

## **Faktorianalyysi**

- Faktorianalyysi on monimuuttujamenetelmä, jonka avulla pyritään löytämään muuttujajoukosta yhteisiä piirteitä tai ulottuvuuksia. Eksploratiivisen faktorianalyysin avulla etsitään keskenään eniten korreloivat muuttujat ja muodostetaan niistä faktoreita eli muuttujien taustalla olevia ns. **piilomuuttujia**. Menetelmän avulla pyritään selvittämään mahdollisimman paljon muuttujien välistä vaihtelua. Se mahdollistaa useiden vähintään hyvällä järjestysasteikolla (esimerkiksi Likertin asteikolla) mitattujen muuttujien antaman informaation tiivistämisen muutamaankin faktoriin.
- Saadussa ratkaisussa muuttujien saamat **lataukset** kertovat, kuinka paljon faktorien avulla pystytään selittämään kyseisten muuttujien vaihtelua. Lataukset saavat arvoja välillä -1 ja 1. Mitä lähempänä latauksen itseisarvo on yhtä (1) sitä vahvemmin muuttuja latautuu faktorilla (eli sitä paremmin faktori selittää muuttujan vaihtelua). Jos muuttujan lataus on arvoltaan negatiivinen, kertoo se ainoastaan sen, että muuttujan arvot korreloivat negatiivisesti faktorin arvojen kanssa.
- Kun faktorit on saatu, tutkijan tulee vielä **tulkita** ne. Tulkinta tapahtuu tutkimalla, mitkä alkuperäisistä muuttujista ovat eniten korreloituneita kyseisen faktorin kanssa, ja antamalla faktorille nimi sen perusteella.

### **Esimerkki**

Etsitään faktorianalyysin avulla kirjaston asiakkaita koskevasta aineistosta sopivia muuttujaryhmiä, jotka voitaisiin yhdistää kuvaamaan samaa asiaa. Tutkimuslomakkeessa oli 28 asioiden tärkeyttä kuvaavaa kysymystä, jotka oli valmiiksi ryhmitelty asiakokonaisuuksiksi. Kokeillaan minkälaisia asiakokonaisuuksia faktorianalyysi antaa.

Ennen faktorianalyysia voisi tarkastella alkuperäisten muuttujien välisiä korrelaatioita. Mikäli korrelaatiokertoimien itseisarvot ovat hyvin pieniä, on selvää, ettei löydy faktoreita, joille muuttujien vaihtelut keskittyvät. Tässä esimerkissä korrelaatiokertoimien joukossa on joitakin lähellä nollaa olevia kertoimia, mutta toisaalta useita 0,5:n ja 0,6:n välillä olevia, muutamia jopa yli 0,6.

Otetaan faktorianalyysiin aluksi mukaan kaikki 28 asioiden tärkeyttä kuvaavaa muuttujaa, jotka liittyvät seuraaviin kysymyksiin:

## Kuinka tärkeitä seuraavat seikat ovat Teille asioidessanne yleensä kirjastossa?

Ympäröikää lähinnä mielipidettänne oleva vaihtoehto jokaiselta riviltä.

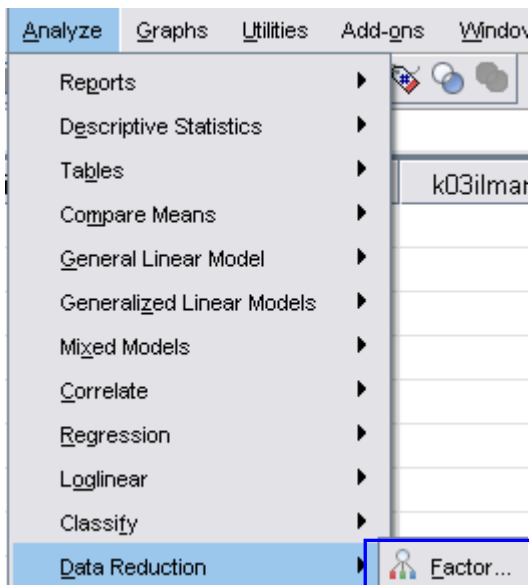
tärkeä	Ei merkitse mitään	Vähäpätöinen	Tärkeä	Erittäin
1. Yleinen viihtyvyys	1	2	3	4
2. Hyvät pysäköintitilat	1	2	3	4
3. Hyvä ilmanvaihto	1	2	3	4
4. Selkeät yleisopasteet	1	2	3	4
5. Rauhalliset lukupaikat	1	2	3	4
6. Aineiston monipuolisuus	1	2	3	4

