

## Johdanto

Tässä oppaassa esitetään SPSS-tilasto-ohjelman alkeita, kuten Excel-tiedoston avaaminen, tunnuslukujen laskeminen ja uusien muuttujien muodostaminen. Lisäksi esitetään esimerkkien avulla muutamia tilastollisia testejä.

## Sisältö

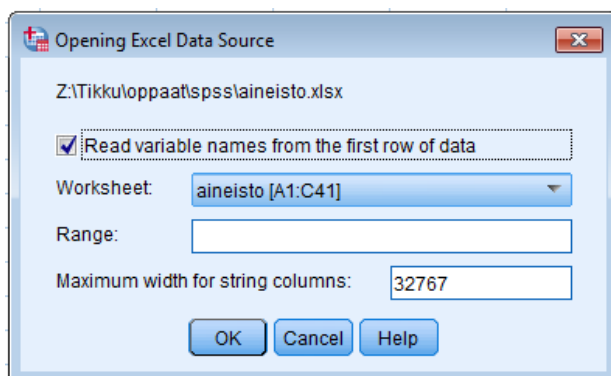
- 1 Excel tiedoston avaaminen
- 2 Aineiston valmistelu
- 3 Tunnuslukujen laskeminen
- 4 Uusien muuttujien laskeminen
- 5 Tilastollisia testejä

## 1 Excel-tiedoston avaaminen

Käynnistä SPSS-ohjelma. Valitse:

File → Open → Data

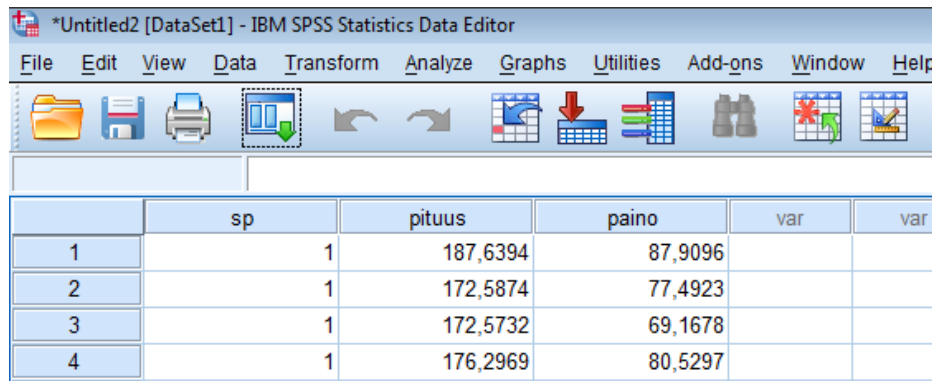
ja etsi Excel-tiedoston sijainti. Valitse "Files of type" -valikosta "Excel", jolloin saat näkyviin vain Excel-tiedostot. Valitse **Open**, jolloin avautuu kuvassa 1 näkyvän kaltainen valikko. Valitse toiminto "Read variable names from the first row of data", mikäli Excel-tiedoston ensimmäisellä rivillä on muuttujien nimet. Valitsemalla tästä valikosta **OK** Excel-tiedosto avautuu SPSS-ohjelmaan.



Kuva 1. Excel-tiedoston avaamiseen liittyvä valikko.

