

# SPSS-ohjelman käyttö tilastollisen aineiston käsittelyssä ja analysoinnissa

## *OSA 1* *SPSS versiolle 19.0*

*Jorma Vainionpää 2012*

### SISÄLTÖ

1. AINEISTON SYÖTTÄMINEN JA MUUTTUIEN MÄÄRITTELY .....	1
2. MUUTTUIEN KUVAILUTIEDOT .....	2
3. FREKVENSSIJAKAUMAT .....	4
4. MUUTTUIEN LUOKITTELU JA KÄÄNTÄMINEN .....	5
5. KAAVOJEN AVULLA MÄÄRITELTÄVÄT MUUTTUIET .....	6
6. KUVIOIDEN LAATIMINEN .....	8
7. RISTIINTAULUKOINTI KAKSIULOTTEISEN JAKAUMAN KUVAAJANA .....	9
8. RISTIINTAULUKOINTIIN LIITTYVÄT TILASTOLLISET TESTIT .....	10
9. KORRELAATIO .....	13
10. KAHDEN KESKIARVON VERTAILU: t-testit.....	15
11. YKSISUUNTAINEN VARIANSSIANALYYSI.....	19
12. JAKAUMAN NORMAALISUUDEN TESTAUS: Kolmogorov-Smirnovin testi.....	23
13. MANN-WHITNEYN U-TESTI.....	24
14. KRUSKAL-WALLISIN TESTI .....	25
15. WILCOXONIN TESTI .....	27
16. RELIABILITEETIN MÄÄRITTÄMINEN MUUTTUJAJOUKOLLE .....	28

# 1. AINEISTON SYÖTTÄMINEN JA MUUTTUJIEN MÄÄRITTELY

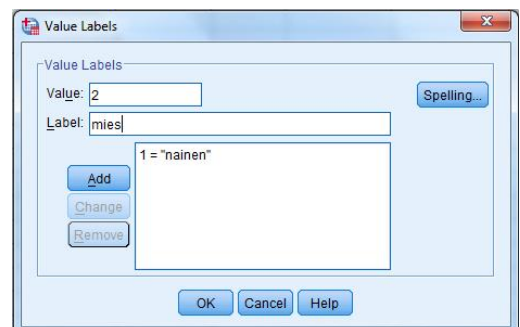
0. Käynnistä **SPSS**-ohjelma. Syötä tai laadi uusi havaintomatriisi ohjelman ollessa **Data View**-tilassa (ks. kuva alla). Kannattaa yleensä syöttää aineisto **numeroina** ja yksi tapaus (eli rivi) kerrallaan. Usein kuitenkin kannattaa tehdä **ensin** muuttujien määrittely eli aloittaa kohdasta 2 ennen aineiston syöttämistä.
1. Tallenna matriisi **File**-valikon **Save as** -toiminnolla. Mieti ja katso tarkkaan, minne tallennat. Tee aina varmuuskopio(t) matriisista työskentelyn loppuksi tai jo sen aikana. Jatkossa saat aineiston auki valitsemalla **File** -valikosta **Open – Data**.
2. **Nimeä** ja **määrittele** muuttujat. Muuttujia pääset määrittelemään valitsemalla **Data Editor** -näytön alanurkasta **Variable View** -ikkunan näkyviin. **TÄMÄN VOIT SIIS TEHDÄ JO ENNEN AINEISTON SYÖTTÄMISTÄ.**

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
sukupuoli	Numeric	8	0	sukupuoli	{1, nainen}...	None	8	Right	Nominal	Input

- Name** = Muuttujan nimi datamatriisin sarakkeessa, ei muuta merkitystä jos Label syötetään. On kuitenkin pakollinen tieto.
- Type** = Muuttujan tyyppi, yleensä numeerinen (numeric), joskus merkkijono (string) tai joku muu.
- Width** = Kuinka monta merkkiä muuttujan arvoksi voi syöttää.
- Decimals** = Kuinka monta desimaalia muuttujan arvossa voi olla, laitetaan nolla jos desimaaleja ei tarvita.
- Label** = Muuttujan tarkempi nimi, selite. Erittäin tärkeä, auttaa analyysien tulkinnessa tai luettavuudessa.
- Values** = Muuttujan arvojen selitteet (Value Labels). Erittäin hyödyllinen, jos muuttujan arvot voidaan nimetä. Auttaa analyysien tulkinnessa tai luettavuudessa. Ei kuitenkaan käytetä, jos muuttuja saa paljon erilaisia arvoja tai niitä ei ole mielekästä nimetä.

Esimerkki:

1. Syötä numero 2 Value-kohtaan.
2. Syötä sana *mies* Label-kohtaan.
3. Paina Add -painiketta.



- Missing** = Puuttuvien arvojen määrittely. Voidaan määrittellä, mitkä muuttujan arvot tulkitaan puuttuviksi.
- Columns** = Havaintomatriisin sarakkeen leveys, ei muuta merkitystä.
- Align** = Havaintomatriisin sarakkeessa arvon sijainti (tasaus), ei muuta merkitystä.
- Measure** = Muuttujan mitta-asteikon määrittely. Opettele mitta-asteikot! Vaihtoehdot:  
**Nominal** = laatueroasteikko    **Ordinal** = järjestysasteikko    **Scale** = Suhde- tai intervalliasteikko
- Role** = Muuttujan rooli analyysissa. Vain jotkut analyysit tukevat role-toimintoa.  
 Esim. Input = selittäjä, Target = selitettävä, Both = voi olla sekä selittäjä että selitettävä.

3. Siirry takaisin **Data View** -ikkunaan (näytön alanurkasta). Syötä aineisto rivi (tapaus) kerrallaan. Aineisto kannattaa yleensä syöttää numeroina, jos se mahdollista, koska se on nopein tapa. Numeroille annoit selitteet edellisessä kohdassa **Value labels** -ikkunassa.

42	mies	41
43	nainen	40
44	nainen	49

4. **View**-valikon **Value labels** -kohdan aktivoimalla näet määrittelemäsi Labelit (= muuttujan saamien arvojen nimet eli selitteet). Voit käyttää myös pikapainiketta työkaluriviltä.

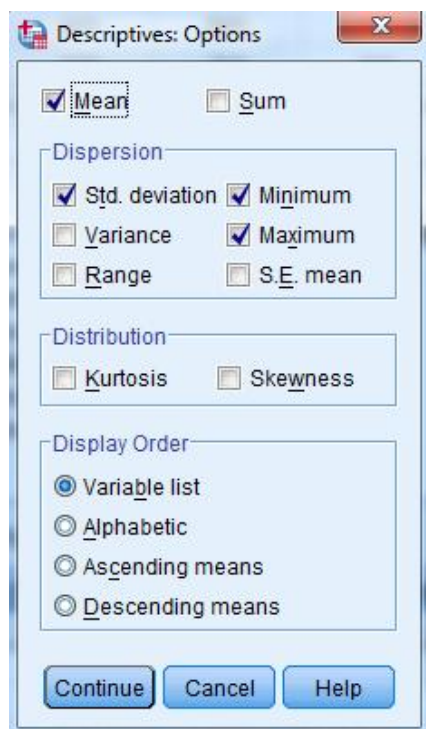
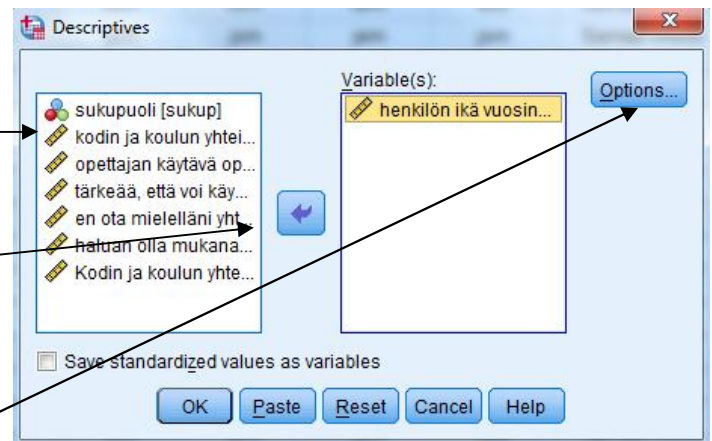
## 2. MUUTTUJIEN KUVAILUTIEDOT

1. Muuttujien tunnuslukujen ja mahdollisten syöttövirheiden tarkastelua varten kannattaa pyytää muuttujien kuvailutietoja **Analyze**-valikosta **Descriptive Statistics – Descriptives**.

Valitse muuttujat näpäyttämällä vasemmalla olevasta muuttujalistasta.

Siirrä nuolta näpäyttämällä valitut muuttujat oikealla olevaan laatikkoon. Nuoli toimii myös toisinpäin, jos tulee virhevalinta

**Options**-painikkeen takaa saat valita tarvitsemasi tunnusluvut.



### DESCRIPTIVES –VAIHTOEHDOT

**Mean** = keskiarvo

**Sum** = muuttujan arvojen summa

#### Hajontaluvut (Dispersion)

**Std. deviation** = keskihajonta

**Variance** = varianssi (keskihajonnan neliö)

**Range** = vaihteluvälin pituus

**Minimum** = muuttujan pienin havaittu arvo

**Maximum** = muuttujan suurin havaittu arvo

**S.E. mean** = keskiarvon keskivirhe

#### Jakauman muodon kuvailu (Distribution)

**Kurtosis** = jakauman huipukkuus suhteessa normaalijakaumaan, positiivinen arvo tarkoittaa huipukkaampaa kuin normaalijakauma

**Skewness** = jakauman vinous suhteessa normaalijakaumaan, positiivinen arvo tarkoittaa oikealle vinoa jakaumaa

#### Tulostusjärjestys

**Display Order** = muuttujien tulostusjärjestys, voi tarvittaessa muuttaa.

2. Valitse haluamasi tunnusluvut. Näpäytä **Continue**-painiketta ja sen jälkeen **Descriptives**-ikkunan **OK**-painiketta.

3. Ohjelmaan avautuu (viimeistään nyt) uusi ikkuna eli **SPSS Statistics Viewerin Output**-ikkuna. Siihen tulostuvat kaikki tehdyt analyysit ja muut objektit. Output-ikkunan voit tallentaa (File – Save as) tai tulostaa (File – Print). Tallentaminen ja tulostaminen eivät tosin ole aina välttämättömiä, koska analyysit voi yleensä toistaa varsin helposti. Tärkeimmät ja hankalimmin toistettavat analyysit voi tallentaa ja/tai tulostaa. Valitse hiirellä näpäyttämällä näytön vasemman reunan navigointi-ikkunasta haluamasi tulosteet ja sen jälkeen **File**-valikosta **Print**. Tarkista, että **Print**-ikkunassa on valittuna Selected output, jos et halua tulostaa koko Output-ikkunan sisältöä.

Navigointi- ja valintaikkuna. Voit liikkua sen avulla edestakaisin analyysien välillä tai esim. poistaa/valita jonkin osan.

Yksittäiset analyysit taulukoina. Esimerkkinä Descriptives

	N	Minimum	Maximum	Mean	Std. Deviation
henkilön ikä vuosina	60	22	63	37,00	8,267
Valid N (listwise)	60				

IBM SPSS Statistics Processor is ready | H: 184, W: 1097 pt.

4. Voit myös siirtää analyysit esim. tekstinkäsittelyohjelmaan valitsemalla hiirellä näpäyttämällä haluamasi taulukon ja sen jälkeen **Edit**-valikosta **Copy**. Kopioinnin voi tehdä myös hiiren oikean näppäimen kautta saatavalla valikolla (näpäytä objektin päällä) tai näppäinkomennolla **Ctrl+C**. Tekstinkäsittelyohjelmassa valitset **Muokkaa/Edit** -valikosta **Liitä/Paste** tai **Liitä määräten/Paste Special**. Kopioinnissa ja liittämässä kannattaa kokeilla erilaisia vaihtoehtoja varsinkin kohdeohjelmassa (esim. tekstinkäsittelyohjelma), kunnes löytää kulloinkin toimivimman ratkaisun.

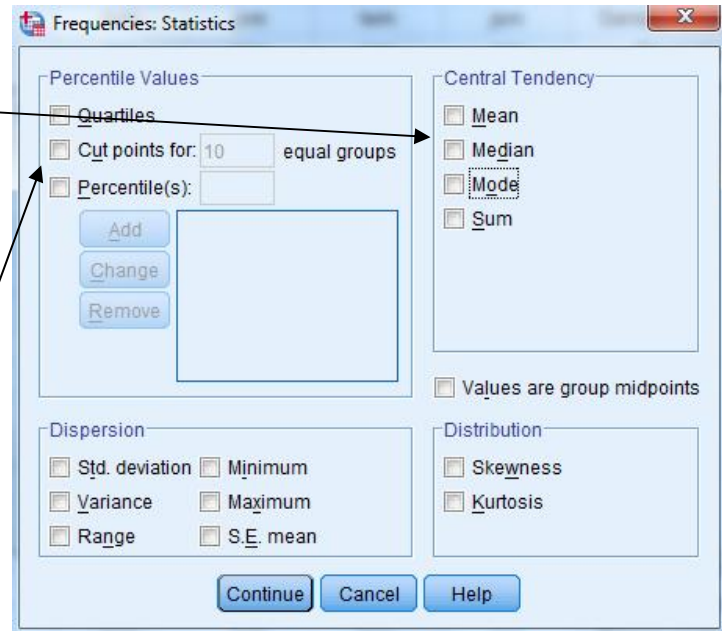
### 3. FREKVENSSIJAKAUMAT

1. Kun haluat aineiston nominaaliasteikollisista tai muuten muutamia arvoja saavista muuttujista tunnuslukuja, valitse **Analyze** -valikosta **Descriptive Statistics – Frequencies** ja siellä muuttujien valinta samoin edellä kuvailutiedoissa.
2. **Statistics**-painikkeen takaa voit valita tarvitsemasi tunnusluvut.

