

PRO GRADU -TUTKIELMA

Risto Nikkanen

**Innovatiivisten työkäytäntöjen vaikutus yrityksen
suorituskykyyn**

TAMPEREEN YLIOPISTO
Matematiikan ja tilastotieteen laitos
Tilastotiede
Toukokuu 2010

Tampereen yliopisto

Matematiikan ja tilastotieteen laitos

NIKKANEN, RISTO: Innovatiivisten työkäytäntöjen vaikutus yrityksen suorituskykyyn, Pro gradu -tutkielma, 46 s., 14 liites.

Toukokuu 2010

Tiivistelmä

Yrityksen tulokseen vaikuttavien erilaisten tuotanto-, työn organisointi- ja tiedon johtamisen käytäntöjen tutkimus on ollut 1990-luvulta lähtien tutkijoiden ja yritysjohtajien keskeinen kiinnostuksen kohde. Tutkimusperinteessä on selvitetty yritysten käytössä olevien erilaisten käytäntöjen vaikutuksia yrityksen suorituskykyyn. Tämän tutkimuksen teoreettisena lähtökohtana ovat Ichniowskin (1990) tutkimustulokset, joiden mukaan yksittäisillä käytännöillä ei ole suurta vaikutusta yrityksen menestymiseen, mutta tietyillä käytäntöjen yhdistelmillä, nipuilla (bundles) sen sijaan on.

Tässä tutkimuksessa selvitetään em. tutkimusperinteen pohjalta 12 eri valmistusprosessin ja työn organisoinnin käytännön vaikutuksia yrityksen suorituskykyyn sekä erillisinä että yhdessä käytettyinä. Tutkimusaineisto on saatu Tampereen yliopiston Työelämän tutkimuskeskuksen keväällä 2007 toteuttamasta kyselytutkimuksesta, jossa suomalaisilta teollisuusyrityksiltä kysyttiin mm. organisatoristen innovaatioiden käytöstä ja niiden käyttöasteesta. Hanke kuului Suomen Akatemian Liike2-tutkimusohjelmaan, ja sitä rahoitti kyseisen ohjelman lisäksi myös Työsuojelurahasto.

Suorituskykymuuttujat muodostettiin kyselylomakkeen osion ”liiketoiminnan mittarit” pohjalta. Muuttujista kolme (kokonaissuorituskyky, työhyvinvointi ja joustavuus) ovat keskiarvomuuttujia ja neljäs, yrityksen tuoteinnovaatiokykyä mittaava muuttuja on binaarinen. Tilastollisina menetelminä tutkimuksessa on käytetty lineaarista ja logistista regressioanalyysia, pääkomponenttianalyysia ja lopuksi rakenneyhtälömalleja lopullisen tilastollisen mallin muodostamiseksi.

Saatujen tulosten mukaan erillisillä käytännöillä oli vain muutamia merkitseviä suorituskykyvaikutuksia, kun taas pääkomponenttianalyysin tuloksena saaduilla käytäntönipuilla niitä oli enemmän. Erityisesti ensimmäinen, työn rajojen purkamiseen tähtäävä komponentti osoittautui merkitseväksi kaikkien neljän suorituskykyä mittaavan muuttujan kohdalla. Analyysitulokset antavat näiltä osin tukea Ichniowskin (1990) tutkimustuloksille.

Asiasanat Lineaarinen regressioanalyysi, logistinen regressioanalyysi, pääkomponenttianalyysi, rakenneyhtälömallit

Sisältö

1	Johdanto	5
1.1	Teoreettinen tausta	5
1.2	Tutkimusongelman määrittely	6
2	Tutkimusaineisto ja aineiston muokkaus	7
2.1	Tutkimusaineisto	7
2.2	Aineiston muokkaus analyysia varten	8
3	Tutkimusmenetelmät	11
3.1	Lineaarinen regressio	11
3.1.1	Lineaarinen regressiomalli ja pienimmän neliösumman estimointi	11
3.1.2	Neliösummahajotelmat ja selitysasteet	13
3.1.3	Hypoteesien testaus	14
3.1.4	Muita mallin arvioinnin kriteerejä	15
3.2	Logistinen regressio	16
3.2.1	Logistinen regressiomalli	17
3.2.2	Suurimman uskottavuuden estimointi	18
3.2.3	Logistisen regressiomallin arviointi	19
3.3	Pääkomponenttianalyysi	22
3.3.1	Pääkomponentit	22
3.3.2	Koordinaatiston kierto	23
3.3.3	Pääkomponenttien estimointi	24
3.4	Rakenneyhtälömallit	25
3.4.1	Usean yhtälön regressiomallit	25
3.4.2	Rekursiiviset ja simultaaniset regressiopolkumallit	26
3.4.3	Mallin arviointikriteerejä	27
4	Aineiston kuvaus	29
4.1	Organisatoristen innovaatioiden käyttö kyselyn toimipaikoissa	29
4.2	Innovatiiviset työkäytännöt yrityksen toimialan ja koon mukaan	30
5	Analyysitulokset	32
5.1	Yksittäiset käytännöt ja yrityksen suorituskyky	32
5.2	Pääkomponenttianalyysi yksittäisille käytännöille	35
5.3	Käytäntökomponentit ja yrityksen suorituskyky	37
5.4	Lopullinen malli	39

5.4.1	Rekursiivinen regressiopolkumalli	39
5.4.2	Mallin arviointi	40
6	Johtopäätökset ja tutkimuksen rajoitukset	42
	Lähdeluettelo	45

1 Johdanto

1.1 Teoreettinen tausta

Yritysjohtajia ja organisaatiotutkijoita on pitkään kiinnostanut kysymys, millaisin organisaation toiminta- ja työkäytännöin on saavutettavissa yrityksen paras mahdollinen suorituskyky. USA:ssa alan tutkimusten lähtökohtana olivat Cappellin ja Neumarkin mukaan (1999, 5) havainnot japanilaisten johtamissopien mahdollisesta paremmuudesta amerikkalaisiin verrattuna ja yleisemminkin huoli USA:n teollisuuden kilpailukyvyistä. Erityisen merkittävä keskustelun vauhdittaja oli japanilaista autoteollisuutta koskenut tutkimus, jonka mukaan japanilaisen teollisuuden parempi suorituskyky amerikkalaisiin verrattuna oli yhteydessä japanilaisten harjoittamiin tiettyihin organisatorisiin käytäntöihin (ks. esim. Kenney & Florida 1993).

Hyviksi, parhaiksi, innovatiivisiksi tai uusiksi nimettyihin työkäytäntöihin on luettu tutkijasta riippuen monia erilaisia käytäntöjä. Usein tällaisiksi mainitaan esimerkiksi laatupiirit, työkierto, tiimityö sekä joustava tuotantoautomaatio. Vaikka alan tutkimuksen juuret ulottuvat noin 20 vuoden taakse, tutkijat eivät ole saavuttaneet yksimielisyyttä siitä, millaisia toimintatapoja voidaan nimittää parhaiksi käytännöiksi. Toinen kiistelty aihe koskettaa sitä, tulisiko tiettyjä käytäntöjä käyttää yhdessä, jotta ne tuottaisivat parhaat mahdolliset tulokset. Toisin sanoen tutkijoilla on erilaisia käsityksiä siitä, tukevatko tietyt työkäytännöt toisiaan muodostaen nippuja, joiden merkitys yrityksen menestymiselle on suurempi kuin osiensa summa.

Työkäytäntöjen ja niistä muodostuvien nippujen sekä yrityksen suorituskyvyn välisen suhteen pohdiskeluissa paljon käytetty käsite on korkeatuottoiset työjärjestelmät eli ”high-performance work systems”, josta käytetään usein lyhennettä HPWS. Käsitettä käyttävässä tutkimusperinteessä on pyritty laajentamaan työkäytäntöjen tarkastelua yksittäisistä käytännöistä kohti laajempia organisatorisia järjestelmiä, systeemejä. HPWS-tutkimus sai alkunsa 1990-luvun alussa USA:ssa, josta se on sittemmin levinnyt mm. Isoon-Britanniaan ja muualle Eurooppaan (ks. Appelbaum ym. 2000). Suomalainen korkeatuottoisten työjärjestelmien käsitteellä operoiva empiirinen tutkimus on jäänyt varsin vähäiseksi (ks. Kauhanen 2007; Alasoini et al. 2008; Järvensivu & Koski 2009).

Lyhyesti sanoen HPWS-mallin ajatuksena on, että työjärjestelmät, jotka lisäävät työntekijöiden osallistumista, ovat joustavia työntekijöiden työn muotoilun suhteen ja hajauttavat päätöksentekoa, saavat aikaan paremman tuottavuuden, laadun, joustavuuden, innovatiivisuuden, työmotivaation ja työhyvin-

voinnin vähentäen samalla myös poissaoloja (Ichniowski et al. 1996; Järvensivu & Koski 2009, 8). Korkeatuottoisten työjärjestelmien innovatiivisina pidetyt käytännöt ajatellaan työn osittamiseen pyrkivälle tayloristiselle työn organisoinnille vaihtoehtoisiksi. Tällaisia työkäytäntöjä ovat mm. tiimityö, työn sisällön rikastaminen ja työkierto, osallistumisjärjestelmät kuten laatupiirit sekä ongelmanratkaisuryhmät.

Ichniowskin (1990) tutkimus oli ensimmäisiä, joissa analysoitiin käytäntöjen ja yritysten taloudellisten mittareiden välisiä yhteyksiä. Tutkimuksen tulosten mukaan yksittäisten käytäntöjen omaksumisella ei ole suorituskykyvaikutuksia. Merkittäviä parannuksia varsinkin laatuun ja tuottavuuteen saadaan, kun kyetään ottamaan käyttöön yhtenäinen ja integroitu käytäntöjen järjestelmä, mutta sen täytyy koostua juuri oikeasta käytäntöjen kombinaatiosta ollakseen tehokas.

1.2 Tutkimusongelman määrittely

Tässä esityksessä jatketaan edellä kuvattua tutkimuslinjaa analysoimalla suomalaisille teollisuusyritysten tuotantopäällikkötason johtajille suunnatun kyselyn pohjalta uudenlaisten toimintakäytäntöjen ja yritysten suorituskyvyn välistä suhdetta. Tutkimuskysymykset ovat:

1. *Kuinka yleisiä uudet työkäytännöt ovat suomalaisilla teollisuustyöpaikoilla?*
2. *Miten työkäytännöt vaikuttavat yrityksen suorituskykyyn erikseen käytettyinä sekä siten, että tietyt käytännöt ovat yrityksen käytössä yhtäaikaaisesti?*

Jälkimmäisen tutkimuskysymyksen kohdalla tarkastellaan neljää eri selitettävää suorituskykymuuttujaa, joista yksi kuvaa yrityksen toiminnan joustavuutta, yksi työhyvinvointia, yksi yrityksen kokonaissuorituskykyä (ns. teknis-taloudellinen suorituskyky) ja yksi yrityksen kykyä tuottaa uusia tuotteita markkinoille (tuoteinnovaatiokyky). Viimeksi mainitun muuttujan mittarina toimii kysymys, oliko kyselyyn vastannut toimipaikka tuonut markkinoille aikavälillä 2003–2005 tuotteita, jotka olivat tuotemarkkinoille täysin uusia (vastausluokat: kyllä/ei). Kyselyn yrityksistä noin 57 prosenttia ilmoitti tuoneensa markkinoille uusia tuotteita ja noin 43 prosenttia ei ollut tuonut markkinoille uusia tuotteita kyseisenä ajanjaksona. Kolme muuta suorituskykymuuttujaa ovat keskiarvomuuttujia, jotka muodostettiin yrityksen toiminnan muutosta kuvaavien liiketoiminnan mittarien pohjalta. Muuttujien muodostamiseen palataan kappaleessa 2.2.

2 Tutkimusaineisto ja aineiston muokkaus

2.1 Tutkimusaineisto

Tampereen yliopiston työelämän tutkimuskeskus toteutti keväällä 2007 kyselytutkimuksen, jossa suomalaisilta teollisuusyrityksiltä kysyttiin mm. organisatoristen innovaatioiden käytöstä ja niiden käyttöasteesta. Hanke kuului Suomen Akatemian Liike2-tutkimusohjelmaan, ja sitä rahoitti kyseisen ohjelman lisäksi myös Työsuojelurahasto. Tämä tutkielma on tehty kyseisen kyselyaineiston pohjalta. Hankkeen työnimi oli ”Organizational innovations and their role in Finnish companies’ renewal processes” (2006–2009).

Tutkimuksessa kysymysten muodoksi valittiin ennalta annettujen käytäntöjen rastittaminen siten, että kutakin käytäntöjen ryhmää koskien vastaajia pyydettiin ilmoittamaan onko kyseinen käytäntö toimipaikassa käytössä virallisesti, epävirallisesti tai ei lainkaan. Kyselyssä vastaajaa pyydettiin ilmoittamaan kaikkiaan 45 käytännön käyttönotosta, minkä arvioitiin antavan hyvän ja kattavan kuvan monien yleisimpien käytäntöjen leviämisestä suomalaisissa teollisuusyrityksissä. Vaihtoehtojen ”virallisesti käytössä” ja ”epävirallisesti käytössä” ajateltiin helpottavan vastaajaa tunnistamaan käytännön olemassaolo sekä valottavan sitä tosiseikkaa, että yrityksestä riippuen käytäntö voi olla osa organisaation virallista ja tunnustettua toimintamallia tai epävirallisemmin ja mahdollisesti vain osittain tiedostaen toteutuvaa toimintaa. Tutkimuksissa on tullut esille, että työpaikoilla toteutetaan esimerkiksi tiimityön kaltaisia työn organisoinnin tapoja, vaikkei tiimityötä ole muodollisesti otettu käyttöön.

Kyselyn kohteiksi tulleet toimipaikat valittiin toimialan ja toimipaikan koon perusteella. Toimipaikan vähimmäiskooksi asetettiin 50 henkilöä. Tämä raja perustui oletukseen, että innovatiivisia käytäntöjä on käytössä useimmin suurissa yrityksissä ja että melko suuri osa kyselyyn mukaan otetuista käytännöistä ei ole käytössä pienissä yrityksissä, ainakaan samalla tavalla muodollisesti organisoituina kuin suuremmissa toimipaikoissa (ks. Neumark & Cappelli 1999).

Toimialojen valinnan perusteena oli ottaa mukaan Suomen talouden teollisuuden ydintä edustavia aloja. Toimialat valittiin ja nimettiin Tilastokeskuksen toimialaluokitusta (TOL 2002) noudattaen. Valitut toimialat olivat seuraavat: sahatavaran ja puutuotteiden valmistus, massan, paperin ja paperituotteiden valmistus, kustantaminen ja painaminen, kemikaalien, kemiallisten tuotteiden ja tekokuitujen valmistus, kumi- ja muovituotteiden valmistus, metallituottei-

Taulukko 2.1. Teollisuuden toimialarakenteen ja kyselyyn vastanneiden vertailu (N=191).

Toimiala	Prosenttia (%)		Koko	Prosenttia (%)	
	Osuus toimialarakenteessa	Vastanneet		Osuus toimialarakenteessa	Vastanneet
Sahatavara ja puutuotteet	16	16	50-99	51	43
Massa ja paperi	13	8	100-249	32	37
Kumi ja muovi	10	17	250-499	12	12
Metallituotteet	24	21	500-999	4	5
Koneenrakennus	29	31	yli 1000	1	3
Kulkuneuvot	8	7			
Yhteensä	100	100		100	100

den valmistus, koneiden ja laitteiden valmistus sekä kulkuneuvojen valmistus. Kulkuneuvojen valmistus on Suomessa pieni toimiala, mutta sen mukaan ottamista voi perustella autoteollisuuden kansainvälisesti suurella merkityksellä ja tulosten kansainvälisen vertailtavuuden mahdollistumisella.

Kyselyn toteuttamiseksi Tilastokeskuksesta tilattiin osoitetiedot toimialan ja toimipaikan koon mukaisista Suomessa sijaitsevista toimipaikoista. Kysely lähetettiin ensimmäisen kerran kaikkiin kriteerit täyttäviin toimipaikkoihin vuoden 2007 maaliskuussa ja kahden uusintakierroksen jälkeen hyväksytyjä lomakkeita palautettiin kaikkiaan 191. Vastausprosentiksi saatiin 30,6, jota voi pitää kohtuullisena tämäntyypisessä yritys-kyselyssä.

Kyselyn palauttaneiden toimipaikkojen toimiala- ja kokorakenne vastaavat melko hyvin perusjoukkoa. Suurimmat erot ovat massa- ja paperiteollisuuden aliedustus sekä kumi- ja muoviteollisuuden yliedustus perusjoukkoon verrattuna. Koon mukaan tarkasteltuina aliedustettuina ovat pienimmät 50–99 henkeä työllistävät toimipaikat, kun taas hieman yliedustettuina ovat 100–249 työllistävät sekä jonkin verran myös yli 500 henkeä työllistävät yritykset (ks. taulukko 2.1.).

2.2 Aineiston muokkaus analyysia varten

Käytännöt luokiteltiin kyselylomakkeessa seuraavasti: 1 = ”ei ole käytössä”, 2 = ”käytössä epävirallisella tavalla” ja 3 = ”käytössä virallisella tavalla”. Yksittäisille työkäytännöille tehtyjä regressioanalyysia varten kaksi ensimmäistä luokkaa yhdistettiin luokaksi 0 = ”ei käytössä” ja viimeinen jätettiin en-

nalleen (1 = ”virallisesti käytössä”), mikä on yksi varsin usein käytetty tapa luokitella yrityksen käytössä olevien käytäntöjen käyttöasteet vastaavanlaisissa tutkimuksissa (ks. esim. Kalmi & Kauhanen 2008).¹ Oletuksena oli, että virallisesti käytössä olevilla työkäytännöillä on suurempi positiivinen vaikutus yrityksen suorituskykyyn kuin jos ne eivät ole käytössä lainkaan tai ovat vain epävirallisella tavalla käytössä. Kuitenkin luokka ”epävirallisesti käytössä” on tässä jonkin verran ongelmallinen, koska suorituskykyä kuvaavien muuttujien vastauskeskiarvot (asteikolla 1–5) eivät kaikkien käytäntöjen kohdalla eronneet kovin paljon luokissa ”epävirallisesti käytössä” ja ”virallisesti käytössä”, kun taas luokassa ”ei käytössä” vastauskeskiarvot olivat yleensä alempia.

Yrityksen suorituskykyä mitattiin kymmenellä eri liiketoiminnan mittarilla. Vastaaajaa pyydettiin arvioimaan kyseisillä mittareilla yrityksessä tapahtunutta muutosta vuosina 2003–2005. Vastausvaihtoehtoina käytettiin viisiportaista Likertin asteikkoa: 1 = ”merkittävästi huonontunut”, 2 = ”jonkin verran huonontunut”, 3 = ”ei muutosta”, 4 = ”jonkin verran parantunut” ja 5 = ”merkittävästi parantunut”.² Näistä muodostettiin jo edellä mainitut kolme selitettävää keskiarvomuuttujaa, jotka on kuvattu taulukossa 2.2. Yrityksen tuoteinnovaatiokykyä kuvaava muuttuja ei ole taulukossa, koska se muodostettiin eri tavalla (ks. kappale 2.1.).

Taulukko 2.2. Tutkimuksessa käytetyt selitettävät suorituskykymuuttujat ”kokonaissuorituskyky”, ”työhyvinvointi” ja ”joustavuus”. N=188-189.

Suorituskykymuuttuja	Keskiarvo	Keskiahajonta	Cronbachin alfa
<i>Kokonaissuorituskyky (tekniis-taloudellinen)</i>	3,74	,50	,733
Toimitusaika	3,59	,86	
Toimitusvarmuus	3,63	,89	
Tuotteiden laatu	3,88	,72	
Myynti	3,92	,83	
Markkinaosuus	3,63	,75	
Tuottavuus	3,84	,73	
Joustavuus	3,69	,81	
<i>Työhyvinvointi</i>	3,18	,77	,724
Työtyytyväisyys	3,36	,87	
Työssä jaksaminen (poissaolot)	3,00	,86	
<i>Joustavuus</i>	3,64	,69	,738
Toimitusaika	3,59	,86	
Toimitusvarmuus	3,63	,89	
Joustavuus	3,69	,81	

Taulukossa on suorituskykymuuttujien ja niihin kuuluvien osamuuttujien

¹Joskus on käytetty käytäntöjen käyttöasteesta myös viisiportaista luokitusta (ks. esim. Laugen & Acur & Boer & Frick 2005).

²Alun perin asteikko oli päinvastainen, joten se on tutkielmaa varten käännetty.

keskiarvot ja hajonnat sekä mittarien reliabiliteettia mittaavat Cronbachin alfan arvot. Suorituskykymuuttuja ”joustavuus” koostuu kolmesta eri liiketoiminnan mittarista, ”työhyvinvointi” kahdesta ja ”kokonaissuorituskyky” seitsemästä eri liiketoiminnan mittarista. Viimeksi mainitussa ovat mukana myös joustavuutta kuvaavat liiketoiminnan mittarit. Kaikki Cronbachin alfa-kertoimien arvot ovat yli 0,7, joten suorituskykymuuttujien reliabiliteetit voi katsoa riittäviksi.

Mittarin yhtenäisyyttä (konsistenssia) kuvaava Cronbachin alfan standardoitu estimaatti lasketaan seuraavasti:

$$(2.1) \quad \alpha = \frac{k\bar{r}}{1 + (k - 1)\bar{r}},$$

missä \bar{r} on mittariin kuuluvien väittämien välinen Pearsonin korrelaatiokertoimien keskiarvo ja k väittämien lukumäärä. Mitä suurempi on alfan estimoitu arvo sitä yhtenäisempänä mittaria voidaan pitää (n. 0,6–0,7:n välillä olevaa arvoa voidaan pitää riittävänä, kun alfan maksimiarvo on 1).

