

Sisällysluettelo

ALKUSANAT	4
ALKUSANAT E-KIRJA –VERSIOON	5
SISÄLLYSLUETTELO	6
LYHYT SANASTO VASTA-ALKAJILLE	7
1. MONIMUUTTUJAMENETELMÄT IHMISTIEETEISSÄ	9
1.1 MONIMUUTTUJA-AINEISTON ERITYISPIIRTEITÄ	10
1.2 AINEISTON ALUSTAVA TARKASTELU	13
1.2.1 Korrelaatio ja käyräviivainen yhteys	13
1.2.2 Outlierit	14
1.2.3 Normaalisuus	16
1.2.4 Multikollinearisuus ja singulaarisuus	17
1.3 KIRJAN RAKENTEESTA	18
2. HAVAINTOJEN RYHMITTELY- JA LUOKITTELUANALYYSIT	19
2.1 EROTTELU- JA LUOKITTELUANALYYSI (DA)	19
2.1.1 Missä tilanteessa toimii parhaiten	20
2.1.2 Rajoitukset ja oletukset	20
2.1.3 Lyhyesti teoriasta ja käsitteistä	20
2.1.4 Lisätestit ja jatkoanalyysit	25
2.1.5 Tekninen suoritus SPSS-ympäristössä ja tulkinta	25
2.2 RYHMITTELYANALYYSI (CA)	37
2.2.1 Missä tilanteessa toimii parhaiten	37
2.2.2 Rajoitukset ja oletukset	37
2.2.3 Lyhyesti teoriasta ja käsitteistä	37
2.2.4 Lisätestit ja jatkoanalyysit	40
2.2.5 Tekninen suoritus SPSS-ympäristössä ja tulkinta	40
2.3 ANSWERTREE -ANALYYSI (ATA)	45
2.3.1 Missä tilanteessa toimii parhaiten	45
2.3.2 Rajoitukset ja oletukset	45
2.3.3 Lyhyesti teoriasta ja käsitteistä	46
2.3.4 Lisätestit ja jatkoanalyysit	51
2.3.5 Tekninen suoritus SPSS-ympäristössä ja tulkinta	51
3. LOPUKSI	57
LIITE A. AINEISTOSSA KÄYTETYT ALKUPERÄISET MUUTTUJAT	58
LIITE B. MATRIISILASKENNASTA KEVYESTI	61
LÄHTEET	64
ASIA- JA HENKILÖHAKEMISTO	66
METODOLOGIA-SARJAN KIRJOITTAJASTA	71

puosalla aineistoa voidaan testata, kuinka hyvin erottelu- ja luokittelufunktiot toimivat muilla kuin analyysissa mukana olleilla havainnoilla.

2.1.4 Lisätestit ja jatkoanalyysit

Muiden monimuuttujamenetelmien tapaan myös erotteluanalyysissa on oleellista tarkistaa, pitävätkö mallin oletukset paikkaansa. **Kovarianssimatriisien yhtäsuuruutta voidaan tutkia Boxin M-testillä.** Valitettavasti M-testi on herkkä multinormaalisuusoletuksen rikoutumiselle ja suurille otosko'uille, mistä syystä se saattaa hylätä yhtäsuuruusoletuksen liian herkästi. Myös **Log-determinantti** (eli determinantin logaritmi) **tutkii ryhmien välisiä kovarianssirakenteita.** Kuvassa 5.1 havainnollistimme kahden muuttujan muodostamaa erotteluavaruutta ja havaitsimme, että havainnot muodostavat eräänlaisia ellipsejä havaintoavaruuteen. Jos lisäämme avaruuteen vielä yhden muuttujan, havainnot ovat ikään kuin amerikkalaisen jalkapallon muotoisessa rykelmässä. Vaikkemme visuaalisesti tiedäkään, miltä 5 tai useampiulotteinen avaruus näyttää, Log-determinantti mittaa niiden soikiopallojen sisään jäävien alkoiden tiheyttä, ts. sitä, mikä on ellipsoidin sisään jäävien alkoiden suhteellinen osuus. Tietenkin **voimme tarkastella myös yksittäisten muuttujien variansseja yksisuuntaisella varianssianalyysillä.** Levenen testi mittaa varianssien yhtäsuuruutta¹⁴. Jos muuttujien variansseissa on tilastollisesti merkitseviä eroja, voidaan muuttujat muuntaa standardoiduiksi muuttujiksi, jolloin varianssiksi tulee 1 ja keskiarvoksi 0. Tämä muunto käy SPSS-ohjelmistossa helpoiten *Analyze – Descriptives – Descriptive Statistics* -komennolla¹⁵. Erotteluanalyysi on herkkä outlieriille. Näitä **poikkeavia havaintoja voidaan etsiä esimerkiksi boxplot** -kuvalla (ks. luku 1.2.2).

