

SPSS-PERUSTEET

SPSS:n versio 14

Maarit Valtari

Helsingin yliopisto
Valtiotieteellinen tiedekunta
Tieto- ja viestintäteknikka

Copyright Maarit Valtari ja HY:n valtiotieteellisen tiedekunnan TVT, 2006

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

Valtiotieteellisen tiedekunnan TVT
Helsinki: 2006

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 SPSS:n perusteet -oppaiden eri versioiden tekemisessä. Tämä SPSS-opas pohjautuu monilta osin valtiotieteellisen tiedekunnan tieto- ja viestintätekniikan SPSS-projektin varhaisimpiin versioihin, joita tekivät pääasiassa Timo Harjo ja Seppo Roponen. Ilman heitä tätä opasta ei olisi, ja siksi arvokkaimmat kiitokseni osoitankin heille. Lisäksi erityiset kiitokset Tero Erkkilälle, Aija Kaartiselle, Kimmo Vehkalahdelle ja Heikki Wileniukselle sekä muille valtiotieteellisen tiedekunnan SPSS-kurssien myöhempien materiaalien täydentämiseen, kommentoimiseen ja parantamiseen osallistuneille. Tämä opas on päivitetty SPSS-ohjelman versioon 14.0.

Espoo, 22.10.2006

Maarit Valtari

SISÄLLYS:

1 JOHDANTO	1
1.1 Yleistä SPSS-ohjelmasta	1
1.1 SPSS:n avaaminen	2
1.2 SPSS:n työskentelyikkunat	3
1.3 Aineistoikkunan valikot ja pikanäppäimet	6
2 HAVAINTOAINEISTON MUODOSTAMINEN, MUOKKAAMINEN JA HALLINTA.....	9
2.1 Havaintoaineiston muodostaminen.....	9
2.1.1 SPSS-tiedoston tallentaminen.....	9
2.1.2 Uuden aineiston tallennus.....	9
2.1.3 Database Wizard.....	11
2.2 Havaintoaineiston muokkaaminen.....	12
2.2.1 Uudelleenluokittelu.....	12
2.2.2 Summamuuuttajat.....	16
2.2.3 Muut muunnokset.....	17
2.2.4 Syntaksi.....	18
2.3 Havaintoaineiston hallinta	20
2.4 Havaintoaineiston tarkastelu ja kuvaaminen	22
3 TAULUKOINTI	23
3.1 Yleistä taulukoinnista	23
3.2 Frekvenssitaulukko	24
3.3 Muuttujan perustunnusluvut	25
3.4 Ristiintaulukko.....	26
3.5 Useampiulotteinen ristiintaulukko.....	30
3.6 Keskiarvotaulukko.....	32
3.7 Korrelaatio	33
3.8 Taulukoiden muokkaus.....	34
3.9 Lopuksi: siis mitä tehdä, kun.....	35
4 GRAFIIKKA.....	37
4.1 Grafiikan yleisiä ehtoja.....	37
4.2 Pylväskuvio	37
4.2.1 Yhden luokitellun muuttujan pylväskuvio.....	38
4.2.2 Yhden muuttujan tarkastelu toisen muuttujan luokissa.....	39
4.2.3 Kahden luokitellun muuttujan pylväskuvio	41
4.2.4 Keskiarvopylväät	44
4.2.5 Useiden erillisten muuttujien samanaikainen kuvaaminen	46
4.3 Histogrammi	49
4.4 Hajontakuviokuva	49
4.5 Viivakuviokuva	51
4.5.1 Kahden muuttujan viivakuviokuva.....	52
4.5.2 Kolmen muuttujan viivakuviokuva	53
4.5.3 Drop-line -kuviokuva	55
4.6 Aluekuviokuva	57
4.7 Piirakkakuviokuva.....	58
4.8 Kuvat niiden käyttötarkoituksen mukaan	59
4.9 Kuvien muokkaus	60
5 TULOSTEN SIIRTÄMINEN MUIHIN OHJELMIIN	63
KIRJALLISUUTTA	64

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. Tässä oppaassa käytetään Windows käyttöympäristöön sovitettua SPSS:n versiota 14. SPSS:stä on olemassa jo versio 15, joka on ilmestynyt Yhdysvalloissa syyskuun puolella välissä. Tässä täytyy todeta, että konkreettiset muutokset analyysien ym. tekemisessä ovat SPSS:n eri versioiden, erityisesti 12 – 15 varsin vähäiset.

SPSS muistuttaa suurelta osin muita Windows-ohjelmia, joten Windowsin käyttöön totunut 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 siis ohjata myös kirjoittamalla komentoja komennoille tarkoitettuun syntaksi-ikkunaan.

Oppaan tavoitteet

Oppaassa ei kerrata erikseen tilastotieteen perustermejä ja/tai teorioita, vaan ne oletetaan osattavan entuudestaan. Jos olet epävarma jostakin termistä tai et ymmärrä sitä, voit tarkastaa sen

Helsingin yliopiston tilastotieteilijä Juha Purasen tilastotieteen sanastosta:

<http://noppa5.pc.helsinki.fi/uudet/dalhtm/sanasto.html>

Laaja-alaisesti tilastollista tietoa löytyy Tilastokeskuksen verkkokoulusta, jota ehdottomasti kannattaa hyödyntää:

<http://www.tilastokeskus.fi/tup/verkkokoulu/index.html>

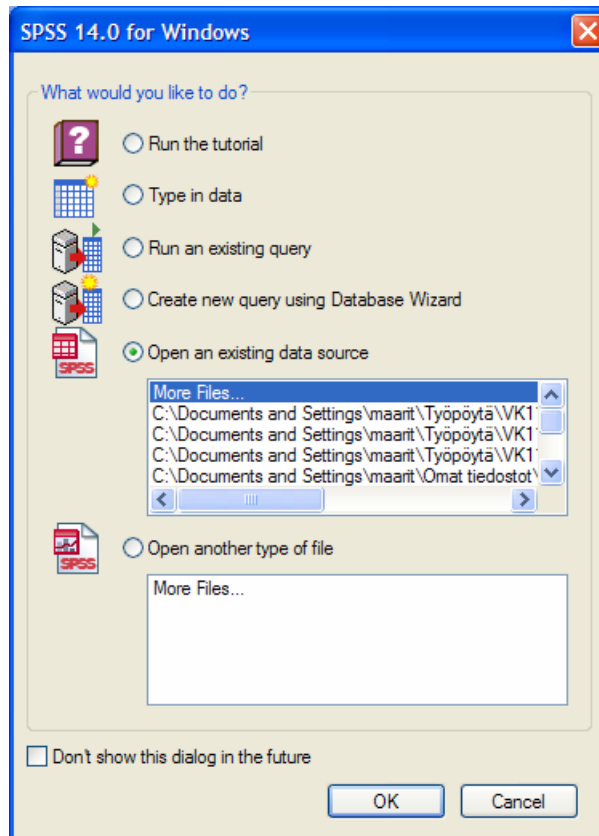
Tämän oppaan tavoitteena on tutustua SPSS-ohjelmaan ja tarjota ohjelman käyttöön tarvittavat perustiedot. Kannattaa palauttaa mieleensä etenkin erilaiset mitta-asteikot (luokitteluasteikko, järjestysasteikko, välimatka- ja suhdeasteikko) sekä käsitteet jatkuva/luokiteltu muuttuja, erilaiset keskiluvut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

Materiaalissa käytetyt aineistot

Materiaalissa esimerkkeinä käytetään kahta aineistoa: 1) Pääsääntöisesti 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. 2) Lisäaineistona grafiikassa käytetään Valtarin vuosina 1998 – 2004 Helsingin yliopiston valtiotieteellisen tiedekunnan SPSS-kurssien yhteydessä opiskelijoilta kerättyjä aineistoja, jotka on yhdistetty yhdeksi aineistoksi. Tutkimusaiheena on opiskelijoiden suhtautuminen ulkonäköönsä. Aineisto sisältää 452 havaintoa.

1.1 SPSS:n avaaminen

SPSS:ää avattaessa ensimmäiseksi ilmestyy seuraavassa nähtävä valikko. Ohjelma kysyy, mitä haluat tehdä.



Valitsemalla:

Run the tutorial –toiminnon ohjelma näyttää SPSS:n toimintaa ajoittelijalle

Type in data –toiminnon avulla siirrytään suoraan havaintopohjan ja -aineiston tallentamiseen

Create new query using Database Wizard –toiminnolla alat tuoda uutta aineistoa SPSS:ään

Erityisen toimivia ovat toiminnot Open an existing data source ja Open another type of file, sillä ne näyttävät lukuisia viimeisempiä käytössä olleita tiedostoja.

Halutessasi voit pyytää ettei ohjelma enää tulevaisuudessa näytä valikkoa ruksimalla kohta Don't show this dialog in the future.

Mikäli et halua valita mitään edellä mainituista toiminnoista, paina Cancel-painiketta ja siten siirryt SPSS:ään.

1.2 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, analyysit ym. tehdään aineistoikkunassa. Aineistoikkunassa on kaksi näkymää: Data view eli datasivu ja variable view eli muuttujasivu.

Aineistoikkunan Data view

	havl	keräysvuosi	sv	asla	maakunta	asym	pk	ak	tyo	tyomuu	ammatti	tulot	pari	y
1	1003	1	44	1	.	1	4	4	1			6	1	
2	1004	1	42	1	.	1	4	2	1			6	1	
3	1006	1	61	1	.	2	4	2	1			9	1	
4	1008	1	63	1	.	1	4	2	4			5	2	
5	1009	1	71	1	.	1	3	1	1			4	2	
6	1010	1	23	1	.	1	4	2	6			5	2	
7	1011	1	26	1	.	1	4	3	6			6	1	
8	1015	1	76	1	.	1	4	4	1			4	1	
9	1016	1	30	1	.	1	4	3	6			4	2	
10	1017	1	73	1	.	2	3	2	7			2	2	
11	1018	1	70	1	.	1	4	3	1			5	2	
12	1019	1	75	1	.	1	4	1	5			1	2	
13	1020	1	60	1	.	2	2	1	1			4	1	
14	1024	1	54	1	.	1	4	2	4			3	1	
15	1025	1	45	1	.	1	3	4	10			3	1	
16	1027	1	40	1	.	1	2	4	1			5	1	
17	1029	1	68	1	.	1	4	2	1			3	2	
18	1030	1	77	1	.	2	4	2	2			2	2	
19	1031	1	60	1	.	1	3	4	1			5	1	
20	1032	1	48	1	.	1	4	3	1			7	2	
21	1033	1	71	1	.	1	3	4	8			2	1	
22	1034	1	74	1	.	1	4	3	1			2	1	
23	1037	1	63	1	.	1	4	2	1			6	1	
24	1038	1	58	1	.	1	3	2	1			7	1	
25	1039	1	74	1	.	1	3	1	1			4	1	
26	1041	1	69	1	.	1	3	4	2			3	1	
27	1043	1	50	1	.	1	2	1	1			6	1	
28	1044	1	67	1	.	1	3	1	1			5	1	
29	1045	1	60	1	.	1	3	4	1			10	2	
30	1046	1	49	1	.	2	4	4	1			6	1	
31	1049	1	62	1	.	1	4	2	2			4	1	

Aineistoikkunan datasivusta (Data View) pääsee muuttujasivulle (Variable View) ikkunan vasemmassa alalaidassa olevan välilehden kautta.

Aineistoikkunan Variable view

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	havt	Numeric	4	0	havaintotunnus	None	None	8	Right	Scale
2	keräysvuosi	Numeric	8	0	Vuosi	{1, 1997}...	None	8	Right	Nominal
3	sv	Numeric	2	0	syntymävuosi	None	None	8	Right	Scale
4	asla	Numeric	2	0	Asuinlääni (vuoden 1997	{1, Uudenmaan lääni}...	None	8	Right	Ordinal
5	maakunta	Numeric	2	0	Maakunta (vuoden 2005	{1, Etelä-Karjala}...	None	8	Right	Nominal
6	asym	Numeric	1	0	Asuinympäristö	{1, Kaupunkimainen}...	None	8	Right	Nominal
7	pk	Numeric	1	0	Peruskoulutus	{1, Osa kansa- tai peruskoulua tai	None	8	Right	Ordinal
8	ak	Numeric	1	0	Ammatillinen koulutus	{1, Ammatillinen koulu}...	None	8	Right	Ordinal
9	tyo	Numeric	2	0	Työllisyystilanne	{1, Kokopäiväinen palkansaaja}...	None	8	Right	Nominal
10	tyomu	String	20	0	Muu vaihtoehto, mikä	None	None	8	Left	Nominal
11	ammatti	String	100	0	Ammatti	None	None	8	Left	Nominal
12	tulot	Numeric	2	0	Kuukausitulot (brutto)	{1, Alle 340 €}...	None	8	Right	Ordinal
13	pari	Numeric	1	0	Vakituinen parisuhde	{1, Kyllä}...	None	8	Right	Nominal
14	yhdessä	Numeric	4	1	Yhdessäolo aika vuosin	{0, Ei parisuhdetta}...	.0	8	Right	Scale
15	pat	Numeric	1	0	Tytyväisyys parisuhtee	{0, Ei parisuhdetta}...	0	8	Right	Ordinal
16	ihm	Numeric	1	0	Tytyväisyys ihmisuhte	{1, Erittäin tytyväinen}...	None	8	Right	Ordinal
17	ss	Numeric	1	0	Siviilisäät	{1, Naimaton}...	None	8	Right	Nominal
18	ruokak	Numeric	1	0	Ruokakunnan koko	None	None	8	Right	Scale
19	raskaus	Numeric	1	0	Raskaana	{1, Kyllä}...	None	8	Right	Nominal
20	lapset	Numeric	1	0	Lapsia	{1, Kyllä}...	None	8	Right	Nominal
21	lluku	Numeric	2	0	Lasten lukumäärä	{0, Ei lapsia}...	0	8	Right	Scale
22	lkoti	Numeric	1	0	Lasten kotona asuminen	{0, Ei lapsia}...	0	8	Right	Nominal
23	elat	Numeric	1	0	Tytyväisyys elämään	{1, Erittäin tytyväinen}...	None	8	Right	Ordinal
24	terveys	Numeric	1	0	Terveydentila	{1, Erittäin hyvä}...	None	8	Right	Ordinal
25	ter	Numeric	1	0	Rakastaa jotakuta	{1, Kyllä}...	None	8	Right	Ordinal
26	rakastaa	Numeric	1	0	Rakastaako Teitä joku	{1, Kyllä}...	None	8	Right	Ordinal
27	k24a	Numeric	1	0	Työ	{1, Erittäin tärkeä}...	None	8	Right	Scale
28	k24b	Numeric	1	0	Perhe	{1, Erittäin tärkeä}...	None	8	Right	Scale
29	k24c	Numeric	1	0	Terveys	{1, Erittäin tärkeä}...	None	8	Right	Scale
30	k24d	Numeric	1	0	Itsensä toteuttaminen, v	{1, Erittäin tärkeä}...	None	8	Right	Scale
31	k24e	Numeric	1	0	Yhteiskunnallinen toimin	{1, Erittäin tärkeä}...	None	8	Right	Scale
32	k24f	Numeric	1	0	Kanss	{1, Erittäin tärkeä}...	None	8	Right	Scale

2. Tulostusikkuna (*Output*): kaikkia tehtyjen analyysien tuloksia, taulukoita, kuvia ja muita tarkasteluita voidaan katsella ja tulostaa tästä tulostusikkunasta. Tulostusikkuna sisältää myös kapean istuntopolkuikkunan, jossa näkyy koko istunnon kulku. Tulostusikkunat voidaan tallentaa omiksi tiedostoiksi. Uutuutena SPSS:n 14 version tulostusikkunassa näkyy ennen tehtyjä analyysijä käytetyn aineiston nimi ja sen sijainti koneella.

[DataSet1] C:\Documents and Settings\maaric\Työpöytä\TK_perusteet versio 14\ulkonäkö.sav

Statistics

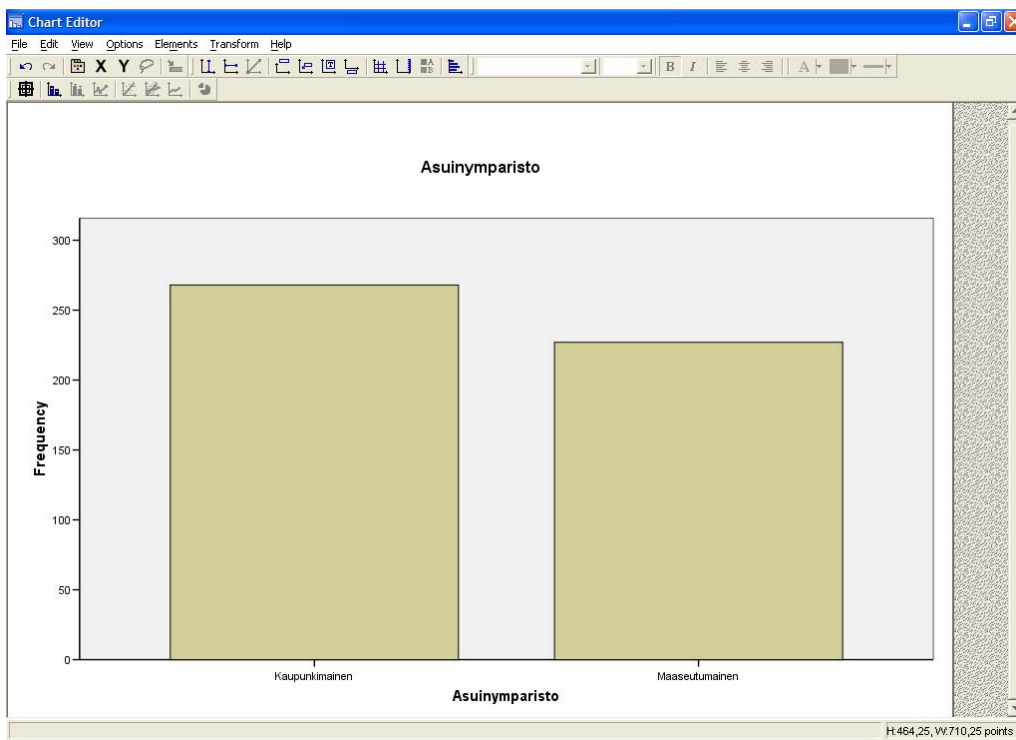
Asuinympäristö		N	Valid	Missing
		495		1

Asuinympäristö		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Kaupunkimainen	268	54,0	54,1	54,1
	Maaseutumainen	227	45,8	45,9	100,0
	Total	495	99,8	100,0	
Missing	System	1	,2		
	Total	496	100,0		

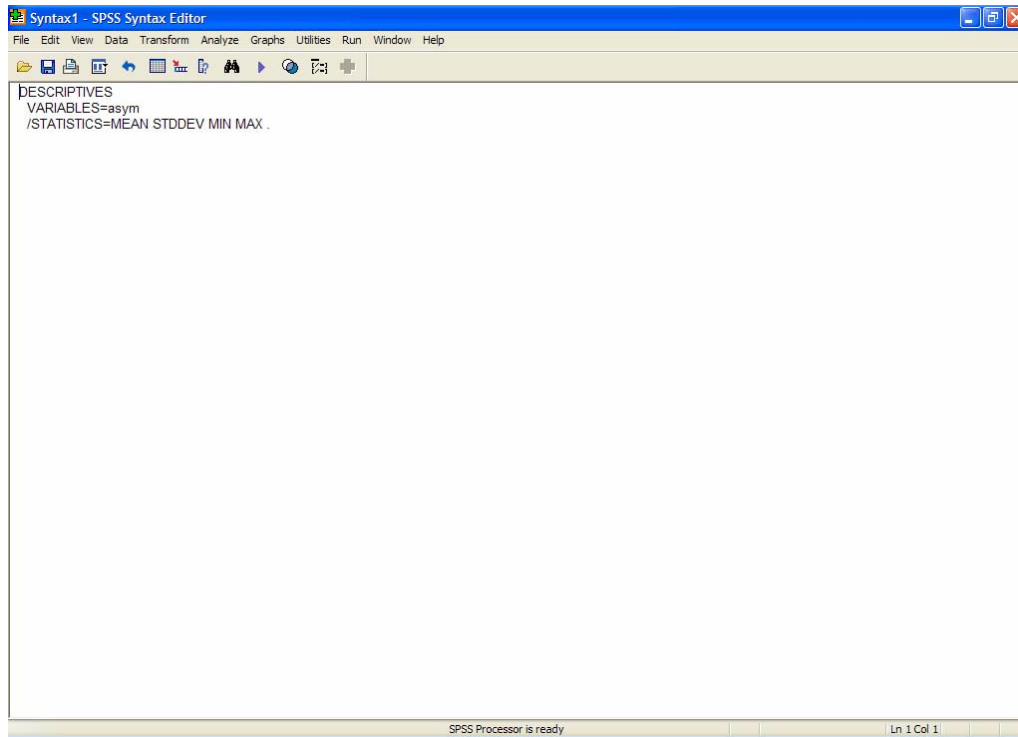
3. **Pivot-taulukkoikkuna (*Pivot table editor*):** taulukoiden editointi (tekstin muuttaminen, sarakkeiden leventäminen tai kaventaminen, taulukoiden kääntäminen, tekstin värin muuttaminen ym.). Kuvankäsittelyikkunaan pääsee tuplaklikkaamalla muokattavaa taulukkoa.