Analyyseissa oli mukana varsin paljon taustamuuttujia, joiden avulla pyrittiin selvittämään, vaikuttavatko pelkkien käytäntöjen (tai käytäntönippujen) lisäksi myös tietyt yrityksen rakenteelliset tekijät (kuten yrityksen toimiala ja koko) tutkittavaan suorituskykymuuttujaan. Eräiden taustamuuttujien kohdalla oli varsin paljon puuttuvaa tietoa, joten niiden mukanaolo vähensi regressioanalyyseissa mukana olevien yritysten kokonaismäärää. Puuttuvan tiedon korvaamista (imputointia) ei kuitenkaan tehty, koska suuri osa mukana olevista muuttujista on binaarisia ja jatkuvia muuttujia on varsin vähän (ks. liite 2). Yrityksiä voi kuitenkin katsoa olevan analyyseissa riittävästi ($N = 110\text{--}111$) tuloksista saatujen alustavien johtopäätösten tekemistä varten.

3 Tutkimusmenetelmät

3.1 Lineaarinen regressio

Seuraavissa kappaleissa käsitellään lyhyesti tavallisen lineaarisen regressiomallin (OLS) peruskäsitteitä. Pääpaino on mallin estimoinnissa, ei niinkään mallin hyvyyden tarkastelussa. Tämä sen vuoksi, että tässä tutkielmassa regressioanalyysia käytetään lähinnä tilastollisesti merkitsevien muuttujien (työkäytäntöjen) etsimiseen eikä optimaalisen tilastollisen mallin löytämiseksi.

3.1.1 Lineaarinen regressiomalli ja pienimmän neliösumman estimointi

Usean selittävän muuttujan lineaarisessa regressiossa selittäjiä käytetään mallintamaan yhtä selitettävää satunnaista vastemuuttujaa. Jos vaste on Y ja selittäjät, jotka oletetaan ei-satunnaisiksi, ovat X_1, X_2, \dots, X_p (p on selittäjien lukumäärä), niin vastetta mallinnetaan lineaarisella yhtälöllä

$$(3.1) \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon,$$

missä β :t ovat tuntemattomia parametreja ja ε :t tilastollisia virhetermejä (Weisberg 1985, 33).

Lineaarinen regressioyhtälö voidaan esittää matriisimuodossa

$$(3.2) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

missä

$$(3.3) \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

ja

$$(3.4) \quad \mathbf{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix},$$

missä n on havaintojen lukumäärä ja p selittäjien lukumäärä.

Jos $\boldsymbol{\varepsilon}$ on virhetermien satunnaisvektori, niin oletetaan, että $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ja $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$, missä $\mathbf{0}$ on $n \times 1$ nollavektori ja \mathbf{I}_n on $n \times n$ identiteettimatriisi (n on havaintojen lukumäärä). Virhetermit oletetaan riippumattomiksi ja korreloimattomiksi eli $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$. Lisäksi virhetermien oletetaan olevan normaalijakautuneita eli (emt., 42)

$$(3.5) \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n).$$

Virhetermien varianssit ovat näin ollen $\text{Var}(\varepsilon_i) = \sigma^2$. Regressiokertoimien ja vakiotermin tapaan σ^2 on mallin tuntematon parametri, joka on aineiston perusteella estimoitavissa (Weisberg 1985, 42; Puntanen 2007).

Pienimmän neliösumman estimaatit $\hat{\boldsymbol{\beta}}$ minimoivat jäännöseliösumman eli funktion (Weisberg 1985, 42) ¹

$$(3.6) \quad SSE(\boldsymbol{\beta}) = \sum_{i=1}^n (y_i - \mathbf{x}'_i \boldsymbol{\beta})^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

missä \mathbf{x}'_i on mallimatriisin \mathbf{X} i . rivi. Selitettävän muuttujan Y sovitearvojen vektori saadaan kaavasta $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, jolloin vektorin i . elementti on $\hat{y}_i = \mathbf{x}'_i \hat{\boldsymbol{\beta}}$. Virhetermien vektori saadaan havaittujen y_i :den ja y :n sovitearvojen erotuksena eli $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}}$, jolloin i . virhetermi on $\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \mathbf{x}'_i \boldsymbol{\beta}$. Funktion (kaava 3.6) ratkaisuna saadaan jäännöseliösummaksi (emt., 44)

$$(3.7) \quad SSE = \hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}} = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}).$$

Pienimmän neliösumman ratkaisu $\boldsymbol{\beta}$:lle eli $\hat{\boldsymbol{\beta}}$ saadaan yhtälöstä (emt., 43)

$$(3.8) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y},$$

edellyttäen, että $\mathbf{X}'\mathbf{X}^{-1}$ on olemassa. Ratkaisuna saatujen regressiokertoimien estimoitu varianssi on (emt., 44)

$$(3.9) \quad \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}, \quad \hat{\sigma}^2 = \frac{SSE}{(n - p - 1)},$$

missä $p - 1$ on estimoitavien parametrien lukumäärä eli regressiokertoimien lukumäärä - vakiotermi (jos sellainen on mallissa). Vakiotermi $\hat{\beta}_0$ saadaan puolestaan ratkaistua yhtälöstä (emt., 44)

$$(3.10) \quad \hat{\beta}_0 = \bar{y} - \hat{\boldsymbol{\beta}}^* \bar{\mathbf{x}}, \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}}^* \end{pmatrix}.$$

Estimoinnin tuloksena saatujen parametrien arvot ovat harhattomia (olettaen, että on voimassa $E(\boldsymbol{\varepsilon}) = \mathbf{0}$) eli $E(\hat{\sigma}^2) = \sigma^2$, $E(\hat{\beta}_0) = \beta_0$ ja $E(\hat{\beta}_i) = \beta_i$, $i = 1, \dots, p$ (emt., 12–13; Vehkalahti 2009, 2).

¹Weisberg käyttää lyhennettä $RSS(\boldsymbol{\beta})$, joka tulee englanninkielisestä termistä ”residual sum of squares”.

3.1.2 Neliösummahajotelmat ja selitysasteet

Taulukossa 3.1. on kuvattu lineaarisen regressiomallin keskeiset neliösummahajotelmat (Puntanen 2007; Vehkalahti 2009, 3):

Taulukko 3.1. Neliösummahajotelmat mallissa $\mathbf{X} = (\mathbf{1} : \mathbf{x}_1 : \dots : \mathbf{x}_p)$.

Neliösumma (SS)	df	$MS = SS/df$
$SSR = \sum(\hat{y}_i - \bar{y})^2$	p	
$SSE = \sum(y_i - \hat{y}_i)^2$	$n - p - 1$	$MSE = SSE/(n - p - 1) = s^2$
$SST = \sum(y_i - \bar{y}_i)^2$	$n - 1$	$MST = SST/(n - 1) = \text{Var}(\mathbf{y})$

Mallin kokonaisvariانسsi ($SST = \text{Var}(\mathbf{y})$) saadaan summaamalla yhteen regressioneliösumma (SSR) ja jäännöseliösumma (SSE). Mallin arvioimisen kannalta neliösummat ovat tärkeitä, koska niiden avulla saadaan laskettua mallin selitysaste (selityskerroin), joka on:

$$(3.11) \quad R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}.$$

Mallin selitysaste on näin ollen mallin systemaattisen vaihtelun (SSR) ja kokonaisvaihtelun ($SST = \text{Var}(\mathbf{y})$) suhde. Selitysasteen positiivinen neliöjuuri on puolestaan mallin yhteiskorrelaatiokerroin eli

$$(3.12) \quad R = \sqrt{\frac{SSR}{SST}}$$

Yhteiskorrelaatiokerroin on havaitun y :n ja mallista lasketun sovitetun korrelaatiokerroin:

$$(3.13) \quad R = \text{kor}(\mathbf{y}, \hat{\mathbf{y}}), \quad 0 \leq R = \text{kor}(\mathbf{y}, \hat{\mathbf{y}}) \leq 1.$$

Yhden selittäjän regressiomallissa yhteiskorrelaatiokerroin on sama kuin selittävän ja selitettävän muuttujan välinen korrelaatiokerroin eli $R^2 = r_{xy}^2$ (Puntanen 2007).

Tilasto-ohjelmien tulostuksissa on usein laskettu myös ns. sovitettu tai korjattu selitysaste (*adjusted R-Square*), joka ottaa huomioon selittäjien ja havaintojen lukumäärät. Sovitettu selitysaste lasketaan kaavalla

$$(3.14) \quad 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1},$$

missä p on selittäjien lukumäärä, n havaintojen lukumäärä ja R^2 mallin selitysaste. Jos havaintojen lukumäärä on suhteellisen pieni ja mallissa on mukana useita ei-merkitseviä selittäjiä, voi sovitettu selitysaste olla negatiivinen.

Esimerkki 3.1.1. Liitteessä 3 on tulostukset yksittäisille käytännöille tehdyistä regressioanalyyseistä. Muuttujalle ”kokonaissuorituskyky” on saatu selitysasteeksi $R^2 = 0,249$. Selittäviä muuttujia on 30, joista yksikään ei ole merkitsevä 5 prosentin merkitsevyystasolla (ja vain kaksi muuttujaa on merkitseviä 10 prosentin merkitsevyystasolla). Havaintojen lukumäärä on 111. Sovitetuksi selitysasteeksi saadaan tällöin:

$$1 - \frac{(1 - 0,249)(111 - 1)}{111 - 30 - 1} = -0,0326.$$

3.1.3 Hypoteesien testaus

Regressiomallissa voidaan testata yksittäisiä regressiokertoimia tai kaikkien selittävien muuttujien kertoimien yhteisvaikutusta. Yksittäisten kertoimien tapauksessa testataan hypoteeseja:

$$\begin{aligned} H_0 &: \beta_i = 0 \\ H_1 &: \beta_i \neq 0, \quad (i = 0, 1, \dots, p). \end{aligned}$$

H_0 :n ollessa tosi testisuure noudattaa studentin t -jakaumaa vapausastein $n - p - 1$

$$(3.15) \quad t = \frac{\hat{\beta}_i}{\hat{\sigma}(\hat{\beta}_i)} \sim t_{n-p-1},$$

missä p on selittäjien lukumäärä ja $\hat{\sigma}(\hat{\beta}_i)$ on regressiokertoimen β_i estimoitu hajonta. Jos mallissa ei ole vakiotermiä vapausasteet ovat $n - p$. Regressiokertoimien yhteistestauksessa (olettaen, että mallissa on vakio-termi mukana) hypoteesit ovat:

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \dots = \beta_p = 0 \\ H_1 &: \text{ainakin jokin } \beta_i \neq 0, \quad (i = 0, 1, \dots, p). \end{aligned}$$

Vakiotermiä β_0 ei siis testata. Testisuure (ns. overall- F -test value) voidaan lausua seuraavasti käyttämällä apuna selitysasteen laskemisessa käytettyjä neliö- ja keskineliösummia (Puntanen 2007):

$$\begin{aligned} (3.16) \quad F = F(\beta_1, \dots, \beta_p) &= \frac{MSR}{MSE} = \frac{SSR/p}{SSE/(n-p-1)} \\ &= \frac{SSR/p}{(SST - SSR)/(n-p-1)} \\ &= \frac{\frac{SSR}{SST}/p}{\frac{SST-SSR}{SST}/(n-p-1)} \\ &= \frac{R^2/p}{(1-R^2)/(n-p-1)}. \end{aligned}$$

Kyseisen mallin tapauksessa H_0 :n ollessa tosi testisuure noudattaa F -jakaumaa vapausastein p ja $n - p - 1$

$$(3.17) \quad F(\beta_1, \dots, \beta_p) \sim F(p, n - p - 1).$$

Nollahypoteesi hylätään riskitasolla α , jos aineistosta laskettu F -testisuureen arvo on suurempi kuin $F_{\alpha; p, n-p-1}$, missä viimeksi mainittu tarkoittaa F -jakauman kriittistä arvoa riskitasolla α eli (Puntanen 2007)

$$(3.18) \quad P(F_{p, n-p-1} > F_{\alpha; p, n-p-1}) = \alpha.$$

Esimerkki 3.1.2. Esimerkissä 3.1.1. käytetystä yksittäisille käytännöille tehdystä analyysistä (liite 3) voidaan edelleen laskea F -testisuure muuttujalle ”kONAISSUORITUSKYKY”:

$$F = F(\beta_1, \dots, \beta_p) = \frac{0,249/30}{(1 - 0,249)/(111 - 30 - 1)} = 0,884.$$

R-ohjelmalla saadaan p -arvoksi $(1 - \text{pf}, 0.884, 30, 80)$ noin 0,639, joka poikkeaa hieman SPSS-ohjelman laskemasta arvosta (0,641), mutta johtopäätös on kuitenkin selvä eli regressiomallia ei voi pitää testin mukaan käyttökelpoisena.

3.1.4 Muita mallin arvioinnin kriteerejä

Jos regressiomallissa olevat selittävät muuttujat korreloivat voimakkaasti keskenään siten, että tämä aiheuttaa ongelmia saatujen tulosten luotettavuudelle, sanotaan näiden muuttujien olevan keskenään multikollineaarisia. Usean selittäjän regressiomallissa selittäjät X_1, X_2, \dots, X_p ovat multikollineaarisia, jos vakioille c_1, c_2, \dots, c_p on voimassa (Weisberg 1985, 197)

$$(3.19) \quad c_1 X_1 + c_2 X_2 + \dots + c_p X_p = C_0.$$

Tällöin ainakin yksi selittäjä X_k voidaan (ainakin likimain) lausua toisten lineaarikombinaationa eli (ibid.)

$$(3.20) \quad X_k \cong \left(c_0 - \sum_{i \neq k} c_i X_i \right) / C_k, \quad i = 1, \dots, p.$$

Multikollineaarisuustilanne voi syntyä vahingossa, jos mallissa on kaksi muuttujaa, jotka mittaavat käytännössä samaa asiaa. Multikollineaarisuutta voi mitata ns. VIF-arvoilla (*variance inflation factor*), jotka saadaan laskettua kaavalla (emt, 198; Vehkalahti 2009, 14)

$$(3.21) \quad \text{VIF}_i = \frac{1}{1 - R_i^2}, \quad i = 1, \dots, p,$$

missä R_i^2 on selitysaste mallista, jossa R_i :tä selitetään mallin muilla selittäjillä.

Jos mallille on voimassa oletus (3.5), niin tällöin on voimassa myös (Vehkalahti 2009, 17)

$$(3.22) \quad \varepsilon_i \sim N[0, \sigma^2(1 - h_{ii})], \quad i = 1, \dots, n,$$

missä h_{ii} on matriisin $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ i . lävistäjäalkio. Tämän seurauksena (ibid.)

$$(3.23) \quad \frac{\varepsilon_i}{\sigma\sqrt{1 - h_{ii}}} \sim N(0, 1), \quad i = 1, \dots, n,$$

minkä perusteella saadaan tietoa residuaalien teoreettisesta käyttäytymisestä. Lävistäjäalkioita h_{ii} kutsutaan vetovoima-arvoiksi (*leverages*), jolloin suuren vetovoima-arvon omaavia alkioita kutsutaan vaikutusvaltaisiksi havainnoiksi. Vetovoima-arvoja voi käyttää poikkeavien havaintojen etsimisessä esimerkiksi ulkoisesti studentoitujen residuaalien (r_i^*) avulla, jotka lasketaan mallin estimoitujen arvojen perusteella seuraavasti (Weisberg 1985, 116)

$$(3.24) \quad r_i^* = \frac{y_i - \hat{y}_i}{\hat{\sigma}_{(i)}\sqrt{1 - h_{ii}}}, \quad \hat{\sigma}_{(i)} = s_{(i)} = \sqrt{SSE_{(i)}/(n - p - 2)}, \quad i = 1, \dots, n,$$

missä $\hat{\sigma}_{(i)}$ on jäännöshajonnan estimaatti mallissa, josta on poistettu i . havainto (Vehkalahti 2009, 18).

Residuaalitarkastelujen lisäksi usein käytetty mitta poikkeavien havaintojen löytämiseksi on Cookin etäisyys (*Cook's distance*). Cookin etäisyysmitat voidaan ratkaista yhtälöstä (Emt., 18; Weisberg 1985, 120)

$$(3.25) \quad c_i = \frac{1}{p'} r_i^2 \left(\frac{h_{ii}}{1 - h_{ii}} \right), \quad r_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}, \quad i = 1, \dots, n,$$

missä r_i :t ovat standardoituja residuaaleja, $\hat{\sigma}$ jäännöshajonnan estimaatti (s) ja p' mallimatriisin \mathbf{X} aste $\text{rank}(\mathbf{X})$, joka on vakiotermin sisältävässä mallissa $p + 1$.

Matriisin \mathbf{H} ominaisuuksien vuoksi vetovoima-arvot h_{ii} ovat keskimäärin p/n , jolloin vaikutusvaltaisina havaintoina pidetään tavallisesti arvoja $h_{ii} > (2p)/n$, $0 \leq h_{ii} \leq 1$. Residuaalitarkasteluissa huomio kannattaa kiinnittää havaintoihin, joilla standardoidun (r_i) tai studentoidun (r_i^*) residuaalin arvot ovat itseisarvoltaan suurempia kuin 2. Cookin etäisyyksissä puolestaan on hyödyllistä yleensä tutkia arvot, joissa $c_i > 0,5$ (Vehkalahti 2009, 17-18). VIF-arvoille ei ole asetettu tarkkoja rajoja, mutta yleensä tilanteessa $VIF \geq 10$ voidaan multikollineaarisuutta epäillä jo melko voimakkaaksi (Hair ym. 1998).

3.2 Logistinen regressio

Logistinen regressioanalyysi (logit-malli) on yleistetty lineaarinen malli, ja se sopii tilanteisiin, joissa selitettävä muuttuja on binaarinen eli saa vain kaksi arvoa. Selittävinä muuttujina voidaan käyttää sekä kvantitatiivisia (jatkuvia) että kvalitatiivisia (kategorisia) muuttujia. Selitettävä muuttuja voi saada myös useampia arvoja (vähintään 3), jolloin menetelmänä käytetään multinomiaalista logistista regressiota (moniluokkaiset logit-mallit).

3.2.1 Logistinen regressiomalli

Todennäköisyyden ja riskin käsitteet ovat keskeisiä logistisessa regressiossa. Menetelmä pyrkii ennustamaan todennäköisyyttä, jolla tarkasteltavana oleva asia tapahtuu tai pätee. Menetelmällä saaduista tuloksista nähdään, vaikuttavatko selittävät muuttujat tapahtuman toteutumistodennäköisyyteen ja kuinka suuri vaikutus on. Yleensä tapahtumien vaihtoehdot koodataan siten, että selitettävä muuttuja Y saa arvon 1, jos tapahtuma toteutuu ja muuten 0. Jos Y on Bernoullin jakaumaa noudattava selitettävä muuttuja ja X selittävä muuttuja, niin merkitään (ks. esim. Isotalo 2009 ja 2010)

$$(3.26) \quad \pi(\mathbf{x}_i) = \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}}}, \quad i = 1, \dots, n,$$

missä vektori \mathbf{x}_i sisältää havaintomatriisin i :nnen rivin kaikki arvot. Näin ollen muuttujan Y tulosvaihtoehdon 1 todennäköisyys $\pi(\mathbf{x}_i)$ on epälineaarisesti riippuva selittävien muuttujien arvoista. Logistisessa regressiomallissa vedonlyöntikerroin (*odds*) $\gamma(\mathbf{x}_i)$ määritellään (emt.)

$$(3.27) \quad \gamma(\mathbf{x}_i) = \frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} = e^{\mathbf{x}'_i \boldsymbol{\beta}}, \quad i = 1, \dots, n.$$

Kun vedonlyöntikertoimesta otetaan logaritmi, saadaan logistisen regressiomallin yhtälö (emt.)

$$(3.28) \quad \log(\gamma(\mathbf{x}_i)) = \log\left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right) = \text{logit}(\pi(\mathbf{x}_i)) = \mathbf{x}'_i \boldsymbol{\beta}, \quad i = 1, \dots, n.$$

Yleistetyissä lineaarisissa malleissa kuten logistisessa regressiossa mallinnetaan näin ollen selitettävän muuttujan Y odotusarvosta $E(Y) = \mu$ riippuvan linkkifunktion $g\mu$ arvoa selittävien muuttujien X_1, X_2, \dots, X_p avulla käyttämällä lineaarista yhtälöä (Isotalo 2010)

$$(3.29) \quad g(\mu) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} = \mathbf{x}'_i \boldsymbol{\beta}, \quad i = 1, \dots, n.$$

Linkkifunktio $g(\mu)$ yhdistää selittävät muuttujat X_1, X_2, \dots, X_p selitettävän muuttujan (satunnaiskomponentin) Y arvoihin. Koska logistisessa regressiossa odotusarvot ovat todennäköisyyksiä eli $0 \leq \mu \leq 1$, niin käytetty linkki on tällöin logit-linkki, joka on muotoa (emt.)

$$(3.30) \quad g(\mu) = \log\left(\frac{\mu(\mathbf{x}_i)}{1 - \mu(\mathbf{x}_i)}\right), \quad i = 1, \dots, n.$$

Estimoidussa logistisessa regressiomallissa vasteen Y estimoitu toteutumistodennäköisyys ($Y = 1$) saadaan seuraavasti:

$$(3.31) \quad \hat{\pi}(x) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p)},$$

missä p on selittäjien lukumäärä.