tärkeä	Ei merkitse mitään	Vähäpätöinen	Tärkeä	Erittäin
7. Aineiston löydettävyys	1	2	3	4
8. Aineiston aiheenmukainen järjestys	1	2	3	4
9. Aineiston esillepano	1	2	3	4
10. Aineistonäyttelyt	1	2	3	4
11. Aineiston varausmahdollisuus	1	2	3	4
12. Kaukopalvelu	1	2	3	4
13. Henkilökunnan riittävyys	1	2	3	4
14. Henkilökunnan palveluvalttius	1	2	3	4
15. Henkilökunnan ammattitaito	1	2	3	4
16. Yksilöllinen palvelu	1	2	3	4
17. Tiedonhakupalvelu	1	2	3	4
18. Yleisöpäätteet	1	2	3	4
19. Internet-päätteet	1	2	3	4
20. Palvelujen monipuolisuus	1	2	3	4
21. Palvelujen laatu	1	2	3	4
22. Lainaustapahtuman nopeus	1	2	3	4
23. Lainauksen maksuttomuus	1	2	3	4
24. Yleisötilaisuudet	1	2	3	4
25. Satutunnit	1	2	3	4
26. Itsepalvelulainaus	1	2	3	4
27. Itsepalvelupalautus	1	2	3	4
28. Kahvilapalvelut	1	2	3	4