Ikä luokitettuna viiteen luokkaan	asuinymparisto	Kaupunkimainen	Count	Painoindeksi kahdessa luokassa		Total
				All-tai normaali-painoinen	Ylipainoinen	
18-29	Kaupunkimainen	Count	33	5	38	
			% within asuinymparisto	86,8%	13,2%	100,0%
	Maaseutumainen	Count	16	6	22	
			% within asuinymparisto	72,7%	27,3%	100,0%
Total	Count	49	11	60		
		% within asuinymparisto	81,7%	18,3%	100,0%	
30-39	Kaupunkimainen	Count	28	7	35	
			% within asuinymparisto	80,0%	20,0%	100,0%
	Maaseutumainen	Count	21	9	30	
			% within asuinymparisto	70,0%	30,0%	100,0%
Total	Count	49	16	65		
		% within asuinymparisto	75,4%	24,6%	100,0%	
40-49	Kaupunkimainen	Count	23	10	33	
			% within asuinymparisto	69,7%	30,3%	100,0%
	Maaseutumainen	Count	22	6	28	
			% within asuinymparisto	78,6%	21,4%	100,0%
Total	Count	45	16	61		
		% within asuinymparisto	73,8%	26,2%	100,0%	
50-59	Kaupunkimainen	Count	7	9	16	
			% within asuinymparisto	43,8%	56,3%	100,0%
	Maaseutumainen	Count	6	23	29	
			% within asuinymparisto	20,7%	79,3%	100,0%
Total	Count	13	32	45		
		% within asuinymparisto	28,9%	71,1%	100,0%	
60-74	Kaupunkimainen	Count	8	4	12	
			% within asuinymparisto	66,7%	33,3%	100,0%
	Maaseutumainen	Count	4	15	19	
			% within asuinymparisto	21,1%	78,9%	100,0%
Total	Count	12	19	31		
		% within asuinymparisto	38,7%	61,3%	100,0%	

4. **Kuvankäsittelyikkuna (*Chart editor*):** kuviin voidaan tehdä muutoksia tässä ikkunassa. Kuvankäsittelyikkunaan pääsee tulostusikkunasta vastaavasti kuin taulukonkäsittelyikkunaan eli tuplaklikkaamalla kuvaa.



5. Syntaksi-ikkuna (*Syntax editor*): 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. Syntaksista ja sen toiminnoista sekä sen käytön tarpeellisuudesta lisää luvussa 2.2.4.



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 (topics/hakemisto: hakusanalla script(s)). Scriptien opiskelu ei sisälly tähän materiaaliin.

1.3 Aineistoikkunan valikot ja pikanäppäimet

Valikot

Seuraavassa esitetään aineistoikkunan kymmenen valikkoa käyttötarkoitusten mukaan.

1. File: valikkoa käytetään aineiston avaamiseen, tallentamiseen, tulostamiseen, tulostuksen oletusasetusten muuttamiseen ja lopetukseen. File-valikosta voi myös avata uusia syntaksi- tai tulostusikkunoita. Open database –toiminnolla voidaan tuoda eri muodossa olevia tiedostoja SPSS:ään ja Read text file –toiminnolla ns. ascii-muodossa olevia tiedostoja. Mark file read only –toiminnolla saadaan aineisto suojattua. Rename dataset –toiminnolla, voit halutessasi vaihtaa käytössäsi olevien aineistojen nimeä. Nykyisin voi samanaikaisesti olla useampi aineistoikkuna auki, mikä on varsin hyödyllinen toiminto, tosin sen kanssa tulee olla varovainen. Toiminnolla Display datafile information saadaan kaikki käytössä olevan aineiston tai jonkin toisen aineiston muuttajat arvoselitteineen listattua tulostusikkunaan. Cache data –toiminnon avulla voi tehdä väliaikaisen osakopion koko aineistosta (tämän tarpeellisuus tulee

yleensä ainoastaan tilanteessa, jossa koneen muistikapasiteetti ei riitä koko aineiston käsittelyyn). 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 haluttuja arvoja ja tietyn havainnon etsimiseen `Go to case` -toimintoa. `Insert variable` -toiminto on varsin hyödyllinen monissa tilanteissa, esim. silloin kun tallennuspohjaa tehdessä on unohdettu välistä tallentaa jokin muuttuja. Vastaavasti havaintojen syöttämiseen voi käyttää `Insert cases` -toimintoa. Lisäksi tässä valikossa voi muuttaa ohjelman oletusasetuksia (`Options`). Kun haluaa tehdä ohjelman oletusasetuksiin muutoksia, kannattaa oletusasetusikkunassa painaa Ohje-painiketta. Se kertoo, mitä muutoksia on mahdollista tehdä. Tässä materiaalissa ei perehdytä ohjelman perusasetusten muuttamiseen.

3. View: valikosta voi vaihtaa tarpeidesi mukaisesti käytössä olevat pikapainikepalkit (`Toolbar`) sekä 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: Tässä valikossa on kaksi varsin hyödyllistä toimintoa aineiston tallennusta helpottamaan, `Define variable properties` -toiminto sekä `Copy data properties`, jolla voidaan kopioida olemassa olevien muuttujien tietoja, kuten esimerkiksi vastausvaihtoehdot uusiin muuttujiin. Havaintoja voi järjestää `Sort cases` -toiminnolla. Toisesta aineisto(i)sta voidaan yhdistää havaintoja tai muuttujia käsiteltävään aineistoon `Merge` -toiminnolla. Tässä valikossa voidaan aineisto jakaa osiin `Split file` -toiminnolla ja suorittaa ainoastaan tiettyjen havaintojen valinta analyysiin `Select cases` -toiminnolla. Lisäksi muuttujia voidaan painottaa tarpeen mukaan eri tavoilla `Weight cases` -toimintoa käyttäen.

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 analyysiin. Valikko sisältää muun muassa mahdollisuuden erilaisten tarkastelujen tekemiseen (`Reports`), frekvenssit (`Descriptive statistics/Frequencies`), tilastolliset tunnusluvut (`Descriptive statistics/Descriptives`), muuttujan ominaisuuksien tutkimiseen (`Descriptive statistics/Explore`), 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 peruslisenssi 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`), viiksikuvio (`Boxplot`), hajontakuvat (`Scatter/Dot`) ja histogrammit (`Histogram`). `Chart builder` -valikossa on erilaisia graafeja, halutun graafin voi valita kenttään ja sen jälkeen täytyy vain määrittää käyttämänsä muuttujat. Toimintoon 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 ns. ”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 erittäin tärkeä opetella käyttämään!

Helpin tärkeimmät toiminnot:

Topics:

- Topics/Sisälllys: 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 sisällystoiminnolla ei tunnu löytyvän mitään.

Tutorial: Näyttää kohta kohdalta SPSS:n peruskäyttöä. Tämän avulla aloittelijan on hyvä tutustua ohjelmaan.

Case studies: SPSS:n käytön esimerkkejä. Suosittelen tutustumaan.

Statistics Coach eli ”tilastollinen valmentaja”, joka opettaa sinua analyysimenetelmien käytössä.

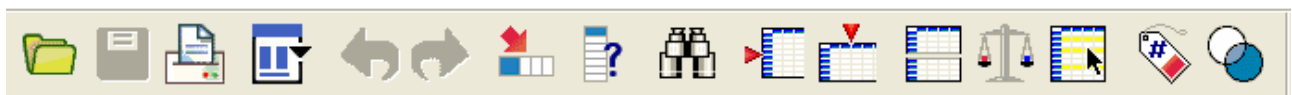
Command Syntax Reference: PDF-muodossa oleva opas SPSS-komentokielen käyttöön.

Algorithms: PDF-muodoissa runsaasti tilastollisten menetelmien matematiikkaa.

Helpistä löytyvät myös linkki SPSS-ohjelman kotisivuille (SPSS Home Page), joilta löytyy lisätietoja ohjelmasta.

Aineistoikkunan pikanäppäimet

Kuten edellä lyhyesti tuli jo esille, aineistoikkunan, kuten myös muidenkin ikkunoiden pikanäppäimiä voi vaihtaa valikosta View. Tässä esitellään muokkaamattomat aineistoikkunan pikanäppäimet ja niiden toiminnot:



Avaus Tallennus Tulostus

Undo Redo

Muuttujat

Lisää
havainto/muuttuja

Split file

Select
cases

Käytä ”muut-
tujasettiä”

Edelliset
toiminnot

Siirry
havaintoon

Etsi

Havaintojen
painottaminen

Näytä
muuttujien
arvoselitteet

2 HAVAINTOAINEISTON MUODOSTAMINEN, MUOKKAAMINEN JA HALLINTA

2.1 Havaintoaineiston muodostaminen

Tässä luvussa tarkastellaan SPSS-havaintoaineiston muodostamista. Nykypäivänä internetissä tehtävät kyselyt ovat yleistyneet, ja ko. kyselyihin saadut vastaukset ovat niistä suoraan siirrettävissä tilastollisiin tietojenkäsittelyohjelmiin. Kuitenkin näissäkin tapauksissa havaintoaineiston muuttujapohjan tekeminen vaatii yleensä käsityötä.

2.1.1 SPSS-tiedoston tallentaminen

SPSS-ikkunoista aineisto-ikkunan, tulostusikkunan ja syntaksi-ikkunan sisällöt **pitää kaikki tallentaa erikseen**. HUOM. Aineistoikkuna ja muuttujaikkuna kuitenkin tallentuvat samanaikaisesti kummasta tahansa ikkunasta. SPSS antaa tallennetulle 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-tiedostotyytit voi kuitenkin avata mistä tahansa ikkunasta, kunhan muistaa valita avaus-valintaikkunassa oikean tiedostotyytin tai All Files -vaihtoehdon).

2.1.2 Uuden aineiston tallennus

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

Aineistopohja tulee suunnitella tarkkaan. Tulee miettiä ja ottaa huomioon seuraavat asiat:

- Millä tavalla muuttuja kannattaa koodata.
- Muuttujat tallennetaan tallennuspohjaan yksi kerrallaan.
- Mikä nimi (Name) muuttujalle on kuvaavin ja helpoin muistaa (muuttujan nimen tulee alkaa kirjaimella). 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 (Label) ja määrittää luokitellun muuttujan mahdolliset arvot selitysteksteineen (Values).
- Measure-kenttään tulee kertoa muuttujan mittaustaso: luokiteltu muuttuja: nominaaliasteikollinen (Nominal) tai ordinaaliasteikollinen (Ordinal), jatkuva muuttuja: (Scale). Kenttä on tärkeä täyttää oikein, sillä enenevässä määrin SPSS käyttää ko. tietoa muuttujan mittaustasosta hyödykseen analyysejä tekemisessä. Toisin sanoen ohjelma auttaa ohjelman käyttäjää esim. tilanteissa, joissa käyttäjä on ottamassa nominaaliasteikollisesta muuttujasta keskiarvoa.

Aineiston tallennuksen peruseriaatteen:

- Vasta sitten, kun tallennuspohja on valmis, voidaan alkaa syöttää dataa.
- Data on järkevintä syöttää yksi rivi eli yhden vastaajan tiedot kerrallaan. Perinteisissä kyselytutkimuksissa (vs. web-lomaketutkimukset) kullekin vastauslomakkeelle kannattaa antaa havaintotunnus (eli numeroida ne), joka syötetään havaintotunnusmuuttujan arvoksi. Tämä helpottaa mahdollisten tallennusvirheiden korjaamista. Havaintotunnuksen perusteella löytää nimittäin kätevästi vastauslomakkeen, jota on tallennettu virheellisesti.

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 mahdolliseksi 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 jälkepäin on hyvin vaivatonta. Tästä puhutaan myöhemmin uudelleenluokittelun yhteydessä luvussa 2.2.1 sekä 2.2.4.

Siviilisäät

Valmiiksi luokitettavat vaihtoehdot (naimaton, naimisissa, eronnut, leski) koodataan yhden muuttujan alle mahdolliseksi vastausvaihtoehdoiksi (vastaaja voi valita ainoastaan yhden vaihtoehdon). Tällöin esim. arvo yksi vastaa naimatonta, kaksi naimisissa olevaa, kolme eronnutta ja neljä leskeä.

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. Arvolle voi antaa ja kannattaakin antaa selitysteksti normaaliin tapaan, esim. arvo "99", selite "aviossa". Tällöin analyyseissä huomioidaan erikseen ne vastaajat, joiden ei ole edes oletettu kysymykseen vastaavan. Jos henkilö on jättänyt vastaamatta kysymykseen jostakin tuntemattomasta syystä, kyseinen kohta jätetään tyhjäksi ilmentämään puuttuvaa tietoa.

Ammatti

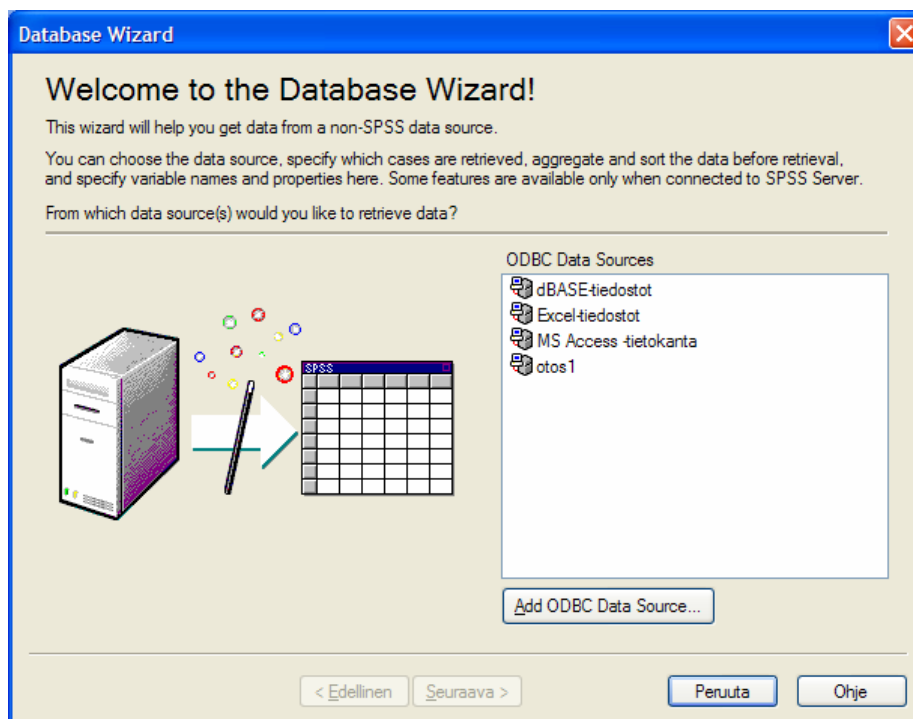
Yleensä vastaajalta tiedustellaan kyselylomaketutkimuksessa avoimella kysymyksellä hänen ammattinsa, jolloin vastaukset ovat tekstiä; eli eri ammatteja. Muuttuja tulee tallennuspohjaa tehdessä määritellä `String`-muuttujaksi (jolloin muuttuja voi sisältää joko tekstiä tai numeroita tai molempia, tässä esimerkissä tekstiä). Tallennusvaiheessa vastaajien eri ammatit tallennetaan siis teksteinä ja aineiston tallennuksen jälkeen SPSS muuttaa haluttaessa `Automatic recorde`-toiminnolla vastaukset ammattiluokkiin antaen niille numeeriset arvot, ja näin moninaisten analyyseiden tekeminen mahdollistuu. Tämähän ei ole mahdollista tekstimuuttujien kohdalla.

2.1.3 Database Wizard

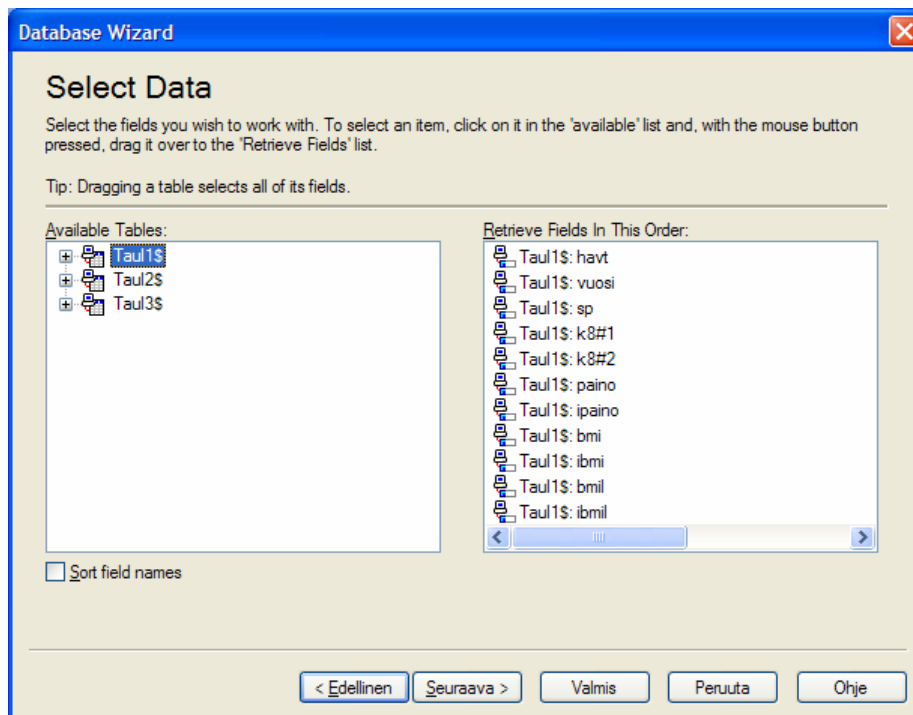
SPSS:n versiot ennen versiota 14 lukevat Excelin 5 tai myöhempää versiota suoraan. Aineiston lukeminen tapahtuu seuraavasti: avaa Excel-tiedosto SPSS:n File-valikosta komennolla Open data. Huom. tiedostotyyppiä valitaan Excel (*.xls). Siirrettävän Excel-tiedoston tulee SPSS:ään avaamisen aikana olla kiinni, muutoin SPSS ei pysty sitä lukemaan! 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 määrittää niiden muuttujien solutiedot (eli sijaintitiedot), jotka SPSS:ään halutaan ottaa käytettäviksi. Tarvittaessa löydät eri tiedostojen tuomisesta SPSS:ään lisätietoa SPSS:n Helpistä (Tutorial/Reading data).

HUOM! Syystä tuntemattomasta toisissa SPSS:n 14 versioissa Excel-tiedosto luetaan edellä esitettyyn tapaan ja toisissa seuraavassa esitettyyn tapaan.

SPSS:n versiossa 14 Excel-tiedoston (kuten myös esim. Access-tiedostojen) lukeminen tapahtuu hieman eri tavalla. File-valikosta valitaan toiminto Open database, josta valitaan New query. Tällöin aukeaa seuraava ikkuna:



Valitaan Excel-tiedostot ja painetaan Seuraava-painiketta, jolloin SPSS antaa hakea (Browse) halutun tiedoston. Tiedosto avataan, jonka jälkeen ohjelma kysyy, mitkä taulukot halutaan ottaa mukaan. Tässä esimerkissä halutaan ainoastaan Taulukko 1, jonka jälkeen siirtoikkuna näyttää seuraavalta:



Valmis-painiketta painamalla Excel-aineisto muokkaantuu SPSS-dataksi ja se aukeaa tavallisen SPSS-aineiston tapaan suoraan käytettäväksi, tosin muuttujien selitteet ym. tulee tehdä uudelle SPSS-datalle.

2.2 Havaintoaineiston muokkaaminen

Tutkimusaineistoa ei koskaan tule ottaa annettuna. Valmiit muuttujat ja niiden luokitukset ja asteikot eivät ole pyhiä: tutkija saa muuttaa ja useimmiten hänen onkin muutettava niitä omien tarpeidensa mukaan. Tämä luku tulee avautumaan huomattavasti paremmin sen jälkeen, kun on lukenut myös luvut 3 ja 4. Muuttujamuunnoksia ovat esim.

- Uudelleenluokittelu
- Summamuuttujien tekeminen
- Aritmeettiset muunnokset
- Jne.

2.2.1 Uudelleenluokittelu

Miksi pitäisi uudelleenluokitella?

Muuttujien luokkia yhdisteltäessä katoaa aina informaatiota, joten yhdistelyä ei pidä tehdä turhaan. Yhdistelylle on aina löydettävä perustelu. Toisaalta, joskus muuttujan 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 voidaan tehdä luotettavia johtopäätöksiä.
- Muuttuja voidaan joskus normalisoida (=muuttujan jakauma muistuttaa paremmin normaalijakaumaa) yhdistämällä ääriluokkia.

Tutkija pyrkii tiiviiseen esitykseen ja hävittää aineistosta turhana pitämänsä 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!

On hyödyllistä kuitenkin tarkistaa, onko kadotettava informaatio olennaista. Tutkija voi esimerkiksi osoittaa, että selitettävän muuttujan vaihtelu on samanlaista kahdessa pienehkössä luokassa, ja näin ollen, ne voidaan yhdistää jatkoanalysejä varten.

Usein on niin, että analyysivaiheessa luokkia kannattaa varmuuden vuoksi pitää erillään, mutta raportissa yhdistelyä kannattaa tehdä kovemmallalla 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 (ks. halutessasi tarkemmin kyselylomakkeen tekemisestä FSD:n nettisivuilta).

Uudelleenluokittelun teossa huomioitavaa:

- Pidä huolta, ettei luokkien frekvensseistä tule liian pieniä analyysejä varten.
- Ainakin kolme luokkaa on hyvä olla dikotomian estämiseksi. Kun muuttujalla on ainakin kolme luokkaa, voidaan osoittaa mahdollinen epälineaarinen yhteys. Samalla vältytään myös liialta maailman yksinkertaistamiselta.
- Esimerkiksi ikäryhmiä koskevia luokituksia tehdessä kannattaa miettiä, mitkä ovat ko. tutkimusongelman kannalta mielekkäitä ryhmiä esim. nuoret ja aikuiset
- Esimerkki siviilisäädyn eri luokkien yhdistämisen ongelmasta: useasti sekä eronneet, että erityisesti lesket, ovat havaintoaineistossa frekvenssiltään pieniä ryhmiä, mutta niiden yhdistäminen ei ole järkevää, koska ko. ryhmät poikkeavat toisistaan merkittävästi.

Muuttujan uudelleenluokittelu

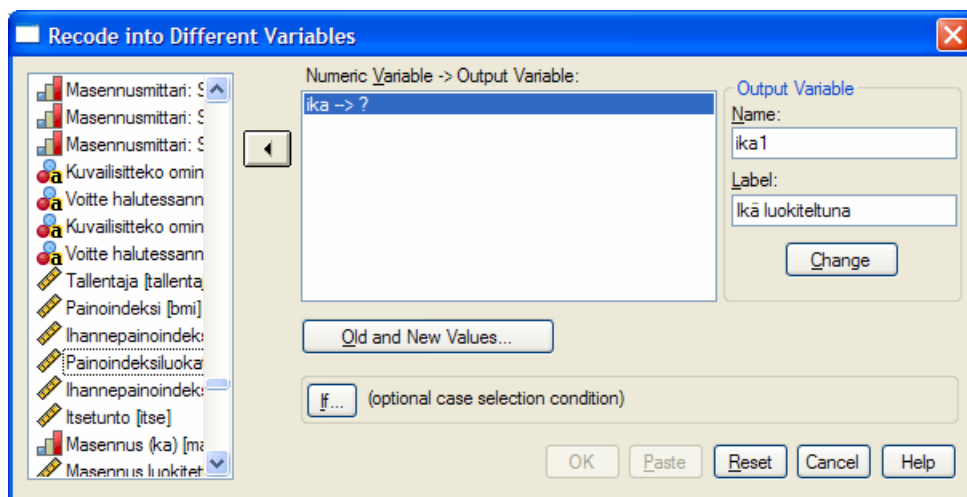
Yleensä kannattaa selostaa raporteissa, mitä uudelleenluokitteluja on käyttämilleen muuttujille tehnyt ja perustella, minkä vuoksi. Kaikkein helpointa ja nopeinta muuttujien uudelleenluokittelu on syntaksi-ikkunassa, josta luvussa 2.2.4. Siitä huolimatta tässä esitetään, miten uudelleenluokittelu tehdään valintaikkunoiden kautta.

