

Aapo Loukaskorpi

VERENPAINEN JA SYKKEEN MALLINTAMINEN MANCOVA-MALLILLA

Tiivistelmä

Aapo Loukaskorpi: Verenpaineen ja sykkeen mallintaminen MANCOVA-mallilla

Kandidaattitutkielma

Tampereen yliopisto

Matematiikan ja tilastollisen data-analyysin tutkinto-ohjelma

Toukokuu 2024

Monen muuttujan kovarianssianalyysi eli MANCOVA on yleinen lineaarinen malli. Se on kovarianssianalyysin eli ANCOVAn laajennus. Toisin kuin ANCOVA:ssa selitettäviä muuttujia on nyt vähintään kaksi. MANCOVA antaa näin tehokkaamman tavan tutkia useamman muuttujan yhteisvaikutusta samanaikaisesti.

Tämä kandidaatintyö voidaan jakaa kahteen osioon. Ensimmäisessä osassa esitellään MANCOVA-malli teoreettisesti, sen parametrien estimointi sekä hypoteesien testaus. Tämän lisäksi esitellään pääkomponenttianalyysi sekä kuinka pääkomponentit määritellään. Toisessa osassa tutkitaan käytännön aineistoa MANCOVA-mallin avulla. Aineistona käytetään dataa ambulatorisesta verenpaine tutkimuksesta. Ensimmäiseksi esitellään aineistoa ja analysoidaan testauksessa käytettyjä muuttujia. Tämän jälkeen suoritetaan pääkomponenttianalyysi, jolla saadaan aineistosta tiiviimpi. Testauksessa tutkitaan, kuinka mallin kovariaatit vaikuttavat selitettäviin muuttujiin.

Aineiston koehenkilöiden keskiverenpaine sekä syke toimivat mallissa selitettävänä muuttujina. Selittävinä muuttujina eli kovariaatteina toimivat koehenkilöiden tunnetilat sekä heidän verenpaineryhmänsä. Koehenkilöt on jaettu kolmeen eri verenpaineryhmään ja nämä verenpaineryhmät toimivat mallissa kategorisena muuttujan eli faktorina. Näistä muuttujista muodostetaan MANCOVA-malli ja tutkitaan, onko koehenkilöiden tunnetiloilla tilastollista vaikutusta heidän verenpaineeseensa sekä sykkeeseen. Tutkitaan myös graafisesti, kuinka mallin residuaalit käyttäytyvät.

Pääkomponenttianalyysin tuloksista saadut stressi- ja kiinnostuneisuusmuuttujat eivät vaikuta merkittävästi keskipaineen ja sykkeen vaihteluihin. Verenpaineryhmien vaikutus mallissa on hyvinkin selkeä. Tutkitaan vielä toista mallia, josta verenpaineryhmät on poistettu. Saaduissa tuloksissa on eroavaisuuksia, mutta tunnetilojen vaikutus ei vielä ole tilastollisesti merkitsevä. Koehenkilöiden verenpaineeseen ja sykkeeseen vaikuttavat enemmän muut tekijät.

Avainsanat: MANCOVA, kovarianssianalyysi, parametrien estimointi,
pääkomponenttianalyysi, monimuuttujaisuus

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

Sisällys

1	Johdanto	5
2	Tilastolliset menetelmät	6
2.1	MANCOVA-malli	6
2.2	Parametrien estimointi	6
2.3	Hypoteesien testaus	7
2.4	Pääkomponenttianalyysi	9
2.4.1	Pääkomponenttien määrittäminen	9
3	Aineiston analysointi ja testaus	10
3.1	Aineiston kuvaus	10
3.2	Muuttujien valinta	10
3.3	Muuttujien analysointi	12
3.4	Pääkomponenttianalyysi aineistolle	13
3.4.1	Pääkomponenttien analysointi	15
3.5	MANCOVA aineistolle	16
3.5.1	Mallin residuaaliesitykset	18
3.6	Malli pelkästään kovariaateilla	20
4	Yhteenveto	22
	Lähteet	23
	Liite: R-koodi	24

1 Johdanto

Tässä kandidaatintyössä esitellään monen muuttujan kovarianssianalyysi eli MANCOVA-malli, joka perustuu yleiseen lineaariseen malliin. MANCOVA-mallia käytetään useissa eri tieteenaloissa, kuten lääke- sekä terveystieteissä. Mallin avulla on tarkoitus tutkia jatkuvien sekä luokiteltujen muuttujien vaikutusta riippuviin muuttujiin.

Oletetaan, että mallin kovariaatit vaikuttavat riippuviin muuttujiin. Kontrollomalla kovariaattien vaihteluväliä, voidaan saada päteviä tuloksia testattaessa mahdollisia vaikutuksia. Tässä kandidaatintutkielmassa esitellään tarkemmin yleinen MANCOVA-malli ja sovelletaan sitä tutkielmassa käytetyn aineiston tutkimiseen.

Kandidaatintyö rakentuu siten, että ensimmäisessä luvussa johdatellaan aiheeseen ja kerrotaan, miten työ etenee. Luvussa kaksi esitellään MANCOVA-malli sekä sen teoreettista taustaa. Teoriaosuudessa tarkastellaan mallin parametrien estimointia sekä hypoteesin testausta. Samassa luvussa todistetaan myös pääkomponenttianalyysi sekä se, kuinka pääkomponentit määritetään.

Tutkielman kolmannessa luvussa esitellään aineisto. Tämän jälkeen testataan MANCOVA-mallilla muuttujien riippuvuutta tutkielmaan valitun esimerkkiaineiston avulla. Tutkielmassa tutkitaan verenpaineryhmittäin koehenkilöiden verenpainetta sekä sykettä, kun kovariaatteina toimii pääkomponenttianalyysin tulokset tunnetilojen vaikutuksesta. Valituista pääkomponenteista muodostetaan kovariaatit MANCOVA-malliin ja tarkastellaan niiden merkitystä selitettäviin muuttujiin. Tämän jälkeen analysoidaan saatuja tuloksia.

Tutkielman lukijalta oletetaan tilastollisten menetelmien sekä matriisilaskennan tietämystä. Lukijalta oletetaan myös varianssi- sekä kovarianssianalyysin perusteiden ymmärtämistä.

2 Tilastolliset menetelmät

Kandidaatintutkielmassa sovelletaan usean muuttujan kovarianssianalyysiä (MANCOVA), joka on kovarianssianalyysin (ANCOVA) laajennus. Nämä molemmat menetelmät liittyvät regressioanalyysiin sekä varianssianalyysiin (ANOVA). Malli sisältää kovarianssianalyysin tapaan vähintään yhden selittävän kategorisen muuttujan eli faktorin sekä vähintään yhden jatkuvan muuttujan eli kovariaatin. Erona kovarianssianalyysiin selitettäviä muuttujia on nyt vähintään kaksi. MANCOVA antaa tällöin tehokkaamman tavan tutkia usean riippuvan muuttujan yhteisvaikutusta samanaikaisesti. MANCOVA on yleinen lineaarinen malli, jolla voidaan tutkia, onko ryhmien välillä tilastollisesti merkittäviä eroja, samalla kun taustamuuttujien vaikutusta pyritään kontrolloimaan kovariaattien avulla. (Tabachnick & Fidell, 2014, s.287).

Tässä luvussa esitellään MANCOVA-malli, sen parametrien estimointi ja hypoteesin testaus. Lisäksi esitellään pääkomponenttianalyysi, jota hyödynnetään tutkielmassa MANCOVA-mallin kovariaattien valitsemiseen.