## 2 Aineiston valmistelu

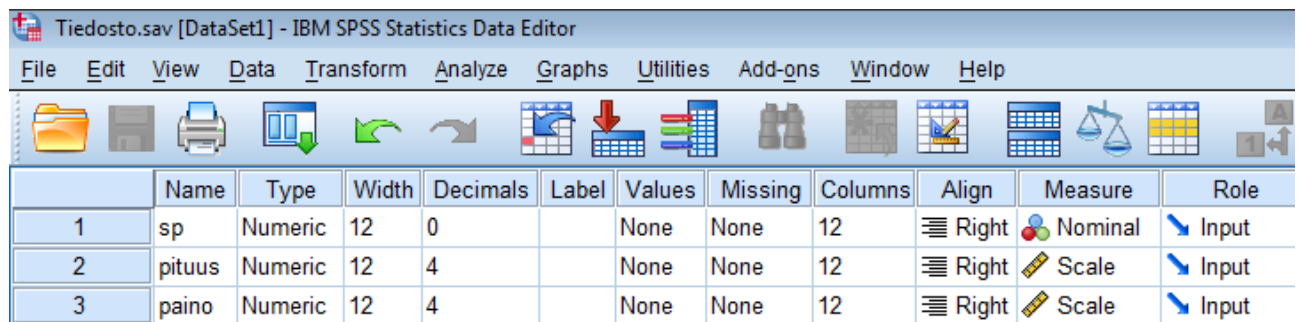
Kuvassa 2 esitetään avatun esimerkkiaineiston neljä ensimmäistä havaintoa ”Data View” -näkyvässä. Aineistossa on kerätty ihmisten pituuksia ja painoja sukupuolittain. Ennen laskelmien tekoa kannattaa varmistaa ”Variable View” -näkyvästä, että muuttujien mitta-asteikot ovat oikein. Näiden kahden näkymän valinta tapahtuu ohjelman vasemmasta alareunasta.



	sp	pituus	paino	var	var
1	1	187,6394	87,9096		
2	1	172,5874	77,4923		
3	1	172,5732	69,1678		
4	1	176,2969	80,5297		

Kuva 2. Esimerkkiaineiston neljä ensimmäistä havaintoa. Muuttujina sp (sukupuoli), pituus ja paino.

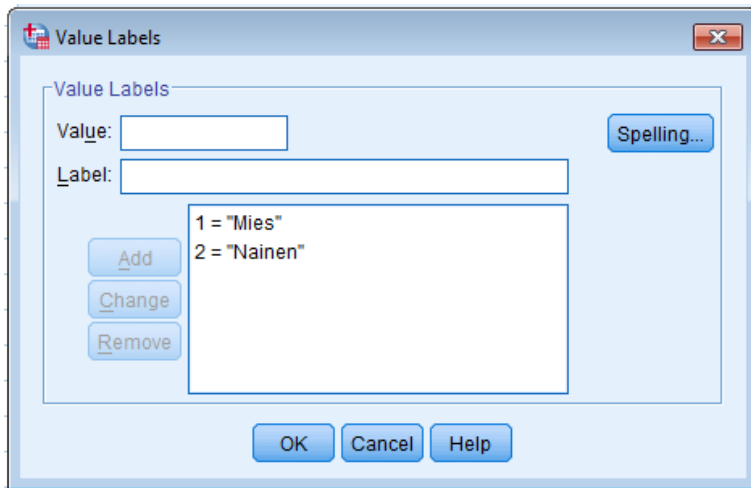
Kuvassa 3 esitetään ”Variable View” -näkyvä. Muuttujien mitta-asteikot ovat menneet oikein tiedoston avaamisessa. Sukupuoli (sp) on luokitteleva muuttuja. Muuttujat pituus ja paino ovat jatkuvia. Tässä kohdassa voidaan vähentää jatkuvien muuttujien desimaalimäärä esimerkiksi kahteen, kohdasta [Decimals](#). Näin tehden aineiston näkyvä on hieman siistimpi.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	sp	Numeric	12	0		None	None	12	Right	Nominal	Input
2	pituus	Numeric	12	4		None	None	12	Right	Scale	Input
3	paino	Numeric	12	4		None	None	12	Right	Scale	Input

Kuva 3. Tiedot aineiston muuttujista.

Muuttujan sukupuolen arvot kannattaa määrittellä. Tällä hetkellä se sisältää arvoja 1 ja 2. Nämä on hyvä määrittellä tarkasti [Values](#)-laatikossa vastaamaan sukupuolia. Aineistossa numero 1 tarkoittaa miestä ja numero 2 naista. Kuvassa 4 on esitetty muuttujien arvojen määrittely.



Kuva 4. Näkymä valikosta, jossa voi määrittellä luokittelevan muuttujan arvot.

