tag{RCR}$$

Todistus. Matriisitulon astesäännön (9.158) nojalla on voimassa

$$\mathrm{r}(\mathbf{AY}) = \mathrm{r}(\mathbf{A}) - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{N}(\mathbf{Y}'), \tag{9.167}$$

joten jos $\mathrm{r}(\mathbf{AY}) = \mathrm{r}(\mathbf{A})$, niin

$$\mathscr{C}(\mathbf{A}') \cap \mathscr{N}(\mathbf{Y}') = \{\mathbf{0}\}. \tag{9.168}$$

Oletus $\mathbf{LAY} = \mathbf{MAY}$ voidaan kirjoittaa tietenkin yhtäpitävästi muodossa

$$\mathbf{Y}'(\mathbf{A}'\mathbf{M}' - \mathbf{A}'\mathbf{L}') = \mathbf{0}. \tag{9.169}$$

Yhtälö (9.169) merkitsee, että jokainen matriisin $\mathbf{A}'\mathbf{M}' - \mathbf{A}'\mathbf{L}'$ sarakkeiden lineaarikombinaatio kuuluu matriisin $\mathbf{Y}'$ nolla-avaruuteen eli

$$\mathscr{C}(\mathbf{A}'\mathbf{M}' - \mathbf{A}'\mathbf{L}') \subset \mathscr{N}(\mathbf{Y}'). \tag{9.170}$$

Toisaalta on voimassa

$$\mathscr{C}(\mathbf{A}'\mathbf{M}' - \mathbf{A}'\mathbf{L}') = \mathscr{C}[\mathbf{A}'(\mathbf{M}' - \mathbf{L}')] \subset \mathscr{C}(\mathbf{A}'). \tag{9.171}$$

Ehdot (9.170) ja (9.171) yhdessä merkitsevät että

$$\mathscr{C}(\mathbf{A}'\mathbf{M}' - \mathbf{A}'\mathbf{L}') \subset \mathscr{N}(\mathbf{Y}') \cap \mathscr{C}(\mathbf{A}'), \tag{9.172}$$

mutta koska (9.168):n perusteella $\mathscr{C}(\mathbf{A}') \in \mathscr{N}(\mathbf{Y}') = \{\mathbf{0}\}$, on välttämättä

$$\mathscr{C}(\mathbf{A}'\mathbf{M}' - \mathbf{A}'\mathbf{L}') = \{\mathbf{0}\}. \tag{9.173}$$

Matriisin sarakeavaruus muodostuu pelkästään nollavektorista täsmälleen silloin kuin kyseinen matriisi on nollamatriisi, joten on oltava $\mathbf{A}'\mathbf{M}' - \mathbf{A}'\mathbf{L}' = \mathbf{0}$, minkä transponointi antaa yhtälön $\mathbf{LA} = \mathbf{MA}$. □

9.13 Kaksi projektorihajotelmaa

Mainitsemme vain tässä kaksi poikkeuksellisen hyödyllistä hajotelmaa ortogonaaliprojektoreille (kun sisätulomatriisina on $\mathbf{I}$). Yksityiskohtia ovat selvittäneet mm. Puntanen, Styan & Isotalo (2011, Th. 7, Th. 8).

Olkoot $\mathbf{P_A}$ *ja* $\mathbf{P_B}$ *ortogonaaliprojektoreita* $\mathscr{C}(\mathbf{A})$*:lle ja* $\mathscr{C}(\mathbf{B})$*:lle. Tällöin*

$$\mathbf{P_A} + \mathbf{P_B} \textit{ on ortogonaaliprojektori} \iff \mathbf{A}'\mathbf{B} = \mathbf{0}\,, \tag{9.174}$$

jolloin

$$\mathbf{P_A} + \mathbf{P_B} = \mathbf{P}_{(\mathbf{A}\,:\,\mathbf{B})}\,. \tag{9.175}$$

Matriisi

$$\mathbf{P} := \mathbf{P_A} + \mathbf{P_B} \tag{9.176}$$

on ortogonaaliprojektori jos ja vain jos se on idempotentti ja symmetrinen. Koska $\mathbf{P_A}$ ja $\mathbf{P_B}$ ovat symmetrisiä, $\mathbf{P}$ on tietenkin symmetrinen. Idempotenttisuus on voimassa jos

$$\mathbf{P}^2 = \mathbf{P_A} + \mathbf{P_B} + \mathbf{P_A}\mathbf{P_B} + \mathbf{P_B}\mathbf{P_A} = \mathbf{P_A} + \mathbf{P_B} = \mathbf{P}\,, \tag{9.177}$$

ts. $\mathbf{P}$ on idempotentti (ja täten ortogonaaliprojektori) jos ja vain jos

$$\mathbf{P_A}\mathbf{P_B} = -\mathbf{P_B}\mathbf{P_A}, \tag{9.178}$$

Kun (9.178) kerrotaan oikealta matriisilla $\mathbf{P_B}$, saadaan $\mathbf{P_A}\mathbf{P_B} = -\mathbf{P_B}\mathbf{P_A}\mathbf{P_B}$, joten $\mathbf{P_A}\mathbf{P_B}$ on symmetrinen: $(\mathbf{P_A}\mathbf{P_B})' = \mathbf{P_A}\mathbf{P_B}$, ts. $\mathbf{P_B}\mathbf{P_A} = \mathbf{P_A}\mathbf{P_B}$. Täten välttämättä

$$\mathbf{P_A}\mathbf{P_B} = -\mathbf{P_A}\mathbf{P_B}\,, \tag{9.179}$$

mikä edelleen seuraa, että

$$\mathbf{P_A}\mathbf{P_B} = \mathbf{0}\,. \tag{9.180}$$

Kertomalla (9.180) vasemmalta $\mathbf{A}'$:lla ja oikealta $\mathbf{B}$:llä saadaan $\mathbf{A}'\mathbf{B} = \mathbf{0}$, ja näin on osoitettu, että (9.178):sta seuraa $\mathbf{A}'\mathbf{B} = \mathbf{0}$. Käänteisimplikaatio $\mathbf{A}'\mathbf{B} = \mathbf{0} \implies$ (9.178) on ilmeinen.

Jos $\mathbf{P_A} + \mathbf{P_B}$ on ortogonaaliprojektori, niin se on ortogonaaliprojektori omalle sarakeavaruudelleen $\mathscr{C}(\mathbf{P_A} + \mathbf{P_B})$. Täten (9.175):n osoittamiseksi meidän on näytettävä että

$$\mathscr{C}(\mathbf{P_A} + \mathbf{P_B}) = \mathscr{C}(\mathbf{A} : \mathbf{B})\,, \tag{9.181}$$

mikä seuraa välittömästi seuraavasta:

$$\mathscr{C}(\mathbf{P_A} + \mathbf{P_B}) = \mathscr{C}\left[(\mathbf{P_A} : \mathbf{P_B})\begin{pmatrix}\mathbf{P_A}\\ \mathbf{P_B}\end{pmatrix}\right] = \mathscr{C}(\mathbf{P_A} : \mathbf{P_B}) = \mathscr{C}(\mathbf{A} : \mathbf{B})\,. \tag{9.182}$$

Ortogonaaliprojektori sarakeavaruudelle

$$\mathscr{C}(\mathbf{A}_{n\times a} : \mathbf{B}_{n\times b}) \tag{9.183}$$

voidaan esittää muodossa

$$\mathbf{P}_{(\mathbf{A}\,:\,\mathbf{B})} = \mathbf{P}_{\mathbf{A}} + \mathbf{P}_{(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{B}} = \mathbf{P}_{\mathbf{A}} + \mathbf{P}_{\mathscr{C}(\mathbf{A}\,:\,\mathbf{B})\cap\mathscr{C}(\mathbf{A})^{\perp}}\,. \tag{9.184}$$

Havaitsemme heti, että summa $\mathbf{P}_{\mathbf{A}} + \mathbf{P}_{(\mathbf{I}-\mathbf{P}_{\mathbf{A}})\mathbf{B}}$ on ortogonaaliprojektori koska $\mathbf{A}'(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B} = \mathbf{0}$. Vielä on osoitettava, että

$$\mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}[\mathbf{A} : (\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{B}]\,, \tag{9.185}$$

mikä seuraa siitä että

$$\mathscr{C}[\mathbf{A} : (\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{B}] \subset \mathscr{C}(\mathbf{A} : \mathbf{B})\,, \tag{9.186}$$

ja ositetun matriisin astesäännöstä (9.129) (s. 325):

$$\mathrm{r}(\mathbf{A}_{n\times a} : \mathbf{B}_{n\times b}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}[(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{B}]\,. \tag{9.187}$$

Mainittakoon, että (9.131):n mukaan (s. 325) on voimassa

$$(\mathbf{A} : \mathbf{B})\mathbf{L} = (\mathbf{A} : \mathbf{B})\begin{pmatrix} \mathbf{I}_a & -\mathbf{A}^{+}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_b \end{pmatrix} = [\mathbf{A} : (\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{B}]\,. \tag{9.188}$$

Matriisi $\mathbf{L}$ on kääntyvä ja täten (9.188) implikoi välittömästi (9.185):n.

Tuloksen (9.184) toinen yhtälö saadaan, kun osoitetaan, että

$$\mathscr{C}[(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B}] = \mathscr{C}(\mathbf{A} : \mathbf{B}) \cap \mathscr{C}(\mathbf{A})^{\perp}. \tag{9.189}$$

Tämä jätetään harjoitustehtäväksi.

Mainittakoon vielä, että kun merkitään $\mathbf{Q}_{\mathbf{A}} = \mathbf{I} - \mathbf{P}_{\mathbf{A}}$, saamme hajotelman

$$\mathbf{I} - \mathbf{P}_{(\mathbf{A}\,:\,\mathbf{B})} = \mathbf{Q}_{\mathbf{A}} - \mathbf{P}_{\mathbf{Q}_{\mathbf{A}}\mathbf{B}} = \mathbf{Q}_{\mathbf{A}}(\mathbf{I} - \mathbf{P}_{\mathbf{Q}_{\mathbf{A}}\mathbf{B}})\,. \tag{9.190}$$

Lähteitä projektoreihin

Baksalary (1987),
Baksalary & Kala (1979),
Baksalary & Trenkler (2009),
Ben-Israel & Greville (2003, §2.7),
Bryant (1984),
Chipman & Rao (1964),
Galántai (2004, 2008),
Groß & Trenkler (1998b),
Halmos (1951, 1958),
Kala (1981, 2008),
Kala & Pordzik (2006),
Mitra (1993),
Mitra & Rao (1974),
Mäkeläinen (1970b),
Rao (1974),
Rao & Mitra (1971, §5.1),
Rao & Yanai (1979),
Seber (1980, 2015), (1984, App. B),
Seber & Lee (2003, App. B),
Sibuya (1970),
Takane & Yanai (1999),
Takeuchi, Yanai & Mukherjee (1982, §1.3),
Yanai, Takeuchi & Takane (2011),
Trenkler (1994, 2006).

Harjoitustehtäviä

9.1. Olkoon $\mathbf{A}^2 = \mathbf{A}$. Osoita että tällöin $\mathbf{A} = \mathbf{A}' \iff \mathscr{C}(\mathbf{A}) = \mathscr{C}(\mathbf{A}')$.

Groß, Trenkler & Troschke (1997).

9.2. Osoita:

$$\begin{gathered} \mathrm{r}(\mathbf{A}+\mathbf{B}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) \\ \iff \\ \dim \mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}) = 0 = \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B}'). \end{gathered}$$

Puntanen, Styan & Isotalo (2011, §17.2).

9.3. Olkoot $\mathbf{A}$ ja $\mathbf{B}$ $n \times n$-matriiseja, jotka toteuttavat ehdon

$$\mathbf{A} + \mathbf{B} = \mathbf{I}_n.$$

Osoita, että tällöin seuraavat ehdot ovat yhtäpitäviä:

(i) $\mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) = n$,

(ii) $\mathbf{A}^2 = \mathbf{A}$ ja $\mathbf{B}^2 = \mathbf{B}$,

(iii) $\mathbf{AB} = \mathbf{0}$.

Puntanen, Styan & Isotalo (2011, §17.3).

9.4. Jos $\mathbf{V}$ on ei-negatiivisesti definiitti, niin

$$\mathscr{C}(\mathbf{A}'\mathbf{V}\mathbf{A}) = \mathscr{C}(\mathbf{A}') \iff \mathrm{r}(\mathbf{A}'\mathbf{V}) = \mathrm{r}(\mathbf{A}) \iff \mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{V})^{\perp} = \{\mathbf{0}\}.$$

9.5. Olkoon $\mathbf{Q_B} = \mathbf{I} - \mathbf{P_B}$. Tällöin

$$\begin{aligned} \mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}) &= \mathscr{C}[\mathbf{A}(\mathbf{A}'\mathbf{B}^{\perp})^{\perp}] = \mathscr{C}[\mathbf{A}(\mathbf{A}'\mathbf{Q_B})^{\perp}] \\ &= \mathscr{C}[\mathbf{A}(\mathbf{I} - \mathbf{P}_{\mathbf{A}'\mathbf{Q_B}})]. \end{aligned}$$

Puntanen, Styan & Isotalo (2011, §5.11).

9.6. Olkoot $\mathbf{P_A}$ ja $\mathbf{P_B}$ ortogonaaliprojektoreita $\mathscr{C}(\mathbf{A})$:lle ja $\mathscr{C}(\mathbf{B})$:lle. Osoita seuraavien ehtojen yhtäpitävyys:

(a) $\mathbf{P_A} - \mathbf{P_B}$ on ortogonaaliprojektori,

(b) $\mathbf{P_A}\mathbf{P_B} = \mathbf{P_B}\mathbf{P_A} = \mathbf{P_B}$,

(c) $\mathscr{C}(\mathbf{B}) \subset \mathscr{C}(\mathbf{A})$,

(d) $\|\mathbf{P_A}\mathbf{x}\| \geq \|\mathbf{P_B}\mathbf{x}\|$ kaikilla $\mathbf{x}$,

(e) $\mathbf{P_A} - \mathbf{P_B} \geq_{\mathsf{L}} \mathbf{0}$.

Jos jokin em. ehdoista on voimassa, niin

$$\mathbf{P_A} - \mathbf{P_B} = \mathbf{P}_{(\mathbf{I}-\mathbf{P_B})\mathbf{A}} = \mathbf{P}_{\mathscr{C}(\mathbf{A})\cap\mathscr{C}(\mathbf{B})^\perp}\,.$$

Puntanen, Styan & Isotalo (2011, s. 152).

9.7. Tarkastellaan ositettua lineaarista mallia

$$\mathscr{M}_{12}:\ \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}\,,$$

ts.

$$\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \sigma^2\mathbf{I}\} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\}\,,$$

missä $\mathbf{X}_i \in \mathbb{R}^{n\times p_i}$, $p = p_1 + p_2$, ja $\mathrm{r}(\mathbf{X}) = p$. Merkitään

$$\mathscr{M}_i = \{\mathbf{y}, \mathbf{X}_i\boldsymbol{\beta}_i, \sigma^2\mathbf{I}\}\,,\ i = 1, 2.$$

Tällöin siis

$$\mathrm{OLSE}(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}} = \begin{pmatrix}\hat{\boldsymbol{\beta}}_1\\ \hat{\boldsymbol{\beta}}_2\end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^+\mathbf{y},$$

missä $\hat{\boldsymbol{\beta}}(\mathscr{M}_{12})$ korostaa, että $\hat{\boldsymbol{\beta}}$ on laskettu mallista $\mathscr{M}_{12}$. Osoita:

(a) $\mathbf{H} = \mathbf{P}_{(\mathbf{X}_1:\mathbf{X}_2)} = \mathbf{P}_{\mathbf{X}_1} + \mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}$, missä $\mathbf{M}_1 = \mathbf{I} - \mathbf{P}_{\mathbf{X}_1}$,

(b) $\mathbf{H}\mathbf{y} = \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{P}_{\mathbf{X}_1}\mathbf{y} + \mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}\mathbf{y}$,

(c) $\mathbf{M}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}\mathbf{y} = \mathbf{M}_1\mathbf{X}_2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y}$,

(d) $\mathbf{X}_1\hat{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) = \mathbf{X}_1\hat{\boldsymbol{\beta}}_1(\mathscr{M}_1) - \mathbf{P}_{\mathbf{X}_1}\mathbf{X}_2\hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12})$,

(e) $\hat{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}_1(\mathscr{M}_1) - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12})$.