2.1 MANCOVA-malli

Tarkastellaan ensiksi yleistä MANCOVA-mallia. Malli voidaan esittää matriisimuodossa:

$$(2.1) \quad \mathbf{Y} = \mathbf{XB} + \mathbf{Z}\boldsymbol{\Gamma} + \mathbf{U},$$

Mallissa on esitettynä matriisi \mathbf{Y} ($n \times m$) selitettävistä muuttujista, suunnittelumatriisi \mathbf{X} ($n \times p$), ryhmäkeskiarvojen parametrimatriisi \mathbf{B} ($d \times m$), kovariaatit \mathbf{Z} ($n \times t$), kovariaattien kertoimet $\boldsymbol{\Gamma}$ sekä virhetermien matriisi \mathbf{U} . Virhetermin sarakkeet ovat riippumattomia ja noudattavat normaalijakaumaa $N(0, \boldsymbol{\Sigma})$. (Seber, G. A. F., 1984, s.465).

2.2 Parametrien estimointi

Tässä luvussa parametrien estimointien merkinnät pohjautuvat pääosin G.A.F. Seberin kirjan *Multivariate observations* lukuun 9.5 *Analysis of covariance*.

Mallista 2.1 voidaan muodostaa parametreille \mathbf{B} ja $\boldsymbol{\Gamma}$ niin kutsuttu normaaliyhtälö

$$(2.2) \quad \begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z \end{pmatrix} \begin{pmatrix} B \\ \Gamma \end{pmatrix} = \begin{pmatrix} X'Y \\ Z'Y \end{pmatrix}.$$

Olkoon $(X'X)^{-1}$ matriisin $X'X$ käänteismatriisi. Korvataan Γ estimaatilla $\widehat{\Gamma}$, jolloin ratkaisemalla normaaliyhtälö 2.2 parametrin B estimaatiksi saadaan

$$(2.3) \quad \widehat{B} = (X'X)^{-1} X'(Y - Z\widehat{\Gamma}) = \widehat{B}_0 - (X'X)^{-1} X'Z\widehat{\Gamma},$$

missä $\widehat{B}_0 = (X'X)^{-1} X'Y$. Parametrin Γ estimaatiksi saadaan

$$(2.4) \quad \widehat{\Gamma} = (Z'RZ)^{-1} Z'RY,$$

missä $R = I - X(X'X)^{-1} X'$.

2.3 Hypoteesien testaus

Kandidaatintyössä on käytetty hieman eri merkintöjä kuin kirjassa *Multivariate observations*. Esimerkiksi kokonaisneliösummaa merkitään SST , jotta se on helpommin ymmärrettävissä.

Tarkastellaan ensiksi hypoteesia $H_0 : \Gamma = 0$. Tällöin hypoteesin kokonaisneliösumma $SST = Y'RY$. Saadaan mallin 2.1 jäännöseliösummaksi

$$SSE = (Y - X\widehat{B} - Z\widehat{\Gamma})'(Y - X\widehat{B} - Z\widehat{\Gamma}).$$

Kun hyödynnetään malleja 2.3 ja 2.4, saadaan jäännöseliösummaksi

$$(2.5) \quad SSE = (Y - Z\widehat{\Gamma})'R(Y - Z\widehat{\Gamma}) = Y'RY - \widehat{\Gamma}'Z'RY.$$

Kokonaisneliösumman ja jäännöseliösumman erotukseksi saadaan nyt

$$(2.6) \quad SST - SSE = \widehat{\Gamma}'Z'RY.$$

Nollahypoteesin $\Gamma = 0$ voimassa ollessa F-testisuure on

$$F = \frac{(SST - SSE)/t}{SSE/(n - p - t)} = \frac{\widehat{\Gamma}'Z'RY/t}{s^2},$$

missä s^2 on jäännösvarianssi ja testisuure noudattaa F-jakaumaa $F_{t, n-p-t}$. (G.A.F. Sefer, 1984, s.464).

Testattaessa hypoteesia $H_0 : \mathbf{AB} = 0$, missä annettu \mathbf{A} on $q \times p$ matriisi. Käytetään apumatriiseja \mathbf{H} ja \mathbf{E} merkitsemään kokonaisneliösumman SST ja jäännösneliösumman SSE erotusta sekä pelkkää jäännösneliösummaa SSE .

$$(2.7) \quad \mathbf{H} = SST - SSE = \widehat{\mathbf{F}}' \mathbf{Z}' \mathbf{R} \mathbf{Y}$$

$$(2.8) \quad \mathbf{E} = SSE = \mathbf{Y}' \mathbf{R} \mathbf{Y} - \widehat{\mathbf{F}}' \mathbf{Z}' \mathbf{R} \mathbf{Y}$$

Esitetään hypoteesin $H_0 : \mathbf{AB} = 0$ testaus esimerkin avulla.

Voidaan kirjoittaa $\mathbf{Y}' \mathbf{R} \mathbf{Y}$ muotoon $\mathbf{Y}' \mathbf{R} \mathbf{Y} = \mathbf{E}_{yy}$. Tällöin yhtälö 2.4 voidaan esittää muodossa

$$(2.9) \quad \widehat{\mathbf{F}} = \mathbf{E}_{ZZ}^{-1} \mathbf{E}_{ZY}.$$

Samanlaista merkintätapaa apuna käyttäen, voidaan esittää yhtälö 2.8 muodossa

$$(2.10) \quad \mathbf{E} = \mathbf{E}_{YY} - \mathbf{E}_{YZ} \mathbf{E}_{ZZ}^{-1} \mathbf{E}_{ZY}.$$

Merkitään $\mathbf{E}_H = \mathbf{E} + \mathbf{H}$. Jotta saadaan ratkaistua \mathbf{E}_H , voidaan aluksi kirjoittaa $\mathbf{Y}' \mathbf{R}_H \mathbf{Y} = \mathbf{E}_{HYY}$. Nyt saadaan kaavoja 2.8 ja 2.10 hyödyntäen

$$(2.11) \quad \mathbf{E}_H = \mathbf{E}_{HYY} - \mathbf{E}_{HZY} \mathbf{E}_{HZZ}^{-1} \mathbf{E}_{HZY}.$$

Tästä saadaan nollahypoteesille Wilks'in Λ -testin ("lambda-testi") kaavaksi

$$(2.12) \quad \Lambda = \frac{|\mathbf{E}|}{|\mathbf{E}_H|} = \frac{|\mathbf{E}_{YY} - \mathbf{E}_{YZ} \mathbf{E}_{ZZ}^{-1} \mathbf{E}_{ZY}|}{|\mathbf{E}_{HYY} - \mathbf{E}_{HZY} \mathbf{E}_{HZZ}^{-1} \mathbf{E}_{HZY}|}.$$

Tämä kaava noudattaa tasajakaumaa $U_{d,q,n-p-t}$, kun H_0 on tosi. (G.A.F. Sefer, 1984, s.466) Wilks'in Λ -testi on yleisimmin käytetty monimuuttujainen versio ANOVA-testin F-testisuureesta. Testin tulos on tärkeässä osassa MANCOVA-testiä suorittaessa. Mitä suuremman arvon testi saa, niin sitä todennäköisemmin siitä tulee tilastollisesti merkitsevä.