”Muuttujien kuvailutiedot” -kohdassa esiteltyjen tunnuslukujen lisäksi **Statistics** -ikkunassa voi pyytää mediaanin (Median) ja moodin (Mode).

Lisäksi voit pyytää katkaisupisteitä esim. muuttujien luokittelua varten **Percentile Values** -kohdassa.

Esim. **Cut points for** – toiminnolla voit määrittellä katkaisupisteet haluamallesi määrälle ryhmiä



3. Tarkastele aineiston muuttujien frekvenssijakaumia (suorat jakaumat) ja pyytämiäsi tunnuslukuja.

Esimerkkitulostus:

Statistics		
sukupuoli		
N	Valid	60
	Missing	0

Mukana olevien tapausten lukumäärä ja pyydetyt tunnusluvut tulostuvat ylempään taulukkoon (tässä ei ole pyydetty tunnuslukuja.)

sukupuoli					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	nainen	30	50,0	50,0	50,0
	mies	30	50,0	50,0	100,0
	Total	60	100,0	100,0	

Alempaan taulukkoon tulostuvat muuttujan arvojen:

- frekvenssit
- prosenttiosuudet

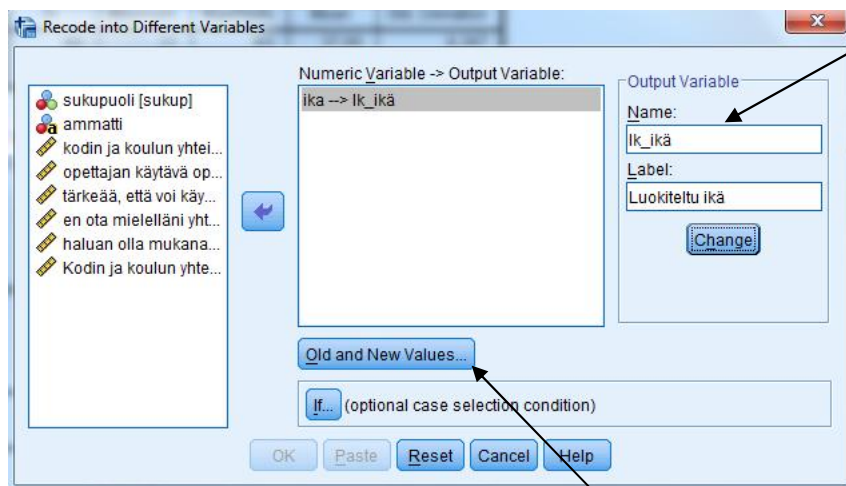
huom. Valid Percent = puuttuvat havainnot poistettu

- kertymäprosentit (Cumulative Percent)

## 4. MUUTTUJAN LUOKITTELU JA KÄÄNTÄMINEN

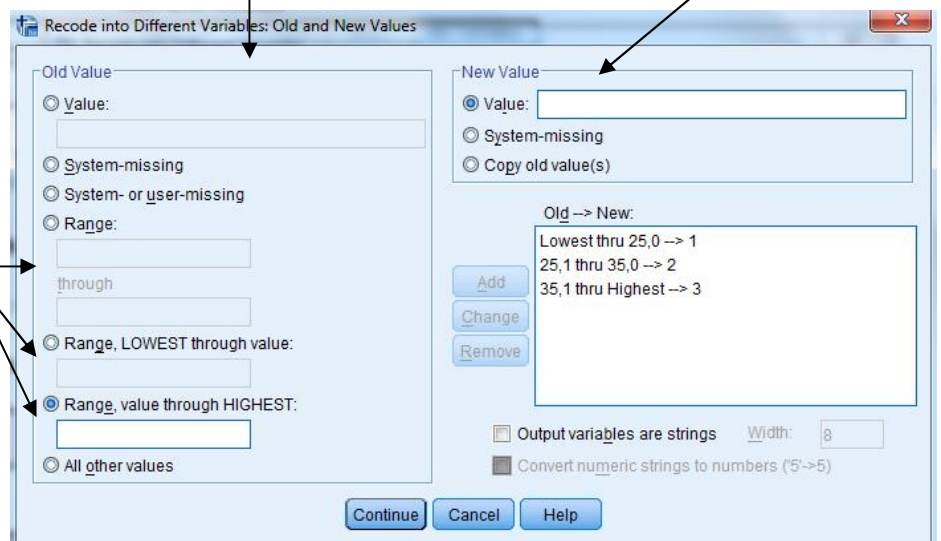
### LUOKITTELU

- Varsin usein mittauksessa saatuja muuttujan arvoja on syytä ryhmitellä luokkiin käsittelyn selkeyttämiseksi ja tutkimusongelmiin vastauksen saamiseksi. Etsi ennen luokittelua muuttujan luokkien katkaisupisteet frekvenssijakauman avulla tai anna koneen määrittellä ne (3 equal groups tai joku muu katkaisu). Katso Frekvenssijakaumat -kohdassa olevaa kuvaa.
- Luokittele löytämäsi muuttuja (muuttujat) toiminnolla **Transform – Recode into Different Variables**, joka siis luo uuden luokitellun muuttujan matriisiin ja jättää myös alkuperäisen luokittelemattoman muuttujan matriisiin. Kannattaa yleensä jättää alkuperäinen muuttuja talteen. Valitse avautuvasta ikkunasta haluamasi muuttuja, joka valinnan jälkeen siirtyy viereiseen määritysikkunaan. **Anna uudelle muuttujalle nimi ja kuvaus** *Output Variable: Name, Label*.



- Ilmoita luokkarajat ja uudet arvot painikkeella **Old and New Values**.
- Avautuvaan määritysikkunaan määrittelet **Old Values**-osastossa vanhat arvot, joille annetaan **New Value**-osastossa uusi arvo. Lisää saatu uusi arvo **Old → New**-listaan **Add**-painikkeella. Toista Old–New-työvaiheet vaiheet niin monta kertaa kuin on kulloinkin tarve. Tarkista, että yksikään alkuperäinen muuttujan arvo ei jää luokittelun ulkopuolelle.

**HUOM.** Käytä Range-painikkeita, kun haluat luokitella muuttujia tiettyjen vaihteluvälien avulla. Kuvassa on käytetty kyseisiä painikkeita.



- Nimeä lopuksi luokitellun muuttujan uudet luokat *Variable View*-ikkunassa.

## MUUTTUJAN PISTEYTYKSEN KÄÄNTÄMINEN

Varsin usein pitää joku muuttujista kääntää eli sen pistemäärä muuttua päinvastaiseksi yhteenlaskun tai vertailun vuoksi. Siis jos esim. yksi väittämä on negatiivinen ja muut positiivisia. Toiminto tehdään samalla lailla kuin luokittelu tehtiin yllä, mutta nyt uudet muuttujan arvot ovat päinvastaisia alkuperäisille. (esimerkiksi  $5 \rightarrow 1$ ,  $4 \rightarrow 2$ ,  $3 \rightarrow 3$ ,  $2 \rightarrow 4$ ,  $1 \rightarrow 5$ )

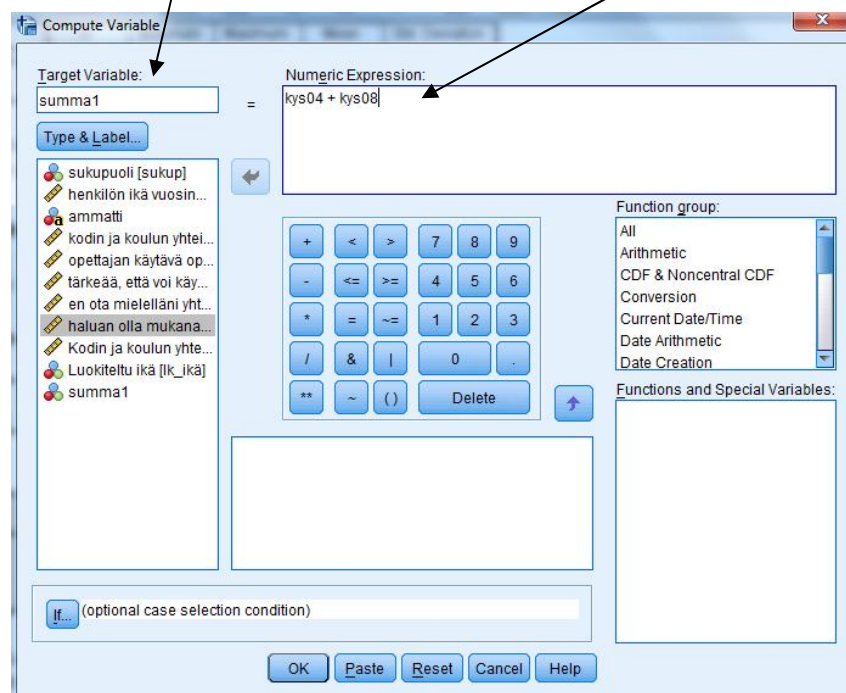
## 5. KAAVOJEN AVULLA MÄÄRITELTÄVÄT MUUTTUJAT

### Summamuuttuja

Laskemalla yhteen samaa asiakokonaisuutta mittaavia muuttujia saadaan se etu, että muuttujien lukumäärä vähenee, mutta ennen kaikkea se, että muodostuneiden summamuuttujien luotettavuus (reliabiliteetti) on parempi kuin yksittäisten muuttujien luotettavuus. Mittarin pidentäminen siis parantaa sen reliabiliteettia. Useampi muuttuja kykenee kattamaan ilmiöstä laajemman osan kuin yksittäinen muuttuja.

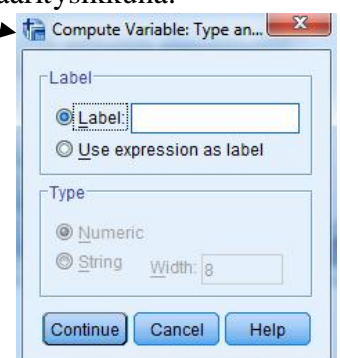
### TOIMENPITEET

1. Valitse **Transform** -valikosta **Compute Variable**.
2. Siirrä nuolipainiketta näpäyttämällä summaan tulevat muuttujat **Numeric Expression** -laatikkoon yksitellen ja laita +-merkki muuttujien väliin.
3. Anna summamuuttujalle nimi **Target Variable** laatikossa.



3. Näpäytä **Type & Label** -painiketta. Tällöin avautuu vieressä oleva määritysikkuna.

4. Näpäytä **Use expression as label**, jolloin summamuuttujan selitteeksi (label) tulee antamasi laskukaava. Ohjelma ei tallenna laskukaavaa mihinkään muualle! Jos et halua laskukaavaa selitteeksi, näpäytä Label-buttonia ja kirjoita haluamasi nimi summamuuttujalle. Hyväksy valinnat näpäyttämällä **Continue** ja **Compute Variable** -ikkunan **OK**-painiketta. Muuttuja ilmestyy matriisiin viimeiseksi muuttujaksi.

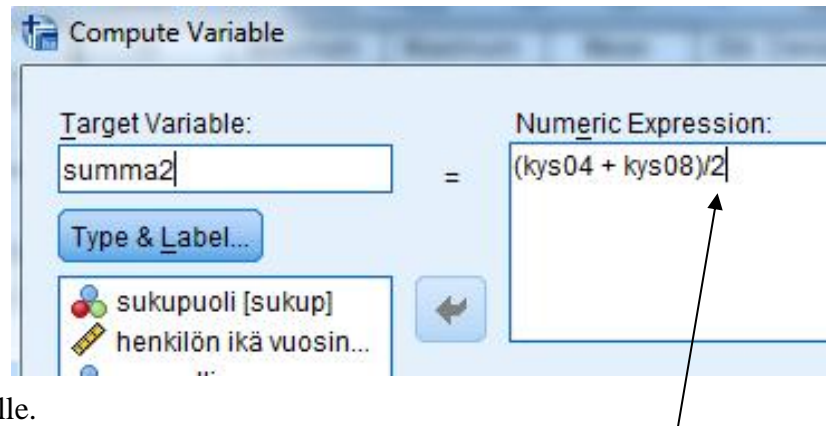


## Keskiarvosummamuuttuja

Keskiarvosummamuuttujissa on se etu, että niitä voi vertailla keskenään riippumatta siinä olevien muuttujien lukumäärästä. Työvaiheet ovat osin samoja kuin ”tavallisessa” summamuuttujassa.

### TOIMENPITEET

1. Valitse **Transform** – valikosta **Compute Variable**.
2. Siirrä nuolipainiketta näpäyttämällä summaan tulevat muuttujat **Numeric Expression** -laatikkoon yksitellen ja laita +-merkki muuttujien väliin.
3. Kun olet syöttänyt kaikki muuttujat kaavaan, laita sulkumerkit lausekkeen ympärille.
4. Laita /-merkki (jakolasku) lausekkeen loppuun ja kirjoita sen perään muuttujien lukumäärä.
5. Tee summamuuttujan nimeäminen ja valintojen hyväksyminen samalla periaatteella kuten edellä ”Summamuuttujat”-kohdassa.