### 3 Tunnuslukujen laskeminen

Frekvenssien eli lukumäärien katsominen onnistuu valikon

Analyze → Descriptive Statistics → Frequencies

kautta. Lisää muuttuja sp nuolipainikkeella **Variable**-laatikkoon. Frekvenssit saa näkyviin laittamalla ruksin kohtaan **"Display frequency tables"**. Valikosta **Statistics...** voit valita muita tilastollisia tunnuslukuja.

SPSS näyttää muuttujan sp frekvenssit taulukossa, kuva 5. Miehiä on 18 ja naisia 22, osuuksina 45 % ja 55 %.

**sp**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Mies	18	45,0	45,0	45,0
	Nainen	22	55,0	55,0	100,0
	Total	40	100,0	100,0	

Kuva 5. Muuttujan sp frekvenssit.

Keskiarvojen laskeminen tapahtuu valikon

Analyze → Descriptive Statistics → Descriptives

kautta. Valikosta **Options...** voi valita eri tunnuslukuja. Kuvassa 6 näkyvät muuttujan pituus minimi, maksimi, keskiarvo ja -hajonta.

**Descriptive Statistics**

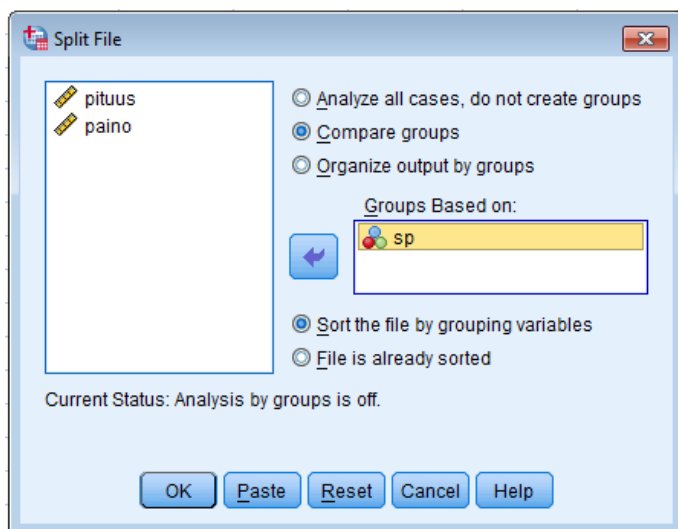
	N	Minimum	Maximum	Mean	Std. Deviation
pituus	40	156,62	195,68	171,3877	8,68621
Valid N (listwise)	40				

Kuva 6. Muuttujan pituus tunnuslukuja.

Usein on tarve katsoa tunnuslukuja jonkin ryhmittelevän muuttujan suhteen. Tässä aineistossa ryhmittelevä muuttuja on sukupuoli. Asetetaan aineiston ositus päälle valikon

Data → Split file

kautta. Kuvassa 7 on esitetty, kuinka aineiston osittaminen tapahtuu. Ruksi on laitettu kohtaan "Compare groups" ja muuttuja sp on siirretty laatikkoon "Groups Based on".



Kuva 7. Aineiston osittaminen muuttujan sp suhteen.

Lasketaan muuttujan paino tunnusluvut, kuten muuttujan pituus tunnusluvut, kun aineiston ositus on päällä. Saadaan tulokset jaoteltuna muuttujan sp suhteen, kuva 8.

Descriptive Statistics						
sp		N	Minimum	Maximum	Mean	Std. Deviation
Mies	paino	18	69,17	98,33	81,1441	7,70493
	Valid N (listwise)	18				
Nainen	paino	22	56,41	75,50	64,9143	5,40904
	Valid N (listwise)	22				

Kuva 8. Tilastollisia tunnuslukuja muuttujasta paino sukupuolittain.

Korrelaatiokertoimen voi laskea valikosta:

Analyze → Correlate → Bivariate

Korrelaatiokerroin mittaa kahden järjestys- tai jatkuva-asteikollisen muuttujan lineaarisen riippuvuuden voimakkuutta. Valitaan muuttujat pituus ja paino **Variables**-laatikkoon ja painetaan **OK**. Pituuden ja painon välinen korrelaatio on 0,86 tässä aineistossa. SPSS näyttää korrelaatiot, kuten kuvassa 9 on esitetty.