Jatkuvan muuttujan luokittelu analyysejä varten

Tässä käytetään jatkuvan muuttujan luokittelussa esimerkkinä vuoden tarkkuudella mitattua ikämuuttujaa (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 esimerkkitapauksessa saatiin selville, että minimiarvo on 18 ja maksimiarvo 75. Seuraavassa vaiheessa mietitään, minkälainen luokittelu olisi oman tutkimuksen kannalta paras. On havainnollista myös tulostaa tarkemmat tiedot muuttujasta, jotka saa tilattua Analyze/Descriptive statistics /Frequencies -valikosta sekä lisäksi histogrammi (jonka saa tilattua samasta valikosta Chart-painikkeen takaa), josta myös näkee, miten ikä on jakautunut. Tehdyistä tarkasteluista huomataan vaikkapa, että vanhoja ihmisiä on paljon vähemmän kuin keski-ikäisiä. Jos muuttujaa aiotaan käyttää ristiintaulukossa, kannattaa ehkä ottaa vähän isompi luokitusväli "vanhuksille", jotta ko. luokan frekvenssit pysyvät riittävän suurina. Toinen asia on sitten se, jos ollaan kiinnostuneita erityisesti vanhoista ihmisistä ja halutaan tarkempaa informaatioita heistä. Tässä kuvitteellisessa esimerkissä tutkija haluaa oman tutkimusongelmansa takia yhdistää luokat seuraavasti: 18–30v luokaksi 1, 31–45v luokaksi 2, 46–55 luokaksi 3, 56–65 luokaksi 4 ja 66–75 luokaksi 5. Huomaa, että sama ikä ei saa olla yhtä aikaa kahdessa eri luokassa!

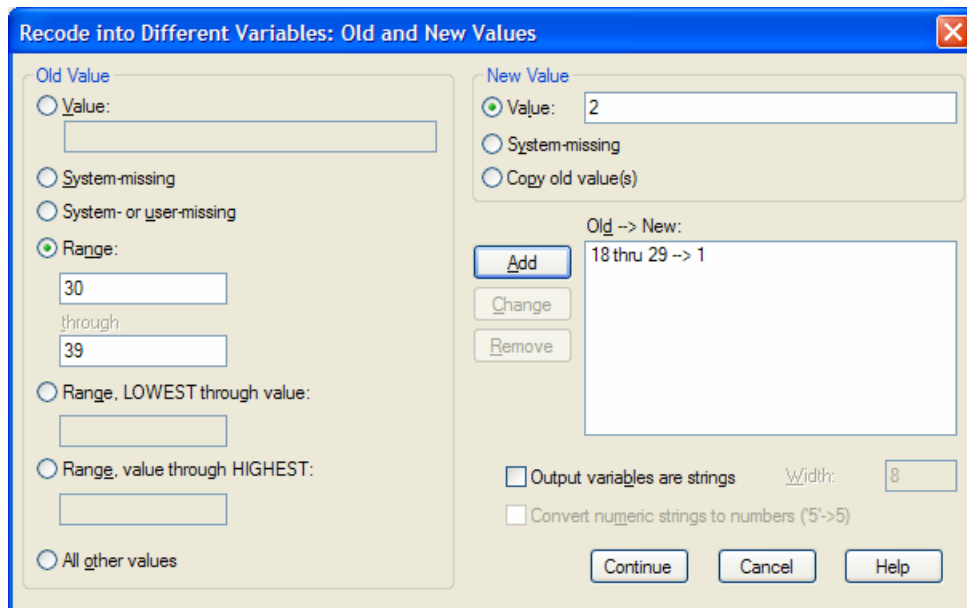
Uudelleenluokittelun tekeminen valintaikkunoiden kautta


Uudelleenluokittelua pääsee tekemään Transform-valikossa (Transform/Recode/Into different variables). Into different variables-toiminto luo uuden, luokitellun muuttujan ja säilyttää alkuperäisen luokittelemattoman ikämuuttujan, toisin kuin Into Same variable -toiminto, joka korvaa alkuperäisen muuttujan uudella muuttujalla. En tiedä, mistä syystä kannattaisi valita toiminto Into same variable..., joka kadottaa alkuperäisen informaation! Uudelleenluokittelua tehdessä Into different variable, saadaan näkyville seuraava ikkuna:



Valitse muuttujaluettelosta muuttuja eli tässä tapauksessa ika, jonka arvot haluat uudelleenluokitella Numeric variable -> Output variable -kenttään. Output Variable -kenttään anna uuden muuttujan nimi (Name) ja sen selitysteksti (Label). Uuden muuttujan nimeämisessä on noudatettava samoja sääntöjä, kuin muuttujien nimeämisessä muutenkin (eli muuttujan tulee 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 –painiketta. Näin avautuu seuraava ikkuna:



Valitse tilannekohtaisesti Old value –kentässä sopivin vanhoja arvoja koskeva kohta, ja täytä haluamasi uudelleenluokitettavan muuttujan arvovälit ja/tai arvo uudelleenluokitettavaksi. Yksittäisiä lukuarvoja voi luokitella uudestaan value-kohdassa. 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 seuraavat vanhat arvot ja uuden arvon ja painat Add jne. Kun kaikki arvot on uudelleenluokiteltu, paina **Continue** ja Recode-valintaikkunassa paina OK painikkeen sijasta **Paste**-painiketta, jolloin uudelleenluokittamasi muuttujaa koskevat tiedot tulostuvat syntaksi-ikkunaan. Tämän jälkeen maalaa syntaksi-ikkunassa tekemäsi komento ja paina RUN-painiketta . Tällöin ohjelma tekee uuden muuttujan aineistoosi. **Mikäli et käytä Paste-toimintoa, sinulle ei jää mihinkään tietoa siitä, miten olet muuttujan uudelleenluokitellut!**

Huom! Uudelleenluokitellun muuttujan arvoille tulee myös antaa niiden selitystekstit (Values) Data Editor –ikkunassa ennen kuin alkaa käyttää sitä analyyseissä, taulukoissa ja grafiikassa. Tällöin selitystekstit tulostuvat luokkien numeroiden tilalle, mikä on paljon havainnollisempaa! Uudet muuttujat asettuvat aina aineistoikkunan (eli Data Editorin Data view –välilehdelle) 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 voisi periaatteessa antaa, mitkä numeeriset arvot tahansa, esim. juu voisi olla vaikka 55 ja ei 99 jne. Loogisinta kuitenkin on (etenkin jos haluat jonkun muunkin kuin sinun ymmärtävän työstäsi jotakin), että uusille luokille annetaan numerot numerojärjestyksessä yhdestä eteenpäin.

2.2.2 Summamuuttujat

Yksinkertaisimpien summamuuttujien perusajatuksena on ajatus yksiulotteisesta ilmiöstä, jota on mitattu usealla mittarilla (esimerkiksi kysyty 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.

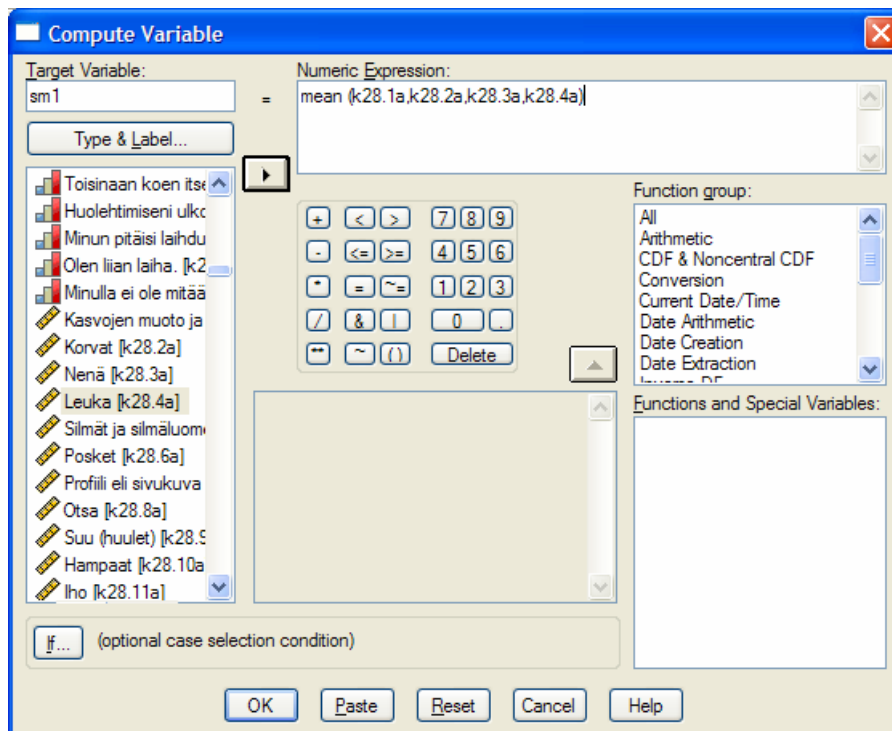
Pääsääntöisesti summamuuttujan osioiden välisten korrelaatioiden on oltava nimenomaan positiivisia. Summamuuttujan osiot onkin usein käännettävä samansuuntaisiksi ennen summaamista. Esimerkiksi jos summamuuttujaan kerätään asennemuuttujia, joissa osa asenneväittämistä on esitetty myönteisessä ja osa 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 taitoineet, vaikka esimerkiksi matematiikan ja liikunnan taidot korreloisivat oppilasjoukossa negatiivisesti.

Käytännön esimerkki summamuuttujasta: kyselyssä on kysyty 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 valintaikkunoiden kautta

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



Target value –kenttään annetaan 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. Tällöin esimerkki kannattaa lisätä jakotoimitus (kissa+koira+hamsteri)/3. Esimerkiksi likert-asteikolliset muuttujat: vastausvaihtoehdot: täysin samaa mieltä, jokseenkin samaa mieltä, ei osaa sanoa, jokseenkin eri mieltä, täysin eri mieltä. Vastausvaihtoehtojen vastausskaala vaihtelee siis yhdestä viiteen. Uuden muuttujan arvo pysyy tämän skaalan sisällä, mikäli jakotoimitus tehdään. Tämä helpottaa ajattelua siitä, mitä tarkoittaa esimerkiksi se, kun keskiarvo uudella summamuuttujalla on 1.8. 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). Vastaavasti kuin uudelleenluokittelun yhteydessä tuotiin esille, uuden summamuuttujan määrittelyn jälkeen tulee OK-painikkeen sijasta painaa Paste-painiketta, jotta syntaksiin jää tieto siitä, miten uusi muuttuja on tehty. Tämä tietenkin edellyttää sitä, että syntaksi-ikkuna tallennetaan.

2.2.3 Muut muunnokset

Aritmeettiset muunnokset:

Mittayksikön vaihtaminen

Mittayksikön vaihtaminen tapahtuu samassa valikossa kuin summamuuttujien tekeminen eli Transform/Compute –valikossa. Tässä muutamia esimerkkejä tilanteista, joissa mittayksikköä voi olla tarpeellista muuttaa: aika vuorokausista tunneiksi, etäisyys maileista metreiksi, markat euroiksi tai energia kilokaloreista kilojouleiksi. Mittayksikön vaihto esimerkiksi seuraavassakin

tilanteessa helpottaa analyysien tekemistä: Koodataan koulutusastemuuttuja, joka saa arvoja 'perusaste', 'keskiaste' ja 'korkea-aste' koulutusvuosiksi. Etuna tämän muunnoksen tekemisessä on se, että koulutusvuosimuuttujaa voidaan käyttää välimatka-asteikollisena muuttujana, jolloin käytettävissä olevien analyysimenetelmien määrä lisääntyy.


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 yksinasuvalle kuin suurelle lapsiperheelle. Edessä on aritmeettisen muunnoksen tarve.

2.2.4 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 **RUN**-toiminnolla (pikapainike ).

Uuteen syntaksi-ikkunaan päästään valitsemalla *File*-valikosta toiminto *New/Syntax*. Aiemmin luotuun tiedostoon pääsee valitsemalla *File*-valikosta *Open/Syntax*.

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 Khi^2 -testi (erojen tilastollisen merkitsevyyden -testi). Muuttujat ovat koti (koti) ja sukupuoli (sp).

Uudelleenluokittelu syntaksi-ikkunassa

Luvussa 2.2.1 käytiin läpi muuttujan uudelleenluokittelu valikkojen kautta. 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 75=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-painiketta. Eikö olekin paljon selkeämpää kuin monimutkaisten valintaikkunoiden käyttö?

Esimerkki **Helpin** käytöstä syntaksikomentojen tekemiseen: Help/Topics/Hakemisto, johon kirjoita Crosstabs ja valitse näkyville tulevista vaihtoehdoista Command Syntax. Älä myöskään unohda Helpissä PDF-muodossa olevaa Syntaksiopasta Help/Command Syntax Reference. Oppaassa on yli 2 000 sivua, joten syntaksin käyttöön tarpeellista tietoa on yllin kyllin tarjolla. Opasta kannattaa käyttää ennen kaikkea hakuteoksena tiettyyn ongelmaan.

- Helpissä vapaavalintaiset määrittelyt tai alakomennot esitetään hakasulkeissa.
- Keskenään vaihtoehtoiset määrittelyt aaltosulkeissa.
- Komennot isoilla kirjaimilla, käyttäjän antamat määreet pienillä.
- Käytännössä etenkin aloittelijan on helpompaa opetella syntaksikomentoja tekemällä niitä valikkojen kautta esim. Analyze/Descriptive statistics/Crosstabs ja painamalla Paste-painiketta, jolloin valikkojen kautta tehty komento saadaan syntax-ikkunaan. Saman mallin mukaan voi jatkossa kirjoittaa samantapaisia komentoja itse.
- Kirjoitetut komennot toteutetaan painamalla syntaksi-ikkunassa olevaa RUN-painiketta.
- Komentojen toteuttaminen tapahtuu kätevimmin niin, että toteuttamiskelpoiset komennot maalataan hiirellä valituksi alueeksi, ja jonka jälkeen painetaan RUN-painiketta.
- Jokaisen komennon tulee päättyä EXECUTE.-komentoon.

Miksi käyttää syntaksi-ikkunaa?

- Transform-valikon toiminnot (muuttujamuunnokset) ovat helpompia tehdä syntaksissa. Lisäksi kaikkia muuttujamuunnoksia tai muita käskyjä ei pysty tekemään ollenkaan valikkojen avulla.
- Tässä täytyy vielä erikseen korostaa sitä, että muuttujamuunnokset eivät tallennu SPSS-ohjelmaan millään tavalla, mikäli niitä ei tallenna syntaksi-ikkunaan.
- Jos haluaa tehdä toistoja esim. ristiintaulukoita eri muuttujille, on nopeinta vaihtaa vain muuttujien nimet syntaksi-ikkunassa.
- Useat samanlaiset selitystekstit muuttujien arvoille (Values) on kätevämpi tehdä syntaksissa.
- Jos haluaa **dokumentoida** (joka on aina kannattavaa!) sen, mitä on tehnyt, "lokityedosto" Edit/Options/General/Session Journal/Browse näyttää kaiken, mitä SPSS-ohjelmalla on tehnyt. Tämä toiminto kannattaa ottaa käyttöön. Toiminnosta on hyötyä neuvontatilanteissa sekä oman muistin paikkaamisessa sen suhteen, mitä ja miten on tehnyt muuttujamuunnoksia ja analyysyjä. "Lokityedosto" avataan syntaksi-ikkunaan.

”Lokitiedoston” ongelmana on se, että se tallentaa kaikki tehdyt toiminnot, kun taas syntaksiin voi tallentaa ainoastaan ne analyysit etc., joita lopulta on käyttänyt.

- Loppujen lopuksi kun syntaksin oppii, minkä tahansa homman tekeminen on sillä nopeampaa kuin valikkojen avulla.

Osa-aineiston tekeminen

Edellä lyhyesti tuotiin esille, että kaikkia tarvittavia ja varsin hyödyllisiä toimintoja ei voi tehdä valikkojen kautta vaan ainoastaan syntaksissa, yksi niistä on osa-aineiston tekeminen. Tilanteessa, jossa jostain isommasta aineistosta halutaan muodostaa pienempi aineisto, tehdään jokin seuraavista komennoista syntaksi-ikkunassa:

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

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

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

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

2.3 Havaintoaineiston hallinta

SPSS:n keskeisten toimintojen esittely tässä vaiheessa opasta voi tuntua vieraalta, ja siksi suosittelenkin niihin tutustumista erityisesti sen jälkeen, kun SPSS-ohjelman perusteet ovat hallinnassa. Esiteltävät toiminnot ovat kuitenkin erittäin keskeisiä ja helpottavat suunnattomasti analyysien tekemisessä.

Data valikossa:

Split file:

Käytetään tilanteessa, jossa halutaan tehdä erilliset analyysit tietyille havaintoryhmille esim. miehille ja naisille. Tämä on erittäin hyödyllinen toiminto.

Select cases:

Tämä 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 cases* -toiminnon ero *Split file* -toimintoon: tässä toiminnossa 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 jokaiselle luokalle erikseen, kuten *Split file* -toimintoa käyttäessä.

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 response:

SPSS-ohjelmassa muuttujat voivat saada ainoastaan yhden arvon. Multiple response -toiminnon avulla havainnoille 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 helppoa 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 ns. ”muuttujasetti” (Analyze /Multiple response/Define sets). Kun ”setti” on tallennettu, voidaan tarkastella ”setin” frekvenssejä sekä tehdä ristiintaulukoita.

Utilities-valikossa:

Define sets:

Tällä toiminnolla (Utilities/Define Sets) voi tehdä haluamistaan muuttujista (muuttujaryhmistä) ”muuttujasetin”. Kun ”setti” on määritelty, se valitaan

tarkasteluun toiminnolla `Utilities/Use sets`. Tämä toiminto on käytännöllinen silloin, kun tarkastelee joitakin muuttujia ryppäänä. Tällöin ei tarvitse aina klikata kaikkia muuttujia erikseen muuttujaluetteloon.

2.4 Havaintoaineiston tarkastelu ja kuvaaminen

Seuraavissa luvuissa (luvut 3 ja 4) käsitellään havaintoaineiston tarkastelua ja kuvaamista. Tässä puhutaan lyhyesti tulevien lukujen johdantona siitä, mikä niiden merkitys ja asema tutkimuksessa on.

Tutkimusaineiston tärkeimmät tarkastelukeinot ovat:

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

Erlaisia 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 muuttujan luokat näkyviin.

Tutkimustulosten esittämistapaa tutkimusraportissa kannattaa pohtia tarkoin. Onko mahdollista esittää tulokset tekstinä ilman, että tekstistä tulee liian raskasta, pitkää tai sekavaa? Olisiko taulukoiden tai kuvien käyttö ko. tutkimusongelman kohdalla sittenkin 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 TAULUKOINTI

3.1 Yleistä taulukoinnista

- 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 esitetään yleensä liitteessä.

Taulukon käytön ehtoja

Frekvenssitaulukoiden etu on sen riippumattomuus mitta-asteikosta. Ristiintaulukoilla voidaan kuvata myös laatueroasteikollisten muuttujien välisiä yhteyksiä: Esimerkiksi havaintoaineistossa olevien ihmisten jakautumista siviilisäätyluokkiin sosiaaliluokan mukaan. Kolmiulotteisessa taulukossa mukaan voidaan ottaa esimerkiksi sukupuoli, 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 sen sijaan on 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.

Ristiintaulukoinnin ongelmaksi voi muodostua tarvittavien **havaintojen määrä**. Laajojen ja moniulotteisten ristiintaulukoiden tekemiseen tarvitaan runsaasti havaintoja.

Taulukoiden raportointi

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

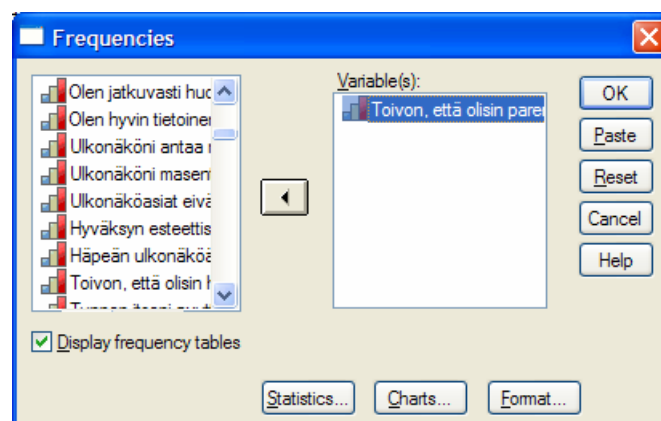
- Esitetään vain rivi-, sarake- tai soluprosentit (ja tietenkin se oikea vaihtoehto on valittu!), eli kussakin solussa on vain yksi prosenttiluku (mutta ei prosenttimerkkiä). Solufrekvenssejä ei ole tarpeen esittää.
- Esitetään myös luvut, joista prosentit on laskettu, eli rivi- ja sarakesummat ja niiden 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 koettaa 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. Tilanteesta riippuen joko yhden tai kahden desimaalin tarkkuus on paras mahdollinen esittämistapa.
- 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 tulee olla taulukossa sanallinen selitys. Tämä onnistuu helposti varmistamalla ennen taulukoiden tekoa, että kaikille käytetyille muuttujille on annettu aineistoikkunassa arvojen selitystekstit (*values*). Jos muuttujan/muuttujien luokkien kuvaukset eivät mahdu taulukon riveille ja sarakkeille, käytetään niistä lyhennettä ja selostusta jatketaan alaviitteessä/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.
- Paras mahdollinen vinkki: Näytä taulukoitasi työkaverillesi, opiskelukaverillesi 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. Siksi otsikosta tulee tulla ilmi, mistä taulukosta on kyse.
- Jos on kyse prosenttiluokasta, tulee havaintojen lukumäärä laittaa esille.

