

Christina Gustafsson

**Tilastollinen tietojenkäsittely STAT2100
IBM SPSS Statistics 20 for Windows
Osa 2**

Kevät 2013

SISÄLLYSLUETTELO

5. YKSIULOTTEISET JAKAUMAT	2
5.1. Frequencies-proseduuri	2
5.2. Descriptives-proseduuri	4
5.3. Explore-proseduuri	5
5.4. Ratio Statistics -proseduuri	8
6. RISTIINTAULUKOINNISTA	9
7. KORRELAATIOISTA	12
8. KESKIARVOJEN VERTAILUA JA KESKIARVOTESTEJÄ	14
8.1. Keskiarvojen ja muiden tunnuslukujen laskentaa ryhmille	14
8.2. Yhden otoksen keskiarvotesti	15
8.3. Kahden riippumattoman otoksen keskiarvotesti	16
8.4. Kahden riippuvan otoksen keskiarvotesti	17
8.5. Varianssianalyysi	19

5. YKSIULOTTEISET JAKAUMAT

Kun tilastollinen analyysi aloitetaan, on ensimmäinen tehtävä aineiston kuvaaminen *frekvenssijakaumien* ja *tilastollisten tunnuslukujen* avulla. Kuvaamisen jälkeen aineistoa voidaan tarpeen tullen tiivistää muodostamalla summamuuttujia ja/tai tekemällä faktorianalyysin (luku 11). Lopuksi tehdään riippuvuustutkimukset ja selittäminen käyttämällä erilaisia menetelmiä (ristiintaulukointi, korrelaatiot, keskiarvotarkastelut, varianssianalyysi, regressioanalyysi jne)

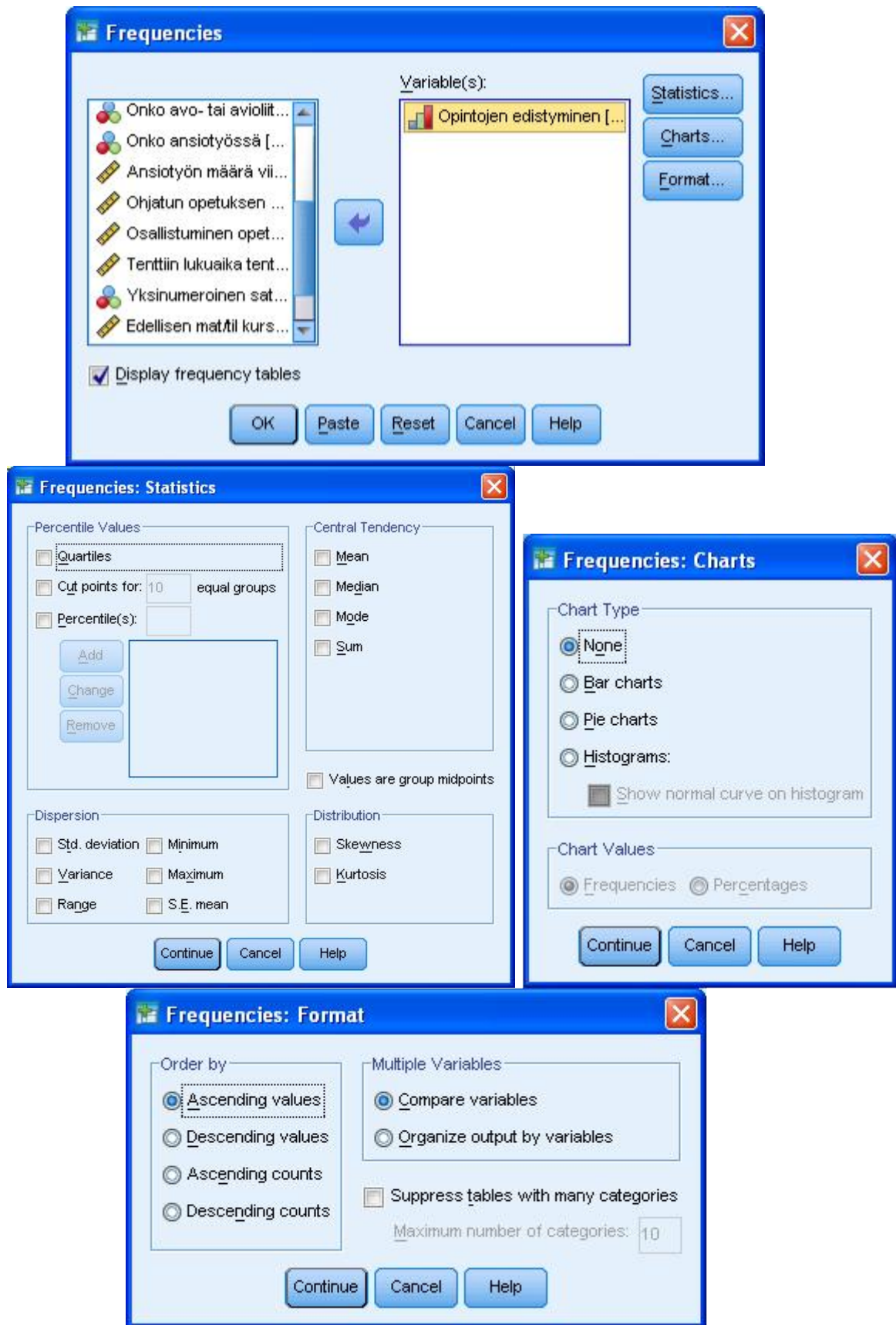
Varsinaiset yksiulotteisen jakauman kuvailuun liittyvät tilastolliset proseduurit ovat **Analyze**-valikon valinnan **Descriptive Statistics** proseduurit **Frequencies**, **Descriptives**, **Explore** ja **Ratio**.

