UNIVERSITY OF JYVÄSKYLÄ
SCHOOL OF BUSINESS AND
ECONOMICS

Heikki Karjaluoto

SPSS opas markkinatutkijoille

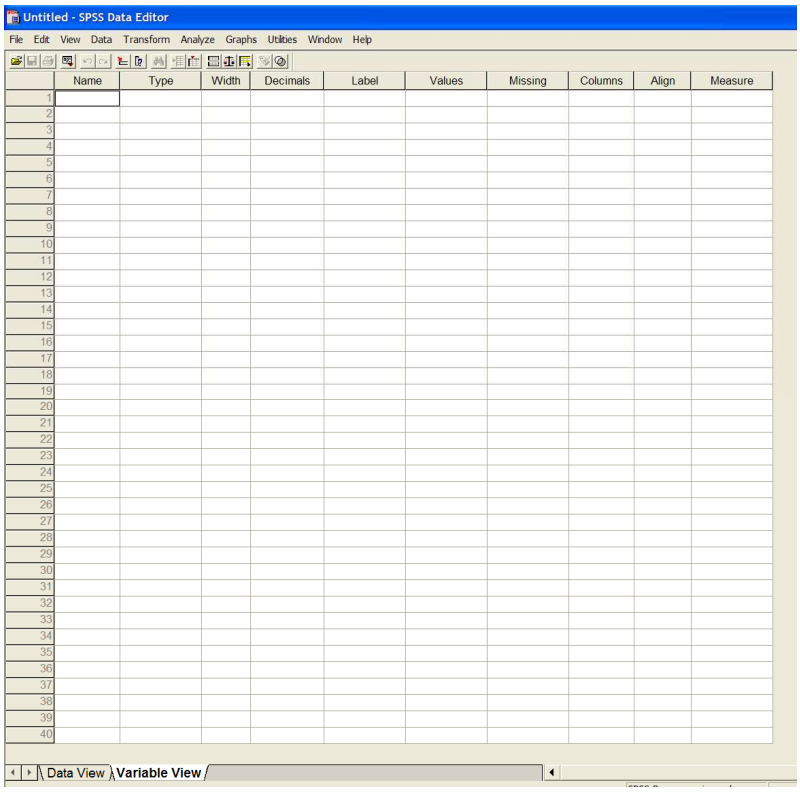
Working paper N:o 344 / 2007

SISÄLTÖ

1 SPSS ohjelman perusteet

SPSS –ohjelman (v.10.1 eteenpäin) käynnistyttyä avautuu **Data Editor** näkymä (Kuva 1), joka koostuu kahdesta välilehdestä (**Data View** sekä **Variable View**). Variable View –lehdelle nimetään aineiston muuttujat sekä määritellään ne. Data Editor –näkymiä ei voi olla samanaikaisesti käytössä kuin yksi, joten on tärkeätä että heti kun aloittaa aineiston syöttämisen ohjelmaan tallentaa sen kiintolevylle (File / Save as).

Data View –näkymään syötetään lomakkeista vastaukset edellyttäen että kullekin vastaukselle on annettu numeerinen arvo eli aineisto on koodattu numeeriseen muotoon. SPSS kykenee käsittelemään myös sanallisia vastauksia, mutta niiden analysointi ohjelmalla on lähes mahdotonta. Niinpä järkevä tapa onkin pyrkiä antamaan kaikille vastauksille numeerinen koodiarvo, joitakin poikkeuksia lukuun ottamatta (esim. katuosoite, vastaajan nimi jne.).



KUVA 1. Data Editor näkymä

2 Kyselylomakkeen koodaus ja muuttujien määrittely ja aineiston syöttö

Kun aineisto on saatu kerättyä kyselylomakkeilla pitää siinä olevat muuttujat määritellä SPSS muotoon. Kyselylomakkeen muuttujat kannattaa määritellä SPSS -ohjelmaan samassa järjestyksessä kuin ne ovat itse lomakkeessa ja antaa niille samat numerot (esim. kysymys yksi määritellään SPSS -ohjelmaan koodilla k1). Ohessa esimerkki kyselylomakkeesta ja sen koodauksesta SPSS -ohjelmaan.

6. Arvioikaa suhtautumistanne seuraaviin teknologioihin/palveluihin ja asioihin.
Asteikko 1 (en pidä lainkaan) – 7 (pidän erittäin paljon).

	1	2	3	4	5	6	7	En tiedä
a) Tietokoneet	☐	☐	☐	☐	☐	☐	☐	☐
b) Internet vapaa-ajan käytössä	☐	☐	☐	☐	☐	☐	☐	☐
c) Internet hyötykäytössä (esim. tiedon etsintä)	☐	☐	☐	☐	☐	☐	☐	☐
d) Verkko-ostaminen	☐	☐	☐	☐	☐	☐	☐	☐
e) Sähköposti	☐	☐	☐	☐	☐	☐	☐	☐
f) Itsepalvelut verkossa (esim. pankkipalvelut)	☐	☐	☐	☐	☐	☐	☐	☐
g) Matkapuhelin puhekäytössä	☐	☐	☐	☐	☐	☐	☐	☐
h) Matkapuhelimen tekstiviestipalvelut	☐	☐	☐	☐	☐	☐	☐	☐
i) Matkapuhelimen datapalvelut	☐	☐	☐	☐	☐	☐	☐	☐
j) Internet-palveluiden personointi (esim. oma nimi näkyy sivustolla)	☐	☐	☐	☐	☐	☐	☐	☐
k) Internet-palveluissa reaaliaikainen viestintä myyjän/asiantuntijan kanssa	☐	☐	☐	☐	☐	☐	☐	☐
l) Video- ja ääniyhteyden käyttö verkkopalveluissa	☐	☐	☐	☐	☐	☐	☐	☐
m) Henkilökohtainen, kasvotusten tapahtuva palvelu	☐	☐	☐	☐	☐	☐	☐	☐
n) Kaupungilla ostoksilla käynti, "shoppailu"	☐	☐	☐	☐	☐	☐	☐	☐

8. Mikä matkapuhelin Teillä on tällä hetkellä pääasiallisesti käytössänne?

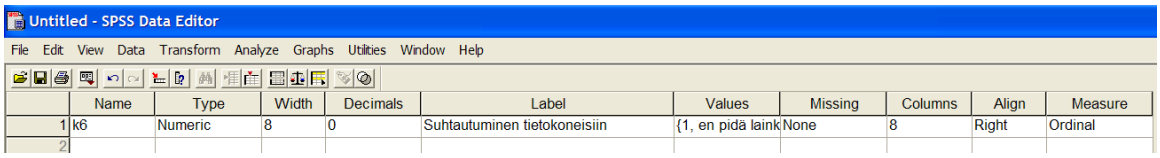
a) Nokia ☐
b) Sony-Ericsson ☐
c) Siemens ☐
d) Motorola ☐
e) Panasonic ☐
f) Muu ☐

Koodataan
1=Nokia,
2=SonyE,
3=Siemens jne.

KUVA 2. Kyselylomakkeen koodaus

Kysymys 6a (tietokoneet) kannattaa nimetä SPSS Data View näkymässä koodilla k6a (*Name*). Muuttujan tyypiksi määritellään *Numeric*, eli muuttuja saa tällöin numeerisia arvoja. Muita tyyppejä joita jonkin verran käytetään ovat päivämäärä (*Date*) jos sitä on kysytty sekä *String* –muuttuja, jota käytetään mikäli vastaus on kirjoitetussa muodossa (esim. henkilön nimi tai

katuosoite). *Width* –sarakkeessa määritellään muuttujan saamat lukuarvot (oletuksena 8, mikä tarkoittaa sitä että muuttujan leveys *Data View* –valikossa voi olla 8 merkkiä. Tässä tapauksessa (k6a) leveys on max. 1 merkki (voi saada vain arvot 1-7), joten halutessaan tutkija voi sen määritellä arvoksi 1. *Decimals* –välilehti kertoo sen kuinka monta desimaalia muuttujalle määritellään *Data View* valikossa. Oletuksena on kaksi desimaalia, joskin tämäntyyppisissä kysymyksissä ei juuri koskaan tarvita desimaaleja joten suositellaan arvoksi nolla. *Label* –sarakkeeseen syötetään muuttujan nimi kokonaisuudessaan. Se tulee näkyviin kun aineistosta ajetaan ajoja. Muuta merkitystä sillä ei ole. *Values* –sarakkeeseen kannattaa määritellä muuttujan saamat arvot eli kysymyksen 6a arvot ovat 1-7, jossa 1=en pidä lainkaan ja 7=pidän erittäin paljon. Arvot 2-6 pitää myös nimetä eli tässä tapauksessa 2 olisi esimerkiksi ”pidän hyvin vähän”, 3 ”pidän vähän” ja 4 ”neutraali” jne. HUOM! Kysymyksessä on myös vastausvaihtoehto ”en tiedä” ja se kannattaa jättää koodaamatta eli EI anneta sille arvoa 8, koska se vääristää tuloksia.



KUVA 3. Muuttujan nimeäminen

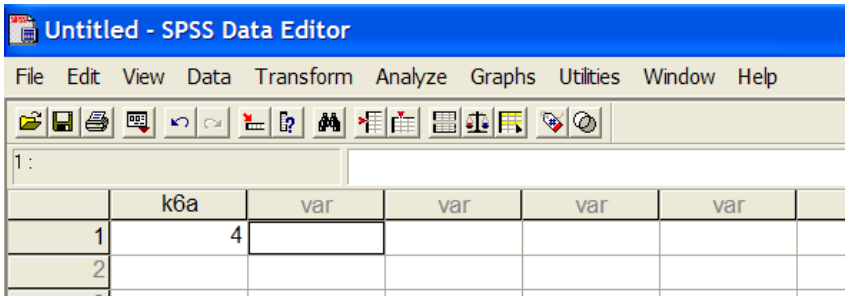
Missing –sarakkeeseen voidaan määritellä mitä tehdään jos muuttuja ei saa mitään arvoa eli lomakkeessa on tyhjä kohta vastauksen kohdalla. Suositellaan jätettävän oletusarvo *None. Columns* –sarakkeessa voidaan määritellä muuttujan leveys *Data View* –näkymään. Lähinnä kosmeettinen määritelmä eli jos muuttujia halutaan saada kerralla näkymään paljon voi tämän sarakkeen oletusarvoa pienentää. *Align* –sarake on myös kosmeettinen määrittely, eli sen avulla voi määritellä haluaako muuttujien arvot vasempaan tai oikeaan reunaan solua vaiko keskelle *Data view* –näkymässä. *Measure* –sarakkeessa määritellään muuttujan mitta-

asteikko. Yleisimmin käytettyjä ovat *Nominal* (eli nominaalisasteikolliset muuttujat kuten sukupuoli, siviilisääty; tai kuten kysymys 8.) ja *Ordinal* (eli järjestysasteikollinen muuttuja kuten k6 kokonaisuudessaan). Kolmas vaihtoehto on *Scale*, jota käytetään silloin kun kyseessä on suhde- tai intervalliasteikollinen muuttuja (esim. tulotaso avoimena kysymyksenä tai ikä avoimena kysymyksenä).

Muuttujien nimeäminen noudattaa samaa logiikkaa kuin Windows –ohjelmat yleensäkin eli jos on saatu määriteltyä muuttuja 6a voidaan sen tietoja kopioida seuraaviin määrittelyihin (eli 6b, 6c, 6d jne.) klikkaamalla esim. hiiren oikeaa nappia *Values* –solun kohdalla ja liittämällä se seuraaviin soluihin.

2.1. Aineiston syöttäminen

Kun lomakkeen kaikki muuttujat on saatu nimettyä syötetään lomakkeiden tiedot horisontaalisesti vasemmalta oikealle *Data View* –näkymään (KUVA 4).



KUVA 4. Data View –näkymä

Ensimmäinen lomake syötetään riville 1. Eli kuten kuvassa näkyy on vastaaja vastannut kysymykseen 6a 4 (”neutraali) ja tutkija on syöttänyt sitä vastaavan arvon soluun. Toinen lomake syötetään riville 2 ja näin jatketaan kunnes kaikki lomakkeet on syötetty. Muista tallentaa

aineistoa aina syöttämisen aikana, koska SPSS -ohjelmalla on tapana syystä tai toisesta välillä kaatua, varsinkin jo aineisto on suuri eli tuhansia vastauksia.

Mikäli lomakkeita on satoja (tai tuhansia) kannattaa sen syöttäminen jakaa esim. kahdelle henkilölle. SPSS.sav tiedostot voi myöhemmin sitten yhdistää käyttämällä hyväksi toimintoa **Data / Merge Files / Add Cases**.

2.2. Muuttujien muokkaaminen

Kun aineisto on saatu koodattua ja syötettyä SPSS -ohjelmaan voidaan muuttujille tehdä monenlaisia asioita. Yleisimmin törmää kolmeen tapaukseen. Nämä ovat muuttujien uudelleen koodaus eli rekoodaus, suodatin- eli filtterimuuttujan tekeminen sekä kolmantena summamuuttujan tekeminen.

2.2.1. Muuttujan rekoodaus

Muuttuja voidaan jälkikäteen luokitella uudelleen eli esim. esimerkin muuttuja 6a voidaan tehdä kaksiluokkaiseksi siten että se saa vain arvot 1 tai 2. Rekoodaus tehdään käskyllä **Transform / Recode / Into Different Variables** (tai into same variables mutta tällöin uusi muuttuja korvaa vanhan, ei suositella). Avautuvassa rekoodaus –ikkunassa pitää uudelle muuttujalle antaa nimi sekä kuvaava selite (*Label*). Muista painaa **Change** –laatikkoa, muuten muuttuja ei vaihdu. Laatikosta ”**Old and New Values**” annetaan muuttujalle uudet arvot vanhojen tilalle.

2.2.2. Suodatin- eli filtterimuuttujan tekeminen

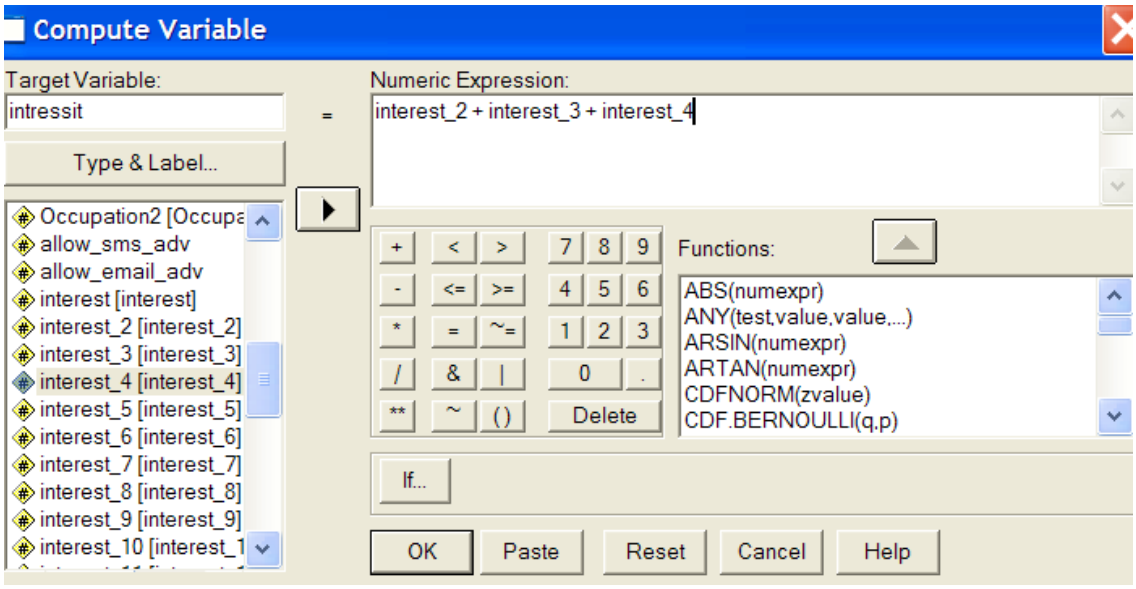
Suodatinmuuttujaa käytetään kun halutaan aineistosta vain tietynlaiset vastaajat (esim. vain naiset). Suodatinmuuttuja tehdään valitsemalla **Data / Select Cases** jolloin avautuvasta ikkunasta valitaan tapa jolla suodatus tehdään. Yleensä valitaan vaihtoehto ***If condition is satisfied***, jolloin valitaan vain tietyntyyppiset vastaajat. Avautuneessa **Select Cases: If** –ikkunassa valitaan vasemmasta laatikosta muuttuja jonka perusteella suodatus halutaan tehdä eli tässä esimerkkitapauksessa valitaan muuttuja sukupuoli ja siirretään se oikeanpuoleiseen soluun. Määritellään että sp=2 (sp koodattu 1=mies ja 2=nainen) eli otetaan mukaan vain naiset. Painetaan OK ja suodatus on valmis.

HUOM. Kun suodatuksen jälkeen halutaan jälleen analysoida kaikkia vastaajia pitää suodatinmuuttuja käydä ottamassa pois eli valitaan **Select Cases** valikosta **All cases** jälleen.