3.2 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 grafiikkaa.

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.

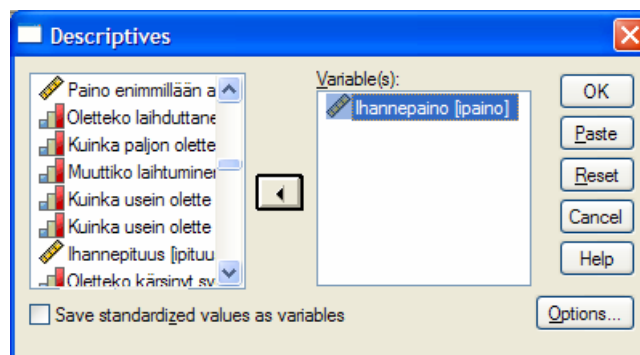
Taulukko 1. Vastaukset väittämään ”Toivon, että olisin paremmannäköinen”.

		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		

3.3 Muuttujan perustunnusluvut

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

Keskeisimmät 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 raportoimisen yhteydessä esitetään yleensä myös keskihajonta.

Taulukko 2. Naisten ihannepaino.

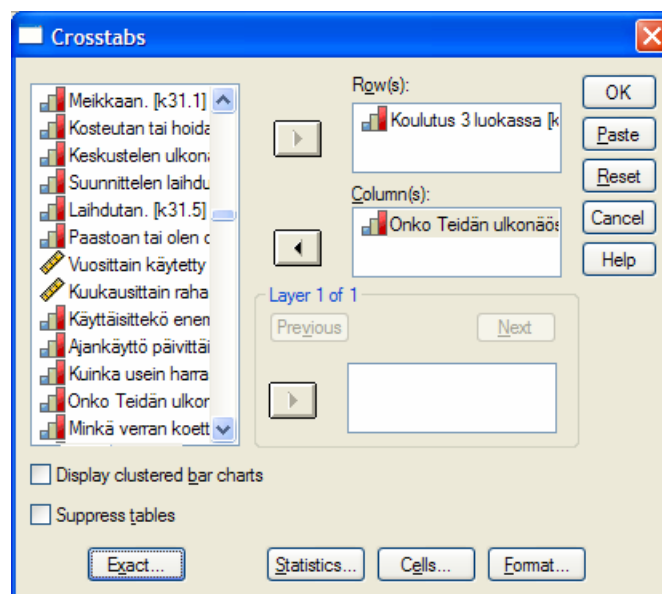
	N	Minimum	Maximum	Mean	Std. Deviation
Ihannepaino	251	40,0	80,0	60,118	6,2507
Valid N (listwise)	251				

3.4 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 mainituin ominaisuuksin varustettua yksilöä aineistossa on. Melkein aina tilastollisessa tutkimuksessa on kuitenkin havainnollisempaa tarkastella prosentteja kuin pelkkiä frekvenssejä. Kuitenkin vastausten lukumäärän tulee ilmetä taulukosta tai taulukon otsikosta.

Ristiintaulukkoa pääsee tekemään analyysivalikossa (Analyze/Descriptive statistics/Crosstabs). Näin saadaan näkyville seuraava 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. Lyhyesti testin käyttämisessä on huomioitava sen käytön edellytykset, jotka ovat seuraavat:

- Korkeintaan 20 % odotetuista frekvensseistä saa olla pienempiä kuin 5.
- Jokaisen odotetun frekvenssin on oltava vähintään 1.

HUOM. Odotetut frekvenssit antavat vertailukohdaksi sellaiset frekvenssit, jotka syntyisivät, jos ristiintaulukon luokkien väliset erot johtuisivat pelkästä sattumasta. Mikäli ristiintaulukossa on yksikin tyhjä solu, testin käytön edellytykset eivät täyty. SPSS-ohjelma tulostaa kuitenkin aina testisuureet, vaikka käyttöedellytykset eivät täyty. Kun edellytykset eivät ole voimassa, tieto on luettavissa testitaulukon alalaidasta, ja siihen tulee aina kiinnittää huomio.

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

Format: täällä voidaan vaikuttaa hieman taulukon ulkoasuun.

Seuraavassa taulukossa (taulukko 3) on esitetty naisten tyytymättömyys johonkin omassa ulkonäössään koulutusasteen mukaan tarkasteltuna. Taulukossa on esitetty frekvenssit, riviprocentit ja sarakeprosentit. Riviprocentit on esitetty jokaisessa solussa toisella rivillä ja sarakeprosentit kolmannella rivillä. 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?		Total		
		Kyllä	Ei			
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) tarkastellaan asuinympäristön ja kavennetun painoindeksin välistä yhteyttä. Lisäksi taulukossa on esitetty absoluuttisten frekvenssien lisäksi rivi- ja sarakeprosentit sekä odotetut frekvenssit (expected count). Mitä taulukosta on nähtävissä, onko asuinympäristö ja painoindeksi yhteydessä toisiinsa?

Taulukko 4. Painoindeksi asuin ympäristön mukaan.

		Painoindeksi kahdessa luokassa			
			Ali- tai normaali-painoinen	Ylipainoinen	Total
asuin ympäristö	Kaupunkimainen	Count	99	35	134
		Expected Count	86,8	47,2	134,0
		% within asuin ympäristö	73,9%	26,1%	100,0%
		% within Painoindeksi kahdessa luokassa	57,2%	37,2%	50,2%
	Maaseutumainen	Count	74	59	133
		Expected Count	86,2	46,8	133,0
		% within asuin ympäristö	55,6%	44,4%	100,0%
		% within Painoindeksi kahdessa luokassa	42,8%	62,8%	49,8%
Total		Count	173	94	267
		Expected Count	173,0	94,0	267,0
		% within asuin ympäristö	64,8%	35,2%	100,0%
		% within Painoindeksi kahdessa luokassa	100,0%	100,0%	100,0%

- Odotetut frekvenssit antavat vertailukohdaksi sellaiset frekvenssit, jotka syntyisivät, jos ristiintaulukon luokkien väliset erot johtuisivat pelkästä sattumasta.
- 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 prosentteiksikin tulee valita riviprocentit. Joten ei niin, kuin näissä aikaisemmissa taulukoissa on esitetty. Kuitenkin on makuasia, haluaako selittävän muuttujan riville vai sarakkeisiin, kunhan tulostaa taulukkoon selittävän muuttujan prosentit.
- Edellä esitetyt ristiintaulukot ovat liian isoja esitettäväksi tutkimusraportissa. Karsiminen on tarpeen, sillä näin isoja taulukkoja on jo vaikeaa hahmottaa ja tulkita, eivätkä ne vastaa ristiintaulukoille asetettuja normeja. Tosin tutkijalle itselleen on tärkeää tarkastella monimuotoisesti ristiintaulukkojen eri arvoja.

Palataan vielä edellä esitettyyn kysymykseen, eli onko asuin ympäristö ja painoindeksi yhteydessä toisiinsa? Kyllä ne ovat. Kaupunkimaisessa asuin ympäristössä asuvista naisista 74 % on ali- tai normaalipainoisia. Maaseutumaisessa asuin ympäristössä asuvista naisista huomattavasti pienempi osuus, 56 %, on ali- tai normaalipainoisia. Seuraavassa taulukossa (taulukko 4A) on esitetty muiden testien lisäksi χ^2 -riipumattomuustesti, joka on esitetty taulukon ensimmäisellä rivillä. Siitä nähdään, että näiden muuttujien välillä on tilastollisesti merkitsevä yhteys. Lisäksi taulukon alalaidasta on nähtävissä, että χ^2 -testin käytön edellytykset ovat voimassa.

Taulukko 4A. Painoindeksi asuinympäristön mukaan taulukon Chi-square –testin tulostaulukko.

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9,737 ^b	1	,002		
Continuity Correction ^a	8,954	1	,003		
Likelihood Ratio	9,818	1	,002		
Fisher's Exact Test				,002	,001
Linear-by-Linear Association	9,700	1	,002		
N of Valid Cases	267				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 46,82.

3.5 Useampiulotteinen ristiintaulukko

Useampiulotteinen ristiintaulukko 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 johtopäätöksiä.

Tässä otetaan tarkastelun kohteeksi edellinen kahden muuttujan ristiintaulukko (taulukko 4). Nyt haluamme tarkastella asuinympäristön ja painoindeksin välistä yhteyttä kolmannen muuttujan, iän määrittelemissä ryhmissä. Kiinnostuksen kohteena on siis se, tuoko ikä havaitun yhteyden tarkasteluun jotain uutta informaatiota, vai onko asuinympäristön ja painoindeksin välinen yhteys samanlainen kaikissa ikäryhmissä. Käytännössä tämä tapahtuu siten, että samassa valikossa, jossa teimme kahden muuttujan ristiintaulukon, me lisäämme Layer-kohtaan luokitellun ikä-muuttujan. Ristiintaulukon soluihin otetaan riviprosenttien lisäksi mukaan havaintojen lukumäärä kussakin solussa. Seuraavassa on esitetty tämä kolmen muuttujan ristiintaulukko ja sen jälkeen taulukko, jossa on esitetty χ^2 -testi. Mitä niistä on nähtävissä? Toiko iän mukaanotto merkittävää lisätietoa asuinympäristön ja painoindeksin välisen yhteyden tarkasteluun? Vastauksen etsimiseksi kannattaa vertailla kunkin ikäluokan kohdalla kaupunkimaisessa vs. maaseutumaisessa asuinympäristössä asuvien ali- tai normaalipainoisten tai vaihtoehtoisesti ylipainoisten naisten prosentuaalista osuutta.

Taulukko 5. Painoindeksi asuin ympäristön mukaan ikäluokittain tarkasteltuna.

Ikä kolmessa luokassa				Painoindeksi kahdessa luokassa		
				Ali-tai normaali-painoinen	Ylipainoinen	Total
18-34	asuin-ympäristö	Kaupunkimainen	Count	44	12	56
			% within asuin-ympäristö	78,6%	21,4%	100,0%
	Maaseutumainen	Count	25	8	33	
		% within asuin-ympäristö	75,8%	24,2%	100,0%	
	Total	Count	69	20	89	
		% within asuin-ympäristö	77,5%	22,5%	100,0%	
35-54	asuin-ympäristö	Kaupunkimainen	Count	43	15	58
			% within asuin-ympäristö	74,1%	25,9%	100,0%
	Maaseutumainen	Count	38	27	65	
		% within asuin-ympäristö	58,5%	41,5%	100,0%	
	Total	Count	81	42	123	
		% within asuin-ympäristö	65,9%	34,1%	100,0%	
55-74	asuin-ympäristö	Kaupunkimainen	Count	12	8	20
			% within asuin-ympäristö	60,0%	40,0%	100,0%
	Maaseutumainen	Count	11	24	35	
		% within asuin-ympäristö	31,4%	68,6%	100,0%	
	Total	Count	23	32	55	
		% within asuin-ympäristö	41,8%	58,2%	100,0%	

Taulukko 5A. Painoindeksi asuin ympäristön mukaan ikäluokittain tarkasteltuna taulukon Chi-square –testin tulostaulukko.

Ikä kolmessa luokassa		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
18-34	Pearson Chi-Square	,094 ^b	1	,759		
	Continuity Correction ^f	,002	1	,965		
	Likelihood Ratio	,094	1	,760		
	Fisher's Exact Test				,796	,477
	Linear-by-Linear Association	,093	1	,760		
	N of Valid Cases	89				
35-54	Pearson Chi-Square	3,350 ^c	1	,067		
	Continuity Correction ^f	2,689	1	,101		
	Likelihood Ratio	3,387	1	,066		
	Fisher's Exact Test				,087	,050
	Linear-by-Linear Association	3,322	1	,068		
	N of Valid Cases	123				
55-74	Pearson Chi-Square	4,270 ^d	1	,039		
	Continuity Correction ^f	3,177	1	,075		
	Likelihood Ratio	4,272	1	,039		
	Fisher's Exact Test				,050	,037
	Linear-by-Linear Association	4,193	1	,041		
	N of Valid Cases	55				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 7,42.

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 19,80.

d. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8,36.

3.6 Keskiarvotaulukko

Keskiarvotaulukot ovat monella tavalla verrattavissa ristiintaulukkoon. Keskiarvotaulukoissa tarkastellaan jonkun muuttujan keskiarvoa toisen muuttujan luokissa.

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

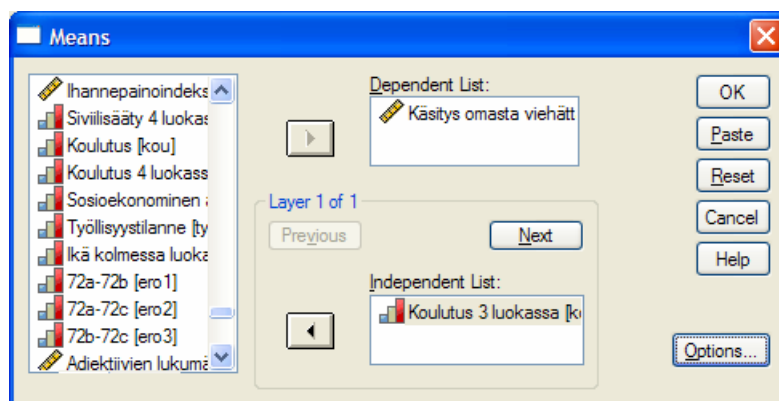
- Numeerinen (eli mitattu joko suhde- tai välimatka-asteikolla)
- Joskus myös järjestysasteikollinen muuttuja käy, etenkin mikäli keskiarvoa pystyy mielekkäästi tulkitsemaan. Tästä hyvänä esimerkkinä Likert-asteikolliset muuttujat, jotka todellisuudessa ovat järjestysasteikollisia muuttujia, mutta yleensä niiden keskiarvolle on olemassa mielekäs tulkinta. Niin ikään esim. läänien keskiarvo silloin, kun arvo 1 on eteläisin ja arvo 6 pohjoisin saattaa olla jotenkin mielekäs. Jos läänit on numeroitu jotenkin muuten, ei niistä ole mielekästä ottaa keskiarvoa. Jos olet epävarma keskiarvon tulkinnasta, käytä mieluummin ko. muuttujaan jotakin muuta analyysimenetelmää.

Selittävän muuttujan tulee olla

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

Keskihajonta on hyvä ottaa mukaan keskiarvotaulukoihin, koska se kuvaa havaintoarvojen keskimääräistä poikkeamaa keskiarvosta. 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 6 on esitetty vastaajien käsitys omasta ulkoisesta viehättävyydestä koulutusasteen mukaan tarkasteltuna. Käsitys omasta viehättävyydestä muuttuja on koodattu siten, että arvon kasvaessa käsitys omasta viehättävyydestä lisääntyy. Mitä siis keskiarvotaulukon perusteella voidaan sanoa? Kyllä vain, perusasteen suorittaneet naiset pitävät omaa ulkonäköään selkeästi keski- tai korkea-asteen koulutuksen suorittaneita naisia viehättävämpänä. Yleinen käytäntö on se, että sekä keskiarvo, että keskihajonta esitetään kahden desimaalin tarkkuudella.

Taulukko 6. Käsitys oman ulkonäön viehättävyydestä koulutusasteen mukaan.

Koulutus 3 luokassa	Mean	N	Std. Deviation
Perusaste	2,5365	69	,68855
Keskiaste	2,2558	159	,69628
Korkea-aste	2,2930	43	,73824
Total	2,3331	271	,70868

3.7 Korrelaatio

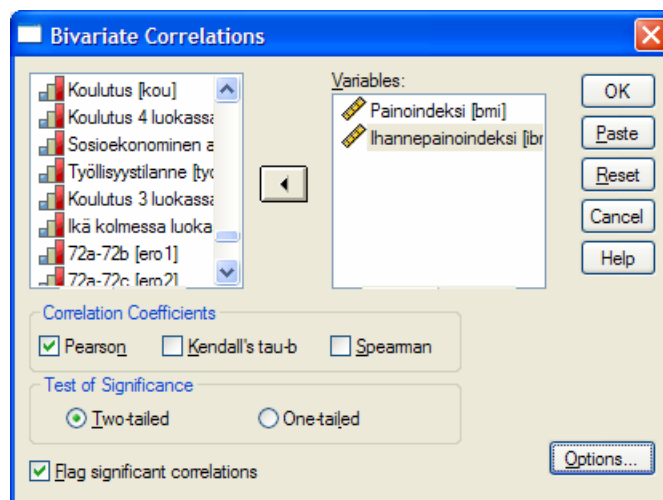
Korrelaatiotermiä käytetään yleiskielessä kuvaamaan kahden asian välistä yhteyttä. Voidaan esimerkiksi sanoa, että asuinpaikka korreloi sydän- ja verisuonitautien määrään. Tilastotieteellisessä mielessä korrelaatiolla on suppeampi merkitys. Korrelaatioilla tarkoitetaan tällöin yleensä numeeristen, jatkuvien 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 usein myös hajonta- eli sirontakuvioiden avulla. Tottunut lukija pystyy arvioimaan korrelaation hajontakuvioiden perusteella. Hajontakuvat paljastavat myös lineaarisista yhteyksistä poikkeavia muuttujien välisiä yhteyksiä. Tähän korrelaatiokerroin ei pysty.

Korrelaatioanalyysin tekeminen

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



Muuttujat, joiden välisestä korrelaatiosta on kiinnostunut, valitaan `Variables`-laatikkoon. Ohjelma tulostaa muuttujien välisen korrelaation eli koko korrelaatiomatriisin. Samanaikaisesti voidaan tarkastella useiden eri muuttujien välisiä korrelaatioita.

Taulukossa 7 on esitetty nykyisen painoindeksin ja ihannepainoindeksin korrelaatio. Siitä nähdään, että näiden muuttujien välillä on vahva positiivinen korrelaatio.

Taulukko 7. Nykyisen painoindeksin ja ihannepainoindeksin välinen 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).

3.8 Taulukoiden muokkaus

Pääset muokkaamaan SPSS-ohjelman Output-ikkunassa olevaa taulukkoa tuplaklikkaamalla sitä. Huom! Et pääse kaikkiin alla mainittuihin toimintoihin käsiksi, jollet ensin tee sitä, koska tarvittavat valikot käskyineen ilmaantuvat Output-ikkunan valikkoriville yleensä vasta taulukon tuplaklikkaamisen jälkeen. Tulostusikkunassa voi muokata taulukkoa, mutta kaikkia muokkaustoimintoja ei siinä pysty tekemään. Mikäli et tuplaklikkaamalla saa Pivot-taulukoiden muokkausikkunaa auki, tee seuraavasti: taulukon päällä ollessasi paina hiiren oikeaa painiketta, jolloin aukeaa valikko, josta valitse hiirellä `SPSS Pivot table object` ja siitä toiminto `Open`.

Valikkoikkunoiden käytön sijasta taulukkoa voidaan muokata klikkaamalla hiirellä taulukon eri osia. Tosin kaikkia valikoissa olevia toimintoja ei voi toteuttaa hiiren avulla. Usein kuitenkin taulukon muokkaamista on helpointa opetella hiiren avulla.

Keskeisimmät toiminnot

Koko taulukkoon vaikuttavia muotoiluja (kaikkien tekstien fontti, tekstin asettelu soluissa, solujen rajat) 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 (eli "maalamalla") ne osat, joita haluaa muuttaa ja valitsemalla `Format/Cell Properties` HUOM! `Cell Properties` -valikkoon ei ole mahdollista päästä, jollet ole ensin valinnut (maalannut) soluja, joita haluat muokata. Useita erillisiä solujahan saa tarvittaessa valittua kerralla pitämällä `Ctrl`-näppäintä pohjassa valinnan ajan.

Prosenttimerkit pois, oikea määrä desimaaleja.

Format/Cell Properties/Value: Format-listasta valitaan oikean näköinen muoto ja sen alapuolella olevasta kohdasta oikea määrä desimaaleja.

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 muuttujan nimen näkyviin valitse kyseinen solu hiirellä, paina hiiren oikeaa painiketta, ja tämän jälkeen valitse toiminto Show Dimension Label. Tällöin teksti tulee taulukossa näkyviin, tosin vasta siinä vaiheessa, kun on siirtynyt taulukon muokkaamisvalikosta pois. Toisaalta erityisesti keskiarvotaulukkojen kohdalla tilanne on eri tavalla harmillinen: keskiarvotaulukon ylälaitaan tulee jatkuvan muuttujan nimi, jota siinä harvemmin tarvitaan. Sen poistaminen tapahtuu valitsemalla Pivot-valikosta toiminto Pivoting Trays. Kyseisen valikon avauduttua menemällä Layerin ”ruusukkeeseen” päälle ja hiiren oikealla painalluksella saa näkyville valikon, josta valitaan Hide All Category labels -toiminto.

Lähdeviitteen lisääminen

Insert/Caption

Alaviitteen lisääminen

Insert/Footnote

Osien tuhoaminen

Valitaan taulukosta tuhottava(t) osa hiirellä ja painetaan delete-painiketta.

Rivit sarakkeiksi ja toisinpäin

Pivot/Transpose Rows and Columns. Useasti tämä on pelkkää kikkailua, mutta siitä huolimatta sitä kannattaa kokeilla.

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.

3.9 Lopuksi: siis mitä tehdä, kun

Yksi muuttuja, joka

- a. jatkuva: valitse descriptives (Analyze/Descriptive statistics/Descriptives), jolloin voit pyytää SPSS:ää tulostamaan muun muassa 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.

HUOM! Muista uudelleenluokittelun antamat mahdollisuudet!

4 GRAFIIKKA

4.1 Grafiikan yleisiä ehtoja

Useat edellä (luvussa 3.1) esille tuodut taulukointia koskevat säännöt koskevat myös grafiikkaa. Tässä kuitenkin hieman lisätietoa hyvän grafiikan tekemiseen.

- **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 kuvan alaviitteessä)
- Kuvan otsikko on aina kuvan alla (vs. taulukon otsikko, joka on aina taulukon yläpuolella)
- Kuvallisessa esityksessä on pyrittävä yksinkertaisuuteen
- Kuvaajan on oltava helppo lukea
- Sisältää ainoastaan oleellisen tiedon
- Yhdenmukainen muiden kuvien kanssa tekstiltään ym.