Ko. proseduureissa valittavia tunnuslukuja ja tuloksissa esiintyviä lukuja ovat ryhmittäin mm.

Percentile Values:	fraktiileja/prosenttiileja
Quartiles	kvartiilit
Cut points for # equal groups	luokkarajat #:lle frekvensseiltään yhtäsuurille ryhmälle
Percentile(s)	halutut prosenttiilit/fraktiilit
Central Tendency:	keskilukuja
Mean	aritmeettinen keskiarvo
Median	keskusarvo, mediaani
Mode	tyyppiarvo, moodi
Sum	havaintoarvojen summa
(5%) Trimmed Mean	(5 % :n) leikattu keskiarvo
Weighted Mean	painotettu keskiarvo
Dispersion:	hajontalukuja
Std. deviation	keskihajonta
Variance	varianssi
Range	vaihteluvälin pituus
Max(imum)	suurin arvo
Min(imum)	pienin arvo
S(tandard) E(rror) (for) mean	keskiarvon keskivirhe
I(nter) Q(uartile) R(ange)	kvartiilivälin pituus
Distribution:	jakauman muotoa kuvaavia tunnuslukuja
Kurtosis	huipukkuus
S E Kurt	huipukkuuden keskivirhe
Skewness	vinous
S E Skew	vinouden keskivirhe
Muita lukuja:	
95% C(onfidence) I(nterval) for Mean	keskiarvon 95 % :n luottamusväli
Valid cases	analyysissä käytettyjen havaintojen lkm
Missing cases	puuttuvien havaintojen lkm
Percent missing	puuttuvien havaintojen prosentuaalinen osuus
Values are group midpoints	muuttuja-arvot ovat luokkakeskuksia

5.1. Frequencies-proseduuri

Frequencies-proseduurilla (kuvio 47) voidaan muodostaa *frekvenssijakauma* (**Display frequency tables**). Proseduuria kannattaa käyttää lähinnä *nominaali- ja järjestysasteikon* muuttujille (esimerkissä *Opintojen edistyminen*). Proseduurin avulla voidaan tuottaa myös tunnuslukuja (**Statistics**-painike) ja kuvioita (**Charts**-painike). Jos jatkuvan muuttujan frekvenssijakauma halutaan esittää taulukkomuodossa, on muuttujan arvot ensin luokiteltava (proseduuri **Transform-Recode**).