### Correlations

		pituus	paino
pituus	Pearson Correlation	1	,861**
	Sig. (2-tailed)		,000
	N	40	40
paino	Pearson Correlation	,861**	1
	Sig. (2-tailed)	,000	
	N	40	40

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

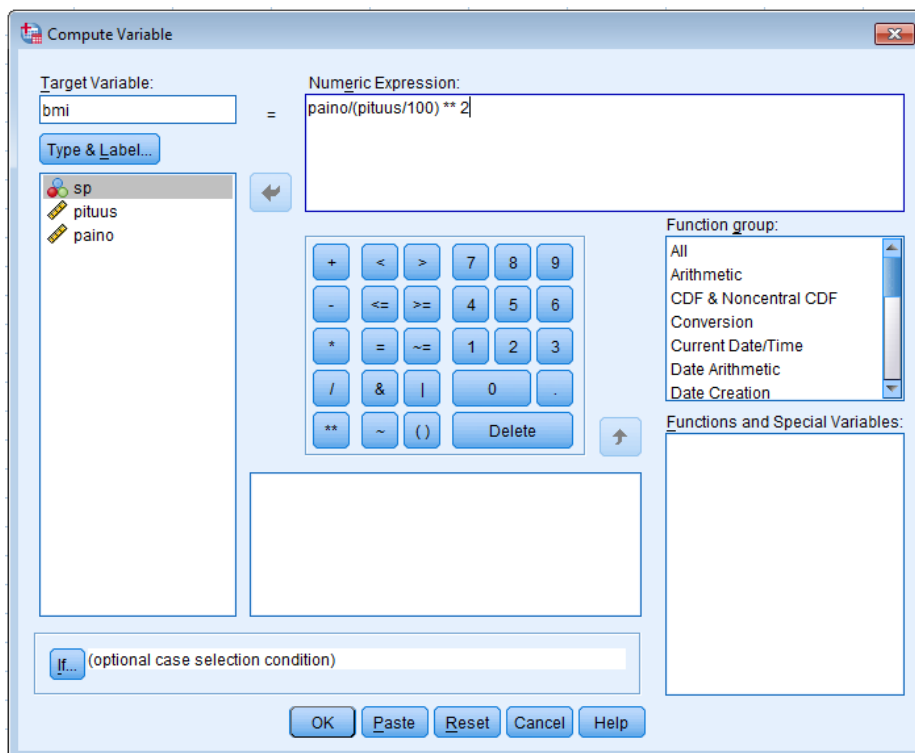
Kuva 9. Pituuden ja painon välinen korrelaatio.

## 4 Uusien muuttujien laskeminen

Lasketaan muuttuja painoindeksi (bmi). Mene valikkoon:

Transform → Compute Variable

Kuvassa 10 esitetään, kuinka painoindeksin laskenta tapahtuu. Laskettavan muuttujan nimi kirjoitetaan kohtaan **Target Variable**; tässä se on bmi. Kaava kirjoitetaan **Numeric Expression** -kohtaan. Painoindeksin laskennassa pituuden on oltava metreissä, aineistossa se on senttimetreissä. Pituusyksikön muuntaminen onnistuu kuvassa 10 esitetyllä tavalla.