Puntanen, Styan & Isotalo (2011, §8.2).

Luku 10

Yleistetty käänteismatriisi

Tunnemme kaikki liiankin hyvin hyvän tohtori P:n heikkoudet ja hänen rehvastelevan esiintymisensä, mutta sitä suurempi syy on meillä hänen ystävillään tehdä voitavamme parantaaksemme hänet noista kuljeksivan elämän aiheuttamista varjopuolista enkä epäile, että hän yliopiston oppituolin saavutettuaan siistii asunsa ja käytöksensä, siivoaa sanansa ja ottaa esimerkkiä säädyllisten ihmisten käyttäytymisestä, sillä emme voi sallia, että näin suuri ja oppinut mies menee hukkaan vain joidenkin vähäisten luonteenvirheiden takia.

Mika Waltari (1948): *Mikael Karvajalka.*

10.1 Johdanto

Aiemmissa luvuissa on jo käsitelty jonkin verran yleistetyn käänteismatriisin käsitettä. Tutustumme nyt siihen täsmällisemmin. Lähteinä yleistettyihin käänteismatriiseihin mainittakoon mm. Rao & Mitra (1971), Ben-Israel & Greville (2003), Piziak & Odell (2007), ja Puntanen, Styan & Isotalo (2011, Ch. 4).

Olkoon $\mathbf{A}$ annettu $n \times m$-matriisi ja $\mathbf{y}$ annettu $n \times 1$-vektori. Tarkastellaan lineaarista yhtälöryhmää

$$\mathbf{Ab} = \mathbf{y}. \tag{10.1}$$

Ongelmana on löytää sellainen vektori $\mathbf{b}$ $(m \times 1)$, että yhtälö $\mathbf{Ab} = \mathbf{y}$ toteutuu. Tunnetusti lineaarisella yhtälöryhmällä on joko

(a) ei yhtään ratkaisua, tai

(b) yksikäsitteinen ratkaisu, tai

(c) äärettömän monta ratkaisua.

Kuten muistamme, sarakevaruuden avulla voidaan ratkaisujen olemassaoloa luonnehtia siten, että

> yhtälöllä $\mathbf{Ab} = \mathbf{y}$ on ratkaisu (ts. yhtälö on ratkeava) täsmälleen silloin kun $\mathbf{y}$ kuuluu $\mathbf{A}$:n sarakevaruuteen $\mathscr{C}(\mathbf{A})$.

Jos ratkaisu on olemassa, niin se on yksikäsitteinen vain jos $\mathbf{A}$:n sarakkeet ovat vapaat. Tällöin ratkaisu on tunnetusti

$$\mathbf{b}_0 = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} := \mathbf{G}_1\mathbf{y}. \tag{10.2}$$

Voimme esittää (10.2):n mukaisen ratkaisun myös muodossa (miksi?)

$$\mathbf{b}_0 = (\mathbf{A}'\mathbf{N}\mathbf{A})^{-1}\mathbf{A}'\mathbf{N}\mathbf{y} := \mathbf{G}_2\mathbf{y}. \tag{10.3}$$

missä $\mathbf{N}$ on mielivaltainen ehdon $\mathrm{r}(\mathbf{A}'\mathbf{N}\mathbf{A}) = m$ toteuttava matriisi. (Huomaa että $\mathbf{G}_1\mathbf{y} = \mathbf{G}_2\mathbf{y}$ mutta mahdollisesti $\mathbf{G}_1 \neq \mathbf{G}_2$.) Jos $\mathbf{A}$ on epäsingulaarinen neliömatriisi, niin ratkaisu on tietenkin $\mathbf{b}_0 = \mathbf{A}^{-1}\mathbf{y}$.

Voimme nyt pyrkiä etsimään $m \times n$-matriisia $\mathbf{G}$, joka käyttäytyisi mahdollisimman paljon $\mathbf{A}^{-1}$:n tapaan; esimerkiksi olisi luontevaa toivoa $\mathbf{G}$:ltä sellaista ominaisuutta, että mikäli $\mathbf{Ab} = \mathbf{y}$ on ratkeava, niin $\mathbf{Gy}$ olisi yksi ratkaisu. Tällöinhän matriisi $\mathbf{G}$ toimisi lineaarisen yhtälöryhmän ratkaisemisessa kuten $\mathbf{A}^{-1}$; lauseketta $\mathbf{A}^{-1}\mathbf{y}$ vastaten on $\mathbf{Gy}$ yhtälön $\mathbf{Ab} = \mathbf{y}$ ratkaisu.

Huomautus 10.1. Korostettakoon vielä (uudestaan), että yhtälöllä $\mathbf{Ab} = \mathbf{y}$ ei välttämättä ole lainkaan ratkaisua, jolloin tietenkään $\mathbf{G}$:n määrittämisestä ei ole apua ("olemattoman") ratkaisun löytämisessä.

Voimmekin määritellä matriisin $\mathbf{A}_{n\times m}$ yleistetyn käänteismatriisin $\mathbf{G}_{m\times n}$ kolmella yhtäpitävällä tavalla; ks. esim. Rao & Mitra (1971, ss. 20–21):

Matriisi $\mathbf{G}_{m\times n}$ on matriisin $\mathbf{A}_{n\times m}$ yleistetty käänteismatriisi, jos jokin seuraavista yhtäpitävistä ehdoista toteutuu:

(a) *Vektori $\mathbf{Gy}$ on (10.1):n ratkaisu aina kun (10.1) on ratkeava eli aina kun $\mathbf{y} \in \mathscr{C}(\mathbf{A})$,*

(b) $\mathbf{AGA} = \mathbf{A}$,

(c) $\mathbf{GA}$ *on idempotentti ja* $\mathrm{r}(\mathbf{GA}) = \mathrm{r}(\mathbf{A})$ *eli yhtäpitävästi* $\mathbf{AG}$ *on idempotentti ja* $\mathrm{r}(\mathbf{AG}) = \mathrm{r}(\mathbf{A})$.

Olemme jo aiemmin käyttäneet $\mathbf{G}$:stä merkintää $\mathbf{A}^-$ (ja lyhempänä nimityksenä käyttäneet *g-inverssiä*). Kaikkien yleistettyjen käänteismatriisien joukkoa merkitsemme symbolilla $\{\mathbf{A}^-\}$:

$$\{\mathbf{A}^-\} = \{\, \mathbf{G} : \mathbf{AGA} = \mathbf{A} \,\}. \tag{10.4}$$

Ben-Israel & Greville (2003) käyttävät merkintää

$$\mathbf{A}\{1\} = \{\, \mathbf{G} : \mathbf{AGA} = \mathbf{A} \,\}. \tag{10.5}$$

Yleistetty käänteismatriisi ei välttämättä ole yksikäsitteinen – se on aina olemassa, mutta se on yksikäsitteinen vain jos $\mathbf{A}^{-1}$ on olemassa. Jos $\mathbf{A}^{-1}$ on olemassa, niin kertomalla (10.1) vasemmalta ja oikealta $\mathbf{A}^{-1}$:llä saadaan yhtälö $\mathbf{G} = \mathbf{A}^{-1}$. On paikallaan korostaa yhtälön $\mathbf{Ab} = \mathbf{y}$ ratkaisun yksikäsitteisyyden ja $\mathbf{A}^{-}$:n yksikäsitteisyyden välistä eroa:

- ratkaisu on yksikäsitteinen $\iff$ $\mathbf{A}$:n sarakeaste täysi,
- $\mathbf{A}^{-}$ on yksikäsitteinen $\iff$ $\mathbf{A}$ on epäsingulaarinen neliömatriisi.

Yleistetyn käänteismatriisin olemassaolon voimme päätellä (esimerkiksi) $\mathbf{A}$:n täysiastehajotelman

$$\mathbf{A}_{n\times m} = \mathbf{UV}' \tag{TAH}$$

avulla; yhtälössä (TAH)

$$r(\mathbf{U}_{n\times r}) = r = \mathrm{r}(\mathbf{A}), \quad \mathrm{r}(\mathbf{V}_{m\times r}) = r. \tag{10.6}$$

Tällöin nimittäin

$$\mathbf{UV}' \cdot \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}' \cdot \mathbf{UV}' = \mathbf{UV}' \tag{10.7}$$

eli

$$\mathbf{G} := \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}' \in \{\mathbf{A}^{-}\}, \tag{10.8}$$

ts.

$$\mathbf{AGA} = \mathbf{A}\,. \tag{mp1}$$

On helppo varmistaa, että (10.8):n mukainen $\mathbf{G}$ toteuttaa lisäksi seuraavat kolme ehtoa:

$$\text{(mp2) } \mathbf{GAG} = \mathbf{G}\,, \quad \text{(mp3) } (\mathbf{AG})' = \mathbf{AG}\,, \quad \text{(mp4) } (\mathbf{GA})' = \mathbf{GA}\,. \tag{10.9}$$

Matriisia $\mathbf{G}$, joka toteuttaa kaikki neljä mp$i$-ehtoa, sanotaan *Mooren-Penrosen käänteismatriisiksi* ja sitä merkitään symbolilla $\mathbf{A}^{+}$. Matriisi $\mathbf{A}^{+}$ osoittautuu yksikäsitteiseksi; hajotelma (TAH) sen sijaan ei ole yksikäsitteinen. Lähteinä Mooren-Penrosen käänteismatriisiin mainittakoon Moore (1920, 1935), Penrose (1955), Ben-Israel (2002) ja Ben-Israel & Greville (2003, Appendix A).

Jos $\mathbf{G}$ toteuttaa ehdon (mp$i$), merkitsemme $\mathbf{G} = \mathbf{A}_i^{-}$ tai $\mathbf{G} \in \{\mathbf{A}_i^{-}\}$, ja jos se toteuttaa ehdot i ja j, merkitsemme $\mathbf{G} = \mathbf{A}_{ij}^{-}$ tai $\mathbf{G} \in \{\mathbf{A}_{ij}^{-}\}$, jolloin $\mathbf{G}$:tä voidaan sanoa $\{ij\}$-inverssiksi.

Jos $\mathbf{A}$:n sarakkeet ovat vapaat, niin havaitsemme välittömästi että

$$(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \in \{\mathbf{A}^{-}\}. \tag{10.10}$$

Itse asiassa $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ on $\mathbf{A}$:n Mooren-Penrosen käänteismatriisi:

$$\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}', \text{ kun } \mathrm{r}(\mathbf{A}_{n\times m}) = m. \tag{10.11}$$

Matriisi $\mathbf{A}_L := (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ on $\mathbf{A}$:n *vasemmanpuoleinen käänteismatriisi* siinä mielessä, että se toteuttaa yhtälön $\mathbf{A}_L\mathbf{A} = \mathbf{I}_m$. Vastaavasti

$$\mathbf{A}^+ = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}, \text{ kun } \mathrm{r}(\mathbf{A}_{n\times m}) = n, \tag{10.12}$$

jolloin $\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ on $\mathbf{A}$:n oikeanpuoleinen käänteismatriisi. Erityisesti jos $\mathbf{A}$ on pystyvektori $\mathbf{a}$, niin

$$\mathbf{a}^+ = (\mathbf{a}'\mathbf{a})^{-1}\mathbf{a}' = \frac{1}{\mathbf{a}'\mathbf{a}}\mathbf{a}' \ \text{(vaakavektori)}, \ \mathbf{a} \in \mathbb{R}^n, \ \mathbf{a} \neq \mathbf{0}. \tag{10.13}$$

Vastaavasti

$$(\mathbf{a}')^+ = \mathbf{a}(\mathbf{a}'\mathbf{a})^{-1} = \frac{1}{\mathbf{a}'\mathbf{a}}\mathbf{a} \ \text{(pystyvektori)}, \ \mathbf{a} \in \mathbb{R}^n, \ \mathbf{a} \neq \mathbf{0}. \tag{10.14}$$

Osoitamme seuraavaksi, että (10.1) ja (10.1) ovat yhtäpitäviä [muiden kohtien todistukset ks. esim. Puntanen, Styan & Isotalo (2011, Ch. 4)] eli

$$\mathbf{AGA} = \mathbf{A} \iff \mathbf{AG} \text{ on idempotentti ja } \mathrm{r}(\mathbf{AG}) = \mathrm{r}(\mathbf{A}). \tag{10.15}$$

Oletetaan ensin, että $\mathbf{AGA} = \mathbf{A}$. Tällöin tietenkin $\mathbf{AGAG} = \mathbf{AG}$ eli $\mathbf{AG}$ on idempotentti. Koska

$$\mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{AGA}) \leq \mathrm{r}(\mathbf{AG}) \leq \mathrm{r}(\mathbf{A}), \tag{10.16}$$

niin $\mathrm{r}(\mathbf{AG}) = \mathrm{r}(\mathbf{A})$. Oletetaan sitten, että $\mathbf{AG}$ on idempotentti ja $\mathrm{r}(\mathbf{AG}) = \mathrm{r}(\mathbf{A})$. Astesupistussäännön (RCR) (s. 330) nojalla voimme tällöin supistaa alleviivatut termit yhtälöstä

$$\mathbf{AGA}\underline{\mathbf{G}} = \mathbf{A}\underline{\mathbf{G}}. \tag{10.17}$$

Täten (10.1) implikoi (10.1):n.

On syytä havaita, että $\mathbf{A}$:n yleistetyn käänteismatriisin aste on aina suurempi tai yhtäsuuri kuin $\mathbf{A}$:n aste eli

$$\mathrm{r}(\mathbf{A}^-) \geq \mathrm{r}(\mathbf{A}). \tag{10.18}$$

Tämä nähdään seuraavasta:

$$\mathrm{r}(\mathbf{A}^-) \geq \mathrm{r}(\mathbf{AA}^-) \geq \mathrm{r}(\mathbf{AA}^-\mathbf{A}) = \mathrm{r}(\mathbf{A}). \tag{10.19}$$

10.2 Yleistetty käänteismatriisi & singulaariarvohajotelma

Osoitamme seuraavaksi, että yleistetty käänteismatriisi voidaan määrittää singulaariarvohajotelman avulla. Olkoon matriisilla $\mathbf{A}_{n\times m}$ seuraava singulaariarvohajotelma:

$$\mathbf{A} = \mathbf{U}\begin{pmatrix}\mathbf{\Delta}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}' = \mathbf{U}\mathbf{\Delta}\mathbf{V}', \tag{SAH}$$

missä $\mathbf{U}_{n\times n}$ ja $\mathbf{V}_{m\times m}$ ovat ortogonaalisia ja

$$\mathbf{U}_{n\times n} = (\mathbf{U}_1 : \mathbf{U}_0), \quad \mathbf{U}_1 \in \mathbb{R}^{n\times r}, \quad r = \mathrm{r}(\mathbf{A}), \tag{10.20a}$$

$$\mathbf{V}_{m\times m} = (\mathbf{V}_1 : \mathbf{V}_0), \quad \mathbf{V}_1 \in \mathbb{R}^{m\times r}, \quad \mathbf{\Delta} \in \mathbb{R}^{n\times m}. \tag{10.20b}$$

Matriisi $\mathbf{\Delta}_1 = \mathrm{diag}(\delta_1, \delta_2, \dots, \delta_r)$ on $\mathbf{A}$:n nollasta poikeavien singulaariarvojen muodostama lävistäjämatriisi:

$$\delta_i = \mathrm{sg}_i(\mathbf{A}) = \sqrt{\mathrm{ch}_i(\mathbf{A}'\mathbf{A})} > 0, \quad i = 1, 2, \dots, r. \tag{10.21}$$

Osoitamme nyt seuraavan tuloksen:

$$\mathbf{G} = \mathbf{V}\begin{pmatrix}\mathbf{\Delta}_1^{-1} & \mathbf{K}\\ \mathbf{L} & \mathbf{N}\end{pmatrix}\mathbf{U}' \Rightarrow \mathbf{G} \in \{\mathbf{A}^-\}. \tag{10.22}$$

missä $\mathbf{K}$, $\mathbf{L}$ *ja* $\mathbf{N}$ *ovat mielivaltaisia matriiseja (sopivan dimensioisia).*

Täten varioimalla matriisien $\mathbf{K}$, $\mathbf{L}$ ja $\mathbf{N}$ elementtejä saamme generoiduksi kaikki $\mathbf{A}$:n yleistetyt käänteismatriisit.