Kuvio 47. Frequencies -proseduuriin liittyvät määrittely- ja lisämäärittelyikkunat

Valittavat kuviot ovat diskreetille muuttujalle sopivat **Bar charts** eli pylväskuvio ja **Pie charts** eli piirakkakuvio, joissa luokkien havaintomäärät voidaan ilmaista absoluuttisina tai prosentuaalisina (**Chart Values: Frequencies/Percentages**) sekä jatkuva-arvoiselle muuttujalle sopiva **Histograms**, jolloin

kuvioon voidaan myös lisätä normaalijakauman tiheysfunktion kuvaaja (**Show normal curve on histogram**). Jos minkäänlaista kuvioita ei haluta, voidaan valita **None**.

Format-määrittelyillä voidaan muuttujan luokat tulostuksessa järjestää (**Order by**) nousevaan (**Ascending**) tai laskevaan (**Descending**) järjestykseen joko muuttuja-arvon (**Values**) tai havaintomäärän (**Counts**) suhteen. Jos proseduurin on valittu useita muuttujia (**Multiple Variables**) voidaan tulokset esittää niin, että muuttujia on helppo vertailla (**Compare variables**) tai muuttujien tiedot esitetään yksitellen (**Organize output by variables**). Jos proseduurilla käsitellään jatkuva-arvoisia muuttujia, voidaan estää frekvenssijakaumien tulostus, jos luokkia on enemmän kuin ## kpl (**Suppress tables with more than n categories. Maximum number of categories: ##**).

Tavallinen frekvenssijakaumataulukko (kuvio 48) sisältää frekvenssit (**Frequency**), prosenttiset frekvenssit, jotka on laskettu kaikista tilastoyksiköistä (**Percent**) tai vain niistä, joiden muuttuja-arvot tiedetään (**Valid Percent**) sekä prosenttisen summajakauman (**Cumulative Percent**).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	selvästi hitaammin	11	3,4	3,5	3,5
	jonkin verran hitaammin	52	16,1	16,4	19,9
	keskimääräisesti	207	64,3	65,3	85,2
	jonkin verran nopeammin	43	13,4	13,6	98,7
	selvästi nopeammin	4	1,2	1,3	100,0
	Total	317	98,4	100,0	
Missing	System	5	1,6		
	Total	322	100,0		