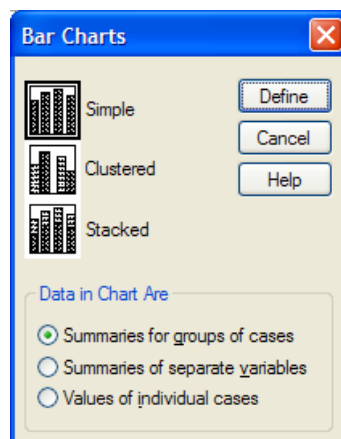
Ohjelmistojen käytön helpottuessa tilastoaineistojen graafinen kuvailu on yleistynyt. Kuvalla voi kertoa nopeasti tarvittavan tiedon lukijalle. Tilastografiikkaa voidaan ryhmitellä erilaisten kuvatyyppeihin mukaan. Yleisimpiä kuvatyyppejä ovat pylväät, hajontakuviot, viiva- ja aluekuviot sekä piirakat.

Kuvatyyppejä valitessa on tärkeää miettiä, mitä kuvan on **tarkoitus viestiä** lukijalle sekä sitä, **millainen kuva viestii asian parhaiten**. Kaikki kuvatyypit eivät sovellu kaikkien asioiden esittämiseen. Tässä luvussa esitetään seuraavat kuvatyypit: pylväskuvio, histogrammi, hajontakuviot, viivakuviot, aluekuviot, piirakkakuviot ja viiksikuviot.

4.2 Pylväskuvio

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.

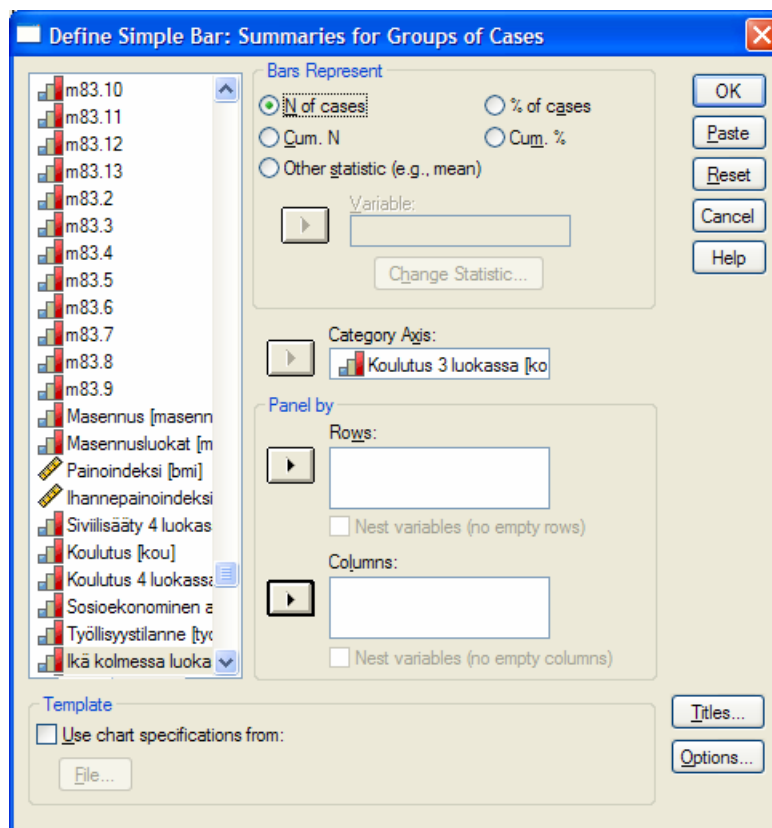
(1) Erilaisia pylväskuvioita on kolme: Simple, Clustered ja Stacked. Oletusvalinnalla Simple voidaan tehdä frekvenssiylväitä ja esimerkiksi keskiarvopylväitä. Valinnoilla Clustered ja Stacked voidaan tarkastella kahden tai kolmen muuttujan pylväskuvioita.

(2) Data in Chart are –kohdassa voidaan valita, mitä pylväät kuviossa esittävät. Summaries for groups of cases: mikäli ollaan kiinnostuneita muuttujan eri luokkien frekvensseistä, pylvään korkeus kertoo havaintojen määrän kussakin luokassa. Tämä on oletusmenetelmä, ja usein parhaiten sopiva esittämistapavaihtoehto, useimmiten kuitenkin havaintojen määrän sijasta tarkastellaan prosentuaalisia osuuksia. Seuraavissa esimerkeissä käytetään tätä valikkoa. Muut vaihtoehdot ovat: Summaries of separate variables, tällöin pylvään korkeus kuvaa kunkin valitun muuttujan tunnuslukua ja Values of individual cases, tällöin pylvään korkeus esittää muuttujan arvoa.

4.2.1 Yhden luokitellun muuttujan pylväskuvio

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

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



Category Axis –kohtaan valitaan muuttuja, josta halutaan tehdä frekvenssiylväs.

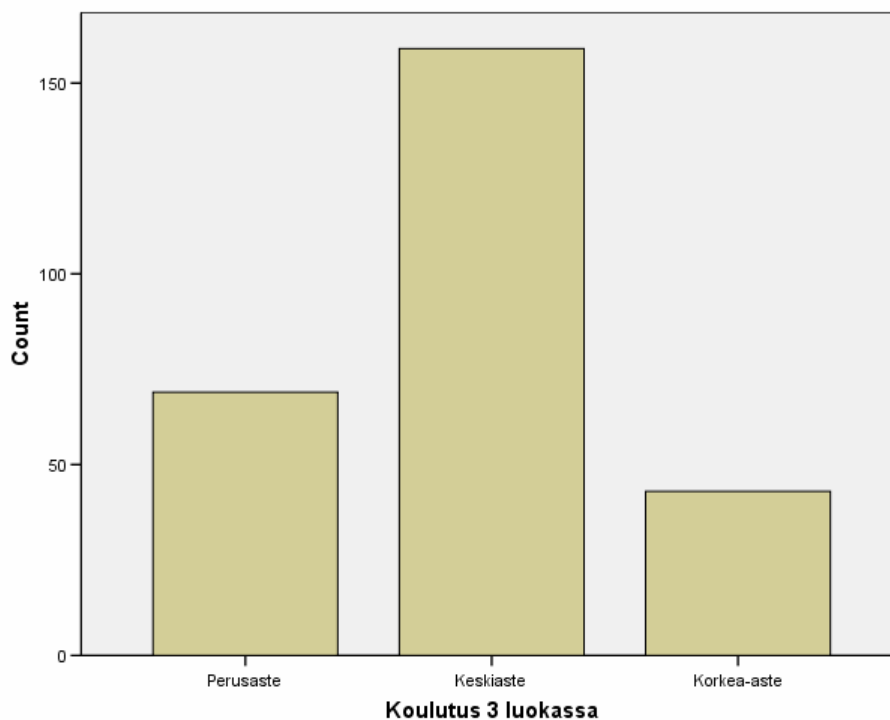
Titles: täällä voi kirjoittaa kuvalle otsikon, otsikon voi kuitenkin kirjoittaa myös kuvankäsittelyikkunassa tai vasta siinä tiedostossa, johon kuva liitetään (tätä suosittelen).

Options: tässä valikossa voi määrittää, esitetäänkö kuvassa puuttuvat tiedot, oletusmenetelmä on että ei esitetä. Muutoksia puuttuvien tietojen esittämiseen 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. Frekvenssiylvästä tehdessä ohjelman oletusarvoa, N of cases (eli number of cases), ei tarvitse vaihtaa.

Panel by –kenttä: tähän voi lisätä ns. paneelimuuttujan, muuttujan, jonka luokissa kiinnostuksen kohteena olevaa muuttujaa tarkastellaan. Paneelimuuttujan luokat voi asettaa joko riveittäin tai sarakkeittain.

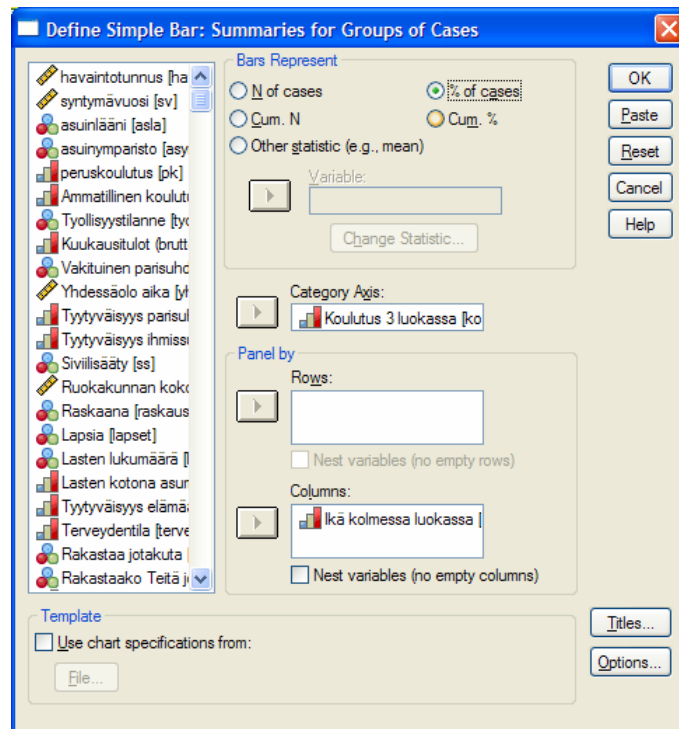
Kuvassa 1 on frekvenssiylväs, jossa on esitetty vastaajien koulutusaste. Jotta kuva olisi esittämiskelpoinen, kuvaa tulee muokata kuvankäsittelyikkunassa (ainakin count-sana täytyy suomentaa).



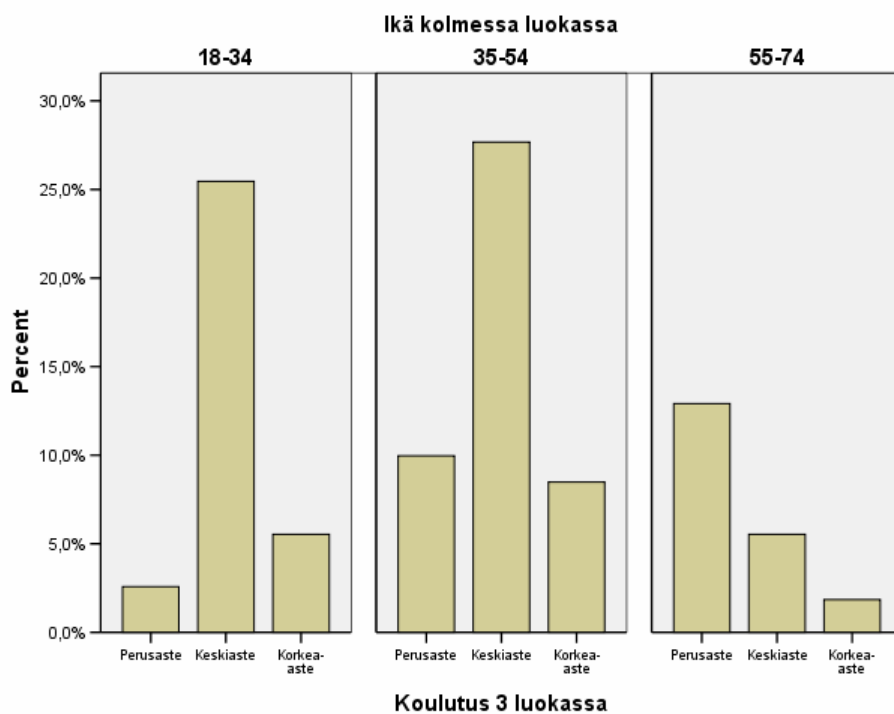
Kuva 1. Vastaajien koulutusaste.

4.2.2 Yhden muuttujan tarkastelu toisen muuttujan luokissa

Kuvassa 1 esitettiin vastaajien koulutusaste pylväskuviona, nyt ollaan kiinnostuneita siitä, mitä lisätietoa iän mukaanotto tarkasteluun tuo. Valintaikkunaan tehtiin seuraavat muutokset:



Esittämistavaksi vaihdettiin prosentit, koska niiden avulla on helpompi tarkastella eroja eri ikäluokissa ja ikä-muuttuja laitettiin **Panel by** – kenttään Columns-kohtaan.

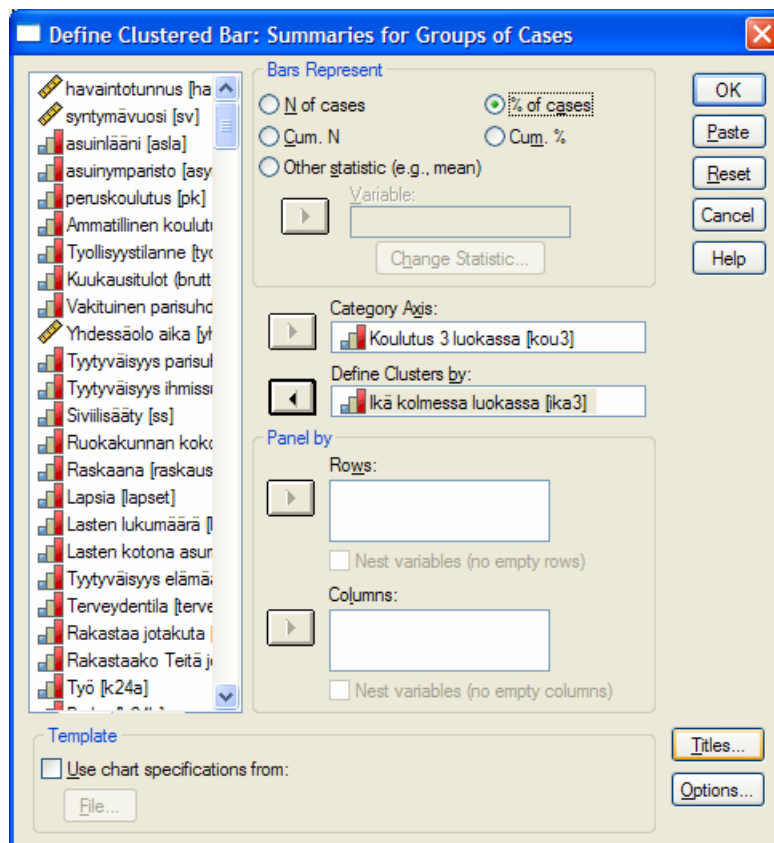


Kuva 2. Naisten koulutusaste ikäryhmittäin tarkasteltuna (%).

4.2.3 Kahden luokitellun muuttujan pylväskuvio

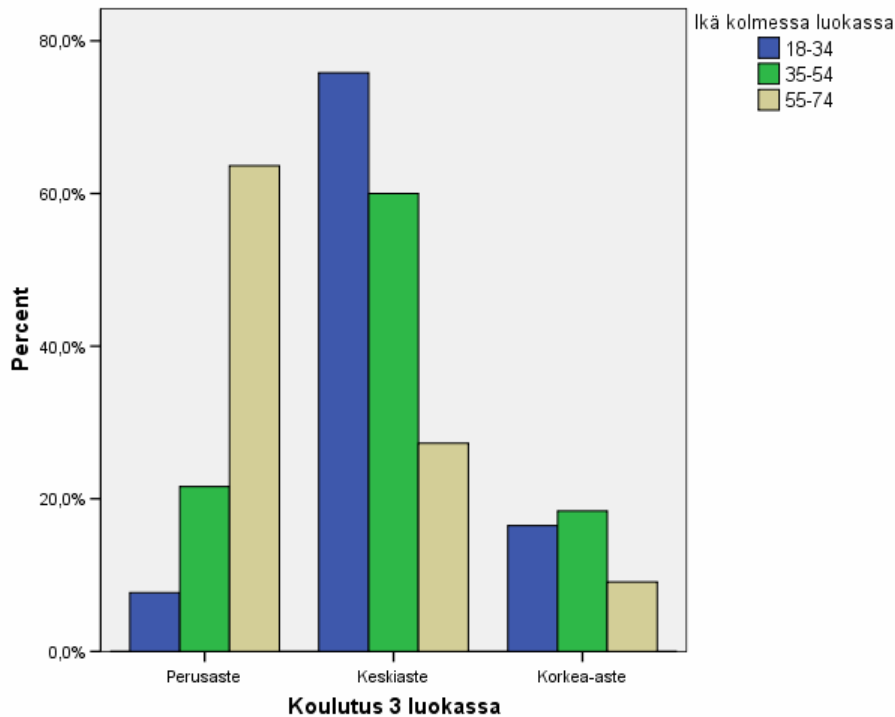
Kahden luokitellun muuttujan pylväskuvioita voidaan tehdä joko **ryhmittelemällä** pylviäitä (Clustered) tai **pinoamalla** niitä (Stacked).

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



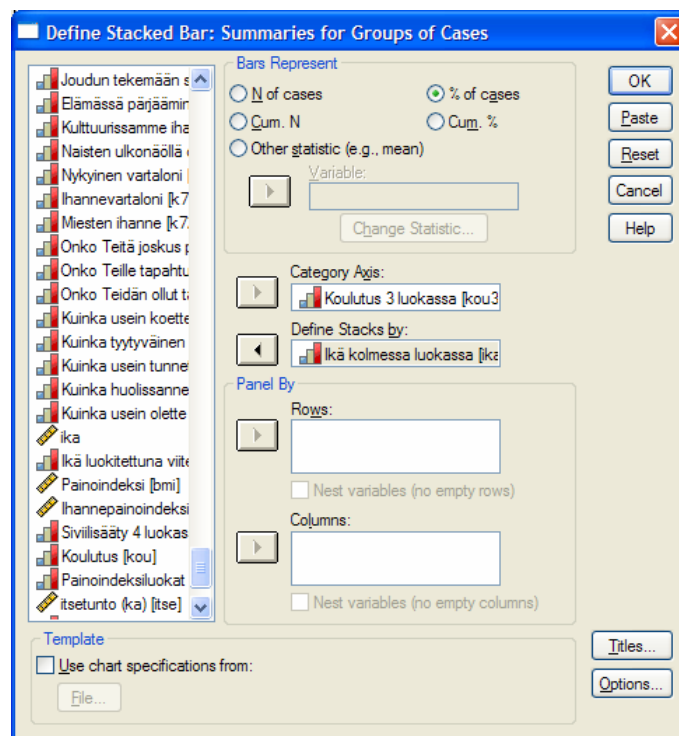
Ikkuna on samanlainen kuin Simple-valinnan ikkuna. Ainoa ero on se, että tässä ikkunassa on varattu myös toiselle muuttujalle kenttä (Define clusters by).

Seuraavassa on esitetty vastaajien koulutusaste ikäluokittain tarkasteltuna (%). Mieti, mitä kuvasta nähdään ja miten kuvaa tulee tulkita.

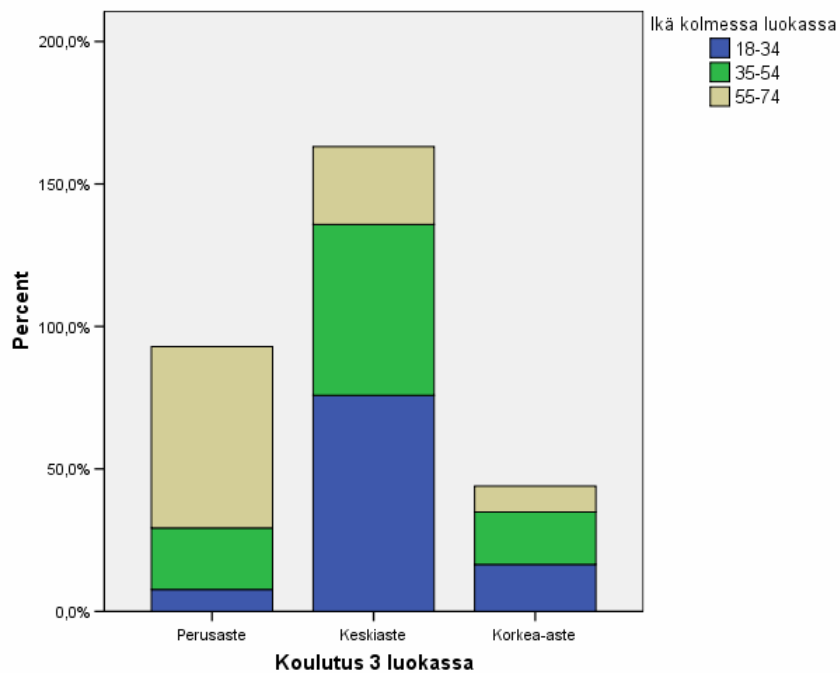


Kuva 3. Naisten koulutusaste ikäluokittain tarkasteltuna (%).

On ilmeistä, että yleensä tutkija on kiinnostunut siitä, miten koulutus jakaantuu eri ikäryhmien sisällä kuin siitä tiedosta, mitä kuva 3 tarjosi. Sen tekeminen saisi olla huomattavasti helpompaa kuin mitä se on. Sitä päästään tekemään seuraavilla valinnoilla Graph/Bar valitsemalla Stacked. Näin saadaan näkyville vastaava ikkuna kuin edellisestäkin kuvaa tehdessä, se esitetään tässä, koska muuttujien sijoittelu on tärkeä, kun ko. kuvan tekeminen vaatii muuttujien uudelleensijoittelua:

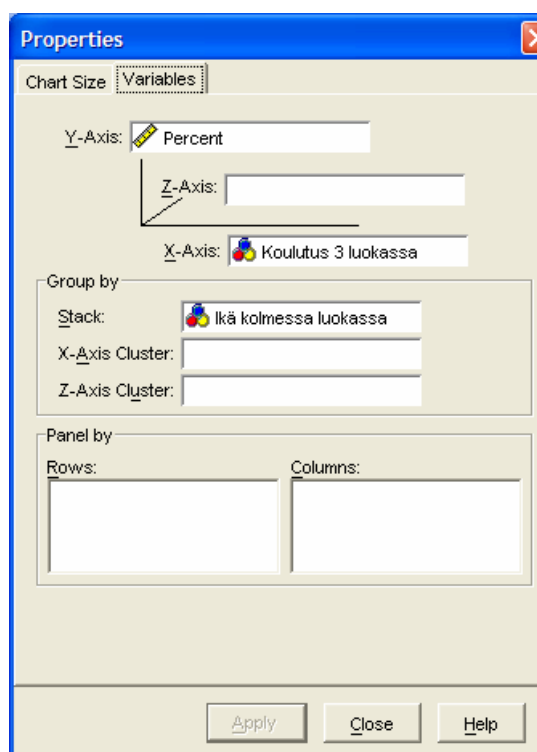


Näillä valinnoilla kuva näyttää seuraavalta:

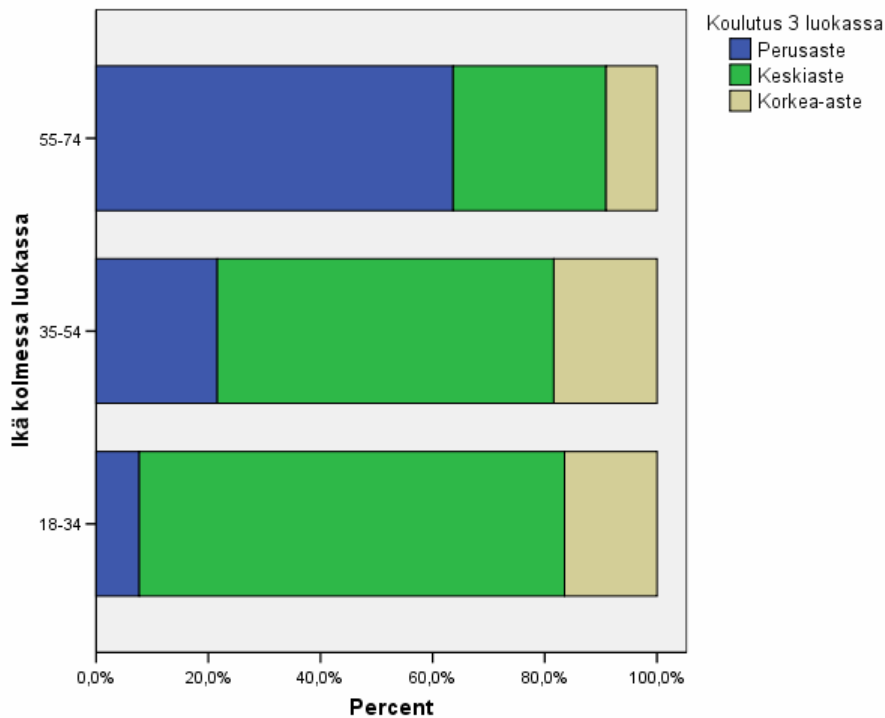


Kuva 4A. Alustava kuva.