Tulos (10.22) seuraa suoraan kertolaskusta

$$\begin{aligned}\mathbf{AGA} &= \mathbf{U}\mathbf{\Delta}\mathbf{V}' \cdot \mathbf{V}\begin{pmatrix}\mathbf{\Delta}_1^{-1} & \mathbf{K}\\ \mathbf{L} & \mathbf{N}\end{pmatrix}\mathbf{U}' \cdot \mathbf{U}\mathbf{\Delta}\mathbf{V}'\\ &= \mathbf{U}\begin{pmatrix}\mathbf{\Delta}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{\Delta}_1^{-1} & \mathbf{K}\\ \mathbf{L} & \mathbf{N}\end{pmatrix}\begin{pmatrix}\mathbf{\Delta}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}'\\ &= \mathbf{U}\begin{pmatrix}\mathbf{\Delta}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}' = \mathbf{A}.\end{aligned} \tag{10.23}$$

Itse asiassa voidaan todistaa seuraavat väiteet:

$$\mathbf{G} \in \{\mathbf{A}^-\} \iff \mathbf{G} = \mathbf{V} \begin{pmatrix} \boldsymbol{\Delta}_1^{-1} & \mathbf{K} \\ \mathbf{L} & \mathbf{N} \end{pmatrix} \mathbf{U}', \tag{10.24}$$

$$\mathbf{G} \in \{\mathbf{A}_{12}^-\} \iff \mathbf{G} = \mathbf{V} \begin{pmatrix} \boldsymbol{\Delta}_1^{-1} & \mathbf{K} \\ \mathbf{L} & \mathbf{L}\boldsymbol{\Delta}_1\mathbf{K} \end{pmatrix} \mathbf{U}', \tag{10.25}$$

$$\mathbf{G} \in \{\mathbf{A}_{13}^-\} = \{\mathbf{A}_\ell^-\} \iff \mathbf{G} = \mathbf{V} \begin{pmatrix} \boldsymbol{\Delta}_1^{-1} & \mathbf{0} \\ \mathbf{L} & \mathbf{N} \end{pmatrix} \mathbf{U}', \tag{10.26}$$

$$\mathbf{G} \in \{\mathbf{A}_{14}^-\} = \{\mathbf{A}_m^-\} \iff \mathbf{G} = \mathbf{V} \begin{pmatrix} \boldsymbol{\Delta}_1^{-1} & \mathbf{K} \\ \mathbf{0} & \mathbf{N} \end{pmatrix} \mathbf{U}', \tag{10.27}$$

$$\mathbf{G} = \mathbf{A}^+ \iff \mathbf{G} = \mathbf{V} \begin{pmatrix} \boldsymbol{\Delta}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{U}', \tag{10.28}$$

missä $\mathbf{K}$, $\mathbf{L}$ ja $\mathbf{N}$ ovat mielivaltaisia (sopivan kokoisia) matriiseja.

Tuloksesta (10.24) ilmenee havainnollisesti yleistetyn käänteismatriisin yksikäsitteisyyden mahdollinen puute: matriisien $\mathbf{K}$, $\mathbf{L}$ ja $\mathbf{N}$ elementit ($nm - r^2$ kpl) voidaan valita miten hyvänsä. On helppo vakuuttautua esimerkiksi siitä, että sopivilla valinnoilla saadaan muodostetuksi sellainen $\mathbf{A}^-$, jolla on haluttu aste k:

$$\mathrm{r}(\mathbf{A}) \leq k = \mathrm{r}(\mathbf{A}^-) \leq \min(n, m). \tag{10.29}$$

Erityisesti jos symmetrisellä $\mathbf{A}$:lla on ominaisarvohajotelma

$$\begin{aligned} \mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}' = (\mathbf{T}_1 : \mathbf{T}_0) \begin{pmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{T}_1' \\ \mathbf{T}_0' \end{pmatrix} &= \mathbf{T}_1\boldsymbol{\Lambda}_1\mathbf{T}_1' \\ &= \lambda_1 \mathbf{t}_1\mathbf{t}_1' + \cdots + \lambda_r \mathbf{t}_r\mathbf{t}_r', \end{aligned} \tag{10.30}$$

niin

$$\begin{aligned} \mathbf{A}^+ = \mathbf{T}\boldsymbol{\Lambda}^+\mathbf{T}' = (\mathbf{T}_1 : \mathbf{T}_0) \begin{pmatrix} \boldsymbol{\Lambda}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{T}_1' \\ \mathbf{T}_0' \end{pmatrix} &= \mathbf{T}_1\boldsymbol{\Lambda}_1^{-1}\mathbf{T}_1' \\ &= \frac{1}{\lambda_1} \mathbf{t}_1\mathbf{t}_1' + \cdots + \frac{1}{\lambda_r} \mathbf{t}_r\mathbf{t}_r', \end{aligned} \tag{10.31}$$

missä $\lambda_1, \dots, \lambda_r$ ovat $\mathbf{A}$:n nollasta poikkeavat ominaisarvot.

Tuloksen (10.28) nojalla havaitsemme välittömästi, että

$$\mathbf{A}\mathbf{A}^+ = \mathbf{U}_1\boldsymbol{\Delta}_1\mathbf{V}_1'\mathbf{V}_1\boldsymbol{\Delta}_1^{-1}\mathbf{U}_1' = \mathbf{U}_1\mathbf{U}_1' = \mathbf{P}_{\mathbf{U}_1} = \mathbf{P}_{\mathbf{A}}, \tag{10.32a}$$

$$\mathbf{A}^+\mathbf{A} = \mathbf{V}_1\boldsymbol{\Delta}_1^{-1}\mathbf{U}_1'\mathbf{U}_1\boldsymbol{\Delta}_1\mathbf{V}_1' = \mathbf{V}_1\mathbf{V}_1' = \mathbf{P}_{\mathbf{V}_1} = \mathbf{P}_{\mathbf{A}'}. \tag{10.32b}$$

10.3 Vino projektori

Luvun 9.7 (s. 318) mukaan idempotentilla matriisilla $\mathbf{P}_{n\times n}$ on seuraava ominaisuus:

$$\mathbf{P}^2 = \mathbf{P} \implies \mathbf{P} \text{ on } \textit{vino projektori } \mathscr{C}(\mathbf{P})\text{:lle suuntaan } \mathscr{N}(\mathbf{P}) = \mathscr{C}(\mathbf{I}_n - \mathbf{P}). \tag{10.33}$$

Täten matriisin $\mathbf{AA}^-$ idempotenttisuus merkitsee, että $\mathbf{AA}^-$ on vino projektori sarakeavaruudelle $\mathscr{C}(\mathbf{AA}^-)$ suuntaan $\mathscr{N}(\mathbf{AA}^-)$; termi ”vino projektori” on siis vain synonyymi idempotentille matriisille. Koska

$$\mathscr{C}(\mathbf{AA}^-) \subset \mathscr{C}(\mathbf{A}) \text{ ja } \mathrm{r}(\mathbf{AA}^-) = \mathrm{r}(\mathbf{A}), \tag{10.34}$$

niin tietenkin

$$\mathscr{C}(\mathbf{AA}^-) = \mathscr{C}(\mathbf{A}), \tag{10.35}$$

ja täten

$$\mathbf{AA}^- \text{ on vino projektori } \mathscr{C}(\mathbf{A})\text{:lle suuntaan } \mathscr{N}(\mathbf{AA}^-) = \mathscr{C}(\mathbf{I}_n - \mathbf{AA}^-). \tag{10.36}$$

Vastaavasti, koska $\mathbf{A}^-\mathbf{A}$ on idempotentti, havaitaan että

$$\mathbf{A}^-\mathbf{A} \text{ on vino projektori } \mathscr{C}(\mathbf{A}^-\mathbf{A})\text{:lle suuntaan } \mathscr{N}(\mathbf{A}^-\mathbf{A})\,. \tag{10.37}$$

Aliavaruus $\mathscr{N}(\mathbf{A}^-\mathbf{A})$ voidaan tietenkin ilmaista muodossa

$$\mathscr{N}(\mathbf{A}^-\mathbf{A}) = \mathscr{C}(\mathbf{I}_m - \mathbf{A}^-\mathbf{A})\,. \tag{10.38}$$

Koska

$$\mathscr{N}(\mathbf{A}) \subset \mathscr{N}(\mathbf{A}^-\mathbf{A}) \subset \mathscr{N}(\mathbf{AA}^-\mathbf{A}) = \mathscr{N}(\mathbf{A})\,, \tag{10.39}$$

niin nolla-avaruus $\mathscr{N}(\mathbf{A})$ voidaan esittää muodossa

$$\mathscr{N}(\mathbf{A}_{n\times m}) = \mathscr{C}(\mathbf{I}_m - \mathbf{A}^-\mathbf{A})\,. \tag{10.40}$$

Matriisien $\mathbf{AA}^+$ ja $\mathbf{A}^+\mathbf{A}$ symmetrisyys merkitsee seuraavia erittäin tärkeitä Mooren-Penrosen käänteismatriisin ominaisuuksia, ks. myös (10.32) (s. 340):

$$\boxed{\mathbf{AA}^+ = \mathbf{P}_\mathbf{A}\,, \quad \mathbf{A}^+\mathbf{A} = \mathbf{P}_{\mathbf{A}'}\,. \qquad (10.41)}$$

Voimme ilmaista $\mathscr{C}(\mathbf{A})^\perp$:n matriisin $\mathbf{A}^-$ avulla:

$$\mathscr{C}(\mathbf{A})^\perp = \mathscr{C}[\mathbf{I}_n - (\mathbf{A}')^-\mathbf{A}'] = \mathscr{C}[\mathbf{I}_n - (\mathbf{A}^-)'\mathbf{A}']\,. \tag{10.42}$$

Tämä seuraa siitä, että

$$\mathscr{C}(\mathbf{A})^\perp = \mathscr{N}(\mathbf{A}')\,. \tag{10.43}$$

Ensimmäinen yhtälö (10.42):ssa seuraa (10.40):sta ja (10.43):sta ja toinen siitä, että

$$\{(\mathbf{A}^-)'\} = \{(\mathbf{A}')^-\}\,. \tag{10.44}$$

Hellä varoitus (10.44):sta: on hiukan kyseenalaista kirjoittaa $(\mathbf{A}^-)' = (\mathbf{A}')^-$, koska (10.44) merkitsee kahden *joukon* identtisyyttä. Aina on kuitenkin voimassa

$$(\mathbf{A}^+)' = (\mathbf{A}')^+, \qquad (\mathbf{A}^-)' \in \{(\mathbf{A}')^-\}, \quad (\mathbf{A}')^- \in \{(\mathbf{A}^-)'\}. \tag{10.45}$$

Olemme aiemmin määritelleet matriisin $\mathbf{A}^\perp$ siten, että

$$\mathbf{A}^\perp = \text{ matriisi, jolla on ominaisuus } \mathscr{C}(\mathbf{A}^\perp) = \mathscr{N}(\mathbf{A}')\,. \tag{10.46}$$

Saamme esimerkiksi seuraavat esitykset $\mathbf{A}^\perp$:lle:

$$\mathbf{I}_n - (\mathbf{A}')^-\mathbf{A}' \in \{\mathbf{A}^\perp\},\ \mathbf{I}_n - (\mathbf{A}^-)'\mathbf{A}' \in \{\mathbf{A}^\perp\},\ \mathbf{I}_n - \mathbf{A}\mathbf{A}^+ \in \{\mathbf{A}^\perp\}\,. \tag{10.47}$$

Lisää ”nastoituksen” ominaisuuksia löytyy artikkelista Markiewicz & Puntanen (2015).

10.4 Lineaarisen yhtälöryhmän ratkaisut

Kuten edellä totesimme, niin jos yhtälö $\mathbf{Ab} = \mathbf{y}$ on ratkeava, niin yksi sen ratkaisu on $\mathbf{A}^-\mathbf{y}$. Ratkeavan yhtälön $\mathbf{Ab} = \mathbf{y}$ *yleinen ratkaisu* eli kaikkien ratkaisujen joukko saadaan tunnetusti lausekkeesta

$$\begin{aligned} \mathbf{b}_0 &= \{\text{yhtälön } \mathbf{Ab} = \mathbf{y} \text{ jokin ratkaisu}\} \\ &\quad + \{\text{yhtälön } \mathbf{Ab} = \mathbf{0} \text{ yleinen ratkaisu}\}, \end{aligned} \tag{10.48}$$

ts.

$$\mathbf{b}_0 = \mathbf{A}^-\mathbf{y} + \mathbf{t}\,, \tag{10.49}$$

missä $\mathbf{t}$ on yhtälön $\mathbf{At} = \mathbf{0}$ yleinen ratkaisu eli $\mathbf{t} \in \mathscr{N}(\mathbf{A}) = \mathscr{C}(\mathbf{I}_m - \mathbf{A}^-\mathbf{A})$. Täten yhtälön $\mathbf{Ab} = \mathbf{y}$ yleinen ratkaisu on esitettävissä muodossa

$$\mathbf{b}_0 = \mathbf{A}^-\mathbf{y} + (\mathbf{I}_m - \mathbf{A}^-\mathbf{A})\mathbf{z}\,, \tag{10.50}$$

missä $\mathbf{A}^-$ on (jokin kiinteä) $\mathbf{A}$:n yleistetty käänteismatriisi ja $\mathbf{z}$ on mielivaltainen m elementin vektori. Kun $\mathbf{z}$:aa varioidaan, saadaan yhtälön $\mathbf{Ab} = \mathbf{y}$ kaikki ratkaisut generoiduksi. Toisaalta on osoitettavissa, että lauseke $\mathbf{A}^-\mathbf{y}$ generoi kaikki (ratkeavan) yhtälön $\mathbf{Ab} = \mathbf{y}$ ratkaisut, kun $\mathbf{A}^-$ käy läpi kaikki mahdolliset $\mathbf{A}$:n yleistetyt käänteismatriisit.

Ratkaisulla $\mathbf{A}^+\mathbf{y}$ on yksi erityispiirre: se on se ratkaisu, jolla on on pienin normi. Tämä nähdään seuraavasti. Olkoon $\mathbf{b}_*$ jokin ratkaisu yhtälölle $\mathbf{Ab} = \mathbf{y}$. Silloin

$$\begin{aligned} \|\mathbf{b}_*\|^2 &= \|\mathbf{A}^+\mathbf{y} + (\mathbf{I}_m - \mathbf{A}^+\mathbf{A})\mathbf{z}\|^2 \\ &= \|\mathbf{A}^+\mathbf{y}\|^2 + \|(\mathbf{I}_m - \mathbf{A}^+\mathbf{A})\mathbf{z}\|^2 + 2(\mathbf{A}^+\mathbf{y})'(\mathbf{I}_m - \mathbf{A}^+\mathbf{A})\mathbf{z} \\ &= \|\mathbf{A}^+\mathbf{y}\|^2 + \|(\mathbf{I}_m - \mathbf{A}^+\mathbf{A})\mathbf{z}\|^2 \\ &\geq \|\mathbf{A}^+\mathbf{y}\|^2 \qquad \text{kaikilla } \mathbf{z} \in \mathbb{R}^m, \end{aligned} \tag{10.51}$$

mikä seuraa vektorien $\mathbf{A}^+\mathbf{y}$ ja $(\mathbf{I}_m - \mathbf{A}^+\mathbf{A})\mathbf{z}$ ortogonaalisuudesta.