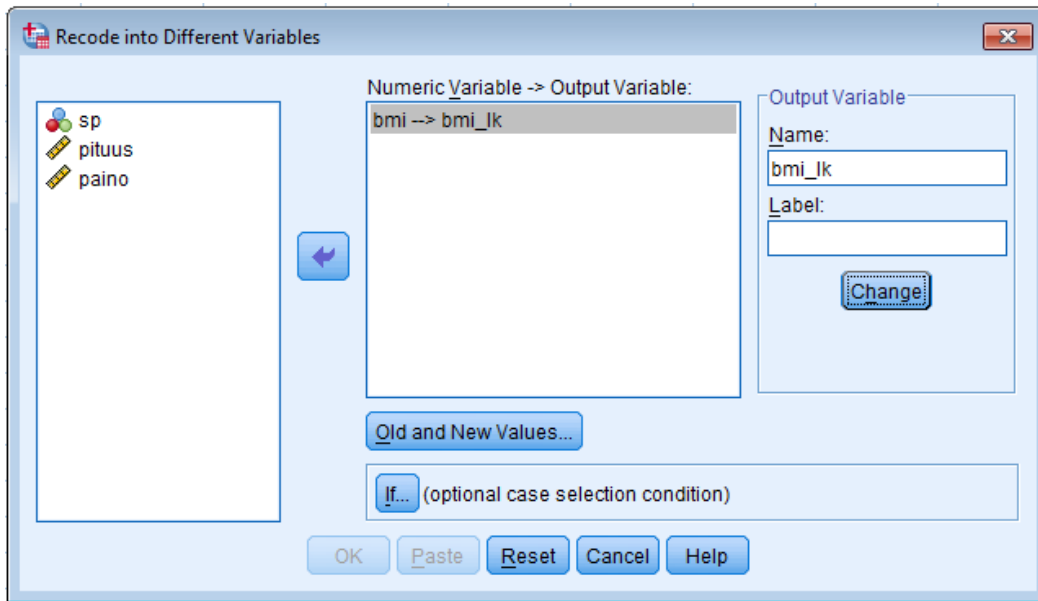


Kuva 10. Muuttujan bmi laskeminen.

Luokitellaan bmi kahteen luokkaan, normaali- ja ylipainoisiin. Mene valikkoon:

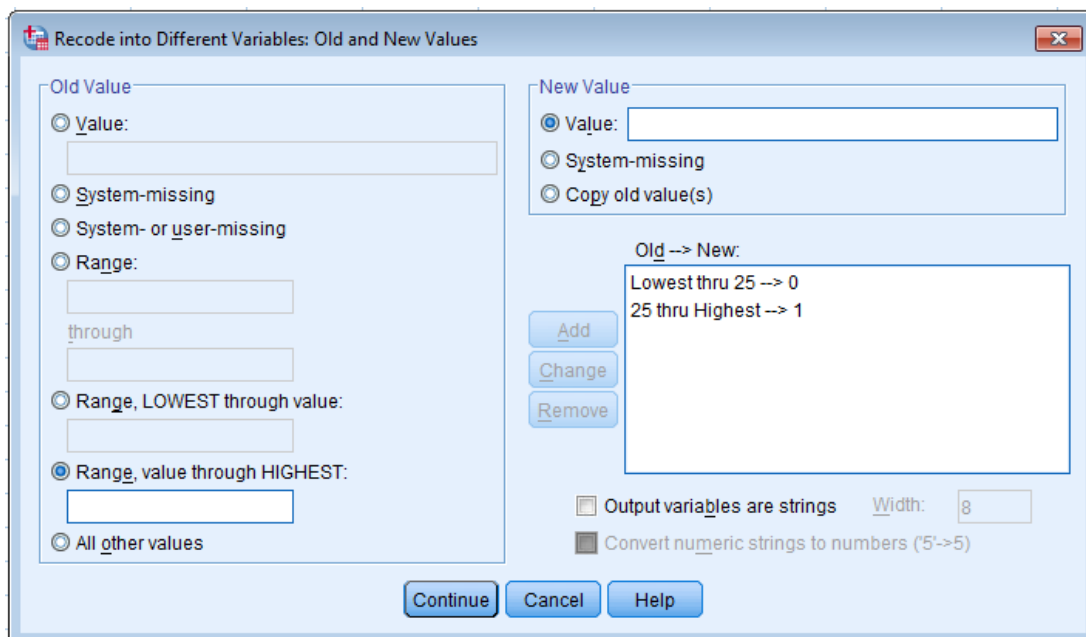
Transform → Recode into Different Variables

Valitse luokiteltava muuttuja ja anna sille nimi laatikkoon **Output Variable** ja paina **Change**. Mene sitten valikkoon **Old and New Values**, kuva 11.