Koska hypoteesi $\mathbf{F} = 0$ on tosi, voidaan kirjoittaa Wilks'in Λ -testin kaava muotoon

$$(2.13) \quad \Lambda = \frac{|\mathbf{E}_{YY}|}{|\mathbf{E}_{HYY}|} = \frac{|\mathbf{E}_{YY}|}{|\mathbf{E}_{YY} + \mathbf{H}_{YY}|}.$$

Tiedetään, että $\mathbf{H} = \mathbf{E} - \mathbf{E}_H$, joten tästä saadaan

$$(2.14) \quad \mathbf{H} = \mathbf{H}_{YY} - (\mathbf{E}_{YZ} + \mathbf{H}_{YZ})(\mathbf{E}_{ZZ} + \mathbf{H}_{ZZ})^{-1}(\mathbf{E}_{ZY} + \mathbf{H}_{ZY}) + \mathbf{E}_{YZ} \mathbf{E}_{ZZ}^{-1} \mathbf{E}_{ZY}.$$

2.4 Pääkomponenttianalyysi

Tämän luku pohjautuu pitkälti Arto Luoman vuonna 2013 järjestämän kurssin Monimuuttujamenetelmät kurssimonisteeseen.

Pääkomponenttianalyysillä pyritään kuvaamaan moniulotteisen aineiston vaihtelua muutaman alkuperäisistä muuttujista muodostetun lineaarikombinaation eli pääkomponentin avulla. Pääkomponentit määritellään siten, että ensimmäiset pääkomponentit selittävät suurimman osan aineiston vaihtelusta. Pääkomponenttianalyysi auttaa myös havaitsemaan sellaisia riippuvuussuhteita, joita ei muuten välttämättä saisi selville.

2.4.1 Pääkomponenttien määrittäminen

Pääkomponentit voidaan esittää muodossa $Y = AX$, missä Y on pääkomponenttivektori, X on alkuperäinen satunnaisvektori ja A on kerroinmatriisi.

Satunnaisvektorin $X = (X_1, \dots, X_p)'$ jakauman pääkomponentteina toimivat lineaarikombinaatiot

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p,$$

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p,$$

⋮

$$Y_p = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p.$$

Pääkomponentit voidaan ratkaista ominaisarvohajotelmasta $\Sigma = T\Lambda T'$, missä $\Sigma = \text{Cov}(X)$, T on ortogonaalinen matriisi ja Λ on diagonaalimatriisi, jonka elementit ovat positiivisia arvoja laskevassa järjestyksessä. Matriisin Λ diagonaalielementit ovat matriisin Σ ominaisarvot. Näiden tietojen pohjalta osoittautuu, että $A = T'$. (Luoma 2013, s.44).

3 Aineiston analysointi ja testaus

Tässä luvussa esitellään aineisto sekä muuttujat. Tämän jälkeen suoritetaan aineistolle pääkomponenttianalyysi, jonka tulosten perusteella suoritetaan usean muuttujan kovarianssianalyysi. Tämän jälkeen analysoidaan siitä saatuja tuloksia.

3.1 Aineiston kuvaus

Tämän kandidaatintyön aineistona toimii M. T. Tuomiston keräämä data Tampereella vuosina 1987–1991 suoritetusta ambulatoorisesta verenpainetutkimuksesta. Ambulatoorisella verenpainetutkimuksella tarkoitetaan verenpaineen jatkuvaa mittausta kannettavan laitteen avulla. Tutkimuksissa seurattiin 95:n 35-, 40- ja 45-vuotiaan perusterveen miehen systolista ja diastolista verenpainetta sekä sykettä päivän ajan. Tämän lisäksi tutkittiin heidän tunnetilojaan päivän jokaisen aktiviteetin aikana.

Kaikki koehenkilöt olivat hyvässä fyysisessä kunnossa, eikä heillä ollut mitään pitkäaikaislääkitystä. Heidän alkoholin sekä tupakkatuotteiden kulutus ei myöskään saanut olla korkea. Kaikki koehenkilöt olivat myös keskenään samaa kansalaisuutta, puhuivat samaa kieltä, eivätkä heidän kulttuurilliset taustansa eronneet toisistaan. Testipäivänä koehenkilöt eivät saaneet tehdä mitään raskasta tai likaista työtä.

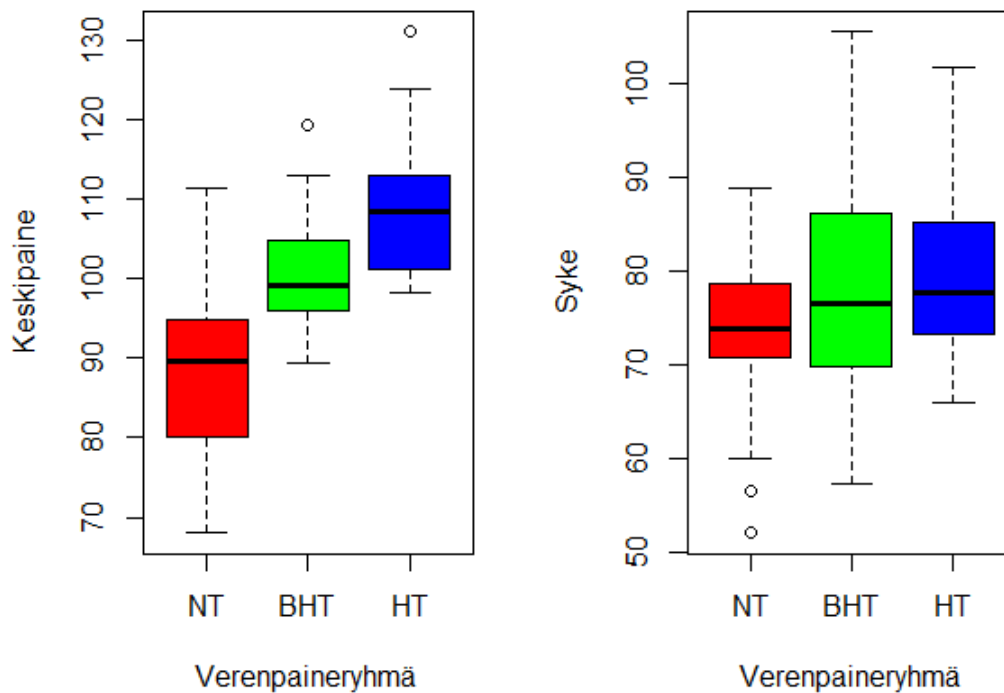
Diastolista ja systolista verenpainetta sekä sykettä mitattiin päivän aikana suunnilleen 100 000 arvoa per koehenkilö. Näistä mittausarvoista laskettiin aluksi 30 sekunnin keskiarvot. Tämän jälkeen laskettiin vielä keskiarvot tunneittain. Tällä tavalla saatiin datasta suppeampi ja helpommin käsiteltävä.

Aineistossa on vähän virheellisiä tai puuttuvia arvoja. Puutteellisten arvojen syytä ei ole tiedossa. Tässä kandidaatintyössä nämä kyseiset arvot on jätetty kokonaan pois, eikä niillä ole suurta vaikutusta lopullisiin tuloksiin. (Tuomisto, 1997).

3.2 Muuttujien valinta

Kandidaatintyössä tarkastellaan henkilön tunnetilojen vaikutusta henkilön verenpaineeseen sekä sykkeeseen. Tunnetilat toimivat tässä mallissa kovariaatteina. Selitettävänä muuttujina toimivat koehenkilön keskipaine sekä syke. Kandidaatintyössä tutkitaan tunnetiloja aikavälillä 15.00–18.00. Verenpainetta sekä sykettä tutkitaan aikavälillä 20.00–22.00. Kategorisena selittävänä muuttujana mallissa toimii henkilön

verenpaineryhmä. Aineistossa koehenkilöt on jaettu lääketieteellisin perustein kolmeen eri verenpaineryhmään. Nämä kyseiset ryhmät ovat, normotensiiviset (NT), prehypertensiiviset (BHT) sekä hypertensiiviset (HT). Verenpaineryhmien koot ovat NT = 33, BHT = 30 ja HT = 32.