Ristitulosuhde (*odds ratio*) määritellään vedonlyöntikertoimien osamääränä

$$(3.32) \quad \theta = \frac{\text{odds}_1}{\text{odds}_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}.$$

Ristitulosuhteen perusteella voidaan tutkia estimoidussa mallissa muuttujien *odds ratio* -suhteiden muutosta $\exp(\hat{\beta}_j)$, missä $j = 1, \dots, p$. Näin voidaan selvittää, kuinka paljon kyseinen suhde muuttuu, kun selittävän muuttujan arvo kasvaa yhden yksikön.

Esimerkki 3.2.1. Liitteessä 4 on tulostus logistisesta regressiosta muuttujalle ”tuoteinnovaatiot” käytäntökomponenttien tapauksessa. Muuttujan ”korkeasti koulutetun henkilöstön prosenttiosuus” (`high_edu3`) *odds ratio* -arvo on $e^{0,039} = 1,039$ (tulostuksen $\text{Exp}(B)$ -sarake). Osuuden kasvaessa 5 prosenttiyksiköllä saadaan uudeksi *odds ratio* -arvoksi $1,039^5 = 1,211$ eli prosenttiosuuden kasvaessa todennäköisyys sille, että yritys on tuonut uusia tuotteita markkinoille kasvaa.

3.2.2 Suurimman uskottavuuden estimointi

Yleistettyjen lineaaristen mallien tapauksessa estimoidaan parametrivektori β , joka liittyy satunnaismuuttujiin Y_i odotusarvon $E(Y_i) = \mu_i$ sekä linkkifunktion $g(\mu) = \eta_i = \mathbf{x}_i' \beta$ kautta. Vektori \mathbf{x}_i sisältää i :n:n havainnon kaikki arvot ja satunnaisotos $\mathbf{y} = (Y_1, Y_2, \dots, Y_n)' = (y_1, y_2, \dots, y_n)'$. Tällöin jokaiselle Y_i :lle on voimassa uskottavuusfunktio l_i , joka on muotoa (Isotalo 2010)

$$(3.33) \quad l_i(\Theta_i, \phi) = \log(f(y_i; \Theta_i, \phi)) = \frac{y_i \Theta_i - b(\Theta_i)}{a(\phi)} + c(y_i, \phi), \quad i = 1, \dots, n$$

ja satunnaisotoksen \mathbf{y} uskottavuusfunktio l on vastaavasti muotoa (ibid.)

$$(3.34) \quad \begin{aligned} l(\Theta, \phi) &= \log(f(\mathbf{y}; \Theta, \phi)) = \sum_{i=1}^n l_i(\Theta_i, \phi) \\ &= \sum_{i=1}^n \frac{y_i \Theta_i - b(\Theta_i)}{a(\phi)} + c(y_i, \phi). \end{aligned}$$

Parametrivektorin $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ suurimman uskottavuuden estimaatit saadaan, kun ratkaistaan logaritmoidun uskottavuusfunktion osittaisderivaattojen nollakohdat yhtäaikaaisesti. Tällöin päädytään suurimman uskottavuuden estimointiyhtälöihin, jotka ovat muotoa (ks. tarkemmin esim. Agresti 1990, 447–449; Isotalo 2010)

$$(3.35) \quad \frac{\partial l(\Theta, \phi)}{\partial \beta_j} = \sum_{i=1}^n \frac{y_i - \mu_i}{\text{Var}(Y_i)} \frac{1}{g'(\mu_i)} x_{ij} = \sum_{i=1}^n \frac{y_i - \mu_i}{\text{Var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) = 0,$$

missä $j = 1, \dots, p$.

Usein ratkaisuyhtälöillä ei ole suljetun muodon ratkaisua, jolloin ne joudutaan ratkaisemaan numeerisesti käyttämällä esimerkiksi Newton-Raphson tai Fisherin scoring-menetelmiä.

Binomijakauman tilanteessa $\text{Var}(Y_i) = \pi_i(1 - \pi_i)/n_i$, jolloin

$$\frac{\partial \mu_i}{\partial \eta_i} = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^{-1} = \left(\frac{1}{\pi_i(1 - \pi_i)} \right)^{-1} = \pi_i(1 - \pi_i).$$

Estimointiyhtälöt binomijakautuneelle satunnaismuuttujalle ovat näin ollen muotoa (Isotalo 2010)

$$\begin{aligned} \sum_{i=1}^n \frac{y_i - \mu_i}{\text{Var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) &= \sum_{i=1}^n \frac{y_i - \pi_i}{\frac{\pi_i(1-\pi_i)}{n_i}} x_{ij} (\pi_i(1 - \pi_i)) \\ &= \sum_{i=1}^n n_i (y_i - \pi_i) x_{ij}, \quad j = 1, \dots, p. \end{aligned}$$

Koska $\pi_i = e^{\mathbf{x}'_i \beta} / (1 + e^{\mathbf{x}'_i \beta})$, saadaan estimointiyhtälöiksi binomijakauman tapauksessa (ibid.)

$$(3.36) \quad \sum_{i=1}^n n_i \left(y_i - \frac{e^{\mathbf{x}'_i \beta}}{1 + e^{\mathbf{x}'_i \beta}} \right) x_{ij} = 0, \quad j = 1, \dots, p.$$

Suurimman uskottavuuden estimointiyhtälöt binomijakauman tilanteessa joudutaan ratkaisemaan numeerisesti käyttämällä em. iteratiivisia menetelmiä (Agresti 1990, 449–451; Isotalo 2010; Myers et al 2002, 101–108).

3.2.3 Logistisen regressiomallin arviointi

Logistisessa regressiossa käytetään Waldin testisuuretta regressiokertoimien testauksessa. Testattaessa hypoteeseja

$$\begin{aligned} H_0 : \beta_j &= 0 \\ H_1 : \beta_j &\neq 0, \quad (j = 0, 1, \dots, p) \end{aligned}$$

Waldin testisuure

$$(3.37) \quad Z = \frac{\hat{\beta}_j}{\hat{\sigma}(\hat{\beta}_j)}$$

noudattaa tällöin asympotoottisesti standardoitua normaalijakaumaa H_0 :n ollessa tosi. Eri mallien vertailussa voidaan käyttää myös mallien devianssien erotuksia. Jos esimerkiksi malli \mathcal{M}_0 on muotoa $\text{logit}(\pi(\mathbf{x})) = \beta_0$ eli sisältää vain vakiotermin ja \mathcal{M}_1 on muotoa $\text{logit}(\pi(\mathbf{x})) = \beta_0 + \beta_1 x$ eli sisältää vakiotermin lisäksi yhden selittäjän, niin mallien devianssien erotus $D(M_0) - D(M_1)$ noudattaa χ^2 -jakaumaa vapausastein $df = 1$, jos H_0 on tosi. Yleisemmin H_0 (malli \mathcal{M}_0 on voimassa) hylätään riskitasolla α , jos erotus ΔD kuuluu kriittiselle alueelle eli jos se on suurempi kuin $\chi^2_{(df(\mathcal{M}_1) - df(\mathcal{M}_0))}$ -jakauman kertymäfunktion $100(1 - \alpha)$ % arvo.

Regressiokertoimelle β_j voidaan muodostaa Waldin testisuureen avulla asymp-
toottinen $100(1 - \alpha) \%$ luottamusväli kaavalla (Isotalo 2010)

$$(3.38) \quad \hat{\beta}_j \pm z_{\alpha/2} \hat{\sigma}(\hat{\beta}_j), \quad j = 0, 1, \dots, p,$$

jolloin $P(Z > z_{\alpha/2}) = \alpha/2$, kun testisuure $Z \sim N(0, 1)$.

Mallin estimoidun varianssin $\hat{\sigma}^2(\text{logit}(\hat{\pi}(\mathbf{x}))) = \mathbf{x}'\hat{\boldsymbol{\beta}} \pm z_{\alpha/2} \sqrt{\mathbf{x}'\widehat{\text{cov}}(\hat{\boldsymbol{\beta}})\mathbf{x}}$ avul-
la voidaan muodostaa vasteen toteutumistodennäköisyyden $\pi(\mathbf{x})$ luottamusvä-
liestimaatti seuraavasti (emt.)

$$(3.39) \quad \left(\frac{e^{\mathbf{x}'\hat{\boldsymbol{\beta}} - z_{\alpha/2} \sqrt{\mathbf{x}'\widehat{\text{cov}}(\hat{\boldsymbol{\beta}})\mathbf{x}}}}{1 + e^{\mathbf{x}'\hat{\boldsymbol{\beta}} - z_{\alpha/2} \sqrt{\mathbf{x}'\widehat{\text{cov}}(\hat{\boldsymbol{\beta}})\mathbf{x}}}}, \frac{e^{\mathbf{x}'\hat{\boldsymbol{\beta}} + z_{\alpha/2} \sqrt{\mathbf{x}'\widehat{\text{cov}}(\hat{\boldsymbol{\beta}})\mathbf{x}}}}{1 + e^{\mathbf{x}'\hat{\boldsymbol{\beta}} + z_{\alpha/2} \sqrt{\mathbf{x}'\widehat{\text{cov}}(\hat{\boldsymbol{\beta}})\mathbf{x}}} \right).$$

Devianssien erotusten $\Delta D = D(\mathcal{M}_0) - D(\mathcal{M})$ avulla voidaan laskea ns.
pseudoselitysaste, jota käytetään yhtenä mallin sopivuuden mittarina. Yleisesti
se määritellään seuraavasti (emt.)

$$(3.40) \quad R_{\mathcal{M}}^2 = \frac{D(\mathcal{M}_0) - D(\mathcal{M})}{D(\mathcal{M}_0)} = 1 - \frac{D(\mathcal{M})}{D(\mathcal{M}_0)},$$

missä \mathcal{M}_0 on pelkän vakiotermin sisältävä malli. Usein logistisessa regressiossa
käytetään Nagelkerken selitysastetta, joka voi saada arvoja 0 ja 1 väliltä. Jos
vertaillaan malleja \mathcal{M}_1 ja \mathcal{M}_2 , niin mallien selitysasteet ovat (emt.)

$$(3.41) \quad R_{\mathcal{M}_1}^2 = \frac{1 - e^{(D(\mathcal{M}_1) - D(\mathcal{M}_0))/n_{++}}}{1 - e^{-D(\mathcal{M}_0)/n_{++}}}$$

$$R_{\mathcal{M}_2}^2 = \frac{1 - e^{(D(\mathcal{M}_2) - D(\mathcal{M}_0))/n_{++}}}{1 - e^{-D(\mathcal{M}_0)/n_{++}}},$$

missä $D(\mathcal{M}_0)$ saadaan mallista, joka sisältää vain vakiotermin eli

$$\mathcal{M}_0 = \log \left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right) = \text{logit}(\pi(\mathbf{x}_i)) = \beta_0.$$

Kaavan n_{++} on puolestaan havaintojen kokonaismäärä N . Muita mallin ar-
vioinnissa käytettyjä selitysasteita ovat McFaddenin ja Cox & Snell -selitysas-
teet, joiden arvoalueet eivät ulotu 1:een asti, jolloin näiden selitysasteiden arvot
jäävät yleensä Nagelkerken selitysastetta pienemmiksi.

Mallien \mathcal{M}_1 ja \mathcal{M}_2 vertailussa voidaan käyttää myös informaatiokriteerejä
kuten Akaiken informaatiokriteeriä (AIC). Malleista sopivampi on tällöin se,
jonka informaatiokriteerin arvo on pienempi.

Estimoidun mallin sopivuutta aineistoon voidaan testata Hosmer–Lemeshow
-testillä. Menetelmää käytettäessä todennäköisyysväli 0–1 jaetaan 10 osaan ja
tutkitaan näiden 10 luokan odotettujen frekvenssien ja havaittujen frekvens-
sien välisiä eroja. Testisuure noudattaa Khi(2)-jakaumaa vapausastein $df =$
 $(10 - 2) = 8$. Jos malli on käyttökelpoinen, testin tulos on ei-merkitsevä.

Eräissä tilasto-ohjelmissa (esim. SPSS) on mahdollista käyttää ns. askelta-
via menetelmiä, joissa malliin lisätään tai siitä vähennetään selittäviä muuttu-
jia. Esimerkiksi eteenpäin askeltavassa menetelmässä käytetään uskottavuus-
suhdetestiä, jossa tutkitaan $-2 \log$ likelihood -arvojen avulla, tuovatko uudet
muuttujat malliin lisäarvoa. Jos Khi(2) -jakaumaa noudattavan testisuureen ar-
vo on merkitsevä, tuodaan uusi muuttuja malliin (ks. esim. Myers et al. 2002,
112).

Mallin arvioinnissa voidaan käyttää lisäksi residuaalitarkasteluja. Logistisen
regressiomallin ns. raakaresiduaalit ovat muotoa $e_i = y_i - \hat{\pi}_i$. Jos $\hat{\pi}(\mathbf{x}_i)$ on mallin
antama sovite todennäköisyydelle $\pi(\mathbf{x}_i)$ selittäjien \mathbf{x}_i arvoilla, niin Pearsonin
residuaalit e_i saadaan osamääränä (Isotalo 2010):

$$(3.42) \quad e_i = \frac{y_i(\mathbf{x}_i) - n_i(\mathbf{x}_i)\hat{\pi}(\mathbf{x}_i)}{\sqrt{n_i(\mathbf{x}_i)\hat{\pi}(\mathbf{x}_i)(1 - \hat{\pi}(\mathbf{x}_i))}},$$

missä muuttujan \mathbf{x}_i arvoilla on toistettu bernoullin koetta $n_i(\mathbf{x}_i)$ kertaa. Täl-
löin $y_i(\mathbf{x}_i)$ on onnistumisten ($Y = 1$) lukumäärä selittävien muuttujien arvoilla
 \mathbf{x}_i . Kuten lineaarisessa regressiomallissa voidaan myös logistisessa regressiossa
määritellä standardoidut residuaalit. Logistisessa regressiomallissa ne määritel-
lään osamääränä (ibid.)

$$(3.43) \quad r_i = \frac{y_i(\mathbf{x}_i) - n_i(\mathbf{x}_i)\hat{\pi}(\mathbf{x}_i)}{\sqrt{n_i(\mathbf{x}_i)\hat{\pi}(\mathbf{x}_i)(1 - \hat{\pi}(\mathbf{x}_i))(1 - \hat{h}_i(\mathbf{x}_i))}},$$

missä $\hat{h}(\mathbf{x}_i)$:t ovat selittävästä muuttujista riippuvia vetovoima-arvoja (*leverage*).

Näiden lisäksi voidaan laskea devianssiresiduaalien arvot (ibid.)

$$(3.44) \quad d_i = \sqrt{q_i} \times \text{sign}(y_i(\mathbf{x}_i) - n_i(\mathbf{x}_i)\hat{\pi}(\mathbf{x}_i)),$$

missä

$$q_i = 2 \left(y_i(\mathbf{x}_i) \log \left(\frac{y_i(\mathbf{x}_i)}{n_i(\mathbf{x}_i)\hat{\pi}(\mathbf{x}_i)} \right) + (n_i(\mathbf{x}_i) - y_i(\mathbf{x}_i)) \log \left(\frac{n_i(\mathbf{x}_i) - y_i(\mathbf{x}_i)}{n_i(\mathbf{x}_i) - n_i(\mathbf{x}_i)\hat{\pi}(\mathbf{x}_i)} \right) \right).$$

Mallin sopivuutta dataan voidaan tarkastella residuaalien ja selittävien
muuttujien tai sovitearvojen $\hat{\pi}(\mathbf{x}_i)$ pisteparvikuvioilla (esimerkiksi R-ohjelman
plot-funktio). Kuitenkin residuaalitarkastelujen käyttökelpoisuutta saattavat
rajoittaa tilanteet, joissa onnistumisten $n_i(\mathbf{x}_i)$ lukumäärä on hyvin pieni (eri-
tyisesti, jos $n_i(\mathbf{x}_i) = 1$).

Esimerkki 3.2.2. Liitteen 3 SPSS-ohjelman tulosteessa on muuttujalle ”tuo-
teinnovaatiot” tehdyn logistisen regression tulokset yksittäisten käytäntöjen
tapauksessa. Tulostuksesta nähdään Nagelkerken selitysaste (0,418) ja Cox &
Snellin selitysaste (0,312). McFaddenin selitysaste voidaan laskea kaavalla

$$(3.45) \quad R^2 = 1 - \left(\frac{-2L}{-2L_0} \right),$$

missä $-2L_0$ on pelkän vakiotermin sisältävän mallin $-2 \log$ likelihood -arvo ja $-2L$ koko mallin $-2 \log$ likelihood -arvo. Tulostuksen perusteella MacFaddenin selitysasteeksi saadaan $1 - (110, 403/151, 582) = 0, 272$. Näin ollen sekä Cox & Snellin että McFaddenin selitysasteet jäävät melko selvästi Nagelkerken selitysastetta alemmiksi.

3.3 Pääkomponenttialyysi

Pääkomponenttialyysia (samoin kuin faktorianalyysia) käytetään usein tutkimusasetelmissa, joissa on tarkoituksena kuvata monimuuttujaisen aineiston vaihtelua siten, että alkuperäisistä muuttujista muodostetaan muutama lineaarikombinaatio eli pääkomponentti (tai faktori). Tarkoitus on näin tiivistää havaintoaineistoa, jolloin pääkomponentteja voidaan käyttää jatkoanalyysissa kuten regressioanalyysissa. Pääkomponentit määritellään keskenään korreloimattomiksi ja ensimmäisten pääkomponenttien tulee selittää suurin osa aineiston vaihtelusta eli varianssista. Seuraavaksi esitellään lyhyesti menetelmän matemaattisia perusteita (ks. esim. Johnson & Wichern 2002, 426–476; Luoma 2008; Sharma 1996, 58–89).

3.3.1 Pääkomponentit

Määritelmä 3.3.1. *Satunnaisvektorin $\mathbf{x} = (x_1, \dots, x_p)'$ jakauman pääkomponentit ovat lineaarikombinaatiot:*

$$\begin{cases} y_1 = \mathbf{a}'_{(1)}\mathbf{x} = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \\ y_2 = \mathbf{a}'_{(2)}\mathbf{x} = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p \\ \vdots \\ y_p = \mathbf{a}'_{(p)}\mathbf{x} = a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p, \end{cases}$$

joita toteuttavat seuraavat ehdot:

1. $\text{Var}(y_1)$ on mahdollisimman suuri ehdolla $\mathbf{a}'_1\mathbf{a}_1 = a_{11}^2 + \dots + a_{1p}^2 = 1$ ja
 2. $\text{Var}(y_i)$ on mahdollisimman suuri ehdoilla $\mathbf{a}'_k\mathbf{a}_k = 1$, $k = 2, \dots, p$
- ja $\text{Cov}(y_k, y_j) = 0$, $j = 1, \dots, k - 1$.

Ensimmäinen kohta asettaa ehdon ensimmäisen pääkomponentin varianssille ja toisessa ilmaistaan pääkomponenttien korreloimattomuusehto. Merkitään $\mathbf{y} = \mathbf{A}\mathbf{x}$, missä \mathbf{y} on pääkomponenttivektori, \mathbf{x} on alussa mainittu satunnaisvektori ja \mathbf{A} kerroinmatriisi.²

Pääkomponentit saadaan ratkaistua ominaisarvohajotelmasta $\Sigma = \mathbf{T}\mathbf{\Lambda}\mathbf{T}'$, missä $\Sigma = \text{Cov}(\mathbf{x})$, \mathbf{T} on ortogonaalinen matriisi, jolle on voimassa $\mathbf{T}\mathbf{T}' = \mathbf{I}$, ja $\mathbf{\Lambda}$ on diagonaalimatriisi, jonka elementit ovat laskevassa järjestyksessä eli $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots > 0$. Näitä diagonaalielementtejä kutsutaan matriisin Σ ominaisarvoiksi. Ominaisarvojen kovarianssimatriisi on $\text{Cov}(\mathbf{y}) = \text{Cov}(\mathbf{T}'\mathbf{x}) = \mathbf{T}'\Sigma\mathbf{T} = \mathbf{T}'\mathbf{T}\mathbf{\Lambda}\mathbf{T}'\mathbf{T} = \mathbf{\Lambda}$ eli i . pääkomponentin varianssi on $\text{Var}(y_i) = \lambda_i$ (Luoma 2008).

²Usein y -n tilalla käytetään kreikkalaisia aakkosia kuten ξ :tä (Sharma 1996, 66).

3.3.2 Koordinaatiston kierto

Pääkomponenttien estimoinnissa tapahtuu ns. koordinaatiston kierto, jonka avulla pyritään saamaan pääkomponentit erottumaan mahdollisimman selvästi toisistaan. On osoitettavissa, että kappaleessa 3.3.1 esitetyn pääkomponentti-vektorin yhtälössä $\mathbf{y} = \mathbf{A}\mathbf{x}$ matriisi $\mathbf{A} = \mathbf{T}'$. Näin ollen saadaan $\mathbf{y} = \mathbf{T}'\mathbf{x}$. Kun yhtälö kerrotaan vasemmalta matriisilla \mathbf{T} , saadaan $\mathbf{T}\mathbf{y} = \mathbf{T}\mathbf{T}'\mathbf{x} = \mathbf{I}\mathbf{x} = \mathbf{x}$. Ortogonaalisella matriisilla \mathbf{T} kertominen vastaa koordinaatiston kiertoa, joten matriisia \mathbf{T} kutsutaan kiertomatriisiksi (Luoma 2008).