Kuva ei lainkaan tyydytä, siksi siirrytään kuvankäsittelyikkunaan muokkaamaan kuvaa (tuplaklikkaamalla kuvaa). Kuvankäsittelyikkunassa ensimmäiseksi käännetään pylväät vaakasentoon valitsemalla Options-valikosta toiminto Transpose chart. Tämän jälkeen sekä muuttujien sijoittelua että prosenttien esittämistä tulee muuttaa. Valitaan Edit-valikosta Properties-valikko ja sen välilehti Variables, jolloin saadaan näkyville seuraava ikkuna:



Täällä raahaamme ensimmäiseksi koulutus-muuttujan esim. X-Axis Cluster -valikkoon hetkeksi ja sitten ikä-muuttujan X-axis -kohtaan, jonka jälkeen koulutus-muuttaja siirretään Stack-valikkoon. Näin saamme seuraavan kuvan:

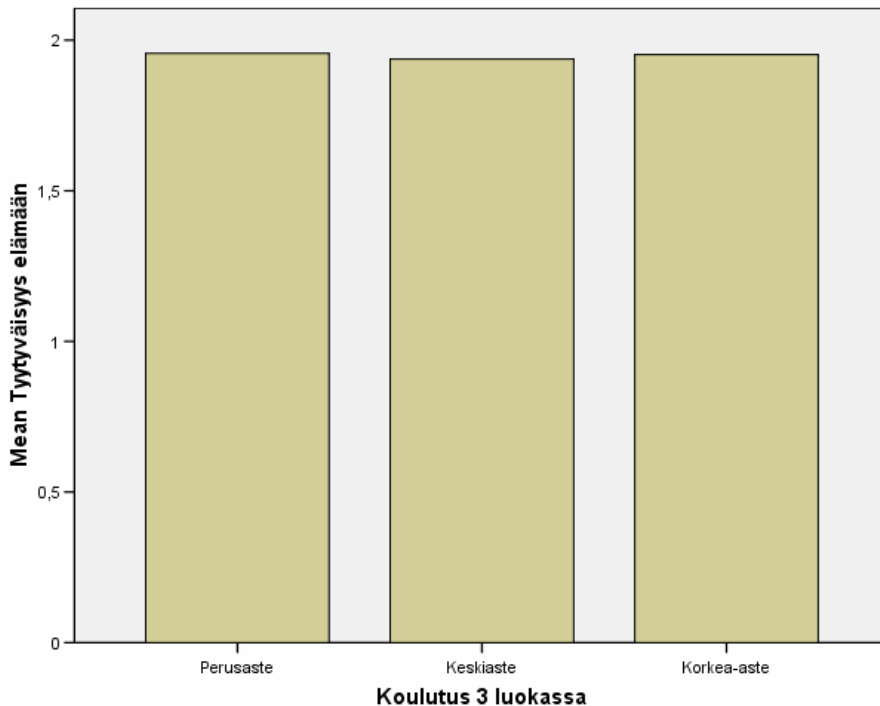


Kuva 4B. Naisten koulutusaste ikäluokittain tarkasteltuna (%).

4.2.4 Keskiarvopylväät

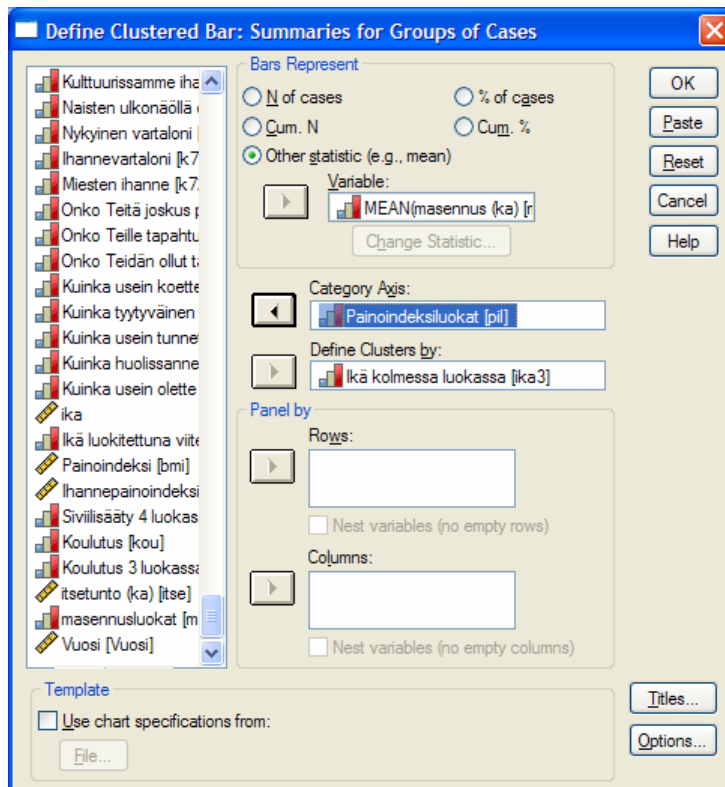
Keskiarvopylväillä esitetään jonkin muuttujan keskiarvoja joko (1) yhden tai (2) kahden muuttujan määrittelemässä ryhmässä. Seuraavassa esitetään, miten näitä tehdään. Keskiarvopylväät ovat yksinkertainen tapa tiivistää tietoa hyvin helposti havainnoitavaan muotoon. Selitettävän muuttujan pitää ainakin periaatteessa olla mitattu joko suhde- tai välimatka-asteikolla (huom. useasti myös järjestysasteikolliset muuttujat kelpaavat).

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

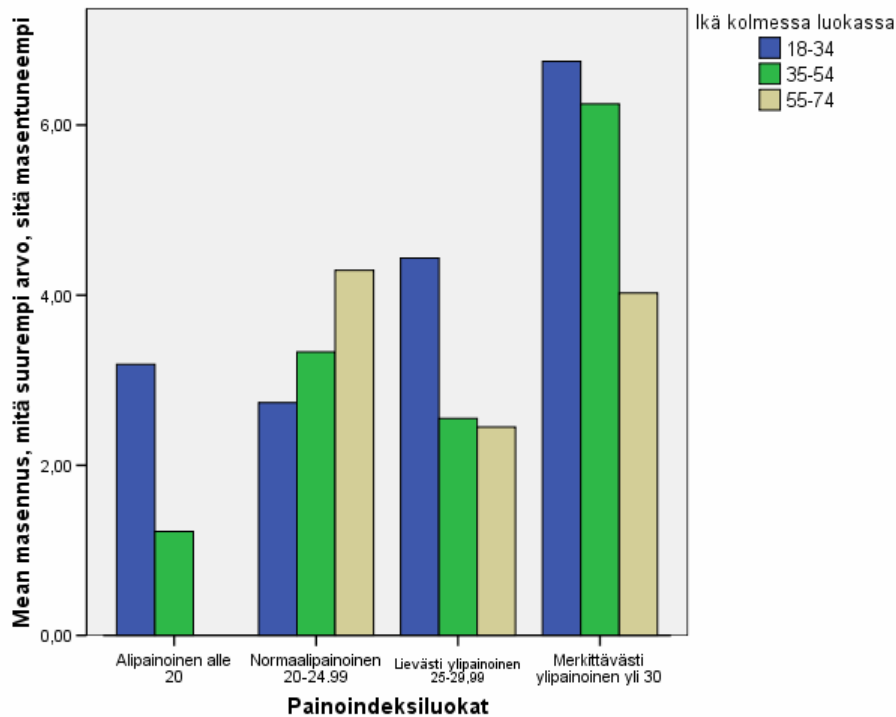


Kuva 5. Tytyväisyys elämään koulutusasteen mukaan.

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



Kuvassa 6 on tehty keskiarvokuva valintaikkunassa tehdyillä valinnoilla. Mitä siitä on nähtävissä? Onko ikä tai painoindeksi tai molemmat yhteydessä naisten masentuneisuuteen?

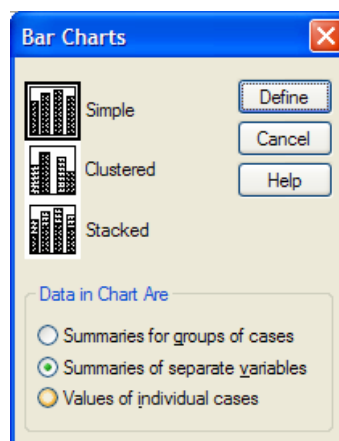


Kuva 6. Masentuneisuuden aste eri painoindeksiluokissa ikäluokittain tarkasteltuna.

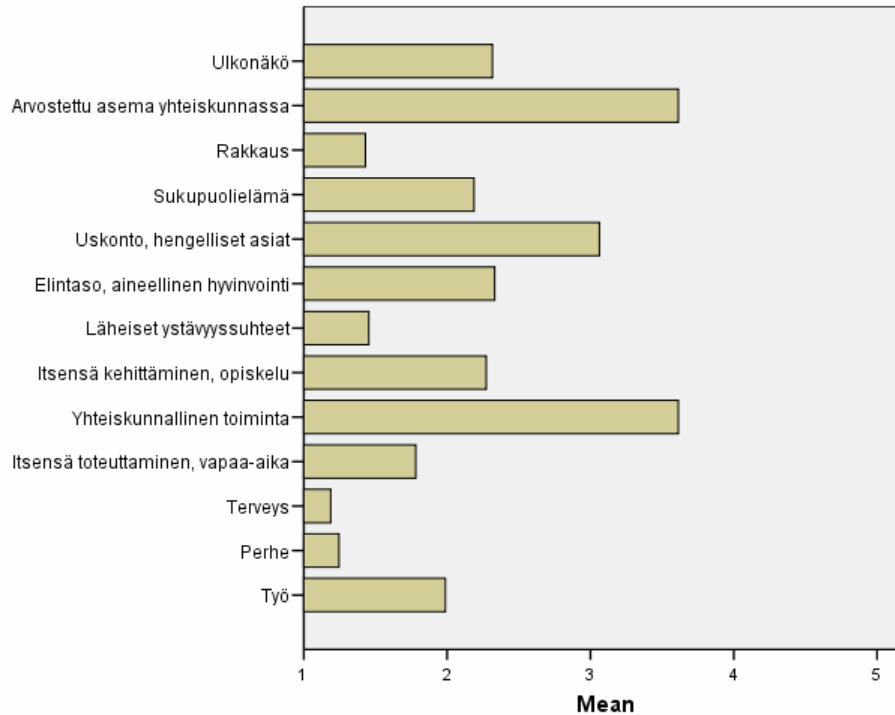
4.2.5 Useiden erillisten muuttujien samanaikainen kuvaaminen

Edellisessä luvussa keskityttiin tarkastelemaan lähinnä yhden muuttujan kuvailua yksinään tai toisen muuttujan luokissa. Nyt tarkastellaan useiden erillisten muuttujien kuvaamista samanaikaisesti (1) yksinään ja (2) toisen muuttujan luokissa. Näin tehdessä käytettyjen muuttujien vastauskaalan tulee olla samanlainen.

(1) Yksittäisten muuttujien pylväskuviota tehdessä kuvien tekovalikossa valitaan Data in chart are -kentästä toiminto Summaries of separate variables, muutoin oletusvalinnan Simple annetaan olla, kuten alla olevasta ikkunasta nähdään:

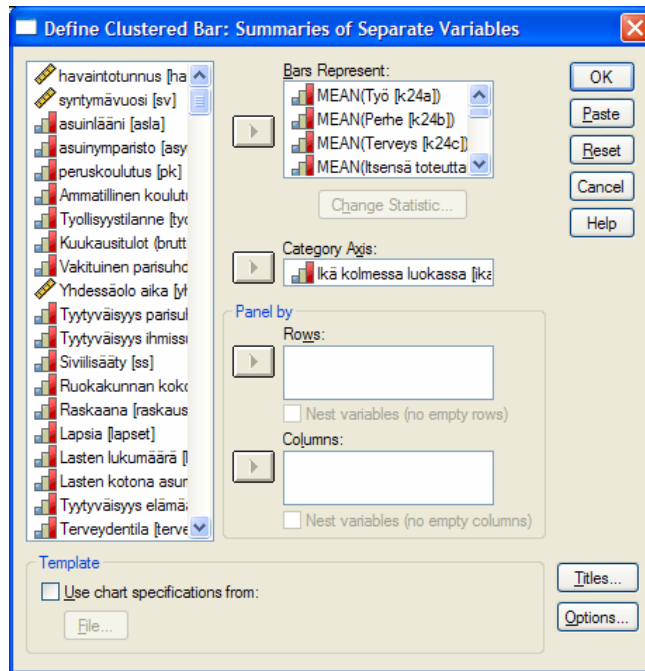


Kuvassa 7 on esitetty naisten eri elämäneläyteen tärkeiksi kokemista koskevat vastaukset keskiarvoina. Vastaukset on koodattu siten, että mitä pienempi arvo, sitä tärkeämpi elämäneläyteen. Tämyntyyppistä kuvaa varten olisi yleensä hyödyllistä kääntää muuttujat toisinpäin, eikö totta? Kuvasta nähdään, että tärkeimmät asiat elämässä ovat terveys, perhe, läheiset ystävyysuhteet ja rakkaus.

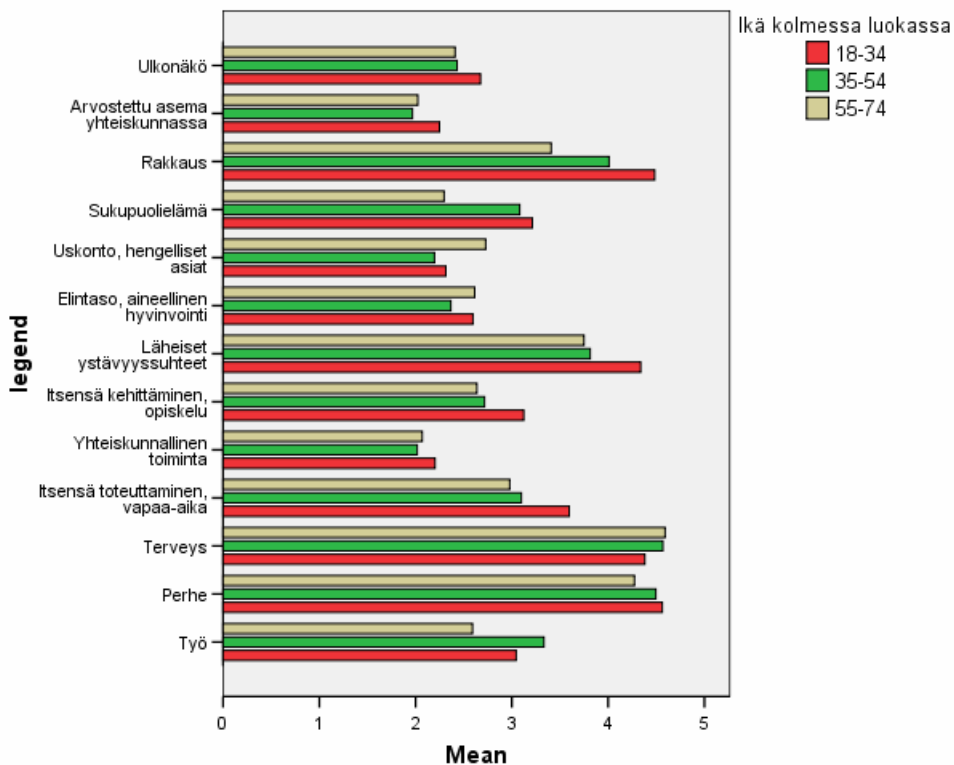


Kuva 7. Eri elämäneläyteen tärkeys.

(2) Useiden yksittäisten muuttujien samanaikaisen tarkastelu toisen muuttujan luokissa pylväs-kuviossa eroaa edellisen pylväs-kuvion tekemisestä siten, että kuvan tilausvalikossa valitaan Clustered tai Stacked. Tässä valittiin Clustered. Alla olevasta valintaikkunasta nähdään, miten muuttujat sijoitetaan. Tarkastelun kohteena on eri elämäneläyteen tärkeys naisille (nyt näiden muuttujien vastauskaala on käännetty toisinpäin) ja niitä tarkastellaan ikäluokittain.



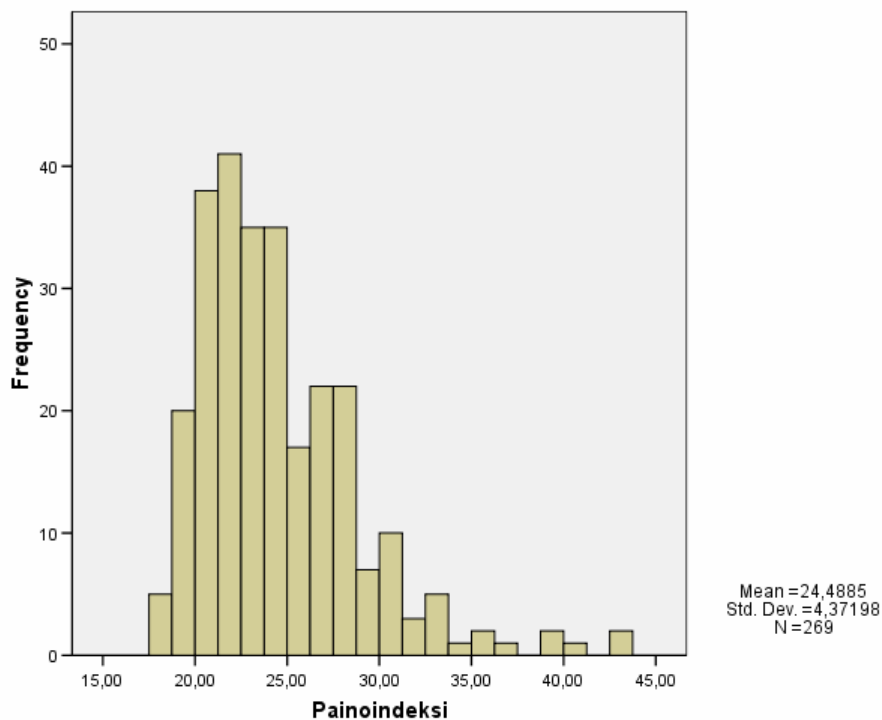
Kuvaa joudutaan käsittelemään kuvankäsittelyikkunassa vastaavaan tapaan kuin luvussa 4.2.3 nähtiin. Lopulta näkyville saadaan seuraava kuva. Mitä siitä nähdään? Onko eri elämäalueiden tärkeäksi kokemisessa eroja ikäluokittain ja jos niin minkä elämäalueiden kohdalla? Kriittisesti kuvaa tarkasteltaessa kuva on niillä rajoilla voidaanko sitä pitää esittämiskelpoisena; kuva pitää sisällään paljon informaatiota eikä ole välttämättä kovin helppolukuinen. Taulukko esittäisi asian paremmin.



Kuva 8. Eri elämäalueiden tärkeys ikäluokittain.

4.3 Histogrammi

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 muuttujat jokseenkin järkevästi. Mieti kuitenkin aina itse, onko luokittelu järkevä.) SPSS:ssä histogrammi tehdään valinnoilla `Graphs/Histogram`, erityisiä lisämääreitä ei yleensä tarvita. Usein on kuitenkin hyödyllistä ruksia kohta `Display normal curve` kuvaan mukaan. Kuvassa 9 on esitetty vastaajien painoindeksin histogrammi.



Kuva 9. Naisten painoindeksi.

4.4 Hajontakuvi

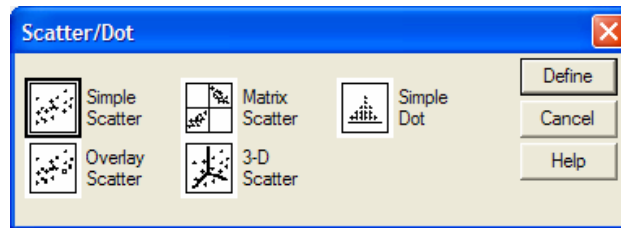
Hajontakuviolla (*scatter*) esitetään **kahden jatkuvan muuttujan yhteisjakauma**. Hajontakuviolla tarkoitetaan muuttujien piirtämistä samaan koordinaatistoon siten, että toisen muuttujan asteikko asettuu x-akselille ja toisen muuttujan y-akselille. Havainnot asettuvat pisteinä tähän koordinaatistoon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

Koordinaatistosta voidaan tutkia muuttujien yhteisvaihtelua, jolloin havaitaan, onko muuttujien välillä keskinäistä riippuvuutta. Useimmiten hajontakuvioita tehdään tutkijan omaan käyttöön, sellaisina ne ovat erittäin hyödyllisiä aineistoon ja sen muuttujiin tutustumisessa.