Kuva 3.1. Laatikkokaaviot keskipaineelle ja sykkeelle

Kuvassa 3.1 on esitetty jokaisen verenpaineryhmän koehenkilöiden keskiverenpaineet sekä keskimääräiset sykkeet laatikkokaavioina. Laatikkokaavio on tilastollinen kuvaaja ja se antaa nopean visuaalisen kuvan aineiston mediaanista, hajonnasta sekä mahdollisista poikkeavista arvoista. Kuvaajassa laatikko kattaa ryhmän keskimäisen 50 % ja poikkiviiva esittää ryhmän mediaania. Laatikosta lähtevät viikset ("whiskers") esittävät ryhmän minimi- ja maksimiarvoja. Jos kuvaajassa esiintyy viiksien ulkopuolisia arvoja, niin kyseessä on todennäköisesti virheellinen tai poikkeuksellisen epätavallinen arvo. Kuvaajissa verenpaineryhmä NT tarkoittaa normotensiivisiä, BHT prehypertensiivisiä ja HT hypertensiivisiä.

Vasemmanpuoleisesta kuvaajasta nähdään selvästi, että normotensiivisillä koehenkilöillä on keskimäärin matalin keskiverenpaine. Tämän ryhmän omaavilla havaitaan myös selkeästi eniten hajontaa verenpaine-arvoissa. Prehypertensiivisillä koe-

henkilöillä havaitaan vähiten hajontaa arvoissa. Tämän ryhmän keskiverenpainearvot ovat keskimäärin korkeampia kuin normotensiivisillä ja matalampia kuin hypertensiivisillä. Odotetustikin kaikkein korkeimmat keskiverenpaineet ovat hypertensiivisen verenpaineryhmän omaavilla koehenkilöillä.

Oikeanpuoleisesta kuvaajasta nähdään selkeästi, että koehenkilön verenpaineryhmän vaikutus hänen sykkeeseensä on pienempi kuin sen vaikutus verenpaineeseen. Kuten keskipaineella havaittiin, niin myös sykkeen mediaaniarvo on pienin normotensiivisillä ja suurin hypertensiivisillä. Arvojen välillä tapahtuu kuitenkin paljon enemmän hajontaa kahdessa jälkimmäisessä ryhmässä. Korkeimmat sykkeen arvot havaitaankin prehypertensiivisillä koehenkilöillä. Tämän perusteella voidaankin todeta, että verenpaineryhmän vaikutus koehenkilön keskimääräiseen sykkeeseen on nähtävissä, mutta se ei kuitenkaan ole aina kovinkaan suurta.

Kummassakaan kuvaajassa ei ole otettu huomioon sitä, millaisia aktiviteetteja koehenkilö on päivän aikana tehnyt. Aktiviteettien fyysisyydessä on varmasti suuriakin eroja koehenkilöiden välillä. Tämän vuoksi kuvasta 3.1 ei siis kannata vielä tehdä liian suuria johtopäätöksiä.

Taulukko 3.1. Verenpaineryhmien keskimääräiset verenpaineet ja sykkeet

Verenpaineryhmä	Keskiverenpaine	Keskisyke
Normotensiiviset	89.43049	73.79131
Prehypertensiiviset	100.8325	77.78435
Hypertensiiviset	108.124	80.1284

Taulukossa 3.1 on vielä esitettyä eri verenpaineryhmien koehenkilöiden keskimääräiset verenpaineet sekä sykkeet numeerisesti. Myös tästä voidaan helposti todeta, että verenpaineryhmän vaikutus henkilön verenpaineeseen on suuri. Myös vaikutus sykkeeseen on nähtävissä selkeästi. Voidaan siis todeta, että kullakin verenpaineryhmällä on selkeästi merkitystä verenpaineita sekä sykkeitä tutkittaessa.

3.3 Muuttujien analysointi

Tutkitaan seuraavaksi tunnetilojen sekä verenpaineiden ja sykkeen korreloitumista keskenään korrelaatiomatriisilla.

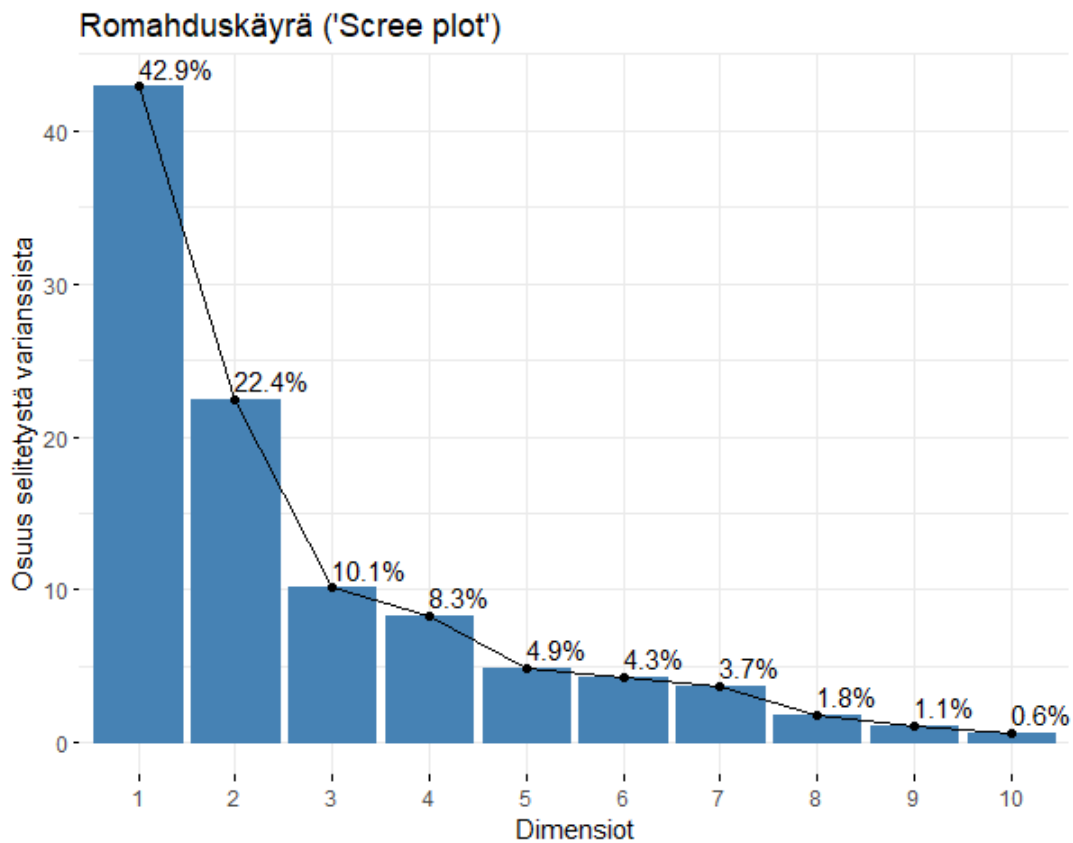
Kuvan 3.2 korrelaatiomatriisista nähdään, että lähes kaikki tunnetilat ovat korre-

tutkittavasta aineistosta tiiviimpi ja vähentää sen moniulotteisuutta. (Luoma 2013, s.41).