## Logaritmimuutokset

Muuttujalle voi tehdä logaritmimuutoksen esim. silloin, kun sen jakauma on hyvin vino. Tällöin sen jakauma saadaan todennäköisesti lähemmäs normaalijakaumaa kuin alkuperäinen. Luvun logaritmi on eksponentti, johon kantaluku on korotettava, jotta saataisiin kyseinen luku. Eli:  $k^x = y \leftrightarrow \log_k y = x$  SPSS-ohjelmassa on käytössä esimerkiksi

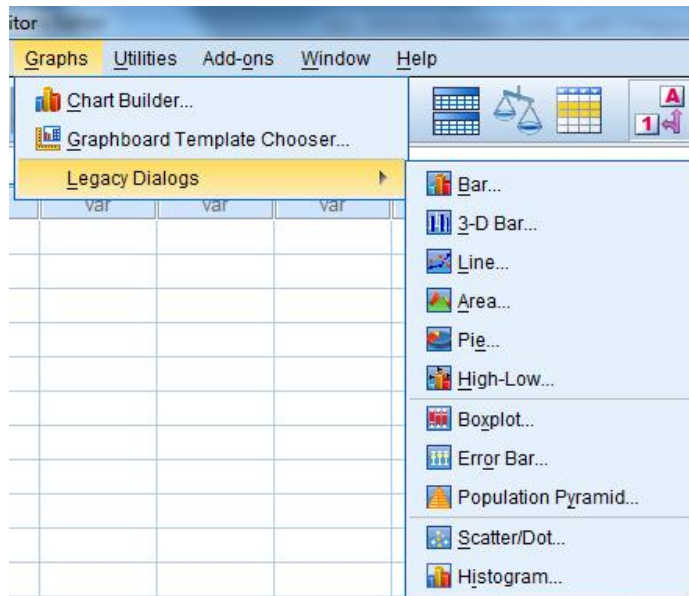
- 10-kantaiset logaritmit (kantalukuna 10, yllä olevassa kaavassa k), kaavan lyhenne **Lg10**
- luonnolliset logaritmit (kantalukuna Neperin luku 2,72, yllä olevassa kaavassa k), kaavan lyhenne **Ln**

### TOIMENPITEET

1. Valitse **Transform** -valikosta **Compute Variable**.
2. Valitse **Function group** -listasta **Arithmetic** ja alla olevaa **Functions and Special Variables** -listaa vierittämällä kaava **Lg10** tai **Ln**.
3. Näpäytä ylöspäin osoittavaa nuolta laatikoiden vieressä, jolloin kaava siirtyy **Numeric Expression**-laatikkoon.
4. Siirrä muuttujalistan vieressä olevaa nuolipainiketta näpäyttämällä kaavaan tuleva muuttuja **Numeric Expression** -laatikossa olevan **Lg10**- tai **Ln**-funktion sulkujen sisälle.
5. Tee muuttujan nimeäminen ja valintojen hyväksyminen samalla periaatteella kuten edellä ”Summamuuttujat”-kohdassa.

## 6. KUVIOIDEN LAATIMINEN

1. Suurin osa analyyseja ja tuloksia havainnollistavista kuvioista saadaan **Graphs**-valikon vaihtoehtoista valitsemalla (myös joidenkin analyysien yhteydessä kuviota saa pyydettyä tai automaattisesti).
2. Kuvioiden laatimiseen on monta päävaihtoehtoa ja kukin niistä pitää sisällään lukuisia alavaihtoehtoja. Halutun kuvion laatiminen on parhaita tehdä kokeilemalla rohkeasti erilaisia vaihtoehtoja, kunnes löytää haluamansa.
3. **Päävaihtoehdot:**



**Chart Builder** = Käyttäjää neuvova käyttöliittymä, joka esittää havainnollisesti millainen kuviosta on tulossa ja mitä määrittämiä vielä tarvitaan. Tämä toiminto vaatii, että muuttujien mitta-asteikot ja kategoristen muuttujien arvojen selitteet (value labels) on merkitty oikein ja tarkoituksenmukaisesti matriisiin (Data Editor) muuttujien määrittelyikkunassa (Variable View, ks. luku 1 tässä ohjeessa). Moneen tilanteeseen sopiva vaihtoehto.

**Graphboard Template Chooser** = Käyttöliittymä, joka tarjoaa kullekin muuttujalle/muuttujille sopivia valmiita kuviomalleja (template), joista käyttäjä voi valita haluamansa. On monessa tilanteessa helpoin tapa kuvioiden laadintaan.

**Legacy Dialogs** = SPSS:n aiemmista versioista tutut perinteiset valintaikkunat (dialogit), joiden avulla kuvioiden laadinta käy myös varsin sujuvasti (varsinkin jos on tottunut tähän tapaan aiemmissa versioissa). Vaatii käyttäjältä eniten omia valintoja verrattuna muihin Graphs-valikon päävaihtoehtoihin, mutta on tottuneelle käyttäjälle varteenotettava toimintatapa.

### **Yleisimmät kuviotyypit:**

**Bar** = pylväsdiagrammi

**Pie** = ympyrädiagrammi

**Boxplot** = laatikko-jana-diagrammi

**Scatter** = korrelaatiodiagrammi

**Line** = viivadiagrammi

**Area** = aluediagrammi

**High-Low** = muuttujan ylä- ja alarajat tai muu haluttu ääriarvo

**Error bar** = keskihajonnan tai keskivirheen kuvaus

**Histogram** = histogrammi

4. Kun kuvio on saatu valmiiksi, voit vielä muokata sen ominaisuuksia kaksoisnäpäyttämällä kuviota **Output**-ikkunassa. Näpäytyksen jälkeen siirryt **Chart Editor** -ikkunaan, jossa on erilaisia muokkaamiseen tarkoitettuja työkaluja. Kun olet saanut muokkaamisen valmiiksi, sulje **Chart Editor** -ikkuna valitsemalla **File**-valikosta **Close** (ei tarvitse erikseen tallentaa). Siirryt tällöin takaisin **Output**-ikkunaan, josta kuvio voidaan tallentaa, tulostaa tai siirtää esimerkiksi tekstinkäsittelyohjelmaan. **HUOM.** Jos teit kuvion **Graphboard Template Chooser** -työkalulla, kaksoisnäpäytys vie **Graphboard Editor** -ikkunaan, jossa kuvion muokkaus tehdään.

## 7. RISTIINTAULUKOINTI KAKSIULOTTEISEN JAKAUMAN KUVAAJANA

### KÄYTTÖTARKOITUS

Muuttujien ristiintaulukointia voidaan käyttää kaksiulotteisen jakauman (kahden muuttujan yhteisen jakauman) kuvaajana. Yleensä sitä käytetään tähän tarkoitukseen, kun halutaan tarkastella muuttujien välisiä yhteyksiä, mutta jompikumpi (tai molemmat) muuttuja on diskreetti eli epäjatkuva, nominaali- eli laatueroasteikollinen, vain muutamia luokkia sisältävä tai dikotominen (vain kaksi arvoa, esim. sukupuoli).

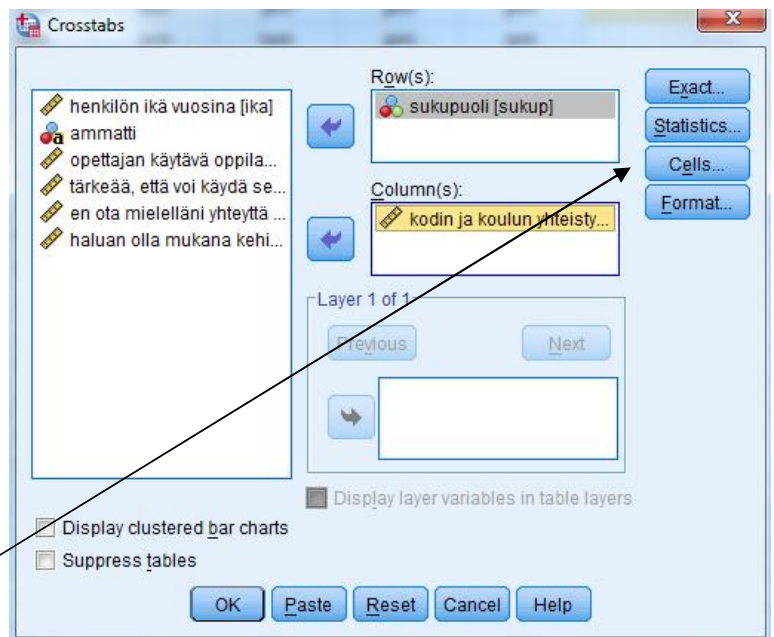
### TOIMENPITEET

1. **Analyze**-valikosta **Descriptive Statistics – Crosstabs**.

2. Valitse toinen muuttuja **riveille** (Rows) ja toinen **sarakkeille** (Columns).

Riveille kannattaa usein valita vähemmän luokkia sisältävä muuttuja ja/tai selittävä muuttuja (jos sellainen voidaan nimetä).

Muuttuja ei yleensä sovi ristiintaulukointiin, jos se saa paljon arvoja laajalla vaihteluvälillä. Tällöin taulukosta tulee liian monisoluinen tulkinnan kannalta. Jos silti haluaa käyttää ristiintaulukointia, voi moniluokkaisen muuttujan luokitella esim. 2–5-luokkaiseksi.



3. Paina **Cells**-painiketta ja valitse avautuvasta ikkunasta ne luvut, jotka haluat tulostaa taulukon soluihin.

4. Kokeile etenkin erilaisia prosenttilukuja eli **Percentages**.

#### Counts

**Observed** = havaitut frekvenssit

**Expected** = odotetut frekvenssit

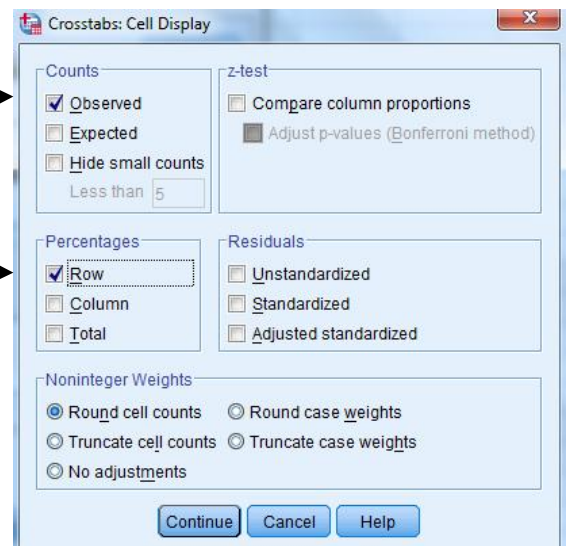
#### Percentages

**Row** = riviprosentit

**Column** = sarakeprosentit

**Total** = kokonaisprosentit

5. Hyväksy valinnat **Crosstabs**-ikkunan **OK**-painikkeella. Analysoi tulosta.



**ESIMERKKITULOSTUS (RIVIPROSENTTIEN KANSSA):**

Mukaan tulostuva Case Processing Summary poistettu  
(ilmaisee kuinka monta tapausta kyseisessä analyysissä on mukana)

**sukupuoli \* Kodin ja koulun yhteistyö on tärkeää (luokiteltu) Crosstabulation**

			Kodin ja koulun yhteistyö on tärkeää (luokiteltu)		Total
			Eri mieltä	Samaa mieltä	
sukupuoli	nainen	Count	7	23	30
		% within sukupuoli	23,3%	76,7%	100,0%
	mies	Count	8	22	30
		% within sukupuoli	26,7%	73,3%	100,0%
Total		Count	15	45	60
		% within sukupuoli	25,0%	75,0%	100,0%

