

SPSS-OPAS

PERUSTEET

Maarit Valtari

Helsingin yliopisto
Valtiotieteellinen tiedekunta
Tieto- ja viestintätekniikka

Copyright Maarit Valtari ja HY:n valtiotieteellisen tiedekunnan TVT, 2004

ISBN 952-10-1509-8 (PDF)
ISBN 952-10-1510-1 (HTML)
ISBN 952-10-1511-X (WORD)

Valtiotieteellinen tiedekunnan TVT
Helsinki: 2004

KIITOKSET

SPSS-ohjelman pitkäjänteisessä oppimisprosessissa kiinnostukseni ohjelmaa kohtaan lisääntyi, ja aikanaan se johti ohjelman opettajana toimimiseen. SPSS:n oppiminen ei olisi ollut mahdollista ilman Timo Harmon kärsivällistä suhtautumista jatkuviin kysymyksiini ja ongelmiini. Suurimmat kiitokseni esitänkin Timo Harmolle. Lämpimät kiitokset Jere Majavalle siitä, että hän on varsin avuliaasti auttanut tämän SPSS-oppaan tekemisessä. Tämä SPSS-opas pohjautuu monilta osin valtiotieteellisen tiedekunnan tieto- ja viestintätekniikan SPSS-projektin varhaisimpiin versioihin, joita tekivät pääasiassa Timo Harmo ja Seppo Roponen. Ilman heitä tätä opasta ei olisi, ja siksi arvokkaimmat kiitokseni osoitankin heille. Lisäksi erityiset kiitokset Tero Erkkilälle, Aija Kaartiselle, Kimmo Vehkalahdelle sekä muille valtiotieteellisen tiedekunnan SPSS-kurssien myöhempien materiaalien täydentämiseen, kommentoimiseen ja parantamiseen osallistuneille. Tämän oppaan versio on päivitetty SPSS-ohjelman versioon 12.0.

Espoo, 6.9.2004

Maarit Valtari

SISÄLLYS:

1 JOHDANTO	1
1.1 Yleistä SPSS-ohjelmasta	1
1.2 SPSS käyttöliittymä	1
1.3 SPSS:n työskentelyikkunat	2
1.4 Aineistoikkunan valikot	5
2 HAVAINTOAINEISTON TALLENNUS	8
2.1 Yleistä SPSS -tiedostojen tallentamisesta	8
2.2 Uuden aineiston tallennus	8
3 HAVAINTOAINEISTON TARKASTELU JA KUVAAMINEN	10
3.1 Taulukointi	11
3.1.1 Frekvenssitaulukko	11
3.1.2 Muuttujan perustunnusluvut	12
3.1.3 Ristiintaulukko	13
3.1.4 Useampiulotteiset frekvenssitaulukot	16
3.1.5 Keskiarvotaulukot	16
3.1.6 Korrelaatio	18
3.1.7 Taulukointi	19
3.1.8 Ohjeita taulukoiden muotoiluun	20
3.1.9 Taulukoiden raportointi	22
3.1.10 Lopuksi: siis mitä tehdä, kun	23
3.2 Graafinen esittäminen	23
3.2.1 Pylväät	24
3.2.2 Histogrammit	30
3.2.3 Hajontakuviot	30
3.2.4 Viiva- ja aluekuviot	32
3.2.5 Piirakat	32
3.2.6 Kuvat niiden käyttötarkoituksen mukaan	33
3.2.7 Kuvien muokkaus SPSS:n kuva-editorilla	33
4 HAVAINTOAINEISTON MUOKKAAMINEN	37
4.1 Uudelleenluokittelu	37
4.2 Summamuuuttujat	41
4.3 Muut muunnokset	43
5 SYNTAKSI	44
6 SPSS:N HALLINTA	46
6.1 Keskeiset toiminnot	46
6.2 Osa-aineiston tekeminen	48
7 SPSS, ASCII, EXCEL JA WORD	49
7.1 ASCII	49
7.2 Excel	49
7.3 Tulosten siirtäminen tekstinkäsittelyohjelmaan	50
7.4 Kuvan tallentaminen eri grafiikkaohjelmiin jatkokäsiteltäväksi	50
KIRJALLISUUTTA	51

1 JOHDANTO

1.1 Yleistä SPSS-ohjelmasta

SPSS (kirjainlyhenne tulee sanoista Statistical Package for Social Sciences) on monipuolinen **tilastollisen tietojenkäsittelyn ohjelmisto**, jolla voidaan täyttää tavallisimmat tilastollisen analyysin tarpeet ja tehdä monia vaativampiakin tarkasteluja. Muita tilastolliseen tietojenkäsittelyyn sopivia tilasto-ohjelmistoja ovat mm. SURVO ja SAS. Tällä kurssilla käytetään Windows käyttöympäristöön sovitettua SPSS 12 versiota.

Kurssin tavoitteet

Kurssin tavoitteena on tutustua SPSS-ohjelmaan ja tarjota ohjelman käyttöön tarvittavat perustiedot. Kannattaa palauttaa mieleensä etenkin erilaiset mitta-asteikot (luokitteluasteikko, järjestysasteikko, suhde- ja välimatka-asteikko) sekä käsitteet jatkuva/luokiteltu muuttuja, erilaiset keskiluvut (ainakin keskiarvo sekä se, millä mitta-asteikolla mitatuista muuttujista niitä saa laskea!) sekä korrelaatio.

Kurssilla käytettävä aineisto

Materiaalissa esimerkkeinä ja kurssin harjoituksissa käytetään Valtarin (1997) keräämää ”Suomalaisten naisten suhtautuminen ulkonäköönsä” -tutkimusaineistoa. Siinä selvitettiin laaja-alaisesti suomalaisten naisten suhtautumista ulkonäköönsä: muun muassa emootioita, kognitioita, käyttäytymistä. Kyselylomake oli 24-sivuinen. Se postitettiin 500:lle satunnaisesti valitulle 18 – 74 -vuotiaalle äidinkieleltään suomenkieliselle suomalaiselle naiselle. Satunnaisotannan suoritti Väestörekisterikeskus. Vastausprosentti oli 54,6 prosenttia. Aineisto sisältää siis 273 havaintoa.

1.2 SPSS käyttöliittymä

SPSS muistuttaa suurelta osin muita Windows-ohjelmia, joten Windowsin käyttöön tottunut käyttäjä tuntee pitkälti sen käyttöön tarvittavat työskentelytavat. Alas vedettävät valikot, valikkojen takaa aukeavat valintaikkunat (*dialog box*) ja painikkeet ovat tuttuja kaikille Windowsin käyttäjille. Merkittävä poikkeus tavalliseen Windows-ohjelmaan verrattuna on ohjelman komentokieli. SPSS:ää voidaan ohjata myös kirjoittamalla komentoja komennoille tarkoitettuun syntaksikkunaan.

1.3 SPSS:n työskentelyikkunat

SPSS -ohjelmassa on **kuusi keskeistä erilaista työskentelyikkunaa**, joista jokaisella on oma käyttötarkoituksensa. Niistä yksi kerrallaan on aktiivinen. Ikkunat ja niiden käyttötarkoitukset ovat seuraavat:

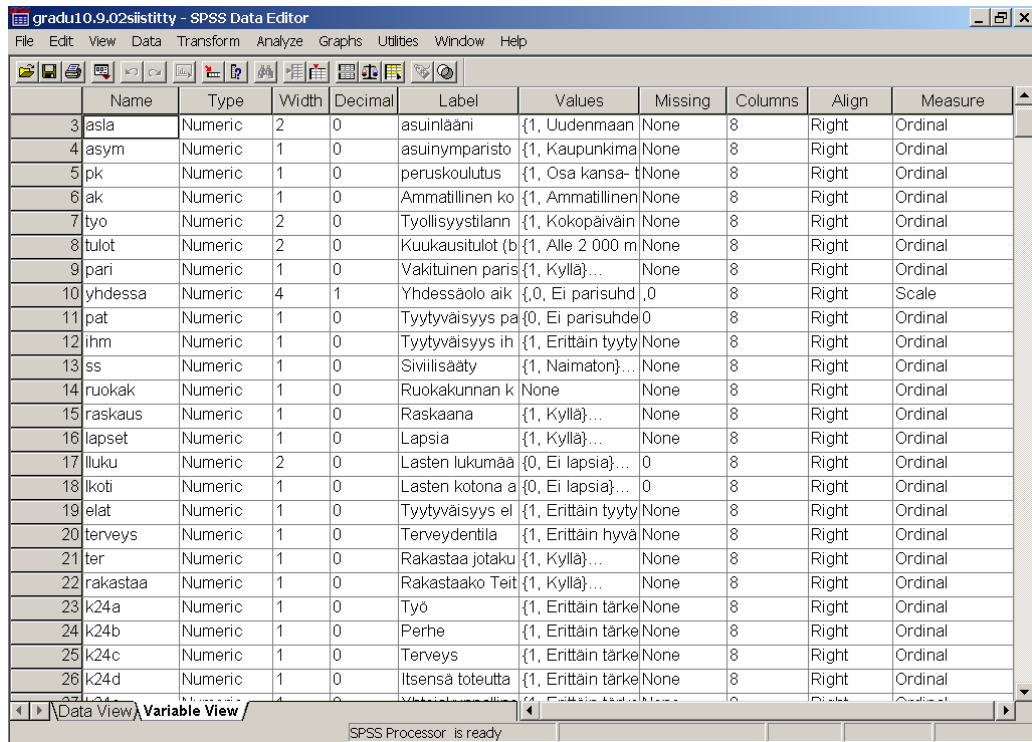
1. Aineistoikkuna (*data editor*): aineiston tallennus, tiedostojen avaaminen, uuden aineiston syöttö, muuttujamuunnokset ym. tehdään aineistoikkunassa. Aineistoikkunassa on kaksi näkymää: Data view eli datasisivu ja variable view eli muuttujasisivu.

Aineistoikkunan Data view

	havl	sv	asla	asym	pk	ak	tyo	tulot	pari	yhdes
1	1303	78	8	1	3	2	7	1	1	
2	1041	69	1	1	3	4	2	3	1	
3	1487	74	11	2	4	2	5	2	1	
4	1327	76	4	2	4	4	1	2	1	
5	1304	72	8	1	4	2	8	4	1	
6	1440	69	10	2	3	1	4	4	1	
7	1160	72	1	1	4	2	5	10	1	
8	1107	78	1	1	4	4	5	1	1	
9	1009	71	1	1	3	1	1	4	2	
10	1251	68	3	1	4	3	7	1	1	
11	1033	71	1	1	3	4	8	2	1	
12	1471	70	10	2	4	2	1	4	1	
13	1074	69	1	2	4	1	1	4	1	
14	1017	73	1	2	3	2	7	2	2	
15	1088	69	1	1	4	3	7	1	1	
16	1075	79	1	1	3	4	7	10	2	
17	1247	69	3	1	4	1	5	2	1	
18	1265	71	3	2	3	2	9	2	1	
19	1356	76	5	1	4	2	7	1	1	
20	1335	68	5	1	4	2	2	4	1	
21	1193	78	2	2	3	1	5	1	1	
22	1463	78	10	1	4	4	5	10	2	
23	1287	77	8	1	4	4	10	1	1	

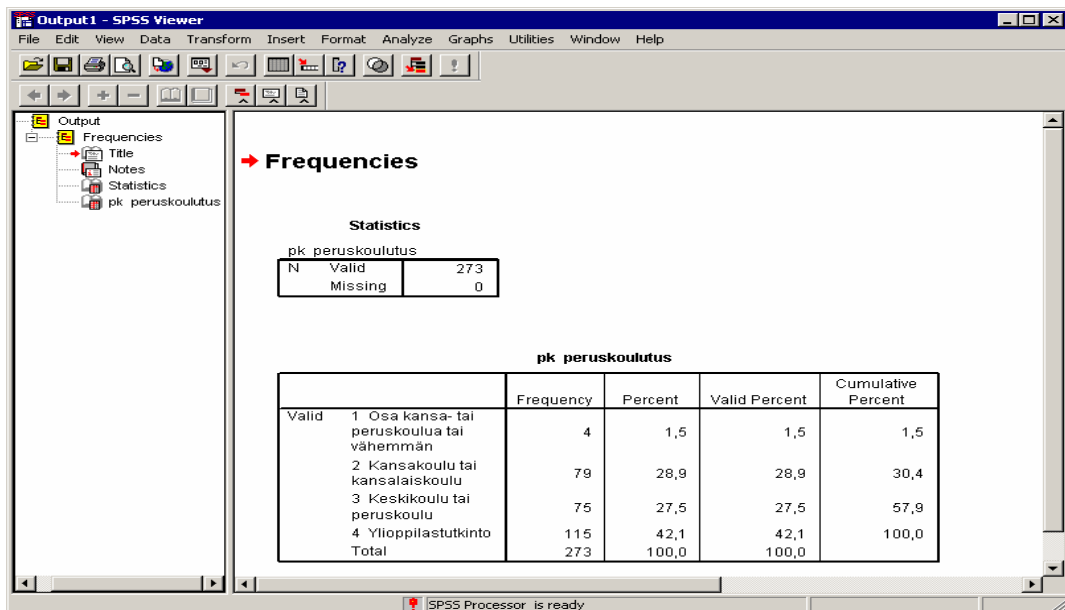
Aineistoikkunan datasisivusta (Data View) pääsee muuttujasisivulle (Variable View) ikkunan vasemmassa alalaidassa olevan välilehden kautta.

Aineistoikkunan Variable view



	Name	Type	Width	Decimal	Label	Values	Missing	Columns	Align	Measure
3	asla	Numeric	2	0	asuinlääni	{1, Uudenmaan	None	8	Right	Ordinal
4	asym	Numeric	1	0	asuinymparisto	{1, Kaupunkima	None	8	Right	Ordinal
5	pk	Numeric	1	0	peruskoulutus	{1, Osa kansa- t	None	8	Right	Ordinal
6	ak	Numeric	1	0	Ammatillinen ko	{1, Ammatillinen	None	8	Right	Ordinal
7	tyo	Numeric	2	0	Työllisyystilann	{1, Kokopäiväin	None	8	Right	Ordinal
8	tulot	Numeric	2	0	Kuukausitulot (b	{1, Alle 2 000 m	None	8	Right	Ordinal
9	pari	Numeric	1	0	Vakituinen paris	{1, Kyllä}...	None	8	Right	Ordinal
10	yhdessa	Numeric	4	1	Yhdessäolo aik	{0, Ei parisuhd	0	8	Right	Scale
11	pat	Numeric	1	0	Tyytyväisyys pa	{0, Ei parisuhte	0	8	Right	Ordinal
12	ihm	Numeric	1	0	Tyytyväisyys ih	{1, Erittäin tyyt	None	8	Right	Ordinal
13	ss	Numeric	1	0	Siviilisääty	{1, Naimaton}...	None	8	Right	Ordinal
14	ruokak	Numeric	1	0	Ruokakunnan k	None	None	8	Right	Ordinal
15	raskaus	Numeric	1	0	Raskaana	{1, Kyllä}...	None	8	Right	Ordinal
16	lapset	Numeric	1	0	Lapsia	{1, Kyllä}...	None	8	Right	Ordinal
17	lluku	Numeric	2	0	Lasten lukumää	{0, Ei lapsia}...	0	8	Right	Ordinal
18	lkoti	Numeric	1	0	Lasten kotona a	{0, Ei lapsia}...	0	8	Right	Ordinal
19	elat	Numeric	1	0	Tyytyväisyys el	{1, Erittäin tyyt	None	8	Right	Ordinal
20	terveys	Numeric	1	0	Terveystilanta	{1, Erittäin hyvä	None	8	Right	Ordinal
21	ter	Numeric	1	0	Rakastaa jotaku	{1, Kyllä}...	None	8	Right	Ordinal
22	rakastaa	Numeric	1	0	Rakastaako Teit	{1, Kyllä}...	None	8	Right	Ordinal
23	k24a	Numeric	1	0	Työ	{1, Erittäin tärke	None	8	Right	Ordinal
24	k24b	Numeric	1	0	Perhe	{1, Erittäin tärke	None	8	Right	Ordinal
25	k24c	Numeric	1	0	Terveys	{1, Erittäin tärke	None	8	Right	Ordinal
26	k24d	Numeric	1	0	Itsensä toteutta	{1, Erittäin tärke	None	8	Right	Ordinal

2. Tulostusikkuna (output): kaikkia tehtyjen analyysien tuloksia, taulukoita, kuvia ja muita tarkasteluja voidaan katsella ja tulostaa tästä tulostusikkunasta. Tulostusikkuna sisältää myös kapean istuntopolkuikkunan, jossa näkyy koko istunnon kulku. Nämä tulostusikkunat voidaan tallentaa omiksi tiedostoiksi.



Statistics

pk peruskoulutus

N	Valid	273
	Missing	0

pk peruskoulutus

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1 Osa kansa- tai peruskoulua tai vähemmän	4	1,5	1,5	1,5
2 Kansakoulu tai kansalaiskoulu	79	28,9	28,9	30,4
3 Keskikoulu tai peruskoulu	75	27,5	27,5	57,9
4 Ylioppilastutkinto	115	42,1	42,1	100,0
Total	273	100,0	100,0	

3. **Pivot-taulukkoikkuna (*pivot table editor*):** taulukoiden editointi (tekstin muuttaminen, sarakkeiden leventäminen tai kaventaminen, taulukoiden kääntäminen, tekstin värin muuttaminen ym.)

Output1 - SPSS Viewer

File Edit View Insert Pivot Format Analyze Graphs Utilities Window Help

Output

- Frequencies
 - Title
 - Notes
 - Statistics
 - K26.1 Olen tyytyväinen

Frequencies

Statistics

Olen tyytyväinen ulkonäkööni.

N	Valid	Missing
	272	
		1

Olen tyytyväinen ulkonäkööni.