Taulukko 3.2. Pääkomponentit ja niiden tärkeys

	1. komponentti	2.	3.	4.	5.	6.	7.
Standardipoikkeama	2.53	1.83	1.23	1.11	0.85	0.80	0.74
Osuus selitetyistä varianssista	0.43	0.22	0.10	0.083	0.049	0.043	0.037

Suoritetaan pääkomponenttianalyysi R-ohjelmistokielen avulla. Taulukossa 3.2 on esitettyä pääkomponenttianalyysin tuloksia seitsemältä ensimmäiseltä pääkomponentilta. Nähdään, että neljän ensimmäisen pääkomponentin standardipoikkeama on suurempi kuin yksi. Tämän perusteella voitaisiin siis valita kahdesta neljään pääkomponenttia.



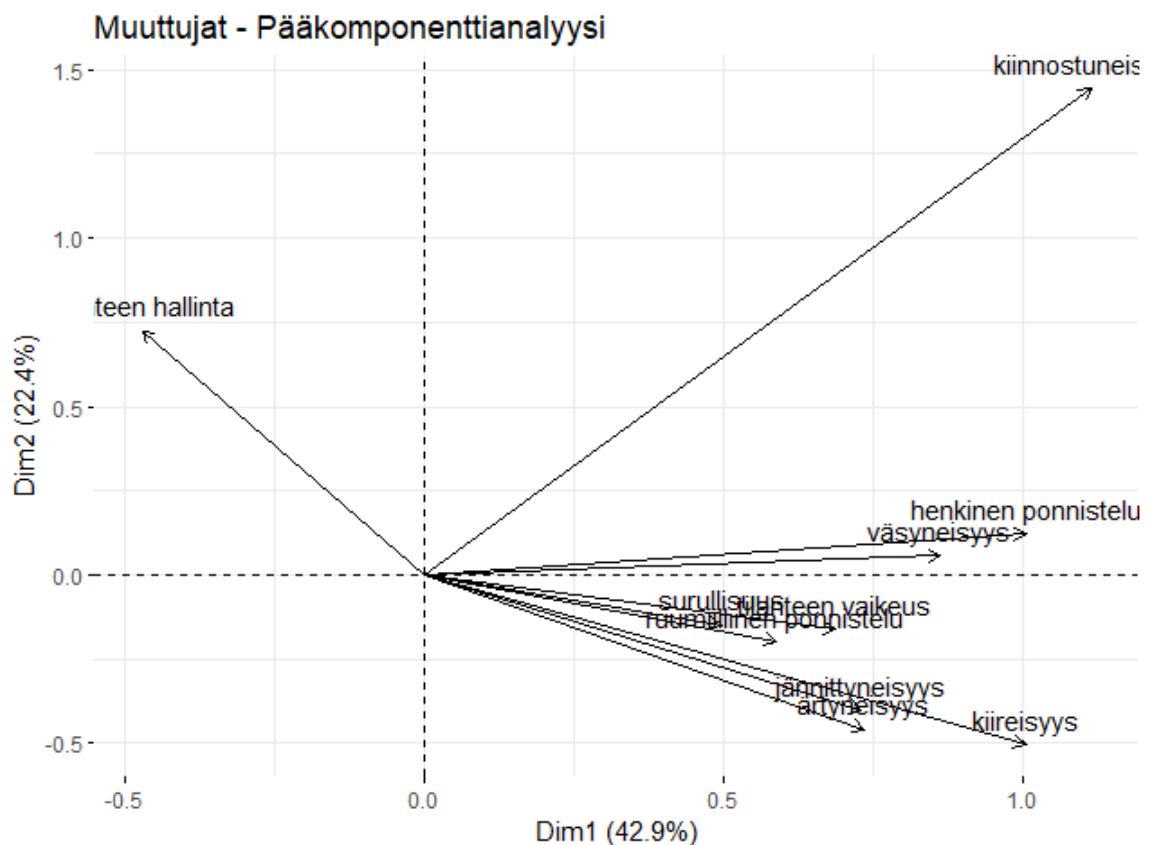
Kuva 3.3. Pääkomponentit kuvaajana

Kuvan 3.3 romahduskäyrästä nähdään hyvin kunkin pääkomponentin tärkeys. Ensimmäinen ja toinen pääkomponentti selittävät yhteensä noin 65,3 %. Pääkomponentteja valittaessa usein halutaan valita niin monta pääkomponenttia, että selitettyä

varianssia olisi yhteensä ainakin 70 % (Nummi, 2018, s.53). Tässäkin tapauksessa voitaisiin valita ainakin kolme pääkomponenttia. Pääkomponenttianalyysi ei kuitenkaan ole tässä kandidaatintyössä avainasemassa. Tarkoitus on vain saada aineistosta tiiviimpi. Täten päädytään valitsemaan vain kaksi pääkomponenttia. Kuvasta 3.3 huomataan myös selkeä pudotus selitettävän varianssin osuudessa toisen pääkomponentin jälkeen. Tämä antaa myös päteväen syyn valita vain kaksi pääkomponenttia.

3.4.1 Pääkomponenttien analysointi

Aikaisemmin valittiin aineistolle kaksi pääkomponenttia. Nyt halutaan selvittää, mitenkä kukin tunnetila vaikuttaa pääkomponenteissa. Tiedetään myös, että monet tunnetilat korreloivat keskenään.



Kuva 3.4. Tunnetilojen merkitsevyys pääkomponenteissa

Kuvasta 3.4 nähdään, kuinka merkitseviä kukin tunnetila on pääkomponenteille. Mitä pidempi viiva tunnetilalla on, sitä enemmän se selittää pääkomponenttia. Kuva kertoo myös sen, kuinka paljon eri tunnetilat korreloivat keskenään. Mitä lähempänä tunnetilat ovat toisiaan, sitä enemmän ne myös korreloivat keskenään.

Kuvasta nähdäänkin tunnetilan *kiinnostuneisuus* selittävän kaikkein vahvimmin sekä ensimmäistä, että toista pääkomponenttia. Se ei myöskään korreloi paljoa muiden tunnetilojen kanssa. *Tilanteen hallinta* on ainoa tunnetila, joka korreloi negatiivisesti ensimmäisen pääkomponentin kanssa. Se selittää kuitenkin paljon toista pääkomponenttia. Kaikki loput tunnetilat ovat isossa ryppäessä. Ne siis korreloivat vahvasti keskenään. Näistä tunnetiloista *henkinen ponnistelu*, *kiireisyys* ja *väsyneisyys* selittävät eniten ensimmäistä pääkomponenttia.

Voidaan nimetä pääkomponentit niistä kerättyjen tietojen perusteella. Annetaan ensimmäiselle pääkomponentille nimeksi "Stressi", koska sitä selittävät useat stressiin vaikuttavat tunnetilat. Toiselle pääkomponentille voidaan suoraan antaa nimeksi "Kiinnostuneisuus", koska tämä tunnetila selittää selkeästi suurimman osan pääkomponenttia.

3.5 MANCOVA aineistolle

Pääkomponenttianalyysin suorittamisen jälkeen, voidaan saatujen tulosten perusteella suorittaa monen muuttujan kovarianssianalyysi eli MANCOVA. Tarkoituksena siis on tutkia MANCOVAa hyväksikäyttäen, onko pääkomponenteiksi valituilla stressillä ja kiinnostuneisuudella vaikutusta koehenkilöiden verenpaineeseen tai sykkeeseen.