Havainnoille voidaan laskea ns. kanoniset muuttujat samaan tapaan kuin pääkomponentti-, faktori- tai regressioanalyysissa laskettiin faktoripisteet tai ennustearvot. Kanonisten muuttujien avulla voidaan havainnollistaa havaintojen sijoittumista eri erottelijoille esimerkiksi sirontakuvioiden (*Scatter-plot*) avulla, mikäli erottelufunktioita on enemmän kuin yksi.

2.1.5 Tekninen suoritus SPSS-ympäristössä ja tulkinta

Erotteluanalyysillä **haluamme tutkia, voidaanko motiivien avulla erotella toisistaan erilaisia harrastusryhmiä ja toisaalta luokitella eri harrastusryhmiin kuuluvia harrastajia oikeisiin harrastusryhmiin löytyneiden erottelufunktioiden perusteella.** Tutkittuja harrastusryhmiä oli kahdeksan taulukon 2.1 mukaisesti urheiluharrastuksesta kirjoittamisharrastukseen. Motiiveina käsitellään 11 motiivien dimensiota, jotka edustivat Madsenin tarveluokituksen eri alueita. Motiivit oli laskettu yhdistäen summaamalla osioita. Summaamisen jälkeen summamuuttujat standardoitiin, että ne olisivat yhteismitallisia (ks. tarkemmin Metsämuuronen 1995). Mukaan tulevat muuttujat olivat seuraavat:

¹⁴ Katso tarkemmin *SPSS aloittelevan tutkijan käytössä*, Metodologia-sarja 5.

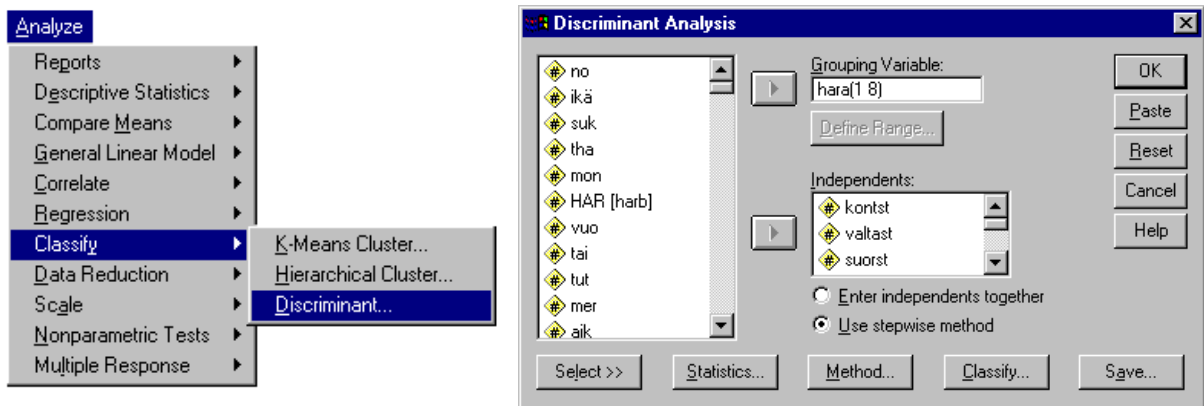
¹⁵ Katso tarkemmin *SPSS aloittelevan tutkijan käytössä*.

Taulukko 2.1 Erotteluanalyysiin mukaan tulevat muuttujat

Mja	selite	Merkitys mallissa
HARA	Harrastusalue 1=urheiluharrastus 5=käsityöharrastus 2=kuvataideharrastus 6=matematiikkaharrastus 3=musiikkiharrastus 7=kieliharrastus 4=luontoharrastus 8=kirjallisuusharrastus	Ryhmittelevä muuttuja
KONTST	Kontaktimotiivi, standardoitu	Selittävä tekijä
VALTAST	Valtamotiivi, standardoitu	Selittävä tekijä
SUORST	Suoritusmotiivi, standardoitu	Selittävä tekijä
HYÖTYST	Hyötymotiivi, standardoitu	Selittävä tekijä
OMAEHTST	Omaehtoisuusmotiivi, standardoitu	Selittävä tekijä
T		
UTELST	Uteliaisuusmotiivi, standardoitu	Selittävä tekijä
MIELIKST	Mielikuvutusmotiivi, standardoitu	Selittävä tekijä
LIKST	Liikunnan tarve, standardoitu	Selittävä tekijä
KOKEST	Kokemusten tarve, standardoitu	Selittävä tekijä
LUOMST	Luomisen tarve, standardoitu	Selittävä tekijä
JÄNNST	Jännityksen tarve, standardoitu	Selittävä tekijä