2.2.3. Summamuuttujan tekeminen

Summamuuttujalla tarkoitetaan muuttujaa joka on muodostettu kahdesta tai useammasta yksittäisestä muuttujasta laskemalla niiden arvot yhteen. Summamuuttuja tehdään valikosta **Transform / Compute**, jolloin avautuu ikkuna *Compute Variable* (Kuva 5). Avautuvaan ikkunaan kirjoitetaan vasemman yläkulman laatikkoon uuden muuttujan nimi eli tässä esimerkkitapauksessa luodaan muuttuja joka koskee vastaajan mielenkiinnonkohteita. Summamuuttujan muodostamisessa on otettu mukaan kolme mielenkiinnon kohdetta (interest_2, interest_3, interest_4 –muuttujat) jotka summaamalla yhteen saadaan muodostettua uusi muuttuja. Muuttujat erotellaan toisistaan matemaattisella yhtälöllä (usein + merkki). *Functions* –valikosta voidaan valita esim. Mean(numexp, numexp) –tyyppinen muuttuja, jos halutaan tehdä esi-

merkiksi kouluarvosana-asteikollisesta muuttujasta keskiarvosummamuuttuja (esim. Mean (mja1, mja2, mja3, mja4).



KUVA 5. Summamuuttujan tekeminen

Painamalle OK summamuuttuja muodostuu. Summamuuttuja ilmestyy *Data View* –näkymään oikeaan reunaan eli sinne minne kaikki uudet muuttujat muodostuvat.

3 Aineiston analysointi

Kun tarvittavat esivalmistelut on tehty niin aineiston analysointia aloitettaessa on syytä tiedostaa muutama asia. Keskeinen päätös on valinta parametrisen ja ei-parametrisen testin välillä. Parametristen testien (esimerkiksi t-testi) käyttöön liittyy olettamuksia perusjoukon tunnusluvuista (parametreista) ja muuttujien jakauman muodosta. Niissä muuttujan on oltava vähintään välimatka-asteikon tasoinen. Jos edellytykset eivät ole voimassa, pitäisi valita ei-parametrinen testi (esim. khin neliö –testi). Yleensä on suositeltavaa käyttää parametrisia testejä jos mahdollista, sillä ne ovat voimakkaampia eli ne suosittelevat helpommin väärän nollahypoteesin hylkäämistä. Nyrkkisääntönä sosiaalitieteissä voidaan sanoa että kun aineisto on

mitattu metrisesti käyttäen vähintäänkin järjestysasteikollisia muuttujia, ja aineiston koko on yli 50 tai toisena raja-arvona käytetty yli 100 voi soveltaa parametrisia testejä. Osa tutkijoista on sitä mieltä että esimerkiksi asennemittauksissa kun käytetään asteikkoja 1-5 tai 1-7 pitäisi aina soveltaa ei-parametrisia menetelmiä, kun taas toiset ovat sitä mieltä että voidaan käyttää myös parametrisia. Tieteellisistä artikkeleista esim. liiketaloustieteen puolella näkee selvästi että pienilläkin aineistoilla käytetään lähes yksinomaan parametrisia menetelmiä kuten t-testiä ja regressioanalyyseja.

Kuitenkin kannattaa muistaa että mikäli aineisto on pieni (vastauksia <50) tarkoittaa se sitä että kovin monimutkaisia tilastollisia analyysejä ei voida tehdä eikä ole järkevää tehdä. Mikäli aineisto on suurempi (esim. <100), voidaan jo tietyin varauksin käyttää huoletta parametrisia tilastollisia menetelmiä kuten faktorianalyysiä, t-testiä, regressioanalyysiä jne. Aineiston koon lisäksi analysointiin vaikuttaa muuttujien tyyppi. Kyselytutkimuksissa muuttujat useimmiten ovat järjestysasteikollisia muuttujia, joilla ei aina ole järkevää tehdä esim. regressioanalyysiä. Useimmiten turvaudutaankin faktorianalyysiin ja selittämiseen ristiintaulukoinnein, keskiarvovertailuihin ja varianssianalyyseihin. Myös tutkimuksen lähtökohta eli tutkimuksen tavoitteet ja tutkimusongelma sekä teoreettinen viitekehys antavat lähtökohdat oikean analysointimenetelmän valintaan. Tieteellisessä tutkimuksessa kannattaa vilkaista mitä menetelmiä muut tutkijat ovat käyttäneet ja enimmäkseen turvautua niihin. Esim. tieteelliset sosiaalitieteiden artikkelit usein perustuvat aineiston tiivistämiseen faktorianalyysin avulla ja sen pohjalta rakennetun mallin testaukselle (esim. regressioanalyysit ja LISREL-mallit).

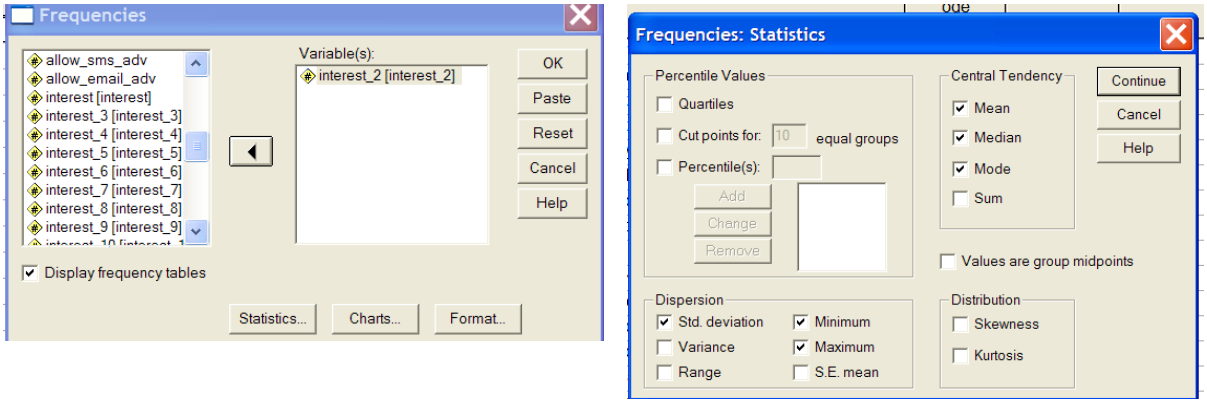
Aineiston analysointi voidaan luokitella analyysitehtävittäin seuraavasti:

1. **Kuvaaminen** (suorat jakaumat ja tilastolliset tunnusluvut)
2. **Tiivistäminen** (faktorianalyysi, ryhmittely- eli klusterianalyysi)

3. **Selittäminen** (ristiintaulukointi, keskiarvovertailut, t-testi, varianssianalyysit, moni-muuttujainen varianssianalyysi (manova), regressioanalyysi, erotteluanalyysi)

3.1. Suorat jakaumat ja tilastolliset perustunnusluvut

Aineiston analysointi kannattaa aloittaa katsomalla aineistosta suorat jakaumat eli frekvenssit sekä kysymyksittäin muutamia tilastollisia tunnuslukuja kuten minimi, maksimi, mediaani, moodi, keskiarvo ja %-jakaumat. Suorat jakaumat toiminto löytyy seuraavasti: **Analyze / Descriptive Statistics / Frequencies**, jolloin avautuvaan soluun valitaan vasemmalta muuttujat joita halutaan tarkastella. Aluksi on hyvä ottaa suorat jakaumat kaikista muuttujista jolloin selviää aineiston luonne ja voidaan havaita myös mahdolliset syöttövirheet (jos esim. 5-asteikollinen muuttuja on saanut maksimiarvoksi 50 on kyseessä koodausvirhe joka on helppo käydä *Data View* –näkymässä korjaamassa).



KUVA 6. Suorat jakaumat

Statistics –painikkeen alta avautuvassa ikkunassa (Kuva 6) voidaan valita tilastollisia tunnuslukuja joista käytetyimpiä kuvataan seuraavaksi (Tilastokeskus 2005):

- **Mean** eli keskiarvo on keskiluvuista kaikkein tavallisin. Se ilmoittaa, mihin kohtaan muuttujan jakauman ***keskikohta*** mitatulla ulottuvuudella sijoittuu.

- **Median** eli mediaani on paljon käytetty keskiluku, joka ilmoittaa jakauman tyypillisen arvon. Täsmällisemmin kyseessä on ***jakauman keskimmäinen havaintoarvo***, kun havainnot on järjestetty suuruusjärjestykseen.
- **Moodi** ilmoittaa on muuttujan jakauman huippukohta eli ***eniten tapauksia sisältävä luokka***.

Nominaali- ja järjestysasteikollisista muuttujista kannatta raportoida mediaani ja moodi. Välimatka- ja suhdeasteikollisista lisäksi keskiarvo, joskin sosiaalitieteissä keskiarvoa käytetään hyvin yleisesti myös järjestysasteikollisten muuttujien kanssa.

Muita käytettyjä tunnuslukuja ovat **hajontaa kuvaavat tunnusluvut** eli **std. deviation** ja **variance**. Std.deviation eli keskihajonta joka on tärkein ns. hajontaluku eli luku, joka mittaa havaintoarvojen hajaantumista muuttujan jakauman keskikohdan ympärille. Keskihajonta siis kuvaa havaintoarvojen keskimääräistä etäisyyttä keskiarvosta. Keskihajonnan neliö on nimeltään **varianssi**, joka ilmoittaa miten suuria keskimäärin ovat neliöidyt poikkeamat keskiarvosta.
Mitä pienempi on keskihajonta (ja varianssi), sitä tiiviimmin havaintoaineisto on keskittynyt keskiarvon ympärille.

----------*-*--*-*- |--*-*--*--*---------- **pieni hajonta**

---*-----*----*-------*---|-*-------*-----*----*------* **suuri hajonta**

Keskihajontaa ja varianssia voidaan käyttää, mikäli muuttaja on välimatka-asteikollinen tai suhdeasteikollinen, mutta raportoidaan myös järjestysasteikollisista muuttujista.

Minimum ja maximum kertovat minimi- ja maksimiarvot muuttujalle. On järkevä käyttää esim. aineiston koodauksen tarkistamisessa.

Charts –painikkeen alta voidaan valita tulosteeseen kysymyksistä diagrammeja jos halutaan. SPSS -ohjelman tuottama grafiikka ei pärjää vertailussa esim. Wordille tai Powerpointille joten on suoriteltavampaa tehdä taulukot ja kuvat näissä ohjelmissa SPSS:n sijaan.

3.2.1. Tulosteen tulkinta

Frekvenssitaulukoista raportoidaan usein frekvenssit eli vastausten määrä ja niiden jakautuminen prosentuaalisesti. Tuloste näyttää seuraavanlaiselta:

Statistics

age_new

N	Valid	183
	Missing	27
Mean		2,5246
Median		3,0000
Mode		3,00
Std. Deviation		1,08850
Minimum		1,00
Maximum		5,00

age_new

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Below 25	46	21,9	25,1	25,1
	26-34	31	14,8	16,9	42,1
	35-49	73	34,8	39,9	82,0
	50-64	30	14,3	16,4	98,4
	65 or over	3	1,4	1,6	100,0
	Total	183	87,1	100,0	
Missing	System	27	12,9		
Total		210	100,0		

KUVA 7. Esimerkki frekvenssitaulukosta

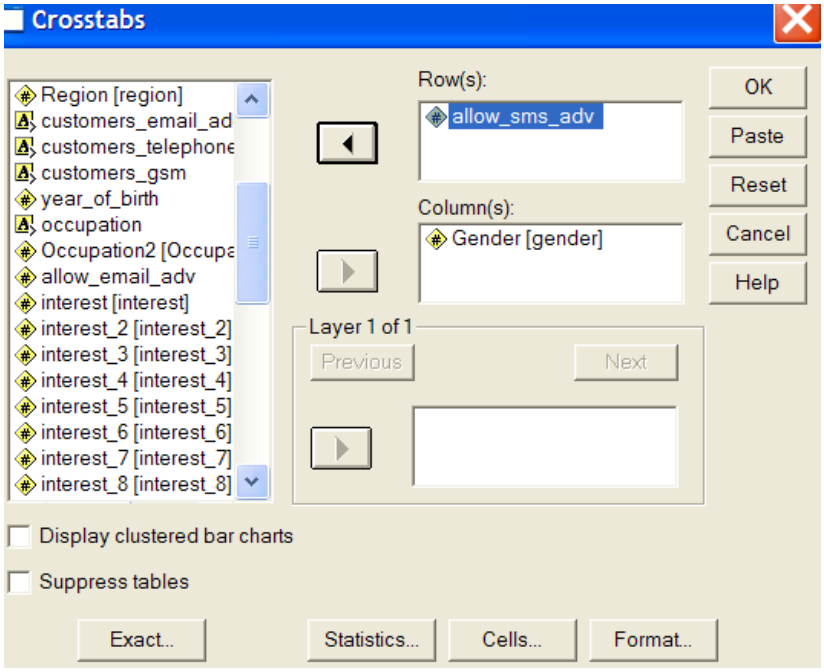
Esimerkkiin on valittu aineistosta ikä-muuttuja (age_new) joka on viisiluokkainen. Ensimmäinen tuloste on yhteenvetotaulukko, joka kuvaa tilastolliset tunnusluvut muuttujasta. Siitä nähdään että esimerkiksi yleisimmin esiintynyt (moodi) ikäryhmä on 3 eli ryhmä 35-49 (koodattu 1-5), heitä on kaikkiaan 73 kappaletta. Ylimmäisestä taulukosta nähdään myös että ky-

symykseen on vastannut 183 vastaajaa (27 on tyhjiä). Koko aineiston koko siis on 210 vastaajaa.

Alapuolella olevasta taulukosta nähdään kunkin luokan kohdalta vastauksien jakauma. Frekvenssitaulukosta pitää raportoida prosenttiluvut sarakkeesta **Valid Percent**, koska silloin mukana ovat vain kyseiseen kysymykseen vastanneet.

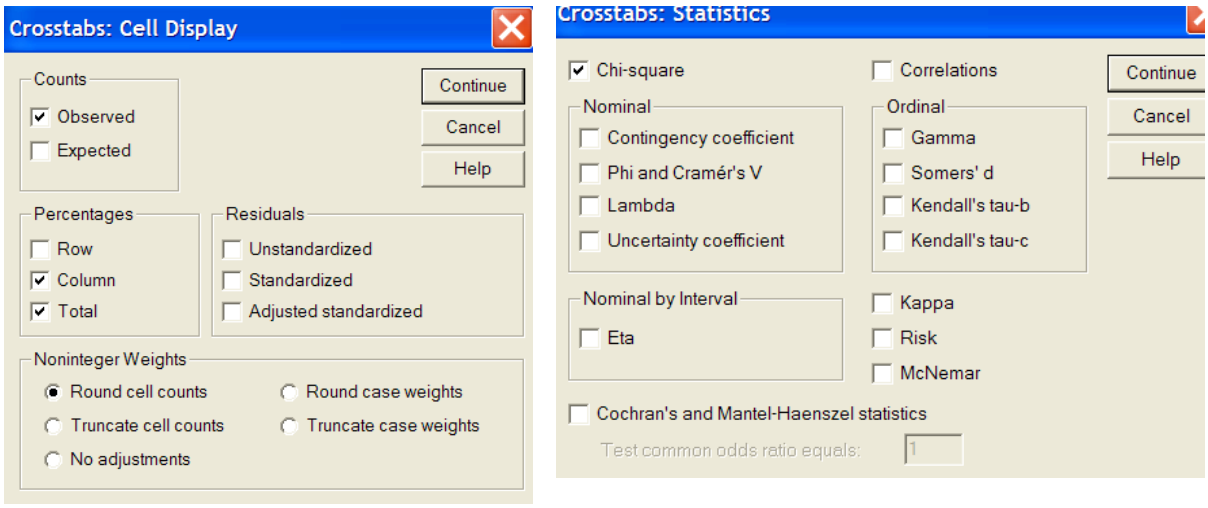
3.2. Ristiintaulukointi

Mikäli halutaan tutkia kahden nominaaliasteikollisen (tai toinen nominaali ja toinen järjestysasteikollinen) muuttujan välistä riippuvuutta käytetään ristiintaulukointia. Ristiintaulukointiin liittyvä nollahypoteesi väittää että muuttujat ovat toisten suhteensa riippumattomia eli ristiintaulukointia ja sen testejä voidaan käyttää hypoteesien testaukseen. Ristiintaulukointi suoritetaan valikosta **Analyze / Descriptive Statistics / Crosstabs**, jolloin avautuu *Crosstabs* –ikkuna (Kuva 8).



KUVA 8. Ristiintaulukoinnin aloitusikkuna

Ristiintaulukointi aloitetaan valitsemalla vasemmalla olevasta muuttujaluettelosta ne muuttujat joita halutaan tarkastella. Lähtökohtana on valita selittävä/riippuva muuttuja (=DEPENDENT) muuttuja, joka asetetaan rivimuuttujaksi (**ROW variable**). Tässä esimerkissä Row-soluun on valittu muuttuja ”allow_sms_adv”, joka kuvaa halukkuutta vastaanottaa SMS -markkinointia. Selittäväksi/riippumattomaksi (=INDEPENDENT) muuttujaksi valitaan muuttuja jolla riippuvaa muuttujaa aiotaan selittää eli tässä tapauksessa on valittu ”Gender” eli sukupuoli. Se siirretään Column(s) soluun[1]. Halutaan siis tutkia *onko sukupuolella vaikutusta halukkuuteen vastaanottaa SMS -markkinointia.* Ristiintaulukkoon tulostetaan näkyville sarakeprosentit, jotka valitaan *Cells* –painikkeen alta sekä monesti ”Total” prosentit (Kuva 9).