Aikaisemmin todettiin koehenkilön verenpaineryhmän vaikuttavan selkeästi verenpaineeseen sekä sykkeeseen. Verenpaineryhmät toimivat tässä MANCOVA-mallissa kategorisena muuttujana eli faktorina. Kovariaatteina toimivat valitut pääkomponentit. Mallia kuvataan seuraavanlaisella esityksellä.

$$\begin{pmatrix} \text{Keskipaine} \\ \text{Syke} \end{pmatrix} = B_0 + B_{1i} \text{Verenpaineryhmä} + B_2 \text{Stressi} + B_3 \text{Kiinnostuneisuus} + U.$$

Verenpaineryhmän luokkakertoimen B_{1i} indikaattori $i = 1, 2, 3$ ja arvot kuvaavat eri verenpaineryhmiä. Estimoitu MANCOVA-malli saadaan taulukon 3.3 estimaattien arvoista. Nämä arvot toimivat muuttujien kertoimina mallissa.

Taulukossa 3.3 on esitettyä estimoidun MANCOVA-mallin estimaattien kertoimet. Vakiotermin \widehat{B}_0 kertoimet ovat 89.40 keskipaineelle ja 73.98 sykkeelle. *Verenpaineryhmä* on luokkamuuttuja ja sen \widehat{B}_{1i} kertoimet on esitetty taulukossa luokittain. Kertoimen indikaattori $i = 1, 2, 3$ ja kukin arvo kuvaa eri verenpaineryhmää. Muuttujan *Stressi* \widehat{B}_2 kertoimet ovat 0.13 sekä -0.16. Muuttujalla *Kiinnostuneisuus* \widehat{B}_3 kertoimet ovat -0.092 ja -0.29.

Taulukko 3.3. MANCOVA-mallin estimaatit

	Estimaatit keskipaineelle	Estimaatit sykkeelle
Vakio	89.4030	73.9835
Stressi	0.1291	-0.1622
Kiinnostuneisuus	-0.0917	-0.2898
Verenpaineryhmä		
NT		
BHT	11.4131	3.7149
HT	18.7648	6.0272

Taulukoissa 3.4 ja 3.5 on esitetty MANCOVA-testin tulokset. Verenpaineryhmien p-arvo keskipaineelle on $5.83 \cdot 10^{-13}$ ja sykkeelle 0.036. Varsinkin keskipaineelle p-arvo on todella pieni, joten voidaan todeta verenpaineryhmien vaikuttavan selkeästi siihen. Myös sykkeelle todetaan tämän arvon olevan tilastollisesti merkitsevä.

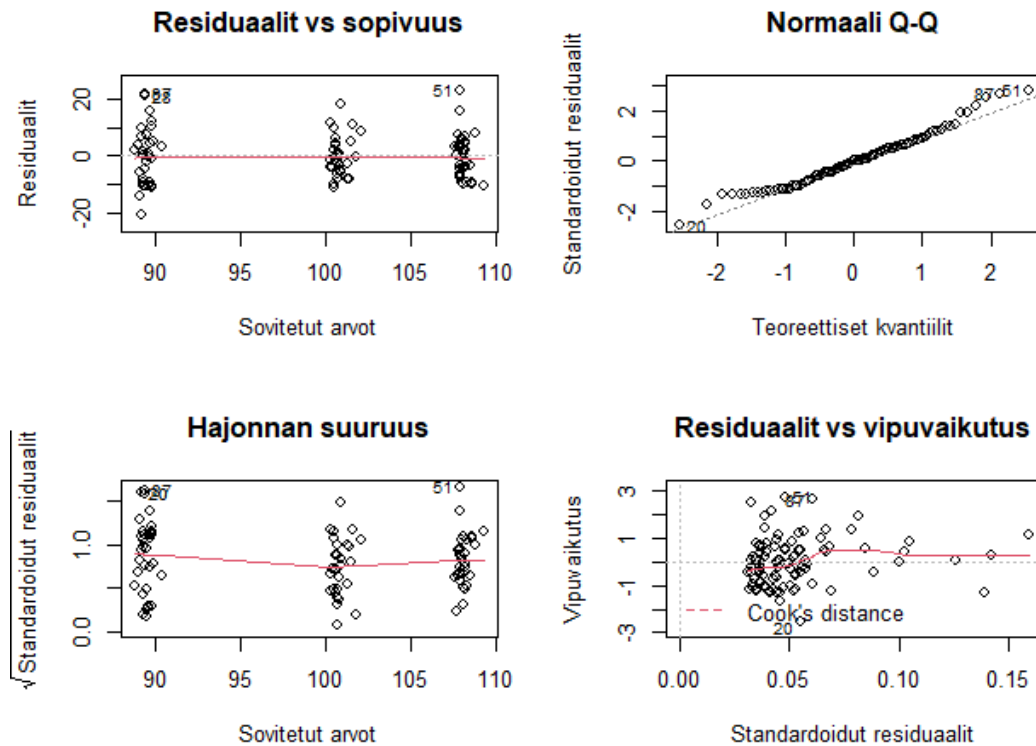
Stressin sekä kiinnostuneisuuden p-arvo ovat liian suuria. Täten niiden vaikutukset sykkeeseen tai verenpaineeseen eivät ole tilastollisesti merkitseviä. Koehenkilön tunnetilojen vaikutus verenpaineisiin sekä sykkeisiin ei siis ole tarpeeksi suuri. Tästä voidaan päätellä, että verenpaineen sekä sykkeen vaihteluihin vaikuttavat enemmän muut tekijät.

Taulukko 3.4. MANCOVA-taulu keskipaineelle

	df	Neliösumma	Keskineliösumma	F-testisuure	p-arvo
Verenpaineryhmä	2	5776.4	2888.22	39.1575	$5.826 \cdot 10^{-13}$
Stressi	1	10.1	10.08	0.1367	0.7125
Kiinnostuneisuus	1	2.6	2.60	0.0352	0.8516
Residuaalit	90	6638.3	73.76		

Taulukko 3.5. MANCOVA-taulu sykkeelle

	df	Neliösumma	Keskineliösumma	F-testisuure	p-arvo
Verenpaineryhmä	2	668.1	334.04	3.4587	0.03572
Stressi	1	14.7	14.72	0.1524	0.69714
Kiinnostuneisuus	1	25.9	25.94	0.2686	0.60555
Residuaalit	90	8692.2	96.58		



Kuva 3.5. MANCOVA-mallin residuaaliesitys keskipaineelle

3.5.1 Mallin residuaaliesitykset

Tutkitaan mallia vielä visuaalisesti. Kuvassa 3.5 on esitettyä MANCOVA-mallin residuaaliesitys keskipaineelle.