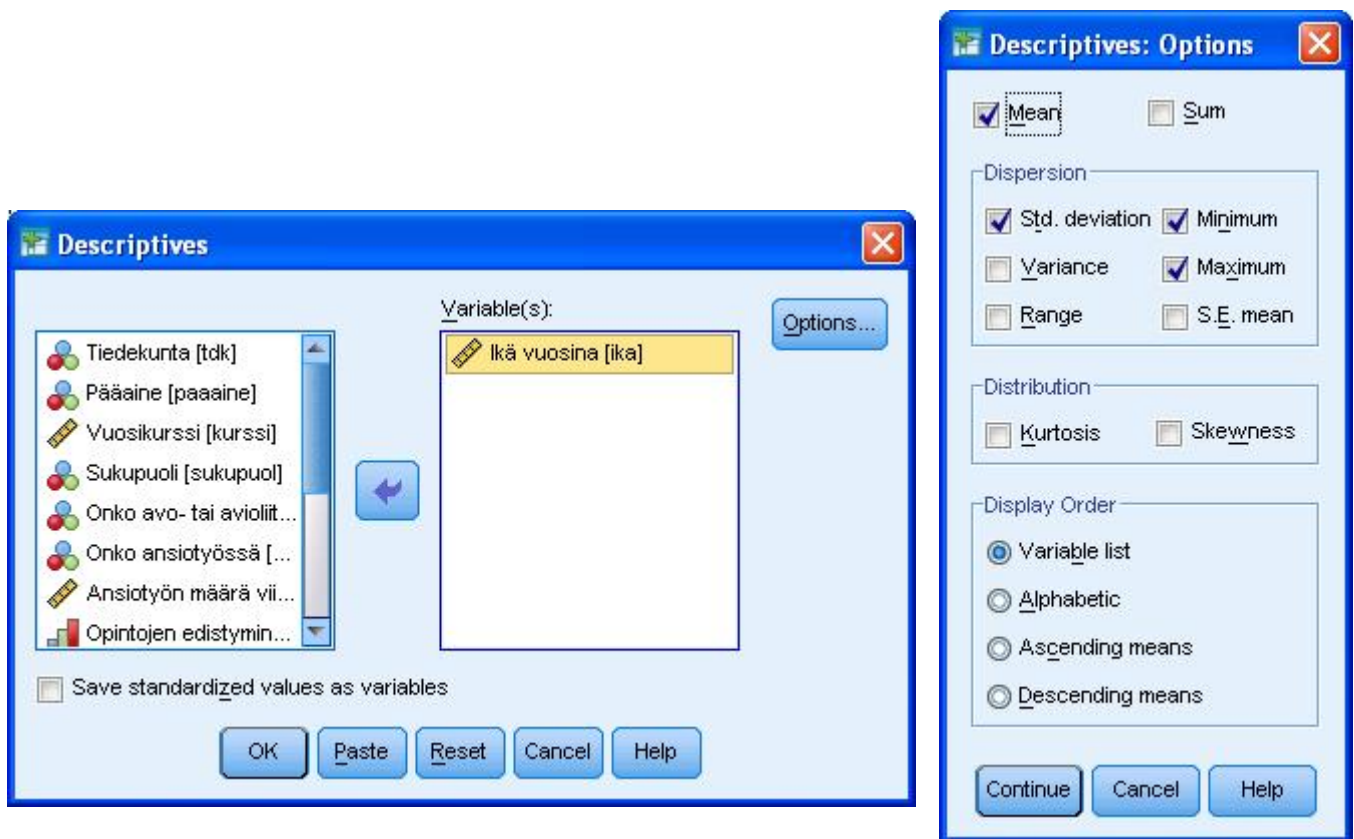
Kuvio 48. Frequencies-proseduurin tulostusta

Taulukon perusteella voidaan todeta, että 322 opiskelijasta suurin osa (65.3 %) kokee opintojensa edistyvän keskimääräisesti (eli opintojen edistymisen moodiarvo on ”keskimääräisesti”). Jos proseduurin **Statistics**-lisävalinnalla olisi määritetty laskettavaksi tilastollisia tunnuslukuja, olisivat niiden arvot olleet käytetyn muuttujan numerokodeista laskettuja suureita eli esim. tällöin moodiarvoksi ja myös mediaaniarvoksi olisi esitetty luku ”3”.

5.2. Descriptives-proseduuri

Descriptives-proseduuri (kuvio 49) sopii käytettäväksi vain **määrällisille** (eli *intervalli-* tai *suhdeasteikon*) muuttujille (esimerkissä *Ikä vuosina*). Tämän proseduurin avulla voidaan määrittää muutamia **tunnuslukuja**, jotka sopivat vain määrällisille muuttujille, ja lisätä havaintoaineistoon muuttujien standardoidut arvot (**Save standardized values as variables**), joita voi käyttää mm. poikkeavien havaintojen tutkimisessa ja vertailutilanteissa.