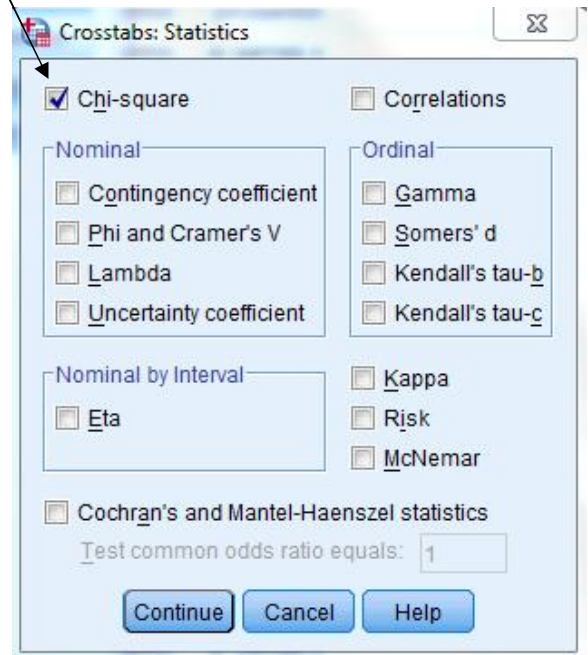
## 8. RISTIINTAULUKOINTIIN LIITTYVÄT TILASTOLLISET TESTIT

### KÄYTTÖTARKOITUS

- Katso edellä luku 7 ”Ristiintaulukointi kaksiulotteisen jakauman kuvaajana”. Usein varsinkin otantatutkimusten yhteydessä ristiintaulukoinnissa ei tyydytä pelkkään kuvaamiseen, vaan halutaan tilastollisen testaamisen avulla selvittää, miten todennäköinen saatu jakauma on perusjoukossa: onko muuttujien välinen yhteys niin selvä ja voimakas, että voimme väittää näin olevan myös perusjoukossa.
- Testaamiseen ryhdyttäessä jokaiseen taulukon soluun sijoitetaan havaitut frekvenssit (Observed Count) ja lasketaan odotetut (Expected Count) frekvenssit.
- Jos suoritettussa ristiintaulukoinnissa havaitaan, että frekvenssit jakautuvat odotusarvojen mukaisesti, ei muuttujien välillä ole yhteyttä. Jos taas arvot jakautuvat lineaarisesti tai kasautuvat, voidaan olettaa muuttujien välillä olevan yhteyttä. Silmämääräistä tarkastelua auttaa esimerkiksi khiin neliötesti ( $\chi^2$ ), joka testaa onko havaittujen ja odotettujen frekvenssien välillä tilastollisesti merkitsevää eroa.
- Jatkuvien, moniluokkaisten (vähintään järjestysasteikollisten) muuttujien välisten yhteyksien tutkimiseen käytetään mieluiten tulomomenttikorrelaatiokerrointa ( $r_{xy}$ ), jos riippuvuus voidaan olettaa tai todeta suoraviivaiseksi.
- Khiin neliötestin ehtona on, että korkeintaan 20 prosenttia odotetuista frekvensseistä saa olla alle 5 ja jokaisen odotetun frekvenssin pitää olla yli 1.**

## TOIMENPITEET: Khiin neliö -testi

1. Katso ristiintaulukointiin liittyviä ohjeita luvusta 7 ”Ristiintaulukointi kaksiuotteisen jakauman kuvaajana” tai aloita suoraan ristiintaulukointi valitsemalla **Analyze**-valikosta **Descriptive Statistics – Crosstabs**.
2. Valitse muuttujat riveille ja sarakkeille.
3. Valitse haluamasi solufrekvenssit ja prosentit **Cells**-painikkeen takaa avautuvasta ikkunasta.
4. Näpätä **Statistics**-painiketta ja valitse **Chi-square** (khiin neliö).



5. Hyväksy valinnat **Crosstabs**-ikkunan **OK**-painikkeella. Analysoi tulosta.

## ESIMERKKITULOSTUS

Ristiintaulukointi perusmuodossaan, mukana riviprozentit

luokiteltu ikä \* Teen mieluummin yhden asian kerrallaan kuin monta yhtä aikaa. Crosstabulation

			Teen mieluummin yhden asian kerrallaan kuin monta yhtä aikaa.		Total
			eri mieltä	samaa mieltä	
luokiteltu ikä	nuorin ryhmä	Count	34	46	80
		% within luokiteltu ikä	42,5%	57,5%	100,0%
	keskiryhmä	Count	25	53	78
		% within luokiteltu ikä	32,1%	67,9%	100,0%
	vanhin ryhmä	Count	17	12	29
		% within luokiteltu ikä	58,6%	41,4%	100,0%
Total		Count	76	111	187
		% within luokiteltu ikä	40,6%	59,4%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6,386 <sup>a</sup>	2	,041
Likelihood Ratio	6,363	2	,042
Linear-by-Linear Association	,602	1	,438
N of Valid Cases	187		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 11,79.

**TARKISTA! Suluissa oleva prosenttiluku ei saa olla yli 20 %.**  
**Jos on, pitää toinen tai molemmat muuttujat luokitella uudestaan tai luopua khiin neliön käytöstä (tulkinasta).**  
**Prosenttiluku ilmaisee kuinka monessa solussa odotettu frekvenssi on alle 5. Se tarkoittaa, että taulukossa on soluja(riivejä tai sarakkeita) liikaa jommankumman muuttujan jakauman muodon (vino) tai aineiston pienen koon vuoksi.**

Khiin neliön arvo (Value), vapausasteet (df) ja sen merkitsevyys (Sig.).

Tässä tulos on tilastollisesti merkitsevä, koska p-arvo (Sig) on **pienempi** kuin .05.

Johtopäätös on, että ikäluokat eroavat kysytyssä mielipiteessä. Frekvenssien ja prosenttilukujen avulla voidaan kuvailla tulosta tarkemmin.

### Muut ristiintaulukointiin liittyvät testit

Kun SPSS-ohjelmaa pyytää tekemään khiin neliötestin, se tekee samalla muitakin testejä, vaikka käyttäjä ei sitä pyydäkään. Esimerkkejä:

**Likelihood Ratio** = uskottavuussuhteen testi, käytetään log-lineaarille malleille

**Linear-by-Linear Assosiation** = mittaa muuttujien lineaarista riippuvuutta

**Continuity Correction** = khiin neliö ns. jatkuvuuskorjauksen kanssa, lasketaan vain 2x2- taulukoille

**Fisher's Exact Test** = Fisherin uskottavuustesti, jota käytetään vain 2x2- taulukoissa. Voidaan tulkita, vaikka mukana olisi alle viiden suuruisia odotettuja frekvenssejä. Tätä voidaan siis käyttää myös vinoissa jakaumissa ja pienillä aineistoilla.

## 9. KORRELAATIO

### KÄYTTÖTARKOITUS

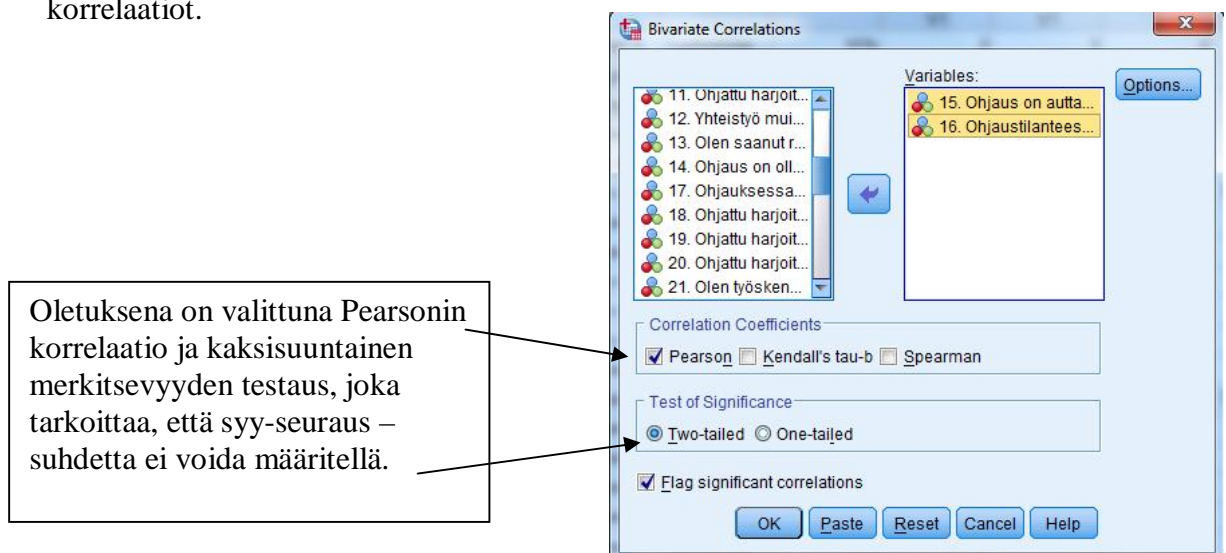
- Korrelaatiokerroin (lyhenne  $r_{xy}$ ) ilmoittaa kahden muuttujan välisen yhteyden voimakkuuden ja suunnan. Se voi vaihdella välillä  $[-1,1]$ .
- Negatiivinen korrelaatio ilmaisee sitä, että toisen muuttujan arvojen kasvaessa toisen arvot pienyvät. Positiivinen korrelaatio taas ilmaisee sitä, että toisen muuttujan arvojen kasvaessa myös toisen arvot kasvavat.
- Se ilmaisee **lineaarisen** yhteyden voimakkuuden kahden muuttujan välillä, se **ei** siis ilmaise **käyräviivaisen** yhteyden voimakkuutta.
- Korrelaatio ilmaisee muuttujien yhteistä vaihtelua, mutta se ei ole absoluuttisen samankaltaisuuden mitta (toinen muuttuja voi saada arvoja väliltä 1–10, toinen 100–500).
- Korrelaatiokertoimen neliö ilmaisee, kuinka suuri osuus muuttujien varianssista (vaihtelusta) on yhteistä, selitettävissä toisen avulla, esim.  $r = .60$ , joten  $r^2 = .36$  eli 36% yhteistä vaihtelua. SPSS-laskee korrelaatiokertoimen neliön automaattisesti mm. regressioanalyysin yhteydessä, mutta ei korrelaatioanalyysissä (pitää laskea ”käsillä”).
- Korrelaatiokertoimen tulkinta:
  - yhteyden voimakkuuden ja suunnan tarkastelu
  - merkitsevyyden tarkastelu
  - kertoimen neliön tarkastelu
- Korrelaation eli  $r_{xy}$ :n arvo riippuu mm. mittauksien reliabiliteetista, populaatioista (*niiden homogeenisuus madaltaa saatuja riippuvuuksia*) ja valitusta otoksesta.

### Pearsonin tulomomenttikorrelaatiokerroin

Pearsonin tulomomenttikorrelaatiokerroin soveltuu muuttujien välisen yhteyden tarkasteluun, jos molemmat muuttujat ovat vähintään intervallasteikollisia ja niiden jakauma noudattaa normaalijakaumaa (mieluiten – ainakin likimäärin).

### TOIMENPITEET

- Valitse **Analyze**-valikosta **Correlate – Bivariate**.
- Valitse muuttujat **Variables**-laatikkoon. Voit valita samalla kertaa useita muuttujia, jolloin saat ns. korrelaatiomatriisin. Korrelaatiomatriisista näet kaikkien valittujen muuttujien väliset korrelaatiot.



## ESIMERKKITULOSTUS

		15. Ohjaus on auttanut minua kehittymään opettajana	16. Ohjaustilanteessa olen voinut tuoda esille omia näkemyksiäni
15. Ohjaus on auttanut minua kehittymään opettajana	Pearson Correlation	1	,624**
	Sig. (2-tailed)		,000
	N	1674	1674
16. Ohjaustilanteessa olen voinut tuoda esille omia näkemyksiäni	Pearson Correlation	,624**	1
	Sig. (2-tailed)	,000	
	N	1674	1674

korrelaatiokerroin

p-arvo

tapausten lukumäärä

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

Oletusmuodossaan SPSS tulostaa kahdelle muuttujalle neljäsoluisen taulukon, josta vain yhtä tarvitsee tulkita. Tässä korrelaatiokerroin on ,624, joka tarkoittaa positiivista riippuvuutta muuttujien välillä (kun toisen muuttujan arvot kasvavat, myös toisen kasvavat). Yhteys on myös merkitsevä (Sig = p = ,000). Korrelaatiokertoimen voidaan siis väittää poikkeavan nolasta ilmoitetulla erehtymisriskillä. Nolasta poikkeaminen tarkoittaa, että jotain yhteyttä on olemassa. Alimpana solussa on mukana olleiden tapausten määrä (= N).

Korrelaatiokertoimen voi siis korottaa toiseen potenssiin eli kertoa itsellään, jolloin saada selville kuinka suuri osuus muuttujien varianssista (vaihtelusta) on yhteistä, selitettävissä toisen avulla.

Tässä esimerkissä  $r^2 = 0,624 * 0,624 = 0,389 \rightarrow 0,389 * 100 \% = 38,9 \%$

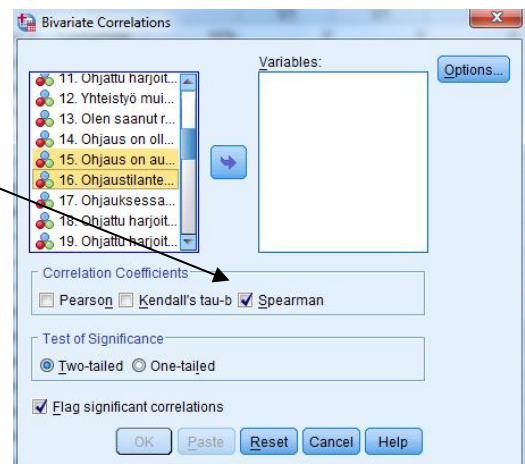
Esimerkissä on kaksi muuttujaa, jotka liittyvät opetusharjoittelun ohjaukseen ja niiden sisältöä tarkastelemalla voitiin odottaakin, että ne ehkä korreloivat – varmistus tuli kuitenkin tässä. Korrelaatiokertoimen tulkinnassa pitää kuitenkin olla tarkkana, koska asiat voivat korreloida, vaikka ne asiallisesti ottaen eivät liity toisiinsa mitenkään. Toisena vaarana ovat itsestäänselvytykset, jolloin saadaan korkeita korrelaatiokertoimia muuttujien välille, koska ne ovat mitanneet lähes samaa asiaa tai ovat muuten riippuvuussuhteessa (joka ei ole tutkimuksellisesti kiinnostava).

### Spearmanin korrelaatiokerroin

Spearmanin korrelaatio soveltuu muuttujien välisen yhteyden tarkasteluun, jos molemmat muuttujat ovat vähintään järjestysasteikollisia. Muuttujien jakaumille ei aseteta oletuksia, joten se sopii myös vinoille jakaumille (joille Pearson ei oikein sovi).