KUVA 9. Ristiintaulukoinnin välilehdet

Ristiintaulukosta lasketaan khin neliö -testin (Chi-square) arvo (*Statistics* painikkeen alta avautuu ikkuna Kuva 9). Khin neliö -testiä käytetään paljon jakaumien erojen testauksessa laadullisten (luokiteltujen) muuttujien tapauksessa.

3.2.1. Ristiintaulukoinnin tulosteiden tulkinta

Kun tarvittavat esimäärittelyt on tehty painetaan ok ja avautuu oheinen tulostusnäkymä (KUVA 10).

[1] Voi laittaa myös toisinpäin eli ROW / COLUMN jako selittävään / selitettävään ei ole yksiselitteinen

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
allow_sms_adv * Gender	198	94,3%	12	5,7%	210	100,0%

allow_sms_adv * Gender Crosstabulation

			Gender		
			Male	Female	Total
allow_sms_adv	0	Count	109	30	139
		% within Gender	71,7%	65,2%	70,2%
		% of Total	55,1%	15,2%	70,2%
	1	Count	43	16	59
		% within Gender	28,3%	34,8%	29,8%
		% of Total	21,7%	8,1%	29,8%
Total		Count	152	46	198
		% within Gender	100,0%	100,0%	100,0%
		% of Total	76,8%	23,2%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,712[b]	1	,399		
Continuity Correction[a]	,435	1	,509		
Likelihood Ratio	,698	1	,404		
Fisher's Exact Test				,462	,252
Linear-by-Linear Association	,708	1	,400		
N of Valid Cases	198				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 13.71.

KUVA 10. Ristiintaulukoinnin tuloste

Case Processing Summary taulukko kertoo, kuinka monen havainnon pohjalta ristiintaulukko on laskettu (Valid N ja Percent). Taulukosta ilmenee myös puuttuvien havaintojen määrä eli esimerkkitapauksessa puuttuu 12 havaintoa (5.7 %). **Crosstabulation** –taulukko on varsinainen ristiintaulukko, josta ilmenee havaintojen jakautuminen luokkiin. Kuvassa 10 on esitetty vastaajien jakautuminen sukupuolen (selittävä muuttuja) ja halukkuuteen vastaanottaa SMS -markkinointia (allow_sms_adv; selitettävä muuttuja) suhteen. Taulukkoa tutkimalla nähdään että tarkastelemalla esimerkiksi miesten kohdalta halukkuutta vastaanottaa SMS -

markkinointia että 71.7% miehistä kuuluu luokkaan 0 (eli koodauksessa 0=ei halua ja 1=haluaa), eli he eivät halua vastaanottaa. Vastaavasti naisista 65.2% kuuluu tähän ”ei halua” luokkaan. Eroa näyttäisi sukupuolten välillä olevan mutta vielä ei kyetä sitä tilastollisesti perustelemaan vaan täytyy katsoa khin neliö –testin tulosta, joka kertoo tilastollisen eron jos sitä siis on.

Khin neliö –testin arvo on pieni (.712) ja merkitsevyys (sig-arvo[2]) on > .05 joten voidaan todeta että sukupuolella ja halukkuudella vastaanottaa SMS -markkinointia ei ole riippuvuutta. Havaitun merkitsevyystason lukuarvo [Asymp.Sig. (2-sided)] kertoo, että on olemassa 39.9% riski sille, että hylätään tosi nollahypoteesi epätotena eli toisin sanoen on olemassa 39.9% riski sille, että muuttujien välinen riippuvuus johtuu sattumasta. Kun arvo siis on < .05 niin riski sille että ollaan ”väärässä” on alle 5%.

Chi-Square testitaulukosta pitää lisäksi katsoa viite viite b, joka kertoo täyttyvätkö testin edellytykset [1) max 20%:ia odotusarvoista < 5, ja 2) pienin odotusarvo > 1]. Tässä tapauksessa 0% odotusarvoista < 5 ja pieninkin odotusarvo on 13,71 (>1) eli testin edellytykset täyttyivät ja tehty testi on siis pätevä.

Khin neilö –testin arvo (*Value*) voi olla välillä 0 ja ääretön. Mitä suurempi sen arvo on sen vahvempi on muuttujien välinen riippuvuus.

Kontingenssikerroin mittaa kahden nominaaliasteikollisen muuttujan välistä riippuvuutta ja sitä voidaan myös käyttää riippuvuutta tutkittaessa. Mitä suurempi kerroin on lukuarvoltaan,

[2] Havaittu merkitsevyystaso eli ns. p-arvo kertoo tilastollisen merkitsevyyden. Käytetyt raja-arvot ovat
p<.001 (tilastollisesti erittäin merkitsevä)
p<.01 (tilastollisesti merkitsevä)
p<.05 (tilastollisesti melkein merkitsevä)

sitä voimakkaampaa on muuttujien välinen riippuvuus. Arvot vaihtelevat nollan ja ykkösen välillä (karkeat ohjearvot: C < 0.3 heikko riippuvuus; C > 0.6 voimakas riippuvuus). Valitaan rastittamalla **Contingency coefficient** *Statistics* -painikkeen alta.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	,060	,399
N of Valid Cases		198	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

KUVA 11. Kontingenssikerroin (nominal by nominal)

YHTEENVETO TULOKSESTA

Eli kuten todettiin ristiintaulukon ja Khin neliö –testin valossa pitää todeta että sukupuolella ja halukkuudella vastaanottaa SMS -markkinointia ei ole tämän aineiston valossa riippuvuutta (eli nollahypoteesi H_0 jää voimaan). Tässä yhteydessä on hyvä esittää seuraavat kolme perustelua:

1. Testisuureen arvo on pieni (Pearson Chi-Square 0.712)
2. Sig arvo on suuri (.399), kun pitäisi olla <.05
3. Kontingenssiarvo on pieni (.060)

3.3. Muuttujan jakauman testaus

Joissakin tilanteissa on tärkeätä selvittää muuttujan jakauma, esimerkiksi monet parametriset testit edellyttävät muuttujan jakauman noudattavan normaalijakaumaa. Yleisimmin käytetty normaalisuustesti on **Kolmogorov-Smirnovin (KS)** testi ,jota käytetään aineiston ollessa >

50. Alle 50 otoksen aineistoille suositellaan käytettäväksi **Shapiro Wilks** –testiä. KS –testissä aineiston jakaumaa siis verrataan johonkin ennalta tunnettuun jakaumaan kuten normaalijakaumaan.

3.3.1. Kolmogorov-Smirnovin testi

KS testi tehdään valitsemalla **Analyze / Nonparametric Tests / 1-Sample K-S**. SPSS -ohjelmistossa voidaan valita neljä mahdollista vertailujakaumaa (normaali, Poisson, Uniform, ja Exponential). Yleensä valitaan ”Normal”. Testaamme erään muuttujan normaalijakautuneisuutta (kuinka usein käytät tekstiviestipalvelua asteikko 1=en koskaan ja 5=päivittäin). Testituloste on seuraavanlainen:

One-Sample Kolmogorov-Smirnov Test

		Tekstiviestit
N		194
Normal Parameters[a,b]	Mean	4,38
	Std. Deviation	1,017
Most Extreme Differences	Absolute	,368
	Positive	,272
	Negative	-,368
Kolmogorov-Smirnov Z		5,120
Asymp. Sig. (2-tailed)		,000

a. Test distribution is Normal.

b. Calculated from data.

KUVA 12. Normaalijakautuneisuuden testaus

Tuloksissa kerrotaan muuttujan keskiarvo (4.38) ja keskihajonta (1.017). Näiden lukujen avulla K-S testi muodostaa normaalijakauman kumulatiivisen todennäköisyysjakauman, johon aineistomme jakaumaa verrataan. Taulukosta nähdään että merkitsevyys (Asymp.Sig) hylkää nollahypoteesin (p = .000) jakaumamme normaalisuudesta. Eli K-S testin kohdalla merkitsevyysarvon ollessa pieni se on **HUONO UUTINEN** jakauman normaalisuuden kan-

nalta. Tulos siis on että kun nollahypoteesi (jakauma noudattaa normaalijakaumaa) hylätään jakauma ei noudata normaalijakaumaa. On tosin todettava että hyvin harvoin asennetyyppiset muuttujat jotka saavat arvoja 1-7 noudattavat normaalijakaumaa, joten normaaliuden tarkistus jää sitten tutkijoitten harkittavaksi eli kannattaako sitä ylipäätään testata.

3.3.2. Shapiro-Wilk -testi

Shapiro Wilk –testiä siis käytetään kun aineisto on pieni (<50). Se suoritetaan komennolla: **Analyze / Descriptive Statistics / Explore / Plots / Normality plots with tests** (rastitetaan). Muuttuja jota halutaan tarkastella syötetään oikeaan ylimmäiseen soluun ja painetaan ok. Tulostetta tulee aika paljon ja oheisesta tulosteen tehdään tulkinta:

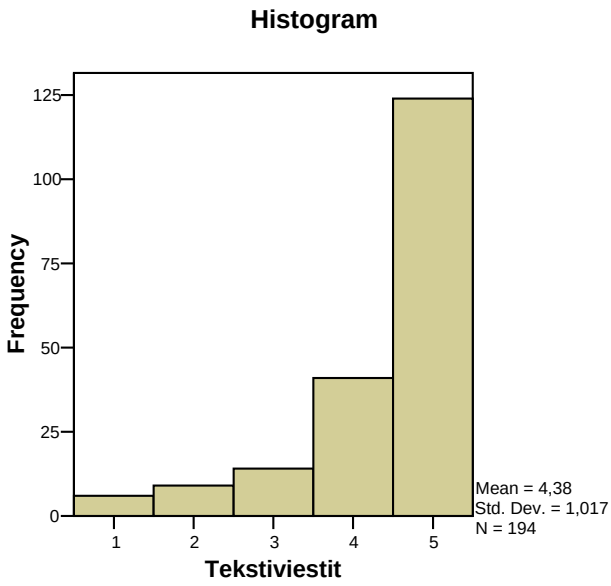
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Tekstiviestit	,368	194	,000	,656	194	,000

a. Lilliefors Significance Correction

KUVA 13. Shapiro-Wilk testisuure

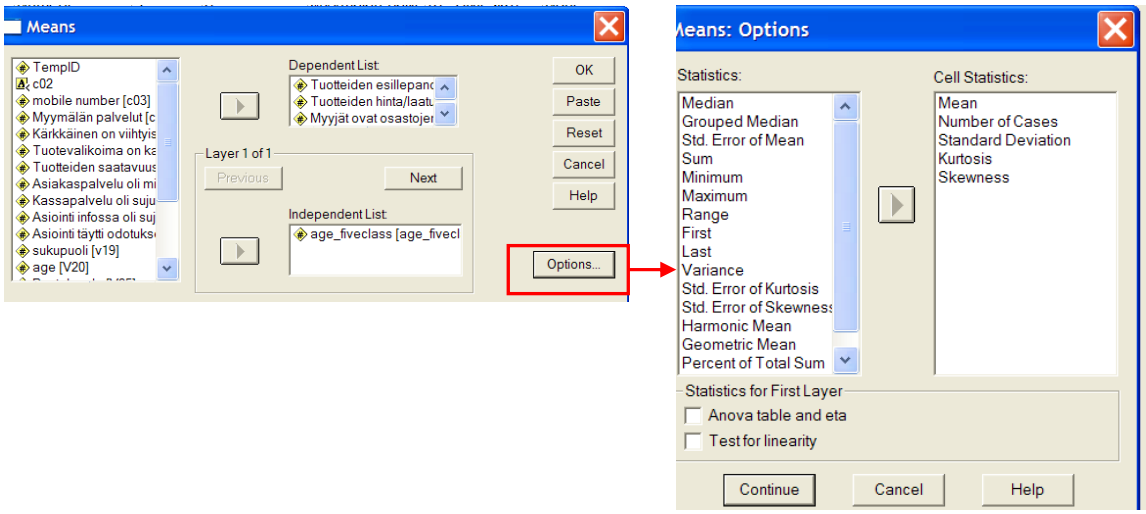
Shapiro-Wilkin testisuure saa arvon 0,656 (p = .000) eli tulos tulkitaan samalla tavalla kuin Kolmogorov-Smirnovinkin testissä eli aineiston jakauma ei noudata normaalijakaumaa. Tämä ei ole mikään ihme kun katsotaan miten tarkastelussa ollut muuttuja on jakautunut (Kuva 14). Kuvasta näkyy selvästi että jakauma on vahvasti painottunut oikealle (eli arvon 5 saavia vastauksia on eniten).



KUVA 14. Esimerkki jakaumasta joka ei noudata normaalijakaumaa

3.3.3. Jakauman normaalisuuden testaus varianssianalyysien yhteydessä

Jakauman normaalisuutta voidaan tutkia varianssianalyysien yhteydessä vinous (*Skewness*)- ja huipukkuuskertoimien (*Kurtosis*) avulla. Komennon **Analyze / Compare Means / Means** kautta avautuvassa ikkunassa valitaan *Options* –painikkeen alla olevasta *Statistics* –listasta *Cell Statistics* –kenttään vaihtoehdot *Skewness* (vinous) ja *Kurtosis* (huipukkuus):



KUVA 15. Vinous- ja huipukkuuskertoimet ANOVA:ssa

Oheisessa esimerkissä tutkitaan ikäryhmittäin (*Independent List* –muuttujaksi on valittu ikä viisiluokkaisena muuttujana) asiakastyytyväisyyttä J.Kärkkäisen tavaratalon palveluihin (tyytyväisyyskysymykset *Dependent List* –kohtaan). Aineisto on kohtuullisen suuri (n=2050), joten myös kysymyksien jakaumat noudattavat kohtuullisen hyvin normaalijakaumaa kuten tulosteesta nähdään:

Report

age_fiveclass		Tuotteiden esillepanoon on kiinnitetty huomiota	Tuotteiden hinta/laatu suhde on kohdallaan	Myyjät ovat osastojensa asiantuntijoita	Sain myyjältä apua tarvittaessa	Liikenneyhteydet toimivat hyvin
18 and below	Mean	3,77	3,80	3,81	3,67	3,86
	N	159	159	159	159	159
	Std. Deviation	,969	,855	1,022	1,133	1,016
	Kurtosis	-,365	-,463	-,271	-,354	,509
	Skewness	-,447	-,337	-,609	-,595	-,820
19-29	Mean	3,57	3,75	3,53	3,37	3,57
	N	476	475	476	476	475
	Std. Deviation	,921	,877	,914	1,157	1,064
	Kurtosis	-,256	,118	,052	-,704	-,406
	Skewness	-,304	-,455	-,412	-,281	-,454
30-49	Mean	3,67	3,60	3,63	3,37	3,60
	N	803	803	803	804	799
	Std. Deviation	,906	,867	,876	1,166	1,085
	Kurtosis	-,251	-,099	,002	-,851	-,558
	Skewness	-,314	-,255	-,327	-,192	-,463
50-64	Mean	3,94	3,67	3,84	3,57	3,90
	N	488	487	488	485	483
	Std. Deviation	,835	,825	,912	1,116	1,006
	Kurtosis	,073	-,262	,138	-,587	,122
	Skewness	-,512	-,190	-,581	-,416	-,777
65 and over	Mean	3,95	3,91	4,12	3,91	3,99
	N	120	119	119	119	117
	Std. Deviation	,887	,902	,894	,965	1,004
	Kurtosis	,147	-,117	,268	-,471	,078
	Skewness	-,635	-,520	-,814	-,502	-,814
Total	Mean	3,73	3,69	3,70	3,47	3,71
	N	2046	2043	2045	2043	2033
	Std. Deviation	,908	,864	,919	1,148	1,063
	Kurtosis	-,210	-,142	-,054	-,723	-,330
	Skewness	-,393	-,303	-,431	-,322	-,574

KUVA 16. Vinous ja huipukkuuskertoimien tarkastelu

Täydellisesti normaalijakaumaan noudattavan muuttujan vinous- ja huipukkuuskertoimien arvot = 0. Oikealle vinon jakauman vinouskerroin on positiivinen (eli jakauma, jossa valtaosa havainnoista saa keskiarvoja pienempiä arvoja). Vasemmalle vinon jakauman vinouskerroin on negatiivinen. **Jos vinouskerroin (*Skewness)* on > 1**, on kyseessä siinä määrin normaalista poikkeava jakauma että se kannattaa jättää pois varianssianalyysistä. Mitä suuremman arvon huipukkuuskerroin (*Kurtosis*) saa, sitä korkeammasta jakaumasta on kyse. Negatiivinen huipukkuuskerroin viittaa joko normaalia laakeampaan jakaumaan tai jakaumaan, jolla on useita

huippuja. Yllä olevasta esimerkistä nähdään että kaikki vinouskertoimen arvot < -1 tai 1, joten kaikki muuttujat voidaan ottaa huoletta mukaan varianssianalyysiin (ks. kpl 3.4.4.)