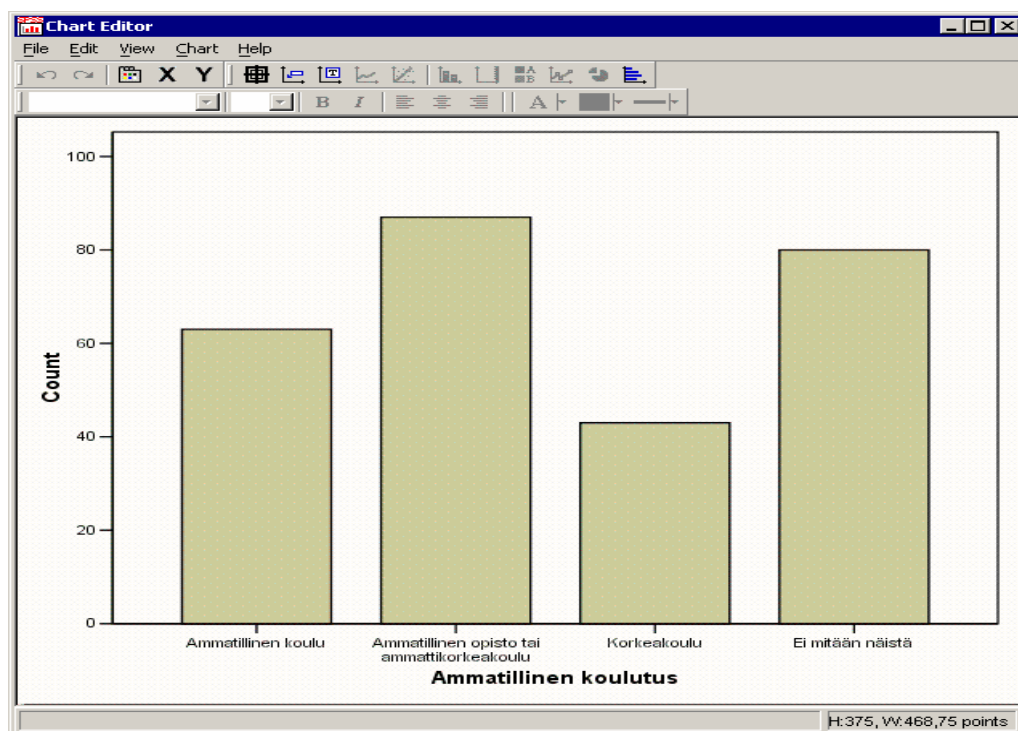
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Täysin samaa mieltä	47	17,2	17,3	17,3
	Jokseenkin samaa mieltä	157	57,5	57,7	75,0
	Ei samaa eikä eri mieltä	36	13,2	13,2	88,2
	Jokseenkin eri mieltä	28	10,3	10,3	98,5
	Täysin eri mieltä	4	1,5	1,5	100,0
	Total	272	99,6	100,0	
Missing	System	1	,4		
	Total	273	100,0		

SPSS Processor is ready

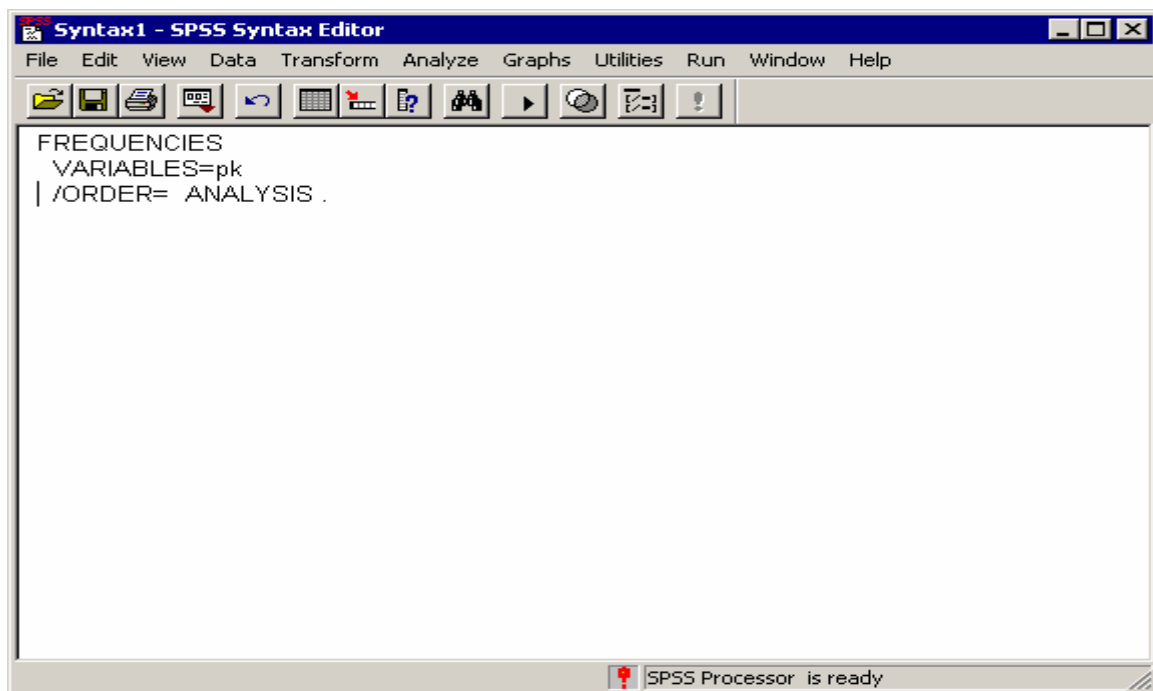
H: 157, W: 345 pt

Start | gradu10.9.02sisitt... | Output1 - SPSS ... | spss12.9.2002 - Mi... | 7:51

4. **Kuvankäsittelyikkuna (*chart editor*):** kuviin voidaan tehdä muutoksia tässä ikkunassa.



5. Syntaksi-ikkuna (*syntax*): ohjelmaa ohjataan tässä ikkunassa komentokielen avulla. Syntaksi-ikkunan käyttöä suositellaan mm. muuttujamuunnosten tekemiseen ja silloin, kun käyttäjä kokee komentokielen itselleen helpompikäyttöiseksi kuin valikkojen käytön. Syntaksi-ikkunassa on kätevämpi tehdä toistuvat analyysikomennot. Lisäksi syntaksikomentojen avulla on tarvittaessa helpompi neuvoa muita.



6. Script-ikkuna (*script editor*): ohjelmassa on valmiina joitakin käyttöä helpottavia scriptejä (muun muassa taulukoiden total-summien lihavointi). Itsekin voi luoda lisää scriptejä. Script-toimintoihin kannattaa tutustua Helpissä ([tutorial/sisälllys/customizing spss/customizing output with scripts](#)). Scriptien opiskelu ei sisälly tämän kurssin aiheisiin.

1.4 Aineistoikkunan valikot

Seuraavassa esitetään kymmenen valikkoa niiden käyttötarkoituksen mukaan:

1. File: valikkoa käytetään aineiston avaamiseen, tallentamiseen, tulostamiseen, tulostuksen oletusasetusten muuttamiseen ja lopetukseen. File-valikosta voi myös avata uusia syntaksi- tai tulostusikkunoita. Mark file read only -toiminnolla saadaan aineisto suojattua. Toiminnolla Display datafile information saadaan kaikki aineiston muuttujat

arvoselitteineen listattua. Toiminnoilla `Recently used data` ja `Recently used files` saadaan näkyville aiemmin käytetyt aineistot ja tiedostot.

2. Edit: käytetään tekstin ja muuttujien kopioimiseen, leikkaamiseen ja uudelleensijoittamiseen sekä poistamiseen (`Clear`). Täältä löytyvät myös aina niin tärkeät palauta (`Undo`) ja tee uudelleen (`Redo`) –toiminnot. Etsi (`Find`) –komennolla voidaan hakea valitun muuttujan arvoja. Lisäksi tässä valikossa voi muuttaa ohjelman oletusasetuksia (`Options`). Kun haluaa tehdä ohjelman oletusasetuksiin muutoksia, kannattaa oletusasetus-ikkunassa painaa Ohje-painiketta. Se kertoo, mitä muutoksia on mahdollista tehdä.

3. View: valikosta voi valita erilaisia pikapainikepalkkeja (`Toolbar`) ja vaihtaa aineistoikkunan fonttia. Täältä voi vaihtaa myös aineistoikkunan näkymän eli vaihtaa muuttujien arvokoodit arvoselitteiksi (`Value label`). `Variables` -toiminnolla voi siirtyä aineistoikkunasta muuttujaikkunaan ja vastaavasti muuttujaikkunasta voi siirtyä `Data` –toiminnolla aineistoikkunaan.

4. Data: käytetään muun muassa uusien muuttujien (`Insert variable`) ja havaintojen syöttämiseen (`Insert cases`), havainnon etsimiseen (`Go to case`) ja järjestämiseen (`Sort cases`). Tässä valikossa on varsin hyödyllinen toiminto, `Copy data properties`, jolla voidaan kopioida olemassa olevien muuttujien tietoja, kuten esimerkiksi vastausvaihtoehdot uusiin muuttujiin. Toisesta aineisto(i)sta voidaan yhdistää havaintoja tai muuttujia käsiteltävään aineistoon `Merge` –toiminnolla. Tässä valikossa voidaan aineisto jakaa osiin (`Split file`) ja suorittaa tiettyjen havaintojen valinta (`Select cases`) analyysihin. Lisäksi muuttujia voidaan painottaa tarpeen mukaan eri tavoilla (`Weight cases`).

5. Transform: käytetään muuttujamuunnosten (`Compute`), muuttujien uudelleenluokittelun (`Recode`) tekemiseen sekä muun muassa puuttuvien tietojen korvaamiseen (`Replace missing values`). `Count`-toiminnolla voidaan muodostaa uusi muuttuja jonkin muuttujan tietyistä arvoista. `Automatic recode` on varsin hyödyllinen toiminto käsiteltäessä tekstivastauksia, sillä toiminto koodaa vastaukset numeroarvoiksi.

6. Analyze: käytetään aineiston tarkasteluihin ja analyysihin. Valikko sisältää muun muassa mahdollisuuden erilaisten tarkastelujen tekemiseen (`Reports`), frekvenssit (`Descriptive statistics/Frequencies`), tunnusluvut (`Descriptive statistics/ Descriptives`), ristiintaulukoinnin (`Descriptive statistics /Crosstabs`), keskiarvotestit (`Compare means`), yleistetyt lineaariset mallit (`General linear model`), korrelaatiot (`Correlate`), regressioanalyysit (`Regression`), faktorianalyysin (`Data reduction/ Factor`), reliabiliteettianalyysin (`Scale`), ei-parametriset testit (`Non-parametric tests`) ja

monivastausmuuttujien käsittelyn (Multiple response). HUOM. Analyysivalikon toimintojen määrä on riippuvainen lisenssistä: SPSS:n peruslissenssi ei sisällä esimerkiksi toistettuja varianssianalyysijä.

7. Graphs: käytetään grafiikan tekemiseen joko perinteisellä tai interaktiivisella (kolmiulotteinen grafiikka) tavalla: muun muassa pylväskuvat (Bar), viivakuvat (Line), aluekuvat (Area), piirakat (Pie), viiksilaatikkokuvat (Boxplot), hajontakuvat (Scatter) ja histogrammit (Histogram). Gallery-toiminnolla saa näkyville erilaisia graafeja, ja opastuksen niiden tekemiseen. Tämä on hyödyllinen toiminto, johon kannattaa tutustua.

8. Utilities: erilaisia aputoimintoja sisältävä valikko. Variables-toiminnolla voidaan yksitellen tarkastella muuttujia ja niiden koodausta. Aineistoon voi kirjata kommentteja toiminnolla Data file comments. Tässä valikossa voidaan myös luoda muuttujasetti (Define sets) ja käyttää sitä (Use sets). Lisäksi scriptin ajaminen tapahtuu tässä valikossa (Run script).

9. Window: valikkoa käytetään ikkunasta toiseen siirtymiseen.

10. Help: SPSS:n käytön opastus.

Tätä on tärkeä opetella käyttämään! Tässä Helpin tärkeimmät toiminnot:

- **Topics/Sisällys:** Vastaa ohjekirjaa. Tätä kannattaa lueskella esim. SPSS -käyttöä aloittaessa, kun ei oikein tiedä vielä tarkalleen, mitä haluaa tehdä.
- **Topics/Hakemisto:** Hakusanalista. Jos esim. mietit, miten teet ristiintaulukon, nopeimmin löydät avun täältä hakusanalla Crosstabs.
- **Topics/Etsi:** Eroaa sisällyksestä siten, että tämä toiminto etsii kaikki maininnat antamastasi hakusanasta. Kokeile tätä, jos sisällys-toiminnolla ei tunnu löytyvän mitään.
- **Tutorial:** Näyttää kohta kohdalta SPSS:n peruskäyttöä. Hyvä tutustumiskeino ohjelmaan aloittelijalle. Aloita liikkuminen sivustolla nuolten avulla, tai aloita sisällysluettelon katsomisesta, johon pääset ylöspäin osoittavasta nuolesta.
- **Case studies:** SPSS:n käytön esimerkkejä. Suosittelen tutustumaan.

Helpistä löytyvät myös linkki SPSS -ohjelman kotisivuille (SPSS Home Page), joilta löytyy lisätietoja ohjelmasta sekä PDF -muodossa oleva opas SPSS -komentokielen käyttöön (Command Syntax Reference) sekä ”tilastollinen valmentaja” (Statistics Coach). Näitä kannattaa hyödyntää.

2 HAVAINTOAINEISTON TALLENNUS

2.1 Yleistä SPSS -tiedostojen tallentamisesta

SPSS -ikkunoista aineisto-ikkunan, tulostusikkunan ja syntaksi-ikkunan sisällöt **pitää kaikki tallettaa erikseen**. HUOM. Aineisto-ikkuna ja muuttujaikkuna kuitenkin tallentuvat samanaikaisesti kummasta tahansa ikkunasta. SPSS antaa talletetulle aineistoikkunalle päätteen **.sav**, tulostusikkunalle **.spo** ja syntaksi-ikkunalle **.sps**.

Tallennukset ja avaukset tapahtuvat samoin kuten muissakin Windows-ohjelmissa. Avaamisessa on kuitenkin otettava huomioon se, mistä SPSS -ikkunasta avaaminen tehdään (aineistot avataan Data Editoriin, tulostukset Output Vieweriin jne. SPSS -tiedostotyyppit voi kuitenkin avata mistä tahansa ikkunasta, kunhan muistaa valita avaus-valintaikkunassa oikean tiedostotyyppin tai All Files -vaihtoehdon).

2.2 Uuden aineiston tallennus

Kuten edellä tuotiin esille uuden aineiston tallennus tapahtuu aineistoikkunassa. Ennen kuin uutta aineistoa voidaan alkaa tallentaa, tulee määrittää **tallennuspohja**.

Peruseriaatteet aineiston tallennuksessa:

1. Määritellään muuttujat tallennuspohjaan yksi kerrallaan.
2. Vasta sitten, kun tallennuspohja on valmis, voidaan alkaa syöttää dataa.
3. Data on järkevintä syöttää yksi rivi eli yhden vastaajan tiedot kerrallaan. Yleensä vastauslomakkeissa on havaintotunnukset (eli ne on numeroitu), jotka syötetään havaintotunnusmuuttujan arvoiksi. Tämä helpottaa mahdollisten tallennusvirheiden korjaamista. Havaintotunnuksen perusteella löytää nimittäin kätevästi vastauslomakkeen, jota on tallennettu virheellisesti.

Aineistopohja kannattaa suunnitella tarkkaan. Tulee miettiä

- Mikä muuttuja kannattaa koodata milläkin tavalla.
- Mikä nimi (*Variable name*) muuttujalle on kuvaavin ja helpoin muistaa (muuttujan nimi saa olla 1 - 8 merkkiä pitkä ja sen tulee alkaa kirjaimella, älä käytä "ääkkösiä"). Huomaa, että

muuttujat kannattaa nimetä niin, että nimestä selviää sekin, kuinka päin jotakin asiaa on kysytty. Voitaisiin esim. kysyä: "En hyväksy aborttia" tai "Hyväksyn abortin". Muuttujien nimet voisivat olla vastaavasti "noabort"/ "yabort".

- Lisäksi jokaiselle muuttujalle kannattaa kirjoittaa selitystekstit eli tunnisteet (*labels*) ja määrittää luokitellun muuttujan mahdolliset arvot selitysteksteineen (*value labels*).
- Millaisia arvoja (*value*) muuttuja voi saada.

Esimerkkejä erilaisten muuttujien koodaamisesta tallennuspohjaan:

Ikä

Ihmisten ikä kannattaa kysyä syntymävuotena, koska silloin saadaan rehellisempää tietoa kuin ikää kysyttäessä. Tallennuspohjaa tehdessä tälle muuttujalle voi antaa esim. nimen sv. Muuttujalle kannattaa antaa selitysteksti (syntymävuosi) ja määrittää muuttujan mahdollisiksi arvoiksi ainoastaan numerot (Numeric). Kun aineisto on tallennettu, voidaan helposti tehdä sv:stä uusi, ikää ilmentävä muuttuja vähentämällä kyselyn suorittamisvuodesta vastaajan syntymävuosi. Yleisesti ottaen ikä kannattaa kysyä ilman valmiiksi tehtyjä ikäluokkia. Näin ikää voidaan käyttää jatkuvana muuttujana analyyseissä. Iän luokittaminen jälkeinpäin on hyvin vaivatonta. Tästä myöhemmin uudelleenluokittelun yhteydessä.

Siviilisäät

Valmiiksi luokitetut vaihtoehdot (naimaton, naimisissa, asumuserossa eronnut, leski) koodataan yhden muuttujan alle mahdollisiksi vastausvaihtoehdoiksi (vastaaja voi valita ainoastaan yhden vaihtoehdon). Tällöin esim. numero yksi vastaa naimatonta.

Vain naimattomille tarkoitettu kysymys

Jos tiedetään syy, miksi johonkin kysymykseen ei ole vastattu (esim. henkilö on avioliitossa, ja kysymys on tarkoitettu vain naimattomille) kannattaa tästä syystä vastaamatta jätetty kysymys merkitä puuttuvaksi arvoksi jollakin valitulla luvulla. Luvulle voi antaa selitystekstin normaaliin tapaan, esim. value "99", value label "aviossa". Jos henkilö on jättänyt vastaamatta kysymykseen jostakin tuntemattomasta syystä, jätetään kyseinen kohta tyhjäksi.

Ammatti

Jos vastaajalta on kyselylomakkeessa tiedusteltu avoimella kysymyksellä hänen ammattinsa. Vastaukset ovat siis tekstiä; eli eri ammatteja. Näin ollen muuttuja tulee määritellä String-muuttujaksi (muuttuja voi olla tekstiä tai numeroita). Tallennusvaiheessa vastaajien eri ammatit tallennetaan siis teksteinä, aineiston tallennuksen jälkeen SPSS muuttaa halutessa

Automatic recorde –toiminnolla vastaukset ammattiluokkiin, ja näin moninaisten analyysien tekeminen mahdollistuu. Tämähän ei ole mahdollista tekstimuuttujien kohdalla.

3 HAVAINTOAINEISTON TARKASTELU JA KUVAAMINEN

Tutkimusaineiston tärkeimmät tarkastelukeinot ovat:

- frekvenssien ja prosenttien laskeminen
- tunnuslukujen (keskiarvo, keskihajonta, vaihteluväli ym.) laskeminen
- graafinen kuvailu (pylväät, piirakat)
- ristiintaulukointi ja
- korrelaatioiden laskeminen

Erilaisia tunnuslukuja ja jakaumia tarkastelemalla tutkija saa itselleen yleiskuvan aineistosta ja mahdolliset tallennusvirheet tulevat esiin.

Vinkki aineiston analyysien, tarkastelujen, uudelleenluokittelun ym. tekemiseen: kaikissa valikoissa valitsemalla muuttujalistasta jokin muuttuja hiirellä (muuttujan tausta on tällöin tumma), saat hiiren oikealla painalluksella näkyviin `Variable info` –toiminnon, jolla saat näkyviin muuttujan luokat näkyviin.