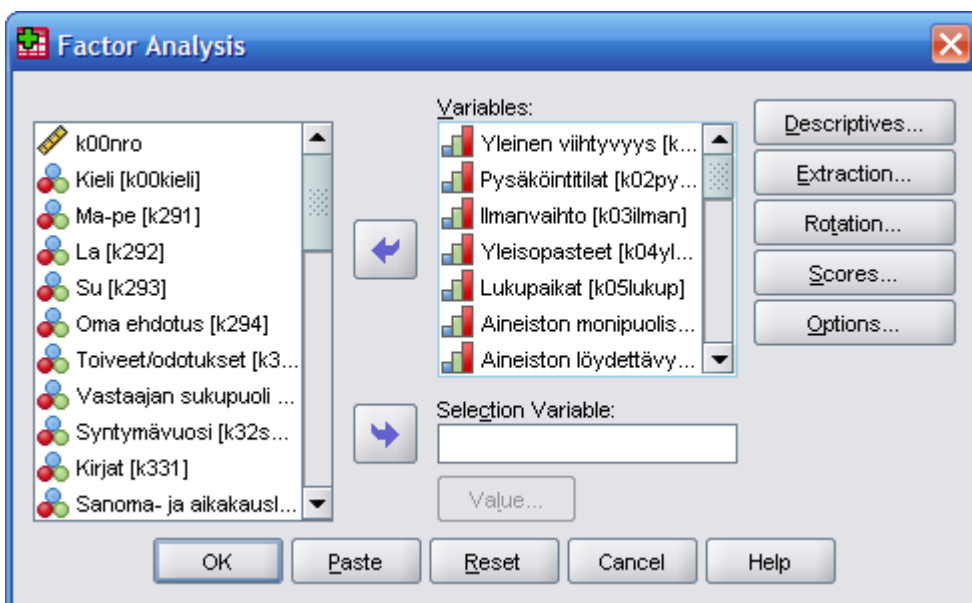
## Faktorianalyysin suorittaminen

Datastiedosto: DATAT\Kirjasto.sav

SPSS: Analyze/Data Reduction/Factor

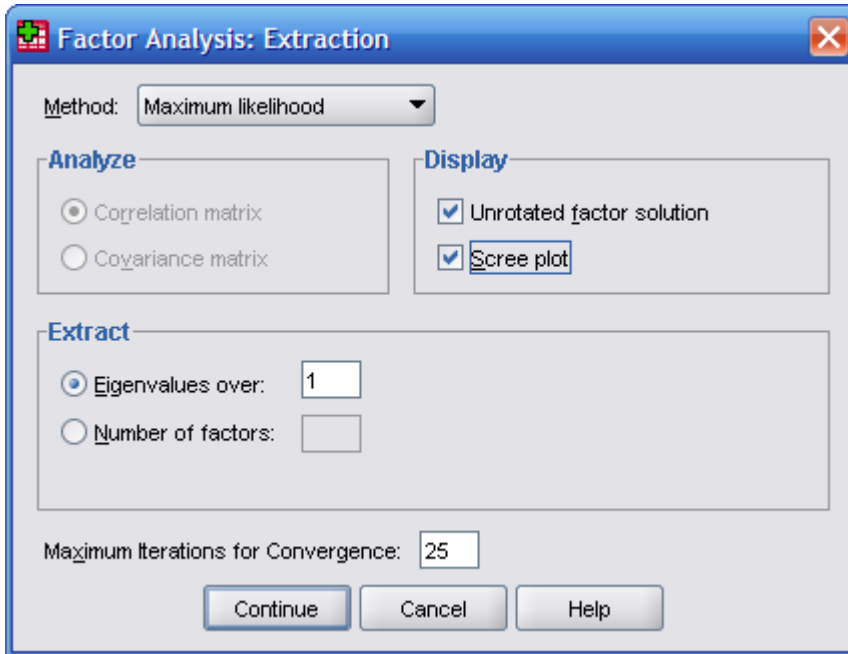


- Valitaan kaikki 28 muuttujaa *Variables*-ruutuun.

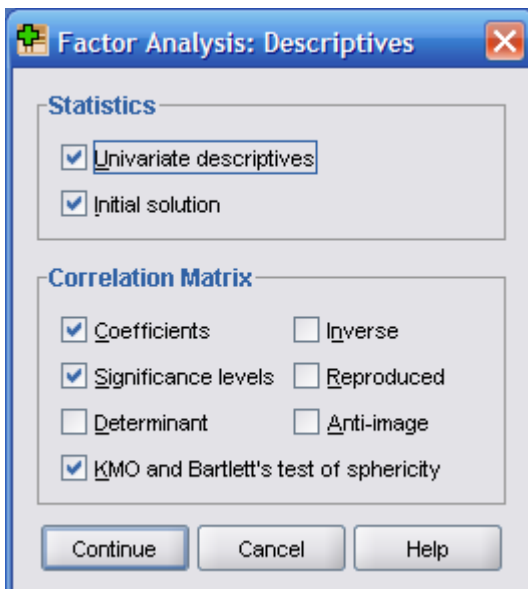


- Painamalla *Extraction*-painiketta voidaan valita, millä menetelmällä muuttujien lataukset etsitään. Ensimmäinen näistä, (joka on oletusmenetelmä), ei ole varsinainen

faktorointimenetelmä. Eri menetelmät voivat tuottaa hyvinkin erilaisen ratkaisun. *Maximum likelihood* -vaihtoehto on suositeltava, jos havaintoja on riittävästi (yli 100). Käytetään tässä sitä.



- Tehdään muihin ikkunoihin seuraavien kuvioiden mukaiset valinnat.

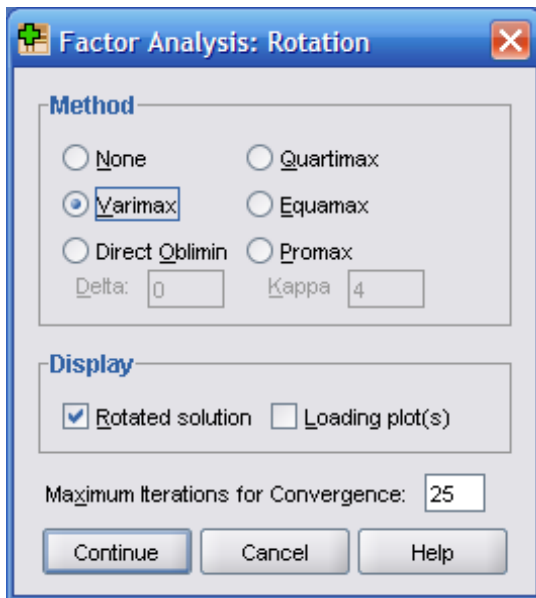


*Descriptives*-ikkunan valinnalla *Univariate descriptives* saadaan tulostaulukko, jossa on muuttujien keskiarvot, keskihajonnat ja vastanneiden määrät.