Options-lisämäärittelyillä saadaan valittua jakaumaa kuvaavia tilastollisia tunnuslukuja sekä voidaan määrätä muuttujien esittämisjärjestys (**Display Order**) tulostuksessa. Oletusarvoisesti tunnusluvuista tulostuvat keskiarvo, keskihajonta, pienin arvo ja suurin arvo. Useimmiten määrällisen muuttujan kuvailussa halutaan kuitenkin esille vielä mediaaniarvokin, jota ei kuitenkaan tämän proseduurin vaihtoehtoisista löydy.



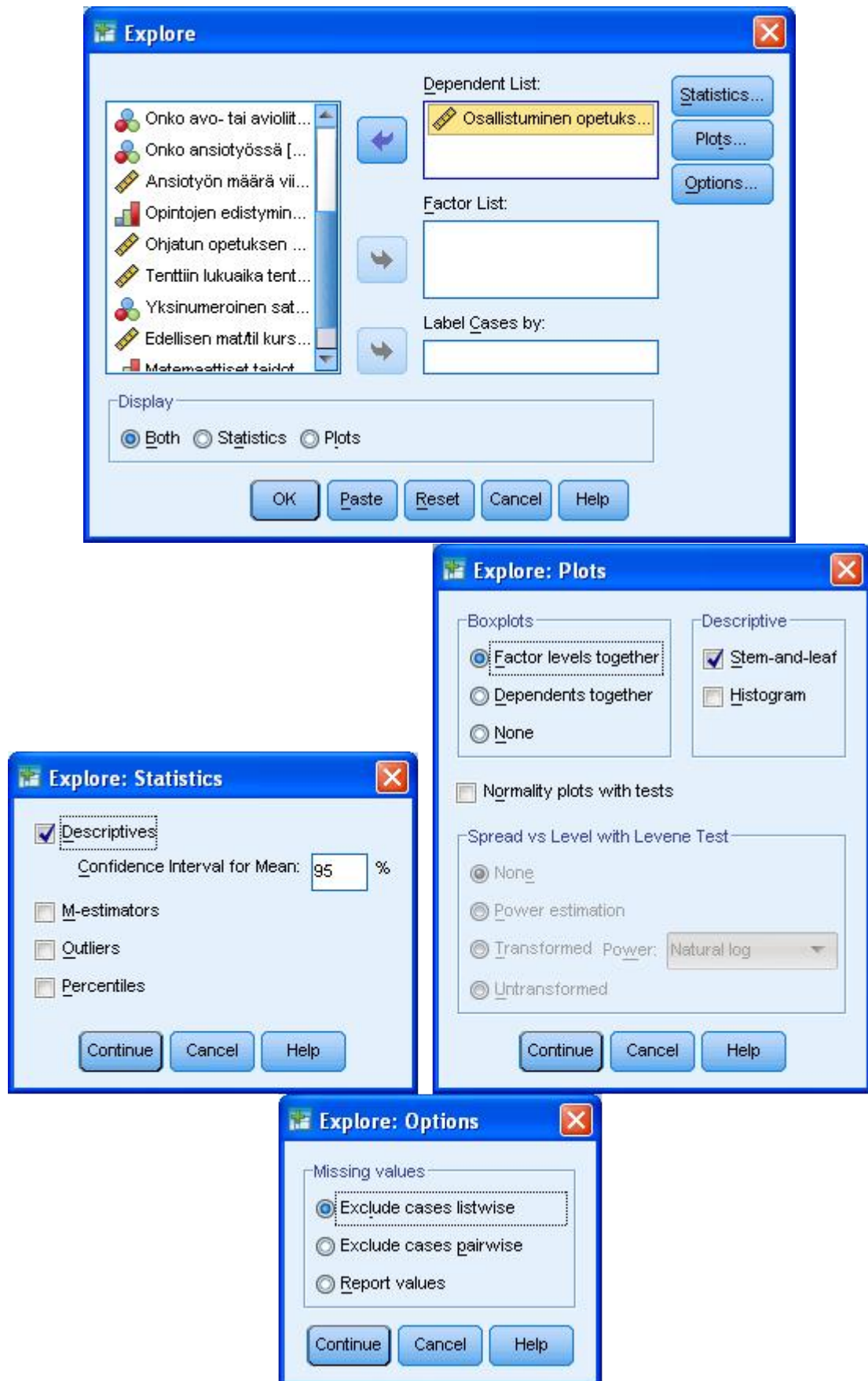
Kuvio 49. Descriptives-proseduurin määrittely- ja lisämäärittelyikkunat