Tutkimustulosten esittämistapaa tutkimusraportissa kannattaa pohtia tarkoin. Onko mahdollista esittää tulokset tekstinä ilman että tekstistä tulee liian raskasta, pitkää tai sekavaa? Vai olisiko taulukoiden tai kuvien käyttö tämän tutkimusongelman kohdalla parempi esittämistapa? Kuvat ovat hyviä muuttujien välisen interaktion kuvaamisessa sekä muuttujien välisissä vertailuissa. Suhteessa taulukoihin niiden haittapuolena on se, että ne eivät ole aivan yhtä tarkkoja. Parasta tutkimustulosten esittämistapaa miettiessä tulee siis miettiä, mitä haluaa kertoa ja mikä on selkein ja taloudellisin tapa esittää se.

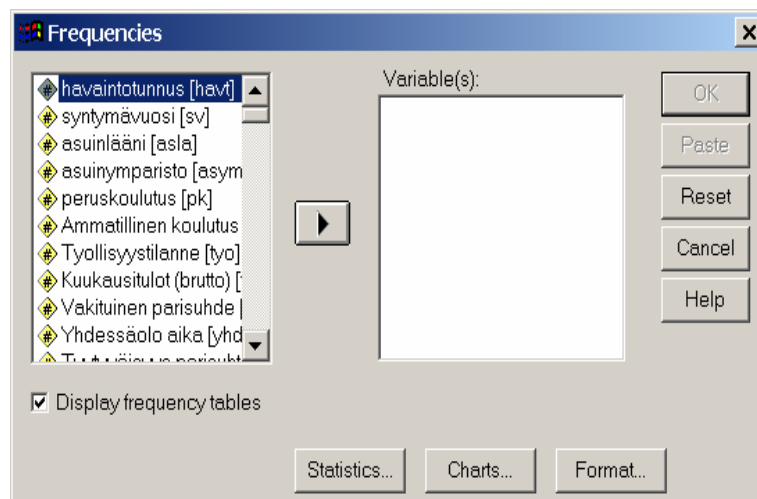
Jos halutaan kertoa, että miehet ovat keskimäärin 10 cm naisia pidempiä kuvaa ei tarvita esityksen tueksi. Jos halutaan vertailla Suomen kuntien elinkeinorakenteita tulee kuvasta helposti aivan liian monimutkainen jotta se tukisi esitystä.

3.1 Taulukointi

3.1.1 Frekvenssitaulukko

Yksisuuntainen frekvenssijakauma eli suora jakauma kertoo muuttujan eri luokkien yleisyyden havaintoaineistossa (eli vastausten *lukumäärän* eri luokissa, engl. *frequency* tai *count*). Vain kategoristen (eli luokiteltujen) muuttujien frekvenssijakaumia on järkevää tarkastella, koska jatkuvalla muuttujalla voi olla joka havainnolla eri arvo ja siten yhtä monta luokkaa kuin havaintojakin. Monesti suorien jakaumien tarkastelu on aineistoon tutustumisen ensimmäinen vaihe, ja kuuluu olennaisena osana myös aineiston koodauksen tarkistamiseen.

Frekvenssitaulukon pääsee tekemään analyysivalikosta (Analyze/Descriptive statistics/Frequencies). Näin saadaan näkyville seuraava ikkuna.



Variable(s)-kohtaan valitaan muuttuja(t). Valikon alalaidassa on kolme painiketta, joiden toiminnot ovat seuraavat:

Statistics: tässä valikossa voi pyytää ohjelmaa tulostamaan muuttujan perustunnuslukuja.

Charts: voidaan pyytää ohjelmaa tulostamaan taulukon lisäksi pylväskuva.

Format: tässä valikossa voi vaikuttaa frekvenssitaulukon esittämistapaan.

Frekvenssitaulukon ensimmäinen sarake kertoo muuttujan luokan ja toinen sarake kertoo kussakin luokassa olevien havaintojen määrän. Osuudet voidaan laskea tarkoituksesta riippuen joko koko aineistosta (*percent*) tai vain vastanneista (*valid percent*). *Cum percent* tarkoittaa kumulatiivista osuutta, eli tämä sarake kertoo, kuinka monta havaintoa kyseisessä ja sitä edeltävissä luokissa on yhteensä.

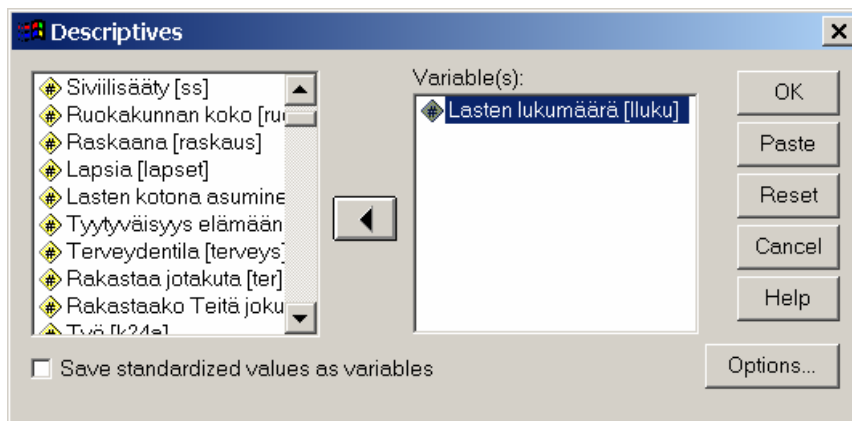
Taulukko 1. Vastaukset väittämään "Toivon, että olisin paremmannäköinen".

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Täysin samaa mieltä	46	16,8	17,1	17,1
	Jokseenkin samaa mieltä	87	31,9	32,3	49,4
	Ei samaa eikä eri mieltä	61	22,3	22,7	72,1
	Jokseenkin eri mieltä	34	12,5	12,6	84,8
	Täysin eri mieltä	41	15,0	15,2	100,0
	Total	269	98,5	100,0	
Missing	System	4	1,5		
Total		273	100,0		

3.1.2 Muuttujan perustunnusluvut

Jatkuvan muuttujan jakaumaa voidaan tarkastella erilaisilla tunnusluvuilla (keskiarvo, keskihajonta etc.) ja histogrammilla. SPSS luokittelee automaattisesti jatkuvan muuttujan histogrammia varten. Mikäli et ole tyytyväinen siihen, voit tehdä oman luokituksen ja käyttää sitä.

Muuttujan perustunnusluvut saa tulostettua analyysivalikosta (Analyze/Descriptive statistics/Descriptives). Näin saadaan näkyville seuraava ikkuna:



Variables-kohtaan valitaan muuttuja(t). Valikon oikeassa alalaidassa on Options-painike. Options-valikossa voi valita, mitkä muuttujan perustunnusluvut esitetään taulukossa.

Keskeisemmät perustunnusluvut:

Keskiarvo (mean), keskihajonta (std. deviation), laskettujen arvojen määrä (valid), puuttuvien arvojen määrä (missing), keskiarvon keskivirhe (std. error of mean, S.E mean), varianssi (variance), vaihteluväli (range), minimi, maksimi, havaintoarvojen summa, alakvartiili (percentile 25), mediaani (percentile 50) ja yläkvartiili (percentile 75). Lisäksi täältä voi pyytää ohjelmaa tulostamaan muuttujan jakaumaa kuvaavat tunnusluvut: huipukkuus (kurtosis) ja vinous (skewness).

Taulukossa 2 on esitetty vastaajien iän keskiarvo, keskihajonta ja iän vaihteluväli. Keskiarvon raportoinnin yhteydessä esitetään yleensä myös keskihajonta.

Taulukko 2. Naisten iän keskiarvo.

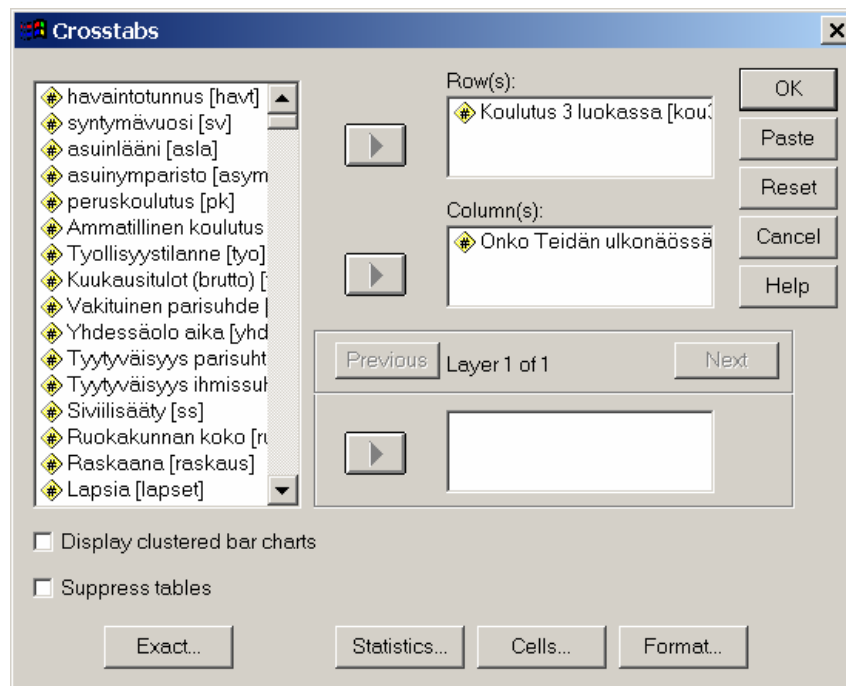
	N	Minimum	Maximum	Mean	Std. Deviation
Iän keskiarvo	251	40,0	80,0	60,118	6,2507
Valid N (listwise)	251				

3.1.3 Ristiintaulukko

Kaksiulotteisessa ristiintaulukossa esitetään muuttujan frekvenssijakauma (havaintojen lukumäärät muuttujan eri luokissa) toisen muuttujan luokissa joko frekvensseinä tai prosentteina (tai kumpinakin). Ristiintaulukon avulla voidaan siis tarkastella kahden muuttujan välistä yhteyttä.

Ristiintaulukossa kaksi muuttujaa esitetään samassa taulukossa siten, että toinen asettuu sarakkeille (*columns*) ja toinen riveille (*rows*). Taulukon soluissa esitetään solufrekvenssit, jotka kertovat, kuinka monta mainitun ominaisuuden varustettua yksilöä aineistossa on. Melkein aina tilastollisessa tutkimuksessa on kuitenkin havainnollisempaa tarkastella prosentteja kuin pelkkiä frekvenssejä.

Ristiintaulukkoa pääsee tekemään analyysivalikossa (Analyze/Descriptive statistics/Crosstabs). Näin saadaan näkyville seuraavalla sivulla oleva ikkuna:



Row(s) ja column(s) kohtaan valitaan muuttujat. Valikon alalaidassa on kolme painiketta, joiden toiminnot ovat seuraavat:

Statistics: täältä voi käydä ruksimassa kohdan Chi-square. Tällöin ohjelma testaa muuttujien välisen riippuvuuden tilastollisen merkitsevyyden.

Cells: tässä valikossa käydään valitsemassa rivi- ja sarakeprosentit sekä odotetut frekvenssit.

Format: täällä voidaan vaikuttaa hieman taulukon ulkoasuun.

Seuraavassa taulukossa (taulukko 3) on esitetty naisten tyytymättömyys johonkin omassa ulkonäössä koulutusasteen mukaan tarkasteltuna. Taulukossa on esitetty frekvenssit, riviprocentit ja sarakeprosentit. Taulukosta nähdään, että (1) perusasteen suorittaneista 43 prosenttia, keskiasteen suorittaneista 67 prosenttia ja korkea-asteen suorittaneista 79 prosenttia on johonkin ulkonäössään tyytymättömiä, (riviprocentit) ja (2) johonkin ulkonäössään tyytymättömistä 17 prosenttia on perusasteen suorittaneita, 63 prosenttia keski-asteen suorittaneita ja 20 prosenttia korkea-asteen suorittaneita (sarakeprosentit).

Taulukko 3. Tyytymättömyys johonkin omassa ulkonäössä koulutusasteen mukaan.

			Onko Teidän ulkonäössänne jotain, mihin ette ole tyytyväinen ja joka häiritseen Teitä?		
			Kyllä	Ei	Total
Koulutus 3 luokassa	Perusaste	Count	29	38	67
		% within Koulutus 3 luokassa	43,3%	56,7%	100,0%
		% within Onko Teidän ulkonäössänne jotain, mihin ette ole tyytyväinen ja joka häiritseen Teitä?	17,1%	38,4%	24,9%
	Keskiaste	Count	107	52	159
		% within Koulutus 3 luokassa	67,3%	32,7%	100,0%
		% within Onko Teidän ulkonäössänne jotain, mihin ette ole tyytyväinen ja joka häiritseen Teitä?	62,9%	52,5%	59,1%
	Korkea-aste	Count	34	9	43
		% within Koulutus 3 luokassa	79,1%	20,9%	100,0%
		% within Onko Teidän ulkonäössänne jotain, mihin ette ole tyytyväinen ja joka häiritseen Teitä?	20,0%	9,1%	16,0%
Total	Count	170	99	269	
	% within Koulutus 3 luokassa	63,2%	36,8%	100,0%	
	% within Onko Teidän ulkonäössänne jotain, mihin ette ole tyytyväinen ja joka häiritseen Teitä?	100,0%	100,0%	100,0%	

Seuraavassa taulukossa (taulukko 4) on esitetty absoluuttisten frekvenssien lisäksi rivi- ja sarakeprosentit sekä odotetut frekvenssit (expected count).

Taulukko 4. Vanhemmuus parisuhteen mukaan.

			Lapsia		
			Kyllä	Ei	Total
Vakituinen parisuhde	Kyllä	Count	173	45	218
		Expected Count	158,7	59,3	218,0
		% within Vakituinen parisuhde	79,4%	20,6%	100,0%
		% within Lapsia	87,4%	60,8%	80,1%
	Ei	Count	25	29	54
		Expected Count	39,3	14,7	54,0
		% within Vakituinen parisuhde	46,3%	53,7%	100,0%
		% within Lapsia	12,6%	39,2%	19,9%
Total	Count	198	74	272	
	Expected Count	198,0	74,0	272,0	
	% within Vakituinen parisuhde	72,8%	27,2%	100,0%	
	% within Lapsia	100,0%	100,0%	100,0%	

- Odotetut frekvenssit antavat vertailukohdaksi sellaiset frekvenssit, jotka syntyisivät, jos ristiintaulukon luokkien väliset erot johtuisivat pelkästä sattumasta.
- Soluissa olevista arvoista toista kutsutaan riviprosentiksi ja kolmatta sarakeprosentiksi.

- Prosentteja ei kannata laskea yllä olevan taulukon tyyliin joka suuntaan, sillä *taulukko on aina sitä helppolukuisempi, mitä vähemmän siinä on numeroita*. Kannattaa pelkistää. Tapana on valita selittävä muuttuja esim. riville, jolloin prosenteiksikin kannattaa valita riviprocentit. Eli ei niin kuin näissä aikaisemmissa taulukoissa on. Kuitenkin on makuasia, haluaako selittävän muuttujan riville vai sarakkeisiin, kunhan tulostaa taulukkoon selittävän muuttujan prosentit.
- Edellä esitetyt ristiintaulukot on liian iso esitettäväksi tutkimusraportissa, pitäisi karsia asiaa, sillä näin isoa taulukkoa on jo vaikeaa hahmottaa ja tulkita.

3.1.4 Useampiulotteiset frekvenssitaulukot

Useampiulotteinen frekvenssitaulukko tarkoittaa käytännössä sitä, että tuotetaan useita frekvenssitaulukkoja - erikseen kolmannen (ja neljännen jne.) muuttujan luokalle. Jos on vaikkapa taulukoitu ristiin kaksi muuttujaa, kolmannen ulottuvuuden lisääminen taulukkoon tarkoittaa sitä, että tuotetaan kaksiulotteiset taulukot kaikille kolmannen muuttujan luokille.

Esimerkiksi jos on taulukoitu ristiin koulutus ja tulot, voidaan tehdä vastaavat taulukot erikseen miehille ja naisille. Tekemällä kaikki nämä taulukot erikseen vaikkapa eri maille tai vuosille, on taulukossa (tai oikeastaan rykelmässä taulukoita) jo neljäs ulottuvuus. Jos käytössä on pienehkö aineisto, solufrekvenssit voivat olla tässä vaiheessa jo kovin pieniä... ja niistä on mahdotonta esittää yleispäteviksi tarkoitettuja päätelmiä.

3.1.5 Keskiarvotaulukot

Keskiarvotaulukot ovat melko lailla verrattavissa ristiintaulukkoon. Keskiarvotaulukoilla tarkastellaan jonkun muuttujan keskiarvoja toisen muuttujan luokissa.

Keskiarvotaulukon laatimista varten **selitettävän** muuttujan (muuttujan, josta lasketaan keskiarvot) tulee olla

- Numeerinen (eli mitattu joko suhde- tai välimatka-asteikolla)
- Joskus harvoin myös järjestysasteikollinen muuttuja käy, jos keskiarvoa pystyy mielekkäästi tulkitsemaan. Esim. läänien keskiarvo silloin, kun 1 eteläisin ja 6 pohjoisin voi olla jotenkin mielekäs, mutta jos läänit on numeroitu jotenkin muuten, ei ole mieltä ottaa niistä

keskiarvoa. Jos olet epävarma keskiarvon tulkinnasta, käytä mieluummin ko. muuttujaan jotakin muuta analyysimenetelmää.

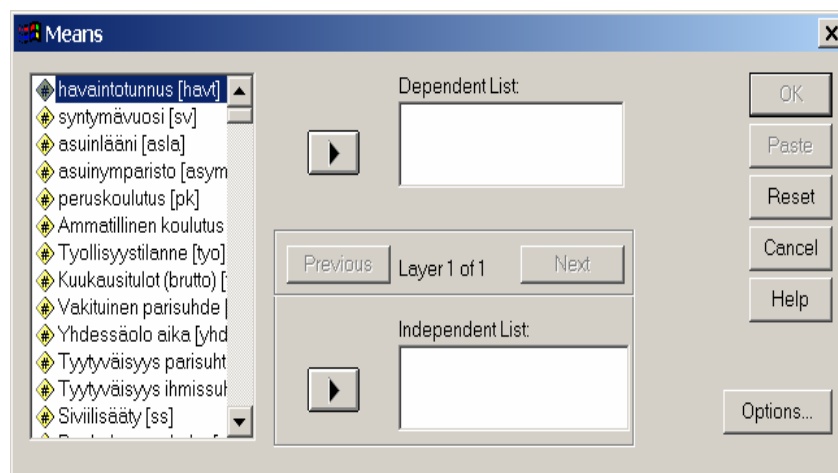
Selittävän muuttujan tulee olla

- luokiteltu, esim. ei ole mielekästä käyttää ikä-muuttujaa sellaisenaan, vaan se on syytä luokitella ikäluokkiin.