Kuva 11. Muuttujan bmi luokittelu.

Asetetaan ihmiset, joiden bmi on alle 25, normaalipainoisiksi, ja ihmiset, joiden bmi on suurempi kuin 25, ylipainoisiksi. Kuvassa 12 on esitetty painoindeksin luokittelu.



Kuva 12. Painoindeksin luokittelu.

Tämän uuden kaksiluokkaisen muuttujan arvot voi nimetä **Variable View** -näkylässä siten, että 0 tarkoittaa normaalipainoista ja 1 tarkoittaa ylipainoista.

## 5 Tilastollisia testejä

Tilastollisilla testeillä testataan muuttujien välistä riippuvuutta. P-arvo kuvastaa muuttujien välisen riippuvuuden voimakkuutta. Jos p-arvo on alle lukeman 0,05, niin muuttujien välillä on riippuvuutta.

### Ristiintaulukointi

Ristiintaulukoinnilla voidaan tutkia kahden luokittelevan muuttujan yhteyttä. Katsotaan, onko sukupuoli ja ylipaino yhteyttä. Mene valikkoon:

Analyze → Descriptive Statistics → Crosstabs

Valitse muuttujat riveille ja sarakkeille. Valikosta **Statistics** valitse **Chi-square**. Tällä valinnalla saat p-arvon muuttujien väliselle riippuvuudelle. Kuvassa 13 on muuttujien sukupuoli ja ylipaino välinen ristiintaulukko ja riippuvuustestin tulokset.

Count

		bmi_ik		Total
		normaalipainoinen	ylipainoinen	
sp	Mies	8	10	18
	Nainen	17	5	22
Total		25	15	40

#### Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4,552 <sup>a</sup>	1	,033		
Continuity Correction <sup>b</sup>	3,259	1	,071		
→ Likelihood Ratio	4,612	1	,032		
Fisher's Exact Test				,050	,035
Linear-by-Linear Association	4,438	1	,035		
N of Valid Cases	40				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6,75.

b. Computed only for a 2x2 table

Kuva 13. Ristiintaulukko ja p-arvo luokitteleville muuttujille sukupuoli ja painoindeksi.

Riippuvuus testin p-arvo on luettavissa riviltä **Pearson Chi-Square**. Tässä  $p=0,033 < 0,05$  eli ylipainolla ja sukupuolella on yhteys, miehet ovat useammin ylipainoisia kuin naiset.

### Keskiarvojen vertailu

Keskiarvojen vertailu suoritetaan usein t-testillä. Tästä aineistosta esimerkkinä voisi olla: eroavatko pituudet ylipainon mukaan? Mene valikkoon:

Analyze → Compare Means → Independent-Samples T Test

Valitse tutkittava muuttuja laatikkoon **Test Variable(s)**. Määrittele ryhmittelevä muuttuja kohtaan **Grouping Variable** ja aseta painikkeesta **Define Groups** numerot, jotka kuvastavat ryhmiä. Kuvassa 14 esitetään t-testin tulokset.

### Group Statistics

bmi_ik	N	Mean	Std. Deviation	Std. Error Mean
pituus normaalipainoinen	25	170,2947	9,71810	1,94362
ylipainoinen	15	173,2093	6,52646	1,68513

### Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
pituus	Equal variances assumed	1,271	,267	-1,028	38	,310	-2,91455	2,83483
	Equal variances not assumed			-1,133	37,408	,264	-2,91455	2,57241

Kuva 14. T-testin tulokset.

Testissä  $p=0,31$  eli normaali- ja ylipainoisia voidaan pitää yhtä pitkinä. Normaalipainoisten pituus on keskimäärin 170 cm ja ylipainoisten 173 cm.

### Lisätietoja

Internetistä on saatavilla useita käyttökelpoisia ja kattavia [SPSS oppaita](#) (Google).

Tämän tutkimusoppaan on laatinut Tuomas Selander, biostatistikko, KYS Tiedepalvelukeskus, tuomas.selander (at) kuh.fi