Se tehdään samalla tavalla kuin Pearsonin korrelaatio, mutta **Bivariate Correlations** -ikkunassa laitetaan rasti kohtaan **Spearman**.

Spearmanin korrelaatio tulkitaan samalla tavalla kuin Pearsonin korrelaatio (ks. ohje yllä).



## 10. KAHDEN KESKIVERTAILU: t-testit

t-testillä testataan kahden keskiarvon eron tilastollista merkitsevyyttä jollakin muuttujalla eli se testaa, onko ryhmien (tai muuttujien) välille saatu ero niin suuri, että voidaan olettaa eroa olevan myös perusjoukossa vai niin pieni, että se johtuu sattumasta.

t-testejä on kolmenlaisia: (1) yhden otoksen t- testi, (2) riippumattomien tapausten t-testi ja (3) riippuvien tapausten t-testi. Seuraavassa esitellään tarkemmin kukin niistä.

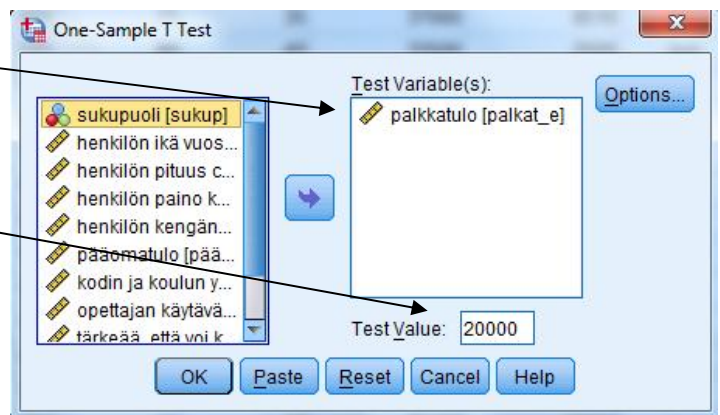
### Yhden otoksen t-testi (One-Sample T Test)

#### KÄYTTÖTARKOITUS

Yhden otoksen t-testiä käytetään silloin, kun halutaan verrata jonkun vähintään intervalliasteikollisen muuttujan otoskeskiarvoa (siis koko otoksen) perusjoukon keskiarvoon. Ennen testin suorittamista pitää siis olla selvillä perusjoukon keskiarvo. Se on harvoin tiedossa ja tästä syystä tämä on t-testeistä harvinaisin.

#### TOIMENPITEET

1. Valitse **Analyze**-valikosta **Compare Means – One-Sample T Test**.
2. Sijoita testimuuttuja **Test Variable(s)**-laatikkoon.
3. Kirjoita perusjoukon keskiarvo **Test Value**-laatikkoon (tässä 20000 = perusjoukon palkkatulo).
4. Hyväksy valinnat **OK**-painikkeella.



#### ESIMERKKITULOSTUS

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
palkkatulo	60	21079,72	8277,655	1068,641

Ylempään taulukkoon tulostuvat muuttujan keskiarvo, keskihajonta ja keskiarvon keskivirhe.

One-Sample Test

	Test Value = 20000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
palkkatulo	1,010	59	,316	1079,722	-1058,62	3218,07

#### PÄÄTULOS

Jos Sig eli  $p > .05$ , otoksen keskiarvo ei poikkea merkitsevästi perusjoukon keskiarvosta. Tässä  $p = .316$ , joten otoksen keskiarvo ei poikkea tilastollisesti merkitsevästi perusjoukon keskiarvosta.

## Riippumattomien tapausten t-testi (Independent-Samples T Test)

### KÄYTTÖTARKOITUS

Tämä on t-testeistä yleisin. Riippumattomat tapaukset tarkoittavat sitä, että on mitattu kahta ERI ryhmää, esimerkiksi kahta ihmisryhmää jossakin ominaisuudessa. Testiä varten pitää valita ryhmittelevä muuttuja (Grouping Variable), jossa on kaksi ryhmää (saa kaksi arvoa) ja valita testimuuttuja (Test Variable) eli riippuva muuttuja (ominaisuus, jossa vertaillaan). Riippuvan muuttujan pitää olla vähintään intervalliasteikollinen ja sen pitäisi (mielummin) noudattaa normaalijakaamaa.

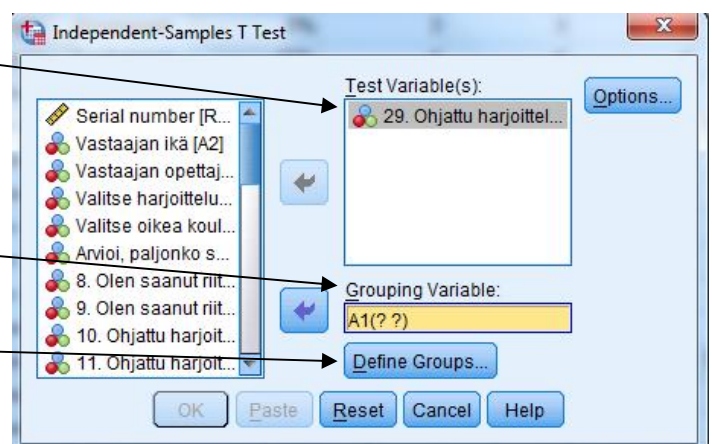
### TOIMENPITEET

1. Valitse **Analyze** –valikosta **Compare Means – Independent-Samples T Test**.

2. Sijoita testimuuttuja **Test Variable** -kohtaan. Voit valita useita testimuuttujia yhdellä kertaa.

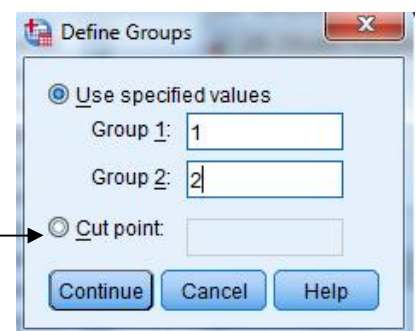
3. Valitse ryhmittelevä muuttuja **Grouping Variable** kohtaan.

4. Määrittele ryhmittelevän muuttujan luokat **Define Groups** -painikkeella.



5. Kirjoita **Group 1** -laatikkoon ensimmäisen ryhmän numerotunnus (muuttujan arvo, esim. 1) ja **Group 2** -laatikkoon toisen ryhmän numerotunnus (muuttujan arvo, esim. 2).

6. Halutessasi voit käyttää ryhmittelijänä jotakin jatkuvaa muuttujaa katkaisemalla sen jakauman, jolloin saat muodostettua kaksi ryhmää. Kirjoita **Cut point** -laatikkoon jakauman katkaisupiste.



7. Hyväksy valinnat **Continue** ja pääikkunan **OK**-painiketta näpäyttämällä.

## ESIMERKKITULOSTUS

Ensin tulosteeseen tulee taulukko, joka kertoo ryhmien ominaisuudet kyseisessä muuttujassa.

	Sukupuoli	N	Mean	Std. Deviation	Std. Error Mean
29. Ohjattu harjoittelu on avartanut näkemystäni opettajan työn moninaisuudesta	Nainen	1258	5,65	1,330	,037
	Mies	416	5,28	1,489	,073

↑                      ↑                      ↑                      ↑                      ↑                      ↑  
 Muuttujan          ryhmien          ryhmien          ryhmien          ryhmien          keskiarvojen  
 nimi                  nimet              koko              keskiarvot      keskihajonnat      keskivirheet

Tarkista alemmasta tulosteesta ensin, onko **Levenen F**-testin tulos tilastollisesti merkitsevä eli eroavatko ryhmien keskihajonnat (variانسit) toisistaan tilastollisesti merkitsevästi ( $p < .05$ ). Jos keskihajonnat (variانسit) eroavat merkitsevästi, t-testissä käytetään toista laskutapaa. Tosin, jos ryhmät ovat jokseenkin yhtä suuret kooltaan, ei vertailtavien ryhmien variانسien merkitsevä ero ole tärkeä, mutta silloinkin kannattaa katsoa tulos alariviltä Levenen F:n ollessa merkitsevä.

jos Levenen F-testin p-arvo (sig) **ei** ole merkitsevä → tulkitse t-testin tulos **yläriviltä**  
 jos Levenen F-testin p-arvo (sig) **on** merkitsevä → tulkitse t-testin tulos **alariviltä**

	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	95% Confidence Interval of the Difference	
29. Ohjattu harjoittelu on avartanut näkemystäni opettajan työn moninaisuudesta	Equal variances assumed	7,197	4,789	1672	,000	,371	,078	,219	,523
	Equal variances not assumed		4,525	647,945	,000	,371	,082	,210	,533

**Tässä tulkitaan alariviä!**

Levenen F-testi ja sen p-arvo

t-testisuure ja sen vapausasteet (df)

Sig = p-arvo, = t-arvon merkitsevyys  
  
**PÄÄ-TULOS!**  
 Tässä on merkitsevä ero keskiarvojen välillä ( $p < .05$ )

ryhmien keskiarvojen erotus ja sen keskivirhe

Luottamusväli, jolla ryhmien keskiarvojen erotus on 95 prosentin todennäköisyydellä. Jos nolla sisältyy väliin, ryhmien välillä ei ole merkitsevää eroa. Tässä nollaa ei ole välissä (sama näkyy päätuloksesta eli p-arvosta).

## Riippuvien tapausten t-testi (paired samples t-test)

### KÄYTTÖTARKOITUS

Riippuvat tapaukset tarkoittavat sitä, että on mitattu **SAMOJA** tapauksia, esimerkiksi ihmisryhmiä kaksi eri kertaa jossakin ominaisuudessa. Testiä varten pitää valita muuttujapari (Paired Variables), esimerkiksi muuttujapari alkutesti–lopputesti. Muuttujien pitää olla vähintään intervalliasteikollisia ja niiden pitäisi (mieluummin) noudattaa normaalijakaumaa.

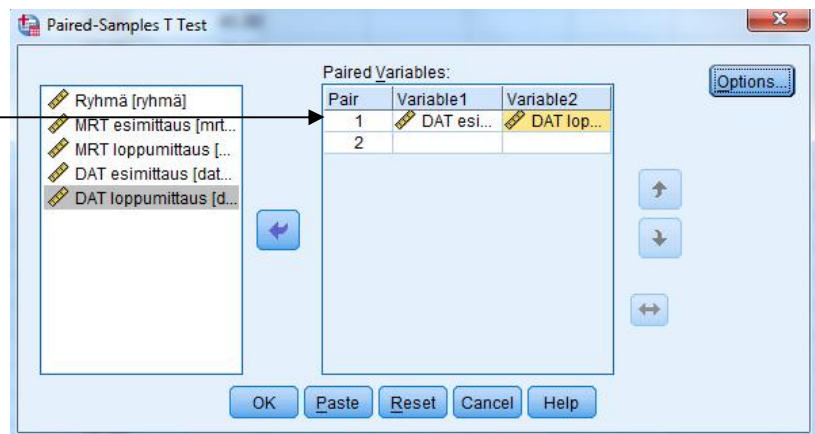
### TOIMENPITEET

1. Valitse **Analyze**-valikosta **Compare Means – Paired-Samples T Test**.

2. Valitse muuttujat hiiren avulla avulla näpäyttämällä. Valitut muuttujat ilmestyvät **Paired Variables** -kohtaan.

3. Voit laatia useita pareja samalla tekniikalla samaan analyysiin.

4. Hyväksy valinnat **OK**-painiketta näpäyttämällä.



### ESIMERKKITULOSTUS

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	DAT esimittaus	42,2281	57	9,79799	1,29778
	DAT loppumittaus	47,0702	57	8,68879	1,15086

Ylimpään taulukkoon tulostuvat muuttujien keskiarvot, keskihajonnat ja keskiarvon keskivirheet.

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	DAT esimittaus & DAT loppumittaus	57	,857	,000

Toiseen taulukkoon tulostuu muuttujien välinen korrelaatio ja sen merkitsevyys. Tässä korrelaatio on varsin korkea ja tilastollisesti merkitsevä.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	DAT esimittaus - DAT loppumittaus	-4,84211	5,05254	,66923	-6,18273	-3,50149	-7,235	56	,000

Alimpaan taulukkoon tulostuu varsinainen testitulos ja sen merkitsevyys. Jos saatu Sig-arvo (= p-arvo) on pienempi kuin .05, on muuttujien keskiarvojen välinen erotus tilastollisesti merkitsevä. Tässä esimerkissä loppumittauksen keskiarvo on tilastollisesti merkitsevästi suurempi kuin esimittauksen keskiarvo.

# 11. YKSISUUNTAINEN VARIANSSIANALYYSI

## KÄYTTÖTARKOITUS

Yksisuuntaisen varianssianalyysin käyttötarkoitus ja -periaate on samantapainen kuin riippumattomien tapausten t-testin. Yksisuuntaisessa varianssianalyysissä on kuitenkin vertailtavia ryhmiä **enemmän kuin kaksi** (yleensä 3–5). t-testiä **ei saa** käyttää useamman ryhmän keskinäiseen vertailuun eli ei saa tehdä useita ristikkäisiä t-testejä.

Yksisuuntainen varianssianalyysi testaa ns. Fisherin F-testillä ryhmien sisäisen ja ryhmien välisen vaihtelun suhdetta. F-testin merkitsevyys on varianssianalyysin tulos.