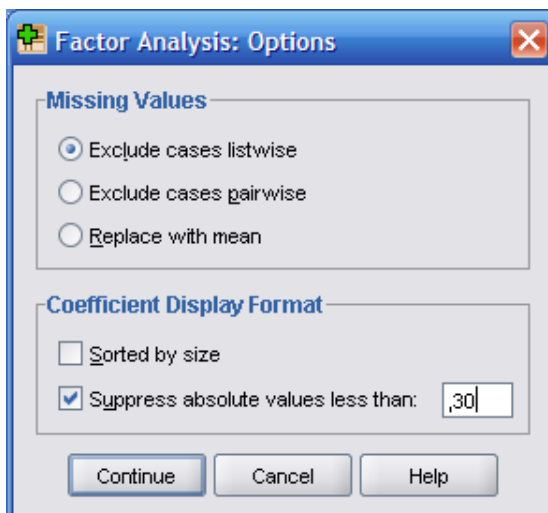
Valitsemalla *Coefficients* saadaan korrelaatiomatriisi.

Rastittamalla *Significance levels* saadaan korrelaatiokertoimien yksisuuntaiset merkitsevyydet.

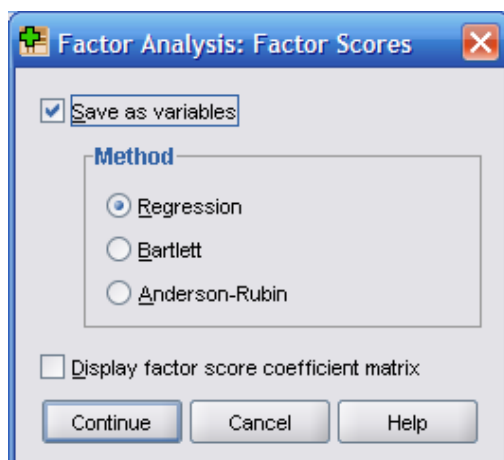
Viimeinen valinta *KMO and...* testaa, onko korrelaatiomatriisi sovelias faktorianalyysiin.



*Rotation*-ikkunassa voidaan valita faktoreiden rotaatiomenetelmä. Saatua faktoriratkaisua pyritään yleensä rotatoimaan tulkinnan helpottamiseksi, esim. nk. *Varimax*-rotaatiolla pyritään ratkaisuun, jossa muuttujien lataukset faktoreilla ovat mahdollisimman suuria tai pieniä.



Kun *Options*-ikkunassa asetetaan kertoimen rajaksi 0,3, ei lopullisessa ratkaisussa ole näkyvissä alle 0,3:n suuria latausarvoja ja tulkinta on helpompi tehdä.



*Factor Scores*-ikunan *Save as variables* tallentaa faktoripistemäärät tiedostoon. Jokaisesta faktorista luodaan oma muuttuja.

- Ensimmäiseksi tulostuu tunnuslukutaulukko.

#### Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Yleinen viihtyvyys	3,19	,609	354
Pysäköintitilat	2,71	1,014	354
Ilmanvaihto	3,29	,649	354
Yleisopasteet	3,36	,656	354
Lukupaidat	3,21	,819	354
Aineiston monipuolisuus	3,73	,506	354
Aineiston löydettävyys	3,66	,514	354
Aineiston järjestys	3,48	,608	354
Aineiston esillepano	3,13	,691	354
Aineistonäyttelyt	2,37	,769	354
Aineiston varausmahdollisuus	3,16	,770	354
Kaukopalvelu	2,62	,942	354
Riittävä henkilökunta	2,98	,641	354
Palveluaitis henkilökunta	3,46	,625	354
Ammatitaitoinen henkilökunta	3,54	,602	354
Yksilöllinen palvelu	3,10	,735	354
Tiedonhakupalvelu	3,13	,755	354
Yleisöpäätteet	2,97	,923	354
Internet-päätteet	2,49	1,044	354
Monipuolinen palvelu	3,23	,623	354
Palvelujen laatu	3,38	,590	354
Nopea lainaus	3,02	,771	354
Maksuton lainaus	3,55	,729	354
Yleisötilaisuudet	2,06	,824	354
Satutunnit	1,95	1,031	354
Itsepalvelulainaus	2,64	,887	354
Itsepalvelupalautus	2,92	,844	354
Kahvilapalvelut	2,01	,967	354