Esimerkiksi kahden pääkomponentin ja kaksiulotteisen aineiston tapauksessa saadaan yhtälöstä $\mathbf{y} = \mathbf{T}'\mathbf{x}$

$$(3.46) \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} t_{11} & t_{21} \\ t_{12} & t_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

jolloin pääkomponentit y_1 ja y_2 ovat

$$\begin{cases} y_1 = t_{11}x_1 + t_{21}x_2 \\ y_2 = t_{12}x_1 + t_{22}x_2 . \end{cases}$$

Vastaavasti x_1 ja x_2 voidaan esittää pääkomponenttien avulla yhtälöllä $\mathbf{x} = \mathbf{T}\mathbf{y}$ eli yhtälöstä (3.46) saadaan

$$\begin{cases} x_1 = t_{11}y_1 + t_{12}y_2 \\ x_2 = t_{21}y_1 + t_{22}y_2 . \end{cases}$$

Satunnaisvektori \mathbf{x} voidaan esittää myös muodossa

$$(3.47) \quad \mathbf{x} = \mathbf{T}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{-1/2}\mathbf{y} = \mathbf{L}\mathbf{u},$$

missä $\mathbf{L} = \mathbf{T}\mathbf{\Lambda}^{1/2}$ on pääkomponenttimatriisi ja $\mathbf{u} = \mathbf{\Lambda}^{-1/2}\mathbf{y}$ standardoitujen pääkomponenttien vektori. Muuttujien x_i ja pääkomponenttien y_j (i ja $j = 1, \dots, p$) väliset korrelaatiot saadaan standardoidun jakauman tapauksessa (jolloin $\text{Cov}(\mathbf{x}) = \mathbf{I}$) matriisista \mathbf{L} eli $\text{Cor}(x_i, y_j) = l_{ij}$ (Luoma 2008).

Standardoidun aineiston tilanteessa voidaan havaintomatriisin i :nnes havaintovektori \mathbf{x}_i esittää muodossa (kun ensimmäinen rivi kerrotaan vasemmalta matriisilla \mathbf{T})

$$(3.48) \quad \begin{aligned} \mathbf{y}_i &= \mathbf{T}'(\mathbf{x}_i - \boldsymbol{\mu}) \\ \mathbf{x}_i - \boldsymbol{\mu} &= \mathbf{T}\mathbf{y}_i \\ \mathbf{x}_i &= \mathbf{T}\mathbf{y}_i + \boldsymbol{\mu} \\ &= \mathbf{T}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{-1/2}\mathbf{y}_i + \boldsymbol{\mu} \\ &= \mathbf{T}\mathbf{\Lambda}^{1/2}\mathbf{u}_i + \boldsymbol{\mu} \\ &\approx u_{i1}\sqrt{\lambda_1}\mathbf{t}_1 + u_{i2}\sqrt{\lambda_2}\mathbf{t}_2, \quad i = 1, \dots, n, \end{aligned}$$

missä $\boldsymbol{\mu}$ on vektori, joka sisältää eri muuttujien keskiarvot, \mathbf{y}_i i :n havainnon pääkomponenttivektori, \mathbf{u}_i i :n havainnon standardoitu pääkomponenttivektori sekä \mathbf{t}_1 ja \mathbf{t}_2 kiertomatriisiin \mathbf{T} kaksi ensimmäistä saraketta. Näin ollen havaintovektorin \mathbf{x}_i komponentti x_{ik} on

$$(3.49) \quad x_{ik} \approx t_{k1} \sqrt{\lambda_1} u_{i1} + t_{k2} \sqrt{\lambda_2} u_{i2} + \bar{x}$$

ja muuttujan x_k koordinaatit ovat $(t_{k1} \sqrt{\lambda_1}, t_{k2} \sqrt{\lambda_2})$ pääkomponenttien u_1 ja u_2 koordinaatistossa (Luoma 2008).

3.3.3 Pääkomponenttien estimointi

Pääkomponenttien estimointi voidaan tehdä joko käyttämällä kovarianssimatriisia \mathbf{S} tai tekemällä singulaariarvohajotelma keskistetyille havaintomatriisille. Suositeltavaa on kuitenkin käyttää korrelaatiomatriisia, jos tutkittavia muuttujia on paljon ja jos ne eivät ole yhteismitallisia (Luoma 2008).

Pääkomponenttianalyysin tuloksena saatuja estimoituja arvoja kutsutaan pääkomponenttipistemääriksi. Ne lasketaan yleensä keskistystä tai standardoidusta aineistosta. Keskistetyn aineiston tapauksessa havaintovektoria \mathbf{x}_i vastaava pistemäärävektori on $\mathbf{y}_i = \mathbf{T}'(\mathbf{x}_i - \bar{\mathbf{x}})$, jos oletetaan, että $\mathbf{S} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}'$ on kovarianssimatriisin ominaisarvohajotelma ja $\tilde{\mathbf{X}} = \mathbf{X} - \bar{\mathbf{X}}$ on keskistetty havaintomatriisi. Keskistetyn aineiston tapauksessa kaikkien pistemäärien matriisi $\mathbf{Y} = \tilde{\mathbf{X}}\mathbf{T}$ (emt.).

Standardoidun aineiston tilanteessa erotusmatriisi on $\mathbf{Z} = (\mathbf{X} - \bar{\mathbf{X}})\mathbf{D}^{-1/2}$, missä $\mathbf{D} = \text{diag}(s_{11}, \dots, s_{pp})$ ja $\mathbf{R} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}$ on vastaavasti korrelaatiomatriisin ominaisarvohajotelma, jolloin standardoitua havaintovektoria $\mathbf{z}_i = \mathbf{D}^{-1/2}(\mathbf{x}_i - \bar{\mathbf{x}})$ vastaava pistemäärävektori on $\mathbf{y}_i = \mathbf{T}'\mathbf{z}_i$ ja kaikkien pistemäärien matriisi on $\mathbf{Y} = \mathbf{Z}\mathbf{T}$ (emt.).

Satunnaisvektorin \mathbf{x} kokonaisvarianssi saadaan summaamalla sen komponenttien varianssit eli

$$(3.50) \quad \sum_{i=1}^p \text{Var}(x_i) = \text{tr}(\boldsymbol{\Sigma}) = \text{tr}(\boldsymbol{\Lambda}) = \sum_{i=1}^p \text{Var}(y_i),$$

joka voidaan merkitä edelleen muodossa $\sigma_{11} + \dots + \sigma_{pp} = \lambda_1 + \dots + \lambda_p$. Näin ollen i ensimmäistä pääkomponenttia selittävät

$$(3.51) \quad \frac{\lambda_1 + \lambda_1 + \dots + \lambda_i}{\lambda_1 + \lambda_1 + \dots + \lambda_p} \cdot 100\%$$

kokonaisvarianssista (Johnson & Wichern, 428–429; Luoma 2008).

Pääkomponenttien estimoinnissa voidaan käyttää ortogonaalisena kierto menetelmänä Varimax-rotatiota kuten tässä tutkielmassa on tehty. Menetelmässä pääkomponentit on estimoitu korrelaatiomatriisin pohjalta. Ortogonaalisessa kierrossa pääkomponentit ovat suorassa kulmassa toisiinsa nähden. Eri-tyisesti faktorianalyyseissa voidaan käyttää myös ns. vinoja kierto menetelmiä (Promax, Oblimin). Kiertomenetelmiä ei käsitellä tässä tarkemmin.

Yleensä jatkoanalyysiin valitaan mukaan ne pääkomponentit, joiden varianssi on suurempi tai yhtäsuuri kuin 1. Toinen vaihtoehto on valita ne pääkomponentit, joiden jälkeen niiden varianssissa tapahtuu selvä pudotus.

3.4 Rakenneyhtälömallit

Rakenneyhtälömalleihin (*Structural Equation Models, SEM*) luetaan useantyyppiset regressioyhtälö- ja regressiopolkumallit, konfirmatorinen faktorianalyysi (CFA) sekä faktoreiden polkumallit ja latentit kasvukäyrämallit. Seuraavissa kappaleissa käsitellään lyhyesti regressioyhtälömalleja sekä regressiopolkumalleja. Aineiston lopullinen mallinnus on tässä tutkielmassa tehty käyttämällä rakenneyhtälömalleja ja LISREL-ohjelmaa.

3.4.1 Usean yhtälön regressiomallit

Kahden tai useamman yhtälön regressiomallit ovat muotoa

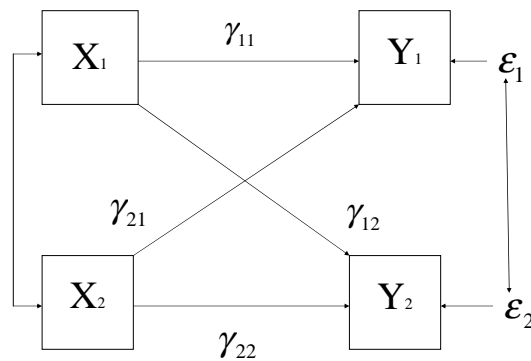
$$(3.52) \quad \mathbf{y} = \Gamma \mathbf{x} + \boldsymbol{\varepsilon},$$

missä \mathbf{y} ja $\boldsymbol{\varepsilon}$ ovat $p \times 1$ vektoreita, \mathbf{x} on $q \times 1$ vektori ja Γ sisältää mallin estimoitavat regressiokertoimet (p on yhtälöiden lukumäärä mallissa).

Jos kahden yhtälön mallissa on kummassakin yhtälössä kaksi selittävää muuttujaa eli $p = 2$ ja $q = 2$, saadaan esimerkiksi seuraavat yhtälöt:

$$\begin{cases} Y_1 = \gamma_{11}X_1 + \gamma_{21}X_2 + \varepsilon_1 \\ Y_2 = \gamma_{12}X_1 + \gamma_{22}X_2 + \varepsilon_2 \end{cases} .$$

Kyseiset yhtälöt estimoitavine parametreineen on kuvattu kuviossa 3.1.



Kuvio 3.1. Kahden yhtälön regressiomalli.

Kyseessä on saturoitu malli, koska mallissa on mukana kaikki muuttujien väliset vaikutukset eli $df = 0$ ja $\chi^2 = 0$. Saturoidussa mallissa estimoidaan kaikki mallin regressiokertoimet (γ_i -parametrit), virhetermit (ε_i tai ψ_{ii} -parametrit) sekä kovarianssimatriisin ψ_{ij} -parametrit.³ Usean yhtälön regressiomalli on aina identifioituva eli yksilöityvä. Jos malli ei ole saturoitu, sen estimointi on iteratiivista, jolloin saadaan regressiokertoimien estimaatit, t -arvot sekä virhetermien estimaatit. Lisäksi saadaan käyttöön χ^2 -testi mallin hyvyystarkastelua varten (Leskinen 2009).

3.4.2 Rekursiiviset ja simultaaniset regressiopolkumallit

Rekursiivinen polkumalli on muotoa

$$(3.53) \quad \mathbf{y} = \mathbf{B}\mathbf{y} + \mathbf{\Gamma}\mathbf{x} + \boldsymbol{\varepsilon}.$$

Kuvion 3.1. mallista saadaan rekursiivinen malli lisäämällä yhdensuuntainen nuoli selitettävien muuttujien Y_1 ja Y_2 välille. Jos ajatellaan esimerkiksi Y_1 :llä olevan vaikutusta Y_2 :een, niin saadaan regressioyhtälöiksi

$$\begin{cases} Y_1 = \gamma_{11}X_1 + \gamma_{21}X_2 + \varepsilon_1 \\ Y_2 = \beta_{12}Y_1 + \gamma_{12}X_1 + \gamma_{22}X_2 + \varepsilon_2, \end{cases}$$

missä kerroin β ilmaisee selitettävien Y -muuttujien välisen vaikutussuhteen. Rekursiivisen mallin tapauksessa estimoidaan kaikki mallin regressiokertoimet, sekä virhetermit, mutta kovarianssimatriisin ψ_{ij} -parametreja (jotka siis oletetaan nolliksi) ei estimoida.⁴ Jos ψ_{ij} -parametrit estimoidaan, voi syntyä tilanne, jossa $df < 0$, jolloin mallissa on liikaa estimoitavia parametreja. Rekursiivinen malli on muuten identifioituva.

Regressiopolkumalleissa saadaan muuttujien suorien vaikutusten lisäksi selville niiden epäsuorat vaikutukset. Kahden muuttujan epäsuora vaikutus mallissa on niiden välisten polkujen suorien vaikutusten tulojen summa. Jos esimerkiksi kuvion 3.1. tilanteessa Y_1 vaikuttaisi Y_2 :een, niin X_1 :n epäsuora vaikutus Y_2 :een olisi $\gamma_{12} + \gamma_{11}\beta_{12}$.

Simultaaninen polkumalli saadaan kuvioista 3.1., jos selitettävien muuttujien Y_1 ja Y_2 välillä on molemminpuolinen vaikutussuhde (kuviossa nuolet molempiin suuntiin). Simultaaninen malli ei kuitenkaan ole aina identifioituva. Mallin identifioituvuudelle on voimassa seuraavat ehdot, jos oletetaan, että mallissa on Y -muuttujia p kappaletta, X -muuttujia q kappaletta ja yhtälöitä i kappaletta (Leskinen 2009):

- $p_i + q_i > q$: i :nnes yhtälö on ali-identifioituva (ei-identifioituva), koska parametreja on liikaa
- $p_i + q_i = q$: i :nnes yhtälö on identifioituva

³Rakenneyhtälömalleissa käytetään virhetermeistä usein merkintää ψ_{ii} .

⁴LISREL-ohjelma jättää oletusarvoisesti ψ_{ij} -parametrit estimoimatta.

- $p_i + q_i < q$: i :nnes yhtälö on yli-identifioituva (on siis identifioituva)

Jos kuvion 3.1. tilanteessa lisättäisiin Y_1 ja Y_2 välille molemminpuoliset vaikutussuhteet, saataisiin kaksi yhtälöä, jotka eivät olisi identifioituvia:

$$\begin{cases} Y_1 = \beta_{21}Y_2 + \gamma_{11}X_1 + \gamma_{21}X_2 + \varepsilon_1 \\ Y_2 = \beta_{12}Y_1 + \gamma_{12}X_1 + \gamma_{22}X_2 + \varepsilon_2 . \end{cases}$$

Tässä tapauksessa kummassakin yhtälössä $p_i + q_i > q$, koska mallissa on X -muuttujia 2, jolloin saadaan $1 + 2 = 3 > q$. Mallista tulee identifioituva, jos kummastakin yhtälöstä poistetaan yksi selittävä muuttuja.

3.4.3 Mallin arviointikriteerejä

Regressioyhtälö- ja regressiopolkumallien hyvyttä kokonaisuutena voidaan arvioida χ^2 -testillä (paitsi saturoidun mallin tilanteessa). Testin tulos on ei-merkittävä, jos malli on käyttökelpoinen. Usean mallin vertailussa voidaan käyttää informaatiokriteerejä (esim. AIC), jolloin valitaan se malli, jonka informaatiokriteerin arvo on pienin sillä edellytyksellä, että malli on riittävä muiden tarkastelujen perusteella.

Normeerattu yhteensopivuusideksi NFI (*normed fit index*) arvioi suuren otoskoon vaikutusta χ^2 -testissä. Yleensä riittävänä pidetään arvoa $NFI \geq 0,90$. Kuitenkin suuren otoskoon ($N > 500$) tapauksessa voi syntyä tilanne, jolloin χ^2 -testin p -arvo $< 0,05$, mutta $NFI \geq 0,90$. Tällöin otoskoon voi olettaa vaikuttaneen liikaa χ^2 -testin arvoon. Mallia voidaan pitää riittävänä, jos osamäärä $\chi^2/df < 2$ (Leskinen 2009).

Muita mallin arvioinnin kriteerejä (sopivuusindeksejä) ovat mm. RMSEA, GFI- ja CFI-kriteerit sekä SRMR. RMSEA (*root mean square error of approximation*) mittaa mallin yksinkertaistamisesta johtuvaa approksimointivirhettä. Yleensä riittävänä pidetään, jos $RMSEA \leq 0,05$. Jos $RMSEA \geq 0,08$, on mallia modifioitava. GFI (*goodness-of-fit index*) vertaa teoreettista ja havaittua mallia toisiinsa. Mallia pidetään riittävänä, jos $GFI \geq 0,95$. CFI (*comparative fit index*) eli suhteellinen yhteensopivuusindeksi on samantyyppinen kuin GFI, mutta se huomioi samalla myös mallien vapausasteet. Myös tässä tapauksessa mallia voi pitää hyvänä, jos $CFI \geq 0,95$. SRMR (*standardized root mean square residual*) eli standardoitujen residuaalien keskiarvo -indeksi vertaa otosvariansseja ja -kovariansseja estimoituihin populaatiota koskeviin variansseihin ja kovariansseihin ja ilmoittaa luvut standardoituina. Mallia voi pitää riittävänä, jos $SRMR \leq 0,05$. (Leskinen 2009).

Edellisten koko mallia koskevien tarkastelujen lisäksi voidaan tehdä parametri- ja havaintokohtaisia tarkasteluja kuten tavallisessa regressioanalyysissä. LISREL-ohjelman tulostuksissa on luettavissa regressiokertoimien t -arvot, jolloin tilanteessa $|t| \geq 2$ parametri on merkitsevä 5 prosentin merkitsevyystasolla ($\alpha = 0,05$). Parametrikohtaisissa tarkasteluissa LISREL-ohjelma tutkii nollakiinnitykset (esim. rekursiivisten polkumallien tapauksessa parametrit ψ_{ij}

asetetaan nolliksi). Ohjelma laskee ns. modifikaatioindeksin (MI) arvot kaikille nollakiinnityksille. Tarkoituksena on tutkia ovatko nollakiinnitykset ”oikeita”. MI-arvot noudattavat Khi(2)-jakaumaa vapausastein $df = 1$. Koska $\chi_{0.95;1}^2 = 3,84$, on mallia modifioitava siten, että nollakiinnitettyjä parametreja vapautetaan estimoitaviksi yksi kerrallaan, jos $MI \geq 4$. Tässä tapauksessa on kuitenkin huolehdittava mallin tulkinnallisuuden säilymisestä (emt.).

Havaintokohtaisissa tarkasteluissa voidaan tutkia standardoituja jäännöksiä SR_{ij} . Jos malli on riittävä, tulisi jäännösten olla normaalijakautuneita eli

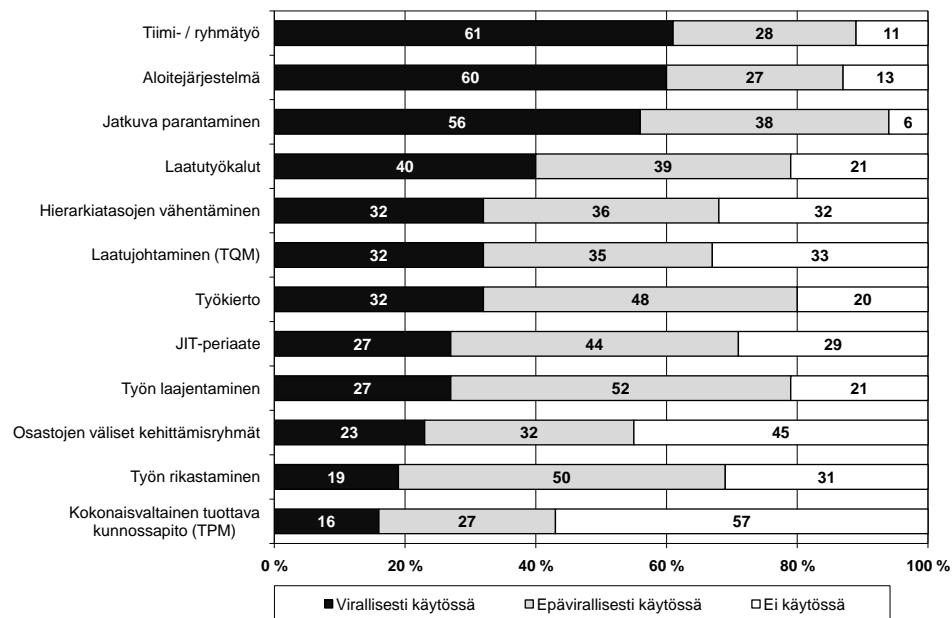
$$(3.54) \quad SR_{ij} \sim N(0, 1), \quad i, j = 1, \dots, p.$$

Niissä tapauksissa, joissa $|SR_{ij}| \geq 2$, malli ei selitä ko. havaintoja riittävän hyvin. Jos tällaisia havaintoja on paljon, voidaan mallin sopivuutta yrittää parantaa lisäämällä siihen parametreja (emt.).

4 Aineiston kuvaus

4.1 Organisatoristen innovaatioiden käyttö kyselyn toimipaikoissa

Alla olevassa kuviossa 4.1. on esitetty tutkimukseen valittujen 12 käytännön yleisyys ja käyttöaste kyselyn toimipaikoissa. Tutkimukseen valittiin tarkas-



Kuvio 4.1. Innovatiiviset työkäytännöt kyselyn toimipaikoissa. N=183-190.

teltaviksi ne valmistusprosessin ja työn organisoinnin käytännöt, jotka olivat vähintään 15 prosentilla kyselyyn vastanneista yrityksistä virallisesti käytössä. Näin varmistettiin, että kyseinen ryhmä on riittävän suuri yksittäisille käytännöille tehtyjä jatkoanalyysseja varten. Tiedon johtamisen käytäntöjä (18 kpl) ei otettu mukaan, koska ne muodostavat selvästi oman käytäntöryhmänsä. Mukaan valikoitui lopuksi 12 käytäntöä.

Tehdyn kyselyn mukaan yleisimmät, yli puolella vastanneista virallisesti käytössä olleet tuotantoprosessin ja työn organisoinnin innovatiiviset käytännöt olivat tiimityö, työntekijöiden aloitejärjestelmät ja jatkuvan parantamisen ohjelmat.