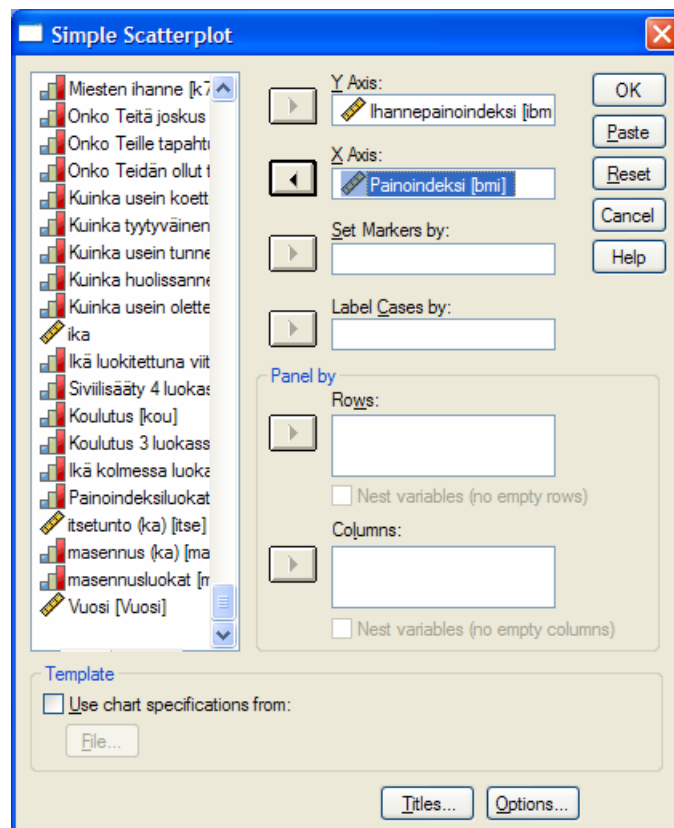
Kun hajontakuvioiden pisteet ovat täysin hajallaan, muuttujien välillä ei ole riippuvuutta. Hajontakuvioita tehdessä voidaan periaatteessa valita muuttujien x ja y akselit kummin päin tahansa, mutta jos on luonnollista ajatella toisen muuttujan selittävän toisen muuttujan

käyttäytymistä, valitaan selittävän muuttujan akseliksi vaaka-akseli. Erilaisilla hajontakuvion päälle piirretyillä regressiokäyrillä voidaan mm. osoittaa riippuvuuksia myös harjaantumattomille. Hajontakuvioiden laatiminen auttaa tekemään ylimalkaisia johtopäätöksiä muuttujien välisistä yhteyksistä. Tämä ei kuitenkaan yleensä riitä, ja siksi tehdään jatkoanalyysyjä, joiden avulla muuttujien mahdollinen riippuvuus voidaan esittää tiiviimmin; numeerisessa muodossa.

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

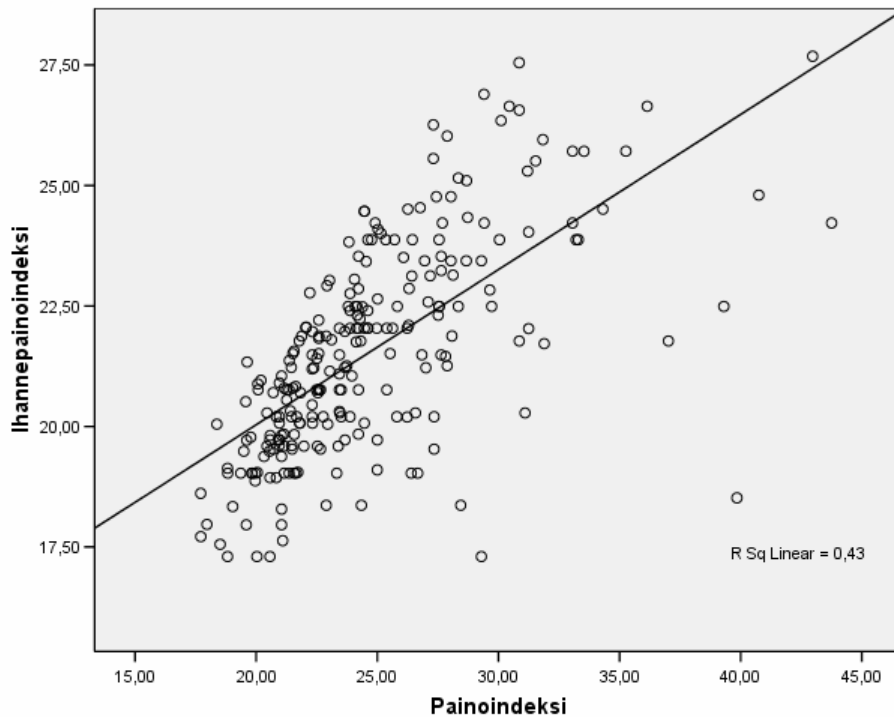


Oletusvalinnalla **Simple Scatter**, saadaan seuraava ikkuna näkyville:



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

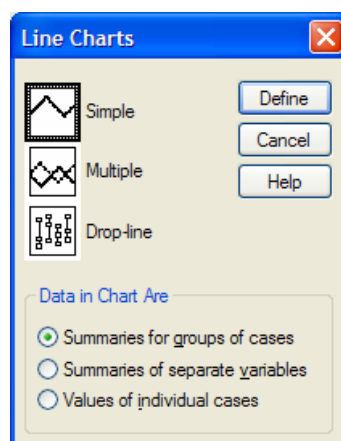
Kuvassa 10 on hajontakuvi, jossa on esitetty naisten nykyisen painoindeksin ja ihannepainoindeksin yhteisjakauma. Siitä nähdään, että näiden muuttujien välillä on riippuvuutta: Nykyisen painoindeksin noustessa ihannepainoindeksi nousee. Kuvaan on lisätty riippuvuuden voimakkuutta ilmentävä regressiosuora. Hajontakuviin regressiosuoran voi lisätä kuvankäsittelyikkunassa. Se tapahtuu valitsemalla Element-valikosta Fit line at total.



Kuva 10. Nykyisen painoindeksin ja ihannepainoindeksin hajontakuvio.

4.5 Viivakuvio

Viivakuvioilla kuvataan yleensä aikasarjoja. Viiva antaa vaikutelman siitä, että sitä pitkin voi liikkua, niinpä sen kuvaaman muuttujan tulisi olla intervalliasteikollinen ja periaatteessa jatkuva. Viivakuvioita pääsee tekemään valikosta *Graph/Line*. Näin näkyville saadaan seuraava ikkuna:



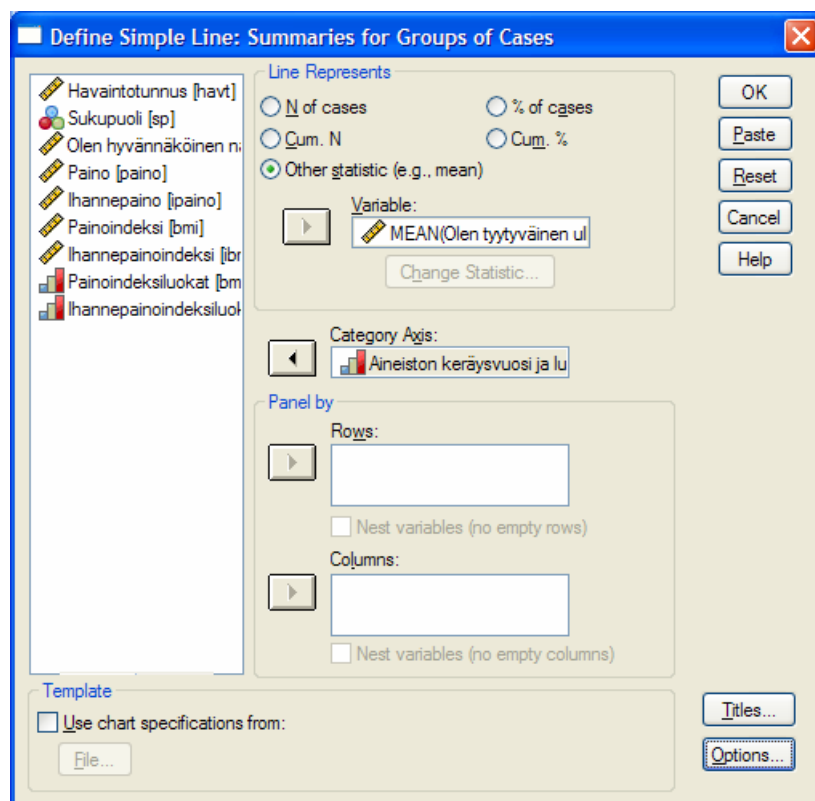
Tässä valikossa määritetään se, (1) Millaisia viivakuvioita halutaan tehdä ja (2) Mitä pylväät kuviossa esittävät.

(1) Viivakuviovaihtoehtoja on kolme: Simple, Multiple ja Drop-Line. Oletusvalinnalla Simple voidaan tarkastella yhden muuttujan viivan esiintymistä esim. keskiarvoina eri ajankohtina. Multiple toiminnolla voidaan tarkastella samanaikaisesti usean viivan esiintymistä kuviossa. Drop Line –toiminnolla sen sijaan voidaan verrata muuttujan arvoja ryhmittäin.

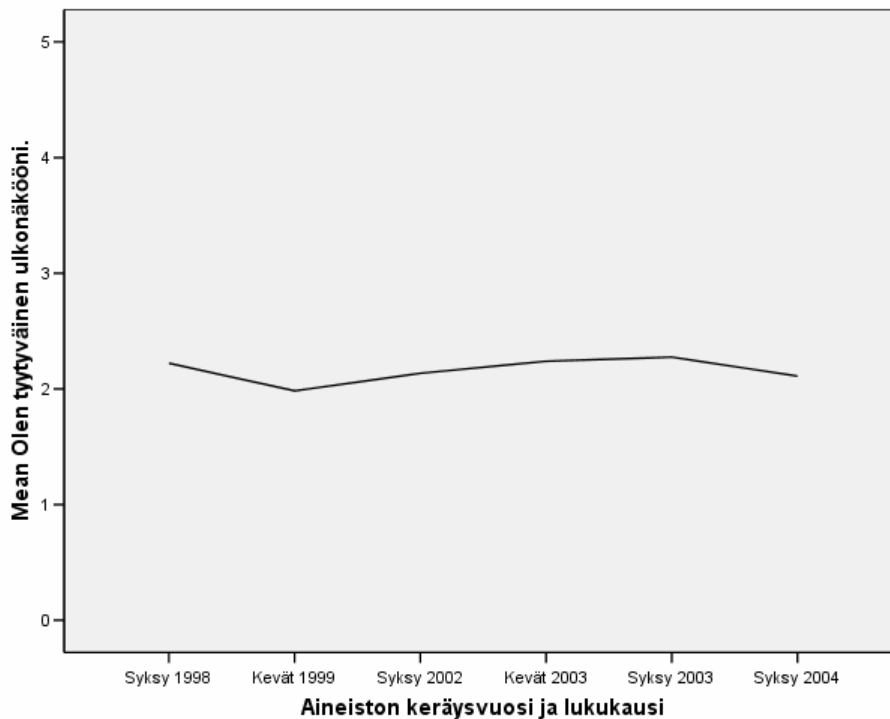
(2) Data in Chart are –kohdassa voidaan valita, mitä viivat kuviossa esittävät. Summaries for groups of cases: viivan piste kohta määrittää havaintojen tunnusluvun ko. luokassa. Tämä on oletusmenetelmä, ja usein parhaiten sopiva esittämistapavaihtoehto. Seuraavassa esimerkissä käytämme sitä. Muut vaihtoehdot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

4.5.1 Kahden muuttujan viivakuvi

Valinnalla Simple, saadaan seuraava ikkuna näkyville, joka vastaa pylväskuvion valikkoa:



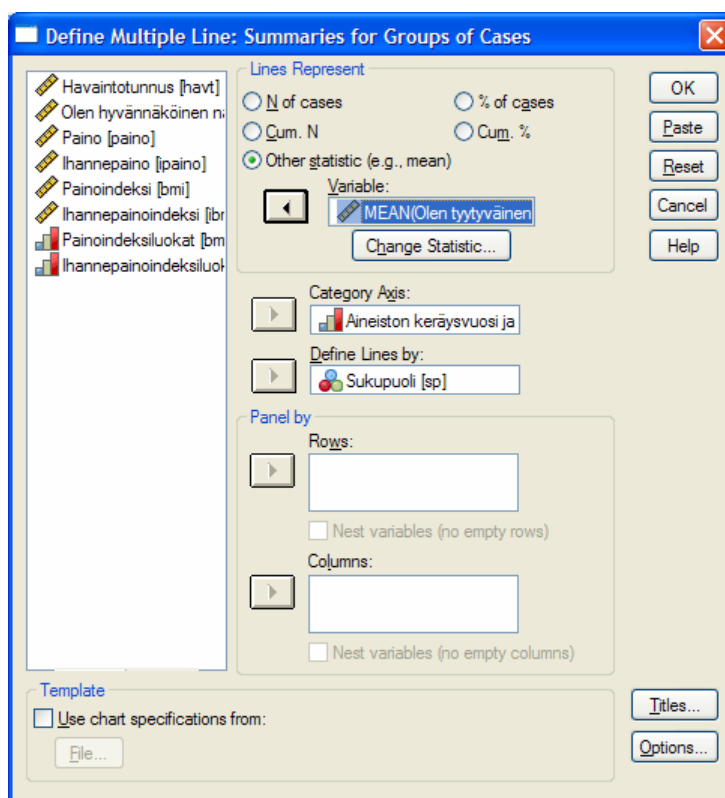
Näillä valinnoilla (ohjelmaa pyydettiin viivan esittävän vastanneiden ulkonäkötyytyväisyyden keskiarvo aineiston keräysvuosittain) saadaan kuva 11. Kuten edellä on jo havaittu, SPSS-ohjelmalla on taipumus kuvissa suurennella eroja, eli muuttujan koko asteikko ei ole näkyvissä. Näin tapahtuu viivakuvioidessakin, joten kannattaa aina tarkastaa kuvan skaalaus ja tarvittaessa, siis oikeastaan näin käy aina, tehdä siihen muutoksia kuvankäsittelyikkunassa. Ulkonäkötyytyväisyysmuuttuja on koodattu siten, että arvon kasvaessa ulkonäkötyytyväisyys vähenee.



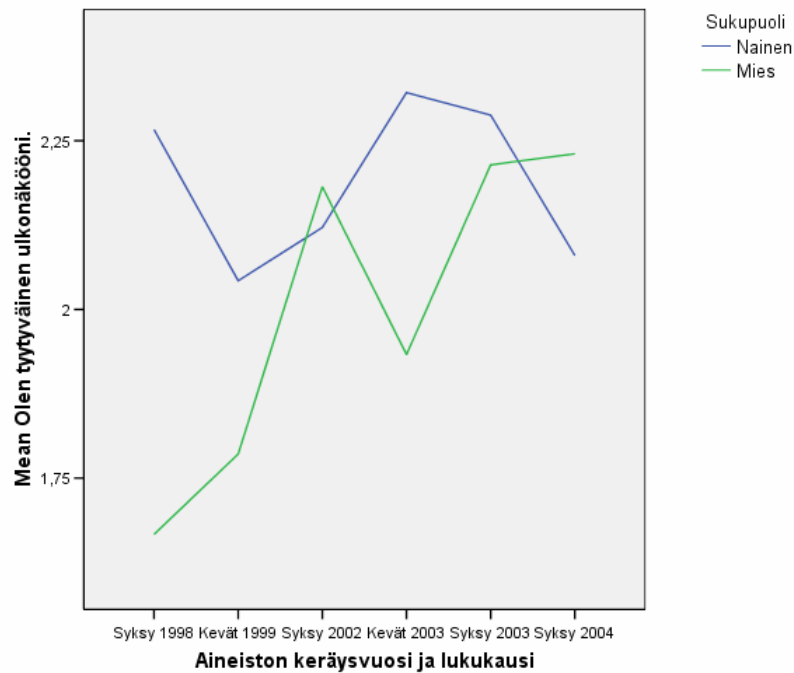
Kuva 11. Ulkonäkötyytyväisyyden keskiarvo eri ajankohtina.

4.5.2 Kolmen muuttujan viivakuvi

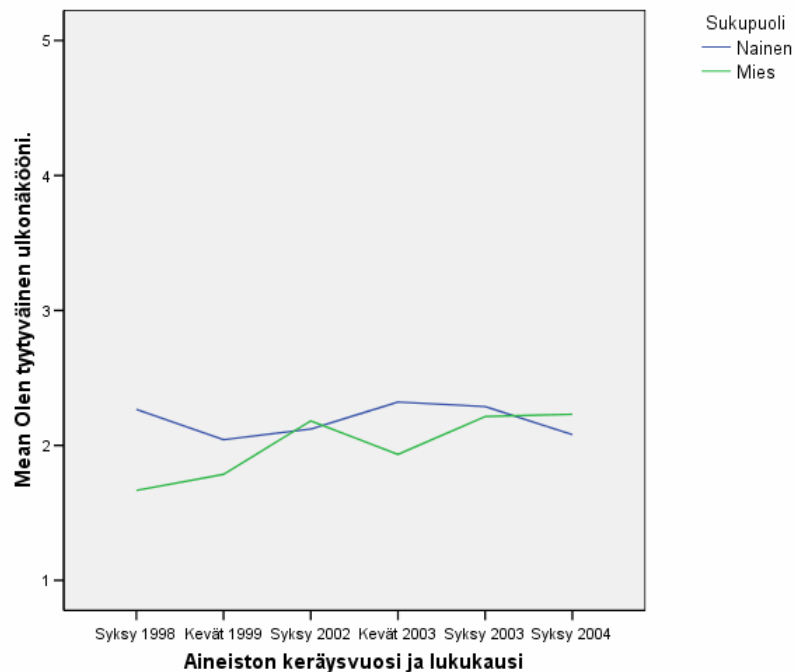
Kolmen muuttujan viivakuviota pääsee tekemään viivavalikosta (Graph/Line) valinnalla Multiple. Näin saadaan näkyville seuraava ikkuna:



Tämän valikon ainoa ero suhteessa Simple-toiminnon valikkoon, on yksi lisäkenttä; Define Lines by. Seuraavassa kuvassa (Kuva 12) on esitetty erikseen naisten ja miesten keskimääräinen ulkonäkötyytyväisyys eri ajankohtina. Vertaa tätä kuvaa kuvaan 13. Mitä havaitset?



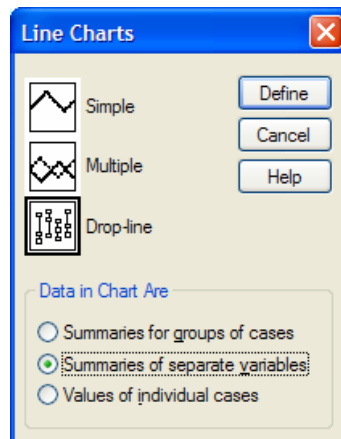
Kuva 12. Ulkonäkötyytyväisyys eri ajankohtina sukupuolen mukaan.



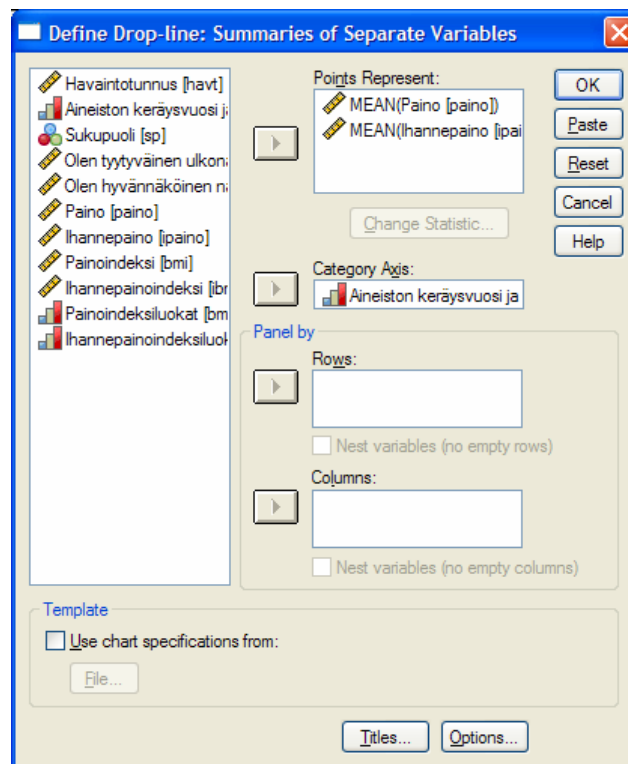
Kuva 13. Ulkonäkötyytyväisyys eri ajankohtina sukupuolen mukaan.