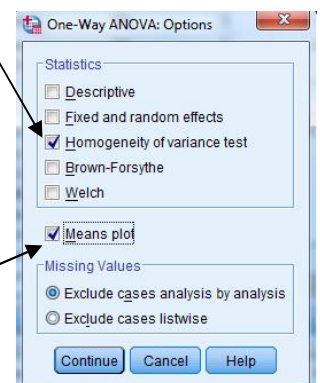
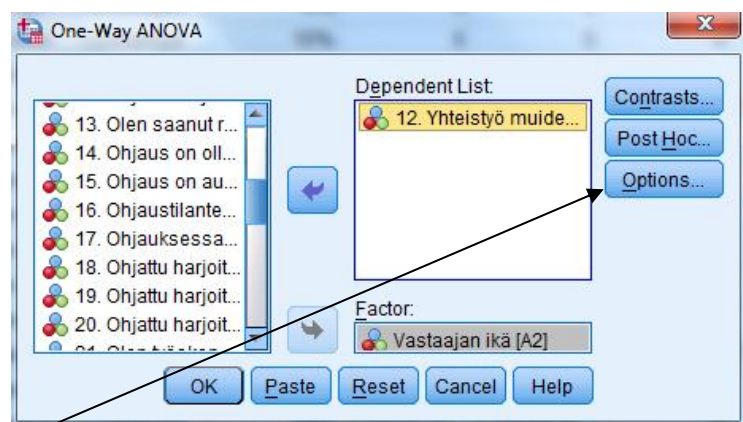
Ennen testiä pitää olla mietittynä:

- Mikä on riippuva muuttuja (Dependent) eli testimuuttuja. Sen pitää olla vähintään intervalliasteikolla mitattu ja noudattaa normaalijakaumaa (mieluummin).
- Mikä on ryhmittelevä muuttuja eli riippumaton muuttuja (Factor), yleensä nominaaliasteikkoinen tai luokiteltu muuttuja.

Jos varianssianalyysi on osoittanut ryhmien välillä olevan eroja, käytetään ryhmäkeskiarvojen keskinäiseen vertailuun Post Hoc -testejä (esim. Bonferroni tai Tamhane's T2). Ryhmävertailuissa huomioidaan yleensä ryhmien varianssien erilaisuus eli tehdään Levenen F-testi, jonka tulos vaikuttaa ryhmävertailutestin valinnassa. Alla olevassa esimerkissä yksisuuntainen varianssianalyysi tehdään kahdessa vaiheessa, mutta sen voi tehdä myös yksivaiheisesti tekemällä jo ensimmäisellä kerralla kaikki valinnat, joista osa saattaa olla turhia, mutta ne voi jättää huomiotta.

## TOIMENPITEET (vaihe 1)

1. Valitse **Analyze**-valikosta **Compare Means – One Way ANOVA**.
2. Valitse avautuvan ikkunan **Dependent List** -kohtaan riippuva muuttuja(t) eli se ominaisuus, jossa ryhmiä vertaillaan.
3. Valitse ryhmittelevä muuttuja (yl. 3–5 luokkainen) **Factor**-kohtaan.
4. Näpätä **Options**-painonappia.
5. Valitse avautuvasta ikkunasta **Levenen F-testi** (Homogeneity-of-variance), joka testaa eroavatko ryhmien keskihajonnat (variانسsit) toisistaan tilastollisesti merkitsevästi ( $p < .05$ ). Jos keskihajonnat (variانسsit) eroavat merkitsevästi, se vaikuttaa jatkossa tehtäviin valintoihin.
6. Samasta **Options**-ikkunasta voit valita myös muuttujien kuvailutiedot (Descriptive), jolloin saat muuttujien kuvailutiedot kullekin vertailtavalle ryhmälle erikseen.
7. Voit myös valita graafisen kuvaajan **Means Plot**, joka havainnollistaa keskiarvojen eroja. Hyväksy valinnat **Continue** ja pääikkunan **OK**-painiketta näpäyttämällä.



## ESIMERKKITULOSTUS (vaihe 1)

### Test of Homogeneity of Variances

12. Yhteistyö muiden opettajaopiskelijoiden kanssa on ollut minulle hyödyllistä harjoittelun aikana

Levene Statistic	df1	df2	Sig.
16,867	3	1670	,000

Levenen F-testin tulos. Tässä Sig < .000 eli ryhmien keskihajonnat (variانسsit) eroavat toisistaan merkitsevästi. Tätä tietoa tarvitaan vain Post Hoc -testin valinnassa, jota taas ei tarvita, jos alla oleva päätulos ei ole merkitsevä.

### ANOVA

12. Yhteistyö muiden opettajaopiskelijoiden kanssa on ollut minulle hyödyllistä harjoittelun aikana

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	44,125	3	14,708	8,341	,000
Within Groups	2945,003	1670	1,763		
Total	2989,128	1673			

Yllä nähdään ryhmien sisäiset ja väliset ns. neliösummat, keskineliösummat, vapausasteet ja F-testin tulos.

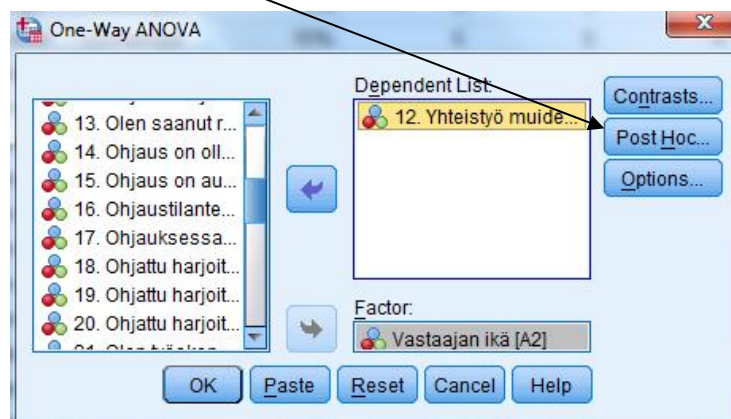
### PÄÄTULOS

Varsinainen varianssianalyysin tulos on ANOVA-taulukon F-testin merkitsevyys. Tässä ryhmien välillä **ON** eroa, koska Sig < .05.

**JOHTOPÄÄTÖS:** ryhmien välillä **ON** eroa mitatussa ominaisuudessa. **MUTTA** emme vielä tiedä **MINKÄ** ryhmien välillä ero on tilastollisesti merkitsevä. Etene kohtaan **TOIMENPITEET** (vaihe 2). **HUOMAA**, että jos ANOVA-taulukon F-testin tulos ei ole merkitsevä, ei ole tarvetta jatkaa vaiheeseen 2. Tällöin todetaan, ettei ryhmien välillä ole eroa, joka sinänsä voi olla tärkeä ja joskus yllättäväkin tulos.

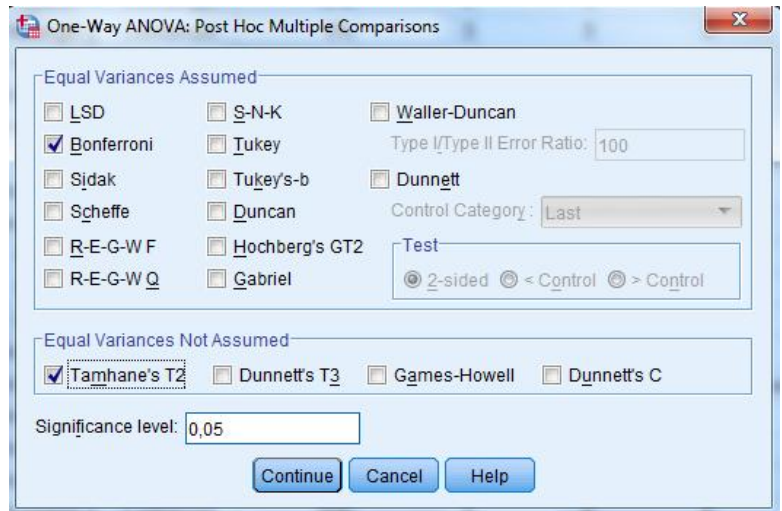
## TOIMENPITEET (vaihe 2)

- Valitse uudelleen **Analyze** -valikosta **Compare Means – One Way ANOVA**.
- Anna äsken tehtyjen valintojen (vaiheessa 1) olla paikallaan kussakin kohdassa.
- Näpäytä **Post Hoc** -painiketta, jolloin saat valittua ryhmien tarkempaan vertailuun soveltuvan testin.



11. Avautuvan ikkunan yläosassa osastossa **Equal Variances Assumed** ovat testit, joita käytetään ryhmävertailuun silloin, kun edellä tehty Levenen F -testin tulos **ei ole** merkitsevä eli ryhmien keskihajonnat (varianssit) **eivät eroa** toisistaan merkitsevästi.

12. Avautuvan ikkunan alaosassa osastossa **Equal Variances Not Assumed** ovat testit, joita käytetään ryhmävertailuun silloin, kun Levenen F **on** merkitsevä eli ryhmien keskihajonnat (varianssit) **eroavat** toisistaan merkitsevästi.



13. Tässä esimerkkikuvassa on valittuna yläosasta testi nimeltään **Bonferroni**, jota tulkittaisiin, jos Levenen F ei olisi merkitsevä. Alaosasta on valittuna Tamhane's T2, jota tulkittaisiin, jos Levene F olisi merkitsevä.

14. Edellä vaiheessa 1 saadun tuloksen perusteella olisi tarpeen vain valinta Tamhane's T2. Päätulos ANOVA-taulukossa oli merkitsevä, joten Post Hoc -testit ovat tarpeen.

15. Tee kuhunkin tilanteeseen sopiva valinta ja hyväksy valinnat **Continue**-painonapilla ja pääikkunan **OK**-painiketta näpäyttämällä.

## ESIMERKKITULOSTUS (vaihe 2)

**Tässä ei toisteta vaiheessa 1 olleita tulosteita, vaan ainoastaan ne, jotka tulivat uutena mukaan Post Hoc-valinnan jälkeen.**

**Multiple Comparisons**

12. Yhteistyö muiden opettajaopiskelijoiden kanssa on ollut minulle hyödyllistä harjoittelun aikana  
Tamhane

(I) Ikä	(J) Ikä	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
20-25	26-30	,003	,074	1,000	-,19	,20
	31-35	,137	,126	,860	-,20	,47
	36+	,573*	,151	,001	,17	,98
26-30	20-25	-,003	,074	1,000	-,20	,19
	31-35	,134	,134	,900	-,22	,49
	36+	,571*	,158	,002	,15	,99
31-35	20-25	-,137	,126	,860	-,47	,20
	26-30	-,134	,134	,900	-,49	,22
	36+	,436	,188	,121	-,06	,94
36+	20-25	-,573*	,151	,001	-,98	-,17
	26-30	-,571*	,158	,002	-,99	-,15
	31-35	-,436	,188	,121	-,94	,06

\*. The mean difference is significant at the 0.05 level.

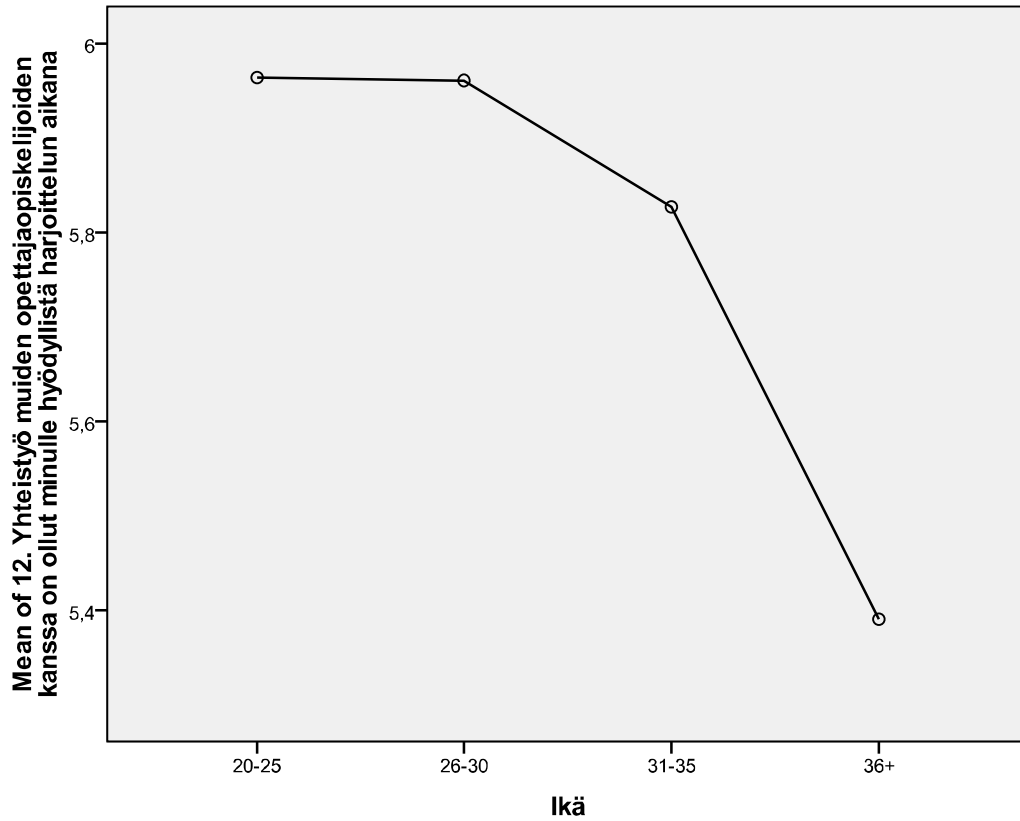
Tässä nähdään kaikkien ryhmien väliset keskinäiset vertailut. Merkitsevät erot löytyvät Sig-sarakkeen perusteella, mutta ne on myös merkitty keskiarvojen erotuksen (Mean Difference) perään tähdellä (\*).