Keskihajonta on hyvä ottaa mukaan keskiarvotaulukoihin, koska sillä saadaan selville, kuinka paljon keskiarvoissa on vaihtelua. Lisäksi tutkittavien määrän tulee selvitä taulukosta.

Keskiarvotaulukkoa pääsee tekemään analyysivalikossa (Analyze/Compare means/Means).

Näin saadaan näkyville seuraava ikkuna:



Dependent list –kenttään valitaan selitettävä muuttuja ja Independent list –kenttään selittävä muuttuja.

Options: täällä voi valita mitä tilastollisia tunnuslukuja haluaa taulukossa esitettävän.

Taulukossa 5 on esitetty vastaajien käsitys omasta ulkoisesta viehättävyydestä koulutusasteen mukaan tarkasteltuna. Käsitys omasta viehättävyydestä muuttuja on koodattu siten, että mitä suurempi arvo, sitä viehättävämpänä itseään pitää. Eli mitä keskiarvotaulukon perusteella voidaan sanoa? Kyllä vain, keski- ja korkea-asteen suorittaneet naiset pitävät omaa ulkonäköään perusasteen koulutuksen saaneita naisia viehättävämpänä.

**Taulukko 5. Käsitys omasta viehättävyydestä
koulutusasteen mukaan.**

Koulutus 3 luokassa	Mean	N	Std. Deviation
Perusaste	3,4635	69	,68855
Keskiaste	3,7442	159	,69628
Korkea-aste	3,7070	43	,73824
Total	3,6669	271	,70868

3.1.6 Korrelaatio

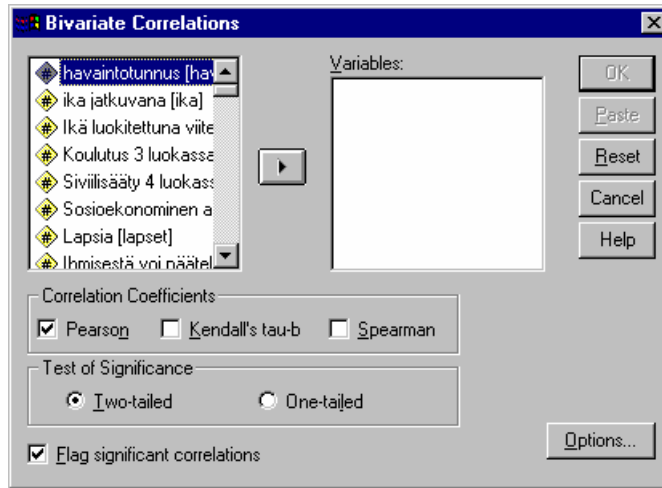
Korrelaatiotermiä käytetään yleiskielessä kuvaamaan kahden asian välistä yhteyttä. Voidaan esimerkiksi sanoa, että asuinpaikka korreloi sydän- ja verisuonitautien määrään. Tilastotieteellisessä mielessä korrelaatiolla on suppeampi merkitys. Korrelaatioilla tarkoitetaan tällöin yleensä numeeristen muuttujien lineaarista yhteyttä kuvaavaa, (Pearson) korrelaatiokerrointa. Korrelaatiokerroin saa arvoja väliltä [-1, 1].

Positiivinen korrelaatio merkitsee, että toisen muuttujan arvojen kasvaessa myös toisen muuttujan arvot kasvavat. Samoin havainnolla, joka saa pienen arvon toisella muuttujalla, on taipumus saada pieni arvo myös toisella. Lähellä nollaa oleva korrelaatio kertoo, ettei muuttujien välillä ole (juurikaan) lineaarista yhteyttä. Tällöin toinen muuttuja voi saada mitä tahansa arvoja toisen muuttujan arvojen kasvaessa. Korrelaatiokerroin saa negatiivisia arvoja silloin, kun toisella muuttujalla on taipumus saada suuria arvoja toisen saadessa pieniä.

Kahden tekijän välistä korrelaatiota havainnollistetaan hajonta- eli sirontakuviin. Tottunut lukija pystyy arvioimaan korrelaation hajontakuvion perusteella. Hajontakuvat paljastavat myös lineaarisesta poikkeavia yhteyksiä muuttujien välillä. Tähän korrelaatiokerroin ei pysty.

Korrelaatioanalyysin tekeminen

Korrelaatioanalyysiä pääsee tekemään analyysivalikossa (Analyze/Correlate/Bivariate) . Näin saadaan näkyville seuraavalla sivulla oleva ikkuna:



Muuttujat, joiden väliltä halutaan laskea korrelaatiokerroin, valitaan Variables-laatikkoon. Ohjelma tulostaa kaikkien muuttujaparien välisen korrelaation eli koko korrelaatiomatriisin.

Taulukossa 6 on esitetty nykyisen painoindeksin ja ihannepainoindeksin korrelaatio. Siitä nähdään, että näiden muuttujien välillä on vahva positiivinen korrelaatio.

Taulukko 6. Nykyisen painoindeksin ja ihannepainoindeksin korrelaatio.

		Painoindeksi	Ihannepainoindeksi
Painoindeksi	Pearson Correlation	1	,656**
	Sig. (2-tailed)	,	,000
	N	269	245
Ihannepainoindeksi	Pearson Correlation	,656**	1
	Sig. (2-tailed)	,000	,
	N	245	249

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

3.1.7 Taulukointi

- Vaikka taulukot olisikin esitetty mahdollisimman selkeästi, ne ovat silti melko hitaita lukea. Useimmiten niiden sisältö saadaan esitettyä pylväskuvioina niin, että nopea vilkaisu selvittää lukijalle mistä on kyse.
- Taulukkojen etuna on se, että niiden avulla voidaan parantaa monimutkaisten tulosten lukemista ja vielä melko pienessä tilassa.
- Käytännössä taulukon tulisi mahtua tutkimusraportissa kokonaan yhdelle sivulle, tätä isommat taulukot voidaan esittää erillisessä liitteessä.

Taulukoiden käytön ehtoja

Frekvenssitaulukoiden etu on sen riippumattomuus mitta-asteikosta. Ristiintaulukoilla voidaan kuvata myös laatueroasteikollisten muuttujien välisiä yhteyksiä. Esimerkiksi otokseen kuuluvien ihmisten jakautumista siviilisäätyluokkiin sosiaaliluokan mukaan (ja kolmiulotteisessa taulukossa lisäksi vaikkapa sukupuolen mukaan, jolloin kaksiulotteinen taulukko tehdään miehille ja naisille erikseen).

Jos jatkuvia tai tarkkaan luokiteltuja muuttujia halutaan esittää taulukossa, muuttujat kannattaa **luokitella ja luokkia yhdistellä** ennen taulukointia. Ei esimerkiksi ole järkevää taulukoida ristiin tuloja sukupuolen mukaan, kun tulot on mitattu euron tarkkuudella. Miesten ja naisten jakautuminen järkevällä tavalla muodostettuihin tuloluokkiin on sen sijaan kiinnostavaa.

Liian moniluokkaisten muuttujien käyttämistä pitää välttää, etenkin jos kyseinen taulukko esitetään raportissa. Tämä koskee sekä selittävää että selitettävää muuttujaa. Jos muuttujalla on liian monta luokkaa, solufrekvenssit jäävät usein liian pieniksi, jotta niistä voitaisiin päätellä mitään. Taulukoita varten aineistoa kannattaa tiivistää.

Jos on tarvetta esittää suurempia taulukoita, joiden sisältö ei kuulu tutkimuksen keskeisiin tuloksiin, taulukot voi ehkäpä sijoittaa raporttiin liitteeksi. ("Tutkimusväestön jakautuminen analyysissä käytettyjen selittävien muuttujien luokkiin" - selittäjiä on kymmeniä - tai "Perustietoa Lapin läänin kunnista".)

Ristiintaulukoinnin ongelmaksi voi muodostua tarvittavan **tutkimusaineiston määrä**. Laajojen ja moniulotteisten ristiintaulukoiden tekemiseen tarvitaan erittäin paljon havaintoja.

3.1.8 Ohjeita taulukoiden muotoiluun

Pääset muokkaamaan SPSS -ohjelman Output-ikkunassa olevaa taulukkoa tuplaklikkaamalla sitä. Huom! Et pääse mihinkään alla mainittuihin toimintoihin käsiksi, jollet ensin tee sitä, koska tarvittavat valikot käskyineen ilmaantuvat Output-ikkunan valikkoriville vasta taulukon tuplaklikkaamisen jälkeen. Vaihtoehtoisesti voit muokata taulukkoa klikkaamalla hiirellä taulukon eri osia.

Keskeisimmät toiminnot

Koko taulukkoon vaikuttavia muotoiluja (kaikkien tekstien fontti, tekstin asettelu soluissa, solujen rajoja) pääsee muokkaamaan **Format/Table properties** -valintaikkunassa.

Tekstien muotoilu

Taulukon tekstejä (otsikkoa, lukuja, muuttujien nimiä) voi muuttaa tuplaklikkaamalla ko. solua tai kohtaa ja kirjoittamalla korjatun tekstin aikaisemman tilalle.

Yksittäisiä taulukon osia pääsee muotoilemaan valitsemalla hiirellä ("maalamalla") ne osat, joita haluaa muuttaa ja valitsemalla **Format/Cell Properties**. Useita erillisiä solujahan saa tarvittaessa valittua kerralla pitämällä ctrl-näppäintä pohjassa valinnan ajan. Koko taulukon saa kätevästi valittua **Edit/Select Table**.

Prosenttimerkit pois, oikea määrä desimaaleja.

Format/Cell Properties/Value: Format-listasta valitaan oikean näköinen muoto.

Muuttujan nimi taulukon vasempaan ylänurkkaan

Valitse hiirellä oikea kohta ja kirjoita siihen muuttujan nimi. Ikävä yllätys, teksti ei kuitenkaan näy taulukossa. Tämä on SPSS:n yksi harmillinen erikoisuus. Jotta saat muuttajan nimen näkyviin valitse kyseinen solu hiirellä, paina hiiren oikeaa painiketta, ja tämän jälkeen valitse toiminto **Show Dimension**. Tällöin teksti tulee taulukossa näkyviin.

Lähdeviitteen lisääminen

Insert/Caption

Alaviitteen lisääminen

Insert/Footnote

Osien tuhoaminen

Valitaan tuhottava osa hiirellä ja painetaan delete-näppäintä.

Rivit sarakkeiksi ja toisinpäin

Pivot/Transpose Rows and Columns

Muuta...

Muitakin temppuja taulukolle voi tehdä. Kokeile **Format** -valikosta löytyviä käskyjä ja katso mitä tapahtuu. Muistathan, että **Edit** -valikosta löytyy tuo aina niin tarpeellinen **Undo**-käsky.

3.1.9 Taulukoiden raportointi

Taito esittää selkeitä taulukoita on tärkeä ja hallussa harvalla. Tässä ohjeita etenkin esittämiskelpoisen ristiintaulukon laadintaan:

- Esitetään vain rivi-, sarake- tai soluprosentit (ja tietenkin se oikea vaihtoehto on valittu!), eli kussakin solussa on vain yksi prosenttiluku (mutta ei %-merkkiä). Solufrekvenssejä ei yleensä ole tarpeen esittää.
 - Esitetään myös luvut, joista prosentit on laskettu, eli rivi- ja sarakesummat ja niiden summa. Jos lukija on kiinnostunut solufrekvensseistä eikä niitä ole esitetty, hän voi laskea ne prosenttien ja rivi- tai sarakeprosenttien avulla.
 - Yksi hyvä tapa varmistaa, että prosentit on laskettu oikein päin, on koittaa sanallisesti selittää taulukon keskeistä antia. Esimerkki lukemisharrastusta koskevasta tutkimuksesta: "Narujärveläisistä aikuisista 88 % ja uppojokelaisista aikuisista vain 32 % lukee vähintään kolme kaunokirjallista teosta vuodessa." (Olisi oudompaa tutkia, kuinka suuri osa kaunokirjallisuuden lukijoista on uppojokelaisia.)
 - Ei esitetä liikaa desimaaleja; ne vain haittaavat luettavuutta.
 - Esitetään prosenttien summa eli rivi satasia joko ala- tai oikeanpuoleisessa marginaalissa riippuen prosenttien laskusuunnasta.
 - Muuttujien luokkien sisältö on kuvattava selkeästi. Siis pelkät luokkien numerot eivät käy, vaan kullekin luokalle tulisi olla itse taulukossa sanallinen selitys. Tämä onnistuu helposti varmistamalla ennen taulukoiden tekoa, että kaikille käytetyille muuttujille on annettu aineistoikkunassa arvojen selitystekstit (*variable labels*). Jos luokkien kuvaukset eivät mahdu taulukon riveille ja sarakkeille, selostusta jatketaan alaviitteissä.
 - Taulukko on otsikoitava selkeästi ja tarvittaessa lisättävä sisältöä kuvaavia alaviitteitä.
- Lukijan on pystyttävä ymmärtämään taulukon sisältö etsimättä selityksiä raportin varsinaisesta tekstistä.** Tieteellisessä raportissa otsikon ei välttämättä tarvitse olla kovin

raflaava tai iskevä. Esimerkki otsikosta: Suomalaisten 18 - 35-vuotiaiden naisten jakautuminen tuloluokkiin vuonna 1997.

- Lähde merkitään usein taulukon alalaitaan. Ja parhaana vinkkinä: Näytä taulukoitasi työkaverillesi tai ystävällesi ja pyydä häntä kertomaan, mitä taulukoissa kuvataan ja mitä luvut osoittavat.
- Taulukon lukujen ja yksityiskohtien liiallista toistoa tekstissä tulee välttää; se on varma tapa saada lukijan mielenkiinto herpaantumaan.
- Taulukolla on otsikon lisäksi oltava numero. Taulukon tulee olla itsenäisesti ymmärrettävä. Ja siksi otsikosta tulisi tulla ilmi, mistä taulukosta on kyse.
- Jos on kyse prosenttilukusta, tulee havaintojen lukumäärä laittaa esille.

3.1.10 Lopuksi: siis mitä tehdä, kun

Yksi muuttuja, joka

- a. jatkuva: valitse descriptives (Analyze/Descriptive Statistics/Descriptives), jolloin voit pyytää SPSS:ää tulostamaan muuttujan keskiarvon, keskihajonnan, minimi- ja maksimiarvon, vaihteluvälin, varianssin, summan, vinouden (*skewness*) ja huipukkuuden (*kurtosis*).
- b. luokiteltu: valitse frekvenssit (Analyze/Descriptive Statistics/Frequencies), jolloin SPSS tulostaa suoran jakauman.

Kaksi muuttujaa, joista

- a. molemmat luokiteltuja muuttujia: valitse ristiintaulukko.
- b. selitettävä jatkuva muuttuja ja selittävä luokiteltu muuttuja: valitse keskiarvotaulukko.
- c. selittävä ja selitettävä muuttuja jatkuvia muuttujia: korrelaatio.

3.2 Graafinen esittäminen

Ohjelmistojen käytön helpottuessa tilastoaineistojen graafinen kuvailu on yleistynyt. Kuvalla voi kertoa nopeasti tarvittavan tiedon lukijalle. Tilastografiikkaa voidaan ryhmitellä erilaisten kuvatyyppejen mukaan. Yleisimpiä kuvatyyppejä ovat pylväät, hajontakuviot, viiva- ja aluekuviot sekä piirakat.

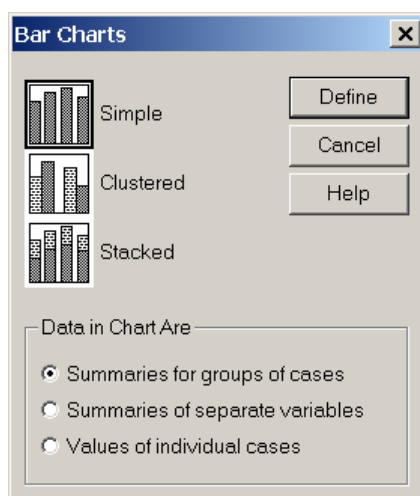
Kuvatyyppiä valitessa on tärkeää miettiä, mitä kuvan on **tarkoitus viestiä** lukijalle sekä sitä, **millainen kuva viestii asian parhaiten**. Kaikki kuvatyyppit eivät sovellu kaikkien asioiden esittämiseen.

KUVATYYPIT

3.2.1 Pylväät

Pylväät ovat yleisin, ja usein havainnollisin tilastografiikan kuvaajatyyppejä. Ne soveltuvat monen erilaisen asian esittämiseen. Pylväät voidaan piirtää pysty- tai vaakasuoraan. Frekvenssit, prosentuaaliset osuudet, keskiarvot ym. voidaan kirjoittaa näkyviin pylväiden yhteyteen tai piirtää pylväät koordinaatistoon, josta ne voidaan lukea.

Pylväitä päästään tekemään valinnalla **Graphs/Bar**. Näin saadaan näkyville seuraava ikkuna:



Tässä valikossa määritetään se, (1) millaisia pylväskuvioita halutaan tehdä ja (2) mitä pylväät kuviossa esittävät.

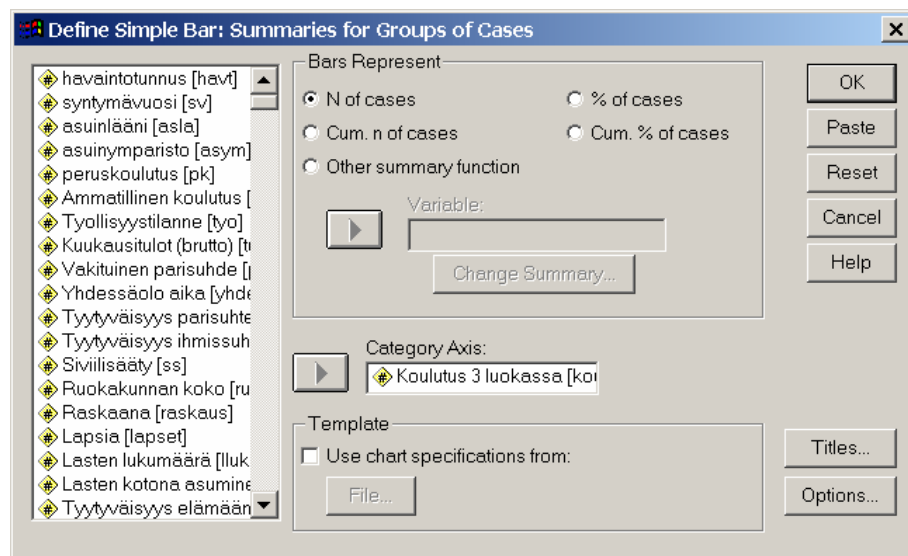
(1) Pylväskuvio vaihtoehtoja on kolme: Simple, Clustered ja Stacked. Oletusvalinnalla Simple voidaan tehdä frekvenssipylväitä ja esimerkiksi keskiarvopylväitä. Valinnoilla Clustered ja Stacked voidaan tarkastella kahden tai kolmen muuttujan pylväskuvioita.