Esimerkki 10.1. Tarkastellaan yhtälöä $b_2 = b_1 + 1$, mikä matriisimuodossa on

$$\mathbf{Ab} = \mathbf{y}\,, \quad \mathbf{A} = \mathbf{a}' = (1, -1)\,, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad \mathbf{y} = 1\,. \tag{10.52}$$

Merkitään (10.52):n ratkaisujen joukkoa

$$\mathcal{V} = \left\{ \mathbf{b} \in \mathbb{R}^2 : (1, -1) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = 1 \right\} = \{\, \mathbf{b} \in \mathbb{R}^2 : b_2 = b_1 - 1 \,\}\,. \tag{10.53}$$

Jos valitaan

$$(\mathbf{a}')^- = (1, -1)^- = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \tag{10.54}$$

niin sitä vastaava ratkaisu on $\mathbf{b}_0 = \binom{2}{1} \cdot 1 = \binom{2}{1}$. Mooren–Penrosen käänteismatriisi

$$(\mathbf{a}')^+ = \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix} = \mathbf{a}(\mathbf{a}'\mathbf{a})^{-1} \tag{10.55}$$

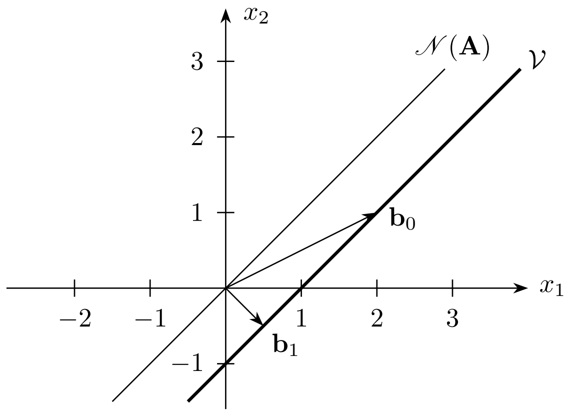
antaa kuvion 10.1 mukaisen ratkaisun $\mathbf{b}_1 = \mathbf{A}^+\mathbf{y}$. Matriisin $\mathbf{A} = \mathbf{a}'$ nolla-avaruus on

$$\mathscr{N}(\mathbf{A}) = \{\, \mathbf{b} \in \mathbb{R}^2 : \mathbf{Ab} = \mathbf{0} \,\} = \{\, \mathbf{b} \in \mathbb{R}^2 : b_2 = -b_1 \,\} = \mathscr{C} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{10.56}$$

Täten yhtälön $\mathbf{Ab} = \mathbf{y}$ kaikkien ratkaisujen joukko on

$$\mathcal{V} = \mathbf{b}_0 + \mathscr{N}(\mathbf{A})\,, \tag{10.57}$$

kuten kuviostakin ilmenee.



Kuvio 10.1. Joukko $\mathcal{V}$ on yhtälön $\mathbf{Ab} = \mathbf{y}$ kaikkien ratkaisujen joukko. Vektori $\mathbf{b}_1$ on ratkaisu, jolla on lyhin normi: $\mathbf{b}_1 = \mathbf{A}^+\mathbf{y}$.

10.5 Yleistetyn käänteismatriisin ominaisuuksia

Seuraavassa on yksinkertaisesti lueteltu joitakin matriisin $\mathbf{A}_{n\times m}$ yleistetyn käänteismatriisin ominaisuuksia

(G1) Ratkeavan yhtälön $\mathbf{Ax} = \mathbf{y}$ yleinen ratkaisu on

$$\mathbf{A}^-\mathbf{y} + (\mathbf{I}_m - \mathbf{A}^-\mathbf{A})\mathbf{z}, \text{ missä } \mathbf{z} \text{ on mielivaltainen } \mathbb{R}^m\text{:n vektori.} \tag{10.58}$$

(G2) Yhtälö $\mathbf{AX} = \mathbf{Y}$ on ratkeava $\iff$ $\mathscr{C}(\mathbf{Y}) \subset \mathscr{C}(\mathbf{A})$ ts. on olemassa $\mathbf{A}^-$ siten että $\mathbf{AA}^-\mathbf{Y} = \mathbf{Y}$, jolloin yleinen ratkaisu on

$$\mathbf{A}^-\mathbf{Y} + (\mathbf{I}_m - \mathbf{A}^-\mathbf{A})\mathbf{Z}, \tag{10.59}$$

missä $\mathbf{Z}$ on mielivaltainen sopivan-dimensioinen matriisi.

(G3) Jos on olemassa jokin $\mathbf{A}^-$ siten että $\mathbf{AA}^-\mathbf{Y} = \mathbf{Y}$, niin $\mathbf{AA}^-\mathbf{Y} = \mathbf{Y}$ kaikilla $\mathbf{A}^- \in \{\mathbf{A}^-\}$.

(G4) Yhtälö $\mathbf{AXB} = \mathbf{C}$ on ratkeava $\iff$ $\mathbf{AA}^-\mathbf{CB}^-\mathbf{B} = \mathbf{C}$, jolloin yleinen ratkaisu on

$$\mathbf{X} = \mathbf{A}^-\mathbf{CB}^- + \mathbf{Z} - \mathbf{A}^-\mathbf{AZBB}^-, \tag{10.60}$$

missä $\mathbf{Z}$ on mielivaltainen sopivan-dimensioinen matriisi.

(G5) Matriisin $\mathbf{A}$ yleistetty käänteismatriisi voidaan esittää seuraavasti:

(i) $\mathbf{G} = \mathbf{A}^- + \mathbf{U} - \mathbf{A}^-\mathbf{AUAA}^-$,

(ii) $\mathbf{G} = \mathbf{A}^- + \mathbf{V}(\mathbf{I}_n - \mathbf{AA}^-) + (\mathbf{I}_m - \mathbf{A}^-\mathbf{A})\mathbf{W}$, missä $\mathbf{A}^-$ on mielivaltainen (mutta kiinteä) $\mathbf{A}$:n yleistetty käänteismatriisi ja $\mathbf{U}$, $\mathbf{V}$ ja $\mathbf{W}$ mielivaltaisia matriiseja. Erityisesti jos valitaan $\mathbf{A}^- = \mathbf{A}^+$, niin

(iii) $\mathbf{G} = \mathbf{A}^+ + \mathbf{U} - \mathbf{P}_{\mathbf{A}'}\mathbf{U}\mathbf{P}_{\mathbf{A}}$,

(iv) $\mathbf{G} = \mathbf{A}^+ + \mathbf{V}(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}}) + (\mathbf{I}_m - \mathbf{P}_{\mathbf{A}'})\mathbf{W}$.

(G6) Olkoon $\mathbf{A} \neq \mathbf{0}$, $\mathbf{C} \neq \mathbf{0}$. Tällöin $\mathbf{AB}^-\mathbf{C}$ on riippumaton $\mathbf{B}^-$:n valinnasta jos ja vain jos

$$\mathscr{C}(\mathbf{C}) \subset \mathscr{C}(\mathbf{B}) \ \& \ \mathscr{C}(\mathbf{A}') \subset \mathscr{C}(\mathbf{B}'). \tag{10.61}$$

(G7) $\mathbf{AA}^-\mathbf{C} = \mathbf{C} \iff \mathscr{C}(\mathbf{C}) \subset \mathscr{C}(\mathbf{A})$, $\mathbf{AB}^-\mathbf{B} = \mathbf{A} \iff \mathscr{C}(\mathbf{A}') \subset \mathscr{C}(\mathbf{B}')$.

(G8) $\mathscr{C}(\mathbf{AA}^-) = \mathscr{C}(\mathbf{A})$.

(G9) $\mathscr{C}(\mathbf{A}^+) = \mathscr{C}(\mathbf{A}')$.

(G10) $\mathscr{C}(\mathbf{I}_m - \mathbf{A}^-\mathbf{A}) = \mathscr{N}(\mathbf{A})$, $\quad \mathscr{N}(\mathbf{I}_n - \mathbf{AA}^-) = \mathscr{C}(\mathbf{A})$.

(G11) $\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^+\mathbf{A}' = \mathbf{A}'(\mathbf{AA}')^+ = (\mathbf{A}^+)^+$.

(G12) $(\mathbf{A}^+)' = (\mathbf{A}')^+$.

(G13) $(\mathbf{A}^-)' \in \{(\mathbf{A}')^-\}$, $\quad (\mathbf{A}')^- \in \{(\mathbf{A}^-)'\}$.

(G14) $(\mathbf{AB})^+ = \mathbf{B}^+\mathbf{A}^+ \iff \mathscr{C}(\mathbf{BB}'\mathbf{A}') \subset \mathscr{C}(\mathbf{A}') \ \& \ \mathscr{C}(\mathbf{A}'\mathbf{AB}) \subset \mathscr{C}(\mathbf{B})$.

Matriisilausekkeiden riippumattomuus niissä esiintyvistä yleistetyistä käänteismatriiseista on monesti erittäin keskeinen kysymys ja tällöin tulos (G6) on välttämätön apuväline. Jos oletamme, että (10.61) on voimassa, niin silloin on olemassa $\mathbf{U}$ ja $\mathbf{Z}$ siten, että $\mathbf{C} = \mathbf{BU}$, $\mathbf{A}' = \mathbf{B}'\mathbf{Z}$ ja täten

$$\mathbf{AB}^-\mathbf{C} = \mathbf{Z}'\mathbf{BB}^-\mathbf{BU} = \mathbf{Z}'\mathbf{BU}\,, \tag{10.62}$$

mikä ei siis mitenkään riipu $\mathbf{B}^-$:sta. Todistus toiseen suuntaan on hiukan vaativampi, ks. Puntanen, Styan & Isotalo (2011, Ch. 12).

Näitä invarianttisuustuloksia tarvitaan tavantakaa, kun ollaan tekemisissä yleistettyjen käänteismatriisien kanssa. Yleistetyn käänteismatriisin käyttökelpoisuus on paljolti siinä, että sen avulla saadaan "mukavasti" esitetyksi tiettyjä matriisilausekkeita, mutta on tietenkin selvää, että näiden lausekkeiden käyttäjän on oltava täsmällisesti selvillä lausekkeiden mahdollisesta yksikäsitteisyyden puutteesta. Esimerkiksi yhtälön $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ (joka on ratkeava kaikilla $\mathbf{X}$:n ja $\mathbf{y}$:n arvoilla) yksi ratkaisu on

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y} \tag{10.63}$$

ja kaikkien ratkaisujen joukko voidaan esittää muodossa

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} + [\mathbf{I}_p - (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}]\mathbf{z} \\ &= (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{y} + [\mathbf{I}_p - (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{X}]\mathbf{t} \\ &= (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{y} + [\mathbf{I}_p - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{+}]\mathbf{t} \\ &= (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{y} + (\mathbf{I}_p - \mathbf{P}_{\mathbf{X}'})\mathbf{t}, \end{aligned} \tag{10.64}$$

missä $\mathbf{z}$ ja $\mathbf{t}$ ovat mielivaltaisia $\mathbb{R}^p$:n vektoreita. Vektorin $\hat{\beta}$ lauseke on $(\mathbf{X}'\mathbf{X})^{-}$:n valinnasta riippumaton täsmälleen silloin kun $\mathbf{X}$:n sarakkeet ovat vapaat (miksi?), jolloin $(\mathbf{X}'\mathbf{X})^{-} = (\mathbf{X}'\mathbf{X})^{-1}$. Kuitenkin $\hat{\beta}$:n elementtien tietty lineaarikombinaatio $\mathbf{k}'\hat{\beta}$ voi olla silti yksikäsitteinen. On helposti pääteltävissä seuraava tulos:

$$\mathbf{k} \in \mathscr{C}(\mathbf{X}') \text{ eli } \exists\, \mathbf{a} : \mathbf{k} = \mathbf{X}'\mathbf{a} \text{ eli } \mathbf{k}'\beta \text{ on estimoituva} \tag{10.65a}$$

$$\Longleftrightarrow$$

$$\mathbf{k}'\hat{\beta} = \mathbf{k}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} \text{ on riippumaton } (\mathbf{X}'\mathbf{X})^{-}\text{:n valinnasta.} \tag{10.65b}$$

10.5.1 Muuntaminen lohkolävistäjämuotoon: $\mathbf{A} \in \mathrm{NND}_n$

Sivuilla 145 ja 148 tarkasteltiin ositetun positiivisesti definiitin matriisin $\mathbf{A}$ muuntamista lohkolävistäjämuotoon. Seuraavaksi tehdään samoja tarkasteluja ei-negatiivisesti definiitin matriisin tapauksessa, jolloin siis $\mathbf{A}$ saattaa olla singulaarinen.

Olkoon ei-negatiivisesti definiitti $\mathbf{A}_{n\times n}$ ositettu siten että

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \tag{10.66}$$

missä $\mathbf{A}_{11}$ on neliömatriisi. Tällöin

$$\boxed{\begin{aligned}\mathbf{A} &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{=} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-}\mathbf{A}_{12} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{A}_{11}^{\sim}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \\ &:= \mathbf{BCD}\,, \qquad (10.67)\end{aligned}}$$

missä $\mathbf{A}_{11}^{=}$, $\mathbf{A}_{11}^{-}$ ja $\mathbf{A}_{11}^{\sim}$ ovat $\mathbf{A}_{11}$:n yleistettyjä käänteismatriiseja, ja

$$\begin{aligned}\mathbf{A}_{22\cdot 1} = \mathbf{A}/\mathbf{A}_{11} &= \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-}\mathbf{A}_{12} \\ &= \mathbf{A}_{11}\text{:n Schurin komplementti } \mathbf{A}\text{:ssa.}\end{aligned} \tag{10.68}$$

Lisäksi

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{=} & \mathbf{I} \end{pmatrix} \cdot \mathbf{A} \cdot \begin{pmatrix} \mathbf{I} & -\mathbf{A}_{11}^{\sim}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-}\mathbf{A}_{12} \end{pmatrix} = \mathbf{B}^{-1}\mathbf{A}\mathbf{D}^{-1}. \tag{10.69}$$

Em. tulosten todistus: ks. esim. Puntanen, Styan & Isotalo (2011, Th. 13). Muistutamme lukijaa siitä,että jos $\mathbf{A}$ on ei-negatiivisesti definiitti, niin on olemassa $\mathbf{L}$ siten että $\mathbf{A} = \mathbf{L}'\mathbf{L}$, missä $\mathbf{L} = (\mathbf{L}_1 : \mathbf{L}_2)$ ja täten

$$\mathbf{A} = \mathbf{L}'\mathbf{L} = \begin{pmatrix} \mathbf{L}_1'\mathbf{L}_1 & \mathbf{L}_1'\mathbf{L}_2 \\ \mathbf{L}_2'\mathbf{L}_1 & \mathbf{L}_2'\mathbf{L}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}. \tag{10.70}$$

Seuraavan tuloksen ovat esittäneet Tällöin seuraavat väitteet ovat yhtäpitäviä, ks. Marsaglia & Styan (1974b):

(a) $\mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{A}_{11}) + \mathrm{r}(\mathbf{A}_{22})$, i.e., $\mathscr{C}(\mathbf{L}_1) \cap \mathscr{C}(\mathbf{L}_2) = \{\mathbf{0}\}$,

(b) $\mathbf{A}^+ = \begin{pmatrix} \mathbf{A}_{11\cdot 2}^+ & -\mathbf{A}_{11\cdot 2}^+\mathbf{A}_{12}\mathbf{A}_{22}^+ \\ -\mathbf{A}_{22}^+\mathbf{A}_{21}\mathbf{A}_{11\cdot 2}^+ & \mathbf{A}_{22}^+ + \mathbf{A}_{22}^+\mathbf{A}_{21}\mathbf{A}_{11\cdot 2}^+\mathbf{A}_{12}\mathbf{A}_{22}^+ \end{pmatrix}$,

(c) $\mathbf{A}^+ = \begin{pmatrix} \mathbf{A}_{11}^+ + \mathbf{A}_{11}^+\mathbf{A}_{12}\mathbf{A}_{22\cdot 1}^+\mathbf{A}_{21}\mathbf{A}_{11}^+ & -\mathbf{A}_{11}^+\mathbf{A}_{12}\mathbf{A}_{22\cdot 1}^+ \\ -\mathbf{A}_{22\cdot 1}^+\mathbf{A}_{21}\mathbf{A}_{11}^+ & \mathbf{A}_{22\cdot 1}^+ \end{pmatrix}$.

Harjoitustehtäviä

10.1. Osoita että ei-negatiivisesti definiitillä singulaarisella $\mathbf{A}$:lla on olemassa sekä symmetrinen että epäsymmetrinen yleistetty käänteismatriisi.

10.2. Olkoon

$$\mathbf{A}_{n\times m} = \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{pmatrix}, \quad \mathbf{G}_{m\times n} = \begin{pmatrix} \mathbf{B}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

missä $\mathrm{rank}(\mathbf{B}_{r\times r}) = r = \mathrm{rank}(\mathbf{A})$. Osoita että $\mathbf{G} \in \{\mathbf{A}^-\}$.

10.3. Osoita että $\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^+\mathbf{A}' = \mathbf{A}'(\mathbf{A}\mathbf{A}')^+$ ja että $(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ on $\{123\}$-inverssi matriisille $\mathbf{A}$.

10.4. Osoita:

$$\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^+ = \mathbf{X}'\mathbf{V}(\mathbf{X}'\mathbf{V})^+ = (\mathbf{X}'\mathbf{V}\mathbf{X})^+(\mathbf{X}'\mathbf{V}\mathbf{X}) = (\mathbf{V}\mathbf{X})^+\mathbf{V}\mathbf{X}\,.$$

10.5. Osoita:

(a) $\begin{pmatrix} \mathbf{0}_{p\times n} \\ \mathbf{A}^- \end{pmatrix} \in \{(\mathbf{0}_{n\times p} : \mathbf{A}_{n\times m})^-\}$,

(b) $\det(\mathbf{P}) \neq 0$, $\det(\mathbf{Q}) \neq 0$
$\implies \mathbf{Q}^{-1}\mathbf{A}^{-}\mathbf{P}^{-1} \in \{(\mathbf{PAQ})^{-}\}$, $\mathbf{Q}^{-1}\mathbf{A}^{+}\mathbf{P}^{-1} = (\mathbf{PAQ})^{+}$.