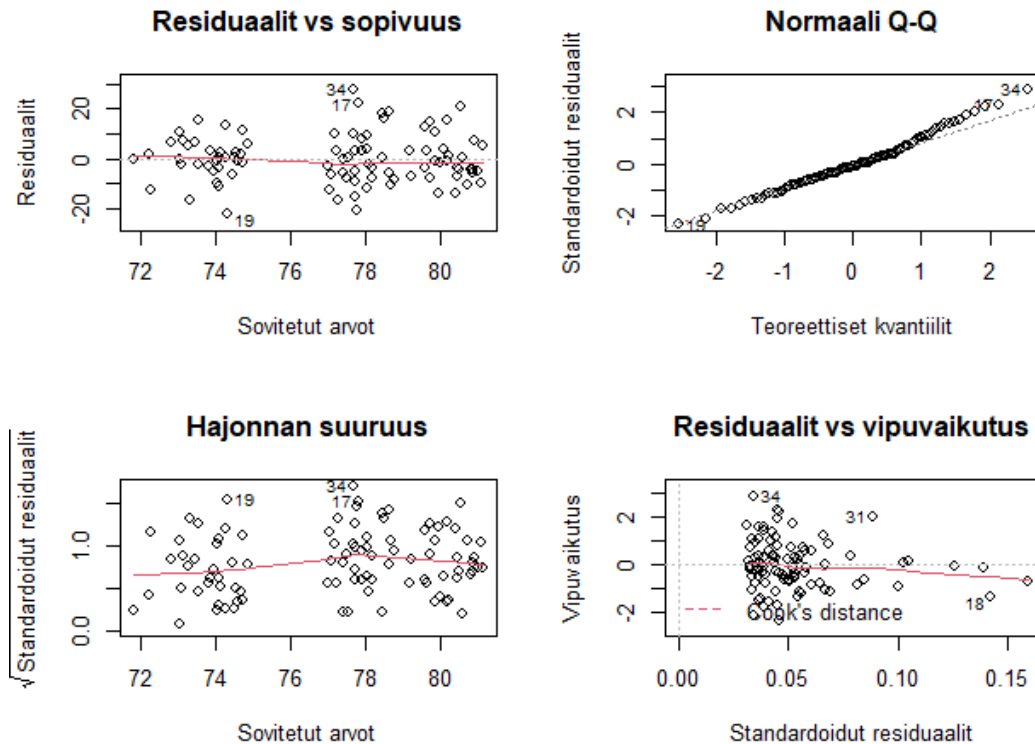
Vasemman yläkulman kuvaaja näyttää mallin sopivuuden. Kuvasta voidaan todeta residuaalien jakautuneen melko tasaisesti. MANCOVA-mallissa ei siis ole juuri yhtään poikkeavia arvoja.

Oikean yläkulman kuvaajassa tarkastellaan, kuinka normaalisti residuaalit ovat jakautuneet. Kuvasta todetaan, etteivät residuaalit ole jakautuneet kovinkaan normaalisti, koska standardoidut residuaalit kaartuvat viivan ala- ja yläpäissä.

Vasemman alakulman kuvaajasta saadaan selville, leviävätkö residuaalit tasapuolisesti ennustajien vaihteluvälille. Kuvaajassa punainen viiva on melko suora, joka osoittaa residuaalien leviävän tasaisesti. Residuaalien hajonta on myös tällöin melko tasainen kaikilla ennustetuilla arvoilla.

Oikean alakulman kuvaajassa tutkitaan sitä, onko yksittäisellä pisteellä vaikutusta MANCOVA-malliin. Jos jokin piste on Cooksin etäisyyden ("Cook's distance") ulkopuolella, niin kyseessä on todennäköisesti poikkeava havainto, jonka poistami-

nen mallista saattaisi vaikuttaa saatuihin tuloksiin. Huomataan, ettei Cooksin etäisyys mahdu edes kuvaan, joten kaikki havainnot ovat sen sisäpuolella. Mallista ei siis löydy selkeästi poikkeavia arvoja.



Kuva 3.6. MANCOVA-mallin residuaaliesitys sykkeelle

Tutkitaan seuraavaksi kuvasta 3.6 MANCOVA-mallin residuaaliesitystä sykkeelle. Vasemman yläkulman kuvaajasta voidaan todeta punaisen viivan olevan suhteellisen suora pienestä kaarevuudesta huolimatta. Residuaalit ovat siis jakautuneet melko tasaisesti, eikä mallissa ole juurikaan poikkeavia arvoja.

Oikean yläkulman kuvaajasta nähdään, että residuaalit ovat jakautuneet melko normaalisti. Vasta ihan kuvan yläosassa standardoidut residuaalit alkavat kaartua hieman ylöspäin.

Vasemman alakulman kuvaajasta saadaan selville, että residuaalit leviävät melko tasaisesti ennustajien vaihteluvälille. Residuaalien hajonta on siis suhteellisen tasainen kaikilla ennustetuilla arvoilla.

Oikean alakulman kuvaajasta huomataan heti, että Cooksin etäisyys ei mahdu edes kuvaan. Tällöin kaikki kuvan pisteet eli havainnot ovat sen sisäpuolella. Mallista ei siis löydy yhtäkään selkeästi muista poikkeavaa arvoa, joka voisi vaikuttaa saatuihin tuloksiin.

Kuvista 3.5 ja 3.6 saatujen tietojen perusteella, voidaan todeta MANCOVA-mallin olevan suhteellisen sopiva tälle aineistolle. Täten mallia voidaan pitää myös luotettavana.

3.6 Malli pelkästään kovariaateilla

Edellä todettiin ainoastaan mallin verenpaineryhmien olevan tilastollisesti merkitseviä. Nyt halutaan selvittää, tapahtuuko mallissa merkittäviä muutoksia poistamalla siitä verenpaineryhmät. MANCOVA-mallia voidaan kuvata nyt seuraavanlaisella esityksellä

$$\begin{pmatrix} \text{Keskipaine} \\ \text{Syke} \end{pmatrix} = B_0 + B_1 \text{Stressi} + B_2 \text{Kiinnostuneisuus} + U.$$

Estimoidun MANCOVA-mallin kertoimet saadaan taulukosta 3.6. Vakiotermin \widehat{B}_0 kertoimet ovat 99.33 keskipaineelle ja 77.19 sykkeelle. Kovariaatin *Stressi* \widehat{B}_1 kertoimet ovat -0.35 sekä -0.32. Kovariaatilla *Kiinnostuneisuus* \widehat{B}_2 kertoimet ovat -0.67 ja -0.48.

Taulukko 3.6. MANCOVA-mallin estimaatit

	Estimaatit keskipaineelle	Estimaatit sykkeelle
Vakio	99.3279	77.1869
Stressi	-0.3493	-0.3160
Kiinnostuneisuus	-0.6694	-0.4766

Kun verrataan taulukoita 3.3 ja 3.6, niin havaitaan, että MANCOVA-mallien estimaatit eroavat toisistaan. Esimerkiksi vakiotermin \widehat{B}_0 kertoimet olivat ensimmäisessä mallissa 89.40 sekä 73.98 ja toisessa mallissa ne ovat 99.33 sekä 77.19. Voidaan siis olettaa mallien tuloksissa olevan eroja.

Taulukoissa 3.7 ja 3.8 on esitettyä MANCOVA-testauksen tulokset. Stressin p-arvot ovat nyt 0.46 sekä 0.44 ja kiinnostuneisuuden p-arvot ovat 0.30 sekä 0.40. Vaikka p-arvot ovat nyt pienempiä kuin ensimmäisessä mallissa, niin ne kaikki ovat jälleen suurempia kuin 0.05, ja täten ne eivät ole tilastollisesti merkitseviä. Poistamalla mallista verenpaineryhmät tulokset eivät siis juurikaan muuttuneet. Jälleen voidaan todeta, että koehenkilöiden verenpaineen sekä sykkeen vaihteluihin vaikuttavat enemmän muut tekijät kuin heidän tunnetilansa.

Taulukko 3.7. MANCOVA-taulu keskipaineelle

	df	Neliösumma	Keskineliösumma	F-testisuure	p-arvo
Stressi	1	73.9	73.937	0.5570	0.4574
Kiinnostuneisuus	1	142.0	141.997	1.0698	0.3037
Residuaalit	92	12211.5	132.734		

Taulukko 3.8. MANCOVA-taulu sykkeelle

	df	Neliösumma	Keskineliösumma	F-testisuure	p-arvo
Stressi	1	60.5	60.506	0.6006	0.4403
Kiinnostuneisuus	1	72.0	71.969	0.7144	0.4002
Residuaalit	92	9268.5	100.745		

4 Yhteenveto

Kandidaatintutkielmassa tutkittiin koehenkilöiden tunnetilojen vaikutusta heidän verenpaineeseensa sekä sykkeeseensä MANCOVA-mallia apuna käyttäen.