5.3. Explore-proseduuri

Explore-proseduuri (kuviokuva 50) on laajin yksiulotteisen jakauman tutkimisproseduuri, mutta se soveltuu lähinnä *määrällisten* muuttujien *jakaumien kuvailuun* ja *tutkimiseen*. Laadullisia muuttujia voidaan kuitenkin ottaa proseduriin mukaan ryhmitteleviksi muuttujiksi. Päämäärittelyikkunassa varsinaiset tutkivat muuttujat valitaan kohtaan **Dependent List** (esimerkissä *Osallituminen opetuks...*), mahdolliset ryhmittelevät muuttujat kohtaan **Factor List** ja havainnot nimeävä muuttuja kohtaan **Label Cases by**.

Tulostukseen voi valita oletusarvoisten tunnuslukujen (**Descriptives**) lisäksi **Statistics**-lisämäärittelyillä robusteja keskilukuja (**M-estimators**), viisi pienintä ja suurinta havaintoarvoa (**Outliers**) sekä fraktiileja (**Percentiles**).

Plots-lisämäärittelyillä voidaan muodostaa laatikko-viikset -kuvioita (**Boxplots**) muuttujittain (**Factor levels together**) tai ryhmittäin (**Dependents together**), runko-lehti-kuvio (**Stem-and-leaf**) sekä frekvenssihistogrammi (**Histogram**). Valinnalla **Normality plots with tests** muodostuu kuvioita: Normal Q-Q Plot ja Detrended Q-Q Plot ja testit: Kolmogorov-Smirnov ja Shapiro-Wilk, joilla voidaan tutkia muuttujan normaalijakautuneisuutta. Kohdan **Spread vs. Level with Levene Test** avulla voidaan tutkia ryhmien varianssien yhtäsuuruutta. Levenen testi voidaan suorittaa vain, jos päämäärittelyikkunassa on valittu ryhmittelevä muuttuja kohtaan **Factor List**. Valinnalla **Power estimation** ohjelma ehdottaa muuttujalle sopivaa muunnosta, jotta ryhmien varianssit saataisiin yhtäsuuriksi, valinnalla **Transformed** voi itse valita sopivan muunnoksen ja valinnalla **Untransformed** ei muunnoksia muodostu.



Kuvio 50. Explore-proseduurin määrittely- ja lisämäärittelyikkunat

Explore-proseduurin tulostaulukoista (kuvio 51) nähdään mm. seuraavaa. Muuttujan keskiarvo on 16.92 tuntia ja mediaani 16.00 tuntia(, joten muuttujan jakauma on hieman oikealle loiveneva eli positiivisesti vino.) Keskihajonta on 5.23 (, joten opiskelijoiden osallistumistuntimäärä eroaa keskimäärin 5.23 tuntia

keskiarvosta). Kvartiilivälin pituus 6.00 tuntia on lyhyt verrattuna koko vaihteluvälin pituuteen 31 tuntia (, joten muuttujan arvot ovat keskittyneet voimakkaasti mediaanin ympärille). Vinouden 0.402 suhde vinouden keskivirheeseen 0.137 on 2.9 (> 2), joten jakaumaa voidaan pitää oikealle loiveneva eli positiivisesti vinona. (Jos suhde olisi välillä $(-2, +2)$ olisi jakauma symmetrinen, ja jos suhde olisi pienempi kuin -2 , olisi jakauma vasemmalle loiveneva eli negatiivisesti vino). Huipukkuuden 0.545 suhde huipukkuuden keskivirheeseen 0.273 on 1.9 (siis välillä $(-2, +2)$), joten jakauma on mesokurtinen. (Jos suhde olisi pienempi kuin -2 , sanottaisiin jakaumaa platykurtiseksi eli normaalijakaumaa laakeammaksi, ja jos suhde olisi suurempi kuin $+2$, olisi jakauma leptokurtinen eli normaalijakaumaa huipukkaampi.)