10.6. Let $\mathbf{V} \in \mathrm{NND}_n$ and $\mathbf{G} \in \{\mathbf{V}^{-}\}$. Johda riittävät ja välttämättömät ehdot sille, että

(a) $\mathbf{G} \in \mathrm{NND}_n$,

(b) $\mathbf{G} \in \mathrm{NND}_n$ ja $\mathbf{G} \in \{\mathbf{V}_{12}^{-}\}$ eli $\mathbf{G}$ on $\mathbf{V}$:n ei-negatiivisesti definiitti refleksiivinen yleistetty käänteismatriisi.

Ohje: käytä $\mathbf{V}^{-}$:n esitystä

$$\mathbf{V}^{-} = \mathbf{T}\begin{pmatrix} \mathbf{\Lambda}_1^{-1} & \mathbf{L} \\ \mathbf{M} & \mathbf{N} \end{pmatrix}\mathbf{T}',$$

ja Albertin (1969) tulosta s. 197.

10.7. Olkoot $\mathbf{A}_{n\times n}$ ja $\mathbf{B}_{n\times n}$ ei-negatiivisesti definiittejä. Osoita seuraavien väitteiden yhtäpitävyys:

1. $\mathbf{A} \leq_{\mathsf{L}} \mathbf{B}$,
2. $\mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{B})$ and $\mathrm{ch}_1(\mathbf{AB}^{+}) \leq 1$.

Liski & Puntanen (1989), Puntanen, Styan & Isotalo (2011, §14.2).

10.8. Olkoon

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

(a) Määritä jokin sellainen matriisi $\mathbf{G}$, joka toteuttaa ehdon $\mathbf{AGA} = \mathbf{A}$.

(b) Mikä ehto $\mathbf{G}$:n elementtien on toteutettava, jotta $\mathbf{G}$ olisi $\mathbf{A}$:n g-inverssi?

(c) Määritä $\mathbf{A}^{+}$,

(d) ... $\mathbf{A}$:n SAH ja sen avulla esimerkit tapauksista $\mathbf{A}_{12}^{-}$, $\mathbf{A}_{13}^{-}$, $\mathbf{A}_{14}^{-}$,

(e) ... ortonormaali kanta $\mathscr{C}(\mathbf{A})$:lle,

(f) ... täysiastehajotelma matriisille $\mathbf{A}$,

(g) ... kaksi eri valintaa matriisille $\mathbf{A}^{\perp}$,

(h) ... ortonormaali kanta $\mathscr{C}(\mathbf{A})^{\perp}$:lle.

10.9. Tarkastellaan matriisia

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Määritä

(a) $\mathbf{X}^{+}$,

(b) ortonormaali kanta $\mathscr{C}(\mathbf{X})$:lle,

(c) täysiastehajotelma matriisille $\mathbf{X}$,

(d) $(\mathbf{X}'\mathbf{X})^{+}$,

(e) kaksi muuta $\mathbf{X}'\mathbf{X}$:n yleistettyä käänteismatriisia,

(f) ortonormaali kanta $\mathscr{C}(\mathbf{X}'\mathbf{X})$:lle,

(g) $\mathbf{H}, \mathbf{M}$,

(h) kaksi eri valintaa matriisille $\mathbf{X}^{\perp}$,

(i) ortonormaali kanta $\mathscr{C}(\mathbf{X})^{\perp}$:lle.

Notation-mtxkirja

$\mathbb{R}$	real numbers
$\mathbb{R}^{n\times m}$	set of $n \times m$ real matrices
$\mathbb{R}_r^{n\times m}$	subset of $\mathbb{R}^{n\times m}$ consisting of matrices with rank r
$\mathbb{R}_s^n$	subset of $\mathbb{R}^{n\times n}$ consisting of symmetric matrices
NND_n	subset of $\mathbb{R}_s^n$ consisting of nonnegative definite (nnd) matrices: $\mathbf{A} \in \mathrm{NND}_n \iff \mathbf{A} = \mathbf{LL}'$ for some $\mathbf{L}$; instead of nnd, the term positive semidefinite is often used
PD_n	subset of NND_n consisting of positive definite (pd) matrices: $\mathbf{A} = \mathbf{LL}'$ for some nonsingular $\mathbf{L}$
$\mathbf{0}$	null vector, null matrix; denoted also as $\mathbf{0}_n$ or $\mathbf{0}_{n\times m}$
$\mathbf{1}_n$	column vector of ones, shortened $\mathbf{1}$
$\mathbf{I}_n$	identity matrix, shortened $\mathbf{I}$
$\mathbf{i}_j$	the jth column of $\mathbf{I}$; the jth standard basis vector
$\mathbf{A} = \{a_{ij}\}$	matrix $\mathbf{A}$ with its elements a_{ij}
$\mathbf{A}_{n\times m}$	$n \times m$ matrix $\mathbf{A}$
$\mathbf{a}$	column vector $\mathbf{a} \in \mathbb{R}^n$
$\mathbf{A}'$	transpose of the matrix $\mathbf{A}$
$(\mathbf{A} : \mathbf{B})$	partitioned (augmented) matrix
$\mathbf{A} = (\mathbf{a}_1 : \ldots : \mathbf{a}_m)$	$\mathbf{A}_{n\times m}$ represented columnwise
$\mathbf{A} = \begin{pmatrix} \mathbf{a}'_{(1)} \\ \vdots \\ \mathbf{a}'_{(n)} \end{pmatrix}$	$\mathbf{A}_{n\times m}$ represented row-wise
$\mathbf{A}^{-1}$	inverse of the matrix $\mathbf{A}$
$\mathbf{A}^{-}$	generalized inverse of the matrix $\mathbf{A}$: $\mathbf{AA}^{-}\mathbf{A} = \mathbf{A}$, also called $\{1\}$-inverse, or inner inverse

$\{\mathbf{A}^-\}$	the set of generalized inverses of $\mathbf{A}$
$\mathbf{A}^-_{12}$	reflexive generalized inverse of $\mathbf{A}$: $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$, $\mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-$, also called $\{12\}$-inverse
$\mathbf{A}^+$	the Moore–Penrose inverse of $\mathbf{A}$: the unique matrix satisfying the four Moore–Penrose conditions: $$\begin{aligned} &(\text{mp1})\ \mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}, && (\text{mp2})\ \mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-, \\ &(\text{mp3})\ (\mathbf{A}\mathbf{A}^-)' = \mathbf{A}\mathbf{A}^-, && (\text{mp4})\ (\mathbf{A}^-\mathbf{A})' = \mathbf{A}^-\mathbf{A} \end{aligned}$$
$\mathbf{A}^-_{ij}$	generalized inverse of $\mathbf{A}$ satisfying the Moore–Penrose conditions (mp$i$) and (mp$j$)
$\mathbf{A}^{1/2}$	symmetric nnd square root of $\mathbf{A} \in \mathrm{NND}_n$: $\mathbf{A}^{1/2} = \mathbf{T}\boldsymbol{\Lambda}^{1/2}\mathbf{T}'$, where $\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}'$ is the eigenvalue decomposition of $\mathbf{A}$
$\mathbf{A}^{+1/2}$	$(\mathbf{A}^+)^{1/2}$
$\mathrm{In}(\mathbf{A}) = (\pi, \nu, \delta)$	inertia of the square matrix $\mathbf{A}$: π, ν, and δ are the number of positive, negative, and zero eigenvalues of $\mathbf{A}$, respectively, all counting multiplicities
$\langle \mathbf{a}, \mathbf{b} \rangle$	standard inner product in $\mathbb{R}^n$: $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}'\mathbf{b}$; can denote also a general inner product in a vector space
$\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbf{V}}$	inner product $\mathbf{a}'\mathbf{V}\mathbf{b}$; $\mathbf{V}$ is the inner product matrix (ipm)
$\mathbf{a} \perp \mathbf{b}$	vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal with respect to a given inner product
$\\|\mathbf{a}\\|$	Euclidean norm (standard norm, 2-norm) of vector $\mathbf{a}$, also denoted $\\|\mathbf{a}\\|_2$: $\\|\mathbf{a}\\|^2 = \mathbf{a}'\mathbf{a}$; can denote also a general vector norm in a vector space
$\\|\mathbf{a}\\|_{\mathbf{V}}$	$\\|\mathbf{a}\\|^2_{\mathbf{V}} = \mathbf{a}'\mathbf{V}\mathbf{a}$, norm when the ipm is $\mathbf{V}$ (ellipsoidal norm)
$\langle \mathbf{A}, \mathbf{B} \rangle$	standard matrix inner product between $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times m}$: $\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{tr}(\mathbf{A}'\mathbf{B}) = \sum_{i,j} a_{ij}b_{ij}$
$\\|\mathbf{A}\\|_F$	Euclidean (Frobenius) norm of the matrix $\mathbf{A}$: $\\|\mathbf{A}\\|^2_F = \mathrm{tr}(\mathbf{A}'\mathbf{A}) = \sum_{i,j} a^2_{ij}$
$\\|\mathbf{A}\\|_2$	matrix 2-norm of the matrix $\mathbf{A}$ (spectral norm): $$\\|\mathbf{A}\\|_2 = \max_{\\|\mathbf{x}\\|_2=1} \\|\mathbf{A}\mathbf{x}\\|_2 = \mathrm{sg}_1(\mathbf{A}) = +\sqrt{\mathrm{ch}_1(\mathbf{A}'\mathbf{A})}$$
$\\|\mathbf{A}^{-1}\\|_2$	matrix 2-norm of nonsingular $\mathbf{A}_{n\times n}$: $\\|\mathbf{A}^{-1}\\|_2 = \frac{1}{\mathrm{sg}_n(\mathbf{A})}$
$\mathrm{cond}(\mathbf{A})$	condition number of nonsingular $\mathbf{A}_{n\times n}$: $\mathrm{cond}(\mathbf{A}) = \\|\mathbf{A}\\|_2\\|\mathbf{A}^{-1}\\|_2 = \mathrm{sg}_1(\mathbf{A})/\,\mathrm{sg}_n(\mathbf{A})$

$\cos(\mathbf{a}, \mathbf{b})$	$\cos \angle(\mathbf{a}, \mathbf{b})$, the cosine of the angle, θ, between the nonzero vectors $\mathbf{a}$ and $\mathbf{b}$: $\cos(\mathbf{a}, \mathbf{b}) = \cos\theta = \cos\angle(\mathbf{a}, \mathbf{b}) = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\\|\mathbf{a}\\| \\|\mathbf{b}\\|}$
$\angle(\mathbf{a}, \mathbf{b})$	the angle, θ, $0 \leq \theta \leq \pi$, between the nonzero vectors $\mathbf{a}$ and $\mathbf{b}$: $\theta = \angle(\mathbf{a}, \mathbf{b}) = \cos^{-1}(\mathbf{a}, \mathbf{b})$
$\mathbf{A}[\alpha, \beta]$	submatrix of $\mathbf{A}_{n \times n}$, obtained by choosing the elements of $\mathbf{A}$ which lie in rows α and columns β; α and β are index sets of the rows and the columns of $\mathbf{A}$, respectively
$\mathbf{A}[\alpha]$	$\mathbf{A}[\alpha, \alpha]$, principal submatrix; same rows and columns chosen
$\mathbf{A}_i^{\mathrm{L}}$	ith leading principal submatrix of $\mathbf{A}_{n \times n}$: $\mathbf{A}_i^{\mathrm{L}} = \mathbf{A}[\alpha, \alpha]$, where $\alpha = \{1, \dots, i\}$
$\mathbf{A}(\alpha, \beta)$	submatrix of $\mathbf{A}$, obtained by choosing the elements of $\mathbf{A}$ which do not lie in rows α and columns β
$\mathbf{A}(i, j)$	submatrix of $\mathbf{A}$, obtained by deleting row i and column j from $\mathbf{A}$
$\mathrm{minor}(a_{ij})$	ijth minor of $\mathbf{A}$ corresponding to a_{ij}: $\mathrm{minor}(a_{ij}) = \det(\mathbf{A}(i, j))$, $i, j \in \{1, \dots, n\}$
$\mathrm{cof}(a_{ij})$	ijth cofactor of $\mathbf{A}$: $\mathrm{cof}(a_{ij}) = (-1)^{i+j} \mathrm{minor}(a_{ij})$
$\det(\mathbf{A})$	determinant of the matrix $\mathbf{A}_{n \times n}$: $\det(a) = a$, $a \in \mathbb{R}$, $\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij} \mathrm{cof}(a_{ij})$, $i \in \{1, \dots, n\}$: the Laplace expansion by minors along the ith row
$\det(\mathbf{A}[\alpha])$	principal minor
$\det(\mathbf{A}_i^{\mathrm{L}})$	leading principal minor of order i
$\lvert\mathbf{A}\rvert$	determinant of the matrix $\mathbf{A}_{n \times n}$
$\mathrm{diag}(\mathbf{A})$	diagonal matrix formed by the diagonal entries of $\mathbf{A}_{n \times n}$
$\mathrm{diag}(d_1, \dots, d_n)$	$n \times n$ diagonal matrix with listed diagonal entries
$\mathrm{diag}(\mathbf{d})$	$n \times n$ diagonal matrix whose ith diagonal element is d_i
$\mathbf{A}_\delta$	diagonal matrix formed by the diagonal entries of $\mathbf{A}_{n \times n}$
$\mathrm{r}(\mathbf{A})$	rank of the matrix $\mathbf{A}$
$\mathrm{rank}(\mathbf{A})$	rank of the matrix $\mathbf{A}$
$\mathrm{tr}(\mathbf{A})$	trace of the matrix $\mathbf{A}_{n \times n}$: $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$
$\mathrm{trace}(\mathbf{A})$	trace of the matrix $\mathbf{A}_{n \times n}$
$\mathrm{vec}(\mathbf{A})$	vectoring operation: the vector formed by placing the columns of $\mathbf{A}$ under one another successively