Perusnäkökulma

Erottelu- ja luokitteluanalyysi alkaa valinnoilla *Analyze – Classify – Discriminant...*



Valinnat

Erotteluanalyysissä voidaan tehdä viidenlaisia valintoja: voidaan valita muuttuja, jonka perusteella tehdään ristiinvaldointi (*Select*), määrätä tunnuslukuja, kertoimia ja tulostettavia matriiseja (*Statistics*), määrittää analyysimenetelmiä (*Method*) ja luokitteluanalyysiin liittyviä valintoja (*Classification*) sekä tallentaa analyysituloksia (*Save*). Muuttujien valinta voidaan tehdä joko pakotetusti (*Enter independents together*) tai askeltavasti (*Use stepwise method*). Valitaan päävalikosta askeltava menettely, sillä ei ole teoriaa tai kokemusta siitä, mitkä motiividimen-

sioista todella olisivat erotteluanalyysin kannalta oleellisia tekijöitä; askeltava menettely karsii muuttujien joukosta erottelun kannalta epäoleelliset tekijät pois.

Mikäli halutaan – kuten suotavaa olisi, jos aineistoa on riittävästi – ristiinvalidoida aineisto eli luoda erottelufunktio vain osalla aineistoa ja testata lopuilla, toimivatko funktiot myös muilla kuin koulutushavainnoilla, tarvitaan muuttuja, joka kertoo, mitkä havainnot kouluttavat ja mitkä testaavat erottelufunktiot. Tätä varten luotiin dataan uusi muuttuja nimeltä CROSSV50. Tässä muuttujassa havainto saa arvon 1, kun halutaan sen kuuluvan koulutusryhmään ja 0, kun halutaan sen kuuluvan testiryhmään. Muuttujaa rakennettaessa on ensin luotu apumuuttuja RANDOM, jossa kullekin havainnolle on tuotettu satunnaisluku väliltä 0 - 1. Sitten on *Analyze – Descriptives – Frequencies* -valinnalla etsitty mediaaniarvo. Mediaaniarvon perusteella on tehty muuttujamuunnos siten, että mikäli satunnaisluku on ollut suurempi kuin mediaani, havainto saa arvon 1, muussa tapauksessa se saa arvon 0. Näin siis puolet aineistosta käytetään mallin rakentamiseen. SPSS:n erotteluanalyysissä mallin testaamiseen käytetään kuitenkin koko aineistoa, eli sekä arvon 1 että arvon 0 saaneet havainnot ovat testiryhmässä. Tätä uutta muuttujaa CROSSV50 hyödynnetään päänäköymän *Select*-muuttujana. *Select*-näppäimen painaminen aktivoi päävalikkoon uuden osan, johon lisätään kyseinen muuttuja ja annetaan *Value*-komennolla se arvo (eli tässä tapauksessa arvo 1), joka ohjaa valitsemaan koulutusaineiston käyttöön. Teemme seuraavat valinnat oletusten lisäksi:

- **Select>>**
Selection Variable: CROSSV50
Value: 1
[Valitaan muuttuja ja sen arvo, jonka avulla aineisto jaetaan ristiinvalidointiosiin.]
- **Statistics**
Descriptives: Box's M
Function Coefficient: Unstandardized
Matrices: Within-Groups correlation
[Testataan soveltuuko korrelaatiomatriisi erotteluanalyysiin, valitaan standardoimattomat kertoimet ja halutaan tietää, muinkaa suuret korrelaatiot ryhmien välillä on.]
- **Method**
Oletusarvot
[Oletuksena on, että teemme analyysin käyttäen Wilksin Lambda-kriteeriä.]
- **Classification**
Prior Probabilities: Compute from group sizes
Display: Summary table
Plots: Combined-groups, Separate-Groups, Territorial map
[Haluaamme painottaa havaintoja niiden ryhmien koon suhteessa – pienet ryhmät saavat vähemmän painoarvoa. Haluaamme tulostuksiin yhteenvototaulukon sekä kuvia, joissa on kukin ryhmä erikseen kahden ensimmäisen erottelijan suhteen kuvattuna.]