Varianssianalyysin yhteydessä esitetään myöhemmin myös toinen reunaehto, jonka tulee täyttyä että varianssianalyysin suorittamiseen on edellytykset. Kyseessä on muuttujien varianssien yhtäsuuruus jokaisessa selittävien muuttujien muodostamassa havaintoluokassa eli yllä olevassa esimerkissä pitää tarkastella ikäluokittain keskihajontaa (varianssi=keskihajonta^2, joten välttämättä ei tarvitse tulostaa erikseen varianssien arvoja, joskin se on helppoa ottamalla se analyysiin mukaan kuvan X taulukosta). Esimerkistämme näemme että ryhmäkohtaisten varianssien yhtäsuuruus toteutuu kaikkien muuttujien kohdalla (keskihajonnan vaihtelu pientä 0.1-0.2).

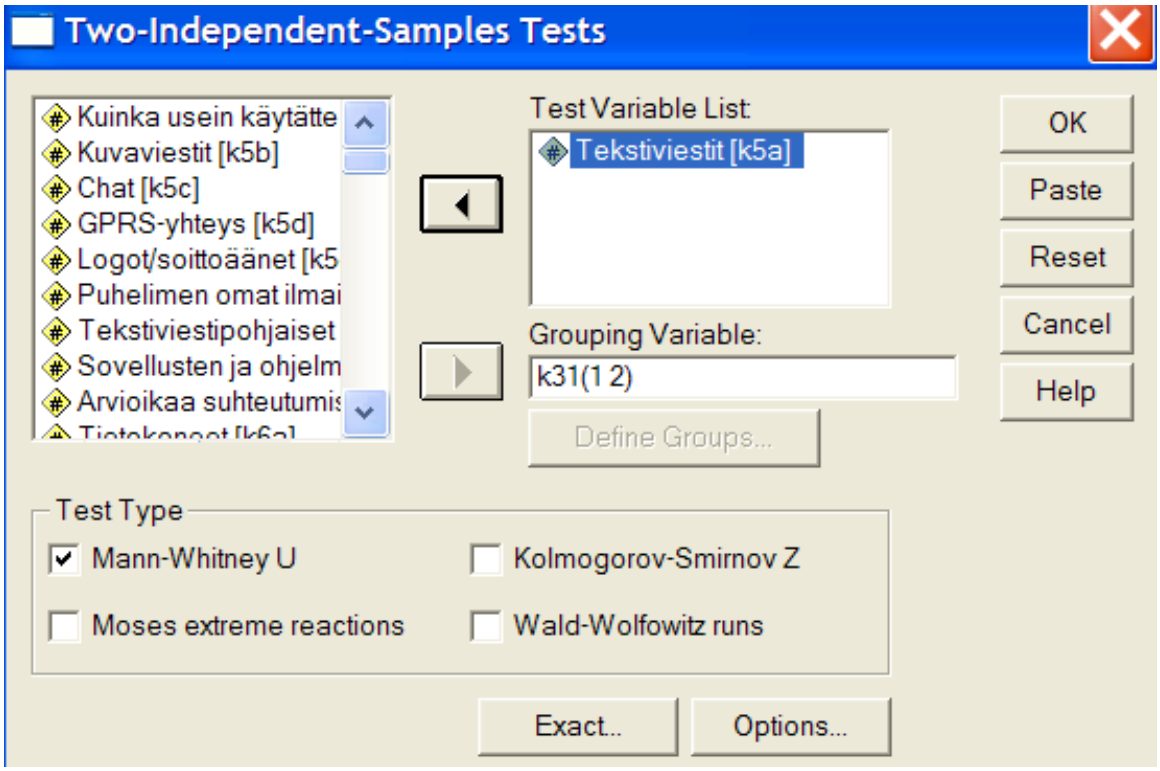
3.4. Keskiarvovertailut

Kun halutaan verrata kahden riippumattoman otoksen keskiarvoja on valittavana joko yleisesti käytetty **t-testi** tai vähemmän käytetty **Mann-Whitney** –testi. Pienille otoskooille soveltuva testi on Mann-Whitney kun taas suurille ja parametrisesti mitatuille käytetään t-testiä. On huomattava että AINA KUN EPÄILLÄÄN T-TESTIN EDELLYTYSTEN OLEMASSA OLOA (ts. vähintään välimatka-asteikollinen mittaus ja oletus muuttujan normaalijakautuneisuudesta) TULISI KÄYTTÄÄ MANN-WHITNEY –U-TESTIÄ. Kuitenkaan esim. sosiaalitieteissä ei käytetä juuri koskaan mitään muuta kuin t-testiä, mikä on ehkä myös virhe (ainakin tilastotieteilijöiden näkökulmasta). Eli on syytä käydä molemmat testit läpi.

3.4.1. Mann-Whitney U-testi

Mann-Whitneyn U-testi on lähes t-testin veroinen testi mittaamaan muuttujan mediaaneissa eli painopisteessä olevaa eroa. Ne soveltuvat tietenkin myös tilanteisiin joissa halutaan verrata

kahden riippumattoman otoksen keskiarvoja toisiinsa. Testi suoritetaan SPSS-ohjelmistossa komennolla **Analyze / Nonparametric Tests / 2 Independent Samples**, jonka jälkeen avautuu seuraavanlainen ikkuna:



KUVA 17. Mann Whitney U-testi

Ikkunaan valitaan kohtaan ”Grouping Variable” ryhmittelymuuttuja, tässä esimerkissä sukupuoli (k31) ja testimuuttujaksi Tekstiviestien käyttö (1-5 asteikollinen muuttuja). Sukupuolimuuttujalle pitää antaa ”Define Groups” painikkeen alta avautuvaan laatikkoon arvot joita muuttuja saa. Tässä tapauksessa arvot ovat 1 ja 2 (eli muuttuja on koodattu 1=mies ja 2=nainen). Kun nämä toimenpiteet on tehty painetaan OK ja avautuu seuraavanlainen tuloste:

Ranks

	Sukupuoli	N	Mean Rank	Sum of Ranks
Tekstiviestit	Mies	124	91,24	11314,00
	Nainen	66	103,50	6831,00
	Total	190		

Test Statistics[a]

	Tekstiviestit
Mann-Whitney U	3564,000
Wilcoxon W	11314,000
Z	-1,710
Asymp. Sig. (2-tailed)	,087

a. Grouping Variable: Sukupuoli

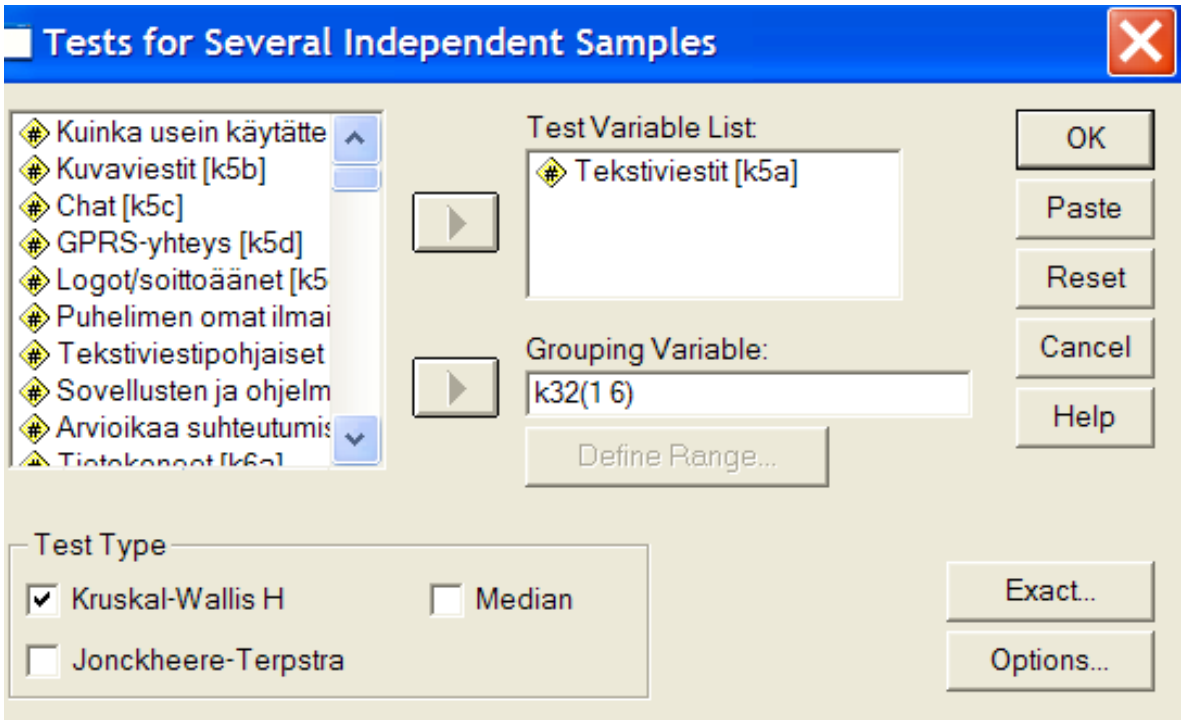
KUVA 18. Mann Whitney tuloste

Vasemmalla oleva taulu kuvaa muuttujien otoskoot sekä järjestyssummat. Varsinainen tarkasteltava taulukko on oikealla ja siitä luetaan testin tulokset. Mann-Whitneyn U-testisuureen arvo on U=3565,000 ja vastaava Wilcoxonin testisuure W=11314,500. Asymp.sig merkitsevyys on .087 eli tuloksena on että ryhmien keskiarvojen välillä ei ole eroja ($p > .05$). Toisin sanoen tulos on että sukupuolella ei tämän testin valossa ole vaikutusta tekstiviestien käyttöön, joskin olemme lähellä merkitsevyystasoa $p < .05$). Hieman myöhemmin testaamme samaa asiaa t-testin avulla niin pystymme myös vertailemaan testien yhdenpitävyyttä.

3.4.2. Kruskall-Wallisin testi

Jos halutaan verrata useamman kuin kahden ryhmän välistä keskiarvoa toisiinsa, tehdään se yleensä varianssianalyysillä (ANOVA) ja F-testillä. Näillä testeillä on kuitenkin samantyyppisiä perusolettamuksia kuin t-testillä joten suositeltava vaihtoehto näille on ei-parametrinen K-W –testi. Useamman kuin kahden keskiarvon testiä voidaan käyttää jos halutaan verrata ryhmien kuten opiskelijat-työntekijät-johtajat tai ei käyttäjät – vähän käyttävät –paljon käyttävät tai alle 18 vuotiaat-19-45 vuotiaat-yli 45 vuotiaat, keskiarvoja toisiinsa jonkin muuttujan suhteen.

KW-testi on itse asiassa suoraan yksisuuntaisen varianssianalyysin eli F-testin parametriton vastine. KW-testin yksi pääolettamus on että muuttuja on mitattu vähintään järjestysasteikolla ja näin usein onkin kun käsitellään esim. asennemuuttujia. Seuraavassa esimerkissä halutaan selvittää tekstiviestin käyttöä eri ikäryhmissä eli verrataan eri ikäluokkien valossa tekstiviestien käyttöä. Ikämuuttuja luokittelee havainnot kuuteen ryhmään ja tekstiviestien käyttö –muuttuja kuvaa käytön frekvenssiä. KW –testi suoritetaan valikosta **Analyze / Nonparametric Tests / K Independent Samples**, jonka jälkeen avautuu oheinen ikkuna:



KUVA 19. Kruskall-Wallis'in -testin aloitusikkuna

Ennen kuin painetaan OK pitää "Grouping Variable" painikkeen alta avautuvaan kenttään syöttää muuttujan arvot (eli tässä esimerkissä ikä muuttuja on kuusiluokkainen ja saanut arvot 1=alle 18v. – 6=65 tai yli), eli syötetään arvot 1 ja 6. Testin tulos on seuraavan näköinen:

Ranks

	Ikä	N	Mean Rank
Tekstiviestit	Alle 18	6	103,17
	18-24	73	105,78
	25-34	77	93,79
	35-49	26	85,42
	50-64	6	57,75
	Yli 65	2	7,25
	Total	190	

Test Statistics[a,b]

	Tekstiviestit
Chi-Square	15,840
df	5
Asymp. Sig.	,007

a. Kruskal Wallis Test

b. Grouping Variable: Ikä

KUVA 20. Kruskall Wallis tulosteen tulkinta

Vasemmalla olevasta taulukosta nähdään ryhmien otoskoot (*Ranks*) ja järjestyslukusummien keskiarvot (*Mean Ranks*). Varsinainen testitulos on seuraavassa taulussa (oikeanpuoleinen). Testisuureen merkitsevyys $p = .007$, joka kertoo sen että ryhmien välillä on tilastollisesti mer-

kitsevä ero ($p < .01$). Tulos ei kylläkään kerro eroavatko kaikki ryhmät toisistaan vai onko vain ehkä yksi ryhmä muista poikkeava. SPSS ei tarjoa mahdollisuutta KW-testille tehdä tähän liittyen jatkotarkasteluita vaan on käytettävä muita menetelmiä. Kruskall-Wallisin testi tehdään usein joukolle muuttujia ja tieteellisessä työssä tuloste voi olla esimerkiksi seuraavanlainen taulukko:

	Frequency of Internet banking						
	Never		Rarely		Often		
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	p-value*
Mobile data services	4.1205	2.10333	5.1356	1.53643	5.1781	1.51903	.000
PCs	3.9450	2.06930	5.8667	1.15666	6.1209	1.03903	.000
Debit cards, charge cards and credit cards	4.5779	2.02400	5.6167	1.15115	5.7505	1.24110	.000
ATMs	4.0983	2.28714	5.0714	1.95269	4.0830	1.86075	.002
e-mail	3.6256	2.12050	6.0000	1.26223	6.2269	1.06188	.000
Internet	3.2488	2.03040	5.8305	1.27513	6.2942	.92491	.000
Personal service	6.3281	1.12512	5.7797	1.24662	5.5165	1.43376	.000
Text-TV	5.1463	1.94857	5.3158	1.59416	4.8928	1.69737	.002
Self-service machines (e.g. stamp machines, train ticket machines etc.)	3.9386	2.11550	4.9322	1.58511	4.5465	1.65106	.000
Electronic ID cards	3.4880	2.12428	5.0769	1.38403	4.8821	1.64548	.000

p-values of Kruskal-Wallis test for equality of means. H0: Means are equal. H0 is rejected if p<0.05

KUVA 21. Esimerkki Kruskall-Wallisin testin raportoinnista[3]

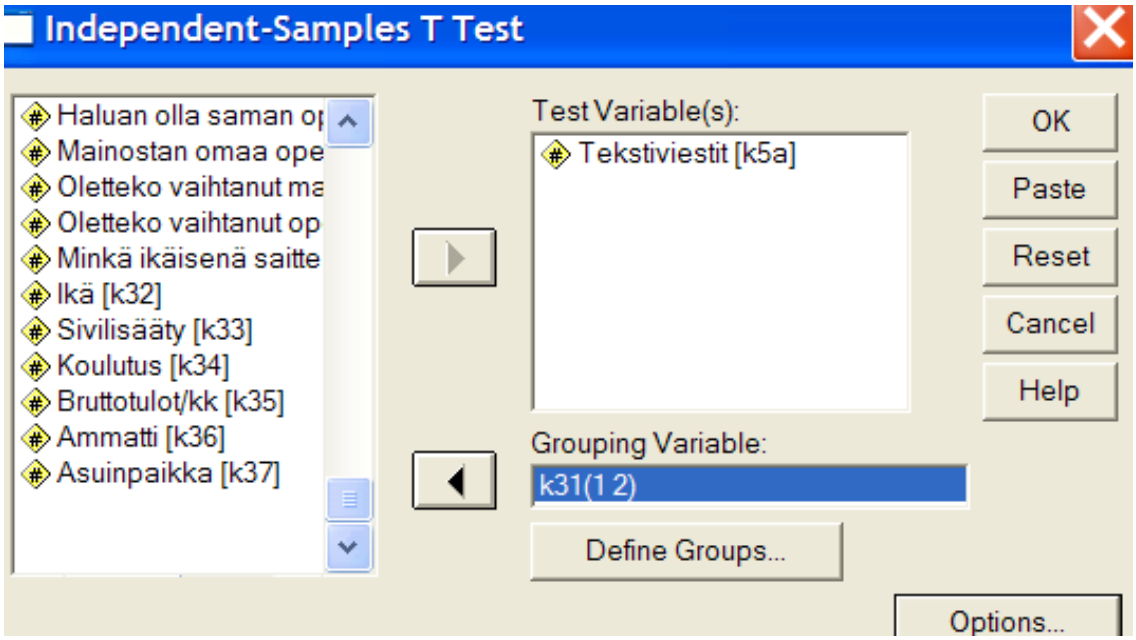
Taulukossa on vertailtu kolmen verkkopankkikäyttäjäryhmän (never-rarely-often) suhtautumista erilaisiin teknologioihin. Kustakin muuttujasta on raportoitu keskiarvo, keskihajonta sekä aivan oikeassa reunassa KW-testistä saatu merkitsevyystaso, joka siis kertoo ryhmien välisen keskiarvon eron. Kyseisessä esimerkissä kaikki muuttujat ovat saaneet merkitsevän arvon eli todetaan että ryhmien välillä on eroja suhtautumisessa teknologioihin.