Kolmogorov-Smirnov –normaalijakaumatestissä (, jota tulkitaan nyt, koska havaintoja on paljon) ja muissakin normaalijakaumatesteissä (esim. Shapiro-Wilk, jota käytetään, kun havaintomäärä on alle 50) hypoteesit ovat

H_0 : populaatiossa ko. muuttujan jakauma on normaalijakauma

H_1 : populaatiossa ko. muuttujan jakauma ei ole normaalijakauma

Kolmogorov-Smirnov –normaalijakaumatestin havaittu merkitsevyytaso eli ns. p-arvo on 0.000, joten muuttuja ei ole normaalijakautunut. Saatu tulos on tilastollisesti erittäin merkitsevä, koska p-arvo < 0.001 . (Jos $p \geq 0.05$, tulos ei ole tilastollisesti merkitsevä; jos $0.01 \leq p < 0.05$, tulos on tilastollisesti melkein merkitsevä; jos $0.001 \leq p < 0.0$, tulos on tilastollisesti merkitsevä.)

Descriptives

			Statistic	Std. Error
Osallistuminen opetukseen (tuntia viikossa)	Mean		16,92	,29
	95% Confidence Interval for Mean	Lower Bound	16,34	
		Upper Bound	17,50	
	5% Trimmed Mean		16,76	
	Median		16,00	
	Variance		27,375	
	Std. Deviation		5,23	
	Minimum		2	
	Maximum		33	
	Range		31	
	Interquartile Range		6,00	
	Skewness		,402	,137
	Kurtosis		,545	,273

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Osallistuminen opetukseen (tuntia viikossa)	,117	316	,000	,976	316	,000

a. Lilliefors Significance Correction

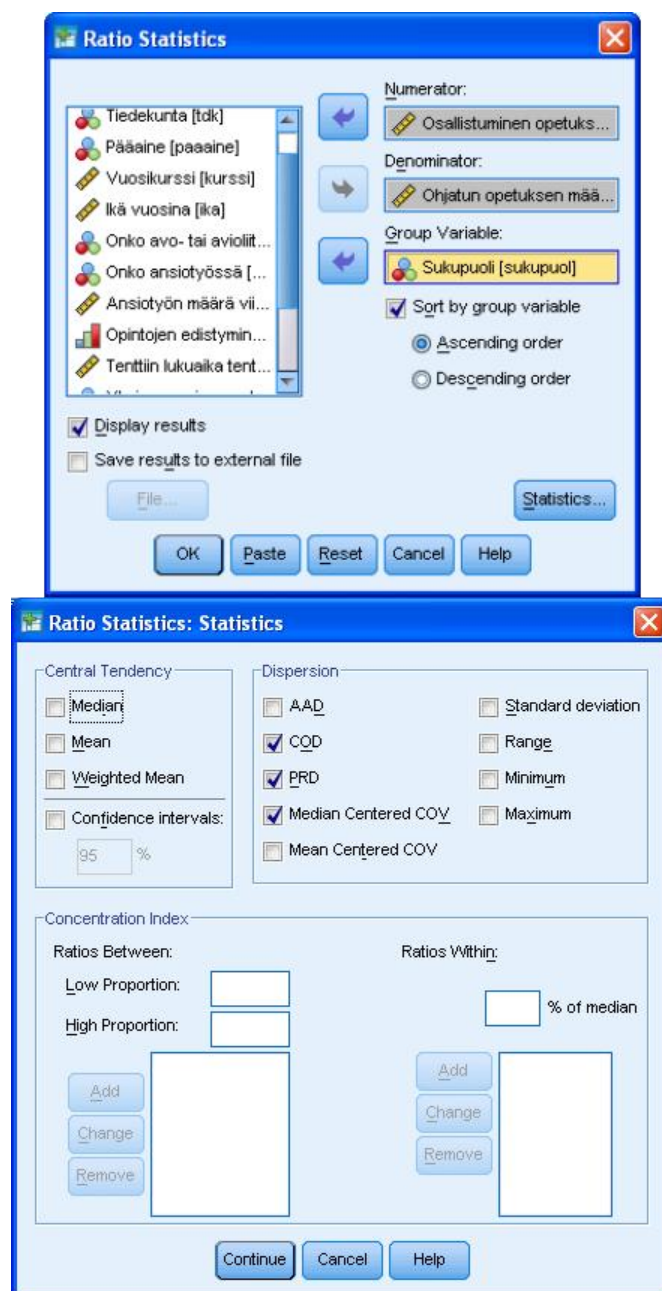
Kuvio 51. Explore-proseduurin tulostusta