Tiimityö on selvästi eniten käyttöönotettu uusi työkäytäntö, kun taas työkierto, hierarkiatasojen vähentäminen ja työn laajentaminen on omaksuttu noin 30 prosentissa vastaajaryyksistä. Työn rikastaminen, jossa työtä pyritään kehittämään liittämällä siihen suunnittelu- ja päätöksentekotehtäviä, on sen sijaan virallisesti käytössä vain noin viidenneksellä vastaajaryyksistä. Vähiten mukaan valituista käytännöistä oli virallisesti käytössä tuotantokatkos- ja koneikkotilanteisiin liittyvä kokonaisvaltainen tuottava kunnossapito (TPM, virallisesti käytössä 16 prosentilla vastaajista eli 29 toimipisteessä).

4.2 Innovatiiviset työkäytännöt yrityksen toimialan ja koon mukaan

Toimialan ja toimipaikan koon on todettu olevan yksi merkittävä innovatiivisten käytäntöjen käyttöönottoon vaikuttava tekijä (Blasi & Kruse 2006). Tässä tutkimuksessa eniten käytössä olevien käytäntöjen kohdalla toimialoitainen analyysi tuottaa melko yhdenmukaisen tuloksen, sillä työntekijöiden aloitejärjestelmien, jatkuvan parantamisen ohjelmien ja tiimityön prosenttiosuudet nousevat korkeimmiksi kaikilla toimialoilla. Nämä ovat tyypillisesti käytäntöjä, jotka voi ajatella olevan sovellettavissa helposti toimialasta riippumatta. Koneenrakennusteollisuudessa tiimityön osuudet ovat kaikkein korkeimmat toimialojen sisäisessä vertailussa (71 prosentilla vastaajista virallisesti käytössä).

Taulukko 4.1. Innovatiiviset käytännöt toimipaikan koon mukaan (prosentteina, virallisesti käytössä; N=183-190).

Käytäntö	50-99	100-249	yli 250	Yhteensä (%)
Tiimi- / ryhmätyö	55	61	74	61
Aloitejärjestelmät	40	71	82	60
Jatkuvan parantamisen ohjelmat	48	62	66	56
Laatutyökalujen käyttö	33	39	55	40
Hierarkiatasojen vähentäminen	27	40	27	32
Laatujohdaminen (TQM)	21	35	54	32
Työkierto	24	40	37	32
Just-in-time eli juuri oikeaan tarpeeseen -periaate (JIT)	23	28	35	27
Työn laajentaminen	22	29	35	27
Osastojen väliset kehittämissryhmät	10	31	39	23
Työn rikastaminen	15	20	26	19
Kokonaisvaltainen tuottava kunnossapito (TPM)	10	19	23	16

Tulokset vahvistavat myös oletusta, että koon mukaan tarkasteltuna suurimmat yritykset ovat ahkerimpia innovatiivisten käytäntöjen soveltajia (ks. taulukko 4.1.). Poikkeuksina ovat hierarkiatasojen vähentäminen ja työkierto, joissa keskisuurten yritysten prosentit ovat korkeimmat. Kokoryhmittäin tarkasteltuna pienimmässä kokoryhmässä (50–99) yleisimpiä innovatiivisia käytäntöjä ovat järjestyksessä tiimityö, jatkuvan parantamisen ohjelmat ja aloitejärjestelmät, keskisuurilla aloitejärjestelmät, jatkuva parantaminen ja tiimityö sekä suurilla aloitejärjestelmät, tiimityö ja jatkuvan parantamisen ohjelmat.

5 Analyysitulokset

5.1 Yksittäiset käytännöt ja yrityksen suorituskyky

Kaikille neljälle suorituskykymuuttujalle tehtiin ensin regressioanalyysit. Suorituskykymuuttujalle ”tuoteinnovaatiot” tehtiin logistinen regressioanalyysi ja muille lineaariset regressioanalyysit. Käytäntömuuttujat ovat analyysissa kaksiluokkaisia (0=epävirallisesti käytössä tai ei käytössä, 1=virallisesti käytössä). Yksittäisten käytäntöjen vaikutus yritysten suorituskykyyn on kuvattu tau-

Taulukko 5.1. Regressioanalyysi: yksittäisten käytäntöjen vaikutus suorituskykymuuttujiin ”kokonaissuorituskyky”, ”työhyvinvointi” ja ”joustavuus”. Käytäntöjen luokitus: 0=epävirallisesti käytössä tai ei käytössä, 1=virallisesti käytössä. Merkitsevyysluokitus: * merkitsevä 0,1: n, ** merkitsevä 0,05: n ja *** merkitsevä 0,01: n merkitsevyystasolla. N=111.

Selitettävä suorituskykymuuttuja	Kokonaissuorituskyky		Työhyvinvointi		Joustavuus	
	Regressio-kerroin	p-arvo	Regressio-kerroin	p-arvo	Regressio-kerroin	p-arvo
KÄYTÄNTÖ						
Työn laajentaminen	,022	,917	-,049	,873	-,028	,923
Työn rikastaminen	,199	,379	,100	,758	,212	,483
Työkierro	,139	,279	,446	,017**	,189	,272
TQM	-,057	,652	-,135	,461	-,005	,978
Laatutyökalut	-,056	,716	,623	,006***	-,242	,240
JIT-periaate	-,043	,735	,005	,978	,026	,880
Osastojen väliset kehittämissryhmät	-,018	,893	-,298	,129	,053	,772
Työntekijöiden aloitejärjestelmät	,004	,976	-,173	,357	,062	,721
TPM	,082	,559	-,110	,590	,033	,862
Jatkuvan parantamisen ohjelmat	,243	,082*	,057	,777	,424	,024**
Tiimityö	-,011	,933	-,054	,769	,043	,801
Hierarkiatasojen vähentäminen	-,035	,801	,046	,821	-,054	,770
Vakiotermi	4,073	,000***	3,917	,000***	4,162	,000***
Yrityksiä /lkm	111		111		111	
Selitysaste	0,249		0,359		0,270	
F-testi	0,882		1,50*		0,99	

lukossa 5.1. Yksittäisistä käytännöistä vain jatkuvan parantamisen ohjelmat osoittautuivat merkitseväksi yrityksen kokonaissuorituskyvyn ja joustavuuden kannalta. Työhyvinvointiin positiivisesti vaikuttavia käytäntöjä olivat työkier-

to ja laatutyökalujen käyttö. Tämän perusteella näyttäisi siltä, että huomion kiinnittäminen toiminnan jatkuvaan parantamiseen olisi joustavuuden kannalta olennaista. Vaikutus kokonaissuorituskykyyn johtuu suurelta osin siitä, että joustavuutta kuvaavat kolme osamittaria ovat myös kokonaissuorituskykyä kuvaavassa muuttujassa mukana (ks. taulukko 2.2.)

Työhyvinvoinnin mittari koostuu työtyytyväisyydestä ja poissaoloista. Tähän mittariin laatutyökalujen käyttämisellä oli positiivisesti erittäin merkittävä vaikutus ja työkierrolla merkittävä vaikutus. Laatutyökalujen positiivisen vaikutuksen tulkitseminen on ongelmallista työkalujen suuren kirjon vuoksi. Niihin voidaan lukea kuuluviksi esimerkiksi laatupiirityyppiset käytännöt, tilastolliset laadunparantamismenetelmät (six sigma), työntekijöiden laatuvas- tuuseen liittyvät käytännöt sekä erilaiset laadunhallinnan työkalut (tarkistus- listat yms.). Työhyvinvoinnin ja laatutyökalujen yhteys voi selittyä osittain sillä, että esimerkiksi laatupiiritoiminnan kautta työntekijät voivat saada konkreettisia parannuksia työoloihinsa sekä osittain siten, että laatuongelmia yhteisesti käsitellessä ymmärrys kokonaisprosessista ja oman työn vaikutuksesta siihen saattaa kasvaa (Godard 2001). Työkierrolla on mahdollista lisätä tehtävien vaihtelevuutta ja monipuolisuutta, millä saattaa teollisuusyrityksissä olla työhyvinvointia kohentava vaikutus. Millään käytännöllä ei ollut selkeästi negatiivista vaikutusta suorituskykyymuuttujiin.

Regressioanalyysissa oli mukana myös kontrollimuuttujia (tai taustamuuttujia), joilla pyrittiin selvittämään, missä määrin tietyt yrityksen rakenteelliset tekijät (kuten toimiala ja yrityksen koko) vaikuttavat yrityksen suorituskykyyn mukaan valittujen käytäntöjen lisäksi. Kontrollimuuttujat on kuvattu liitteessä 2, mutta taulukkoon 5.1. on otettu tilan säästämisen vuoksi mukaan vain käytännöt. Kontrollimuuttujista yrityksen koko vaikutti negatiivisesti merkittävästi työhyvinvointiin. Suurimmissa yrityksissä työhyvinvointi oli pienempiä toimipaikkoja huonommalla tasolla. Ainoa positiivinen vaikutus kontrollimuuttujista oli korkeasti koulutetun henkilöstön osuudella kaikista työntekijöistä. Se vaikutti lievästi positiivisesti yrityksen kokonaissuorituskykyyn.

Suorituskykyymuuttujalle ”tuoteinnovaatiot” tehtiin erillinen regressioanalyysi, jonka tulokset on kuvattu taulukossa 5.2. Suhteessa yrityksen kykyyn tuottaa markkinoille uusia tuotteita ainoa merkittävä selittävä muuttuja yksittäisistä käytännöistä oli työkierto. Työkierron osoittautuminen myönteiseksi uusien tuotteiden kehittämisen näkökulmasta tuntuu varsin erikoiselta tulokselta. Kenties jonkinlaista selityskehikkoa muodostuu siten, että työkierron työntekijöille hahmottuu valmistusprosessista laajempi kokonaisuus ja siten syntyy uusia ajatuksia myös tuotekehityksen kannalta. Toisaalta on mahdollista, että yhteys selittyy muilla tekijöillä ja sellaisista asetelmista, joihin käytetty kysely ei ulotu. Laatutyökalujen käyttö on melko lähellä 10 prosentin merkittävyydestä, mutta sen selitysosuutta ei voi pitää merkittävänä. Muut yksittäiset käytännöt ovat uusien tuotteiden kehittämisen näkökulmasta selvästi ei-merkittäviä (ks. liite 3).

Regressioanalyysien yhteydessä ei ollut tarkoitus kiinnittää paljonkaan huomiota mallin hyvyysarviointeihin, vaan etsiä tilastollisesti merkittäviä käytän-

Taulukko 5.2. Regressioanalyysi: yksittäisten käytäntöjen vaikutus suorituskykymuuttujaan ”tuoteinnovaatiot”. Käytäntöjen luokitus: 0=epävirallisesti käytössä tai ei käytössä, 1=virallisesti käytössä. Merkitsevyysluokitus: * merkitsevä 0,1: n, ** merkitsevä 0,05: n ja *** merkitsevä 0,01: n merkitsevyystasolla. N=110.

Selitettävä suorituskykymuuttuja	Tuoteinnovaatiot		
	Regressio-kerroin	Wald	p-arvo
KÄYTÄNTÖ			
Työn laajentaminen	-,126	,010	,920
Työn rikastaminen	,446	,134	,714
Työkierto	1,618	5,862	,015**
TQM	-1,006	2,285	,131
Laatutyökalut	1,340	2,642	,104
JIT-periaate	-,254	,152	,696
Osastojen väliset kehittämissryhmät	-,102	,023	,880
Työntekijöiden aloitejärjestelmät	-,318	,221	,638
TPM	-,086	,015	,902
Jatkuvan parantamisen ohjelmat	,524	,475	,491
Tiimityö	,337	,280	,597
Hierarkiatasojen vähentäminen	,724	1,119	,290
<i>Vakiotermi</i>	-2,018	,740	,390
<i>Yrityksiä/ lkm</i>	110		
<i>-2 log likelihood</i>	110,403		
<i>Selitysaste (Nagelkerke)</i>	0,418		
<i>Oikein luokiteltuja havaintoja (%)</i>	72,7		

töjä yritysten suorituskyvyn kannalta. Kuitenkin on syytä tarkastella tässä yhteydessä selittävien muuttujien mahdollista multikollinearisuutta, koska se saattaa aiheuttaa epäluotettavuutta mallin estimointituloksiin.

Selitysasteet eri suorituskykymuuttujille lineaarisissa regressioissa ovat noin 0,25:n ja 0,36:n välillä, mutta sovitetut selitysasteet ovat käytännössä nolli, koska analyysitulosten mukaan suurin osa selittävästä muuttujista on ei-merkitseviä. Lisäksi *F*-testin tulokset ovat ei-merkitseviä 5 prosentin merkitsevyystasolla, joten tilastollisina malleina saatuja regressiotuloksia ei voi pitää käyttökelpoisina (ks. esimerkit 3.1.1. ja 3.1.2. sekä liitteen 3 analyysitulokset). Logistisessa regressiossa selitysaste on tyydyttävä (0,418) samoin kuin oikein luokiteltujen havaintojen osuus (72,7 %). Lisäksi myös Hosmer-Lemeshow – testin mukaan malli olisi riittävä, mutta tässäkin tapauksessa suuri osa selittäjistä on ei-merkitseviä.

Lineaarisisissa regressioanalyyseissa suurimmat VIF-arvot olivat työkäytännöllä työn sisällön rikastaminen (VIF: 4,325 ja toleranssiarvo: 0,231) ja työn laajentaminen (VIF: 4,002 ja toleranssiarvo: 0,250).¹ Näiden käytäntöjen välinen Spearmanin korrelaatiokerroin on 0,784. Muut VIF-arvot ovat selvästi täs-

¹ Alle 0,20:n olevaa toleranssiarvoa voidaan pitää jo merkinä mahdollisesta multikollinearisuudesta.

tä alempia. Koska VIF-arvot ovat ≤ 5 , voidaan nämä käytännöt pitää mukana analyysissä ottaen huomioon, että yksittäisille käytännöille tehtyjen analyysien tuloksia ei tässä tukielmassa käytetä tilastollisen mallin rakentamiseen.

Residuaaleissa ei näyttäisi eri suorituskykymuuttujien kohdalla olevan varsinaisia outlier-havaintoja, jos rajana pidetään itseisarvoltaan 3 suuruisia arvoja. Itseisarvoltaan suurin standardoitu residuaali on suorituskykymuuttujalle ”joustavuus” ($-2,634$). Logistisen regression tulostuksissa on lueteltu suorituskykymuuttujalle ”tuoteinnovaatiot” studentoidut residuaalit, jotka saavat itseisarvoltaan suuremman arvon kuin 2. Tällaisia havaintoja on vain kaksi (ks. liitte 3).

5.2 Pääkomponenttialyysi yksittäisille käytännöille

Seuraavaksi tarkastellaan, onko yrityksistä löydettävissä käytäntönippuja ja jos on, niin millainen niiden vaikutus on eri suorituskykymuuttujiin. Tutkimuksen kohteeksi valituille 12 käytännölle tehtiin pääkomponenttialyysi, jonka avulla pyrittiin löytämään yksittäisiä käytäntökimppuja ja näin vähentämään jatkoanalyysissä käytettyjen muuttujien määrää. Pääkomponenttialyysin tulokinnassa teoreettisen ymmärryksen lisäksi tukeudutaan usein myös erilaisiin nyrkkisääntöihin. Yksi tällainen on, että merkitsevien latausten arvojen tulisi olla suurempia kuin 0,5, tosin joidenkin mielestä riittää 0,3 (Hair ym. 1998). Lisäksi pääkomponenttien ominaisarvojen tulisi olla arvoltaan ≥ 1 .

Aineiston analyysissä käytettiin SPSS:n pääkomponenttimenetelmää yhdistettynä Varimax-rotatioon. Menetelmä käyttää pääkomponenttipistemäärien estimoinnissa korrelaatiomatriisia. Saadut pääkomponenttipistemäärät tallennettiin omiksi muuttujikseen (yhteensä neljä), joita käytettiin jatkossa regressioanalyysissä selittävinä muuttujina.

Pääkomponenttialyysin tuloksena saatiin käytännöt jaettua neljälle eri komponentille. Komponenttien reliabiliteettimittauksen perusteella todettiin, että niihin kuuluvat käytännöt liittyivät toisiinsa varsin hyvin. Lisäksi muuttujien saamat korkeimmat latausarvot (taulukossa tummennetut) olivat melko suuria, joten tältä osin neljän pääkomponentin reliabiliteetit² voi katsoa riittäviksi. Pääkomponenttialyysin tulokset on kuvattu seuraavan sivun taulukossa (ks. taulukko 5.3.).

Tarkasteltaessa 0,5 ylittäviä latausarvoja havaittiin toisiinsa läheisesti kyt-

²Mittauksessa on käytetty Cronbachin alfan arvoa, jolla mitataan mittarin yhtenäisyyttä. Alfa-estimoidut arvot ovat yli 0,6:n, joten niitä voi pitää riittävinä. Lisäksi testattiin aineiston sopivuutta pääkomponenttialyysiin KMO-testillä ja Bartlettin testillä ja todettiin aineiston olevan näiltä osin sopiva pääkomponenttialyysia varten. Selitysaste (64%) ilmaisee, kuinka monta prosenttia saadut neljä komponenttia selittävät aineiston kokonaisvarianssista. Kommunaliteetti puolestaan ilmaisee, kuinka suuren osan kunkin yksittäisen käytännön varianssista kyseiset neljä komponenttia pystyvät selittämään, ja se lasketaan kunkin komponentin latausarvojen neliöiden summana. Käytetyllä ortogonaalisella kierto menetelmällä (Varimax) kommunaliteetti voi olla korkeintaan 1. Kiertomenetelmiä ei käsitellä tässä esityksessä tarkemmin

keytyviä ja suhteellisen helposti tulkittavissa olevia käytäntöjä. Työn laajentaminen, työn rikastaminen ja työkierto liittyivät aineistossa vahvasti toisiinsa muodostaen yhden selkeän pääkomponentin. Tämä käytäntöjen ulottuvuus heijastelee välittömään työn organisointiin yhteydessä olevien käytäntöjen samantyyppistä vaihtelua vastaajien joukossa. Kyse on vahvasti työntekijöiden monitaitoisen työskentelyn ja työn rajojen purkamisen käytännöistä, joten se nimettiin *työn rajojen purkamisen komponentiksi* (komponentti 1).

Toiselle komponentille sijoittuivat TQM, laatutyökalut ja JIT-periaate kuvaustaen mitä ilmeisimmin laatuun keskittyvän johtamisen ja turhan työn karsimisen periaatteiden soveltamista, jolloin komponentti voitiin nimetä *laatujohtamisen komponentiksi* (komponentti 2). Kolmannessa komponentissa latautuivat osastojen väliset kehittämissryhmät, aloitejärjestelmät, TPM ja jatkuvan parantamisen ohjelmat. Kyseessä näyttäisi olevan pyrkimys monipuoliseen ja laaja-alaiseen jatkuvaan kehittämistoimintaan, johon myös työntekijät pyritään saamaan mukaan. Komponentti nimettiin *osallistavan kehittämisen komponentiksi* (komponentti 3).

Taulukko 5.3. Pääkomponenttianalyysin tulokset (Varimax-rotatio).

KÄYTÄNTÖ	Komponentit				Kommunaliteetti
	1	2	3	4	
Työn laajentaminen	,846	,327	,152	,058	,849
Työn rikastaminen	,815	,359	,091	,119	,815
Työkierto	,636	-,122	,411	,135	,607
TQM	,091	,727	,227	-,069	,594
Laatutyökalut	,288	,690	,087	,180	,599
JIT-periaate	,145	,597	-,075	,307	,477
Osastojen väliset kehittämissryhmät	,281	,141	,709	,058	,605
Työntekijöiden aloitejärjestelmät	,118	-,098	,687	,184	,530
TPM	,038	,379	,602	-,118	,522
Jatkuvan parantamisen ohjelmat	,060	,487	,549	,236	,598
Tiimityö	,028	,187	,178	,877	,836
Hierarkiatasojen vähentäminen	,491	,090	,063	,616	,633
Ominaisarvo	4,213	1,243	1,192	1,016	
Cronbachin alfa	,789	,619	,658	,612	
Selitysaste 63,9 %					
KMO ,812					
Bartlettin testi ,000					

Neljännelle komponentille sijoittuvat tiimityö ja hierarkiatasojen vähentäminen. Näistä tiimityö alkaa jo olla varsin paljon ja moninaisin tavoin sovellettu käytäntö, joten käytäntökimpun olennaisen elementin löytäminen ei ollut helppoa. Komponentin varsinainen ydin hahmottui vasta tarkempien analyysien perusteella. Niiden perusteella olennaista komponentissa näyttäisi olevan ennen muuta *työntekijöiden päätösvallan lisääminen*, joka annettiin komponentin nimeksi (komponentti 4).

5.3 Käytäntökomponentit ja yrityksen suorituskyky

Regressioanalyysin tulokset käytäntökomponenteille on kuvattu taulukossa 5.4. Myöskään näissä tuloksissa ei ole kontrollimuuttujia mukana (ks. tarkemmin liite 4).

Taulukko 5.4. Regressioanalyysi: käytäntönippujen vaikutus suorituskykyymuuttujiin ”kokonaissuorituskyky”, ”työhyvinvointi” ja ”joustavuus”. Merkitsevyysluokitus: * merkitsevä 0,1: n, ** merkitsevä 0,05: n ja *** merkitsevä 0,01: n merkitsevyystasolla. N=111.