Kuvassa 3.1 sekä taulukossa 3.1 havaittiin verenpaineryhmien vaikuttavan selkeästi keskimääräiseen verenpaineeseen sekä sykkeeseen. Kuvan 3.2 korrelaatiomatriisissa havaittiin tunnetilojen korreloivan sykkeen sekä verenpaineiden kanssa. Haluttiin selvittää, kuinka merkittävästi tunnetilat vaikuttavat verenpaineeseen ja sykkeeseen. Muodostettiin MANCOVA-malli tutkimaan tätä asiaa. Mallin kovariaatit saatiin verenpaineryhmistä sekä tunnetiloille suoritettuna pääkomponenttianalyysin tuloksista.

MANCOVA-mallista saaduista tuloksista havaittiin verenpaineryhmien olevan merkittävässä roolissa varsinkin keskimääräisiä verenpaineita tutkittaessa. Myös vaikutusta keskimääräisiin sykkeisiin oli havaittavissa. Mallista todettiin myös, ettei tunnetilojen vaikutus ole tilastollisesti merkitsevä. Pelkät tunnetilat eivät siis vaikuta tarpeeksi, jotta voitaisiin todeta niiden olevan tärkein syy koehenkilöiden verenpaineen sekä sykkeen vaihteluihin. Kun tutkittiin vielä MANCOVA-mallin residuaaliesityksiä, niin voitiin todeta käytetyn mallin olevan melko luotettava sekä sopiva aineistolle.

Lopuksi testattiin vielä toista MANCOVA-mallia, josta verenpaineryhmät oli poistettu. Havaittiin muutoksia saaduissa tuloksissa, mutta tunnetilojen vaikutus ei vielääkään ollut merkitsevä. Päädyttiin samaan lopputulokseen kuin ensimmäisessä mallissa, eli koehenkilöiden verenpaineeseen sekä sykkeeseen vaikuttavat enemmän muut tekijät. Tällaisia syitä voivat olla muun muassa henkilön ikä sekä tutkimuspäivän aktiiviteettien fyysisyys.

Lähteet

- [1] Seber G.A.F. (1984). *Multivariate observations*
- [2] Tabachnick, Barbara G ja Fidell, Linda S (2014). *Using multivariate statistics*
- [3] T. Nummi (2018). *Multivariate Analysis*, Luentomoniste, Tampereen Yliopisto.
- [4] A. Luoma (2013). *Monimuuttujamenetelmät*, Luentomoniste, Tampereen Yliopisto
- [5] Tuomisto, M. T. (1997). Intra-arterial blood pressure and heart rate reactivity to behavioral stress in normotensive, borderline hypertensive, and mild hypertensive men. *Health Psychology*, 16, 554-565

Liite: R-koodi

```
> tunnetilat <- data_koottu[,c(2:11)]
> tunnetilat.pca <- princomp(tunnetilat)

> Stressi <- tunnetilat.pca$scores[,1]
> head(Stressi)
[1] 0.669263608 0.563534542 1.018395614 4.876670481 -1.528536197 -0.003917487

> Kiinnostuneisuus <- tunnetilat.pca$scores[,2]
> head(Kiinnostuneisuus)
[1] -2.5411859 -0.8211743 1.1885338 -3.1748422 -1.7076927 -2.0303435

> malli <- lm(cbind(keskipaine, syke) ~ factor(vp_ryhma) + Stressi
+           + Kiinnostuneisuus, data = data_koottu)
> malli
```

Call:

```
lm(formula = cbind(keskipaine, syke) ~ factor(vp_ryhma) + Stressi +
    Kiinnostuneisuus, data = data_koottu)
```

Coefficients:

	keskipaine	syke
(Intercept)	89.4030	73.9835
factor(vp_ryhma)Ryhma 2 (BHT)	11.4131	3.7149
factor(vp_ryhma)Ryhma 3 (HT)	18.7648	6.0272
Stressi	0.1291	-0.1622
Kiinnostuneisuus	-0.0917	-0.2898

```
> manova(malli)
```

Call:

```
manova(malli)
```


Terms:

	factor(vp_ryhma)	Stressi	Kiinnostuneisuus	Residuals
keskipaine	5776.441	10.080	2.597	6638.308
syke	668.076	14.723	25.940	8692.243
Deg. of Freedom	2	1	1	90

Residual standard errors: 8.588304 9.827536

Estimated effects may be unbalanced

```
> summary.aov(malli)
```

Response keskipaine :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(vp_ryhma)	2	5776.4	2888.22	39.1575	5.826e-13 ***
Stressi	1	10.1	10.08	0.1367	0.7125
Kiinnostuneisuus	1	2.6	2.60	0.0352	0.8516
Residuals	90	6638.3	73.76		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response syke :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(vp_ryhma)	2	668.1	334.04	3.4587	0.03572 *
Stressi	1	14.7	14.72	0.1524	0.69714
Kiinnostuneisuus	1	25.9	25.94	0.2686	0.60555
Residuals	90	8692.2	96.58		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> keskipaine <- aov(keskipaine ~ factor(vp_ryhma) + Stressi  
+                   + Kiinnostuneisuus, data=data_koottu)
```

```
> summary(keskipaine)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(vp_ryhma)	2	5776	2888.2	39.158	5.83e-13 ***
Stressi	1	10	10.1	0.137	0.712
Kiinnostuneisuus	1	3	2.6	0.035	0.852

```
Residuals      90   6638   73.8
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> syke <- aov(syke ~ factor(vp_ryhma) + Stressi
+             + Kiinnostuneisuus, data=data_koottu)
> summary(syke)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(vp_ryhma)	2	668	334.0	3.459	0.0357 *
Stressi	1	15	14.7	0.152	0.6971
Kiinnostuneisuus	1	26	25.9	0.269	0.6056
Residuals	90	8692	96.6		

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> uusi_malli <- lm(cbind(keskipaine, syke) ~ Stressi +
+                 Kiinnostuneisuus, data = data_koottu)
> uusi_malli
```

```
Call:
```

```
lm(formula = cbind(keskipaine, syke) ~ Stressi + Kiinnostuneisuus,
    data = data_koottu)
```

```
Coefficients:
```

	keskipaine	syke
(Intercept)	99.3279	77.1869
Stressi	-0.3493	-0.3160
Kiinnostuneisuus	-0.6694	-0.4766

```
> manova(uusi_malli)
```

```
Call:
```

```
manova(uusi_malli)
```

```
Terms:
```

```
Stressi Kiinnostuneisuus Residuals
```

keskipaine	73.937	141.997	12211.492
syke	60.506	71.969	9268.507
Deg. of Freedom	1	1	92

Residual standard errors: 11.52101 10.03716

Estimated effects are balanced

> summary.aov(uusi_malli)

Response keskipaine :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Stressi	1	73.9	73.937	0.5570	0.4574
Kiinnostuneisuus	1	142.0	141.997	1.0698	0.3037
Residuals	92	12211.5	132.734		

Response syke :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Stressi	1	60.5	60.506	0.6006	0.4403
Kiinnostuneisuus	1	72.0	71.969	0.7144	0.4002
Residuals	92	9268.5	100.745		