- *Descriptive* -ikkunan valintojen mukaan seuraavaksi tulostuu korrelaatiomatriisi, jonka Sig.-arvoista nähdään, että valtaosa korrelaatiokertoimista poikkeavat nolasta tilastollisesti merkitsevästi. Matriisia ei ole tässä esitetty sen suuren koon vuoksi.
- Sekä Kaiserin testi (*Kaiser-Meyer -Olkin Measure of Sampling Adequacy*), jonka arvo on suurempi kuin 0,6, että Barlettin sväärisyydestä ( $p < 0,001$ ) osoittavat, että korrelaatiomatriisi on sovelias analyysiin.



### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,819
Bartlett's Test of Sphericity	Approx. Chi-Square	3056,603
	df	378
	Sig.	,000

### Communalities<sup>a</sup>

	Initial	Extraction
Yleinen viihtyvyys	,221	,209
Pysäköintitilat	,184	,128
Ilmanvaihto	,309	,845
Yleisopasteet	,292	,221
Lukupaikat	,237	,153
Aineiston monipuolisuus	,256	,246
Aineiston löydettävyys	,508	,733
Aineiston järjestys	,538	,646
Aineiston esillepano	,430	,471
Aineistonäyttelyt	,428	,606
Aineiston varausmahdollisuus	,416	,471
Kaukopalvelu	,333	,335
Riittävä henkilökunta	,415	,419
Palveluaitis henkilökunta	,561	,710
Ammatti tai toinen henkilökunta	,483	,544
Yksilöllinen palvelu	,451	,499
Tiedonhakupalvelu	,481	,516
Yleisöpäätteet	,558	,710
Internet-päätteet	,470	,548
Monipuolinen palvelu	,478	,898
Palvelujen laatu	,471	,479
Nopea lainaus	,298	,261
Maksuton lainaus	,158	,106
Yleisötöläisyydet	,425	,645
Satutunnit	,313	,380
Itsepalvelulainaus	,494	,728
Itsepalvelupalautus	,499	,625
Kahvilapalvelut	,259	,305

Extraction Method: Maximum Likelihood.

a. One or more communalities estimates greater than 1 were encountered during iterations. The resulting solution should be interpreted with caution.

Kommunaliteetti mittaa, kuinka suuri osa muuttujien varianssista pystytään selittämään faktorien avulla. *Initial* antaa alustavan ratkaisun ja *Extraction* lopulliset kommunaliteetit. Osa muuttujien kommunaliteeteista on kovin pieniä, alle 0,3.

Tehdään faktorianalyysi uudelleen poistamalla muuttujat, joiden kommunaliteetti on alle 0,3.

- Kun tämän jälkeen tehdään faktorianalyysi uudelleen, on jälleen muutaman muuttujan kommunaliteetti kovin pieni. Poistetaan alle 0,3:n kertoimen omaavat muuttujat. Näin pyritään helpottamaan lopullisen tuloksen tulkintaa.
- Muuttujien vähentämisen jälkeen saadaan seuraava taulukko kommunaliteeteista:

### Communalities<sup>a</sup>

	Initial	Extraction
Aineiston löydettävyys	,457	,506
Aineiston järjestys	,525	,779
Aineiston esillepano	,407	,425
Aineistonäyttelyt	,419	,466
Aineiston varausmahdollisuus	,334	,336
Riittävä henkilökunta	,376	,420
Palveluallittis henkilökunta	,551	,702
Ammatti tai toinen henkilökunta	,476	,565
Yksilöllinen palvelu	,428	,490
Tiedonhakupalvelu	,455	,487
Yleisöpäätteet	,547	,802
Internet-päätteet	,441	,509
Monipuolinen palvelu	,430	,782
Palvelujen laatu	,438	,505
Yleisötillaisuudet	,373	,564
Satutunnit	,287	,356
Itsepalvelulainaus	,492	,869
Itsepalvelupalautus	,506	,563

Extraction Method: Maximum Likelihood.