Selitettävä suorituskykyymuuttuja	Kokonaissuorituskyky		Työhyvinvointi		Joustavuus	
	Regressio-kerroin	p-arvo	Regressio-kerroin	p-arvo	Regressio-kerroin	p-arvo
KÄYTÄNTÖ-KOMPONENTTI						
1. Työn rajojen purkaminen	,132	,010***	,181	,019**	,150	,031**
2. Laatujohtaminen	,052	,333	,127	,120	,077	,292
3. Osallistava kehittäminen	,097	,080*	-,062	,454	,185	,015**
4. Työntekijöiden päätösvallan lisääminen	,026	,651	,170	,052*	,082	,294
<i>Vakiotermit</i>	4,314	,000***	3,925	,000***	4,385	,000***
<i>Yrityksiä /lkm</i>	111		111		111	
<i>Selitysaste</i>	0,218		0,283		0,226	
<i>F-testi</i>	1,116		1,582*		1,170	

Analyysin keskeisin tulos on 1. työn rajojen purkamisen käytäntönippun merkitsevyys kaikkien kolmen tarkasteltavan suorituskykyymuuttujan kohdalla. Kontrollimuuttujista yrityksen koko vaikuttaa negatiivisesti merkitsevästi kaikkiin kolmeen suorituskykyymuuttujaan. Lisäksi korkeasti koulutettujen työntekijöiden prosenttiosuus vaikuttaa positiivisesti kokonaissuorituskykyyn (ks. liite 4).

Työn rajojen purkamisen komponenttiin kuuluvat käytännöt näyttäisivät tukevan toisiaan myönteisellä tavalla. Tulos on hyvin linjassa sen ajatuksen kanssa, että työkierto, työn laajentaminen ja työn rikastaminen tulevat näkyviin tehokkaina ja joustavuuteen tähtäävinä käytäntöinä. Työn rajojen purkaminen helpottaa arkista työn organisointia ja luo mahdollisuuksia tehtäväjoustoihin, yhtenä esimerkkinä sairaspöissaolojen paikkaamismahdollisuuksien paraneminen. Sen sijaan kyseisen käytäntönippun suora positiivinen vaikutus työhyvinvointiin ei ole itsestään selvää, eikä sille ole helposti löydettävissä olevaa selitystä.

Muista käytäntönipuista osallistava kehittäminen (esim. työntekijöiden aloitejärjestelmät) vaikuttaa positiivisesti joustavuuteen ja lievästi positiivises-

ti kokonaissuorituskykyyn. Vaikutus kokonaissuorituskykyyn voidaan tulkita epäsuoraksi eli se tulee joustavuuden kautta, koska joustavuuden osamittarit ovat mukana kokonaissuorituskykyä kuvaavassa muuttujassa. Lisäksi työhyvinvointiin vaikuttaa positiivisesti työntekijöiden päätösvallan lisäämisen käytännön nippu eli tiimityö ja hierarkiatasojen vähentäminen yrityksissä näyttäisivät lisäävän työhyvinvointia.

Käytäntökomponenttien vaikutus yritysten tuoteinnovaatiokykyyn on kuvattu taulukossa 5.5. Käytäntönipuista tilastollisesti merkitseviksi tulivat sekä työn rajojen purkaminen että työntekijöiden päätösvallan lisääminen. Neljänteen nippuun kuuluvista kahdesta käytännöstä (tiimityö ja hierarkiatasojen vähentäminen) kumpikaan ei ollut yksittäisenä käytäntönä merkitsevä. Kontrollimuuttujista positiivisesti merkitseviä olivat yhden prosentin merkitsevyystasolla liiketoiminnan strategia ”uusien tuotteiden kehittäminen”, viiden prosentin merkitsevyystasolla korkeasti koulutettujen prosenttiosuus yrityksen koko henkilöstöstä sekä 10 prosentin tasolla toimiala kulkuneuvoteollisuus (ks. liite 4).

Taulukko 5.5. Regressioanalyysi: käytäntönippujen vaikutus suorituskykyymuuttujaan ”tuoteinnovaatiot”. Merkitsevyysluokitus: * merkitsevä 0,1: n, ** merkitsevä 0,05: n ja *** merkitsevä 0,01: n merkitsevyystasolla. N=110.

Selitettävä suorituskykyymuuttuja	Tuoteinnovaatiot		
	Regressio-kerroin	Wald	p-arvo
<i>KÄYTÄNTÖKOMPONENTTI</i>			
1. Työn rajojen purkaminen	,954	10,445	,001***
2. Laatujohtaminen	,258	,885	,347
3. Osallistava kehittäminen	,314	1,170	,279
4. Työntekijöiden päätösvallan lisääminen	,734	5,797	,016**
<i>Vakiotermi</i>	-787	,144	,704
<i>Yrityksiä/ lkm</i>	110		
<i>-2 log likelihood</i>	115,179		
<i>Selitysaste (Nagelkerke)</i>	0,377		
<i>Oikein luokiteltuja havaintoja (%)</i>	71,8		

Kontrollimuuttujissa ehkä yllättävin tulos oli, että vaikka korkeasti koulutetun henkilöstön prosenttiosuus koko henkilöstöstä selittää merkitsevästi tuoteinnovaatioita, niin tutkimus- ja tuotekehityshenkilöstön prosenttiosuus kuitenkaan ei. Osittain tämä johtuu muuttujan jakaumasta, koska tutkimus- ja tuotekehityshenkilöstön lukumäärät ovat prosentuaalisesti melko pieniä (mediानी kolme prosenttia) ja yli 10 prosentin arvot ovat harvinaisia eli jakauma painottuu pieniin prosenttilukuihin. Toisaalta voidaan myös tulkita, että muidenkin kuin pelkästään tutkimus- ja tuotekehitystehtävissä toimivien henkilöi-

den korkea koulutustaso vaikuttaa tuoteinnovaatioiden syntyyn.

Käytäntökomponenteilla tehdyssä analyysissä VIF-arvot ovat pienempiä kuin yksittäisille käytännöille tehdyssä analyysissä. Näin tuleekin olla, koska käytäntökomponentit määriteltiin keskenään korreloimattomiksi (määritelmä 3.3.1.). Varsinaisina tilastollisina malleina myös käytäntökomponenteilla saadut tulokset ovat kuitenkin riittämättömiä samoilla perusteilla kuin kapaleen 5.1. loppupuolella todettiin. Regressioanalyysit voitaisiin tehdä uudelleen siten, että mukana olisivat merkitsevät käytäntökomponentit³ ja kontrollimuuttujat, mutta selitysasteet jäisivät tällöin entistäkin vaatimattomammiksi ja muuttujien väliset mahdolliset epäsuorat vaikutussuhteet jäisivät selvittämättä.

5.4 Lopullinen malli

Tilastollisen mallin rakentamista varten LISREL-ohjelman ajojonoon otettiin mukaan käytäntökomponentit, kolme suorituskykymuuttujaa ja tilastollisesti merkitseviksi osoittautuneita kontrollimuuttujia. SPSS:n Select Cases -toiminnon avulla jatkoanalyysiin saatiin ne yritykset, jotka olivat mukana lineaarisissa regressioanalyseeseissa (111 kpl). Tämän jälkeen laskettiin valittujen muuttujien väliset Pearsonin korrelaatiokertoimet ja vietiin korrelaatiomatriisi LISREL-ohjelman ajojonoon.⁴ Yksittäisten käytäntöjen vaikutuksia yritysten suorituskykyyn ei jatkoanalyysissä tutkittu. Seuraavaksi esitetään tältä pohjalta rakennettu regressiopolkumalli sekä arvioidaan mallin käyttökelpoisuutta.

5.4.1 Rekursiivinen regressiopolkumalli

Mallin rakentamisessa selvitettiin käytäntökomponenttien, merkitsevien kontrollimuuttujien ja kolmen suorituskykymuuttujan välisiä vaikutussuhteita. Tarkoituksena oli saada aikaan malli, jossa suorat vaikutukset olisivat tilastollisesti merkitseviä ja lisäksi tutkia mahdollisia muuttujien välisiä epäsuoria vaikutussuhteita. Näin päädyttiin malliin, joka on esitetty kuviossa 5.1. Kyseessä on rekursiivinen regressiopolkumalli, jossa on kaksi selittävää ja kolme selitettävää muuttujaa. Malli koostuu seuraavista kolmesta regressioyhtälöstä:

$$\begin{cases} \text{Joustavuus} = 0,21 \cdot \text{Käytäntökomponentti 1} - 0,21 \cdot \text{Koko} \\ \text{Kokonaissuorituskyky} = 0,82 \cdot \text{Joustavuus} \\ \text{Työhyvinvointi} = 0,28 \cdot \text{Kokonaissuorituskyky}, \end{cases}$$

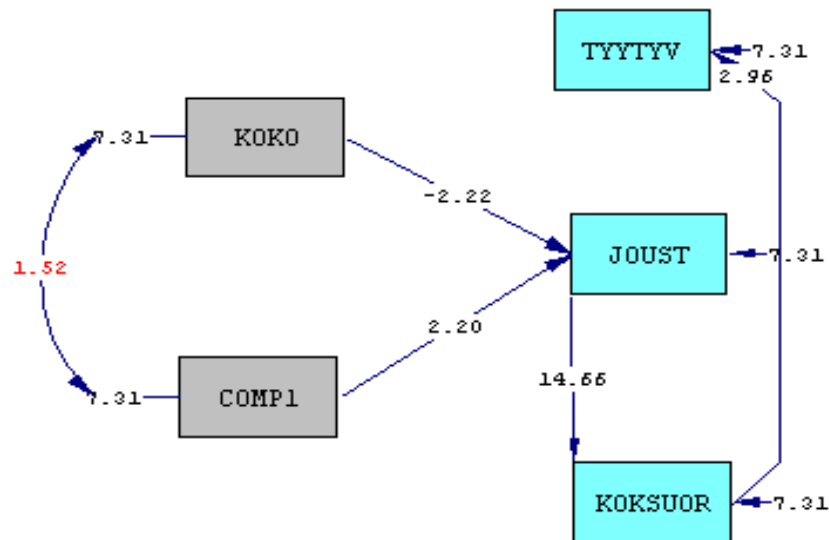
missä regressiokertoimien arvot on saatu mallin estimointituloksista (ks. liitteen 5 beta- ja gammadparametrien arvot).

Yrityksen suuri koko vaikuttaa negatiivisesti ja työn rajojen purkamisen käytäntökomponentti positiivisesti yrityksen toiminnan joustavuuteen. Joustavuus puolestaan vaikuttaa positiivisesti yrityksen kokonaissuorituskykyyn.

³Analyyseissa on käytetty SPSS:n Enter-metodia, joka pitää mukaan tuodut muuttujat mallissa huolimatta niiden merkitsevyysarvoista.

⁴LISREL-ohjelman ajojonossa olevat muuttujat ovat kaikki jatkuvia.

Tämä tulos on sinänsä ilmeinen, koska kolme joustavuutta kuvaavaa osamuuttujaa ovat mukana kokonaissuorituskykymittarissa (ks. taulukko 2.2.). Mielenkiintoinen tulos on sen sijaan se, että joustavuus ei näyttäisi vaikuttavan suoraan positiivisesti työhyvinvointiin, vaan vaikutus tulee kokonaissuorituskyvyn kautta. Vaikka ensimmäinen käytäntökomponentti osoittautui merkitseväksi



Chi-Square=5.55, df=5, P-value=0.35229, RMSEA=0.032

Kuvio 5.1. Rekursiivinen regressiopolkumalli.

lineaarisisissa regressioanalyseissa kaikkien suorituskykymuuttujien kohdalla, niin tässä mallissa nuolien lisääminen käytäntökomponentista 1 työhyvinvointiin ja/tai kokonaissuorituskykyyn aiheuttaa sen, että vaikutuksista yksikään ei tule tilastollisesti merkitseväksi. Siten tässä kuvattu malli on erilaisten kokeilujen pohjalta ainoa, jossa suorat vaikutukset saadaan merkitseviksi. Myös muiden käytäntökomponenttien ja/tai kontrollimuuttujien lisääminen malliin aiheuttaa sen, että kaikki suorat vaikutukset eivät enää ole merkitseviä.

5.4.2 Mallin arviointi

Kuviosta 5.1. nähdään, että kaikki t -arvot ovat itseisarvoltaan suurempia kuin 2, joten suorat vaikutukset ovat kaikki tilastollisesti merkitseviä 5 prosentin merkitsevyystasolla. Epäsuorista vaikutuksista yrityksen suuri koko vaikuttaa negatiivisesti merkitsevästi ja ensimmäinen käytäntökomponentti positiivisesti merkitsevästi kokonaissuorituskykyyn siten että vaikutus tulee joustavuuden kautta (t -arvot $-2, 19$ ja $2, 18$). Sen sijaan näiden muuttujien epäsuora vaikutus

ei ulotu merkitsevästi työhyvinvointiin asti (vastaavat t -arvot $-1,76$ ja $1,75$). Lisäksi selitettävien muuttujien välillä joustavuuden epäsuora vaikutus kokonaissuorituskyvyn kautta työhyvinvointiin on tilastollisesti merkitsevä (t -arvo $2,90$).⁵

Kuviosta 5.1. saadaan myös mallin sopivuutta kuvaavat χ^2 -testisuureen sekä RMSEA:n arvot. Khi(2)-testin tulos on ei-merkitsevä (p -arvo = $0,352$, $df = 5$) ja RMSEA = $0,032$. Mallia kokonaisuudessaan kuvaavat diagnostiikkatulokset on koottu taulukkoon 5.6.

Taulukko 5.6. Mallin sopivuutta kokonaisuudessaan kuvaavat keskeisimmät diagnostiikkatulokset.

χ^2 ($df = 5$)	p -arvo: $0,35$
RMSEA	$0,032$
NFI	$0,95$
GFI	$0,98$
CFI	$0,99$
SRMR	$0,049$

Kappaleessa 3.4.3. käsiteltiin taulukkoon 5.6. koottujen mallin hyvyttä kuvaavien diagnostiikkatulosten kriittisiä arvoja. Näiltä osin mallia voi kokonaisuutena ottaen pitää riittävänä. SRMR-arvo on tosin lähellä kriittistä arvoa, mutta yleisesti käytetyn kriteerin ($SRMR \leq 0,05$) mukaan tämäkin arvo on riittävä (ks. liitteen 5 ”Goodness of Fit Statistics”). Suurin modifikaatioindeksin arvo on $2,88$, joka on pienempi kuin $\chi^2_{0,95;1} = 3,84$, joten malli on myös näiltä osin sopiva eli sitä ei tarvitse modifioida. Standardoitujen residuaalien minimiarvo on $-1,45$ ja maksimiarvo $1,15$ eli myös yksittäisten havaintojen perusteella mallin voi katsoa olevan riittävä. Selitysasteet kahdelle Y -muuttujalle ”joustavuus” ja ”työhyvinvointi” ovat edelleen vaatimattomia ($0,07$ ja $0,08$). Tämä on odotettua, koska selittäviä muuttujia on mallissa vähän. Lisäksi muuttujan ”kokonaissuorituskyky” korkea selitysaste ($0,67$) johtuu pitkälti jo aiemmin todetusta seikasta, että joustavuutta kuvaavat osamittarit ovat mukana kokonaissuorituskykyä kuvaavassa muuttujassa.

Mallin riitävyydestä tarkastelujen perusteella voidaan näin ollen todeta, että edellä kuvattu rekursiivinen regressiopolkumalli on tilastollisena mallina riittävä kokonaisuudessaan (diagnostiikkatulokset) samoin kuin parametrisoitu (regressiokertoimien merkitsevyys) ja yksittäisten havaintojen tasolla (standardoidut residuaalit). Kuitenkin joustavuutta ja työhyvinvointia kuvaavien suorituskykyymuuttujien selitysasteet jäävät mallissa varsin vaatimattomiksi.

⁵ks. liite 5: ”Indirect Effects of X on Y” ja ”Indirect Effects of Y on Y”.

6 Johtopäätökset ja tutkimuksen rajoitukset

Tutkimuksen analyysit perustuvat vuonna 2007 suomalaisille teollisuusyrityksille tehtyyn kyselyyn, jossa selvitettiin, millaisia innovatiivisia käytäntöjä yrityksillä oli käytössään, käytäntöjen käyttöastetta ja niiden vaikutusta yritysten suorituskykyyn. Yrityksen suorituskyvyn mittaamiseksi muodostettiin neljä suorituskykyymmuttujaa. Käytännöissä tarkastelu rajoitettiin koskemaan työn valmistus- ja organisointiprosesseja.

Tarkastelun kohteeksi rajautui 12 tuotantokäytäntöä. Yleisimmiksi käytännöiksi osoittautuivat aloitejärjestelmät, jatkuvan parantamisen ohjelmat ja tiimityö. Käytäntöjen yleisyys vaihteli jonkin verran toimialan ja yrityskoon mukaan. Käytäntöjen ja suorituskykyymmuttujien välisten yhteyksien selvittämiseksi tarkasteltiin ensin yksittäisten käytäntöjen vaikutuksia yritysten suorituskykyyn lineaarisella ja logistisella regressioanalyysillä. Tämän jälkeen muodostettiin pääkomponenttianalyysillä neljä käytäntönippua ja selvitettiin niiden vaikutuksia suorituskykyyn ja myös sitä, missä suhteessa näiden nippujen vaikutukset ovat verrattuna yksittäisten käytäntöjen vaikutuksiin. Myös näissä analyyseissä käytettiin tilastollisina menetelminä lineaarista ja logistista regressiota. Lopuksi merkitsevien käytäntökomponenttien, kontrollimuuttujien ja kolmen suorituskykyymmuttujan välisiä vaikutussuhteita selvitettiin regressiopolkumalleilla, joiden pohjalta rakennettiin lopullinen malli.

Yksittäisiä käytäntöjä tarkasteltaessa vain muutamalla käytännöllä oli positiivinen vaikutus yrityksen suorituskykyyn. Jatkuvan parantamisen ohjelmat paransivat sekä kokonaissuorituskykyä että yrityksen joustavuutta. Työkierro taas osoittautui työhyvinvointia parantavaksi samoin kuin laatutyökalujen käyttö. Minkään käytännön kohdalla vaikutus ei osoittautunut merkitsevästi negatiiviseksi. Hieman yllättäen, mutta toisaalta eräitä tutkimustuloksia vahvasti tukien, parhaat suorituskykyvaikutukset näyttäisivät tulevan varsinkin perinteisistä työn muotoilun käytännöistä (ks. taulukko 5.2.).

Tutkimuksen perusteella oli löydettävissä toisiinsa yhteydessä olevia käytäntöjä. Tutkimukseen valituista 12 käytännöstä saatiin neljä käytäntönippua tai -komponenttia: työn rajojen purkaminen, laatujohtaminen, osallistava kehittäminen ja työntekijöiden päätösvallan lisääminen. Näistä käytäntökomponenteista ensimmäisellä oli tilastollisesti merkitsevä positiivinen vaikutus sekä yrityksen parantuneeseen teknis-taloudelliseen kokonaissuorituskykyyn, työhyvinvointiin että joustavuuteen. Työn rajojen purkaminen näyttäisi niin ikään

lisäävän yrityksen kykyä tuottaa markkinoille uusia tuotteita. Uusien tuotteiden syntyyn vaikutti myönteisesti myös neljäs komponentti, joka sisälsi hierarkioiden purkamisen ja tiimityön käytännöt. Lisäksi osallistavan kehittämisen komponentilla oli positiivinen vaikutus joustavuuteen ja kokonaissuorituskykyyn samoin kuin työntekijöiden päätösvallan lisäämisellä työhyvinvointiin. Negatiivisesti merkitseviä regressiokertoimia ei saatu yhdellekään komponentille (vaikutukset olivat heikoimmillaan lähellä nollaa, ks taulukko 5.4).

Analyyseissa mukana olleista kontrollimuuttujista huomion arvoinen tulos oli yrityksen suuren koon tilastollisesti merkitsevä negatiivinen vaikutus joustavuuteen ja joustavuuden kautta yrityksen kokonaissuorituskykyyn käytäntökomponenteilla saaduissa regressiotuloksissa. Myös työhyvinvointiin vaikutus oli negatiivinen, tosin lopullisessa rakenneyhtälömallissa kyseinen vaikutussuhde oli epäsuora eikä aivan negatiivisesti merkitsevä. Lisäksi kontrollimuuttujista tilastollisesti merkitsevä positiivinen vaikutus yrityksen teknis-taloudelliseen kokonaissuorituskykyyn oli korkeasti koulutetun henkilöstön prosenttiosuudella yrityksen koko henkilöstöstä (ks. liite 4).

Yrityksen tuoteinnovaatiokykyä selittävien tulosten pohjalta voidaan pohdita, liittyykö tiimityön ja hierarkiatasojen vähentämisen yhdistävä komponentti kykyyn tuottaa markkinoille uusia tuotteita työntekijöiden mahdollisesti lisääntyneeseen päätösvaltaan ja kenties heille avautuviin suoriin asiakaskontakteihin eli työntekijä-asiakas-suhteen tiivistymiseen. Käytännössä hierarkioiden purku ja tiimityö yhdessä sovellettuina usein tuovat mukanaan omaa ja muita organisaatioita välittävien toimihenkilöroolien häviämisen kuten erillisten esimiesten ja suunnittelijoiden poistamisen organisaatiosta ja lisäävät samalla työntekijöiden yhteistyötä ja kommunikaatiota sekä mahdollisuuksia hahmottaa yritystoiminnan kokonaisuutta (ks. Järvensivu & Koski 2009). Työn rajojen purkamisen komponentin ja uusien tuotteiden synnyn välinen yhteys saattaisi olla selitettävissä samalla tavoin. Kun työntekijöiden työn rajoja puretaan monin tavoin ja eri suunnista, kokonaisuus hahmottuu entistä paremmin ja kommunikointitilaisuudet lisääntyvät.