3.4.3. T-testi

T-testi on yleisin kahden toisistaan riippumattoman ryhmän keskiarvojen vertailuun. T-testi testaa ensin ovatko varianssit yhtä suuret ja ilmoittaa sen jälkeen tulokset sekä yhtä suurten

[3] Karjaluoto, H., Koivumäki, T., and Salo, J. (2003), "Individual differences in private banking: Empirical evidence from Finland", Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS-36), p. 196, Big Island, Hawaii, January 6-9, 2003.

että erisuurten varianssien tapauksessa. Niistä tulee käyttäjän valita tilanteeseen sopiva. T-testin edellytyksenä on että muuttuja on normaalisti jakautunut ja vähintään välimatka-asteikollinen, joskin näistä olettamuksista usein poiketaan, ainakin sosiaalitieteissä. T-testi suoritetaan valitsemalla **Analyze / Compare Means / Independent Samples T-Test**, jonka jälkeen avautuu oheinen ikkuna:



KUVA 22. T-testin aloitusikkuna

Grouping Variable –kohtaan valitaan muuttuja, jonka perusteella vertailtavat ryhmät muodostetaan. Tässä tapauksessa käytetään samaa esimerkkiä kuin Mann-Whitney –testin kanssa eli valitaan sukupuoli (k31) ja sille *Define Groups* –kohdasta annetaan arvot 1 ja 2. *Test Variable* –kohtaan valitaan tekstiviestin käyttö (asteikko 1-5). *Grouping Variable* –kohtaan voidaan myös valita muu kuin kaksiluokkainen muuttuja, jos luokkia on esim. 6 kuten ikämuuttujamme kohdalla voidaan valita ryhmiksi 1 (alle 18v.) ja vaikkapa 5 (46-64v.). Testituloste on seuraavanlainen:

Group Statistics

	Sukupuoli	N	Mean	Std. Deviation	Std. Error Mean
Tekstiviestit	Mies	124	4,27	1,107	,099
	Nainen	66	4,56	,825	,102

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Tekstiviestit	Equal variances assumed	6,159	,014	-1,846	188	,066	-,286	,155	-,592	,020
	Equal variances not assumed			-2,015	167,786	,045	-,286	,142	-,567	-,006

KUVA 23. T-testin tuloste

Ylimmäisestä taulukosta nähdään yhteenveto vertailtavista ryhmistä eli ryhmäkohtaiset vastausmäärät (N), keskiarvot (Mean), keskihajonnat (Std. Deviation) sekä keskivirhe (Std. Error of Mean). Varsinainen t-testin tulosten tulkinta tehdään alimmaisesta taulukosta. Tulosteesta pitää ensin tulkita varianssien yhtäsuuruustestin tulos (Levenen testisuureen arvo). Levenen testi määrää sen kumpaa riviä t-testin tuloksista tulkitaan.

Levenen testin nollahypoteesi väittää että varianssit ovat yhtäsuuret molemmissa ryhmissä. Tässä esimerkissä voidaan todeta varianssien olevan erisuuret (p = .014) eli merkitsevyystaso p < .05 eli t-testin arvo luetaan riviltä ”**Equal Variances Not Assumed**”. Mikäli Levenen testin Sig-arvo olisi ollut > .05 niin t-testin tulos luettaisiin ylemmältä riviltä ”Equal variances assumed”.

Taulukosta raportoidaan Sig-arvo, joka kyseisessä esimerkissä on p = .045 (eli p < .05) eli sen avulla voidaan varovaisesti todeta että sukupuolella on jonkin verran vaikutusta tekstiviestien käyttöön, joskin tulos ei ole tilastollisesti kovin merkitsevä. Yhteenvetona tuloksia raportoitaessa kannattaa sanoa esim. näin: Naisten ja miesten keskiarvoja tarkasteltaessa (ks. ylempi taulukko) voidaan todeta että naiset käyttävät hieman enemmän tekstiviestipalvelua (n=66,

keskiarvo 4.56) kuin miehet (n=124, keskiarvo=4.27). Ero sukupuolen välillä on tilastollisesti lähellä merkitsevää tasoa (p = .045).

3.4.4. Varianssianalyysi

Varianssianalyysillä tutkitaan yhden tai useamman selitettävän muuttujan riippuvuutta yhdestä tai useammasta selitettävästä muuttujasta. Varianssianalyysi ei nimestään huolimatta testaa ryhmien varianssien välistä eroa, vaan sillä testataan keskiarvojen välisiä eroja joten se kuuluu keskiarvotesteihin. Varianssianalyysi sopii tilanteisiin jossa selitettävä(t) muuttuja(t) on mitattu vähintään intervalliasteikolla, ja selittävä(t) ovat nominaaliasteikollisia tai niitä käytetään nominaaliasteikollisten muuttujien tapaan (ts. muuttujat on luokiteltu). Analyysin lähtöoletuksena eli nollahypoteesina on että kiinnostuksen kohteena olevien luokkien ***keskiarvot ovat yhtä suuret***. Jos varianssianalyysin tuloksena nollahypoteesi voidaan hylätä, selitettävän muuttujan keskiarvojen välillä on eroja selittävän muuttujan eri luokissa.

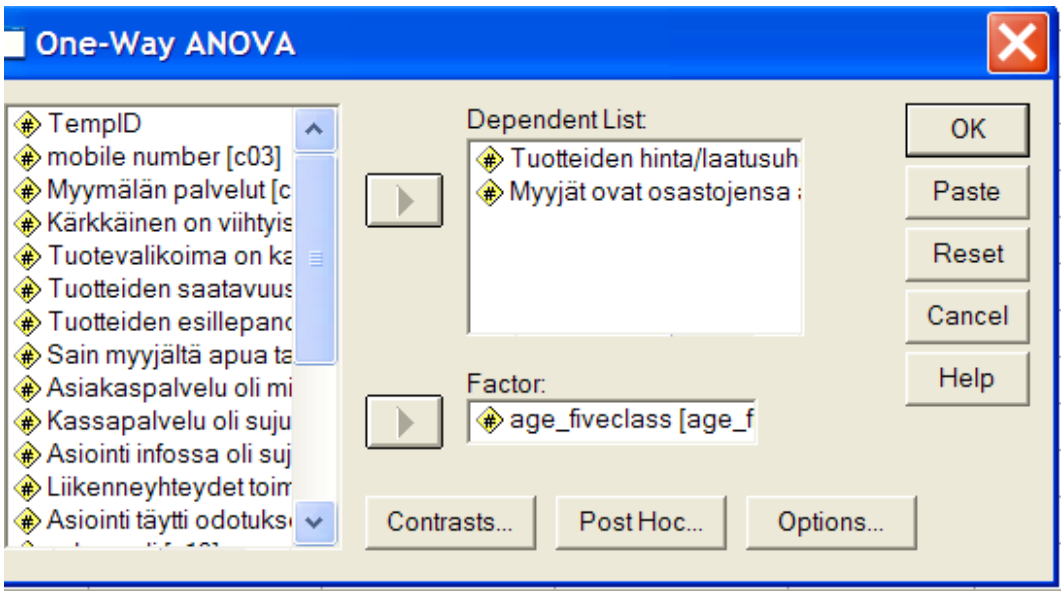
Varianssianalyysin edellyttää että (ks. kpl 3.3.3):

1) muuttujan arvojen tulee olla normaalisti jakautuneita kaikilla vertailtavilla ryhmillä
2) muuttujien varianssien eri ryhmissä tulee olla yhtäsuuret
3) ryhmien tulee olla normaalisti jakautuneita

Jos tilastoyksiköt jaetaan ryhmiin yhden muuttujan perusteella (eli selittävät muuttujat esim. sukupuoli, ikä, osasto, siviilisääty) ja verrataan näiden ryhmien keskiarvoja (esim. selitettävä muuttuja käyttö, asenne, ostovoima tms.), on kyseessä **yksisuuntainen varianssianalyysi** (***One-Way ANOVA***). Mikäli selittäviä muuttujia on useita, käytetään **monisuuntaista va-**

rianssianalyysia (**Multivariate *MANOVA***) ja **kovarianssianalyysia**. Seuraavassa keskitytään yksisuuntaisen varianssianalyysin tekemiseen.

Varianssianalyysi voidaan tehdä kahta eri kautta SPSS -ohjelmassa. Se voidaan tehdä osana t-testiä valitsemalla **Analyze / Compare Means / Means**, ja avautuvasta *Options* –valikosta valitsemalla testiin mukaan *Anova table and eta* (rastitetaan) eli varianssianalyysi. Toinen vaihtoehto on valita komento **Analyze / Compare Means / One-Way Anova**. Mikäli käytetään jälkimmäistä avautuva ikkuna on seuraavanlainen:



Kuva 24. Yksisuuntaisen varianssitestin aloitusikkuna (One-Way ANOVA)

ANOVA –testin perusnäkymä ilman *Post hoc* –testejä on seuraavanlainen:

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Tuotteiden hinta/laatusuhde on kohdallaan	Between Groups	15,545	4	3,886	5,247	,000
	Within Groups	1509,592	2038	,741		
	Total	1525,137	2042			
Myyjät ovat osastojensa asiantuntijoita	Between Groups	50,430	4	12,608	15,342	,000
	Within Groups	1676,414	2040	,822		
	Total	1726,844	2044			

KUVA 25. Yksisuuntainen varianssianalyysi tuloste

Käytännössä varianssianalyysi perustuu siihen, että selitettävän muuttujan varianssi jaetaan kahteen osaan. Näistä ensimmäinen mittaa luokkien sisäistä hajontaa (Kuvassa *Within Groups*) ja toinen luokkakeskiarvojen välistä hajontaa (Kuvassa *Between Groups*). Jos nämä kaksi varianssia eivät eroa kovinkaan paljon toisistaan, on todennäköistä, että eri luokkien saamat keskiarvot ovat peräisin samankaltaisesta jakaumasta. Tällöin niiden välillä ei ole tilastollisesti merkitsevää eroa. Jos taas nämä kaksi varianssia eroavat toisistaan tarpeeksi nollahypoteesi voidaan hylätä. Tilastollisena testinä varianssianalyysissa käytetään ns. **F-testiä**, joka kertoo millä todennäköisyydellä nollahypoteesi ryhmäkeskiarvojen yhtäläisyydestä voidaan hylätä.

Tulosteessa (Kuva 23) pitää siis kiinnittää huomiota F-merkitsevyystestin (Sig.) lisäksi F-arvoon (Kuvassa 5,247 ja 15,342)). Mikäli F-arvo on huomattavasti suurempi kuin 1, selitettävän muuttujan keskiarvot vaihtelevat selittävän/selitettävien muuttujan/muuttujien luokkien välillä enemmän kuin luokkien sisällä, jolloin nollahypoteesi keskiarvojen yhtäsuuruudesta selittävien muuttujien luokassa voidaan hylätä. Tulosteessa on tutkittu ikäryhmittäin (6-luokkaa) asiakastyytyväisyyttä (5-luokkainen muuttuja). Tulosteen perusteella nollahypoteesi hylätään molempien muuttujien kohdalla ja todetaan että ikäryhmien välillä on merkittäviä eroja asiakastyytyväisyydessä. Tähän on kaksi perustelua: 1) F-arvo on molempien muuttujien kohdalla suuri ($F > 1$) ja merkitsevyystestien (Sig.) arvot ovat nolla eli erittäin merkitseviä ($p = .000$ molemmissa).

Mikäli ANOVA tehdään komennolla **Analyze / Compare Means / Means**, ja avautuvasta *Options* –valikosta valitsemalla testiin mukaan *Anova table and eta*, saadaan tulosteeksi edellisen kuvan lisäksi tulostettua seuraavat taulukot:

Report

age_fiveclass		Tuotteiden hinta/laatu suhde on kohdallaan	Myyjät ovat osastojensa asiantuntijoita
18 and below	Mean	3,80	3,81
	N	159	159
	Std. Deviation	,855	1,022
19-29	Mean	3,75	3,53
	N	475	476
	Std. Deviation	,877	,914
30-49	Mean	3,60	3,63
	N	803	803
	Std. Deviation	,867	,876
50-64	Mean	3,67	3,84
	N	487	488
	Std. Deviation	,825	,912
65 and over	Mean	3,91	4,12
	N	119	119
	Std. Deviation	,902	,894
Total	Mean	3,69	3,70
	N	2043	2045
	Std. Deviation	,864	,919

Measures of Association

	Eta	Eta Squared
Tuotteiden hinta/laatusuhde on kohdallaan * age_ fiveclass	,101	,010
Myyjät ovat osastojensa asiantuntijoita * age_ fiveclass	,171	,029

KUVA 26. ANOVA taulukko ja Eta

Vasemmalla olevalla taulukko on jo sinällään käyttökelpoinen esimerkiksi tutkimusraportissa kun siinä näkyy mm. keskiarvot kunkin ikäluokan kohdalta. Oikeanpuoleinen tuloste kertoo **Eta** –kertoimen, joka muodostuu ryhmien välisen vaihtelun ja kokonaisvaihtelun suhteesta. Etan neliö on arvo joka kannattaa raportoida, koska se kuvaa kuinka paljon selitettävän muuttujan vaihtelusta pystytään selittämään selittävän muuttujan avulla. Eta^2 on tunnuslukuna verrattavissa regressioanalyysin yhteydessä käytettävään R^2-lukuun. Se voi saada arvoja nollan ja yhden väliltä ja suuret arvot kuvastavat selittävän muuttujan parempaa selitysvoimaa. Esimerkissä (Kuva 25) eta^2–luku saa arvot 0,010 ja 0,029, jotka molemmat ovat suhteellisen pieniä lukuja. Luku voidaan tulkita niin, että ikäluokkiin jakautumista kuvaavan muuttujan avulla voida selittää 1% ja 2,9% vastaajien asiakastyytyväisyyttä mittaavien muuttujien vaihtelusta. Toisin sanoen tässä esimerkissä Etan neliö on hyvin lähellä nollaa, joten kovin suuria eroja ei eri ikäryhmien välillä näyttäisi olevan.

HUOM! One-Way ANOVA ei kerro vielä minkä ryhmien välillä eroja on. Selvittämiseksi pitää tehdä Post Hoc –testi, joka tehdään ANOVA aloitusikkunasta (Kuva 23) *Post Hoc* –laatikosta. Käytettyjä testejä ovat mm. LSD, Scheffe, ja Tukey.

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) age_fiveclass	(J) age_fiveclass	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tuotteiden hinta/laatusuhde on kohdallaan	18 and below	19-29	,047	,079	,975	-,17	,26
		30-49	,196	,075	,066	-,01	,40
		50-64	,125	,079	,502	-,09	,34
		65 and over	-,109	,104	,835	-,39	,18
	19-29	18 and below	-,047	,079	,975	-,26	,17
		30-49	,149*	,050	,024	,01	,28
		50-64	,078	,056	,623	-,07	,23
		65 and over	-,156	,088	,393	-,40	,08
	30-49	18 and below	-,196	,075	,066	-,40	,01
		19-29	-,149*	,050	,024	-,28	-,01
		50-64	-,071	,049	,607	-,21	,06
		65 and over	-,305*	,085	,003	-,54	-,07
	50-64	18 and below	-,125	,079	,502	-,34	,09
		19-29	-,078	,056	,623	-,23	,07
		30-49	,071	,049	,607	-,06	,21
		65 and over	-,234	,088	,061	-,47	,01
	65 and over	18 and below	,109	,104	,835	-,18	,39
		19-29	,156	,088	,393	-,08	,40
		30-49	,305*	,085	,003	,07	,54
		50-64	,234	,088	,061	-,01	,47
Myyjät ovat osastojensa asiantuntijoita	18 and below	19-29	,276*	,083	,008	,05	,50
		30-49	,177	,079	,160	-,04	,39
		50-64	-,037	,083	,992	-,26	,19
		65 and over	-,313*	,110	,036	-,61	-,01
	19-29	18 and below	-,276*	,083	,008	-,50	-,05
		30-49	-,098	,052	,332	-,24	,04
		50-64	-,313*	,058	,000	-,47	-,15
		65 and over	-,588*	,093	,000	-,84	-,33
	30-49	18 and below	-,177	,079	,160	-,39	,04
		19-29	,098	,052	,332	-,04	,24
		50-64	-,215*	,052	,000	-,36	-,07
		65 and over	-,490*	,089	,000	-,73	-,25
	50-64	18 and below	,037	,083	,992	-,19	,26
		19-29	,313*	,058	,000	,15	,47
		30-49	,215*	,052	,000	,07	,36
		65 and over	-,275*	,093	,025	-,53	-,02
	65 and over	18 and below	,313*	,110	,036	,01	,61
		19-29	,588*	,093	,000	,33	,84
		30-49	,490*	,089	,000	,25	,73
		50-64	,275*	,093	,025	,02	,53

*. The mean difference is significant at the .05 level.

KUVA 27. Esimerkki Post Hoc –testin tulosteesta (Tukey)

Kuten tulosteesta näkyy ikäluokkien välillä on merkittäviä eroja asiakastyytyväisyydessä (Sig.> .05). Esimerkiksi tarkasteltaessa ensimmäisen muuttujan (Tuotteiden hinta/laatusuhde on kohdallaan) ja siitä ikäryhmää 30-49 verrattaessa sitä muihin ikäryhmiin nähdään (ks. tummennettu kohta) että tilastollisesti merkittäviä eroja keskiarvoissa on ikäryhmiin 19-29

sekä 65 ja yli. Toisaalta saman kysymyksen kohdalla jos tarkastellaan ikäryhmää 50-64, ei eroja muihin ryhmiin ole.