<sup>a</sup>. One or more communalities estimates greater than 1 were encountered during iterations. The resulting solution should be interpreted with caution.

- Seuraava taulukko kertoo pääkomponenttien ominaisarvoista ja selitysosuuksista. Kuudella pääkomponentilla ominisarvo (*Initial Eigenvalues*) on suurempi kuin 1,0. Nämä kuusi faktoria pystyvät selittämään noin 70 % muuttujien varianssista.
- Scree Plot -kuviosta nähdään myös ominaisarvot. Kuudennen faktorin jälkeen ominaisarvot ovat alle yhden, joten lisäinformaatiota ei kovin paljon saada, vaikka faktoreiden määrää lisättäisiin.
- Faktoriratkaisun tulkinta on syytä tehdä rotatoidusta faktorimatriisista. Ensimmäisellä faktorilla suuria latauksia saavat henkilökuntaan liittyvät muuttujat: henkilökunnan riittävyys, palveluallittius, ammattitaito ja palvelun yksilöllisyys. Ensimmäistä faktoria voitaisiin kutsua '*henkilökuntafaktoriksi*'. Toisessa faktorissa suuria latauksia saavat tiedonhakuun liittyvät muuttujat, joten nimetään se '*tiedonhakufaktoriksi*' jne.

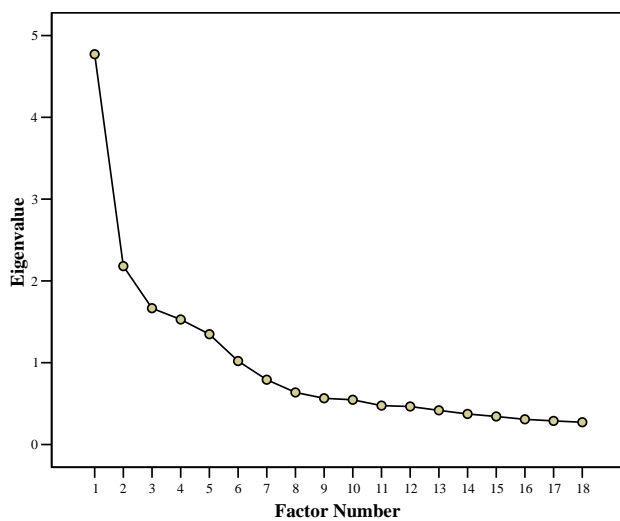


### Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,772	26,509	26,509	4,071	22,615	22,615	2,478	13,768	13,768
2	2,180	12,112	38,621	1,552	8,620	31,234	1,907	10,592	24,360
3	1,666	9,256	47,877	1,623	9,014	40,248	1,770	9,835	34,195
4	1,528	8,491	56,368	1,153	6,408	46,656	1,449	8,050	42,245
5	1,348	7,491	63,859	,720	4,000	50,657	1,425	7,914	50,159
6	1,020	5,667	69,527	1,007	5,592	56,249	1,096	6,090	56,249
7	,792	4,401	73,928						
8	,637	3,538	77,466						
9	,565	3,139	80,605						
10	,547	3,038	83,642						
11	,476	2,643	86,286						
12	,465	2,582	88,867						
13	,418	2,323	91,190						
14	,374	2,076	93,266						
15	,343	1,905	95,172						
16	,308	1,712	96,883						
17	,289	1,603	98,487						
18	,272	1,513	100,000						

Extraction Method: Maximum Likelihood.

### Scree Plot



## Rotatoitu faktorimatriisi

Rotated Factor Matrix<sup>a</sup>

	Factor					
	1	2	3	4	5	6
Aineiston löydettävyys			,673			
Aineiston järjestys			,858			
Aineiston esillepano			,542			
Aineistonäyttelyt				,599		
Aineiston varausmahdollisuus	,381	,321				
Riittävä henkilökunta	,611					
Palvelualtis henkilökunta	,815					
Ammattitaitoinen henkilökunta	,731					
Yksilöllinen palvelu	,649					
Tiedonhakupalvelu		,620				
Yleisöpäätteet		,880				
Internet-päätteet		,661				
Monipuolinen palvelu						,804
Palvelujen laatu	,355					,555
Yleisötilaisuudet				,725		
Satutunnit				,569		
Itsepalvelulainaus					,910	
Itsepalvelupalautus					,697	

Extraction Method: Maximum Likelihood.