On siis mahdollista, että joissakin yrityksissä on kyetty pääsemään Kenneyn ja Floridan (1993) kuvailemaan uuteen tuotantokonseptiin, jonka ytimessä on henkilöstön henkisten ja fyysisten taitojen entistä laajamittaisempi ja kattavampi hyödyntäminen. Tällainen uudentyyppinen tuotantomalli nivoo näiden tutkijoiden mukaan yhteen tuotannon ja innovoinnin koko ketjun tutkimuksesta sekä tuotekehityksestä tuotannolliseen insinööriosuamiseen ja tuotantotason työntekijöihin. Tiimimäinen organisaatio ja muut organisatoriset tekniikat palvelevat henkilöstön osaamisen kehitystä ja hyväksikäyttöä. Näistä tuloksista voisi tehdä varovaisen johtopäätöksen, jonka mukaan työntekijöiden yhteisöllistä toimintaa tukevat laaja-alaiset tehtäväkuvat, suurempien kokonaisuuksien ja yrityksen toimintaidean parempi ymmärrys ja hallinta sekä mahdollisuudet asiakaskontakteihin olisivat tuotekehityksen kannalta hyviä käytäntöjä.

Regressioanalyysien tulosten perusteella tehtiin lopuksi rakenneyhtälömalli, johon saatiin tilastollisesti merkitseviksi selittäviksi muuttujiksi yrityksen koko ja ensimmäinen, työn rajojen purkamisen käytäntökomponentti. Lisäksi

mallissa oli mukana kolme selitettävää suorituskykymuuttujaa (kokonaissuorituskyky, työhyvinvointi ja joustavuus). Muuttujien välille saatiin rakennettua rekursiivinen polkumalli, jossa yrityksen suurella koolla oli negatiivinen ja ensimmäisellä käytäntökomponentilla positiivinen vaikutus joustavuuteen. Joustavuus puolestaan vaikutti positiivisesti kokonaissuorituskykyyn. Epäsuorista vaikutuksista ensimmäisen käytäntökomponentin ja yrityksen koon vaikutukset ulottuivat tilastollisesti merkitsevinä myös kokonaissuorituskykyyn, mutta eivät työhyvinvointiin (tarkastelussa käytettiin viiden prosentin merkitsevyystasoa analyysitulosten t -arvojen perusteella). Lisäksi selitettävien muuttujien (Y) osalta joustavuus vaikutti positiivisesti yrityksen kokonaissuorituskykyyn kautta työhyvinvointiin.

Johtopäätösten tekemisessä on kuitenkin oltava varovainen, koska regressioanalyysissä oli mukana suhteellisen vähän yrityksiä ($N=110-111$) ja kysely koski vain valikoituja toimialoja. Analyysissä oli mukana useita tausta- tai kontrollimuuttujia, joiden mukaan ottaminen aiheutti sen, että puuttuvien havaintojen vuoksi yrityksiä karsiutui pois varsin paljon. Koska mukana oli useita binaarisia muuttujia, ei puuttuvien havaintojen korvaamista eli imputointia tehty. Kuitenkin on todettava, että keskeisimmät merkitsevät tekijät (erityisesti ensimmäisen käytäntökomponentin tilastollinen merkitsevyys) säilyivät samoina, vaikka kontrollimuuttujia jätettiin pois tai yhdisteltiin niitä eri tavoilla eli tälle aineistolle saatuja analyysituloksia voi pitää suhteellisen luotettavina. Vielä on muistettava, että kysely on toteutettu keväällä 2007, jolloin talouden taantuma ei ollut alkanut ja perinteiset teollisuustoimialat menestyivät varsin hyvin. Ongelmana on se, että vastaavanlaista kyselyä ei voi toistaa täsmälleen samanlaisena vuosien kuluttua, koska yrityskentässä tapahtuu muutoksia lyhyelläkin aikavälillä.

Myöskään mallin hyvyysarviointeihin ei (varsinkaan regressioanalyysien yhteydessä) kiinnitetty kovin paljon huomiota. Regressioanalyysissä ei ollut tavoitteena optimaalisen tilastollisen mallin sovittaminen aineistoon, vaan löytää tilastollisesti merkitseviä, hyviä työkäytäntöjä. Mallin arviointiin kiinnitettiin enemmän huomiota lopullisen mallin arvioinnin yhteydessä (kappale 5.4.2). Lisäksi on huomattava, että useissa tämän tyyppisissä tutkimuksissa saadut selitystasteet ovat joka tapauksessa vaatimattomia, vaikka löydettäisiin useita tilastollisesti merkitseviä selittäviä tekijöitä. Yrityksen menestymiseen vaikuttavat niin monet seikat, ettei niitä helposti voi selittää kattavasti yhdellä kyselytutkimuksella. Selitystastetta voitaisiin ehkä jonkin verran nostaa selittävien muuttujien määrää lisäämällä, mutta tämä ei ole järkevää, koska tilastollisen tutkimuksen tarkoituksena on tutkimusdatan olennaisimpien piirteiden etsiminen ja tiivistäminen.

Analyysien keskeisin tulos on varovainen johtopäätös, että tietyllä tavalla yhdisteltyinä käytännöillä on tilastollisesti merkitsevä vaikutus yritysten suorituskykyyn, mutta erikseen käytettyinä ei juurikaan. Tämä tutkimus antaisi siten tapauskohtaista tukea Ichniowskin (1990) vastaavanlaisille tuloksille.

Lähdeluettelo

- Agresti, A. (1990), *Categorical Data Analysis*, New York: Wiley.
- Alasoini, T. & Heikkilä, A. & Ramstad, E. & Ylöstalo, P. (2008), High-Involvement Innovation Practices at Finnish Workplaces, *International Journal of productivity & Performance Management*, Vol. 57 Issue 6, pp. 449–459.
- Appelbaum, E. & Bailey, T. & Berg, B. & Kalleberg, A.K. (2000), *Manufacturing Advantage. Why High-Performance Work Systems Pay Off?* Ithaca and London: Cornell University Press.
- Blasi, J. & Kruse D.L. (2006), U.S. High-Performance Work Practices at Century's End, *Industrial Relations*, 45(4): 547–578.
- Godard, J. (2001), High Performance and the Transformation of Work. The Implications of Alternative Work Practices for the Experience and Outcome of Work. *Industrial and Labor Relations Review* 54(4): 776–805.
- Hair, J.F. & Andersson, R.E. & Tatham, R.L. & Black, W.C.(1998), *Multivariate Data Analysis*, 5th edition, New Jersey: Prentice Hall.
- Ichniowski, C. (1990), *Human Resource Management Systems and the Performance of US manufacturing Business*, Working Paper 3349, Cambridge MA: National Bureau of Economic Research.
- Ichniowski, C. & Kochan, T. & Levine, D. & Olson, O. & Strauss, G. (1996), What Works at Work?, *Industrial Relations* 35(3): 291–313.
- Isotalo, J. (2009), Yleistetyt lineaariset mallit I, Tampereen yliopisto, Matematiikan ja tilastotieteen laitos, kurssimateriaali.
- Isotalo, J. (2010), Yleistetyt lineaariset mallit II, Tampereen yliopisto, Matematiikan ja tilastotieteen laitos, kurssimateriaali.
- Järvensivu, A. & Koski, P. (2009), Hyvä, parempi, innovaatio? Tutkimus organisatorisista innovaatioista, työelämän laadusta ja työn mielekkyydestä, Työelämän tutkimuskeskuksen työraportteja 84/2009, Tampere: Tampereen yliopisto.
- Johnson, R.A. & Wichern D.W. (2002), *Applied Multivariate Statistical Analysis*, 5th edition, New Jersey: Prentice Hall.
- Kalmi, P. & Kauhanen, A. (2008), Workplace Innovations and Employee Outcomes: Evidence from Finland. *Industrial Relations*, Vol. 47, No. 3: 430–459.
- Kauhanen, A. (2007), The Incidence of the High-Performance Work Systems: Evidence from a Nationally Representative Employee Survey, pp. 35–66, teoksessa Kauhanen, A. (ed.), *Essays on Empirical Personnel Economics*, Acta Universitatis Oeconomicae Helsingiensis A-305, Helsinki School of Economics: HSE Print.
- Kenney, M. & Florida, R. (1993), *Beyond Mass Production: The Japanese System and its Transfer to U.S.*, New York: Oxford University Press.

- Laugen, B.T. & Acur, N. Boer, H. & Frick, J. (2005), Best Manufacturing practices. What do the best-performing companies do? *International Journal of Operations & Production Management*, Vol. 25, No. 2/2005, pp. 131–150.
- Leskinen, E. (2009), LISREL–mallit (rakenneyhtälömallit), Yliopistoallianssi: Jyväskylän yliopisto, Tampereen teknillinen yliopisto ja Tampereen yliopisto, kurssimateriaali.
- Luoma, A. (2008), Monimuuttujamenetelmät, Tampereen yliopisto, Matematiikan ja tilastotieteen laitos, kurssimateriaali.
- Myers, R.H., & Montgomery, D.C., & Vining, G.G. (2002), *Generalized Linear Models With Applications in Engineering and the Sciences*, New York: Wiley.
- Neumark, D. & Cappelli, P. (1999), *Do High Performance Work Practices Improve Establishment-Level Outcomes?*, Cambridge MA: National Bureau of Economic Research.
- Puntanen, S. (2007), *Regressioanalyysi*, Tampereen yliopisto, Matematiikan ja tilastotieteen laitos, kurssimateriaali.
- Sharma, S. (1996), *Applied Multivariate Techniques*, New York: Wiley.
- Vehkalahti, K. (2009), *Data-analyysi II*, Helsingin yliopisto, Matematiikan ja tilastotieteen laitos. Saatavilla osoitteessa: <http://www.helsinki.fi/~kvehkala/da2/moniste.pdf>].
- Weisberg, S. (1985), *Applied Linear Regression*, second edition, New York: Wiley.

LIITE 1: Tutkimuksessa käytetyt kyselylomakkeen osiot.

4. Innovatiiviset toimintatavat

A) Onko toimipaikassanne käytössä seuraavia *tiedon johtamisen käytäntöjä*?

	Ei ole käytössä	Käytössä epävirallisella tavalla	Käytössä virallisella tavalla
1. Johdon tehtäväkierto yrityksen sisällä tai sisäryitysten kesken	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Systemaattiset teknologian ennakoitinkäytännöt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Tiekartat (roadmaps) päämäärien välietappien tunnistamiseksi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Systemaattinen benchmarking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Hyvin käytäntöjen levittäminen yritysten osastojen välillä	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Säännöllinen projektien jälkikäteisarviointi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Mentorointikäytännöt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Rinnakkaiset uusien tuotteiden kehittämistiimit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Osaamiskartoitukset ja organisaation osaamiskartan luominen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Henkilökunnan määräaikainen ”vaihtaminen” toimittajien ja asiakkaiden kanssa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Sisäisten tukiryhmien perustaminen (organisaation kehittäminen, talon sisäinen konsultointi jne.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Työntekijöiden osallistuminen yrityksen vision luomiseen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Intranet yrityksen sisäisille asiantuntijaryhmille	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Työntekijöiden aloitejärjestelmät	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Liiketoimintastrategia kirjoitetussa muodossa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Teknologiastrategia kirjoitetussa muodossa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. Yrityksen sisäinen tiedotuslehti	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Systemaattinen tiedon johtamisen strategia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Jokin muu, mikä _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

B) Onko toimipaikassanne käytössä seuraavia *valmistusprosessiin liittyviä käytäntöjä*?

	Ei ole käytössä	Käytössä epävirallisella tavalla	Käytössä virallisella tavalla
1. Kokonaisvaltainen tuottava kunnossapito (TPM)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Solutuotanto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Just-in time (JIT), juuri oikeaan tarpeeseen periaate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Laatujohtaminen (TQM)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Massaräättelöinti	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Jatkuva parantaminen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Osastojen väliset kehittämisryhmät	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. ERP (Sap, Oracle, IFS)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Laatujohtamiset (esim. ISO, EFQM)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Tietokoneintegroitu teknologia (CAD, CAM, CIM, FMS)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. EDI = toimittajien ja asiakkaiden tietojärjestelmien integrointi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Kevyttuotanto (lean-toimintatapa)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Liiketoimintaprosessien uudelleensuunnittelu (BPR)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Jokin muu, mikä _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

C) Toteutetaanko toimipaikassanne joitakin seuraavista *asiakkaisiin* liittyvistä käytännöistä?

	Ei ole käytössä	Käytössä epävirallisella tavalla	Käytössä virallisella tavalla
1. Asiakkaille järjestettävät erityiset tapahtumat (asiakasillat tms.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Asiakkaan tarpeiden systemaattinen arviointi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Asiakastytytyväisyyskysely	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

D) Ovatoiko jotkin seuraavista *työn organisoiminnin* tavoista käytössä toimipaikassanne?

	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1. Tuotannon tiimit/ryhmät	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Hierarkiatasojen vähentäminen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Työkierto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Työn laajentaminen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Työn rikastaminen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Laatu työkalujen käyttäminen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Luovuustekniikoiden käyttäminen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Työryhmien ja johdon väliset sopimukset	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8. Liiketoiminnan mittarit. Arvioikaa seuraavilla mittareilla toimipaikassanne tapahtunutta muutosta vuosien 2003–2005 aikana

	merkittävästi parantunut	jonkin verran parantunut	ei muutosta	jonkin verran huonontunut	merkittävästi huonontunut
1. Kannattavuus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Toimitusaika	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Toimitusvarmuus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Tuotteiden laatu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Työtyytyväisyys	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Työssäjaksaminen (poissaolot)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Myynti	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Markkinaosuus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Tuottavuus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Joustavuus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

LIITE 2: Tutkimuksessa käytettyjen muuttujien kuvaus.

Muuttuja	Muuttujan lyhenne (tunnus)	Kuvaus / luokitus
Yksittäiset käytännöt Käytäntökomponentit	ks. regressio-analyysin tulostus (liite 3) FAC1_1- FAC4_1	0=epävirallisesti tai ei käytössä, 1=virallisesti käytössä pääkomponenttipistemäärät
<i>Toimiala</i> Sahatavara ja puutuotteet Massa- ja paperi Kumi ja muovi Metallituotteet Koneenrakennus Kulkuneuvot	wood pulppape rubberpl metal machine vehicles	0=ei kuulu toimialaan 1=kuuluu toimialaan, vertailuryhmänä koneenrakennus
<i>Liiketoiminnan strategiat</i> Myynnin lisääminen markkinointia vahvistamalla Uusien markkina-alueiden valtaaminen Laajentuminen uusille toimialoille Uusien tuotteiden kehittäminen Nykytilanteen säilyttäminen	k3f1 k3f2 k3f3 k3f4 k3f5	0=ei ole käytössä, 1=on käytössä
<i>Muut kontrollimuuttajat</i> Yrityksen pääkonttori Koulutussuunnitelma kirjoitetussa muodossa Henkilöstön kehittämis- suunnitelma kirjoitetussa muodossa Tutkimus- ja tuotekehitys- henkilöstön määrä Korkeasti koulutetun henkilöstön määrä (opisto-, AMK- tai yliopistotutkinto) Henkilöstökoulutukseen osallistuneiden määrä (vuonna 2005) Kilpailevien yritysten lukumäärä Yrityksen koko (vuonna 2005)	headqu_2 trainpla hrmstrat shrd_2 high_edu3 k6b k3d_3 log_size	0=Suomessa, 1=ulkomailla 0=ei, 1=kyllä 0=ei, 1=kyllä prosenttia koko henkilöstöstä (%) prosenttia koko henkilöstöstä (%) prosenttia koko henkilöstöstä (%) 0=enintään 5 kilpailijaa, 1=enemmän kuin 5 kilpailijaa työntekijämäärän logaritmi

LIITE 3: Regressiotulokset yksittäisille käytännöille.

KOKONAISUORITUSKYKY

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,499	,249	-,033	,49381

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6,455	30	,215	,882	,641
	Residual	19,508	80	,244		
	Total	25,963	110			

Residuals Statistics(a)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	3,2390	4,4370	3,7855	,24224	111
Residual	1,18111	,99819	,00000	,42113	111
Std. Predicted Value	-2,256	2,689	,000	1,000	111
Std. Residual	-2,392	2,021	,000	,853	111

Coefficients

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics	
		B	Std. Error			Tolerance	VIF
1	(Constant)	4,073	,432	9,421	,000		
	wood	,187	,206	,907	,367	,472	2,120
	pulppape	-,054	,231	-,235	,815	,397	2,521
	rubberp	-,085	,161	-,527	,600	,502	1,992
	metal	-,003	,165	-,016	,988	,524	1,910
	vehicles	,022	,260	,085	,932	,754	1,327
	k3f1	,132	,129	1,023	,310	,554	1,806
	k3f2	-,048	,126	-,378	,707	,555	1,803
	k3f3	,030	,158	,187	,852	,678	1,475
	k3f4	-,057	,126	-,454	,651	,702	1,425
	k3f5	,076	,197	,386	,701	,586	1,705
	headqu_2	-,220	,162	-1,364	,177	,763	1,311
	trainpla	,004	,148	,024	,981	,402	2,490
	hmrstrat	-,090	,154	-,583	,561	,376	2,663
	shrd_2	-,013	,016	-,838	,404	,573	1,744
	high_edu3	,007	,004	1,875	,064	,548	1,823
	k6b	,001	,002	,454	,651	,623	1,605
	k3d_3	-,106	,108	-,984	,328	,760	1,316
	log_size	-,109	,084	-1,286	,202	,452	2,214
	k4d4b	,022	,211	,104	,917	,231	4,325
	k4d5b	,199	,224	,885	,379	,250	4,002
	k4d3b	,139	,127	1,090	,279	,584	1,713
	k4b4b	-,057	,126	-,452	,652	,592	1,688
	k4d6b	-,056	,152	-,365	,716	,387	2,581
	k4b3b	-,043	,127	-,340	,735	,644	1,553
	k4b7b	-,018	,135	-,135	,893	,629	1,590
	k4a14b	,004	,129	,030	,976	,562	1,778
	k4b1b	,082	,141	,586	,559	,724	1,381
	k4b6b	,243	,138	1,762	,082	,492	2,033
	k4d1b	-,011	,126	-,084	,933	,631	1,584
	k4d2b	-,035	,138	-,253	,801	,510	1,962

Käytännöt: k4d4b=työn laajentaminen; k4d5b=työn rikastaminen; k4d3b=työkierto; k4b4b=TQM; k4d6b=laatuyläkalut; k4b3b=JIT-periaate; k4b7b=osastojen väliset kehittämissuhteet; k4a14b=työntekijöiden aloitejärjestelmät; k4b1b=TPM; k4b6b= jatkuvan parantamisen ohjelmat; k4d1b=tiimityö; k4d2b=hierarkiatasojen vähentäminen.

TYÖHYVINVOINTI

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,599	,359	,119	,71402

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22,881	30	,763	1,496	,080
	Residual	40,786	80	,510		
	Total	63,667	110			

Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1,9604	4,3813	3,1667	,45608	111
Residual	1,59864	1,48154	,00000	,60892	111
Std. Predicted Value	-2,645	2,663	,000	1,000	111
Std. Residual	-2,239	2,075	,000	,853	111

Coefficients

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	4,162	,625	6,658	,000
	wood	,140	,297	,472	,639
	pulppape	-,397	,335	-1,186	,239
	rubberp	-,087	,232	-,373	,710
	metal	,118	,239	,494	,622
	vehicles	,480	,376	1,274	,206
	k3f1	-,121	,186	-,647	,519
	k3f2	-,172	,182	-,947	,347
	k3f3	-,188	,229	-,821	,414
	k3f4	,045	,182	,245	,807
	k3f5	,029	,285	,103	,918
	headqu_2	-,263	,234	-1,126	,264
	trainpla	,080	,214	,373	,710
	hrmstrat	,311	,223	1,396	,167
	shrd_2	-,004	,023	-,194	,847
	high_edu3	,007	,005	1,348	,182
	k6b	,002	,003	,813	,419
	k3d_3	,120	,156	,769	,444
	log_size	-,301	,122	-2,461	,016
	k4d4b	-,049	,306	-,161	,873
	k4d5b	,100	,325	,309	,758
	k4d3b	,446	,184	2,428	,017
	k4b4b	-,135	,182	-,741	,461
	k4d6b	,623	,220	2,825	,006
	k4b3b	,005	,183	,028	,978
	k4b7b	-,298	,195	-1,532	,129
	k4a14b	-,173	,186	-,926	,357
	k4b1b	-,110	,203	-,541	,590
	k4b6b	,057	,199	,284	,777
	k4d1b	-,054	,182	-,295	,769
	k4d2b	,046	,200	,228	,821

Käytännöt: k4d4b=työn laajentaminen; k4d5b=työn rikastaminen; k4d3b=työkierto; k4b4b=TQM; k4d6b=laatuökalut; k4b3b=JIT-periaate; k4b7b=osastojen väliset kehittämissryhmät; k4a14b=työntekijöiden aloitejärjestelmät; k4b1b=TPM; k4b6b= jatkuvan parantamisen ohjelmat; k4d1b=tiimityö; k4d2b=hierarkiatasojen vähentäminen.