Post Hoc testin yhteyteen voi myös tulostaa luokkien keskiarvot (valitaan *Options* –painikkeen alta *Descriptives*). Tieteellisissä artikkeleissa yksisuuntaisen varianssianalyysin (One-Way ANOVA) voi esittää taulukon muodossa esimerkiksi näin[4]:

Purpose of use	**N**	**Means**	**Mean square between groups**	**F value**	**Sig.**
1 Information seeking			9.765	12.943	.000
34 years and under	104	3.81			
35-49 years	326	3.33			
50-64 years	151	3.19			
65 years and above	32	3.03			
Total	613	3.36			
2 Investments			3.635	2.114	.097
34 years and under	101	2.24			
35-49 years	311	1.87			
50-64 years	138	2.01			
65 years and above	27	1.93			
Total	577	1.97			
3 Banking			.524	.404	.750
34 years and under	104	4.01			
35-49 years	336	4.05			
50-64 years	157	4.06			
65 years and above	34	3.86			
Total	631	4.05			

KUVA 28. One-Way ANOVA esimerkki

Kuvasta nähdään että on tutkittu miten eri ikäryhmät käyttävät internetiä ja kolmen toiminnon kohdalta nähdään että vain ensimmäisen ”*Information seeking*” –kohdalla ikäryhmien välillä on tilastollisesti merkitsevä ero keskiarvoissa (Sig. = .000).

[4] Karjaluoto, H., Mattila, M., and Pento, T. (2002), “A study on Internet usage among bank customers in Finland”, Proceedings of the AMA Winter Marketing Educators’ Conference (Austin, Texas), Vol. 13, pp. 422-429.

3.5. Korrelaatioanalyysi

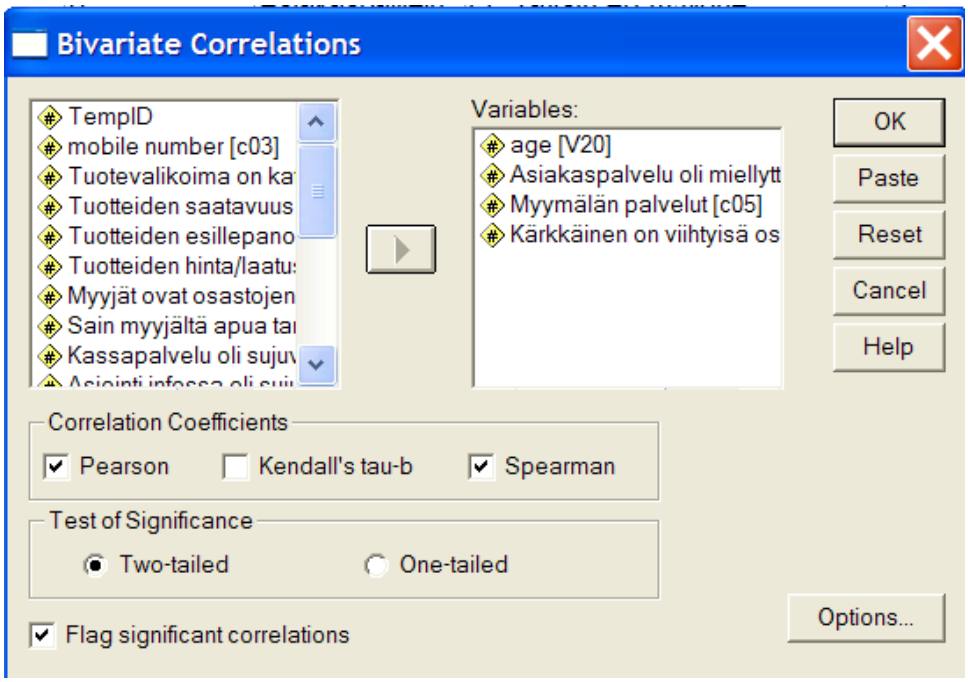
Kahden muuttujan välistä riippuvuutta voidaan lähestyä myös korrelaatiokertoimen kautta. Korrelaatio kuvaa kahden muuttujan (X ja Y) välistä LINEAARISTA riippuvuutta eli yhdysvaihtelun voimakkuutta ja suuntaa. Jos korrelaatio on voimakasta, voidaan toisen muuttujan arvoista päätellä toisen muuttujan arvot melko täsmällisesti. Jos korrelaatio on heikko, ei muuttujien välillä ole yhteisvaihtelua. Kahden muuttujan välinen korrelaatio voi saada minkä tahansa arvon välillä -1 ja +1. Korrelaation saamaa lukuarvoa kutsutaan KORRELAATIOKERTOIMEKSI, ja sitä tulkitaan seuraavasti:

- k < 0 (esim. k= -.15), tarkoittaa että korrelaatio on negatiivinen eli muuttujien X ja Y arvot muuttuvat eri suuntiin, eli kun X (esim. ikä) kasvaa, niin Y (esim. asenne tietokoneisiin) pienenee ja päinvastoin.
- k = 0, muuttujien välillä ei ole lineaarista riippuvuutta
- k > 0 (esim. k= .35), tarkoittaa että korrelaatio on positiivinen eli muuttujien X ja Y arvot muuttuvat samaan suuntaan (eli kun X kasvaa niin Y kasvaa). Esimerkiksi koulutus ja tulot usein.
- k = 1 (tai -1), muuttujien X ja Y lineaarinen riippuvuus on täydellistä (vain teoriassa mahdollinen)

Yleisin käytetty korrelaatiota kuvaava tunnusluku on ***Pearsonin*** tulomomenttikorrelaatiokerroin (r). Se on vähintään kahden **intervalliasteikollisen** muuttujan keskinäisen lineaarisen riippuvuuden voimakkuutta kuvaava tilastollinen tunnusluku.

Järjestysasteikollisten muuttujien korrelaatio pitää laskea käyttämällä ***Spearmanin*** korrelaatiokerrointa (tai Kendall'in Tau-b).

SPSS-ohjelmassa korrelaatioanalyysit löytyvät komennolla: **Analyze / Correlate / Bivariate**, jolloin avautuu oheinen ikkuna:



KUVA 29. Korrelaatioanalyysin aloitusikkuna

Ikkunan vasemmassa reunassa olevasta muuttujalistasta valitaan analyysiin ne muuttujat joiden välistä korrelaatiota halutaan tutkia ja siirretään ne oikealla olevaan Variables –kenttään. *Correlation Coefficients* –kentästä valitaan sopiva korrelaatioanalyysi (nyrkkisääntö: PEARSON silloin kun muuttujat ovat intervalliasteikollisia ja SPEARMAN silloin kun muuttujat ovat järjestysasteikollisia). Usein suuremmilla aineistoilla käytetään Pearsonia vaikka muuttujat onkin mitattu järjestysasteikollisesti.

Oletuksena ohjelma on laittanut rastin kohtaan *Flag significant correlations* joka näkyy tulosteessa korrelaatiomatriisissa siten että tilastollisesti merkittävät korrelaatiot on merkitty asteriskein (* p<0.05; ** p<0.01). *Test of Significance* –valikosta voi valita joko kaksisuuntaisen (*Two-tailed*) testin, joka on oletuksena tai yksisuuntaisen (*One-tailed*). Mikäli tutkimushypoteesia asetettaessa ei ole tehty mitään ennakko-oletuksia korrelaation suunnan suhteen (esim. että tyytyväisyys vaikuttaa onnellisuuden määrään tai onnellisuuden määrä vaikuttaa tyytyväisyyteen) kannattaa tehdä kaksisuuntainen testi, jolloin matriisiin tulostuvat korrelaatiokertoimet ja Sig. –arvot ovat puolet vastaavan kaksisuuntaisen testin arvosta. SPSS tulostaa korrelaatiokertoimista korrelaatiomatriisin (KUVA 30), jossa näkyvät muuttujien väliset korrelaatiokertoimen arvot (esimerkissä r = .143, ja Sig-arvo p = .000). Tulosteeseen tulee myös näkyviin * -merkkinä merkitsevyysarvo, eli kuten todettiin yksi * -merkki tarkoittaa että korrelaatiota on jonkin verran (p < .05) ja kaksi ** -merkkiä tarkoittaa että korrelaatio on tilastollisesti merkitsevä (p < .01).

Correlations

		age	Asiointi infossa oli sujuvaa
age	Pearson Correlation	1	,143**
	Sig. (2-tailed)		,000
	N	2066	2025
Asiointi infossa oli sujuvaa	Pearson Correlation	,143**	1
	Sig. (2-tailed)	,000	
	N	2025	2195

**. Correlation is significant at the 0.01 level (2-tailed).

KUVA 30. Esimerkki korrelaatiomatriisin tulosteesta

Esimerkissä on tutkittu aineistosta kahden muuttujan (AGE ja ASIOINTI INFOSSA OLI SUJUVAA) välistä korrelaatiota jossa ikä –muuttuja on mitattu avoimena kysymyksenä eli on intervalliasteikollinen ja toinen muuttuja on 5-asteikollinen järjestysasteikko. Aineisto on suuri (n=2070) ja data metristä eli voidaan käyttää Pearsonin korrelaatiokerrointa. Spearmanin

käyttö ei myöskään ole virhe ja sen tulos onkin lähes sama Pearsonin kanssa (Spearman = .157, p = .000). Samoin Kendall Tau-b antaa lähes identtisen tulokset (Kendall Tau-b = .120, p = .000).

Yllä olevan esimerkin tulosta tulkitaan siten että ***mitä vanhempi vastaaja sitä sujuvampana asiointia infossa pidettiin***. Tuloksen raportoinnissa on syytä mainita että vaikka riippuvuus on tilastollisesti erittäin merkitsevää (p = .000), on itse korrelaatiokertoimen arvo suhteellisen pieni (r = .143). Suurilla aineistoilla kuten esimerkkiaineisto muuttujien väliset korrelaatiot ovat usein tilastollisesti merkitseviä, joten muuttujien välistä riippuvuutta kannattaa tutkia myös muilla menetelmillä kuten ristiintaulukoinnin avulla.

YHTEENVETONA korrelaatioanalyysistä voidaan todeta että sitä käytetään erityisesti esi-analyyseissä kun halutaan alustavasti selvittää mitkä muuttujat korreloivat keskenään. Näiden korrelaatioiden varaan voidaan sitten rakentaa muuttujista esimerkiksi summamuuttujia tai edetä faktorianalyysin tekemiseen.

3.6. Faktorianalyysi

Faktorianalyysin perusideana on aineiston tiivistäminen eli pyritään kuvaamaan muuttujien kokonaisvaihtelua pienemmällä muuttujien määrällä. Faktorianalyysi voidaan suorittaa vähintään järjestysasteikollisille muuttujille, ja toisena ehtona on aineiston koko (vähintään 70-90 havaintoa, mielellään kuitenkin yli 100). Faktorianalyysi perustuu malliin, jossa etsitään havaittujen muuttujien avulla taustalla olevia tekijöitä eli ns. piilomuuttujia. Kun faktorit on SPSS -ohjelman avulla saatu rakennettua, on tutkijan tehtävä tulkita ne oikealla tavalla. Tulkinta tapahtuu siten että faktorit pitää nimetä tutkimalla mitkä alkuperäisistä muuttujista ovat

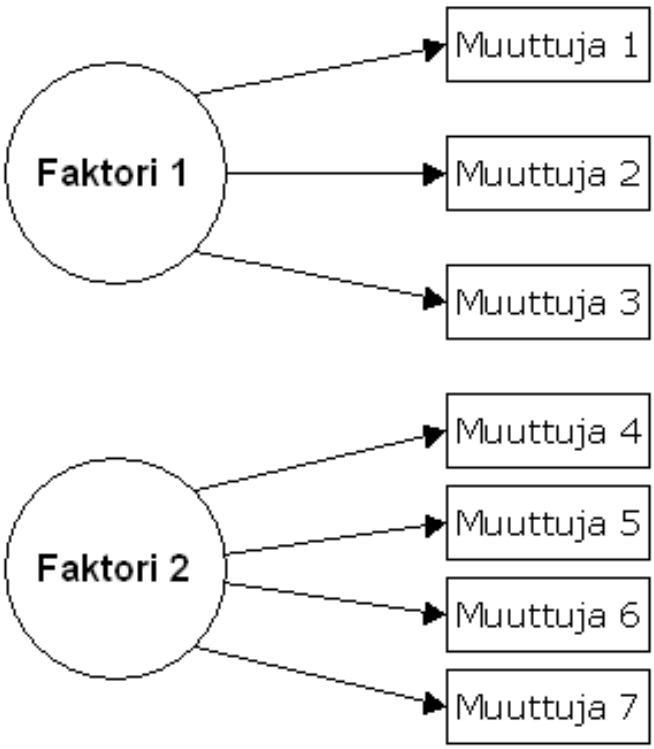
eniten korreloituneita kyseisen faktorin kanssa eli ns. lataavat faktorille eniten. Faktorianalyysia käytetään siis ennen kaikkea:

- Tiedon tiivistämiseen
- Hypoteesien testaamiseen
- Esianalyysinä eli voidaan muodostaa summamuuttujia (faktoripisteet), joita käytetään myöhemmin jatkotarkasteluissa kuten regressiomalleissa

Faktorianalyysin tekemisessä tutkijalla on paljon vaikutusvaltaa tuloksiin. Siksi onkin tärkeätä että faktorianalyysi perustuu valittuun teoreettiseen viitekehykseen eikä puhtaasti keksitä itse faktoreita ja niille nimiä. Faktorianalyysin käyttöä voidaan verrata esimerkiksi mielipidekyselyissä samaan asiaan liittyvien kysymysten ryhmittelyyn, jolloin usein huomataan samaan ryhmään kuuluvien vastausten korrelaatioita.

Faktorianalyysin perusidea on siis muodostaa joukosta muuttujia faktoreita kuvan 30 esittämällä tavalla. Kuvassa on kaksi faktoria ja seitsemän havaittua muuttujaa. Muuttujat voivat olla esimerkiksi kyselylomakkeen seitsemän väittämää, joilla on pyritty mittaamaan asiakastyytyväisyyttä. Faktorit ovat tavallaan ’piilomuuttujia’, koska niitä ei voida suoraan havainnoida, vaan niiden olemassaolo päätellään ainoastaan havaittujen muuttujien avulla. Käytännössä faktorin muodostaa joukko muuttujia, jotka korreloivat vahvasti keskenään, mutta vähän muiden muuttujien kanssa. Kuviossa faktoreista lähtee nuolia havaittuihin muuttujiin. Ne kuvaavat faktorianalyysin pohjana olevaa oletusta, jonka mukaan piilevät faktorit aiheuttavat havaitut ilmiöt, eikä päinvastoin.

Faktorianalyysi tuottaa jokaista kuvion nuolen ’vahvuutta’ kuvaavan arvon eli **faktorilatauksen** (*factor loading*). Latauksen suuruus kertoo kuinka paljon faktorin avulla pystytään selittämään havaitun muuttujan vaihtelusta. Lataukset saavat arvoja -1 ja 1 välillä. 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. Jos faktori kuvaa esimerkiksi asiakastyytyväisyyttä ja yksi muuttuja (esim. positiivinen suhtautuminen kassapalveluun) saa vahvan, mutta negatiivisen latauksen, tarkoittaa tämä sitä, että vastaajilla on huonoja kokemuksia kassapalvelusta (eli ovat vastanneet kysymykseen pienillä arvoilla kuten ”en lainkaan tyytyväinen), kun taas muihin kysymyksiin he ovat vastanneet suurilla arvoilla.



KUVA 31. Faktorianalyysin idea

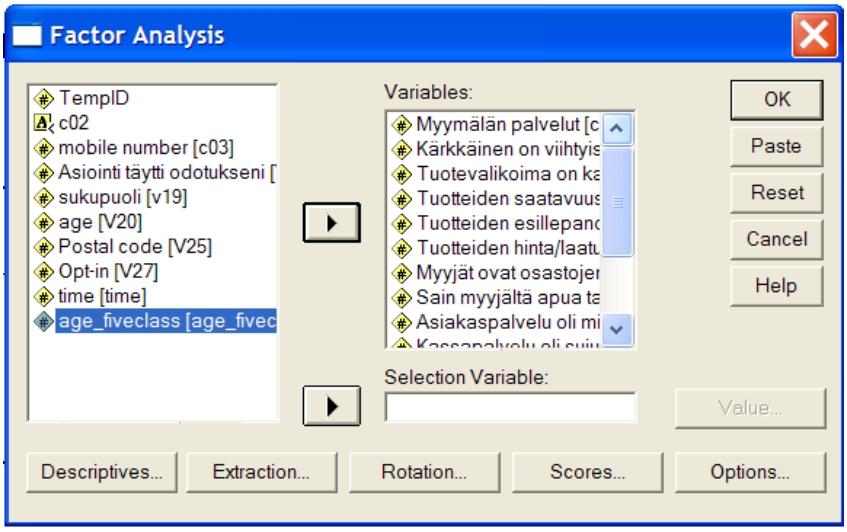
Kullekin havaintoyksikölle voidaan laskea **faktoripisteet,** jotka kuvaavat näiden piilomuuttujien arvoja ja jotka voidaan tallentaa jatkoanalyysejä varten.