Rotation Method: Varimax with Kaiser Normalization.

<sup>a</sup>. Rotation converged in 5 iterations.

Nimetään saadut faktorit, esimerkiksi:

- faktori 1 **henkilökuntafaktori**
- faktori 2 **tiedonhakufaktori**
- faktori 3 **aineistofaktori**
- faktori 4 **yleisötillaisuudet**
- faktori 5 **itsepalvelufaktori**
- faktori 6 **palvelufaktori** ....

- Faktorianalyysi antaa faktoripistemäärät eli laskee jokaiselle tapaukselle näiden faktorien saamat arvot. Faktoripistemääriin vaikuttavat kaikki faktorilla latautuvat muuttujat kukin omalla painollaan. Tilasto- ohjelma normeeraa pistemäärät niin että pistemäärien keskiarvo on 0 ja keskihajonta 1. Näiden pistemäärien avulla voidaan verrata eri ryhmien tuloksia ja käyttää jatkoanalyysiin. Niiden saamat arvot ovat kuitenkin kovin epähavainnollisia.
- Usein faktorianalyysin jatkotoimenpiteenä muodostetaan summa- tai keskiarvomuuttujia, joissa on mukana kuhunkin faktoriin parhaiten latautuneet ns. kärkimuuttujat. Reliabiliteetikertoimen avulla voidaan mitata, kuinka hyvin näihin uusiin muuttujiin yhdistetyt muuttujat mittaavat samaa asiaa. Reliabiliteettikerroin on



välillä [0,1] ja suuret kertoimen arvot ilmoittavat korkeasta reliabiliteetista. Kertoimen arvoa voidaan pitää hyvänä, jos se on yli 0,7. Haastattelututkimukseen sisältyy usein monia satunnaisvirhettä aiheuttavia tekijöitä, niin että käytännössä joudutaan tyytymään alhaisempiinkin kertoimen arvoihin.

**Reliabiliteettikerroin** saadaan SPSS-ohjelmalla:

*Analyze / Scale / Reliability analysis*

Tämän laskemisesta on suoritusohje SPSS-osion esimerkissä *Reliabiliteetti*.

Reliabiliteettikertoimen avulla voidaan edellisen esimerkin tapauksessa mitata, kuinka hyvin esimerkiksi faktoriin 1 parhaiten latautuneet muuttujat mittaavat samaa asiaa. Aluksi voidaan ottaa mukaan kaikki 6 muuttujaa, joiden lataukset ovat yli 0,3 (asetettiin tulostusrajaksi), mutta huomataan, että paras reliabiliteetti saadaan, kun mukaan jätetään vain neljä henkilökuntaan liittyvää muuttujaa. Niinpä faktorille annetaan nimi henkilökuntafaktori.

Jatkotoimenpiteenä voidaan muodostaa summamuuttuja tai keskiarvomuuttuja näistä neljästä kysymyksestä. Itse käytän mieluummin keskiarvomuuttujaa, koska silloin alkuperäisten muuttujien asteikko säilyy. Tämä helpottaa oikean arvion saamista tulosten tasosta. Tämän yhdistetyn muuttujan muodostaminen on selvitetty kirjassa luvun 9 esimerkissä 6.

Tämän yhdistetyn muuttujan avulla voidaan esimerkiksi tutkia, onko sukupuolten välillä eroa siinä, kuinka tärkeänä pitää henkilökuntaan liittyviä asioita kirjastossa. Tuloksena saadaankin, että sukupuolten välillä on tilastollisesti merkitsevä ero ( $p=0,000$ ). Naisille henkilökunta on tärkeämpi (ka 3,3 asteikolla 1- 4) kuin miehille (ka 3,0). Sen sijaan ikäryhmien välillä ei ole tilastollisesti merkitsevää eroa. Sukupuolten tai ikäryhmien välinen ero voitaisiin testata suoraan faktoripistemäärienkin avulla, mutta niiden arvot eivät antaisi havainnollista kuvaa asian tärkeydestä.