(2) Data in Chart are –kohdassa voidaan valita, mitä pylväät kuviossa esittävät. Summaries for groups of cases: pylvään korkeus määrittää havaintojen tunnusluvun ko. luokassa. Tämä on oletusmenetelmä, ja usein parhaiten sopiva esittämistapavaihtoehto. Seuraavassa esimerkissämme käytämme sitä. Muut vaihtoehdot ovat: Summaries of separate variables, tällöin pylvään korkeus kuvaa muuttujan tunnuslukua ja Values of individual cases, tällöin pylvään korkeus esittää muuttujan arvoa.

A YHDEN LUOKITELLUN MUUTTUJAN PYLVÄSKUVIO

Frekvenssiympylvällä esitetään erilaisten osaryhmien kokoja. Pylväät ovat useimmiten havainnollisin tapa esittää tällaisia asioita, vaikka journalistit yleensä leipovatkin niistä piirakoita. Yhden muuttujan frekvenssiympylväs on samannäköinen kuin histogrammi (paitsi että pylväiden välissä on tyhjää esittämässä, että kyse on ei-jatkuvasta eli luokitellusta muuttujasta).

Frekvenssiympylvää pääsee tekemään kuvavalikosta (Graph/Bar) oletusvalinnalla Simple . Näin saadaan näkyville seuraava ikkuna:



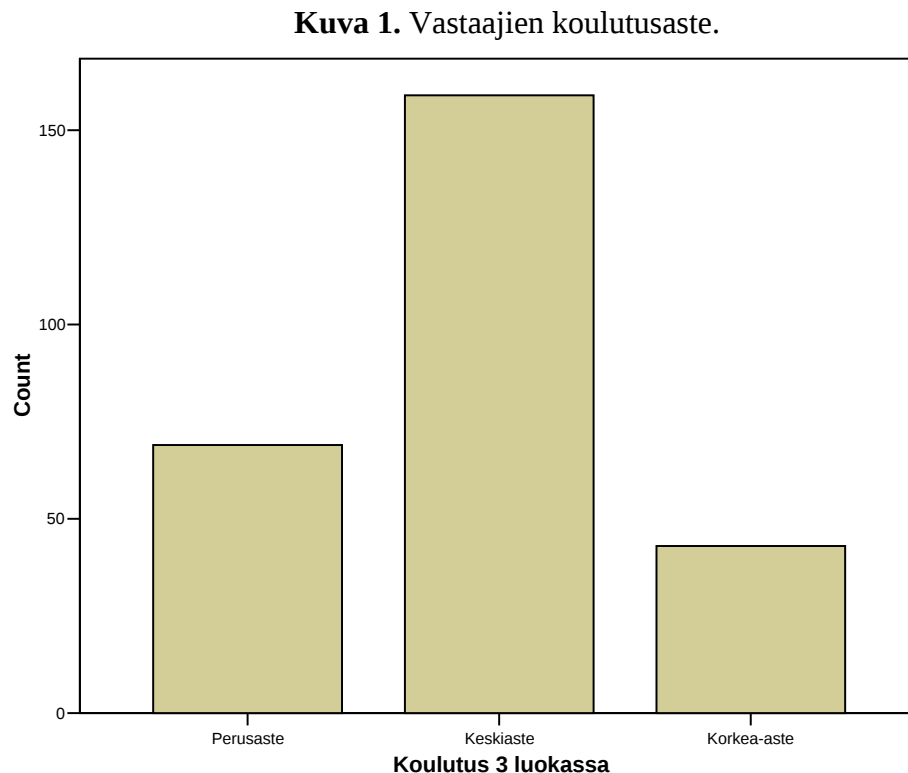
Category Axis- kohtaan valitaan muuttuja, josta halutaan tehdä frekvenssiympylväs.

Titles: täällä voi kirjoittaa kuvalle otsikon, otsikon voi kuitenkin kirjoittaa myös kuvankäsittelyikkunassa tai vasta siinä tiedostossa, johon kuva liitetään (tätä suosittelem).

Options: tässä valikossa voi määrittää, ettei kuvassa esitetä puuttuvia tietoja, tämänkin voi tehdä myöhemmässäkin vaiheessa, kuvankäsittelyikkunassa.

Bars represent –kenttä: tämä kenttä on keskeinen, sillä tässä voi valita, haluaako kuvassa esitettävien pylväiden esittävän lukumääriä, prosentuaalisia osuuksia etc. Frekvenssipylvästä tehdessä ohjelman oletusarvoa, N of cases, ei tarvitse vaihtaa.

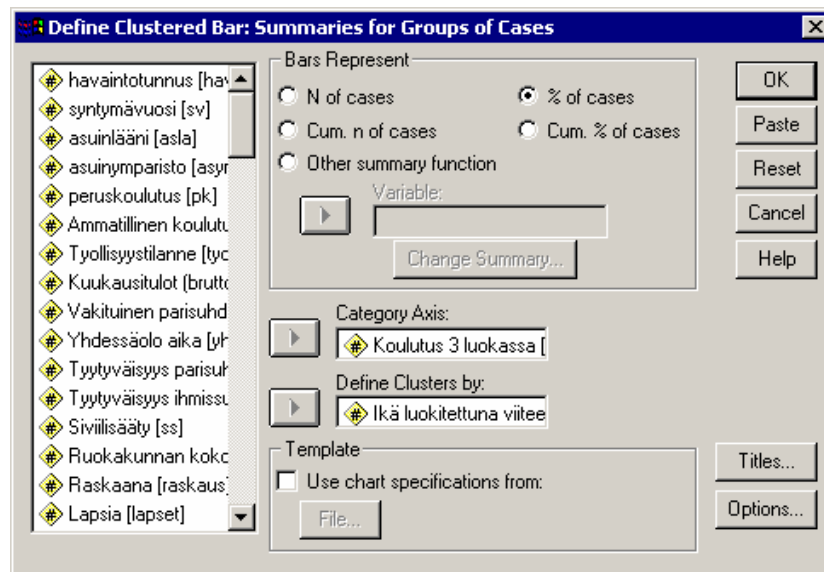
Kuvassa 1 on frekvenssipylväs, jossa on esitetty vastaajien koulutusaste. Jotta kuva olisi esittämiskelpoinen, tulisi kuvaa muokata kuvankäsittelyikkunassa (ainakin count-sana täytyy suomentaa).



B KAHDEN LUOKITELLUN MUUTTUJAN PYLVÄSKUVIOT

Kahden luokitellun muuttujan pylväskuvioita voidaan tehdä joko **ryhmittelemällä** pylväitä (Clustered) tai **pinoamalla** niitä (Stacked).

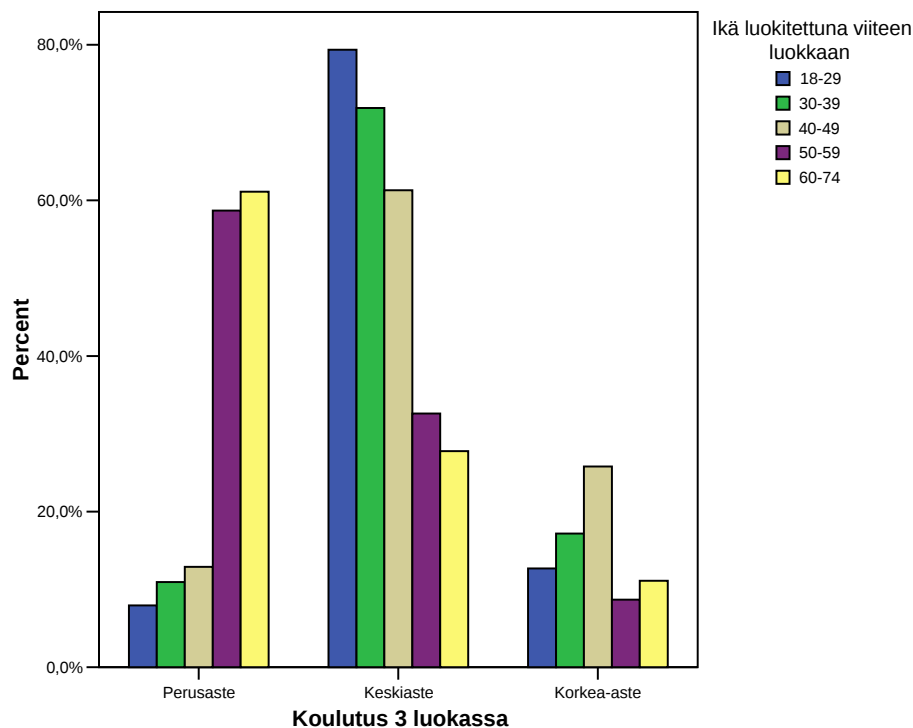
Useamman muuttujan pylväskuvioita pääsee tekemään kuvavalikosta (Graph/Bar) oletusvalinnalla Clustered tai Stacked. Näin saadaan näkyville seuraavalla sivulla oleva ikkuna:



Ikkuna on samanlainen kuin Simple-valinnan ikkuna. Ainoa ero on se, että tässä ikkunassa on varattu myös toiselle muuttujalle kenttä (Define clusters by).

Seuraavassa on esitetty vastaajien koulutusaste ikäluokittain tarkasteltuna (%). Mieti, mitä kuvasta nähdään?

Kuva 2. Naisten koulutusaste ikäluokittain tarkasteltuna (%).

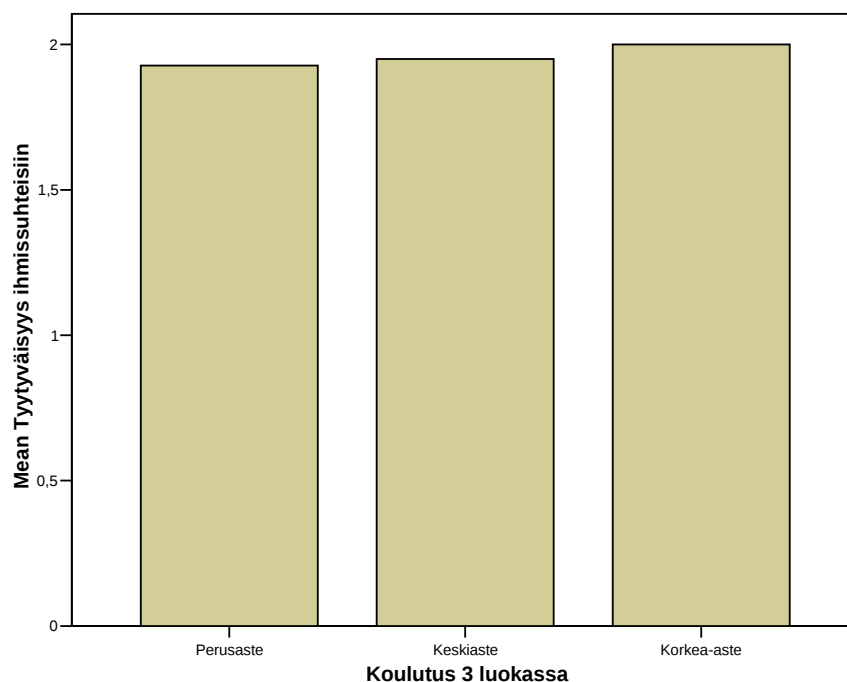


C KESKIARVOPYLVÄÄT

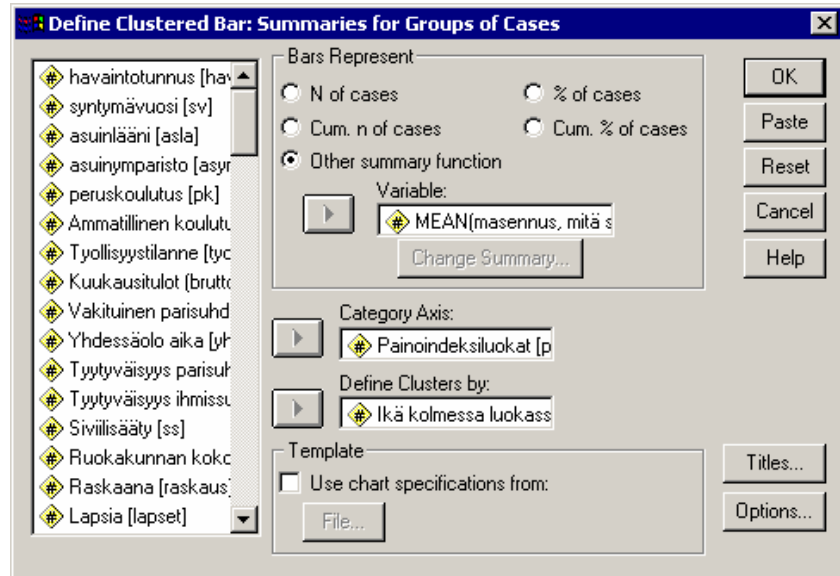
Keskiarvopylväillä esitetään jonkin muuttujan keskiarvoja joko (1) yhden tai (2) kahden muuttujan määrittelemissä ryhmissä. Seuraavassa esitetään, miten näitä tehdään. Keskiarvopylväät ovat yksinkertainen tapa tiivistää tietoa hyvin helposti havainnoitavaan muotoon. Selitettävän muuttujan ainakin periaatteessa pitäisi olla mitattu joko suhde- tai välimatka-asteikolla.

(1) Yhden muuttujan määrittelemissä ryhmissä keskiarvopylväitä pääsee tekemään samassa valikossa, jossa tehtiin frekvenssipylväs (eli valinnoilla Graph/Bar/Simple). **Category axis** – kohtaan valitaan selittävä eli luokiteltu muuttuja, jonka ryhmissä halutaan tarkastella selitettävän muuttujan keskiarvoja. Keskiarvopylväitä tehdessä tulee **Bars represent** –kentästä ruksia kohta **Other summary function**, jonka jälkeen selitettävä muuttuja sijoitetaan **Variable** –kohtaan. Ohjelman oletusarvona on, että tällöin halutaan pylväiden esittävän keskiarvoja jonkun valitun selitettävän muuttujan ryhmissä. Muitakin vaihtoehtoja olisi valittavana painikkeen **Change summary** takaa. Seuraavassa kuvassa (kuva 3) on esitetty vastaajien tyytyväisyys elämään koulutusasteen mukaan. Tyytyväisyyttä ihmissuhteisiin on kysytty 5-portaisella Likert-asteikolla. Vastaukset on koodattu siten, että arvon kasvaessa tyytyväisyys ihmissuhteisiin vähenee. Mitä siitä siis voimme nähdä? Onko kuva sinusta järkevä?

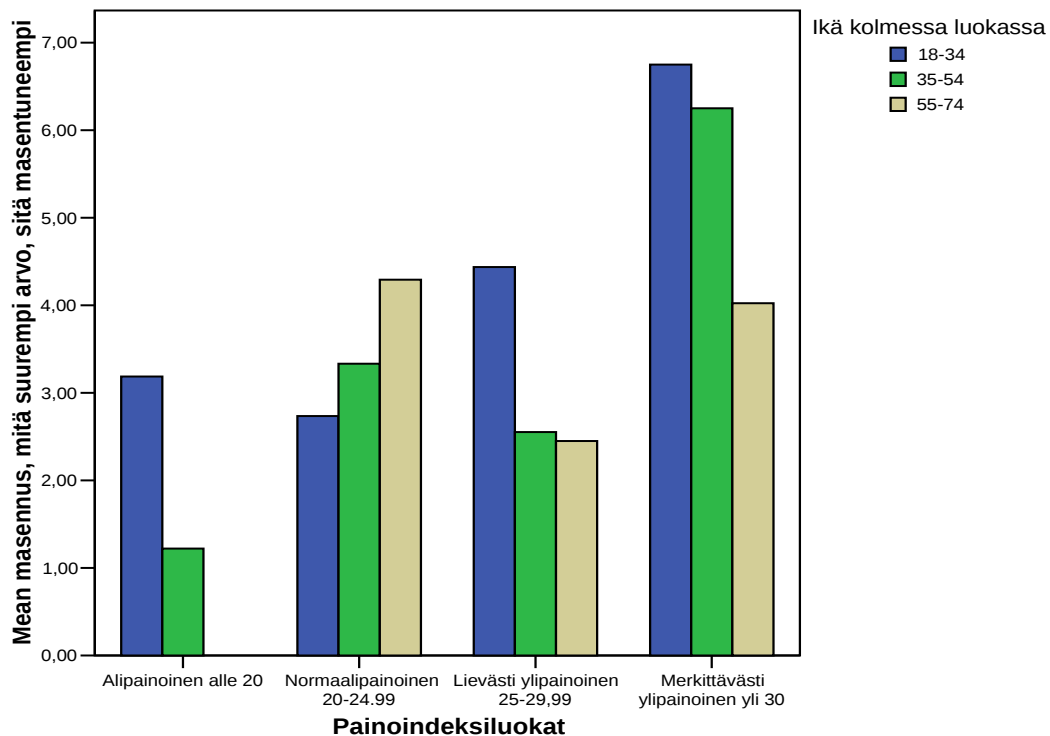
Kuva 3. Tyytyväisyys elämään koulutusasteen mukaan.



2) Kahden muuttujan määrittelemisessä ryhmissä keskiarvopylväitä pääsee tekemään samassa valikossa, jossa tehtiin kahden luokitellun muuttujan pylväskuvioita (eli valinnoilla Graph/Bar/Clustered tai Stacked). Näiden keskiarvopylväiden tekeminen on muutoin samanlainen prosessi kuin yhden muuttujan määrittelemisessä ryhmissä keskiarvopylväiden tekeminen. Ainoa poikkeus on se, että toinen luokiteltu muuttuja valitaan mukaan ja se sijoitetaan Define clusters by -valikkoon, kuten seuraavasta ikkunasta nähdään:



Kuva 4. Masentuneisuus ikäluokittain eri painoindeksiluokissa tarkasteltuna.

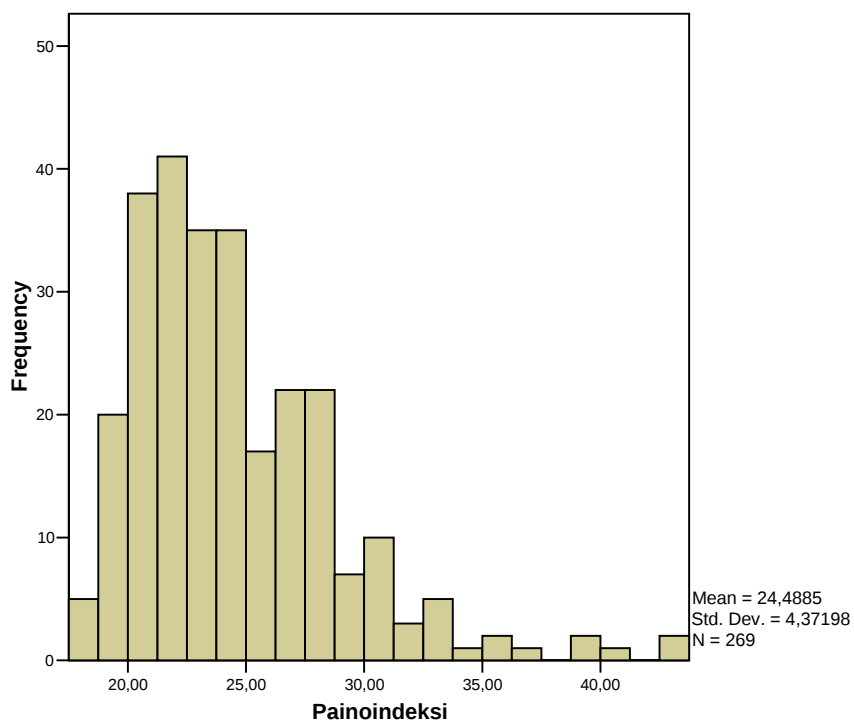


Edellisellä sivulla oleva kuva 4 on tehty keskiarvokuvan valintaikkunassa tehdyillä valinnoilla. Mitä siitä on nähtävissä? Onko ikä tai painoindeksi tai molemmat yhteydessä naisten masentuneisuuteen?