**JOHTOPÄÄTÖS:** Vanhin ikäluokka eroaa 20–25- ja 26–30 -vuotiasta. Suunta on siten, että vanhin ryhmä on kokenut opiskelijoiden välisen yhteistyön vähemmän hyödyllisenä kuin mainitut kaksi muuta ryhmään. Muut erot eivät ole merkitseviä. Tulosta havainnollistaa seuraavalla sivulla näkyvä Means Plot -kuvi.

### Vasemmalta lukien taulukkoon tulostuvat:

- 1) ryhmien tunnukset
- 2) keskiarvojen erotukset
- 3) erotusten keskivirheet
- 4) erotuksen merkitsevyys (Sig) sekä 5) erotuksen luottamusvälin ala- ja yläraja 95:n % todennäköisyydellä.

Alla on edellisellä sivulla esitelty yksisuuntaisen varianssianalyysin tulos havainnollistettuna viivadiagrammilla. Huomaa aina katsoa Y-akselin esitystapa(skaala), koska ohjelma haluaa aina näyttää erot mahdollisimman havainnollisina (hieman ”liioteltuna”), vaikka todellisuudessa erot voivat olla joskus hyvin pieniä. On hieman suhteellista ja asetelmasta riippuvaista, mikä on paljon ja mikä on vähän, mutta merkitsevyyden testaus auttaa tässä (ks. ed. sivu).



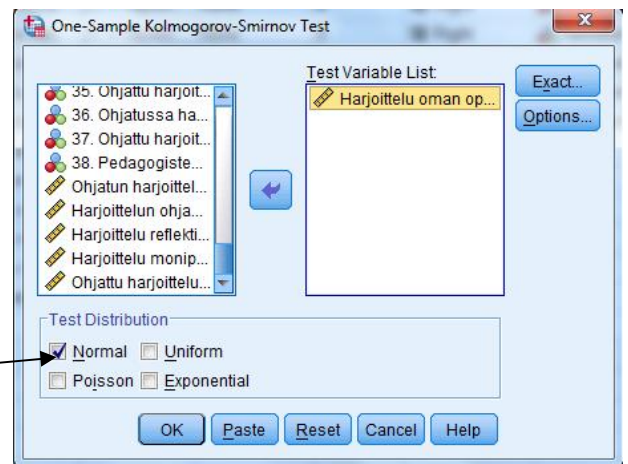
## 12. JAKAUMAN NORMAALISUUDEN TESTAUS: Kolmogorov-Smirnovin testi

### KÄYTTÖTARKOITUS

Monet ns. parametrisistä testeistä sisältävät ehdon, että muuttujien (riippuvien muuttujien) pitäisi mieluummin noudattaa normaalijakaumaa. Esimerkiksi Kolmogorov-Smirnovin testillä voidaan tutkia, onko muuttuja normaalisti jakautunut. Jos testin tulos on tilastollisesti merkitsevä ( $p < .05$ ), muuttujan jakauma **EI** noudata normaalijakaumaa. Tämän lisäksi muuttujan jakaumaa voidaan tarkastella graafisesti esimerkiksi histogrammin avulla, johon saa apukäyräksi normaalijakauman. Normaalisuuden testaus voidaan tehdä *myös* toiminnolla **Analyze – Descriptive Statistics – Explore**. Valitse muuttuja **Dependent List** -kohtaan ja **Plots**-painikkeen takaa **Normality Plots with tests**. Tällöin saat muuttujan tunnuslukujen lisäksi Kolmogorov-Smirnovin testin (mukana myös Shapiro-Wilk -korjaus) ja graafisia kuvaajia jakaumasta.

### TOIMENPITEET

1. Valitse **Analyze**-valikosta **Nonparametric Tests** ja sieltä **Legacy Dialogs → 1-sample K-S**.
2. Valitse muuttuja(t) laatikkoon **Test Variable List**. Voit halutessasi valita samalla kertaa useita muuttujia, jolloin kaikkien niiden jakaumat tulee testattua samalla kertaa.
3. Oletuksena ohjelma tarjoaa vertailua suhteessa normaalijakaumaan (Test Distribution → Normal).
4. Hyväksy valinnat painamalla **OK**-painiketta.



### ESIMERKKITULOSTUS

		Harjoittelun ohjaus
N		1674
Normal Parameters <sup>a, b</sup>	Mean	5,6663
	Std. Deviation	1,07566
Most Extreme Differences	Absolute	,128
	Positive	,108
	Negative	-,128
Kolmogorov-Smirnov Z		5,225
Asymp. Sig. (2-tailed)		,000

a. Test distribution is Normal.  
b. Calculated from data.

Ohjelma tulostaa muuttujan keskiarvon, hajonnan sekä poikkeamat havaitun ja teoreettisen jakauman kertymä-funktioiden välillä

Keskeisintä on kuitenkin tulkita Sig-arvoa (p-arvo). Jos se on **pienempi kuin .05** muuttujan jakauma **ei** noudata normaalijakaumaa. Tässä esimerkissä muuttujan jakauma ei noudata normaalijakaumaa. Huom. K-S-testi ilmoittaa varsin herkästi, että muuttuja ei noudata normaalijakaumaa. Siksi esim. graafinen tarkastelu voi olla tarpeellinen tämän testin rinnalla.

## 13. MANN-WHITNEYN U-TESTI

### KÄYTTÖTARKOITUS

Mann-Whitneyn U-testi kuuluu ns. ei-parametrisiin testeihin, mikä tarkoittaa, ettei siinä aseteta tutkittavalle testimuuttujalle – riippuvalle muuttujalle – jakaumavaatimuksia (oletuksia). U-testiä käytetään samanlaisissa tilanteissa kuin riippumattomien tapausten t-testiä, mutta se soveltuu myös järjestysasteikollisille testimuuttujille ja testimuuttujan ei siis tarvitse noudattaa normaalijakaumaa. t-testin oletusten ei siis tarvitse olla voimassa. U-testi sopii myös pienille aineistoille.

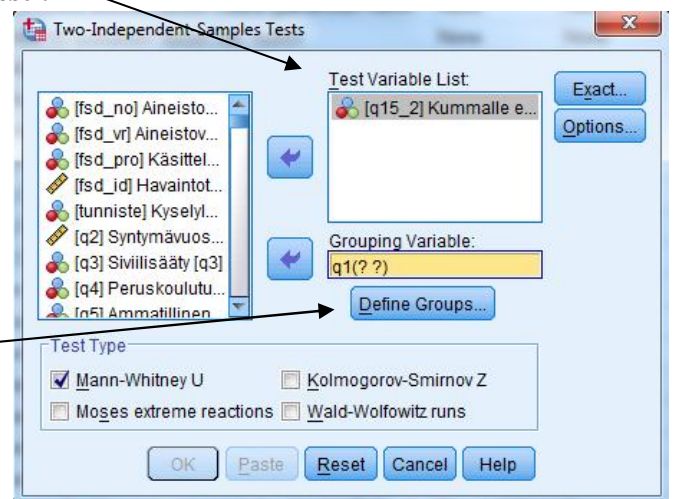
U-testiä käytetään tilanteessa, jossa ryhmittelevä (riippumaton) muuttuja muodostaa kaksi vertailtavaa ryhmää ja on olemassa yksi riippuva muuttuja, joka on vähintään järjestysasteikollinen. Testillä testataan näiden kahden ryhmän välistä eroa riippuvassa muuttujassa (ominaisuudessa).

### TOIMENPITEET

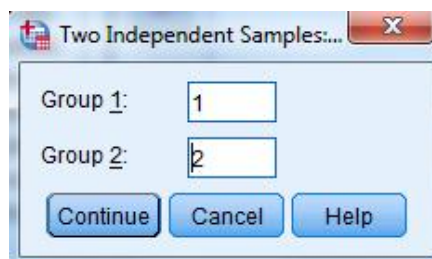
6. Valitse **Analyze**-valikosta **Nonparametric Tests** ja **Legacy Dialogs** → **2 Independent Samples**.
7. Valitse avautuvan ikkunan **Test Variable List** –kohtaan haluamasi testimuuttujan (voit valita samalla useitakin). Oletuksena on valittuna **U-testi**.

8. Valitse ryhmittelevä muuttuja **Grouping Variable** kohtaan.

9. Määrittele ryhmittelevän muuttujan luokat **Define Groups** -painikkeella, jolloin avautuu alla oleva ikkuna.



10. Kirjoita **Group 1**-laatikkoon ensimmäisen ryhmän numerotunnus (muuttujan arvo, esim. 1) ja **Group 2**-laatikkoon toisen ryhmän numerotunnus (muuttujan arvo, esim. 2).



11. Hyväksy valinnat **Continue** ja pääikkunan **OK**-painiketta näpäyttämällä.

## ESIMERKKITULOSTUS

Ranks				
[q15_2]	[q1] Sukupuoli	N	Mean Rank	Sum of Ranks
[q15_2] Kummalle ensisijaisesti, miehelle vai naiselle kuuluu mielestänne perheessä pukemisen opettaminen lapselle.	nainen	1236	798,90	987440,50
	mies	356	788,17	280587,50
	Total	1592		

Ylemmän taulukkoon tulostuvat testin käyttämät järjestyslukujen keskiarvot ja summat kummallekin ryhmälle. Se ryhmä, jolla on suurempi Mean Rank eli järjestyslukujen keskiarvo, saa suuremman arvon muuttujassa (on ”enemmän” sitä ominaisuutta tms.)

Tässä tapauksessa Mean Rank -arvo on suurempi naisilla. Se tarkoittaa, että naisten mielestä pukemisen opettaminen kuuluu enemmän naisille (muuttuja oli koodattu niin päin). Pitää kuitenkin katsoa myös seuraava taulukko.

Test Statistics <sup>a</sup>	
	[q15_2] Kummalle ensisijaisesti, miehelle vai naiselle kuuluu mielestänne perheessä pukemisen opettaminen lapselle.
Mann-Whitney U	217041,500
Wilcoxon W	280587,500
Z	-,485
Asymp. Sig. (2-tailed)	,627

a. Grouping Variable: [q1] Sukupuoli

### PÄÄTULOS

Alimpaan taulukkoon tulostuu varsinainen testitulos ja sen merkitsevyys. Jos saatu Sig-arvo (= p-arvo) on pienempi kuin .05, on ryhmien välinen ero tilastollisesti merkitsevä. Tässä esimerkissä ero ei ole tilastollisesti merkitsevä, vaikka ylemmän taulukon mukaan naisilla on hieman suurempi Mean Rank (ovat enemmän samaa mieltä väitteen kanssa). Miehet ja naiset eivät eroa tässä mitatussa ominaisuudessa.

## 14. KRUSKAL-WALLISIN TESTI

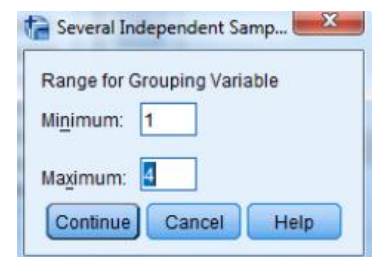
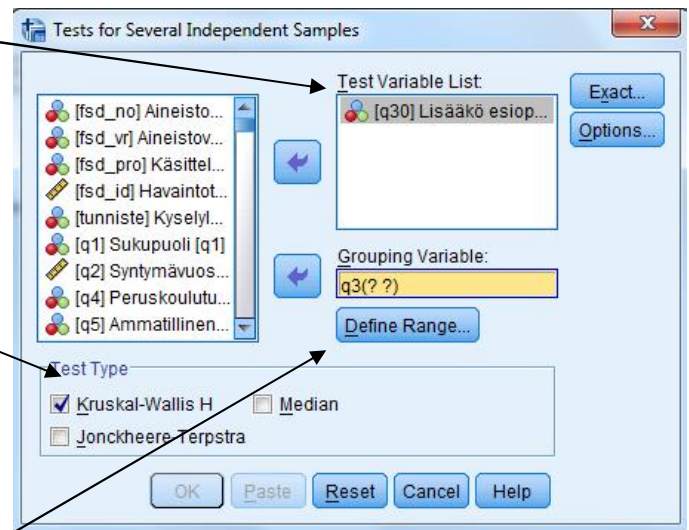
### KÄYTTÖTARKOITUS

Kruskal-Wallis testin (K-W-testi) kuuluu myös ns. ei-parametrisiin testeihin, mikä tarkoittaa, ettei siinä aseteta tutkittavalle testimuuttujalle – riippuvalla muuttujalla – jakaumavaatimuksia (oletuksia). K-W-testiä käytetään samanlaisissa tilanteissa kuin yksisuuntaista varianssianalyysia (ANOVA), mutta se soveltuu myös järjestysasteikollisille testimuuttujille ja testimuuttujan ei siis tarvitse noudattaa normaalijakaumaa. Yksisuuntaisen varianssianalyysin oletusten ei siis tarvitse olla voimassa. K-W-testi sopii myös pienille aineistoille.