JOUSTAVUUS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,519	,270	-,004	,66274

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12,984	30	,433	,985	,501
	Residual	35,138	80	,439		
	Total	48,122	110			

Residuals Statistics(a)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2,6743	4,3907	3,6967	,34357	111
Residual	1,74569	1,35811	,00000	,56518	111
Std. Predicted Value	-2,976	2,020	,000	1,000	111
Std. Residual	-2,634	2,049	,000	,853	111

Coefficients

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	3,917	,580	6,752	,000
	wood	,130	,276	,470	,640
	pulppape	-,396	,311	-1,274	,206
	rubberp	-,154	,216	-,716	,476
	metal	,168	,222	,758	,451
	vehicles	,050	,349	,143	,887
	k3f1	,048	,173	,279	,781
	k3f2	-,001	,169	-,004	,997
	k3f3	,143	,212	,674	,502
	k3f4	-,132	,169	-,781	,437
	k3f5	,283	,265	1,070	,288
	headqu_2	-,069	,217	-,318	,751
	trainpla	,019	,199	,094	,925
	hrmstrat	-,210	,207	-1,018	,312
	shrd_2	-,002	,021	-,097	,923
	high_edu3	,003	,005	,642	,523
	k6b	,001	,002	,551	,583
	k3d_3	-,085	,144	-,591	,556
	log_size	-,104	,113	-,916	,363
	k4d4b	-,028	,284	-,097	,923
	k4d5b	,212	,301	,704	,483
	k4d3b	,189	,171	1,107	,272
	k4b4b	-,005	,169	-,027	,978
	k4d6b	-,242	,205	-1,184	,240
	k4b3b	,026	,170	,152	,880
	k4b7b	,053	,181	,291	,772
	k4a14b	,062	,173	,358	,721
	k4b1b	,033	,189	,175	,862
	k4b6b	,424	,185	2,294	,024
	k4d1b	,043	,169	,252	,801
	k4d2b	-,054	,186	-,293	,770

Käytännöt: k4d4b=työn laajentaminen; k4d5b=työn rikastaminen; k4d3b=työkierto; k4b4b=TQM; k4d6b=laatutyökalut; k4b3b=JIT-periaate; k4b7b=osastojen väliset kehittämissryhmät; k4a14b=työntekijöiden aloitejärjestelmät; k4b1b=TPM; k4b6b= jatkuvan parantamisen ohjelmat; k4d1b=tiimityö; k4d2b=hierarkiatasojen vähentäminen.

TUOTEINNOVAATIOT

Case Processing Summary

Unweighted Cases(a)		N	Percent
Selected Cases	Included in Analysis	110	57,6
	Missing Cases	81	42,4
	Total	191	100,0
Unselected Cases		0	,0
Total		191	100,0

Dependent Variable Encoding

Original Value	Internal Value
no	0
yes	1

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	110,403(a)	,312	,418

Hosmer and Lemeshow Test*

Step	Chi-square	df	Sig.
1	8,105	8	,423

*Hosmer-Lemeshow -testin tulos on ei-merkitsevä eli testin perusteella malli on riittävä.

Classification Table(a)

		Observed	Predicted		Percentage Correct
			Products new to the market?		
			no	yes	
Step 1	Products new to the market?	no	34	16	68,0
		yes	14	46	76,7
Overall Percentage					72,7

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
wood	-,366	1,078	,115	1	,734	,694
pulppape	-1,013	1,193	,721	1	,396	,363
rubberpl	-,296	,834	,126	1	,723	,744
metal	-,804	,860	,874	1	,350	,448
vehicles	1,855	1,406	1,740	1	,187	6,389
k3f1	,343	,627	,300	1	,584	1,409
k3f2	,492	,636	,599	1	,439	1,636
k3f3	,157	,774	,041	1	,839	1,170
k3f4	2,057	,733	7,876	1	,005	7,825
k3f5	,317	1,006	,099	1	,753	1,373
headqu_2	,342	,854	,161	1	,689	1,408
trainpla	-,842	,772	1,188	1	,276	,431
hmrstrat	,936	,810	1,335	1	,248	2,550
shrd_2	-,027	,095	,077	1	,781	,974
high_edu3	,044	,020	4,864	1	,027	1,045
k6b	-,014	,010	2,111	1	,146	,986
k3d_3	-,488	,543	,807	1	,369	,614
log_size	-,195	,435	,200	1	,655	,823
k4d4b	-,126	1,243	,010	1	,920	,882
k4d5b	,446	1,220	,134	1	,714	1,563
k4d3b	1,618	,668	5,862	1	,015	5,044
k4b4b	-1,006	,666	2,285	1	,131	,366
k4d6b	1,340	,824	2,642	1	,104	3,820
k4b3b	-,254	,650	,152	1	,696	,776
k4b7b	-,102	,676	,023	1	,880	,903
k4a14b	-,318	,676	,221	1	,638	,728
k4b1b	-,086	,695	,015	1	,902	,918
k4b6b	,524	,760	,475	1	,491	1,689
k4d1b	,337	,638	,280	1	,597	1,401
k4d2b	,724	,685	1,119	1	,290	2,063
Constant	-2,018	2,346	,740	1	,390	,133

Käytännöt: k4d4b=työn laajentaminen; k4d5b=työn rikastaminen; k4d3b=työkierto; k4b4b=TQM; k4d6b=laatutyökalut; k4b3b=JIT-periaate; k4b7b=osastojen väliset kehittämissryhmät; k4a14b=työntekijöiden aloitejärjestelmät; k4b1b=TPM; k4b6b= jatkuvan parantamisen ohjelmat; k4d1b=tiimityö; k4d2b=hierarkiatasojen vähentäminen.

Block 0: Beginning Block

Iteration History

Iteration	-2 Log likelihood	Coefficients	
		Constant	
Step 1			
0	151,582	,182	
2	151,582	,182	

Casewise List^b

Case	Selected Status ^a	Observed	Predicted	Predicted Group	Temporary Variable	
		products new to the market			Resid	ZResid
21	S	h**	,162	l	,838	2,273
54	S	l**	,847	h	-,847	-2,355

a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

b. Cases with studentized residuals greater than 2,000 are listed.

LIITE 4: Regressiotulokset käytäntökomponenteille.

KOKONAISUORITUSKYKY

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,467	,218	,023	,48029

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5,663	22	,257	1,116	,347
	Residual	20,300	88	,231		
	Total	25,963	110			

Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	3,2046	4,3812	3,7855	,22690	111
Residual	1,1609	,99085	,00000	,42959	111
Std. Predicted Value	-2,560	2,625	,000	1,000	111
Std. Residual	-2,417	2,063	,000	,894	111

Coefficients

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics	
		B	Std. Error			Tolerance	VIF
1	Constant	4,314	,418	10,32	,000		
	wood	,228	,194	1,171	,245	,499	2,004
	pulppape	,004	,212	,021	,983	,448	2,233
	rubberpl	-,001	,147	-,006	,995	,566	1,767
	metal	,077	,150	,515	,608	,606	1,651
	vehicles	,119	,249	,477	,635	,779	1,284
	k3f1	,158	,115	1,376	,172	,656	1,524
	k3f2	-,059	,116	-,507	,613	,617	1,621
	k3f3	,076	,150	,508	,613	,715	1,398
	k3f4	,005	,115	,046	,964	,795	1,258
	k3f5	,135	,186	,725	,470	,620	1,613
	headqu_2	-,224	,155	-1,444	,152	,787	1,271
	trainpla	-,025	,131	-,193	,847	,488	2,050
	hrmstrat	-,059	,133	-,444	,658	,479	2,089
	shrd_2	-,018	,015	-1,244	,217	,621	1,609
	high_edu3	,008	,003	2,510	,014	,649	1,541
	k6b	,001	,002	,341	,734	,622	1,607
	k3d_3	-,100	,101	-,993	,323	,818	1,223
	log_size	-,151	,076	-1,982	,051	,527	1,896
	FAC1_1	,132	,050	2,621	,010	,806	1,241
	FAC2_1	,052	,054	,973	,333	,801	1,249
	FAC3_1	,097	,055	1,771	,080	,655	1,528
	FAC4_1	,026	,058	,454	,651	,795	1,258

Käytäntökomponentit: FAC1_1 – FAC4_1 vastaavat analyysitulosten komponentteja 1.– 4.

TYÖHYVINVOINTI

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,532	,283	,104	,72001

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18,046	22	,820	1,582	,069
	Residual	45,620	88	,518		
	Total	63,667	110			

Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2,0707	4,3978	3,1667	,40504	111
Residual	1,8925	1,6359	,00000	,64400	111
Std. Predicted Value	-2,706	3,040	,000	1,000	111
Std. Residual	-2,628	2,272	,000	,894	111

Coefficients

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	Constant	3,925	,627	6,263	,000
	wood	,270	,291	,925	,357
	pulppape	-,252	,318	-,793	,430
	rubberpl	,091	,221	,414	,680
	metal	,310	,224	1,381	,171
	vehicles	,523	,373	1,402	,164
	k3f1	-,179	,172	-1,037	,303
	k3f2	-,045	,174	-,258	,797
	k3f3	-,048	,224	-,213	,832
	k3f4	-,009	,173	-,053	,958
	k3f5	,136	,280	,488	,627
	headqu_2	-,218	,232	-,940	,350
	trainpla	,222	,196	1,135	,259
	hrmstrat	,201	,199	1,009	,316
	shrd_2	,010	,022	,463	,644
	high_edu3	,006	,005	1,164	,247
	k6b	,001	,003	,433	,666
	k3d_3	,074	,151	,490	,626
	log_size	-,247	,114	-2,163	,033
	FAC1_1	,181	,076	2,396	,019
	FAC2_1	,127	,081	1,570	,120
	FAC3_1	-,062	,082	-,751	,454
	FAC4_1	,170	,086	1,969	,052

Käytäntökomponentit: FAC1_1 – FAC4_1 vastaavat analyysitulosten komponentteja 1.– 4.

JOUSTAVUUS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,476	,226	,033	,65047

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10,889	22	,495	1,170	,296
	Residual	37,234	88	,423		
	Total	48,122	110			

Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2,8760	4,4824	3,6967	,31462	111
Residual	1,5427	1,3416	,00000	,58180	111
Std. Predicted Value	-2,608	2,497	,000	1,000	111
Std. Residual	-2,372	2,062	,000	,894	111

Coefficients

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	Constant	4,385	,566	7,743	,000
	wood	,208	,263	,791	,431
	pulppape	-,311	,287	-1,084	,281
	rubberpl	-,051	,199	-,254	,800
	metal	,258	,203	1,271	,207
	vehicles	,166	,337	,494	,623
	k3f1	,076	,156	,490	,625
	k3f2	-,033	,157	-,208	,836
	k3f3	,181	,203	,895	,373
	k3f4	,004	,156	,024	,981
	k3f5	,361	,253	1,429	,157
	headqu_2	-,082	,210	-,389	,698
	trainpla	-,074	,177	-,416	,679
	hrmstrat	-,117	,180	-,650	,517
	shrd_2	-,014	,020	-,682	,497
	high_edu3	,007	,004	1,539	,127
	k6b	,001	,002	,570	,570
	k3d_3	-,075	,137	-,551	,583
	log_size	-,183	,103	-1,782	,078
	FAC1_1	,150	,068	2,193	,031
	FAC2_1	,077	,073	1,060	,292
	FAC3_1	,185	,074	2,484	,015
	FAC4_1	,082	,078	1,056	,294

Käytännökomponentit: FAC1_1 – FAC4_1 vastaavat analyysitulosten komponentteja 1.– 4.

TUOTEINNOVAATIOT

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	115,179(a)	,282	,377

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6,284	8	,615

Malli on testin mukaan riittävä.

Classification Table(a)

	Observed	Predicted		Percentage Correct
		Products new to the market?	no	
Step 1	Products new to the market?	no	yes	
	no	34	16	68,0
	yes	15	45	75,0
	Overall Percentage			71,8

a The Cut value is .500

Casewise List(b)

Case	Selected Status(a)	Observed products new to the market	Predicted	Predicted Group	Temporary Variable	Resid	ZResid
106	S	**	,841	h	-,841	-2,302	
109	S	**	,857	h	-,857	-2,445	
125	S	**	,829	h	-,829	-2,198	
164	S	**	,906	h	-,906	-3,110	

a S = Selected, U = Unselected cases, and ** = Misclassified cases.
b Cases with studentized residuals greater than 2,000 are listed.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	
Step	wood	-,205	1,006	,041	1	,839	,815
	pulppape	-,531	1,033	,264	1	,607	,588
	rubberpl	,030	,723	,002	1	,967	1,030
	metal	-,413	,751	,302	1	,583	,662
	vehicles	2,282	1,368	2,784	1	,095	9,799
	k3f1	,238	,575	,172	1	,679	1,269
	k3f2	,689	,586	1,382	1	,240	1,991
	k3f3	,564	,740	,582	1	,446	1,758
	k3f4	1,933	,664	8,469	1	,004	6,909
	k3f5	,732	,939	,607	1	,436	2,078
	headqu_2	,408	,807	,256	1	,613	1,504
	trainpla	-,527	,683	,595	1	,441	,591
	hrmstrat	,481	,712	,456	1	,500	1,617
	shrd_2	,007	,089	,007	1	,934	1,007
	high_edu3	,039	,018	4,658	1	,031	1,039
	k6b	-,015	,009	2,536	1	,111	,985
	k3d_3	-,637	,513	1,541	1	,215	,529
	log_size	-,252	,370	,464	1	,496	,777
	FAC1_1	,954	,295	10,445	1	,001	2,597
	FAC2_1	,258	,274	,885	1	,347	1,294
	FAC3_1	,314	,290	1,170	1	,279	1,369
	FAC4_1	,734	,305	5,797	1	,016	2,084
	Constant	-,787	2,074	,144	1	,704	,455

Käytännökomponentit: FAC1_1 – FAC4_1 vastaavat analyysitulosten komponentteja 1.– 4.

LIITE 5: Rekursiivinen regressiopolkumalli.

```

MALLI 1
Observed variables
SHRD HIGHEDU HKOUL KOKO COMP1 COMP2 COMP3 COMP4 TYYTYV JOUST KOKSUOR
Correlation Matrix
1.000
0.423 1.000
-0.032 0.017 1.000
-0.099 0.061 0.245 1.000
0.011 -0.114 0.235 0.149 1.000
0.090 0.086 0.089 0.146 0.010 1.000
-0.123 -0.162 0.006 0.302 -0.002 -0.036 1.000
0.050 -0.045 0.063 0.133 -0.022 -0.118 -0.022 1.000
0.079 0.116 0.124 -0.175 0.147 0.111 -0.115 0.133 1.000
0.001 0.085 0.026 -0.178 0.176 -0.013 0.091 0.044 0.259 1.000
-0.028 0.147 0.012 -0.122 0.207 0.023 0.052 -0.011 0.275 0.817 1.000
Sample Size 111
Relationships
JOUST = COMP1 + KOKO
KOKSUOR = JOUST
TYYTYV = KOKSUOR

LISREL OUTPUT: MR MI RS EF
Path Diagram
End of Problem

```

```

MALLI 1
Correlation Matrix
          TYYTYV      JOUST      KOKSUOR      KOKO      COMP1
-----
TYYTYV      1.00
JOUST      0.26      1.00
KOKSUOR     0.28      0.82      1.00
KOKO      -0.17     -0.18     -0.12      1.00
COMP1      0.15      0.18      0.21      0.15      1.00

```

```

MALLI 1
Number of Iterations = 7

LISREL Estimates (Maximum Likelihood)

```

```

BETA
          TYYTYV      JOUST      KOKSUOR
-----
TYYTYV      - -      - -      0.28
              (0.09)
              2.96
JOUST      - -      - -      - -
KOKSUOR     - -      0.82      - -
              (0.06)
              14.66

GAMMA
          KOKO      COMP1
-----
TYYTYV      - -      - -
JOUST      -0.21     0.21
              (0.09)     (0.09)
              -2.22     2.20
KOKSUOR     - -      - -

```


Covariance Matrix of Y and X

	TYTYTV	JOUST	KOKSUOR	KOKO	COMP1
TYTYTV	1.00				
JOUST	0.22	1.00			
KOKSUOR	0.28	0.82	1.00		
KOKO	-0.04	-0.18	-0.15	1.00	
COMP1	0.04	0.18	0.14	0.15	1.00

PHI

	KOKO	COMP1
KOKO	1.00 (0.14) 7.31	
COMP1	0.15 (0.10) 1.52	1.00 (0.14) 7.31

PSI

Note: This matrix is diagonal.

TYTYTV	JOUST	KOKSUOR
0.92 (0.13) 7.31	0.93 (0.13) 7.31	0.33 (0.05) 7.31

Squared Multiple Correlations for Structural Equations

TYTYTV	JOUST	KOKSUOR
0.08	0.07	0.67

Goodness of Fit Statistics

Degrees of Freedom = 5
 Minimum Fit Function Chi-Square = 5.73 (P = 0.33)
 Normal Theory Weighted Least Squares Chi-Square = 5.55 (P = 0.35)
 Estimated Non-centrality Parameter (NCP) = 0.55
 90 Percent Confidence Interval for NCP = (0.0 ; 10.67)

Minimum Fit Function Value = 0.053
 Population Discrepancy Function Value (F0) = 0.0052
 90 Percent Confidence Interval for F0 = (0.0 ; 0.100)
 Root Mean Square Error of Approximation (RMSEA) = 0.032
 90 Percent Confidence Interval for RMSEA = (0.0 ; 0.14)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.50

Expected Cross-Validation Index (ECVI) = 0.24
 90 Percent Confidence Interval for ECVI = (0.23 ; 0.33)
 ECVI for Saturated Model = 0.28
 ECVI for Independence Model = 1.12

Chi-Square for Independence Model with 10 Degrees of Freedom = 109.55
 Independence AIC = 119.55
 Model AIC = 25.55

Saturated AIC = 30.00
 Independence CAIC = 138.05
 Model CAIC = 62.56
 Saturated CAIC = 85.51

Normed Fit Index (NFI) = 0.95
 Non-Normed Fit Index (NNFI) = 0.99
 Parsimony Normed Fit Index (PNFI) = 0.47
 Comparative Fit Index (CFI) = 0.99
 Incremental Fit Index (IFI) = 0.99
 Relative Fit Index (RFI) = 0.90

Critical N (CN) = 288.22

Root Mean Square Residual (RMR) = 0.049
 Standardized RMR = 0.049
 Goodness of Fit Index (GFI) = 0.98
 Adjusted Goodness of Fit Index (AGFI) = 0.94
 Parsimony Goodness of Fit Index (PGFI) = 0.33

MALLI 1

Fitted Covariance Matrix

	TYITYV	JOUST	KOKSUOR	KOKO	COMP1
TYITYV	1.00				
JOUST	0.22	1.00			
KOKSUOR	0.28	0.82	1.00		
KOKO	-0.04	-0.18	-0.15	1.00	
COMP1	0.04	0.18	0.14	0.15	1.00

Fitted Residuals

	TYITYV	JOUST	KOKSUOR	KOKO	COMP1
TYITYV	0.00				
JOUST	0.03	- -			
KOKSUOR	0.00	0.00	0.00		
KOKO	-0.14	- -	0.02	- -	
COMP1	0.11	0.00	0.06	- -	- -

Summary Statistics for Fitted Residuals

Smallest Fitted Residual = -0.14
 Median Fitted Residual = 0.00
 Largest Fitted Residual = 0.11

Standardized Residuals

	TYITYV	JOUST	KOKSUOR	KOKO	COMP1
TYITYV	- -				
JOUST	0.64	- -			
KOKSUOR	- -	- -	- -		
KOKO	-1.45	- -	0.43	- -	
COMP1	1.15	- -	1.15	- -	- -

Summary Statistics for Standardized Residuals

Smallest Standardized Residual = -1.45
 Median Standardized Residual = 0.00
 Largest Standardized Residual = 1.15

MALLI 1

Modification Indices and Expected Change

Modification Indices for BETA

	TYITYV -----	JOUST -----	KOKSUOR -----
TYITYV	- -	0.42	- -
JOUST	0.09	- -	0.32
KOKSUOR	0.42	- -	- -

Expected Change for BETA

	TYITYV -----	JOUST -----	KOKSUOR -----
TYITYV	- -	0.10	- -
JOUST	-0.05	- -	-0.33
KOKSUOR	-0.05	- -	- -

Modification Indices for GAMMA

	KOKO -----	COMP1 -----
TYITYV	2.41	0.98
JOUST	- -	- -
KOKSUOR	0.19	1.35

Expected Change for GAMMA

	KOKO -----	COMP1 -----
TYITYV	-0.14	0.09
JOUST	- -	- -
KOKSUOR	0.02	0.07

No Non-Zero Modification Indices for PHI

Modification Indices for PSI

	TYITYV -----	JOUST -----	KOKSUOR -----
TYITYV	- -	- -	- -
JOUST	0.06	- -	- -
KOKSUOR	0.42	0.32	- -

Expected Change for PSI

	TYITYV -----	JOUST -----	KOKSUOR -----
TYITYV	- -	- -	- -
JOUST	-0.04	- -	- -
KOKSUOR	-0.04	-0.11	- -

Modification Indices for THETA-EPS

	TYITYV	JOUST	KOKSUOR
	-----	-----	-----
TYITYV	- -		
JOUST	0.11	0.32	
KOKSUOR	0.42	0.40	0.42

Expected Change for THETA-EPS

	TYITYV	JOUST	KOKSUOR
	-----	-----	-----
TYITYV	- -		
JOUST	0.02	0.14	
KOKSUOR	-0.04	-0.09	0.15

Modification Indices for THETA-DELTA-EPS

	TYITYV	JOUST	KOKSUOR
	-----	-----	-----
KOKO	2.88	0.05	0.24
COMP1	1.50	1.21	0.79

Expected Change for THETA-DELTA-EPS

	TYITYV	JOUST	KOKSUOR
	-----	-----	-----
KOKO	-0.15	-0.01	0.03
COMP1	0.11	-0.07	0.05

Maximum Modification Index is 2.88 for Element (1, 1) of THETA DELTA-EPSILON

Indirect Effects of X on Y

	KOKO	COMP1
	-----	-----
TYITYV	-0.05 (0.03)	0.05 (0.03)
	-1.76	1.75
JOUST	- -	- -
KOKSUOR	-0.17 (0.08)	0.17 (0.08)
	-2.19	2.18

Indirect Effects of Y on Y

	TYITYV	JOUST	KOKSUOR
	-----	-----	-----
TYITYV	- -	0.22 (0.08) 2.90	- -
JOUST	- -	- -	- -
KOKSUOR	- -	- -	- -

Time used: 0.000 Seconds