3.2.2 Histogrammit

Histogrammit ovat yksinkertaisia pylväitä, jotka soveltuvat yhden muuttujan jakauman kuvailuun. Pylvään korkeus ilmaisee pylvään luokkaan kuuluvien havaintojen määrän. Histogrammilla voidaan esittää jatkuvien muuttujien jakaumia (SPSS -ohjelma osaa yleensä luokitella muuttujia jokseenkin järkevästi. Mieti kuitenkin aina itse onko luokittelu järkevä.) SPSS:ssä histogrammi tehdään valinnoilla **Graphs/Histogram**, erityisiä lisämääreitä ei yleensä tarvita. Kuvassa 5 on esitetty vastaajien painoindeksin histogrammi.

Kuva 5. Naisten painoindeksi.



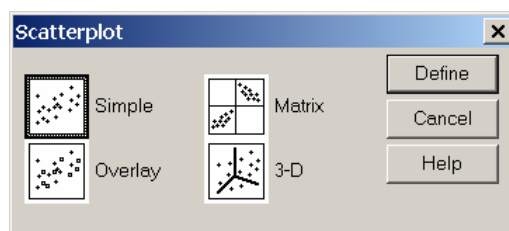
3.2.3 Hajontakuviot

Hajontakuvioilla (scatter) esitetään **kahden jatkuvan muuttujan yhteisjakaumat**. Hajontakuviolla tarkoitetaan muuttujien piirtämistä samaan koordinaatistoon siten, että toisen muuttujan asteikko asettuu x-akselille ja toisen muuttujan y-akselille. Havainnot asettuvat pisteinä tähän koordinaatistoon mittausarvojen mukaisille koordinaattipisteille.

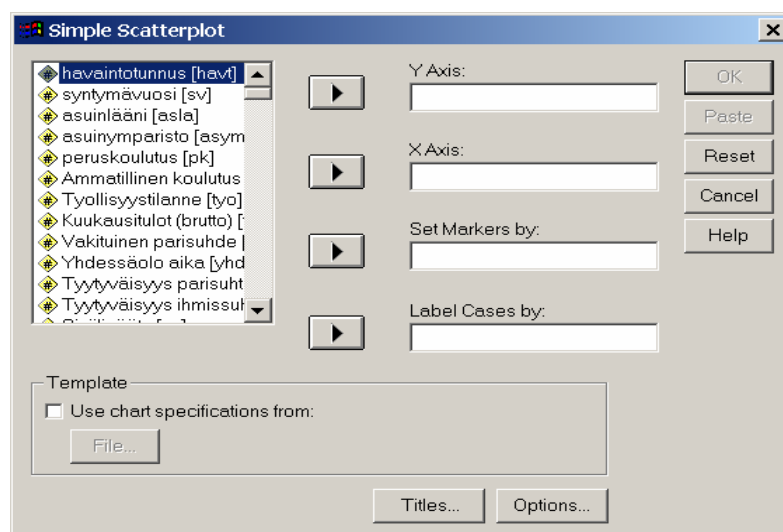
Koordinaatistosta voidaan tutkia muuttujien yhteisvaihtelua, jonka seurauksena havaitaan, onko muuttujilla keskinäistä riippuvuutta. Riippuvuuksien näkemiseen kuvasta tulee harjaantua, joten useimmiten hajontakuvioita tehdään tutkijan omaan käyttöön, sellaisina ne ovat hyvin hyödyllisiä aineistoon tutustumisessa.

Kun hajontakuvioissa ovat pisteet täysin hajallaan, muuttujien välillä ei ole riippuvuutta. Hajontakuvioita tehdessä voidaan periaatteessa valita muuttujien x ja y akselit kummin päin tahansa, mutta jos on luonnollista ajatella toisen muuttujan selittävän toisen muuttujan käyttäytymistä valitaan selittävän muuttujan akseliksi vaaka-akseli. Erilaisilla hajontakuvion päälle piirretyillä regressiokäyrillä voidaan mm. osoittaa riippuvuuksia myös harjaantumattomille. Hajontakuvioden laatiminen auttaa tekemään ylimalkaisia johtopäätöksiä muuttujien välisistä yhteyksistä. Tämä ei kuitenkaan yleensä riitä ja siksi tehdään jatkoanalyysyjä, joiden avulla muuttujien mahdollinen riippuvuus voidaan esittää numeerisessa muodossa.

Hajontakuvioita pääsee tekemään kuvavalikosta (Graph/Scatter). Näin saadaan näkyville seuraava ikkuna:



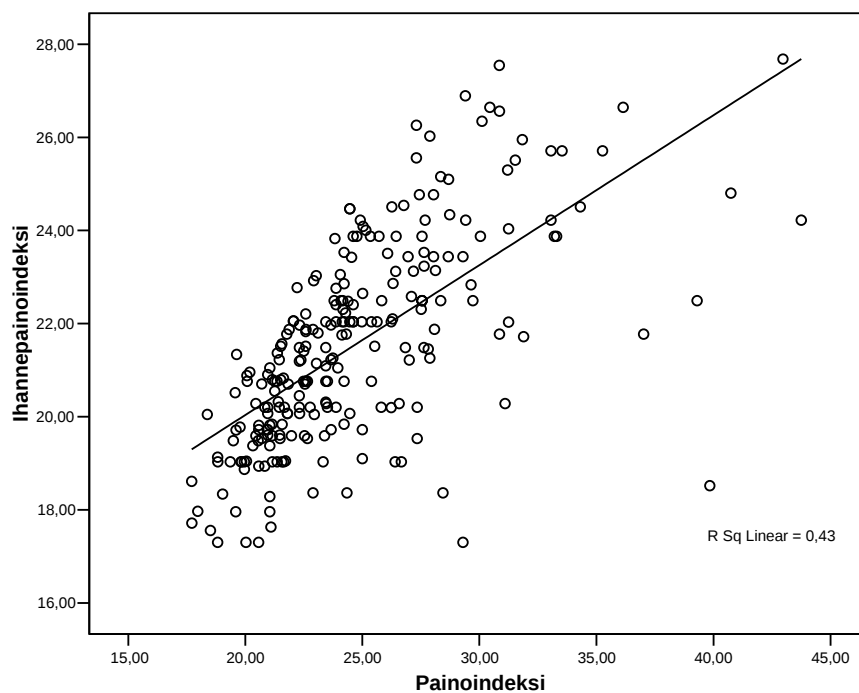
Valinnalla **Simple**, saadaan seuraava ikkuna näkyville:



Y ja X Axis –kenttiin valitaan muuttujat, joiden yhteisvaihtelusta ollaan kiinnostuneita.

Kuvassa 6 on hajontakuvio, jossa on esitetty naisten nykyisen painoindeksin ja ihannepainoindeksin yhteisjakauma. Siitä nähdään, että näiden muuttujien välillä on riippuvuutta: Nykyisen painoindeksin noustessa ihannepainoindeksi nousee.

Kuva 6. Nykyisen painoindeksin ja ihannepainoindeksin hajontakuvio.



3.2.4 Viiva- ja aluekuviot

Viivakuvioilla kuvataan yleensä aikasarjoja. Viiva antaa vaikutelmat siitä, että sitä pitkin voi liikkua, niinpä sen kuvaaman muuttujan tulisi olla intervalliasteikollinen ja periaatteessa jatkuva. Aluekuvioilla kokonaisalueen pinta-alan vaihtelun aika (tms.) -akselilla tulisi kuvata jonkin asian esiintymisen muutosta. Viivakuvioita pääsee tekemään valikosta Graph/Line ja aluekuvioita valikosta Graph/Area.

3.2.5 Piirakat

Piirakkakuvioilla (Pie) on tapana havainnollistaa suhteellisia eli prosentuaalisia osuuksia siten, että kunkin ympyrän sektorin pinta-ala kuvaa tämän sektorin suhteellista osuutta koko aineistossa.

Jotakin luokkaa voidaan erityisesti korostaa irrottamalla tämän luokan sektori hieman erilleen muusta kuviosta. Kahden piirakan vertailu on vaikeaa elleivät erot ole selvät. Parhaiten piirakka sopiikin kuvaamaan jonkin ominaisuuden jakautumista yhdessä joukossa.

3.2.6 Kuvat niiden käyttötarkoituksen mukaan

Tässä tutustutaan tilastografiikkaan sen erilaisten käyttötarkoitusten mukaan, eikä kuvan ominaisuuksien mukaan, kuten edellä.

Yhden muuttujan jakauman kuvailu

Yhden muuttujan jakauman kuvailuun soveltuvat histogrammit (jatkuville muuttujille), pylväskuviot SPSS:ssä (tämänlainen kuva tilataan valitsemalla ensimmäisessä valintaikkunassa "simple" ja muutenkin toiminta on varsin simppeliä) ja viiksikuviot, sekä huonommin piirakat.

Useiden muuttujien kuvailu

Joskus useita muuttujia on hyödyllistä kuvailla samassa kuvassa vertailun helpottamiseksi. Tämä edellyttää useimmiten, että muuttujilla on samanlainen asteikko. Usein taas halutaan samanlaisia, mutta erillisiä kuvaajia useista muuttujista liukuhihnatyönä. SPSS:ssä on monissa kuvaajatyypeissä mahdollisuus valita useiden muuttujien kuvailu. Valikkoja käytettäessä valinta tapahtuu rastimalla ruutu "summaries of several variables" ensimmäisessä valintaikkunassa kuvaajaa tehtäessä. Huom. Kahden muuttujan kuvailua pystyy tekemään myös valitsemalla "summaries for groups of cases". Useita samanlaisia kuvia saa nopeimmin tehtyä SPSS:n komentomuodossa. Ensimmäinen kuvaaja määritellään valmiiksi, ja painetaan `paste` -nappulaa. Tämä kopioi kuvaajan määrittelyt komentomuotoisina syntaksi-ikkunaan. Valintaikkunassa komennot voi *copy-paste* toiminnoilla monistaa ja tehdä niihin haluttuja muunnoksia.

Muuttujan kuvailu toisen luokissa eli yhteisjakauman kuvailu

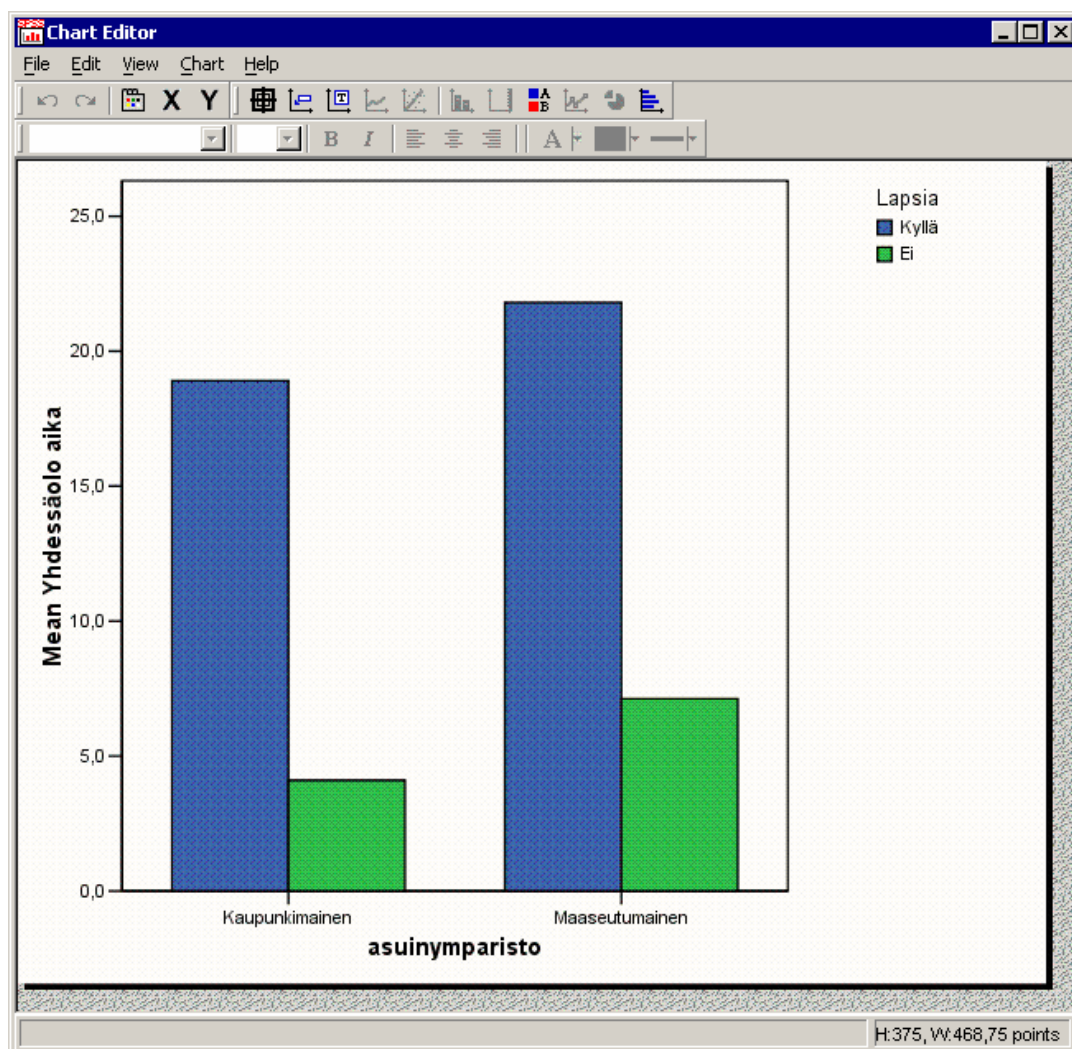
- Frekvenssi-, ristiin- ja keskiarvotaulukoiden kuvalliset vastineet esitetään pylväinä
- Jatkuvien muuttujien yhteisjakaumat hajontakuvina

3.2.7 Kuvien muokkaus SPSS:n kuva-editorilla

SPSS:n kuvaeditorilla voi tehdä monenlaisia temppuja kuville. Kuvatekstejä ja otsikoita voi muokata, vaihtaa esityksen värejä, tuhota osia, muuntaa skaalausta, kääntää kuvioita ja jopa muuntaa piirakan pylvääksi.

Ohjeita kuvaeditorin käyttöön

Kuva-editoriin (*Chart Editor*) pääset tuplaklikkaamalla tulostusikkunassa (*Output*) ko. kuvan päällä. Tällöin kuvasta muodostuu kopio kuvankäsittelyikkunaan, jota voi muokata hiiren (voi olla useissa tapauksissa helpompi tapa muokata kuvaa) tai valikkojen avulla. Näin saadaan näkyville seuraava ikkuna eli kuvankäsittelyikkuna:



Tekstin muotoilu

Olemassa olevaa tekstiä pääset muokkaamaan klikkaamalla hiirellä ko. tekstiä niin että sen ympärille tulee kehykset. Mikäli tuplaklikkaat hiirellä tekstiä, ohjelma avaa kyseistä kenttää koskevan **Properties**-valikon, jossa voi vaikuttaa kyseisen kentän asuun esim. fontin kokoon, väriin, akselin skaalaukseen, kuvan kokoon. Mikäli haluat muuttaa tekstiä, kannattaa ensimmäisen hiiren painalluksen jälkeen odottaa hetki ennenkuin painaltaa hiirtä uudestaan. Näin pääset muuttamaan tekstiä kätevimmin. Jos kuvasta puuttuu jokin tekstiosa, voit lisätä sen **Chart**-valikosta (**Chart/Add chart element/Text box**).

Otsikko

Mikäli haluaa jo SPSS-ohjelmassa otsikoida kuvan, se on kätevintä tehdä siinä vaiheessa kun ”tilaa” eli tekee kuvaa.

Osien tuhoaminen

Jos haluat tuhota kuvista osia, esim. poistaa kuvasta yhden pylvään, aloita tuplaklikkaamalla kuvaajan (esim. pylväät) päällä. Esille tulevasta **Properties**-ikkunan **Categories**-välilehdeltä voit poistaa pylväitä valitsemalla hiirellä **Order**-laatikosta haluamasi muuttujan luokan ja painamalla hiirellä punaista **risti** -painiketta. Tällöin haluamasi muuttujan luokka siirtyy **Excluded**-laatikkoon. Takaisin luokan saa halutessaan valitsemalla se ensin hiirellä ja sitten painamalla vihreää **nuoli** -painiketta.

Akselin skaalauksen muuttaminen

Voit muuntaa kuvan skaalausta tuplaklikkaamalla y-akselia tai painamalla **Y**-pikapainiketta, jolloin sinulle aukeaa y-akselin **Properties**-valikko. Uuden skaalauksen voit syöttää **Scale**-välilehdellä suoraan **Range**-kenttiin (minimum, maximum) ottamalla ruksin pois **Auto**-ruudusta. X-akselin skaalausta ei voi jälkikäteen muuttaa, vaan se määräytyy aineistossa olevien lukujen perusteella.

Värien vaihtaminen

Valitse kuvan osa klikkaamalla ko. osan päällä, jotta pääset **Properties**-valikkoon. Välilehtien nimet, joissa voi vaihtaa väriä vaihtelevat kuvan eri kentissä, esim. pylväiden vaihto tapahtuu **Fill and Border** -välilehdellä, kun taas tekstin värin vaihto tapahtuu **Text**-välilehdellä. **Fill**-ikkuna tarkoittaa pylväiden väriä ja **Border**-ikkuna rajojen väriä. Ikkunasta voit valita värin ja **Apply**-painikkeella näet, miltä muutos näyttäisi. **Close**-painikkeella hyväksyt muutoksen.

Kuvion kääntäminen

Voit kääntää pylväikköä 90 astetta, valikosta **Chart/Transpose chart**, myös pikanäppäin.

Muita

Kannattaa kokeilla rohkeasti mitä eri valikoista löytyy, voit esimerkiksi koettaa vaihtaa pylväikön piirakaksi **Chart/Change Data element type/Pie**. Koko kuvan tulee olla valittuna, jotta tämä onnistuu. Kuvioiden muotoilua oppii parhaiten kokeilemalla eri toimintoja. Samalla kun teet muutoksia kuvaan **Chart Editor** -ikkunassa, samat muutokset tulevat näkyviin **Output**-ikkunassa.