$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of $\mathbf{A}_{n\times m}$ and $\mathbf{B}_{p\times q}$: $$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1m}\mathbf{B} \\ \vdots & \vdots & \vdots \\ a_{n1}\mathbf{B} & \dots & a_{nm}\mathbf{B} \end{pmatrix} \in \mathbb{R}^{np\times mq}$$
$\mathbf{A}/\mathbf{A}_{11}$	Schur complement of $\mathbf{A}_{11}$ in $\mathbf{A} = \left(\begin{smallmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{smallmatrix}\right)$: $\mathbf{A}/\mathbf{A}_{11} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-}\mathbf{A}_{12}$
$\mathbf{A}_{22\cdot 1}$	$\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-}\mathbf{A}_{12}$
$\mathbf{A} \geq_{\mathsf{L}} \mathbf{0}$	$\mathbf{A}$ is nonnegative definite: $\mathbf{A} = \mathbf{L}\mathbf{L}'$ for some $\mathbf{L}$; $\mathbf{A} \in \mathrm{NND}_n$
$\mathbf{A} >_{\mathsf{L}} \mathbf{0}$	$\mathbf{A}$ is positive definite: $\mathbf{A} = \mathbf{L}\mathbf{L}'$ for some invertible $\mathbf{L}$; $\mathbf{A} \in \mathrm{PD}_n$
$\mathbf{A} \leq_{\mathsf{L}} \mathbf{B}$	$\mathbf{B} - \mathbf{A}$ is nonnegative definite; $\mathbf{B} - \mathbf{A} \in \mathrm{NND}_n$; $\mathbf{A}$ lies below $\mathbf{B}$ with respect to the Löwner ordering
$\mathbf{A} <_{\mathsf{L}} \mathbf{B}$	$\mathbf{B} - \mathbf{A}$ is positive definite; $\mathbf{B} - \mathbf{A} \in \mathrm{PD}_n$
$\mathbf{A} \leq_{\mathsf{rs}} \mathbf{B}$	$\mathbf{A}$ and $\mathbf{B}$ are rank-subtractive; $\mathrm{r}(\mathbf{B} - \mathbf{A}) = \mathrm{r}(\mathbf{B}) - \mathrm{r}(\mathbf{A})$; $\mathbf{A}$ lies below $\mathbf{B}$ with respect to the minus ordering
$\mathrm{Sh}(\mathbf{V} \mid \mathbf{X})$	the shorted matrix of $\mathbf{V} \in \mathrm{NND}_n$ with respect to $\mathbf{X}_{n\times p}$, $\mathrm{Sh}(\mathbf{V} \mid \mathbf{X})$ is the maximal element $\mathbf{U}$ (in the Löwner ordering) in the set $\mathcal{U} = \{\, \mathbf{U} : \mathbf{0} \leq_{\mathsf{L}} \mathbf{U} \leq_{\mathsf{L}} \mathbf{V},\ \mathscr{C}(\mathbf{U}) \subset \mathscr{C}(\mathbf{X}) \,\}$
$\mathbf{P}_{\mathbf{A}}$	orthogonal projector onto $\mathscr{C}(\mathbf{A})$ (w.r.t. $\mathbf{I}$): $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}' = \mathbf{A}\mathbf{A}^{+}$
$\mathbf{P}_{\mathbf{A};\mathbf{V}}$	orthogonal projector onto $\mathscr{C}(\mathbf{A})$ w.r.t. $\mathbf{V} \in \mathrm{PD}_n$: $\mathbf{P}_{\mathbf{A};\mathbf{V}} = \mathbf{A}(\mathbf{A}'\mathbf{V}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}$
$\mathbf{P}_{\mathbf{A};\mathbf{V}}$	generalized orthogonal projector onto $\mathscr{C}(\mathbf{A})$ w.r.t. $\mathbf{V} \in \mathrm{NND}_n$: $\mathbf{P}_{\mathbf{A};\mathbf{V}} = \mathbf{A}(\mathbf{A}'\mathbf{V}\mathbf{A})^{-}\mathbf{A}'\mathbf{V} + \mathbf{A}[\mathbf{I} - (\mathbf{A}'\mathbf{V}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}\mathbf{A}]\mathbf{U}$, where $\mathbf{U}$ is arbitrary
$\mathbf{P}_{\mathbf{A}\mid\mathbf{B}}$	projector onto $\mathscr{C}(\mathbf{A})$ along $\mathscr{C}(\mathbf{B})$: $\mathbf{P}_{\mathbf{A}\mid\mathbf{B}}(\mathbf{A} : \mathbf{B}) = (\mathbf{A} : \mathbf{0})$
$\{\mathbf{P}_{\mathbf{A}\mid\mathbf{B}}\}$	set of matrices satisfying: $\mathbf{P}_{\mathbf{A}\mid\mathbf{B}}(\mathbf{A} : \mathbf{B}) = (\mathbf{A} : \mathbf{0})$
$\mathbf{P}_{\mathcal{U}}$	orthogonal projector onto the vector space $\mathcal{U}$ (w.r.t. a given inner product)
$\mathscr{C}(\mathbf{A})$	column space of the matrix $\mathbf{A}_{n\times p}$: $\mathscr{C}(\mathbf{A}) = \{\, \mathbf{y} \in \mathbb{R}^n : \mathbf{y} = \mathbf{A}\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^p \,\}$
$\mathscr{N}(\mathbf{A})$	null space of the matrix $\mathbf{A}_{n\times p}$: $\mathscr{N}(\mathbf{A}) = \{\, \mathbf{x} \in \mathbb{R}^p : \mathbf{A}\mathbf{x} = \mathbf{0} \,\}$

$\mathscr{C}(\mathbf{A})^{\perp}$	orthocomplement of $\mathscr{C}(\mathbf{A})$ w.r.t. $\mathbf{I}$: $\mathscr{C}(\mathbf{A})^{\perp} = \{\, \mathbf{z} \in \mathbb{R}^n : \mathbf{z}'\mathbf{A}\mathbf{x} = \mathbf{0}\ \forall\, \mathbf{x} \in \mathbb{R}^p \,\} = \mathscr{N}(\mathbf{A}')$
$\mathbf{A}^{\perp}$	matrix whose column space is $\mathscr{C}(\mathbf{A}^{\perp}) = \mathscr{C}(\mathbf{A})^{\perp}$
$\mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}}$	orthocomplement of $\mathscr{C}(\mathbf{A})$ w.r.t. $\mathbf{V}$: $\mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}} = \{\, \mathbf{z} \in \mathbb{R}^n : \mathbf{z}'\mathbf{V}\mathbf{A}\mathbf{x} = \mathbf{0}\ \forall\, \mathbf{x} \in \mathbb{R}^p \,\} = \mathscr{N}(\mathbf{A}'\mathbf{V})$
$\mathbf{A}^{\perp}_{\mathbf{V}}$	matrix whose column space is $\mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}}$
$p_{\mathbf{A}}(x)$	the characteristic polynomial of $\mathbf{A}$: $p_{\mathbf{A}}(x) = \det(\mathbf{A} - x\mathbf{I})$
$\mathcal{U} \subset \mathcal{V}$	$\mathcal{U}$ is a subset of $\mathcal{V}$; possibly $\mathcal{U} = \mathcal{V}$
$\mathcal{U} + \mathcal{V}$	sum of the vector spaces $\mathcal{U}$ and $\mathcal{V}$
$\mathcal{U} \oplus \mathcal{V}$	direct sum of the vector spaces $\mathcal{U}$ and $\mathcal{V}$
$\mathcal{U} \boxplus \mathcal{V}$	direct sum of the orthogonal vector spaces $\mathcal{U}$ and $\mathcal{V}$
$\mathcal{U} \cap \mathcal{V}$	intersection of the vector spaces $\mathcal{U}$ and $\mathcal{V}$
$\mathrm{ch}_i(\mathbf{A}) = \lambda_i$	the ith largest eigenvalue of $\mathbf{A}_{n\times n}$ (all eigenvalues being real)
$\mathrm{ch}(\mathbf{A})$	set of all n eigenvalues of $\mathbf{A}_{n\times n}$, including multiplicities, called also the spectrum of $\mathbf{A}$: $\mathrm{ch}(\mathbf{A}) = \{\mathrm{ch}_1(\mathbf{A}), \dots, \mathrm{ch}_n(\mathbf{A})\}$
$\mathrm{ch}(\mathbf{A}, \mathbf{B})$	set of proper eigenvalues of symmetric $\mathbf{A}_{n\times n}$ with respect to $\mathbf{B} \in \mathrm{NND}_n$; $\lambda \in \mathrm{ch}(\mathbf{A}, \mathbf{B})$ if $\mathbf{A}\mathbf{w} = \lambda\mathbf{B}\mathbf{w}$, $\mathbf{B}\mathbf{w} \neq \mathbf{0}$
$\mathrm{nzch}(\mathbf{A})$	set of the nonzero eigenvalues of $\mathbf{A}_{n\times n}$: $\mathrm{nzch}(\mathbf{A}) = \{\mathrm{ch}_1(\mathbf{A}), \dots, \mathrm{ch}_r(\mathbf{A})\}$, $r = \mathrm{rank}(\mathbf{A})$
$\mathrm{chv}_i(\mathbf{A})$	eigenvector of $\mathbf{A}_{n\times n}$ with respect to $\lambda_i = \mathrm{ch}_i(\mathbf{A})$: a nonzero vector $\mathbf{t}_i$ satisfying the equation $\mathbf{A}\mathbf{t}_i = \lambda_i\mathbf{t}_i$
$\mathrm{sg}_i(\mathbf{A}) = \delta_i$	the ith largest singular value of $\mathbf{A}_{n\times m}$: $\mathrm{sg}_i(\mathbf{A}) = +\sqrt{\mathrm{ch}_i(\mathbf{A}'\mathbf{A})} = +\sqrt{\mathrm{ch}_i(\mathbf{A}\mathbf{A}')}$
$\mathrm{sg}(\mathbf{A})$	set of the singular values of $\mathbf{A}_{n\times m}$ ($m \le n$): $\mathrm{sg}(\mathbf{A}) = \{\mathrm{sg}_1(\mathbf{A}), \dots, \mathrm{sg}_m(\mathbf{A})\}$
$\mathrm{nzsg}(\mathbf{A})$	set of the nonzero singular values of $\mathbf{A}_{n\times m}$: $\mathrm{nzsg}(\mathbf{A}) = \{\mathrm{sg}_1(\mathbf{A}), \dots, \mathrm{sg}_r(\mathbf{A})\}$, $r = \mathrm{rank}(\mathbf{A})$
$\rho(\mathbf{A})$	the spectral radius of $\mathbf{A}_{n\times n}$: the maximum of the absolute values of the eigenvalues of $\mathbf{A}_{n\times n}$
$\mathrm{var_s}(y)$	sample variance of the variable y
$\mathrm{var_d}(\mathbf{y}) = s_y^2$	sample variance: argument is the variable vector $\mathbf{y} \in \mathbb{R}^n$: $\mathrm{var_d}(\mathbf{y}) = \frac{1}{n-1}\mathbf{y}'\mathbf{C}\mathbf{y} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
$\mathrm{cov_s}(x, y)$	sample covariance between the variables x and y

$\mathrm{cov_d}(\mathbf{x},\mathbf{y}) = s_{xy}$	sample covariance: arguments are variable vectors $\in \mathbb{R}^n$: $\mathrm{cov_d}(\mathbf{x},\mathbf{y}) = \frac{1}{n-1}\mathbf{x}'\mathbf{C}\mathbf{y} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_i)(y_i - \bar{y})$
$\mathrm{cor_d}(\mathbf{x},\mathbf{y}) = r_{xy}$	sample correlation: $r_{xy} = \mathbf{x}'\mathbf{C}\mathbf{y}/\sqrt{\mathbf{x}'\mathbf{C}\mathbf{x} \cdot \mathbf{y}'\mathbf{C}\mathbf{y}} = \cos(\mathbf{Cx}, \mathbf{Cy})$
$\bar{\bar{\mathbf{x}}}$	projection of $\mathbf{x}$ onto $\mathscr{C}(\mathbf{1}_n)$: $\bar{\bar{\mathbf{x}}} = \mathbf{Jx} = \bar{x}\mathbf{1}_n$
$\tilde{\mathbf{x}}$	centered $\mathbf{x}$: $\tilde{\mathbf{x}} = \mathbf{Cx} = \mathbf{x} - \mathbf{Jx} = \mathbf{x} - \bar{x}\mathbf{1}_n$
$\mathbf{U}$	$n \times d$ data matrix of the variables $u_1, \dots, u_d$:: $$\mathbf{U} = (\mathbf{u}_1 : \dots : \mathbf{u}_d) = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix}$$
$\mathbf{u}_1, \dots, \mathbf{u}_d$	"variable vectors" in "variable space" $\mathbb{R}^n$
$\mathbf{u}_{(1)}, \dots, \mathbf{u}_{(n)}$	"observation vectors" in "observation space" $\mathbb{R}^d$
$\bar{\mathbf{u}}$	vector of means of the variables $u_1, \dots, u_d$: $\bar{\mathbf{u}} = (\bar{u}_1, \dots, \bar{u}_d)'$
$\widetilde{\mathbf{U}}$	centered $\mathbf{U}$: $\widetilde{\mathbf{U}} = \mathbf{CU}$, $\mathbf{C}$ is the centering matrix
$\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_d$	centered variable vectors
$\tilde{\mathbf{u}}_{(1)}, \dots, \tilde{\mathbf{u}}_{(n)}$	centered observation vectors
$\mathrm{var_d}(\mathbf{u}_i) = s_i^2$	sample variance: argument is the variable vector $\mathbf{u}_i \in \mathbb{R}^n$: $\mathrm{var_d}(\mathbf{u}_i) = \frac{1}{n-1}\mathbf{u}_i'\mathbf{C}\mathbf{u}_i = \frac{1}{n-1}\sum_{\ell=1}^{n}(u_{\ell i} - \bar{u}_i)^2$
$\mathrm{cov_d}(\mathbf{u}_i, \mathbf{u}_j) = s_{ij}$	sample covariance: arguments are variable vectors $\in \mathbb{R}^n$: $s_{ij} = \frac{1}{n-1}\mathbf{u}_i'\mathbf{C}\mathbf{u}_j = \frac{1}{n-1}\sum_{\ell=1}^{n}(u_{\ell i} - \bar{u}_i)(u_{\ell j} - \bar{u}_j)$
$\mathrm{ssp}(\mathbf{U}) = \{t_{ij}\}$	matrix $\mathbf{T}$ ($d \times d$) of the sums of squares and products of deviations about the mean: $\mathbf{T} = \mathbf{U}'\mathbf{C}\mathbf{U} = \sum_{i=1}^{n}(\mathbf{u}_{(i)} - \bar{\mathbf{u}})(\mathbf{u}_{(i)} - \bar{\mathbf{u}})'$
$\mathrm{cov_d}(\mathbf{U}) = \{s_{ij}\}$	sample covariance matrix $\mathbf{S}$ ($d \times d$) of the data matrix $\mathbf{U}$: $\mathbf{S} = \frac{1}{n-1}\mathbf{T} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{u}_{(i)} - \bar{\mathbf{u}})(\mathbf{u}_{(i)} - \bar{\mathbf{u}})'$
$\mathrm{cor_d}(\mathbf{u}_i, \mathbf{u}_j) = r_{ij}$	sample correlation: arguments are variable vectors $\in \mathbb{R}^n$
$\mathrm{cor_d}(\mathbf{U}) = \{r_{ij}\}$	sample correlation matrix $\mathbf{R}$ ($d \times d$) of the data matrix $\mathbf{U}$: $\mathbf{R} = \mathrm{cor_d}(\mathbf{U}) = (\mathrm{diag}\,\mathbf{S})^{-1/2}\mathbf{S}(\mathrm{diag}\,\mathbf{S})^{-1/2}$
$\mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S})$	sample Mahalanobis distance (squared) of the ith observation from the mean: $\mathrm{MHLN}^2(\mathbf{u}_{(i)}, \bar{\mathbf{u}}, \mathbf{S}) = (\mathbf{u}_{(i)} - \bar{\mathbf{u}})'\mathbf{S}^{-1}(\mathbf{u}_{(i)} - \bar{\mathbf{u}})$
$\mathrm{MHLN}^2(\bar{\mathbf{u}}_i, \bar{\mathbf{u}}_j, \mathbf{S}_*)$	sample Mahalanobis distance (squared) between two mean vectors: $\mathrm{MHLN}^2(\bar{\mathbf{u}}_i, \bar{\mathbf{u}}_j, \mathbf{S}_*) = (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j)'\mathbf{S}_*^{-1}(\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j)$, where $\mathbf{S}_* = \frac{1}{n_1+n_2-2}(\mathbf{U}_1'\mathbf{C}_{n_1}\mathbf{U}_1 + \mathbf{U}_2'\mathbf{C}_{n_2}\mathbf{U}_2)$