Faktorimallin toimivuutta voidaan arvioida faktoreiden ominaisarvojen ja havaittujen muuttujien kommunaliteettien avulla. **Ominaisarvot** (*eigenvalue*) ilmoittavat, kuinka hyvin faktorit pystyvät selittämään havaittujen muuttujien hajontaa. Mitä suurempi faktorin ominaisarvo on, sitä paremmin se selittää muuttujien hajontaa ja päinvastoin. Kun faktorin ominaisarvo jaetaan havaittujen muuttujien määrällä, saadaan faktorin suhteellinen selitysosuus, joka saa arvoja nollan ja yhden välillä. Selitysosuus kertoo, kuinka suuri osuus kaikkien mallissa mukana olevien havaittujen muuttujien hajonnasta voidaan faktorin avulla selittää. Mitä suurempi osuus on, sitä parempi on faktorin selitysvoima. Kun kaikkien faktoreiden selitysosuudet lasketaan yhteen, saadaan koko analyysin selitysosuus. Se kertoo siis, kuinka suuri osuus kaikkien havaittujen muuttujien hajonnasta voidaan selittää kaikilla löydetyillä faktoreilla. **Kommunaliteetti** puolestaan kertoo, kuinka suuren osan faktorit selittävät muuttujan vaihtelusta. Jos muuttujan kommunaliteetti on lähellä yhtä, pystyvät faktorit selittämään sen vaihtelun lähes kokonaan. Toisaalta mitä pienempiä arvo kommunaliteetti saa, sitä huonommin faktorit muuttujaa selittävät. Jos yksittäisen muuttujan kommunaliteetti on pieni (< 0.3), kannattaa harkita, onko muuttujaa ylipäänsä syytä sisällyttää analyysiin.

Faktorianalyysissa voidaan erottaa kaksi toisistaan poikkeavaa lähestymistapaa. **Eksploratiivinen faktorianalyysi** pyrkii etsimään muuttujajoukosta faktoreita, jotka pystyvät selittämään havaittujen muuttujien vaihtelua ilman, että tutkijalla on etukäteen vahvoja odotuksia löydettävien faktoreiden määrästä tai niiden tulkinnasta. Eksploratiivinen faktorianalyysi on siis aineistolähtöinen tutkimusmenetelmä. Analyysin tuloksena voidaan löytää yksi tai useampia faktoreita, joita käytetään hyväksi tulosten tulkinnassa. **Konfirmatorisessa faktorianalyysissa** tutkijalla on jo etukäteen teorian pohjalta muodostettu käsitys aineiston faktorirakenteesta ja analyysin tehtävänä on joko varmistaa tai kumota tämä käsitys empiirisen aineiston pohjalta. Esimerkki konfirmatorisesta faktorianalyysistä on teorian testaus: -testataan esimerkiksi

Teknologian omaksumisteoriaa (Technology Acceptance Model eli TAM), jossa teoria esittää että kaksi faktoria (helppokäyttöisyys ja koettu hyödyllisyys) selittävät teknologian omaksumista / käyttöä. Teorian pohjalta muodostettaessa faktorianalyysi puhutaan siis konfirmatorisesta analyysista. Eksploratiivinen faktorianalyysi on näistä kahdesta faktorianalyysin muodosta kuitenkin yleisempi, ja esimerkiksi SPSS –ohjelma on periaatteessa rakennettu lähinnä eksploratiivisen faktorianalyysin tekemiseen. Mikäli kuitenkin halutaan käyttää konfirmatorista menetelmää tarkoittaa se lähinnä SPSS:n osalta sitä että asetetaan faktoreiden määrä esimerkiksi kolmeen ja katsotaan toimiiko teoreettinen malli omalla havaintoaineistolla. Konfirmatoriseen faktorianalyysiin liittyy usein rakenneyhtälömallien[5] rakentaminen ja testaus (esim. AMOS ja LISREL –mallit). Seuraavassa keskitytään eksploratiiviseen faktorianalyysiin.

Faktorianalyysi suoritetaan SPSS -ohjelmassa valitsemalla **Analyze / Data Reduction / Factor**, jolloin avautuu seuraavanlainen ikkuna:



KUVA 32. Faktorianalyysin aloitusikkuna

[5] Rakenneyhtälömallien ideana on tarkastella regressioanalyysin avulla faktorien välisiä kausaalisuhteita. Tutkimuksessa on esimerkiksi voitu muodostaa faktorianalyysin avulla vastaajien masentuneisuutta ja itsetuntoa koskevat faktorit. Rakenneyhtälömallien avulla voidaan tutkia, minkälainen kausaalinen vaikutus itsetunnolla on masentuneisuuteen.

Aloitusikkunasta siirretään muuttujanlistasta (vasen kenttä) *Variables* –kenttään ne muuttujat joiden avulla faktorianalyysi halutaan suorittaa. Muuttujien valinta pitää tapahtua tutkimusongelman valossa eli muuttujien valinta suoritetaan aikaisempien tutkimustulosten valossa. Kuitenkin jos kyseessä on täysin uusi aluevaltaus eli tutkittavasta ilmiöstä ei ole olemassa aiempia tutkimuksia voidaan muuttujia valita analyysiin oman teoreettisen pohdinnankin kautta, joskin silloin tulokset eivät ole kovin luotettavia. Muuttujat kannattaa valita faktorianalyysiin myös huomioimalla mukaan ne muuttujat jotka korreloivat vahvasti keskenään. Tämän voi tarkistaa korrelaatioanalyysin avulla ennen faktorianalyysin suorittamista.

Faktorianalyysin aloitusikkuna kätkee taakseen viisi painiketta, joista kuhunkin täytyy tehdä muutamia määrityksiä ennen faktorianalyysin toteutusta. *Descriptives* –painikkeen alta avautuvaan ikkunaan (KUVA 33) rastitaan ainakin Kaiser-Meyer Olkinin (KMO)–testi, joka antaa tuloksena lukuarvon joka kertoo onko faktoroinnille kyseisillä muuttujilla edellytyksiä. KMO-testin tulosta tulkitaan seuraavasti:

> .90 erinomaiset edellytykset

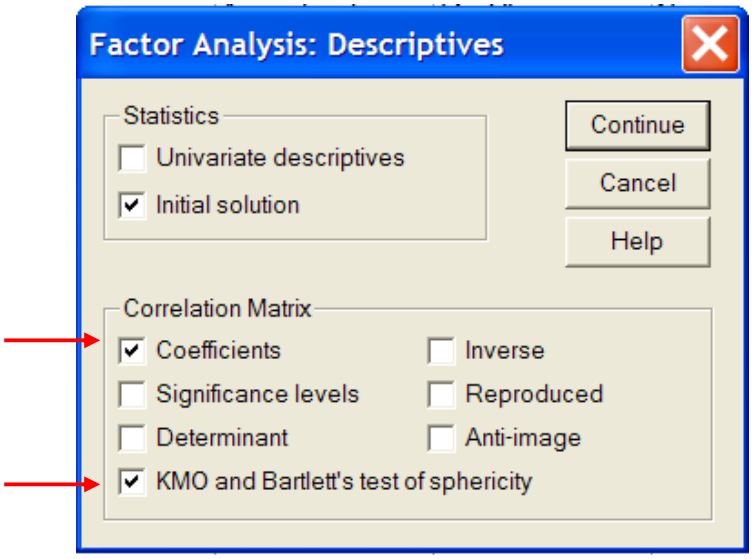
> .80 hyvät edellytykset

> .70 keskinkertainen

> .60 heikot (ei kannata jatkaa)

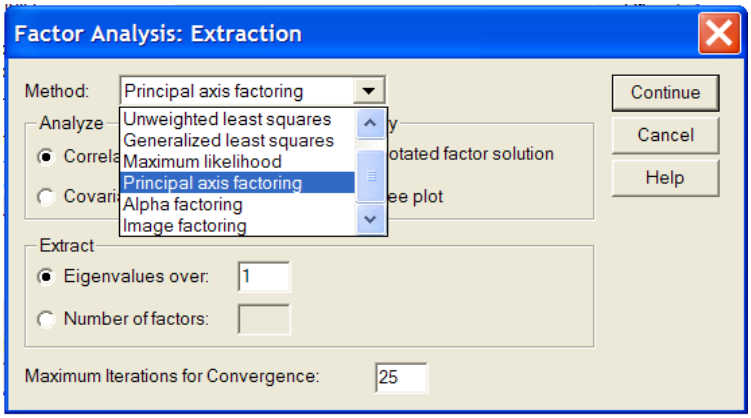
Kuvasta 33 voidaan myös haluttaessa valita tulostettavaksi korrelaatiomatriisi (rastitaan kohta *Coefficients*), jonka avulla nähdään muuttujien väliset korrelaatiot. Bartlett'in testin avulla testataan nollahypoteesia (=muuttujat eivät korreloi keskenään). Bartlett'in testin tulosta tulkitaan siten että mikäli testin Sig. arvo on < .01 (tai < .05) nollahypoteesi hylätään eli todetaan

että faktorianalyysin suorittamiselle on hyvät edellytykset koska muuttujien välillä on riittävästi korrelaatiota.



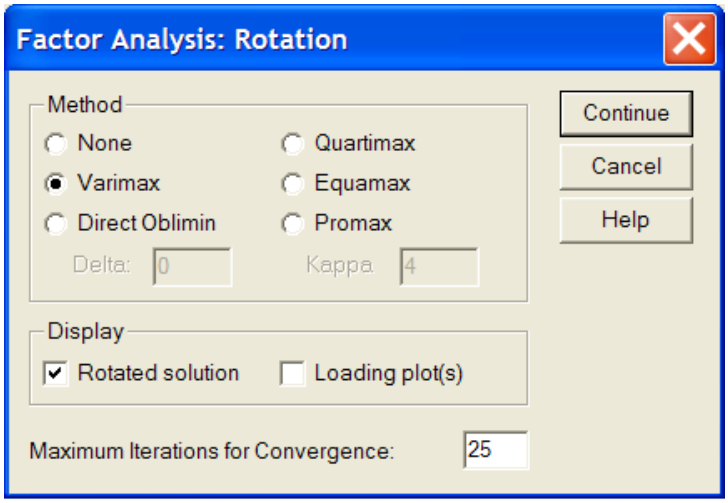
KUVA 33. Faktoroinnin edellytyksien testaaminen

Extractions –painikkeen alta avautuvasta ikkunasta (KUVA 34) valitaan faktoroinnin suorittamisen tapa. Suositeltava on valita ns. pääkomponenttimenetelmä ja erityisesti **Principal Axis Factoring**. *Extract* –menetelmäksi kannattaa valita oletuksena oleva **ominaisarvo** (*Eigenvalue*) eli faktorianalyysiin otetaan mukaan kaikki ne faktorit, joiden ominaisarvokriteeri > 1. Faktoreiden määrä voidaan asettaa myös vakioksi (esim. 4) jos teorian pohjalta testataan suoraan jotain faktorianalyysiä (eli tehdään konfirmatorinen faktorianalyysi).



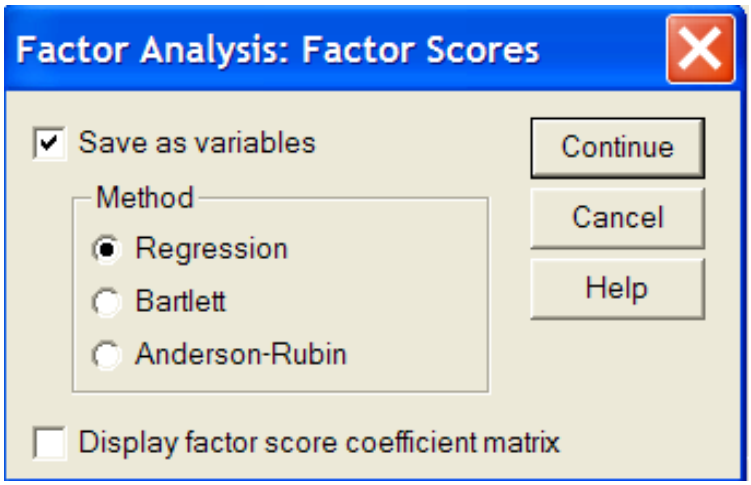
KUVA 34. Faktorointimenetelmän valinta (Extraction ikkuna)

Rotation –painikkeen alta valitaan rotatointimenetelmä (KUVA 35). Rotaatiolla (eli faktoriakselien kiertämisellä) viitataan prosessiin, jonka tarkoituksena on tehdä faktorianalyysin tulosten tulkinta helpommaksi. Rotaatio ei juurikaan muuta tuloksia sisällöllisesti, se tekee niistä vain helpommin tulkittavia. Rotaatiomenetelmät voidaan jakaa kahteen pääluokkaan. **Suorakulmarotaatiomenetelmät** (*orthogonal rotation*) tuottavat sellaisia faktoreita, jotka eivät korreloi keskenään ja **vinokulmarotaatiomenetelmät** (*oblique rotation*) puolestaan faktoreita, jotka voivat korreloida keskenään. Yleisesti rotaation käyttämistä faktorianalyysin yhteydessä voidaan suositella, koska se lähes poikkeuksetta tekee faktorilatausten teoreettisen tulkinnan helpommaksi. Käytetyin on **Varimax** ja sen valintaa suositellaan. Varimax on suorakulmainen rotaatiomenetelmä joka pyrkii minimoimaan vahvasti latautuvien muuttujien määrän yksittäiselle faktorille. Käytännössä tämä tarkoittaa sitä että menetelmä pyrkii faktorianalyysissä siihen, että kukin faktori saisi muutaman 'vahvan' latauksen. Toisin sanoen faktoreiden tulkinta helpottuu ja yksinkertaistuu. Toinen käyttökelpoinen rotaatiomenetelmä on **Promax**, joka on ns. vino rotaatiomenetelmä, joka sallii faktoreiden keskinäisen korrelaation. Suositellaan käytettäväksi erityisesti suurien havaintoaineistojen kohdalla.



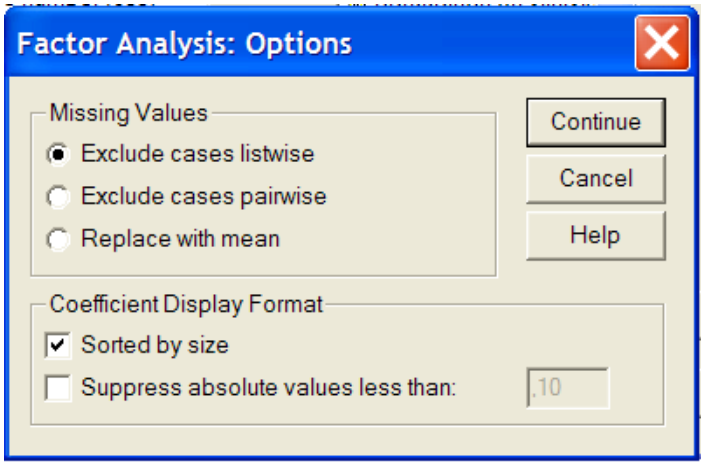
KUVA 35. Rotaatiomenetelmän valinta

Scores –painikkeen alta avautuvasta ikkunasta rastitetaan kohta ”Save as variables”, joka tallentaa **faktoripisteet** muuttujille. Yksi uusi muuttuja (ns. faktorimuuttuja) muodostuu kullekin faktorille. Faktoripisteet tulevat näkyviin SPSS *Data Editor* -ikkunan havaintomatriisissa oikeaan reunaan uusiksi muuttujiksi. Faktoripisteet saadaan laskemalla painotettu keskiarvo alkuperäisten muuttujien standardoiduista arvoista. Painoina käytetään faktorilatauksia. Tällä menetelmällä saatujen uusien faktoripistemuuttujien keskiarvo on aina nolla.



KUVA 36. Faktoripisteiden tallentaminen

Options –painikkeen alta rastitetaan kohta “Sorted by size”, mikä tarkoittaa sitä että tulosteessa faktorilatauksen järjestetään koon mukaan vahvimmasta latauksesta heikoimpaan.



KUVA 37. Latauksien järjestyksen asettaminen

Kun tarvittavat esivalmistetut on tehty painetaan OK ja saadaan näkyviin seuraavanlaisia tulosteita. Ensimmäisen kannattaa katsoa KMO-arvo ja Bartletin testin tulos (KUVA 38):

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,892
Bartlett's Test of Sphericity	Approx. Chi-Square	9200,708
	df	66
	Sig.	,000

KUVA 38. KMO:n ja Bartletin testien tulokset

Tuloksesta nähdään että sekä KMO (.892) että Bartletin testi (p = .000) antavat erinomaiset edellytykset faktorianalyysin tekemiselle. Seuraavaksi tarkistetaan muuttujien kommunaliteettitaulukosta (KUVA 38) muuttujien sopivuuden faktorianalyysiin. Tulkitaan *Extraction* –saraketta ja mikäli kommunaliteetti on pieni (<.3) kannattaa muuttuja pudottaa faktorianalyysista pois (ellei sen mukana olo ole teorian tai tutkimushypoteesien kannalta välttämätöntä) ja aloittaa faktorianalyysin teko alusta. Tässä esimerkissä muuttujajoukko liittyy asiakastyytyväisyyskyselyyn ja kommunaliteettienkin perusteella näyttäisi siltä että kaikki muuttujat sopivat faktorianalyysiin hyvin (kommunaliteetit kaikilla > .3).