Tulokset ja niiden tulkinta

Tulosten tulkinta alkaa kuvailemalla aineisto:

Analysis Case Processing Summary				Group Statistics			
Unweighted Cases		N	Percent	HARA	Valid N (listwise)		
					Unweighted	Weighted	
Valid		350	48,3	1 KONTST	81	81,000	
Excluded	Missing or out-of-range group codes	8	1,1	VALTAST	81	81,000	
	At least one missing discriminating variable	0	,0	SUORST	81	81,000	
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	,0	HY TM TYST	81	81,000	
	Unselected	366	50,6	OMAHTST	81	81,000	
Total	374	51,7	UTELST	81	81,000		
Total	724	100,0	MIELIKST	81	81,000		
				LIKST	81	81,000	
				KOKEST	81	81,000	
				LUOMST	81	81,000	
				J□NNST	81	81,000	
				2 KONTST	56	56,000	
				VALTAST	56	56,000	
				SUORST	56	56,000	
				HY TM TYST	56	56,000	
				OMAHTST	56	56,000	

Havaintojen kokonaismäärä kuvaava taulu (*Analysis Case Processing Summary*) kertoo, että mukana on 350 havaintoa 724:stä, joiden avulla luodaan erottelufunktiot. Loput 374 ja mallituksessa mukana olleet havainnot muodostavat joukon, jolla aineisto lopulta luokitellaan. **Havaintojen määriä ryhmissä** kuvaavasta taulusta (*Group Statistics*) mukaan on otettu vain osa tulostukseen. Havaitsemme, että esimerkiksi liikuntaharrastuksesta (HARA-muuttujan luokka 1) mukaan on tullut 81 havaintoa ja kuvataideharrastuksesta 56. Ohjelmistoa on pyydetty painottamaan (*Weighted*) havaintoja ryhmän lukumäärän suhteessa. Näin tehden mallissa otetaan huomioon se, että tietyistä ryhmistä saattaa olla mukana vain joitakin havaintoja; näiden ryhmien painoarvo ei ole yhtä suuri kuin niiden ryhmien, joista mukana on paljon havaintoja.

Pooled Within-Groups Matrices												
	KONTST	VALTAST	SUORST	HY TM TYST	OMAHTST	UTELST	MIELIKST	LIKST	KOKEST	LUOMST	J□NNST	
Correlation	KONTST	1,000	,257	,361	,261	,281	,106	,256	,255	,503	,344	,256
	VALTAST	,257	1,000	,201	,163	-,015	,017	,218	,136	,344	,219	,205
	SUORST	,361	,201	1,000	,407	,547	,242	,315	,281	,566	,509	,257
	HY TM TYST	,261	,163	,407	1,000	,484	,352	,286	,106	,468	,405	,118
	OMAHTST	,281	-,015	,547	,484	1,000	,352	,311	,134	,453	,441	,116
	UTELST	,106	,017	,242	,352	,352	1,000	,353	-,009	,309	,242	,108
	MIELIKST	,256	,218	,315	,286	,311	,353	1,000	,146	,469	,483	,240
	LIKST	,255	,136	,281	,106	,134	-,009	,146	1,000	,388	,117	,225
	KOKEST	,503	,344	,566	,468	,453	,309	,469	,388	1,000	,506	,400
	LUOMST	,344	,219	,509	,405	,441	,242	,483	,117	,506	1,000	,171
	J□NNST	,256	,205	,257	,118	,116	,108	,240	,225	,400	,171	1,000

Seuraavassa taulussa (*Pooled Within-Groups Matrices*) on esitetty muuttujien välinen **korrelaatiomatriisi**. Näin siis esimerkiksi kontaktimotiivi (KONTST) korreloi valtamotiivin (VALTAST) kanssa 0.257:n verran. Korrelaatiot on laskettu normaalista poikkeavalla tavalla, mistä syystä arvot saattavat erota tavallisesta Pearsonin tulomomenttikorrelaatiokertoimesta. Ensin on laskettu varianssit ja kovarianssit kussakin ryhmässä. Tämän jälkeen varianssit