$\mathrm{MHLN}^2(\mathbf{u}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$	population Mahalanobis distance squared: $\mathrm{MHLN}^2(\mathbf{u}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\mathbf{u} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{u} - \boldsymbol{\mu})$
$\mathrm{E}(\cdot)$	expectation of a random argument: $\mathrm{E}(x) = p_1x_1 + \cdots + p_kx_k$ if x is a discrete random variable whose values are $x_1, \dots, x_k$ with corresponding probabilities $p_1, \dots, p_k$
$\mathrm{var}(x) = \sigma_x^2$	variance of the random variable x: $\sigma_x^2 = \mathrm{E}(x - \mu_x)^2$, $\mu_x = \mathrm{E}(x)$
$\mathrm{cov}(x, y) = \sigma_{xy}$	covariance between the random variables x and y: $\sigma_{xy} = \mathrm{E}(x - \mu_x)(y - \mu_y)$, $\mu_x = \mathrm{E}(x)$, $\mu_y = \mathrm{E}(y)$
$\mathrm{cor}(x, y) = \varrho_{xy}$	correlation between the random variables x and y: $\varrho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x\sigma_y}$
$\mathrm{cov}(\mathbf{x})$	covariance matrix $(d \times d)$ of a d-dimensional random vector $\mathbf{x}$: $\mathrm{cov}(\mathbf{x}) = \boldsymbol{\Sigma} = \mathrm{E}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})(\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})'$
$\mathrm{cor}(\mathbf{x})$	correlation matrix $(d \times d)$ of the random vector $\mathbf{x}$: $\mathrm{cor}(\mathbf{x}) = \boldsymbol{\rho} = (\mathrm{diag}\,\boldsymbol{\Sigma})^{-1/2}\boldsymbol{\Sigma}(\mathrm{diag}\,\boldsymbol{\Sigma})^{-1/2}$
$\mathrm{cov}(\mathbf{x}, \mathbf{y})$	(cross-)covariance matrix between the random vectors $\mathbf{x}$ and $\mathbf{y}$: $\mathrm{cov}(\mathbf{x}, \mathbf{y}) = \mathrm{E}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})(\mathbf{y} - \boldsymbol{\mu}_\mathbf{y})' = \boldsymbol{\Sigma}_{\mathbf{xy}}$
$\mathrm{cov}(\mathbf{x}, \mathbf{x})$	$\mathrm{cov}(\mathbf{x}, \mathbf{x}) = \mathrm{cov}(\mathbf{x})$
$\mathrm{cor}(\mathbf{x}, \mathbf{y})$	(cross-)correlation matrix between the random vectors $\mathbf{x}$ and $\mathbf{y}$
$\mathrm{cov}\binom{\mathbf{x}}{y}$	partitioned covariance matrix of the random vector $\binom{\mathbf{x}}{y}$: $$\mathrm{cov}\begin{pmatrix}\mathbf{x}\\ y\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y}\\ \boldsymbol{\sigma}'_{\mathbf{x}y} & \sigma_y^2\end{pmatrix} = \begin{pmatrix}\mathrm{cov}(\mathbf{x}, \mathbf{x}) & \mathrm{cov}(\mathbf{x}, y)\\ \mathrm{cov}(\mathbf{x}, y)' & \mathrm{var}(y)\end{pmatrix}$$
$\mathbf{x} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$, $\mathrm{cov}(\mathbf{x}) = \boldsymbol{\Sigma}$
$\mathbf{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\mathbf{x}$ follows the p-dimensional normal distribution $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
$n(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	density for $\mathbf{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ pd: $$n(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}\lvert\boldsymbol{\Sigma}\rvert^{1/2}}\mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$
$\mathrm{cc}_i(\mathbf{x}, \mathbf{y})$	ith largest canonical correlation between the random vectors $\mathbf{x}$ and $\mathbf{y}$
$\mathrm{cc}(\mathbf{x}, \mathbf{y})$	set of the canonical correlations between the random vectors $\mathbf{x}$ and $\mathbf{y}$

$\mathrm{cc}_+(\mathbf{x}, \mathbf{y})$ set of the nonzero (necessarily positive) canonical correlations between the random vectors $\mathbf{x}$ and $\mathbf{y}$; square roots of the nonzero eigenvalues of $\mathbf{P_A P_B}$:

$$\begin{aligned}\mathrm{cc}_+(\mathbf{x}, \mathbf{y}) &= \mathrm{nzch}^{1/2}(\mathbf{P_A P_B}) \\ &= \mathrm{nzsg}[(\mathbf{A}'\mathbf{A})^{+1/2}\mathbf{A}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{+1/2}],\end{aligned}$$

$$\mathrm{cov}\begin{pmatrix}\mathbf{x} \\ \mathbf{y}\end{pmatrix} = \begin{pmatrix}\mathbf{A}'\mathbf{A} & \mathbf{A}'\mathbf{B} \\ \mathbf{B}'\mathbf{A} & \mathbf{B}'\mathbf{B}\end{pmatrix}$$

$\mathbf{X} = (\mathbf{1} : \mathbf{X}_0)$ in regression context often the model matrix

$\mathbf{X}_0$ $n \times k$ data matrix of the x-variables:

$$\mathbf{X}_0 = (\mathbf{x}_1 : \ldots : \mathbf{x}_k) = \begin{pmatrix}\mathbf{x}'_{(1)} \\ \vdots \\ \mathbf{x}'_{(n)}\end{pmatrix}$$

$\mathbf{x}_1, \ldots, \mathbf{x}_k$ variable vectors in the variable space $\mathbb{R}^n$

$\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)}$ observation vectors in the observation space $\mathbb{R}^k$

$\mathrm{ssp}(\mathbf{X}_0 : \mathbf{y})$ partitioned matrix of the sums of squares and products of deviations about the mean of data $(\mathbf{X}_0 : \mathbf{y})$:

$$\mathrm{ssp}(\mathbf{X}_0 : \mathbf{y}) = \begin{pmatrix}\mathbf{T_{xx}} & \mathbf{t}_{\mathbf{x}y} \\ \mathbf{t}'_{\mathbf{x}y} & t_{yy}\end{pmatrix} = (\mathbf{X}_0 : \mathbf{y})'\mathbf{C}(\mathbf{X}_0 : \mathbf{y})$$

$\mathrm{cov_d}(\mathbf{X}_0 : \mathbf{y})$ partitioned sample covariance matrix of data $(\mathbf{X}_0 : \mathbf{y})$:

$$\mathrm{cov_d}(\mathbf{X}_0 : \mathbf{y}) = \begin{pmatrix}\mathbf{S_{xx}} & \mathbf{s}_{\mathbf{x}y} \\ \mathbf{s}'_{\mathbf{x}y} & s_y^2\end{pmatrix}$$

$\mathrm{cor_d}(\mathbf{X}_0 : \mathbf{y})$ partitioned sample correlation matrix of data $(\mathbf{X}_0 : \mathbf{y})$:

$$\mathrm{cor_d}(\mathbf{X}_0 : \mathbf{y}) = \begin{pmatrix}\mathbf{R_{xx}} & \mathbf{r}_{\mathbf{x}y} \\ \mathbf{r}'_{\mathbf{x}y} & 1\end{pmatrix}$$

$\mathbf{H}$ orthogonal projector onto $\mathscr{C}(\mathbf{X})$, the hat matrix: $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \mathbf{X}\mathbf{X}^+ = \mathbf{P_X}$

$\mathbf{M}$ orthogonal projector onto $\mathscr{C}(\mathbf{X})^\perp$: $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$

$\mathbf{J}$ the orthogonal projector onto $\mathscr{C}(\mathbf{1}_n)$: $\mathbf{J} = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' = \mathbf{P}_{\mathbf{1}_n}$

$\mathbf{C}$ centering matrix, the orthogonal projector onto $\mathscr{C}(\mathbf{1}_n)^\perp$: $\mathbf{C} = \mathbf{I}_n - \mathbf{J}$

$(\mathbf{X}_1 : \mathbf{X}_2)$ partitioned model matrix $\mathbf{X}$

$\mathbf{M}_1$	orthogonal projector onto $\mathscr{C}(\mathbf{X}_1)^\perp$: $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_1}$
$\hat{\boldsymbol{\beta}}$	solution to normal equation $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$, OLSE($\boldsymbol{\beta}$)
$\mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\mathbf{y}}$	$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ = OLS fitted values, OLSE($\mathbf{X}\boldsymbol{\beta}$), denoted also $\widehat{\mathbf{X}\boldsymbol{\beta}} = \hat{\boldsymbol{\mu}}$, when $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$
$\tilde{\boldsymbol{\beta}}$	solution to generalized normal equation $\mathbf{X}'\mathbf{W}^-\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{W}^-\mathbf{y}$, where $\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{X}'$, $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$
$\tilde{\boldsymbol{\beta}}$	if $\mathbf{V}$ is positive definite and $\mathbf{X}$ has full column rank, then $\tilde{\boldsymbol{\beta}} = \mathrm{BLUE}(\boldsymbol{\beta}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$
$\mathbf{X}\tilde{\boldsymbol{\beta}}$	BLUE($\mathbf{X}\boldsymbol{\beta}$), denoted also $\widetilde{\mathbf{X}\boldsymbol{\beta}} = \tilde{\boldsymbol{\mu}}$
$\bar{y}$	mean of the response variable y: $\bar{y} = (y_1 + \cdots + y_n)/n$
$\bar{\mathbf{x}}$	vector of the means of k regressor variables $x_1, \dots, x_k$: $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_k)' \in \mathbb{R}^k$
$\bar{\bar{\mathbf{y}}}$	projection of $\mathbf{y}$ onto $\mathscr{C}(\mathbf{1}_n)$: $\bar{\bar{\mathbf{y}}} = \mathbf{J}\mathbf{y} = \bar{y}\mathbf{1}_n$
$\tilde{\mathbf{y}}$	centered $\mathbf{y}$, $\tilde{\mathbf{y}} = \mathbf{C}\mathbf{y} = \mathbf{y} - \bar{\bar{\mathbf{y}}}$
$\hat{\boldsymbol{\beta}}_{\mathbf{x}}$	$\hat{\boldsymbol{\beta}}_{\mathbf{x}} = \mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{t}_{\mathbf{x}y} = \mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{s}_{\mathbf{x}y}$: the OLS-regression coefficients of x-variables when $\mathbf{X} = (\mathbf{1} : \mathbf{X}_0)$
$\hat{\beta}_0$	$\hat{\beta}_0 = \bar{y} - \hat{\boldsymbol{\beta}}_{\mathbf{x}}'\bar{\mathbf{x}} = \bar{y} - (\hat{\beta}_1\bar{x}_1 + \cdots + \hat{\beta}_k\bar{x}_k)$: OLSE of the constant term (intercept) when $\mathbf{X} = (\mathbf{1} : \mathbf{X}_0)$
BLP($\mathbf{y}$; $\mathbf{x}$)	the best linear predictor of the random vector $\mathbf{y}$ on the basis of the random vector $\mathbf{x}$
BLUE($\mathbf{K}'\boldsymbol{\beta}$)	the best linear unbiased estimator of estimable parametric function $\mathbf{K}'\boldsymbol{\beta}$, denoted as $\mathbf{K}'\tilde{\boldsymbol{\beta}}$ or $\widetilde{\mathbf{K}'\boldsymbol{\beta}}$
BLUP($\mathbf{y}_f$; $\mathbf{y}$)	the best linear unbiased predictor of a new unobserved $\mathbf{y}_f$
LE($\mathbf{K}'\boldsymbol{\beta}$; $\mathbf{y}$)	(homogeneous) linear estimator of $\mathbf{K}'\boldsymbol{\beta}$, where $\mathbf{K} \in \mathbb{R}^{p\times q}$: $\{\mathrm{LE}(\mathbf{K}'\boldsymbol{\beta};\mathbf{y})\} = \{\, \mathbf{A}\mathbf{y} : \mathbf{A} \in \mathbb{R}^{q\times n} \,\}$
LP($\mathbf{y}$; $\mathbf{x}$)	(inhomogeneous) linear predictor of the p-dimensional random vector $\mathbf{y}$ on the basis of the q-dimensional random vector $\mathbf{x}$: $\{\mathrm{LP}(\mathbf{y};\mathbf{x})\} = \{\, f(\mathbf{x}) : f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{a},\ \mathbf{A} \in \mathbb{R}^{p\times q},\ \mathbf{a} \in \mathbb{R}^p \,\}$
LUE($\mathbf{K}'\boldsymbol{\beta}$; $\mathbf{y}$)	(homogeneous) linear unbiased estimator of $\mathbf{K}'\boldsymbol{\beta}$: $\{\mathrm{LUE}(\mathbf{K}'\boldsymbol{\beta};\mathbf{y})\} = \{\, \mathbf{A}\mathbf{y} : \mathrm{E}(\mathbf{A}\mathbf{y}) = \mathbf{K}'\boldsymbol{\beta} \,\}$
LUP($\mathbf{y}_f$; $\mathbf{y}$)	linear unbiased predictor of a new unobserved $\mathbf{y}_f$: $\{\mathrm{LUP}(\mathbf{y}_f;\mathbf{y})\} = \{\, \mathbf{A}\mathbf{y} : \mathrm{E}(\mathbf{A}\mathbf{y} - \mathbf{y}_f) = \mathbf{0} \,\}$
MSEM($f(\mathbf{x})$; $\mathbf{y}$)	mean squared error matrix of $f(\mathbf{x})$ (= random vector, function of the random vector $\mathbf{x}$) with respect to $\mathbf{y}$ (=

	random vector or a given fixed vector): $\mathrm{MSEM}[f(\mathbf{x}); \mathbf{y}] = \mathrm{E}[\mathbf{y} - f(\mathbf{x})][\mathbf{y} - f(\mathbf{x})]'$
$\mathrm{MSEM}(\mathbf{Fy}; \mathbf{K}'\boldsymbol{\beta})$	mean squared error matrix of the linear estimator $\mathbf{Fy}$ under $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}$ with respect to $\mathbf{K}'\boldsymbol{\beta}$: $\mathrm{MSEM}(\mathbf{Fy}; \mathbf{K}'\boldsymbol{\beta}) = \mathrm{E}(\mathbf{Fy} - \mathbf{K}'\boldsymbol{\beta})(\mathbf{Fy} - \mathbf{K}'\boldsymbol{\beta})'$
$\mathrm{OLSE}(\mathbf{K}'\boldsymbol{\beta})$	the ordinary least squares estimator of parametric function $\mathbf{K}'\boldsymbol{\beta}$, denoted as $\mathbf{K}'\hat{\boldsymbol{\beta}}$ or $\widehat{\mathbf{K}'\boldsymbol{\beta}}$; here $\hat{\boldsymbol{\beta}}$ is any solution to the normal equation $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$
$\mathrm{risk}(\mathbf{Fy}; \mathbf{K}'\boldsymbol{\beta})$	quadratic risk of $\mathbf{Fy}$ under $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}$ with respect to $\mathbf{K}'\boldsymbol{\beta}$: $\mathrm{risk}(\mathbf{Fy}; \mathbf{K}'\boldsymbol{\beta}) = \mathrm{tr}[\mathrm{MSEM}(\mathbf{Fy}; \mathbf{K}'\boldsymbol{\beta})] = \mathrm{E}(\mathbf{Fy} - \mathbf{K}'\boldsymbol{\beta})'(\mathbf{Fy} - \mathbf{K}'\boldsymbol{\beta})$
$\mathscr{M}$	linear model: $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}$: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\mathrm{cov}(\mathbf{y}) = \mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$, $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$
$\mathscr{M}_{\mathrm{mix}}$	mixed linear model: $\mathscr{M}_{\mathrm{mix}} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma},\, \mathbf{D},\, \mathbf{R}\}$: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$; $\boldsymbol{\gamma}$ is the vector of the random effects, $\mathrm{cov}(\boldsymbol{\gamma}) = \mathbf{D}$, $\mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{R}$, $\mathrm{cov}(\boldsymbol{\gamma}, \boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$
$\mathscr{M}_f$	linear model with new future observations $\mathbf{y}_f$:

$$\mathscr{M}_f = \left\{ \begin{pmatrix} \mathbf{y} \\ \mathbf{y}_f \end{pmatrix}, \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{X}_f\boldsymbol{\beta} \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathbf{V} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \right\}$$

Lähteet

Abadir, K. M. & Magnus, J. R. (2005). *Matrix Algebra*. Cambridge University Press. [202, 211]

Aitken, A. C. (1935). On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh, Section A*, 55, 42–49. [206]

Aitken, A. C. (1939). *Determinants and Matrices*. Oliver & Boyd. 2nd–9th editions, 1942–1956; 9th edition, reset & reprinted, 1967. [206]

Albert, A. (1969). Conditions for positive and nonnegative definiteness in terms of pseudoinverses. *SIAM Journal on Applied Mathematics*, 17, 434–440. [197, 348]

Albert, A. (1972). *Regression and the Moore–Penrose Pseudoinverse*. Academic Press. [197]

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, Third Ed. Wiley. [9, 275]

Baksalary, J. K. (1987). Algebraic characterizations and statistical implications of the commutativity of orthogonal projectors. In *Proceedings of the Second International Tampere Conference in Statistics* (T. Pukkila & S. Puntanen, eds.), Dept. of Mathematical Sciences, University of Tampere, pp. 113–142. [332]

Baksalary, J. K. & Kala, R. (1979). Two relations between oblique and Λ-orthogonal projectors. *Linear Algebra and its Applications*, 24, 99–103. [332]

Baksalary, O. M. & Trenkler, G. (2009). A projector oriented approach to the best linear unbiased estimator. *Statistical Papers*, 50, 721–733. [332]

Bapat, R. B. (2000). *Linear Algebra and Linear Models*, Second Ed. Springer. [306]

Bapat, R. B. (2012). *Linear Algebra and Linear Models*, Third Ed. Hindustan Book Agency, Springer. [199]

Ben-Israel, A. (2002). The Moore of the Moore–Penrose inverse. *Electronic Journal of Linear Algebra*, 9, 150–157. [337]