Communalities

	Initial	Extraction
Myymälän palvelut	,282	,448
Kärkkäinen on viihtyisä ostospaikka	,467	,526
Tuotevalikoima on kattava	,362	,328
Tuotteiden saatavuus on hyvä	,440	,441
Tuotteiden esillepanoon on kiinnitetty huomiota	,431	,479
Tuotteiden hinta/laatusuhde on kohdallaan	,296	,339
Myyjät ovat osastojensa asiantuntijoita	,495	,543
Sain myyjältä apua tarvittaessa	,454	,429
Asiakaspalvelu oli miellyttävää	,564	,600
Kassapalvelu oli sujuvaa	,385	,377
Asiointi infossa oli sujuvaa	,423	,429
Liikenneyhteydet toimivat hyvin	,348	,411

Extraction Method: Principal Axis Factoring.

KUVA 39. Kommunaliteetti taulukko

Seuraava tulostettu taulukko (KUVA 40) kertoo faktorianalyysin tuloksen eli kuinka monta prosenttia muuttujien kokonaisvaihtelusta eli varianssista faktoriratkaisu selittää. Taulukon vasen sarake kertoo muuttujien ominaisarvon (*Initial Eigenvalues*). Tarkasteluun on siis otettu mukaan ominaisarvot > 1, joiden pohjalta on päädytty kahden faktorin ratkaisuun (rivit 1-2).

Taulukosta pitää raportoida ”**% of Variance**” kohta, joka kertoo sen kuinka monta prosenttia muuttujien kokonaisvaihtelusta kukin faktori selittää. Taulukossa tulostuu kaksi eri kohtaa josta tulkintaa voidaan tehdä: ”*Extraction Sums of Squared Loadings*”, jossa tulokset on esitetty ennen rotatointia sekä toinen kohta ”*Rotation Sums of Squared Loadings*” kohta jossa tulokset on esitetty rotatoinnin jälkeen. Yleensä tästä tulosteesta raportoidaan jälkimmäinen kohta eli tässä esimerkissä kerrottaisiin että rotatoinnin jälkeen ensimmäinen faktori selittää 26,5% muuttujien kokonaisvaihtelusta ja toinen faktori 18,0%. Tulosteesta nähdään että faktorit selittävät yhteensä 44,6% muuttujien kokonaisvaihtelusta (HUOM. Varianssin selityskyky on molemmissa edellä mainituissa sama eli 44.6%), mikä lisäksi tarkoittaa sitä että 55,6% informaatiosta on faktoroinnin seurauksena kadotettu eli tuloksia kannattaa tulkita tietyin varauksin.

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	4,941	41,177	41,177	4,402	36,680	36,680	3,184	26,536	26,536
2	1,506	12,547	53,724	,948	7,898	44,579	2,165	18,042	44,579
3	,981	8,173	61,897						
4	,804	6,704	68,601						
5	,639	5,324	73,925						
6	,571	4,755	78,680						
7	,495	4,127	82,807						
8	,482	4,021	86,828						
9	,447	3,726	90,554						
10	,423	3,525	94,079						
11	,379	3,158	97,238						
12	,331	2,762	100,000						

Extraction Method: Principal Axis Factoring.

KUVA 40. Faktorianalyysin faktoriratkaisutaulukko

Faktoreiden lopullinen nimeäminen tapahtuu ”*Rotated Factor Matrix*” –taulukosta (KUVA 41). Rotatointi pyrkii siihen että vahvat faktorilataukset pyritään saamaan suuremmiksi ja pie-

nemmät pienemmiksi verrattuna rotatoimattomaan ratkaisuun. Tässä esimerkissä ratkaisu löytyi kolmen rotatointikierroksen jälkeen.

Rotated Factor Matrix[a]

	Factor	
	1	2
Asiakaspalvelu oli miellyttävää	,735	,242
Myyjät ovat osastojensa asiantuntijoita	,700	,230
Sain myyjältä apua tarvittaessa	,645	,114
Tuotteiden hinta/laatusuhde on kohdallaan	,582	-,022
Kassapalvelu oli sujuvaa	,522	,322
Tuotteiden saatavuus on hyvä	,518	,416
Asiointi infossa oli sujuvaa	,465	,461
Tuotevalikoima on kattava	,447	,357
Myymälän palvelut	,204	-,638
Kärkkäinen on viihtyisä ostospaikka	,401	,604
Liikenneyhteydet toimivat hyvin	,254	,589
Tuotteiden esillepanoon on kiinnitetty huomiota	,418	,552

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

KUVA 41. Rotatoitu faktoriratkaisu

Tulosteen (KUVA 41) tulkinta aloitetaan katsomalla vahvimmat faktorilataukset, joiden perusteella faktorin nimeäminen aloitetaan. Faktorille 1 latautuu vahvimmin muuttujat ”Asiakaspalvelu oli miellyttävää”, ”Myyjät ovat osastojensa asiantuntijoita”, sekä ”Sain myyjältä apua tarvittaessa”. Täten faktori 1 näyttäisi selkeästi liittyvän henkilökohtaiseen palveluun ja se voitaisiinkin nimetä ”Henkilökohtainen palvelu” –nimellä. Faktori 2 saa latauksia neljältä muuttujalta, joista vahvimmin lataa muuttuja ”myymälän palvelut” (HUOM! Lataus on nega-

tiivinen, joka tarkoittaa että vastaajat ovat antaneet muuttujalle pieniä arvoja), toiseksi vahvimmin ”Kärkkäinen on viihtyisä ostospaikka”, ja kolmanneksi vahvimmin muuttuja ”Liikenneyhteydet toimivat hyvin”. Näiden latausten valossa faktori liittyy yleiseen viihtyvyyteen ja toimivuuteen joten faktori voitaisiin nimetä ”Viihtyvyys ja toimivuus” –nimellä. Esitetyn faktorianalyysin yksi ongelma on se että muutamat muuttujat (lähinnä ”Tuotteiden saatavuus on hyvä”, Asiointi infossa oli sujuvaa” sekä ”Tuotevalikoima on kattava”) lataavat molemmille faktoreille (lataus > .3) mikä ei ole hyvä asia ja paremman faktorianalyysin saamiseksi kannattaa tutkijan miettiä voisiko ne jättää analyysistä pois. Faktoreiden nimeäminen kannattaa usein tehdä teorian valossa eli käyttää samoja kuin aiemmissa tutkimuksissa on käytetty, jos muuttujat ovat suunnilleen samanlaisia.

3.7. Regressioanalyysi

Regressioanalyysin perusidea on että selitetään yhtä jatkuvaa muuttujaa (DEPENDENT) yhdellä tai useammalla jatkuvalla muuttujalla (INDEPENDENT). Mikäli käytetään useampaa kuin yhtä selittävää muuttujaa puhutaan monimuuttujaisesta regressioanalyysista. Muuttujien riippuvuus oletetaan lineaariseksi sekä jäännökset normaalijakautuneiksi (jäännös eli residuaali on selitettävän muuttujan todellisen arvon ja lasketun ennusteen välinen erotus).

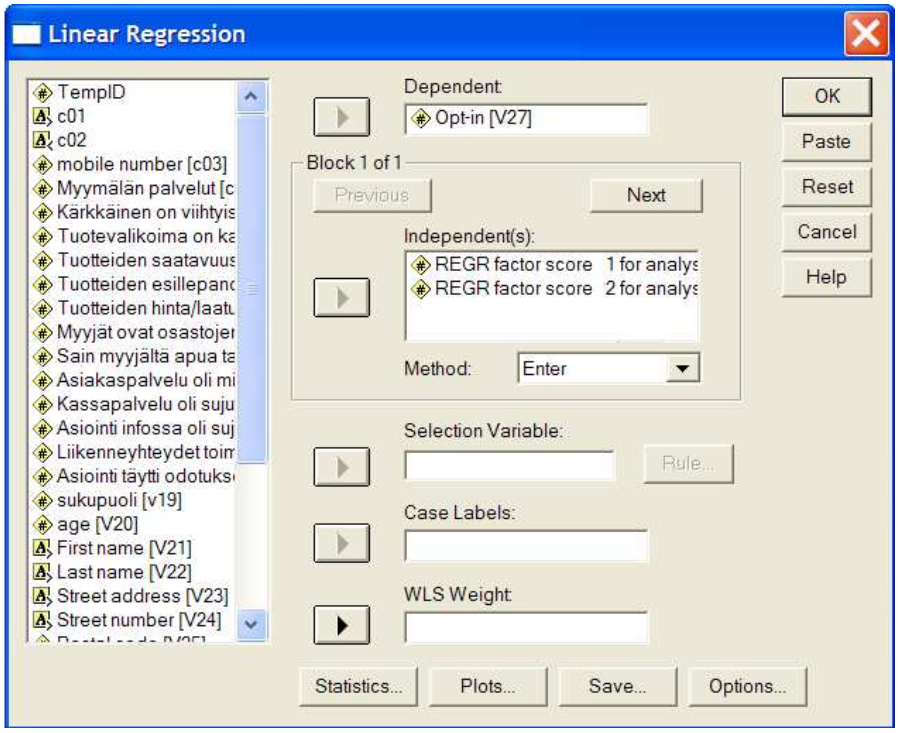
Regressioanalyysin toteuttamiseksi täytyy seuraavien ehtojen täyttyä:

- Muuttujat likimain normaalisti jakautuneita
- Muuttujat mitattu vähintään intervalliasteikolla (järjestysasteikolliset tietyin varauksin)

- **Lineaarisuus.** Muuttujilla pitää olla lineaarinen riippuvuus toisistaan. Regressioanalyysin avulla voidaan tutkia muuttujien välisiä lineaarisia eli suoraviivaisia kausaalisuhteita. Jos regressioanalyysin tulokset osoittavat, että selittävällä muuttujalla ei ole tilastollisesti merkitsevää yhteyttä selitettävään muuttujaan, tarkoittaa tämä tarkasti ottaen ainoastaan sitä, ettei lineaarista yhteyttä esiinny.
- **Poikkeavat havainnot eli outlier-tapaukset** (*outliers*). Joskus yksittäisillä poikkeavilla havainnoilla voi olla suuri vaikutus regressioanalyysiin tuloksiin. Tällaisia havaintoja kutsutaan niiden englanninkielisen nimen mukaan outlier-tapauksiksi. Yksittäinen poikkeava havainto voi muodostua esimerkiksi koodausvirheestä, mutta usein on kyseessä kuitenkin poikkeava havainto johtuen vaikkapa kysymyksen asettelusta.
- **Multikollineaarisuus**. Selittävät muuttujat eivät saisi korreloida keskenään (ainakaan vahvasti)
- **Heteroskedastisuus**. Regressiomallin virhetermien hajonta vaihtelee suuresti ja systemaattisesti x-muuttujien arvojen muuttuessa.
- **Havaintojen aikariippuvuus**. Yksi regressioanalyysin perusolettamuksista on, että havaintojen virhetermit ovat toisistaan riippumattomia. Eli jos tehdään esimerkiksi aikasarja-analyysia tarkoittaa se sitä että tulokset ovat riippuvaisia edellisistä tapahtumista. Havaintojen aikariippuvuuden korjaamiseksi on useita eri tapoja, joita emme kuitenkaan tässä käsittele.

Ennen regressioanalyysin tekoa tarvitaan joukko esianalyysejä kuten faktorianalyysi, josta saatuja faktoreita usein käytetään regressioanalyyseissä selittävinä muuttujina. Esimerkki regressioanalyysistä voisi olla verotuksen ja ostovoiman välinen regressio eli tutkitaan miten verotuksen helpottaminen vaikuttaa kuluttajien ostovoiman kasvuun. Tai miten koulutus vaikuttaa palkkaan.

Regressioanalyysi suoritetaan valitsemalla **Analyze / Regression / Linear**, jolloin avautuu seuraavanlainen ikkuna (KUVA 42). *Dependent* –kenttään valitaan muuttuja jota halutaan selittää, tässä tapauksessa on valittu ”Opt-in”, muuttuja joka kuvaa vastaajien halukkuutta vastaanottaa SMS -markkinointia. Kyseessä on ns. *Dummy* –muuttuja, joka saa vain arvot 1 (en halua) ja 2 (haluan). Selittäviksi muuttujiksi valitsemme edellisen kohdan faktorianalyysin avulla saadut faktorit, joiden faktoripistemuuttujia käytämme.



KUVA 42. Regressioanalyysin aloitusikkuna

Method –kohdasta valitaan *Enter* mallinvalintamenetelmäksi (oletuksena), joka pakottaa kaikki luetellut selittävät muuttujat malliin. Mallin muodostamisessa Method -kohdassa voidaan myös valita erinäisiä **askeltavia analyysejä** joiden käyttö saattaa olla perusteltua joissakin tilanteissa.

Muita esiasetuksia ei välttämättä tarvitse tehdä. Painamalla OK avautuu seuraavanlainen tuloste (KUVA 43):

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,133[a]	,018	,017	,382

a. Predictors: (Constant), REGR factor score 2 for analysis 1 , REGR factor score 1 for analysis 1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5,510	2	2,755	18,881	,000[a]
	Residual	304,828	2089	,146		
	Total	310,338	2091			

a. Predictors: (Constant), REGR factor score 2 for analysis 1 , REGR factor score 1 for analysis 1

b. Dependent Variable: Opt-in

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,819	,008		98,043	,000
	REGR factor score 1 for analysis 1	-,030	,009	-,069	-3,140	,002
	REGR factor score 2 for analysis 1	-,047	,010	-,103	-4,709	,000

a. Dependent Variable: Opt-in

KUVA 43. Regressiomallin tuloste

Model Summary –kohta kertoo mallin yhteenvedon, eli mallin selitysasteen. Mallin selitysvoimaa kuvaavat tunnusluvut. Selitysaste[6] (Adjusted R Square), kertoo, kuinka hyvin kyseinen malli selittää vastaajien halukkuutta vastaanottaa SMS -markkinointia asiakastyytyväisyyden valossa. Selittävät muuttujat (INDENDENT variables) ovat kuten todettiin faktorimuuttujia ja ne selittävät 1,7% halukkuudesta vastaan ottaa SMS -markkinointia. **ANOVA** –taulukko kertoo sopiiko malli aineistoon. Tässä esimerkissä voidaan sanoa mallin sopivan aineistoon (Sig. eli $p < 0.001$). Nollahypoteesina on “Malli ei sovi aineistoon”.

[6] Yhden selittäjän regressiomallissa selitysaste **R Square** on Pearsonin korrelaatiokerroin toiseen potenssiin korotettuna. Selitysaste **R Square** kasvaa aina, kun malliin lisätään muuttujia. Tästä johtuen mallien selitysasteita ei voi verrata keskenään, kun malleissa on eri määrä selittäviä muuttujia. **Adjusted R Square** on korjattu selitysaste. Korjaus on tehty ottaen huomioon mallissa olevien selittäjien määrä sekä havaintojen lukumäärä. Korjattuja selitysasteita voi vertailla keskenään.

Coefficients –taulukko kertoo regressiokertoimet. Standardoituja Beta -kertoimia käytetään muuttujien keskinäiseen vertailuun. Molemmilla selittävillä muuttujilla on merkitsevä vaikutus selitettävään muuttujaan ($p < .01$), ja $t > 2$. Beta –arvot ovat molemmissa tapauksessa negatiiviset mikä tutkimustuloksena viittaa siihen että selittävän ja selitettävien muuttujien välillä on negatiivinen riippuvuus, toisin sanoen mitä korkeampi asiakastyytyväisyys sen vähemmän halutaan SMS -markkinointia. On kuitenkin huomioitava että tämän väitteen todistamiseksi tarvitaan jatkoanalyysejä, koska regressiomallin selitysaste on todella pieni (1.7%). Mallia voisi yrittää parantaa lisäämällä mukaan selittäviä muuttujia kuten ikä ja koulutus, joilla on teorian valossa osoitettu olevan riippuvuutta halukkuuteen vastaan ottaa SMS -markkinointia.

Regressioanalyysin raportointi tieteellisessä tekstissä tehdään yleensä siten että raportoidaan standardoidut *beta* –arvot, *t*-arvot, *p*-arvot ja hypoteesien testaus kuten kuvassa 44:

TABLE 2
Regression Results of Antecedents of Customer Knowledge Development

Independent Variables	Standardized Coefficients	t-Value	Results
Innovation range	.21	(2.11)**	
Customer turbulence	.10	(1.38)*	
Competitive turbulence	.16	(1.65)**	
Technological turbulence	–.01	(–.21)	
Provision of resource slack	.06	(1.11)	H_1: not supported
Intelligent-failure reward system	.21	(2.11)**	
Creation of cross-functional new product development teams	.27	(3.01)***	H_3: supported
Integration mode of conflict resolution	.18	(1.98)**	
Championing the organizational goal of product leadership	.20	(2.09)**	H_5: supported
Project members' goal of product leadership	.07	(1.29)*	
Resource slack × reward system	.19	(2.03)**	H_2: supported
Cross-functional team × integration mode of conflict resolution	.46	(5.92)***	H_4: supported
Organizational goal × project members' goal	.11	(1.48)*	H_6: not supported
Adjusted R^2	.47		

*$p < .10$.
**$p < .05$.
***$p < .01$.
Notes: All significance tests are one-tailed.

KUVA 44. Esimerkki regressiotaulukosta *journal* –artikkelissa