4.5.3 Drop-line -kuvio

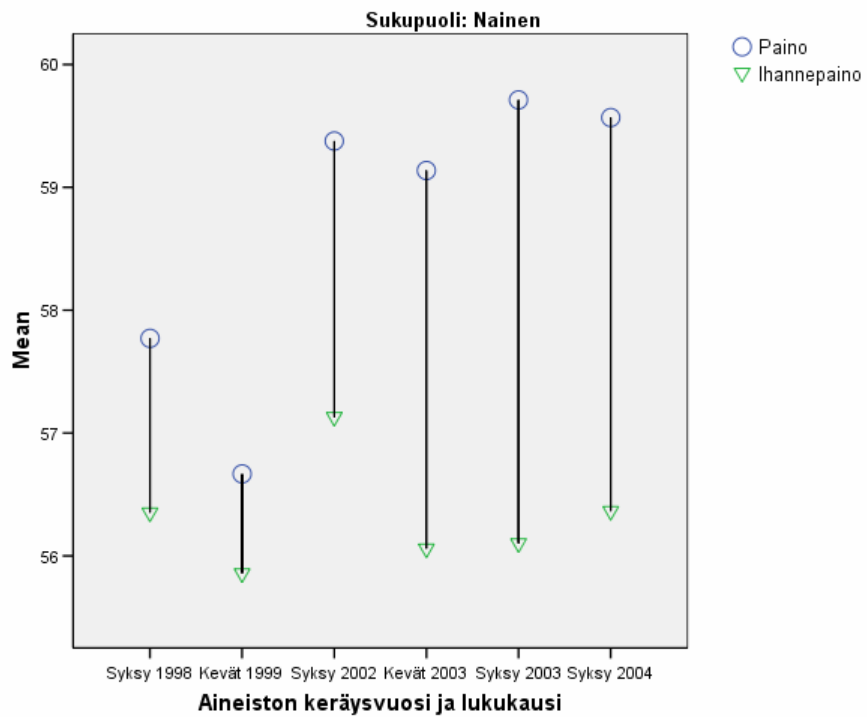
Seuraavia kuvia varten koko aineisto ensiksi splitattiin (Data/Split file) sukupuolen mukaan, jonka jälkeen lähdettiin tekemään kuvaa. Drop-Line -toiminto on yksi viivakuvan muoto. Sillä voi verrata saman muuttujan arvoja eri ryhmillä. Lisäksi nyt haluttiin samanaikaisesti verrata kahden eri muuttujan arvoja, joten kuva tilattiin seuraavilla valinnoilla:



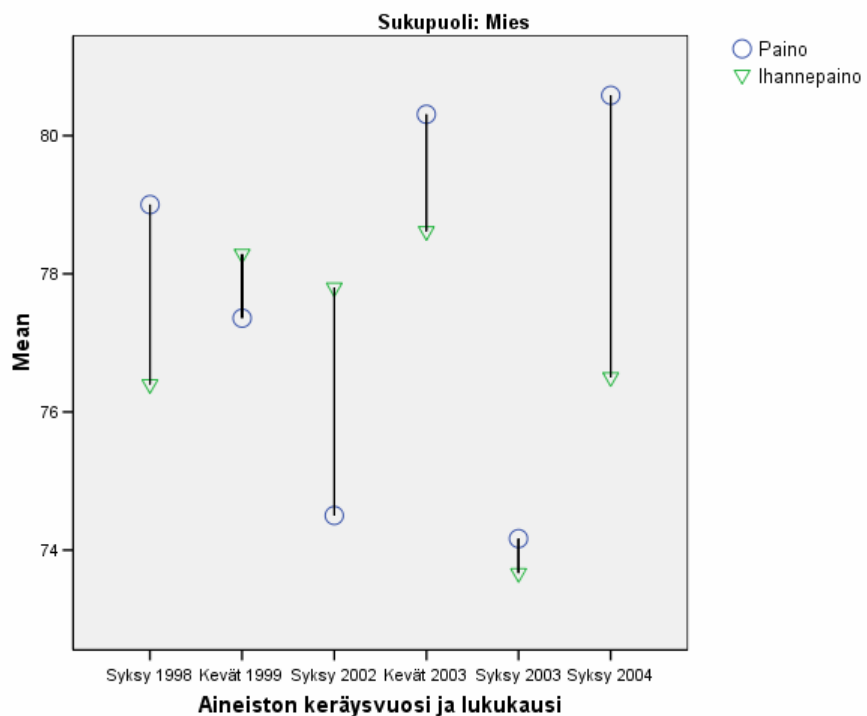
Tämän jälkeen muuttujat sijoitettiin valintaikkunaan seuraavasti: Points represent -kenttään valittiin muuttujat nykyinen paino ja ihannepaino, ja olemme kiinnostuneita niiden keskiarvoista. Change statistics -painikkeen takaa löytyy useita muita mahdollisia tilastollisia tunnuslukuja, joista voi tapauskohtaisesti valita sopivimman vaihtoehdon, kuten esim. mediaani ja moodi.



Mieti, mitä kuvista 14 ja 15 on nähtävissä?



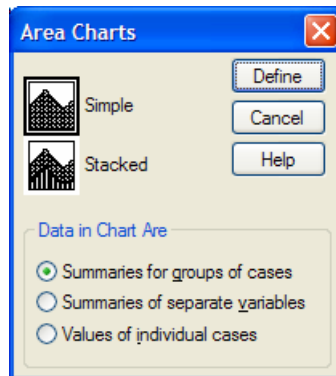
Kuva 14. Naisten nykyisen painon ja ihannepainon keskiarvo eri ajankohtina.



Kuva 15. Miesten nykyisen painon ja ihannepainon keskiarvo eri ajankohtina.

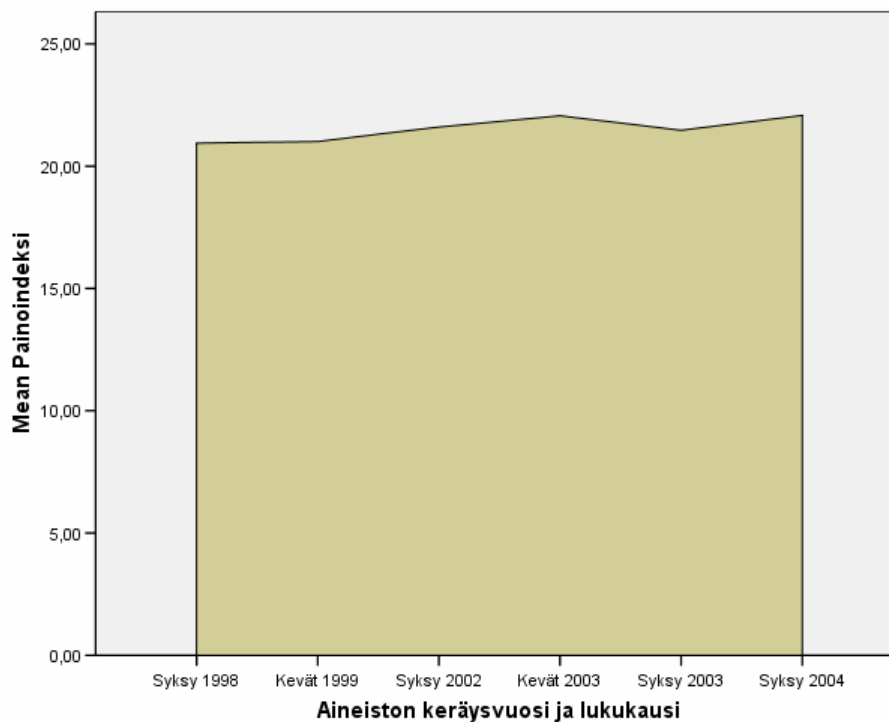
4.6 Aluekuvio

Aluekuvioilla kokonaisalueen pinta-alan vaihtelun aika (tms.) -akselilla tulisi kuvata jonkin asian esiintymisen muutosta. Aluekuvioiden käyttöalueet ovat usein samat kuin viivakuvioiden. Aluekuvioita pääsee tekemään valikosta Graphs/Area. Tällöin avautuu seuraava ikkuna:



Aluekuvan tekemisen vaihtoehdot ovat Simple ja Stacked. Data in Chart are -kentässä on samat vaihtoehdot kuin pylväs- ja viivakuvissa: Summaries for groups of cases on oletusmenetelmä, ja usein parhaiten sopiva esittämistapavaihtoehto. Seuraavassa esimerkissämme käytämme sitä. Muut vaihtoehdot ovat: Summaries of separate variables ja Values of individual cases.

Aluekuvion tilausvalikko on vastaavanlainen kuin viivakuvion, ohjelmaa pyydettiin laskemaan painoindeksin keskiarvo eri keräysajankohtina. Mitä kuvasta 16 on nähtävissä? Eikö olisikin mielenkiintoista katsoa sukupuolittain eriteltynä samaa asiaa?



Kuva 16. Vastanneiden painoindeksi eri ajankohtina.

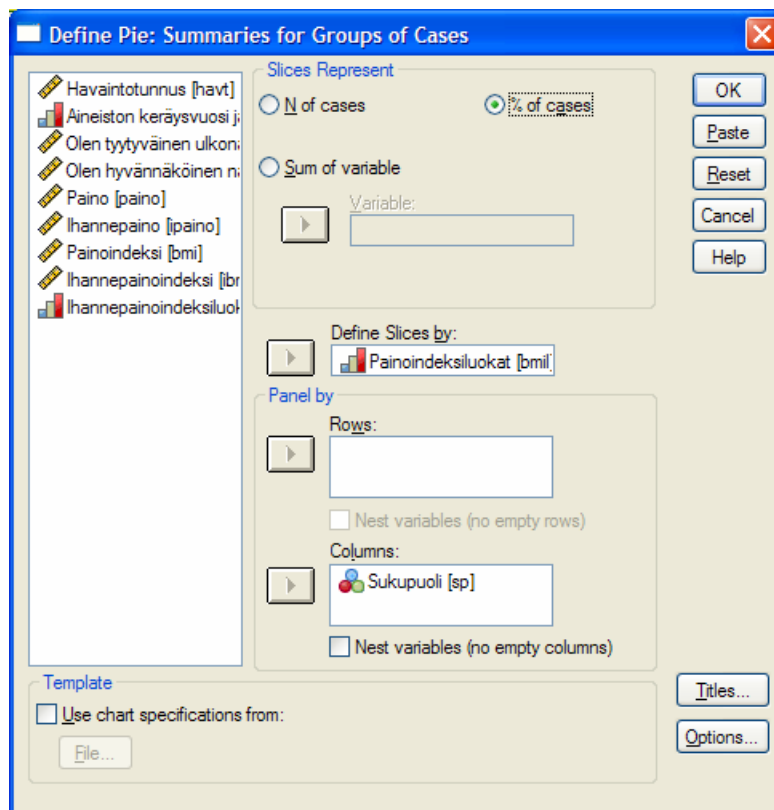
4.7 Piirakkakuvi

Piirakkakuviolla (Pie) on tapana havainnollistaa suhteellisia eli prosentuaalisia osuuksia siten, että kunkin ympyrän sektorin pinta-ala kuvaa tämän sektorin suhteellista osuutta koko aineistossa. Jotain luokkaa voidaan korostaa irrottamalla ko. luokan sektori hieman erilleen muusta kuviosta. Kahden piirakan vertailu on vaikeaa, elleivät erot ole selvät. Parhaiten piirakka sopiikin yleensä kuvaamaan jonkin ominaisuuden jakautumista yhdessä joukossa.

SPSS-ohjelmassa piirakkakuvioita pääsee tekemään Graphs-valikosta toiminnolla Pie. Tällöin avautuu seuraava ikkuna, joka pitää sisällään samat valintavaihtoehdot siitä, miten aineisto piirakkakuvassa esitetään kuin pylväs-, viiva- ja aluekuvioiden valikoissa:

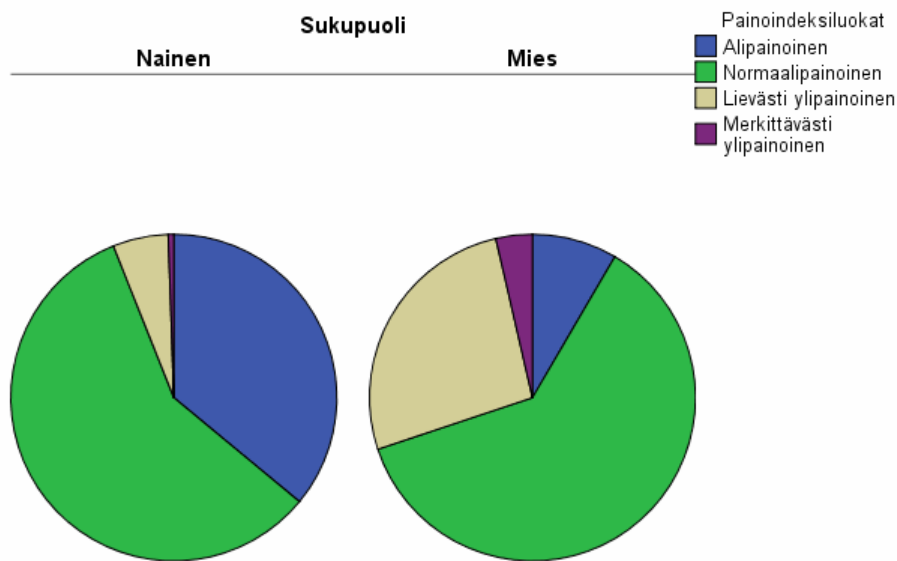


Oletusvalinnan hyväksymisen jälkeen aukeaa piirakanteko-ikkuna:



Perusgrafiikan tekemisen valikossa ei ole paljon erilaisia valinnan mahdollisuuksia. Slices Represent -kenttään valitaan, millä tavalla siivut halutaan esittää. Tässä valittiin prosentit.

Define Slices by -kenttään valitaan muuttuja, josta ollaan kiinnostuneita. Paneelimuuttujaksi (sarake) valittiin sukupuoli.



Kuvasarja 17. Naisten ja miesten sijoittuminen eri painoindeksiluokkiin (%).

4.8 Kuvat niiden käyttötarkoituksen mukaan

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

Yhden muuttujan kuvailu

Yhden jatkuvan muuttujan jakauman kuvailuun soveltuvat histogrammit. Luokitetulle muuttujalle pylväskuviot, viiksikuviot ja 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 syntaksissa. Ensimmäinen kuvaaja määritellään valmiiksi valikkojen kautta, jonka jälkeen painetaan `paste`-painiketta. Tämä kopioi kuvaajan määrittelyt komentomuotoisina syntaksi-ikkunaan. Valintaikkunassa komennot voi `copy-paste` toiminnoilla monistaa ja tehdä niihin haluttuja muutoksia, kuten muuttujien vaihto.

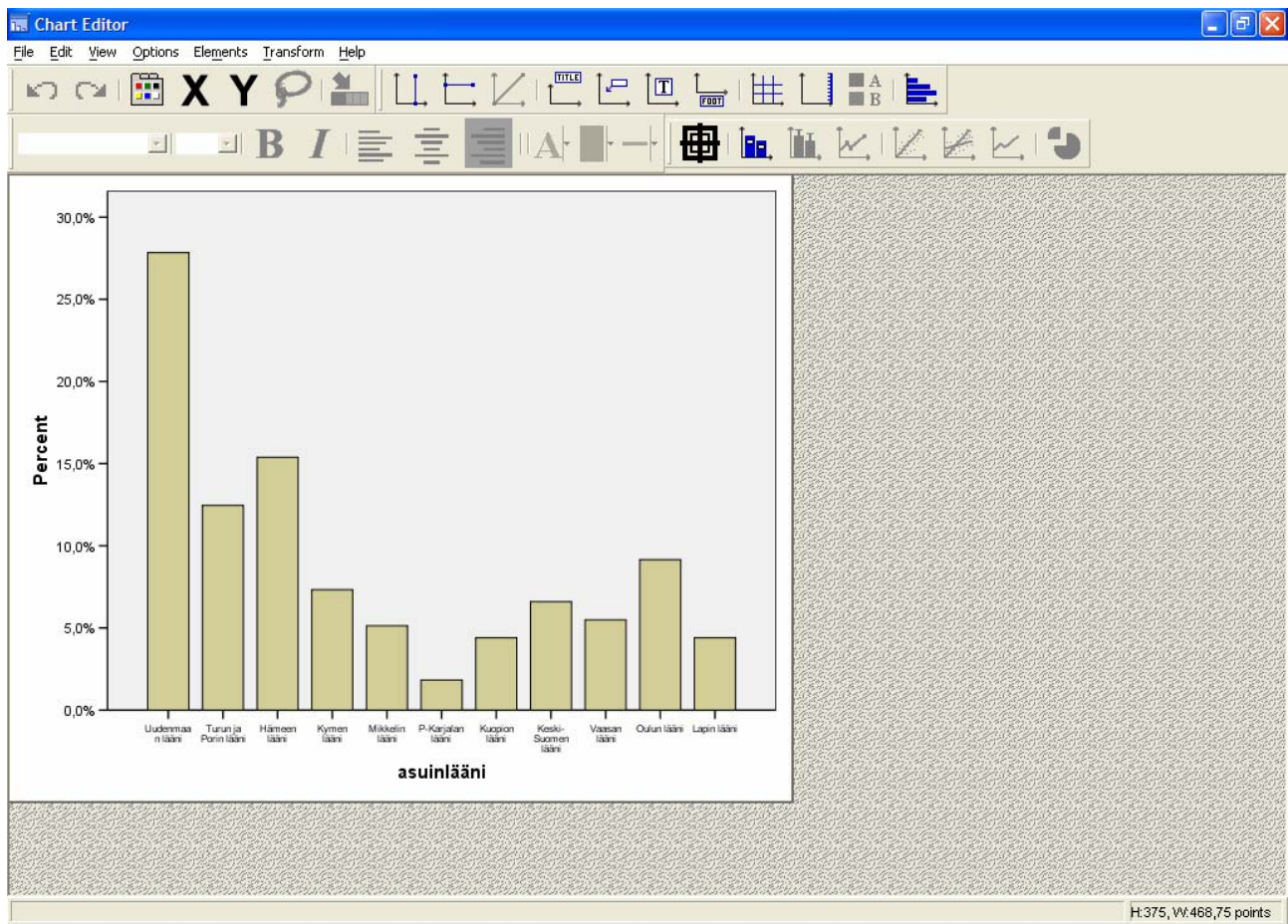
Muuttujan kuvailu toisen luokissa eli yhteisjakauman kuvailu

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

4.9 Kuvien muokkaus

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 esimerkiksi piirakan pylväskuvioksi.

Kuva-editoriin (*Chart Editor*) pääset tuplaklikkaamalla tulostusikkunassa (*Output*) ko. kuvan päällä. Tällöin kuvasta muodostuu kopio kuvankäsittelyikkunaan, jota voi muokata hiiren (aloittelijalle tämä voi olla useissa tapauksissa helpompi tapa oppia muokkaamaan 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 tulevat 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 ennen kuin painaltaa hiirtä uudestaan. Näin pääset muuttamaan tekstiä. Jos kuvasta puuttuu jokin tekstiosa, voit lisätä sen `Options`-valikosta.

Otsikko

Mikäli haluaa jo SPSS-ohjelmassa otsikoida kuvan, se on kätevintä tehdä siinä vaiheessa kun ”tilaa” eli tekee kuvan. Yleensä niin kuvat kuin taulukotkin kannattaa otsikoida vasta tekstinkäsittelyohjelmassa, koska usein tutkimuksen edetessä kuvien ja taulukoiden sijoittelu voi muuttua, jolloin valmiiksi SPSS-ohjelmassa numeroidut kuvat ja taulukot aiheuttavat ongelmia.

Osien tuhoaminen

Jos haluat tuhota kuvista osia, esim. poistaa kuvasta yhden pylvään, aloita työskentely tuplaklikkaamalla kuvaajan (esim. poistettavan pylvään) päällä. Esille tulevasta `Properties`-ikkunan `Categories`-välilehdeltä voit poistaa pylvään/pylväitä valitsemalla hiirellä `Order`-laatikosta haluamasi muuttujan luokka/luokat ja painamalla hiirellä punainen risti -painiketta. Tällöin haluamasi muuttujan luokka siirtyy `Excluded`-laatikkoon. Halutessaan takaisin esim. poistetun pylvään sen saa edellä mainitussa valikossa valitsemalla sen hiirellä ja sitten painamalla vihreä 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 `Auto`-ruudusta pois. X-akselin skaalausta ei voi jälkikäteen muuttaa, vaan se määräytyy aineistossa olevien arvojen perusteella.

Värien vaihtaminen

Valitse haluamasi kuvan osa klikkaamalla sen päällä, jolloin 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 kuvan 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 `Options/Transpose chart`.

Muita

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

Kuvan tallentaminen

A) Tulostusikkunassa: Kuvan voit tallentaa 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 output` -valikossa vaihtoehtoa `Chart(s) only`. Mikäli aiot siirtää kuvan johonkin toiseen ohjelmaan, varmin tapa tulostusikkunassa ollessasi, on tallentaa kuva PNG -muodossa (`File/Export`).

B) Kuvankäsittelyikkunassa: Kuvan voi tallentaa kuvankäsittelyikkunassa `File`-valikon `Export chart XML` -toiminnolla.

C) Kolmas tapa, vaikkakaan ei yhtä tasokas, on klikata kuvaa ja valita ”Copy object”. Tulee kuitenkin muistaa, että julkaisuihin tällä tavoin tehty kuvan siirto, ei ole laadultaan riittävä.

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

5 TULOsten SIIRTÄMINEN MUIHIN OHJELMIIN

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.

SPSS:ssä kuvat voidaan tallentaa sellaiseen muotoon, että ne on mahdollista siirtää jatkokäsiteltäväksi grafiikkaohjelmaan, tai ne voidaan siirtää suoraan tekstinkäsittelyohjelmaan.

Output-ikkunassa valitaan `File`-valikosta toiminto `Export`. `Export`-kentästä valitaan `Charts only` ja `Export Format` -valikosta haluttu tiedostotyyppi. Suositeltavin tiedostomuoto kuvien tallentamiseen on nykyisin `PNG`. `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.

Mellin (1995): *Tilastotieteen johdantokurssi Versio 1.2*. Helsingin yliopisto: Tilastotieteen laitos.
Löytyy useita eri painoksia.

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 (2007): *Using multivariate statistics*. Uusin painos, Boston: Allyn ja Bacon.
Paljon käytetty ja arvostettu teos.

Vehkalahti (2006): Tilastollisista menetelmistä paljon hyödyllistä opiskelumateriaalia.
<http://www.helsinki.fi/~kvehkala/julkaisut.html>.

SPSS:n tietoa/oppaita verkossa:

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

Paljon oleellista infoa SPSS:ää koskien sekä ystävälliset että auttavaiset neuvontahenkilöt SPSS:ää koskevissa ongelmissa.

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

Myös paljon oleellista infoa SPSS:stä.

SPSS, version 14 demo 14 päiväksi:

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

Demon käyttöön saanti vaatii rekisteröitymisen, jonka jälkeen voi siirtyä sivustolle:

http://www.spss.com/registration/premium/consol056.cfm?Demo_ID=37,

hankkimaan uuden version demon käyttöönsä. Syyskuun puolivälistä lähtien on ollut mahdollista saada demo versiosta 15.

Helsingin yliopisto, Antti Nevanlinna:

<http://www.helsinki.fi/atk/tilasto/spsspk13/kurssi.html>

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

Menetelmätietovarasto (FSD): <http://www.fsd.uta.fi/menetelmaopetus/SPSS/spss.html>

Monelta osin hyödyllistä tietoa, vaikkakin monelta osin vielä keskeneräinen.

Tilastokeskuksen verkkokoulusivut

<http://www.tilastokeskus.fi/tup/verkkokoulu/index.html>

Sivuilta löytyy merkittävän paljon hyödyllistä tietoa.