Kuvan tallentaminen

A) Tulostusikkunassa: Kuvan voit tallettaa myöhempää **SPSS** -käyttöä varten tallettamalla tulostusikkunan sisällön (**Output** aktiivisena ikkunana: **File/Save as**). Tällöin ohjelma tallentaa koko tulostusikkunan. Jos haluat tallentaa ainoastaan jonkun kuvan tai kuvat, sinun täytyy käyttää **File**-valikon **Export**-toimintoa. Mikäli aiot siirtää kuvan johonkin toiseen ohjelmaan, varmin tapa tulostusikkunassa ollessasi on tallettaa kuva **Windows Bitmap** -muodossa (**File/Export**).

B) Kuvankäsittelyikkunassa: Kuvan voi tallettaa kuvankäsittelyikkunassa **File**-valikon **Export chart XML** -toiminnoilla.

Kun olet muotoillut ja tallentanut kaavion, sulje kuvankäsittelyikkuna. Tällöin ohjelma palaa takaisin tulostusikkunaan. Mikäli jätät kuvankäsittelyohjelman auki, muokkaamasi kuvan tausta on harmaa tulostusikkunassa.

Hyvän tilastografiikan yleisiä ehtoja

- **Kuva täytyy pystyä lukemaan itsenäisesti.** Kuvalla tulee olla numero, otsikko sekä muuttujat täytyy ilmaista selkokielellä (jos ei mahdu kuvaan voi käyttää lyhenteitä, mutta selitys tulee olla alareunassa)
- Kuvallisessa esityksessä on pyrittävä yksinkertaisuuteen
- Kuvaajan on oltava helppo lukea
- Sisältää ainoastaan oleellisen tiedon.

Yhdenmukainen muiden kuvien kanssa tekstiltään ym.

4 HAVAINTOAINEISTON MUOKKAAMINEN

Tutkimusaineistoa ei koskaan tule ottaa annettuna. Valmiit muuttujat ja niiden luokitukset ja asteikot eivät ole pyhiä: tutkija saa muuttaa ja hänen on muutettava niitä omien tarpeidensa mukaan. Muuttujamuunnoksia ovat esim.

- Uudelleenluokittelu
- Summamuuttujien luominen
- Aritmeettiset muunnokset
- Jne.

4.1 Uudelleenluokittelu

Miksi pitäisi uudelleen luokitella?

Muuttujien luokkia yhdisteltäessä katoaa aina informaatiota, joten yhdistelyä ei pidä tehdä turhaan. Yhdistelylle on aina löydettävä perustelu. Toisaalta, joskus uudelleenluokittelu on pakko tehdä. Perusteluja uudelleenluokittelulle:

- Tarpeettoman informaation hävittäminen.
- Jatkuva muuttuja on pakko luokitella ristiintaulukoita ja pylväskuvioita varten!
- Frekvenssiltään (=havaintojen lukumäärä, engl. *frequency* tai *count*) pienet luokat tulee yhdistää esim. ristiintaulukkoa varten, jotta voitaisiin tehdä johtopäätöksiä.

- Muuttuja voidaan joskus normalisoida (= muuttujan jakauma muistuttaa paremmin normaalijakaumaa) yhdistämällä ääriluokkia.

Tutkija pyrkii tiiviiseen esitykseen ja hävittää aineistosta turhana pitämäänsä informaatiota. Tarpeettoman informaation hävittäminen on yksi tutkimuksen perustehtävistä. Muuttujien uudelleenluokittelulla voidaan yksinkertaistaa turhan yksityiskohtaisia luokitteluja. Informaatiota ei kuitenkaan pidä tahallaan hävittää ilman hyviä perusteita.

Jos jatkuvaa muuttujaa halutaan käyttää luokiteltuna (esim. ristiintaulukoissa), luokat on tietenkin muodostettava (esim. euron tarkkuudella mitatut tulot tuloluokiksi). Se, käytetäänkö muuttujaa jatkuvana vai luokiteltuna, riippuu mm. käytettävästä analyysimenetelmästä ja riippuvuuksien muodosta. Nominaalisille (luokitteluasteikolla mitatuille muuttujille kehitetyt analyysitavat saattavat mahdollistaa joidenkin asioiden esittämisen (esim. epälineaaristen yhteyksien) helpommin ja selkeämmin kuin suhde- tai intervalliasteikollisille muuttujille kehitetyt menetelmät, vaikka niillä menetetäänkin tietoa, joka sisältyy muuttujien järjestykseen. Todella pienet luokat eli sellaiset, joista ei kuitenkaan voida tehdä mitään johtopäätöksiä, voi ehkä yhdistää kylmiltään.

Siis: ristiintaulukointia ja pylväskuvioita varten muuttujat on aina syytä luokitella siten, että jokaisen luokan frekvenssi on riittävän suuri johtopäätösten tekemiseen!

Usein on myös mahdollista tarkistaa, onko kadotettava informaatio olennaista. Tutkija voi esimerkiksi osoittaa, että selitettävän muuttujan vaihtelu on samanlaista kahdessa pienehkössä luokassa: ne voidaan ehkä yhdistää jatkoanalyyseissä. (Mutta myös se, että luokat ovat samanlaisia, on usein kiinnostava ja raportoimisen arvoinen tulos.)

Usein on niin, että analyysivaiheessa luokkia kannattaa varmuuden vuoksi pitää erillään, mutta raportissa yhdistelyä kannattaa tehdä kovemmalla kädellä. Kylmä tosiasia muuttujien uudelleenluokittelusta on, että luokkia voidaan aina yhdistää uudelleen, mutta jos jaottelut ovat alkuperäisessä aineistossa karkeita, niitä ei jälkikäteen saa hienojakoisemmiksi eli suuria luokkia ei saa erilleen toisistaan. Esimerkiksi kyselylomaketta tehtäessä on vältettävä turhia karkeistuksia.

Luokkien teossa huomioitavaa:

- Pidä huolta, ettei luokkien frekvensseistä tule liian pieniä analyysyä varten.
- Ainakin kolme luokkaa on hyvä olla dikotomian estämiseksi. Epälineaarista muuttuvuutta voi ilmaista kun on ainakin kolme luokkaa. Näin vältetään myös liialta maailman yksinkertaistamiselta.
- Esimerkiksi ikäryhmiä tehdessä kannattaa miettiä mielekkäitä ryhmiä esim. nuoret ja aikuiset tutkimusongelmasta riippuen.
- Esimerkki aviosääty: eronneet ja lesket pieniä ryhmiä, mutta niiden yhdistäminen huono asia, koska ryhmät kokonaan eri asioita.

Muuttujien uudelleenluokittelu SPSS –ohjelmalla

Yleensä kannattaa selostaa raporteissa, mitä uudelleenluokitteluja on käyttämilleen muuttujille tehnyt ja perustella, minkä vuoksi. Kaikkein helpointa ja nopeinta muuttujien uudelleenluokittelu on syntaksi-ikkunassa. Seuraavassa on kuitenkin selostettu uudelleenluokittelu valintaikkunoiden avulla.

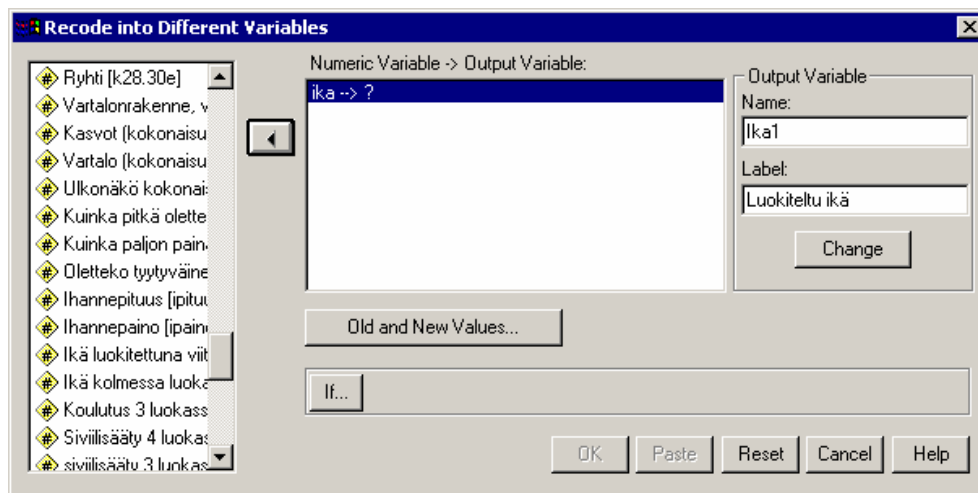
Jatkuvan muuttujan luokittelu ristiintaulukointia varten

Esimerkkinä vuoden tarkkuudella mitattu ikä-muuttuja (IKA). Aluksi on syytä selvittää jatkuvan muuttujan minimi ja maksimiarvot, jotta tiedetään, mistä aloitetaan ja mihin päätetään luokittelu. Nämä saa selville ajamalla esimerkiksi **Analyze/Descriptive Statistics/Descriptives**. Tässä tapauksessa saatiin selville, että minimiarvo on 18 ja maksimi 75. Seuraavassa vaiheessa mietitään, minkälainen luokittelu olisi oman tutkimuksen kannalta paras. Voi olla havainnollista myös tulostaa vaikkapa histogrammi, niin nähdään, miten ikä on jakautunut. Histogrammista huomataan vaikkapa semmoista, että vanhoja ihmisiä on paljon vähemmän kuin keski-ikäisiä. Jos muuttujaa aiotaan käyttää ristiintaulukossa, kannattaa ehkä ottaa vähän isompi luokitusväli "vanhuksille", jotta frekvenssit pysyvät riittävän suurina. Toinen asia on sitten se, jos ollaan kiinnostuneita erityisesti vanhoista ja halutaan tarkempaa informaatioita heistä. Tässä kuvitteellisessa esimerkissä tutkija haluaa yhdistellä luokat seuraavasti: 18-30v luokaksi 1, 31-45v luokaksi 2, 46-55 luokaksi 3, 56-65 luokaksi 4 ja 66- luokaksi 5. Huomaa, että sama ikä ei saa olla yhtä aikaa kahdessa eri luokassa!

Uudelleenluokittelu valintaikkunoiden kautta

Käytännössä luokkien muodostaminen sujuu SPSS:ssä valintaikkunoiden kautta seuraavasti (**Huomhuom! Säästät aikaa ja vaivaa, kun teet uudelleenluokittelun syntaksissa**, syntaksista myöhemmin luvussa 6).

Uudelleenluokittelua pääsee tekemään Transform-valikossa (Transform/Recode/Into different variables) Into different variables -toiminto luo sinulle uuden, luokitellun muuttujan ja säilyttää alkuperäisen luokittelemattoman ikämuuttujan toisin kuin toiminto Into Same variable, joka korvaa alkuperäisen muuttujan uudella. Yleensä on syytä valita into different..., jottei alkuperäinen informaatio häviä! Näin saadaan näkyville seuraava ikkuna:



Valitse muuttujaluettelosta muuttuja eli tässä tapauksessa ika, jonka arvot haluat uudelleenluokitella **Numeric variable** -kenttään. **Output Variable** -kenttään anna uuden muuttujan nimi (Name) ja selitysteksti (Label). Uuden muuttujan nimeämisessä on noudatettava samoja sääntöjä, kuin muuttujien nimeämisessä muutenkin (max. 8 merkkiä, alkaa kirjaimella). Älä anna luokitellulle muuttujalle samaa nimeä kuin jatkuvalle ikämuuttujalle! Paina **Change** nappia, kun olet nimennyt uuden muuttujan. Tämän jälkeen paina **Old and New values** -palkkia. Avautuvassa valikossa klikkaa sopivaa "Range" kohtaa ja täytä haluamasi arvoväli tai arvo ikkunan vasemmalle puolelle. Yksittäisiä lukuarvoja voi luokitella uudestaan kohdassa value. Yksi luokka käsitellään aina kerrallaan: Old Values puolelle vanhan muuttujan arvot ja New value puolelle uuden muuttujan vastaava luokka. Kun olet määritellyt yhden luokan vanhat ja uudet arvot paina Add. Sen jälkeen

syötät seuraavan luokan vanhat ja uuden arvon, painat Add jne. Kun kaikki arvot ovat uudelleenluokiteltu, paina **Continue** ja Recode-valintaikkunassa **OK**.

Huom! Uudelleenluokitellun muuttujan arvoille kannattaa myös antaa arvojen selitystekstit (value labels) Data Editor -ikkunassa ennen kuin alkaa käyttää sitä taulukoissa. Tällöin taulukoita tulostaessa selitystekstit tulostuvat luokkien numeroiden tilalle, mikä on paljon havainnollisempaa! Uudet muuttujat asettuvat aina aineistoikkunan (Data Editor) viimeiseksi muuttujaksi.

Luokitellun muuttujan luokkien yhdistäminen

Kannattaa ensin tarkastella alkuperäisen muuttujan frekvenssejä (Analyze/Descriptive Statistics/Frequencies) ja päätellä niiden perusteella, mitkä luokat kannattaa kenties yhdistää. Muuten uudelleenluokittelu etenee samaan tapaan kuin edellä. Esim. kuviteltu muuttuja "vastaus" joka saa arvot 1 = juu, 2 = ei, 3 = ehkä ja 4 = ei osaa sanoa. Luokat 3 ja 4 halutaan yhdistää, eli vanha luokka = uusi luokka 1 = 1, 2 = 2 ja 3 – 4 = 3.

Uusille luokille muuten voisi antaa mitkä lukuarvot tahansa, esim. juu voisi olla vaikka 55 ja ei 99 jne. Loogisinta kuitenkin lienee (etenkin jos haluat jonkun muunkin kuin sinun ymmärtävän työstäsi jotakin) että uusille luokille annetaan numerot numerojärjestyksessä yhdestä eteenpäin.

4.2 Summamuuttujat

Yksinkertaisimpien summamuuttujien perusajatuksena on ajatus yksiulotteisesta ilmiöstä, jota on mitattu usealla mittarilla (esimerkiksi kysytty usealla kysymyksellä). Tutkija yhdistää summaamalla (ts. laskee yhteen) näitä samaa asiaa kuvaavia muuttujia, jolloin hän saa asialle alkuperäisiä yksittäisiä muuttujia luotettavamman yhdistelmämittarin.

Tavallisesti lähtökohtana todellakin on, että summamuuttujan osiot eli alkuperäiset muuttujat mittaavat samaa asiaa. Summamuuttujien osioiden keskinäisten korrelaatioiden ajatellaan johtuvan siitä, että ne heijastavat samaa asiaa. Kaikkien osioiden ajatellaan siis olevan saman asian indikaattoreita, jolloin niiden varianssista vaihteleva osa on peräisin tuosta piilomuuttujasta, eli niillä on yhteistä varianssia ts. ne korreloivat.

Pääsääntöisesti summamuuttujan osioiden välisten korrelaatioiden on oltava nimenomaan positiivisia. Summamuuttujan osiot onkin usein käännettävä samansuuntaisiksi ennen

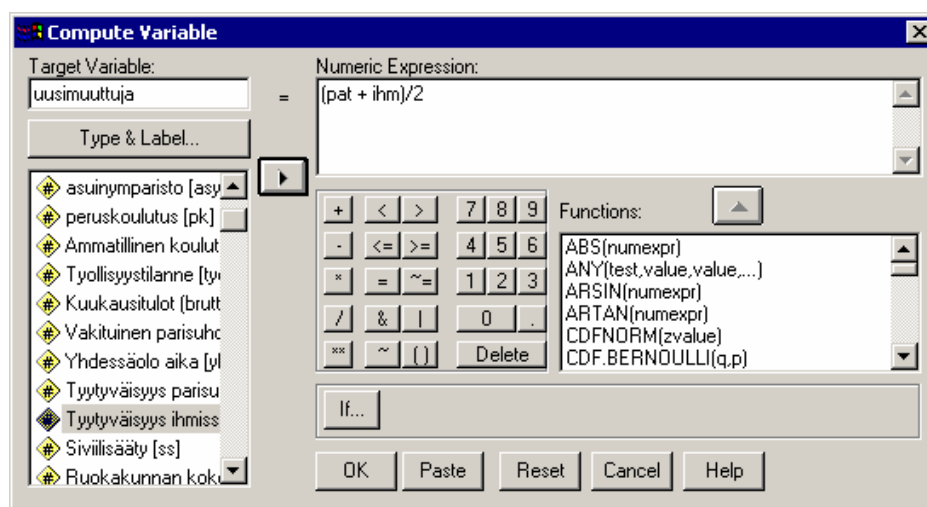
summaamista. Esimerkiksi jos summamuuttujaan kerätään asennemuuttujia, joissa osa asenneväittämistä on esitetty myönteisessä ja osa kielteisessä muodossa, jommatkummat on käännettävä. Kätevä tapa kääntää muuttuja on lisätä muuttujan luokkien määrään ykkönen ja vähentää muuttujan arvo tästä luvusta.

Kun osiot on käännetty, ne yksinkertaisesti summataan uudeksi muuttujaksi. Sitä, mittaavatko summamuuttujan osiot samaa asiaa, voidaan selvittää tarkastelemalla osioiden välisiä korrelaatioita ja faktorianalyysillä. Kaikkien pitäisi olla kohtuullisen isoja. Aina summamuuttujan osioiden ei ole tarkoitus korreloida keskenään positiivisesti. Näin on esimerkiksi, jos halutaan laskea yhteen oppilaiden arvosanapistemääriä, ja halutaan ottaa huomioon kaikki tieto- ja taitoaineet, vaikka esimerkiksi matematiikan ja liikunnan taidot korreloisivat oppilasjoukossa negatiivisesti.

Käytännön esimerkki summamuuttujasta: kyselyssä on kysytty useita kysymyksiä vastaajan suhtautumisesta ulkomaalaisiin, joiden voidaan ajatella mittaavan henkilön "rasistisuutta" (harvempi kuitenkin suoraan myöntää olevansa rasisti, joten tätä ei ehkä kannattaisikaan kysyä suoralla kysymyksellä).