Ben-Israel, A. & Greville, T. N. E. (2003). *Generalized Inverses: Theory and Applications*, Second Ed. Springer. [29, 203, 332, 335, 337]

Bryant, P. (1984). Geometry, statistics, probability: variations on a common theme. *The American Statistician*, 38, 38–48. [332]

Carlson, D. (1986). What are Schur complements, anyway? *Linear Algebra and its Applications*, 74, 257–275. [145]

Casella, G. (2008). *Statistical Design*. Springer. [44]

Chipman, J. S. & Rao, M. M. (1964). Projections, generalized inverses and quadratic forms. *Journal of Mathematical Analysis and Applications*, 9, 1–11. [332]

Christensen, R. (2011). *Plane Answers to Complex Questions: The Theory of Linear Models*, Fourth Ed. Springer. [228]

Cottle, R. W. (1974). Manifestations of the Schur complement. *Linear Algebra and its Applications*, 8, 189–211. See also *Rendiconti del Seminario Matematico e Fisico di Milano*, 45 (1975), 31–40. [145]

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. [201]

Das Gupta, S. (1993). The evolution of the D^2-statistic of Mahalanobis. *Sankhyā, Ser. A*, 55, 442–459. [28]

Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211–218. [203]

Feller, W. (1957). *An Introduction to Probability Theory and Its Applications*, Second Ed. Wiley. [65]

Galántai, A. (2004). *Projectors and Projection Methods*. Kluwer Academic Publishers. [332]

Galántai, A. (2008). Subspaces, angles and pairs of orthogonal projections. *Linear and Multilinear Algebra*, 56, 227–260. [332]

Gelman, A. (2004). A linear regression example, and a question. URL http://andrewgelman.com/2004/11/22/a_linear_regres/. [73]

Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations*, Third Ed. Johns Hopkins University Press. [177]

Groß, J. & Trenkler, G. (1998a). On the equality of OLSE and BLUE in a linear model with partitioned data. In *Frontiers in Probability and Statistics* (S. P. Mukherjee, S. K. Basu & B. K. Sinha, eds.), Narosa Publishing House, pp. 189–194. [251]

Groß, J. & Trenkler, G. (1998b). On the product of oblique projectors. *Linear and Multilinear Algebra*, 44, 247–259. [332]

Groß, J., Trenkler, G. & Troschke, S.-O. (1997). Problem no. 10519. *The American Mathematical Monthly*, 103, 347. [333]

Gustafson, K. (2012). *Antieigenvalue Analysis*. World Scientific. [170]

Halmos, P. R. (1951). *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*. Chelsea Publishing Company. [332]

Halmos, P. R. (1958). *Finite-Dimensional Vector Spaces*, Second Ed. Van Nostrand. Reprinted by Springer, 1974. [332]

Halmos, P. R. (1991). Bad products of good matrices. *Linear and Multilinear Algebra*, 29, 1–20. [126]

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer. [202, 211]

Haynsworth, E. V. (1968a). Determination of the inertia of a partitioned Hermitian matrix. *Linear Algebra and its Applications*, 1, 73–81. [145, 146]

Haynsworth, E. V. (1968b). On the Schur complement. *Basle Mathematical Notes*, #BMN 20, 17 pages. [145, 146]

Henderson, H. V. & Searle, S. R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *The Canadian Journal of Statistics*, 7, 65–81. [211]

Henderson, H. V. & Searle, S. R. (1981a). On deriving the inverse of a sum of matrices. *SIAM Review*, 23, 53–60. [145, 147]

Henderson, H. V. & Searle, S. R. (1981b). The vec-permutation matrix, the vec operator and Kronecker products: a review. *Linear and Multilinear Algebra*, 9, 271–288. [211]

Horn, R. A. & Johnson, C. R. (1990). *Matrix Analysis*. Cambridge University Press. Corrected reprint of the 1985 original. [177, 185, 212]

Horn, R. A. & Olkin, I. (1996). When does $A^*A = B^*B$ and why does one want to know? *The American Mathematical Monthly*, 103, 470–482. [177]

Householder, A. S. & Young, G. (1938). Matrix approximation and latent roots. *The American Mathematical Monthly*, 45, 165–171. [203]

Hubert, L., Meulman, J. & Heiser, W. (2000). Two purposes for matrix factorization: a historical appraisal. *SIAM Review*, 42, 68–82. [203]

Johnson, R. A. & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6th Ed. Pearson Prentice Hall. [275]

Kala, R. (1981). Projectors and linear estimation in general linear models. *Communications in Statistics: Theory and Methods*, 10, 849–873. [332]

Kala, R. (2008). On commutativity of projectors. *Discussiones Mathematicae – Probability and Statistics*, 28, 157–165. [332]

Kala, R. & Pordzik, P. (2006). Two local operators and the BLUE. *Linear Algebra and its Applications*, 417, 134–139. [332]

Liski, E. P. & Puntanen, S. (1989). A further note on a theorem on the difference of the generalized inverses of two nonnegative definite matrices. *Communications in Statistics: Theory and Methods*, 18, 1747–1751. [348]

Löwner, K. (1934). Über monotone Matrixfunktionen. *Mathematische Zeitschrift*, 38, 443–446. [195]

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2, 49–55. [28]

Mäkeläinen, T. (1970a). Extrema for characteristic roots of product matrices. *Commentationes Physico-Mathematicae, Societas Scientiarum Fennica*, 38, 27–53. [202]

Mäkeläinen, T. (1970b). Projections and generalized inverses in the general linear model. *Commentationes Physico-Mathematicae, Societas Scientiarum Fennica*, 38, 13–25. [332]

Markiewicz, A. & Puntanen, S. (2015). All about the $\perp$ with its applications in the linear statistical models. *Open Mathematics (Open access)*, 13, 33–50. [90, 323, 342]

Marsaglia, G. & Styan, G. P. H. (1974a). Equalities and inequalities for ranks of matrices. *Linear and Multilinear Algebra*, 2, 269–292. [329, 330]

Marsaglia, G. & Styan, G. P. H. (1974b). Rank conditions for generalized inverses of partitioned matrices. *Sankhyā, Ser. A*, 36, 437–442. [347]

Marshall, A. W., Olkin, I. & Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*, Second Ed. Springer. [195]

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics (SIAM). [29]

Mitra, S. K. (1993). Transformations of the range space and the direct orthogonal projector. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift* (J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy & B. L. S. Prakasa Rao, eds.), Wiley Eastern, pp. 463–474. [332]

Mitra, S. K. & Rao, C. R. (1974). Projections under seminorms and generalized Moore–Penrose inverses. *Linear Algebra and its Applications*, 9, 155–167. [332]

Moore, E. H. (1920). On the reciprocal of the general algebraic matrix (Abstract). *Bulletin of the American Mathematical Society*, 26, 890–891. [337]

Moore, E. H. (1935). *General Analysis, Part 1* (with editorial collaboration of R. W. Barnard). Memoirs of the American Philosophical Society, Vol. 1 [Cf. pp. 197–209]. [337]

Mustonen, S. (1995). *Tilastolliset monimuuttujamenetelmät* [Statistical Multivariate Methods, in Finnish]. Survo Systems. [9, 251]

Mustonen, S. (1996). *Survo ja minä*. Survo Systems. [66]

Mustonen, S. (1997). A measure for total variability in multivariate normal distribution. *Computational Statistics & Data Analysis*, 23, 321–334. [276]

O'Hagan, T. (2012). A thing of beauty. *Significance*, 9, 26–28. [44]

Olkin, I. (1992). A matrix formulation on how deviant an observation can be. *The American Statistician*, 46, 205–209. [246]

Ouellette, D. V. (1981). Schur complements and statistics. *Linear Algebra and its Applications*, 36, 187–295. [145]

Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51, 406–413. [337]

Piziak, R. & Odell, P. L. (2007). *Matrix Theory: From Generalized Inverses to Jordan Form*. Chapman & Hall/CRC. [335]

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley. Reprinted by SIAM in series Classics in Applied Mathematics, 2006. [197]

Puntanen, S., Seber, G. A. F. & Styan, G. P. H. (2013a). Multivariate statistical analysis. In *Handbook of Linear Algebra* (L. Hogben, ed.), Chapman & Hall/CRC, Chap. 70, pp. 70.1–70.17. [61]

Puntanen, S., Seber, G. A. F. & Styan, G. P. H. (2013b). Random vectors and linear statistical models. In *Handbook of Linear Algebra* (L. Hogben, ed.), Chapman & Hall/CRC, Chap. 71, pp. 71.1–71.16. [61]

Puntanen, S. & Styan, G. P. H. (1989). The equality of the ordinary least squares estimator and the best linear unbiased estimator. *The American Statistician*, 43, 153–164. With comments by Oscar Kempthorne and Shayle R. Searle and a reply by the authors. [206]

Puntanen, S. & Styan, G. P. H. (2005a). Historical introduction: Issai Schur and the early development of the Schur complement. In *The Schur Complement and Its Applications* (F. Zhang, ed.), Springer, Chap. 0 & Bibliography, pp. 1–16, 259–288. [145]

Puntanen, S. & Styan, G. P. H. (2005b). Schur complements in statistics and probability. In *The Schur Complement and Its Applications* (F. Zhang, ed.), Springer, Chap. 6 & Bibliography, pp. 163–226, 259–288. [145]

Puntanen, S., Styan, G. P. H. & Isotalo, J. (2011). *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty*. Springer. [2, 111, 140, 170, 184, 197, 203, 287, 288, 321, 323, 330, 333, 334, 335, 338, 345, 347, 348]

Puntanen, S., Styan, G. P. H. & Isotalo, J. (2013). *Formulas Useful for Linear Regression Analysis and Related Matrix Theory: It's Only Formulas But We Like Them*. Springer. [2]

Rao, A. R. & Bhimasankaram, P. (2000). *Linear Algebra*, Second Ed. Hindustan Book Agency. [329]

Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, California, 1965/66), Vol. I: Statistics* (L. M. Le Cam & J. Neyman, eds.), University of California Press, Berkeley, pp. 355–372. [154]

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, Second Ed. Wiley. [54, 161, 202, 228, 236, 329]

Rao, C. R. (1974). Projectors, generalized inverses and the BLUE's. *Journal of the Royal Statistical Society, Ser. B*, 36, 442–448. [332]

Rao, C. R. (1980). Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In *Multivariate Analysis – V* (P. R. Krishnaiah, ed.), North-Holland, pp. 3–22. [202]

Rao, C. R. (2007). Antieigenvalues and antisingularvalues of a matrix and applications to problems in statistics. *Mathematical Inequalities & Applications*, 10, 471–489. [170]

Rao, C. R. & Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*. Wiley. [332, 335, 336]

Rao, C. R. & Rao, M. B. (1998). *Matrix Algebra and Its Applications to Statistics and Econometrics*. World Scientific. [329]

Rao, C. R. & Yanai, H. (1979). General definition and decomposition of projectors and some applications to statistical problems. *Journal of Statistical Planning and Inference*, 3, 1–17. [332]

Samuelson, P. A. (1968). How deviant can you be? *Journal of the American Statistical Association*, 63, 1522–1525. [246]

Schmidt, E. (1907). Zur Theorie der linearen und nichtlinearen Integralgleichungen: I. Teil: Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener. *Mathematische Annalen*, 63, 433–476. [203]

Schott, J. R. (2005). *Matrix Analysis for Statistics*, Second Ed. Wiley. [162]

Schur, J. (1917). Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind [I]. *Journal für die reine und angewandte Mathematik*, 147, 205–232. [145]

Scott, A. J. & Styan, G. P. H. (1985). On a separation theorem for generalized eigenvalues and a problem in the analysis of sample surveys. *Linear Algebra and its Applications*, 70, 209–224. [202]

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. Wiley. [118, 161, 162, 177]

Searle, S. R., Casella, G. & McCulloch, C. E. (1992). *Variance Components*. Wiley. [228]

Seber, G. A. F. (1980). *The Linear Hypothesis: A General Theory*, Second Ed. Griffin. [332]

Seber, G. A. F. (1984). *Multivariate Observations*. Wiley. [9, 14, 173, 332]

Seber, G. A. F. (2008). *A Matrix Handbook for Statisticians*. Wiley. [38, 54, 211, 236, 242]

Seber, G. A. F. (2015). *The Linear Model and Hypothesis: A General Unifying Theory*. Springer. [332]

Seber, G. A. F. & Lee, A. J. (2003). *Linear Regression Analysis*, Second Ed. Wiley. [332]

Sherman, J. & Morrison, W. J. (1949). Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix (abstract). *The Annals of Mathematical Statistics*, 20, 621. [147]

Sherman, J. & Morrison, W. J. (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21, 124–127. [147]

Sibuya, M. (1970). Subclasses of generalized inverses of matrices. *Annals of the Institute of Statistical Mathematics*, 22, 543–556. [332]

Speed, T. P. (2012a). Terence's stuff: Multiple linear regression, part 1. *The IMS Bulletin*, 41 (5), 13. [12]

Speed, T. P. (2012b). Terence's stuff: Multiple linear regression, part 2. *The IMS Bulletin*, 41 (6), 19. [12]

Speed, T. P. (2013). Terence's stuff: Least-but-not-last squares. *The IMS Bulletin*, 42 (3), 19. [12]

Stewart, G. W. (1993). On the early history of the singular value decomposition. *SIAM Review*, 35, 551–566. [177, 203]

Stewart, G. W. (1998). *Matrix Algorithms. Volume I: Basic Decompositions*. Society for Industrial and Applied Mathematics (SIAM). [177]

Stigler, S. M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press. [12]

Styan, G. P. H. (1985). Schur complements and linear statistical models. In *Proceedings of the First International Tampere Seminar on Linear Statistical Models and their Applications* (T. Pukkila & S. Puntanen, eds.), Dept. of Mathematical Sciences, University of Tampere, pp. 37–75. [145]

Takane, Y. & Yanai, H. (1999). On oblique projectors. *Linear Algebra and its Applications*, 289, 297–310. [332]

Takeuchi, K., Yanai, H. & Mukherjee, B. N. (1982). *The Foundations of Multivariate Analysis: A Unified Approach by Means of Projection onto Linear Subspaces*. Wiley, Halsted Press. [332]

Thompson, W. R. (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *The Annals of Mathematical Statistics*, 6, 214–219. [246]

Trenkler, D. & Trenkler, G. (2001). Magic squares, melancholy and the Moore–Penrose inverse. *Image, Bulletin of the International Linear Algebra Society*, 27, 3–10. [72]

Trenkler, G. (1994). Characterizations of oblique and orthogonal projectors. In *Proceedings of the International Conference on Linear Statistical Inference LINSTAT '93 (Poznań, 1993)* (T. Caliński & R. Kala, eds.), Kluwer Academic Publishers, pp. 255–270. [332]

Trenkler, G. (2006). On oblique and orthogonal projectors. In *Contributions to Probability and Statistics: Applications and Challenges – Proceedings of the International Statistics Workshop held at University of Canberra, 4–5 April 2005* (P. Brown, S. Liu & D. Sharma, eds.), World Scientific, pp. 178–191. [332]

Vehkalahti, K. (2000). Reliability of measurement scales: Tarkkonen's general method supersedes Cronbach's alpha. Ph.D. Thesis, Department of Statistics, University of Helsinki, Statistical Research Reports 17, Finnish Statistical Society. [201]

Vehkalahti, K., Puntanen, S. & Tarkkonen, L. (2009). Implications of dimensionality on measurement reliability. In *Statistical Inference, Econometric Analysis and Matrix Algebra: Festschrift in Honour of Götz Trenkler* (B. Schipp & W. Krämer, eds.), Physica-Verlag, pp. 143–160. [201]

Vehkalahti, K. & Sund, R. (2015). Solving survo puzzles using matrix combinatorial products. *Journal of Statistical Computation and Simulation*, 85, 2666–2681. [72]

Wolkowicz, H. & Styan, G. P. H. (1979). Extensions of Samuelson's inequality. *The American Statistician*, 33, 143–144. [246]

Woodbury, M. A. (1950). *Inverting Modified Matrices*. Princeton University Press. [147]

Yanai, H., Takeuchi, K. & Takane, Y. (2011). *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*. Springer. [332]

Zhang, F., ed. (2005). *The Schur Complement and Its Applications*. Springer. [145]

Zyskind, G. & Martin, F. B. (1969). On best linear estimation and general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure. *SIAM Journal on Applied Mathematics*, 17, 1190–1202. [329]

Hakemisto

E

F

G

H

I

J

K

P

R

Henkilöhakemisto

V

W

Y

Z