K-W-testiä käytetään tilanteissa, jossa ryhmittelevä (riippumaton) muuttuja muodostaa enemmän kuin kaksi vertailtavaa ryhmää (yleensä 3–5) ja on olemassa yksi riippuva muuttuja, joka on vähintään järjestysasteikollinen. Testillä testataan näiden ryhmien välistä eroa riippuvassa muuttujassa (ominaisuudessa).

## TOIMENPITEET

1. Valitse **Analyze**-valikosta **Nonparametric Tests** ja **Legacy Dialogs** → **K Independent Samples**.
2. Valitse avautuvan ikkunan **Test Variable List** -kohtaan haluamasi testimuuttujan (voit valita samalla useitakin). Oletuksena on valittuna **K-W-testi**.
3. Valitse ryhmittelevä muuttuja **Grouping Variable** kohtaan.
4. Määrittele ryhmittelevän muuttujan luokat **Define Range** -painikkeella (tämä kohta pitää tehdä toisin muin ANOVASSA).
5. Kirjoita **Minimum**-laatikkoon ensimmäisen ryhmän numerotunnus (muuttujan arvo, esim. 1) ja **Maximum**-laatikkoon viimeisen ryhmän numerotunnus (muuttujan arvo, esim. 4).
6. Hyväksy valinnat **Continue** ja pääikkunan **OK**-painiketta näpäyttämällä.



## ESIMERKKITULOSTUS

Ranks			
	[q3] Sivillisäätv	N	Mean Rank
[q30] Lisääkö esiopetus mielestänne lapsen kouluvalmiuksia?	avoliitossa	236	796,05
	naimisissa	1135	796,41
	eronnut tai asumerossa	103	745,65
	naimaton	102	726,25
	Total	1576	

Ylempään taulukkoon tulostuvat testin käyttämät järjestyslukujen keskiarvot kullekin ryhmälle. Se ryhmä, jolla on suurin Mean Rank eli järjestyslukujen keskiarvo, saa suurimman arvon muuttujassa.

Tässä tapauksessa naimisissa olevan ryhmän arvo on suurin ja naimattomien pienin. Naimattomat kokevat vähiten, että esiopetus lisää lapsen kouluvalmiuksia.

Test Statistics <sup>a,b</sup>	
	[q30] Lisääkö esiopetus mielestänne lapsen kouluvalmiuksia?
Chi-Square	4,227
df	3
Asymp. Sig.	,238

a. Kruskal Wallis Test

b. Grouping Variable:  
[q3] Sivillisäätv

### PÄÄTULOS

Alimpaan taulukkoon tulostuu varsinainen testitulos ja sen merkitsevyys. Jos saatu Sig-arvo (= p-arvo) on pienempi kuin .05, on ryhmien välinen ero tilastollisesti merkitsevä. Tässä esimerkissä ero ei ole tilastollisesti merkitsevä, vaikka ylempään taulukkoon mukaan eroja jonkun verran onkin. Erot eivät ole kuitenkaan niin suuria, etteivätkö ne olisi voineet johtua sattumasta.

## 15. WILCOXONIN TESTI

### KÄYTTÖTARKOITUS

Wilcoxonin testi kuuluu ns. ei-parametrisiin testeihin, mikä tarkoittaa, ettei siinä aseteta tutkittavalle muuttujille jakaumaoletuksia. Wilcoxonin testiä käytetään samanlaisissa tilanteissa kuin riippuvien tapausten t-testiä, mutta se soveltuu myös järjestysasteikollisille testimuuttujille ja testimuuttujan ei siis tarvitse noudattaa normaalijakaamaa. t-testin oletusten ei siis tarvitse olla voimassa.

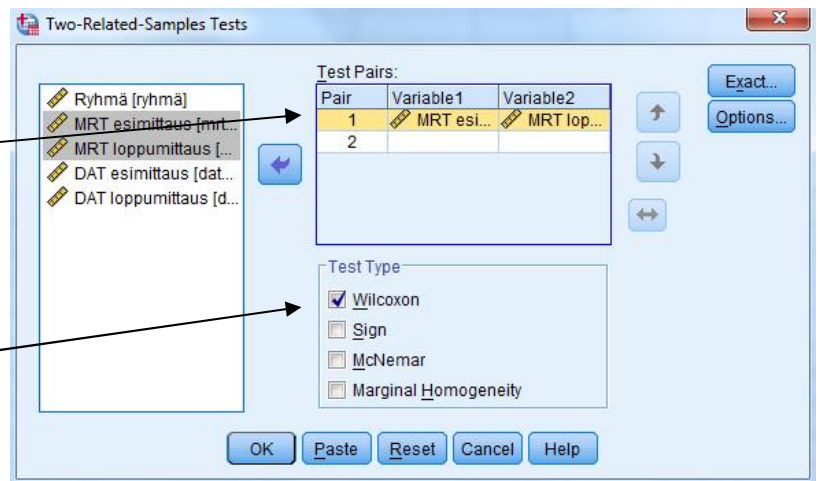
Riippuvat tapaukset tarkoittavat sitä, että on mitattu **SAMOJA** tapauksia, esimerkiksi ihmisryhmiä kaksi eri kertaa jossakin ominaisuudessa. Testiä varten pitää valita siis muuttujapari (2 Related Samples), esimerkiksi muuttujapari alkutesti–lopputesti.

### TOIMENPITEET

1. Valitse **Analyze**-valikosta **Nonparametric Tests** ja **Legacy Dialogs** → **2 Related Samples**.

2. Valitse muuttujat hiiren avulla näpäyttämällä. Muuttujien pitää siis muodostaa pari (esim. alkumittaus – loppumittaus, esimittaus – loppumittaus).

3. Oletusarvoisesti valittuna on **Wilcoxonin** testi.



4. Voit laatia useita pareja samalla tekniikalla (kohta 2) samaan analyysiin.
5. Hyväksy valinnat **OK**-painiketta näpäyttämällä.

### ESIMERKKITULOSTUS

		Ranks		
		N	Mean Rank	Sum of Ranks
MRT loppumittaus - MRT esimittaus	Negative Ranks	8 <sup>a</sup>	13,44	107,50
	Positive Ranks	71 <sup>b</sup>	42,99	3052,50
	Ties	5 <sup>c</sup>		
	Total	84		

- a. MRT loppumittaus < MRT esimittaus
- b. MRT loppumittaus > MRT esimittaus
- c. MRT loppumittaus = MRT esimittaus

Ylempään taulukkoon tulostuvat mm. testin käyttämät järjestyslukuparien erotuksien keskiarvot.

**Positive Rank** = ensimmäisessä muuttujassa on alempi arvo  
**Negative Rank** = toisessa muuttujassa on alempi arvo  
**Ties** = molemmissa muuttujissa sama arvo  
 Vihjeet ovat apuna tulkinnassa.

Test Statistics<sup>b</sup>

	MRT loppumittaus - MRT esimittaus
Z	-7,211 <sup>a</sup>
Asymp. Sig. (2-tailed)	,000

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

**PÄÄTULOS**

Alempaan taulukkoon tulostuu varsinainen testitulos ja sen merkitsevyys. Jos saatu Sig-arvo (= p-arvo) on pienempi kuin .05, on muuttujien ero tilastollisesti merkitsevä. Tässä esimerkissä ero on tilastollisesti merkitsevä eli ylemmän taulukon avulla voidaan sanoa, että loppumittauksessa saatiin parempia pistemääriä kuin esimittauksessa.

## 16. RELIABILITEETIN MÄÄRITTÄMINEN MUUTTUJAJOUKOLLE

### KÄYTTÖTARKOITUS

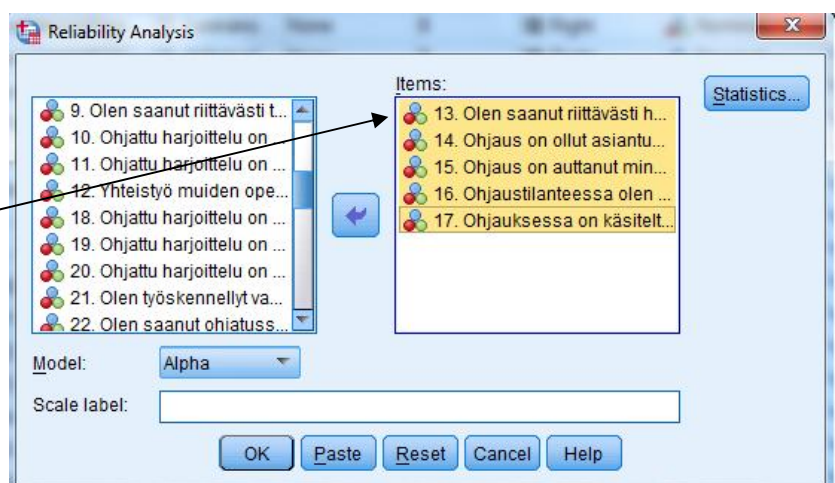
Reliabiliteetti tarkoittaa mittauksen pysyvyyttä, ei-sattumanvaraisuutta, siis mittarin, summamuuttujan tai menetelmän kykyä antaa ei-sattumanvaraisia tuloksia. Reliabiliteetti on sitä suurempi, mitä vähemmän tuloksissa sattumanvaraisuutta. Esimerkiksi kyselytutkimuksissa reliabiliteetin pitäisi olla suurempi kuin 0,50 (= 50 %) jolloin yli puolet muuttujien vaihtelusta on todellista ja alle puolet satunnaisvaihtelua. Joskus rajana pidetään lukua 0,70. Tässä esitellään reliabiliteetin määrittäminen Cronbachin alfan avulla.

Reliabiliteetti määritetään yleensä samaa sisältöaluetta mitanneiden muuttujien kesken. Etenkin summamuuttujien reliabiliteetin määrittäminen on tärkeää ja jopa välttämätöntä **ENNEN** kuin summamuuttujaa tehdään. Liian alhaisen reliabiliteetin omaavaa summamuuttujaa ei saa käyttää, eikä laatia.

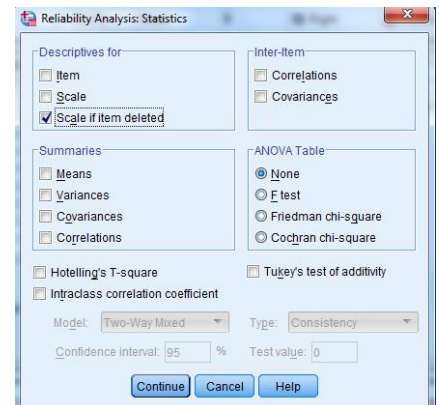
**Ennen reliabiliteetin määrittämistä ja/tai summamuuttujan laatimista pitää eri suuntaan (esim. yksi negatiivinen, muut positiivisia mielipiteitä) mitanneiden muuttujien pistemäärät kääntää samaan suuntaan.** Katso ohjeita kohdasta ”Muuttujan kääntäminen ja luokittelu”. Jos summamuuttujat laaditaan pääkomponentti- tai faktorianalyysin avulla, pitää huomioida minkä suuntaisia latauksia kukin muuttuja sai pääkomponentti- tai faktorianalyysimatriisissa ja kääntää pistemäärät tarvittaessa.

### TOIMENPITEET

1. Valitse **Analyze**-valikosta **Scale ja Reliability Analysis**.
2. Valitse muuttujat **Items**-laatikkoon.



3. Valitse **Statistics**-painonapin takaa ainakin **Scale if Item Deleted** (= miten reliabiliteetti muuttuisi, jos joku osio poistettaisiin ”nipusta”).



4. Hyväksy valinnat näpäyttämällä **Continue** ja **Reliability Analysis** -ikkunan **OK**-painiketta.

## ESIMERKKITULOSTUS

### Reliability

#### Scale: ALL VARIABLES

##### Case Processing Summary

		N	%
Cases	Valid	1674	100,0
	Excluded <sup>a</sup>	0	,0
	Total	1674	100,0

Ylimpään taulukkoon tulostuu mukaan tullut tapausmäärä.

a. Listwise deletion based on all variables in the procedure.

##### Reliability Statistics

Cronbach's Alpha	N of Items
,890	5

Seuraavaan taulukkoon tulostuvat päätulos eli Cronbachin alfa ja muuttujien lukumäärä (N of Items). Tässä alfa on varsin hyvä (selvästi yli .50). Huomaa, että tässä ei ole kyse merkittävyyden testauksesta, vaan muuttujajoukon sisäisen yhtenäisyyden tarkastelusta.

##### Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
13. Olen saanut riittävästi henkilökohtaista ohjausta	22,67	18,983	,723	,868
14. Ohjaus on ollut asiantuntevaa	22,55	18,642	,817	,847
15. Ohjaus on auttanut minua kehittymään opettajana	22,50	18,642	,804	,850
16. Ohjaustilanteessa olen voinut tuoda esille omia näkemyksiäni	22,62	19,414	,676	,878
17. Ohjauksessa on käsitelty opetettavan aineen / aineiden erityiskysymyksiä	22,98	19,435	,649	,885

**Alpha if Item Deleted** -sarakkeen luvut tarkoittavat reliabiliteetin arvoa, **jos** kyseinen muuttuja poistettaisiin summamuuttujasta tai vastaavasta. Tässä ei kannata poistaa yhtäkään muuttujaa, koska poistamalla reliabiliteetti ei parani.