Summamuuttujien tekeminen

Muuttujien välisiä korrelaatioita voi laskea valikossa **Analyze/Correlate/Bivariate**. Havaittuasi joidenkin muuttujien korreloivan hyvin keskenään, voit tehdä niistä summamuuttujan Spss:ssä summamuuttujia pääsee tekemään **Transform-valikossa (Transform/Compute)**. Näin saadaan näkyville seuraava ikkuna:



`Target value` on uuden summamuuttujan nimi, `Numeric expression` -kenttään kirjataan yhteen summattavat muuttujat esim. (kissa+koira+hamsteri). Tässä tapauksessa uuden muuttujan nimi voisi hyvin olla lemmikkieläimet. Hyvin usein yhteenlasketut muuttujat jaetaan muuttujien määrällä, jotta aikaisempi muuttujan skaala pysyy samana. Esimerkiksi likert-asteikolliset muuttujat: vastausvaihtoehdot: täysin samaa mieltä, jokseenkin samaa mieltä, ei osaa sanoa, jokseenkin eri mieltä, täysin eri mieltä. Vastausvaihtoehdot vaihtelevat siis yhdestä viiteen. Uuden muuttujan arvo pysyy tämän skaalan sisällä, jos jakotoimitus tehdään. Tämä helpottaa ajattelua siitä, mitä tarkoittaa esimerkiksi se, kun keskiarvo uudella summamuuttujalla on 1.8. Eli tässä tapauksessa vastanneet ovat olleet lähinnä jokseenkin samaa mieltä kysytyn asian suhteen. Muuttujien yhteen laskeminen edellä esitetyllä tavalla on ongelmallista silloin, kun jossakin muuttujassa on puuttuva tieto. Tämänlaisessa tilanteessa kannattaa käyttää funktio-toimintoa `Mean`. Se laskee keskiarvon niistä muuttujista, joista tieto on olemassa. Tällöin `numeric expression` -kenttään kirjataan: `mean (kissa, koira, hamsteri)`.

4.3 Muut muunnokset

Aritmeettiset muunnokset:

Mittayksikön vaihtaminen

Esim. aika vuorokausista tunneiksi, etäisyys mailleista metreiksi tai energia kilokaloreista kilojouleiksi. Tai vaikkapa: Koodataan koulutusastemuuttuja, joka saa arvoja 'perusaste', 'keskiaste' ja 'korkea-aste' koulutusvuosiksi. Etuna tässä muunnoksessa on, että koulutusvuosimuuttujaa voidaan käyttää välimatka-asteikollisena muuttujana.

Muita muunnoksia

Esimerkkejä muista muunnoksista voisi keksiä loputtomasti.

Muutamia esimerkkejä muuttujien muunnostarpeesta:

- Tutkittavilta on kysytty heidän asuntonsa huonelukua. Koska hyvin harvoilla on enemmän kuin yhdeksän huonetta, luokat 9 + voidaan yhdistää saman tien, etenkin jos muuttujaa tullaan käyttämään luokiteltuna muuttujana.
- Tutkija on kiinnostunut tutkittaviensa asumisahtaudesta. Hän oivaltaa, että asunnon huoneluku kannattaa suhteuttaa asukkaiden määrään. Onhan 20 neliön yksiössä asumisen sosiaalinen

merkitys aivan erilainen sinkulle kuin suurelle lapsiperheelle. Edessä on aritmeettisen muunnoksen tarve.

5 SYNTAKSI

Syntaksi-ikkuna tarjoaa käyttäjälle mahdollisuuden työskennellä komentokielen avulla. Syntaksi-ikkunaa käytettäessä kirjoitetaan SPSS- komentokielen komentoja ruudussa näkyvään tilaan ja kun komento on saatu kirjoitettua käsketään SPSS -ohjelmaa toteuttamaan komento.

Syntaksi-ikkunaan päästään valitsemalla **File**-valikosta toiminto **New/Syntax**. Aiemmin luotuun tiedostoon päästään valitsemalla **File**-valikosta **Open** ja valitsemalla laatikkoon **Files of type: Syntax**. Nyt saadaan kaikki talletetut syntaksi-ikkunat näkyviin ja voidaan valita se, mikä haluttiin avata.

Tärkeää komentojen kirjoittamisessa

- Jokaisen rivin ensimmäinen sarake on pyhitetty SPSS:n komentokielen käskyille, joten rivin alkuun ei voi kirjoittaa mitä tahansa. Esimerkiksi muuttujien nimet eivät koskaan ala rivin alusta vaan aikaisintaan toiselta sarakkeelta.
- **Jokainen varsinainen komento päättyy aina pisteeseen.**
- Monissa komennoissa on sekä pakollisia että vapaavalintaisia määrittelyjä tai alakomentoja. Esimerkiksi tehtäessä ristiintaulukoita on pakko määritellä mitä muuttujia käytetään. Sen sijaan vapaavalintaisia määrittelyjä ovat esim. se, mitä taulukon soluissa halutaan esitettäväksi.

Seuraavassa on esimerkki ristiintaulukon tekemisestä syntaksi-ikkunassa:

CROSSTABS

/TABLES=koti BY sp

/FORMAT= AVALUE TABLES

/STATISTIC=CHISQ

/CELLS= COUNT .

Edellisellä komennolla pyydettiin ohjelmaa tekemään ristiintaulukko sekä χ^2 -testi (erojen tilastollisen merkitsevyyden -testi). Muuttujat ovat koti (koti) ja sukupuoli (sp).

Uudelleenluokittelu syntaksi-ikkunassa

Luvussa 4.1 käytiin läpi muuttujan uudelleenluokittelu valikkojen kautta. Tämän saman saat tehtyä paljon nopeammin ja yksinkertaisemmin kirjoittamalla syntax-ikkunaan:

```
RECODE  
  ika  
  (18 thru 30=1) (31 thru 45=2) (46 thru 55=3) (56 thru 65=4)  
(66 thru  
  Highest=5) INTO ikal .  
VARIABLE LABELS ikal 'ika viidessä luokassa'.  
EXECUTE .
```

Tässä ika on jatkuva ikämuuttuja ja ikal on uuden muuttujan nimi, joka tällä käskyllä tehdään (eli luokiteltu ikä-muuttuja). Käskyn saat toteutettua maalaamalla komennon hiirellä ja painamalla Run-näppäintä. Eikö olekin paljon selkeämpää kuin monimutkaisten valintaikkunoiden käyttö?

Esimerkki *Helpin* käytöstä syntaksikomentojen tekemiseen: [Help/Topics/Index/Crosstabs/Crosstabs Command Syntax](#).

- Helpissä vapaavalintaiset määrittelyt tai alakomennot esitetään hakasulkeissa.
- Keskenään vaihtoehtoiset määrittelyt aaltosulkeissa.
- Komennot isoilla kirjaimilla, käyttäjän antamat määreet pienillä.
- Käytännössä on helpompaa opetella syntaksi-komentoja tekemällä niitä valikkojen kautta esim. [Analyze/Descriptive Statistics/Crosstabs](#) ja painamalla **Paste**-nappia, jolloin valikkojen kautta tehty komento tulee näkyviin syntax-ikkunaan. Saman mallin mukaan voi jatkossa kirjoittaa samantapaisia komentoja itse.
- Kirjoitetut komennot toteutetaan painamalla syntaksi-ikkunassa olevaa **RUN**-painiketta (nuolinäppäin).
- Komentojen toteuttaminen tapahtuu kätevimmin niin, että toteuttamiskelpoiset komennot maalataan hiirellä valituksi alueeksi, ja jonka jälkeen painetaan **RUN**-painiketta.
- **EXECUTE**.-komento tai [Transform/run pending transformations](#) toteuttavat tehdyt muutokset.

Miksi käyttää syntaksi-ikkunaa?

- **Transform**-valikon toiminnot (muuttujamuunnokset) ovat helpompia tehdä syntaksissa. Lisäksi kaikkia muuttujamuunnoksia tai muita käskyjä ei pysty tekemään ollenkaan valikkojen avulla.

- Jos haluaa tehdä toistoja esim. ristiintaulukoita eri muuttujille on nopeinta vaihtaa vain muuttujien nimet syntaksi-ikkunassa.
- Useat samanlaiset selitystekstit muuttujien arvoille (`value labels`) kätevämpiä tehdä syntaksissa.
- Jos haluaa **dokumentoida** (joka on aina kannattavaa!) sen, mitä on tehnyt, "lokitiedosto" `Options/General/ Session Journal/Browse` näyttää kaiken mitä syntaksissa on tehnyt. Tästä toiminnosta on hyötyä neuvontatilanteissa sekä oman muistin paikkaamisessa sen suhteen, mitä ja miten on tehnyt muuttujamuunnoksia ja analyysejä.
- Loppujen lopuksi kun syntaksin oppii, minkä tahansa homman tekeminen on sillä nopeampaa kuin valikkojen avulla.

6 SPSS:N HALLINTA

6.1 Keskeiset toiminnot

Data valikossa:

Split file:

Käytetään tilanteessa, jossa halutaan tehdä erilliset analyysit tietyille havaintoryhmille esim. miehille ja naisille (`Data/Split file`). Erittäin hyödyllinen toiminto.

Select if:

Select if –toiminto on hyödyllinen silloin, kun halutaan tarkastella vain osaa tutkimusaineistosta. Tämän toiminnon avulla analyyseihin voidaan valita esimerkiksi ainoastaan naimattomat tai alle 40–vuotiaat ja niin edelleen. Valikkokäytössä käytetään data-valikon toimintoa `Select cases` ja syntaksissa `SELECT IF` -komentoa. Select if -toiminnon ero `Split file` -toimintoon: tässä ollaan kiinnostuneita ainoastaan yhdestä ryhmästä, esim. henkilöistä, jotka ovat vastanneet tiettyyn kysymykseen tietyllä tavalla. Tämän jälkeen kaikki analyysit ym. tehdään vain ja ainoastaan tälle valitulle ryhmälle, ei kummallekin ryhmälle erikseen, kuten `Split file` -toiminnossa.

Merge files:

Merge files -toiminnon (Data/Merge files/Add cases tai variables) avulla on helppo lisätä aineistoon uusia muuttujia tai havaintoja. Hyödyllinen esim. silloin, kun tehdään jatkokysely tai aineisto on hankittu osissa tai samaa aineistoa on tallentanut useampi henkilö.

Transform-valikossa:**Automatic recode:**

Toimiva tilanteessa, jossa muuttujat on tallennettu merkkijonona eli esimerkiksi tekstinä ja siten niitä ei ole valmiiksi luokiteltu mitenkään. Tämä toiminto (Transform/Automatic/recode) tekee luokittelun. (esim. maat tekstistä numeroiksi).

Count:

Tämän toiminnon avulla voidaan selvittää, kuinka paljon tiettyjä arvoja esiintyy muuttujajoukossa. Esim. kun halutaan laskea, kuinka monen rikoksen uhriksi vastaaja on joutunut viimeisen vuoden aikana. Voidaan käyttää kysymyksiin, joissa on samanlaiset vastausvaihtoehdot.

Analyze-valikossa:**Multiple choices:**

SPSS-ohjelmassa muuttujat voivat saada ainoastaan yhden arvon. Multiple response -toiminnon avulla havainnolle voidaan tallentaa useampia arvoja. Tätä toimintoa tarvitaan esimerkiksi tilanteessa, jolloin ihmisiltä on tiedusteltu, minkä rikosten kohteeksi he ovat joutuneet viimeisen kolmen vuoden aikana. Näiden vastausten tallentaminen on helpointa siten, että jokainen rikos koodataan omaksi muuttujakseen. Esim. muuttuja 1 Moottoriajoneuvo on varastettu, muuttuja 2 Moottoriajoneuvoon on kohdistunut vahingonteko jne. Muuttujien arvot tallennetaan dikotomioiksi siten, että

arvo yksi tarkoittaa jokaisella muuttujalla sitä, että kyseinen henkilö on joutunut ko. rikoksen uhriksi ja arvo 2 sitä, että ei ole joutunut. Kun kaikki havainnot on tallennettu, voidaan näistä muuttujista tehdä muuttuja "setti" (Analyze/Multiple response/Define sets). Kun "setti" on tallennettu voidaan tarkastella "setin" frekvenssejä sekä tehdä ristiintaulukoita.

Utilities-valikossa:

Sets:

Define sets -toiminnolla (Utilities/Define Sets) voi tehdä haluamistaan muuttujista (muuttujaryhmistä) muuttuja "setin". Kun setti on määritelty valitse toiminnolla Use sets se muuttujaryhmä eli setti, jota haluat tarkastella. Käytännöllinen silloin, kun tarkastelee muuttujia jonakin ryppäänä, ei tarvitse aina klikata kaikkia muuttujia erikseen muuttujaluetteloon. Tai silloin, kun haluaa keskittyä jonkin aikaa vain tietyn muuttujaryhmän tarkasteluun ja ei halua koko muuttujalistaa aina näkyville valintaikkunoihin.

6.2 Osa-aineiston tekeminen

Kun jostain isommasta aineistosta halutaan muodostaa pienempi aineisto tehdään joitakin seuraavista komennoista syntaksi-ikkunassa:

`SAVE OUTFILE = 'tiedostonnimi ja polku'` (pakollinen, jos haluaa tallettaa osa-aineiston)

`/KEEP=ALL` (tallettaa kaikki muuttujat uuteen aineistoon)

`/KEEP=eka to vika, ika` (tallettaa muuttujat järjestyksessä eka:sta vika:aan & ika)

`/DROP=eka to vika, ika` (poistaa uudesta aineistosta muuttujat eka:sta vika:an & muuttujan ika).

7 SPSS, ASCII, EXCEL JA WORD

7.1 ASCII

Puhtaana havaintomatriisina koodattu aineisto tallennetaan usein ns. ascii-muodossa ts. tiedosto sisältää ainoastaan numeroita ja välilyöntejä (ja mahdollisesti myös pilkkuja). Tällaisia aineistoja voidaan lukea SPSS:ään komennolla `File/Read text data`.

Kun olet valinnut haluamasi tekstitiedoston, ohjelma käynnistää `Text Import Wizardin`, joka auttaa sinua saamaan tekstitiedostosi siististi SPSS -taulukoksi. Voit koko ajan seurata, miten `Text Import Wizardissa` esitettyihin kysymyksiin antamasi vastauksen vaikuttavat aineiston sijoitteluun taulukoksi.

- Ensimmäisessä ikkunassa kysytään "Does your text file match a predifened format?" johon vastataan yleensä no. Toisessa ikkunassa valitaan, onko luettava tekstitiedosto siististi sarakkeissa (`fixed width`) vai pilkuilla, välilyönneillä tms. eroteltuna (`delimited`) sekä kerrotaan ohjelmalle, ovatko muuttujien nimet tiedoston ensimmäisellä rivillä.
- Ja niin edelleen, vastaile eri valintaikkunoissa oleviin kysymyksiin niin, että aineisto näyttää vähitellen asettuvan sellaiseen muotoon kuin haluat.
- Jos sinulla on monia samalla periaatteella tehtyjä tekstiaineistoja, kannattaa tallettaa tuo malli, millä saatoit ensimmäisen tuota tiettyä muotoa olevan tiedoston SPSS -taulukoksi (Kuutosikkunassa "Would you like to save this format for future use?") tai liimata se syntaksiin (`paste to syntax`) ja tallettaa syntaksi. Seuraava kerta sujuukin sitten nopeammin ja helpommin!

7.2 Excel

SPSS 12.0 lukee Excel 5 tai myöhempää versiota suoraan. Eli lukeminen tapahtuu seuraavasti: avaa Excel-tiedosto (jonka tulee avaamisen aikana olla kiinni, muutoin SPSS ei pysty sitä lukemaan!) SPSS:n file-valikosta komennolla `Open data`. Huom. tiedostotyyppiä valitaan Excel (*.xls). Määriteltäväksi tässä jää ainoastaan se, onko Excel-aineiston ensimmäinen rivi muuttujan nimi vai ei. Mikäli ei tarvitse koko Excel-tiedostoa SPSS:ssä, sitä ei tarvitse sinne ottaa. Tällöin `Range-`kenttään voi laittaa niiden Excelin muuttujien solutiedot (eli sijaintitiedot), jotka SPSS:ään halutaan ottaa käytettäväksi.

7.3 Tulosten siirtäminen tekstinkäsittelyohjelmaan

Tulosten siirtäminen on hyvin helppoa. Valitaan tulostusikkunasta halutut kohdat ja valikosta `Edit/Copy objects`. Etenkin Wordissä, tosin sen versiosta riippuen, kannattaa kopioidut kohdat liittää sinne käyttämällä toimintoa `Edit/Paste special`, eli Liitä määräten. Mikäli haluat siirtää taulukoita ja kuvia, jotka eivät sijaitse tulostuksessa peräkkäin, valinta voidaan tehdä pitämällä CTRL-näppäin alhaalla ja napsauttamalla hiirellä kutakin haluttua kohdetta. Viimeiseksi Wordissä suoritetaan liittäminen (`Paste` tai `Paste special`) siinä kohdassa, mihin tulokset halutaan.

7.4 Kuvan tallentaminen eri grafiikkaohjelmiin jatkokäsiteltäväksi

SPSS for Windows -ohjelmassa voidaan kuvat tallettaa sellaiseen muotoon, että ne on mahdollista siirtää jatkokäsiteltäväksi grafiikkaohjelmaan, tai ne voidaan siirtää vaikkapa jonkin toisen järjestelmän tekstinkäsittelyohjelmaan.

Valitaan kuvankäsittelyikkunassa File-valikosta toiminto `Export`. Valikosta `Save file as type`, valitaan haluttu tiedostotyyppi (tarjolla on muun muassa `.jpg` eli `JPEG File`, `.bmp` eli `Windows bitmap` ja `.wmf` eli `Windows metafile`). `Options...`-painiketta napsauttamalla voi tallennettavalle kuvalle antaa vielä lisämäärittelyjä. Kun kuva on tallennettu `Export`-toiminnolla, se tuodaan esimerkiksi teksturiin Insert-valikon `Picture`-toiminnolla.

KIRJALLISUUTTA

Alkula, Pöntinen, Ylöstalo (1994): *Sosiaalitutkimuksen kvantitatiiviset menetelmät*. WSOY.

Hae tästä kirjasta ymmärtämystä, mutta älä testien detaljeja.

Kanniainen (1999): *SPSS for Windows: Menetelmiä*. Oulun yliopisto, ATK-keskus.

Metsämuuronen (2000): *Monimuuttujamentelmien perusteet SPSS-ympäristössä*. Metodologia-sarja

Kirjan voi tilata osoitteesta www.methelp.com.

Muhli & Kanniainen (2000): *SPSS 10.0 for Windows: Perusteet*. Oulun yliopisto, ATK-keskus.

Ranta, Rita, Kouki (1989): *Biometria. Tilastotiedettä ekologeille*. Yliopistopaino.

Soveltajan käsikirja

Tabachnick & Fidell (2000): *Using multivariate statistics*. 4. painos, Boston: Allyn ja Bacon.

Paljon käytetty ja arvostettu teos.

SPSS:n tietoa/oppaita verkossa:

SPSS Finland: <http://www.spss.fi/>

SPSS: <http://www.spss.com/>

SPSS, versio 12, demo noin 30 päiväksi:

http://www.spss.com/registration/login/login009.cfm?Demo_ID=37

Helsingin yliopisto, Antti Nevanlinna:

<http://www.helsinki.fi/atk/tilasto/spss.html>
<http://www.helsinki.fi/atk/tilasto/spsspk11/>

Jyväskylän yliopisto

<http://www.cc.jyu.fi/~kesonen/spss.html>

Oulun yliopisto:

<http://cc-stat.oulu.fi/spss/spss10Per/>

Tampereen yliopisto:

<http://www.fsd.uta.fi/menetelmaopetus/SPSS/spss.html>