5.4. Ratio Statistics -proseduuri

Ratio Statistics -proseduurilla (kuvio 52) voi määrittää *tilastollisia tunnuslukuja* kahden **määrällisen** muuttujan *osamäärälle*.

Ruutuun **Numerator** valitaan osoittajaksi tuleva muuttuja (esimerkissä *Osallistuminen opetuks...*) ja ruutuun **Denominator** nimittäjäksi tuleva muuttuja (esimerkissä *Ohjatun opetuks...*). Ruutuun **Group Variable** voidaan valita muuttuja, jonka perusteella tilastoyksiköt ryhmitellään (esimerkissä *Sukupuoli*).

Statistics-lisämäärittelyillä saadaan valittua tavallisimpien tunnuslukujen lisäksi hajontaluvuista mm. **AAD** (mediaaniin perustuva keskipoikkeama), **COD** (edellinen tunnusluku suhteutettuna mediaaniin), **PRD** (keskiarvon ja painotetun keskiarvon suhde), **Median Centered COV** (mediaanikeskistetty variaatiokerroin) ja **Mean Centered COV** (keskiarvokeskistetty variaatiokerroin) sekä erilaisia muuttuja-arvojen keskittymistä/hajaantumista kuvaavia suhteita (**Concentration Index**).



Kuvio 52. Ratio Statistics -proseduurin määrittely- ja lisämäärittelyikkunat

6. RISTIINTAULUKOINNISTA

Usean muuttujan samanaikaisen tarkastelun lähtökohtana on usein muuttujien välisen mahdollisen yhteyden/riippuvuuden olemassaolon, voimakkuuden ja luonteen selvittäminen. **Tilastollisen** (epätäydellisen) **riippuvuuden** kuvaamiseen voidaan käyttää mm. tilastollisia riippuvuuslukuja, estimoituja tilastollisia malleja jne. Riippuvuuden tutkimistapa riippuu tarkasteltavien muuttujien mitta-asteikoista. Tutkimistapaan vaikuttaa myös se, millä tavalla riippuvuutta halutaan kuvata ja myös se, minkä tyyppistä riippuvuutta ollaan tutkimassa.

Ristiintaulukointi on **riippuvuuden** tutkimusmenetelmä, joka sopii kaikille muuttujille, kunhan vain muuttujan arvot on luokiteltu, mutta useimmiten menetelmää käytetään kahdelle **nominaaliasteikolliselle** (tai toinen **nominaali-** ja toinen **järjestysasteikollinen**) muuttujalle. Ristiintaulukointi paljastaa minkä luonteisen riippuvuuden tahansa. Kahden (tai useamman) muuttujan ristiintaulukko (kontingenssitaulukko, kaksiulotteinen frekvenssijakauma, yhteisjakauma) saadaan muodostettua **Analyze**-valikon valinnan **Descriptive Statistics** alivalikolla **Crosstabs**.

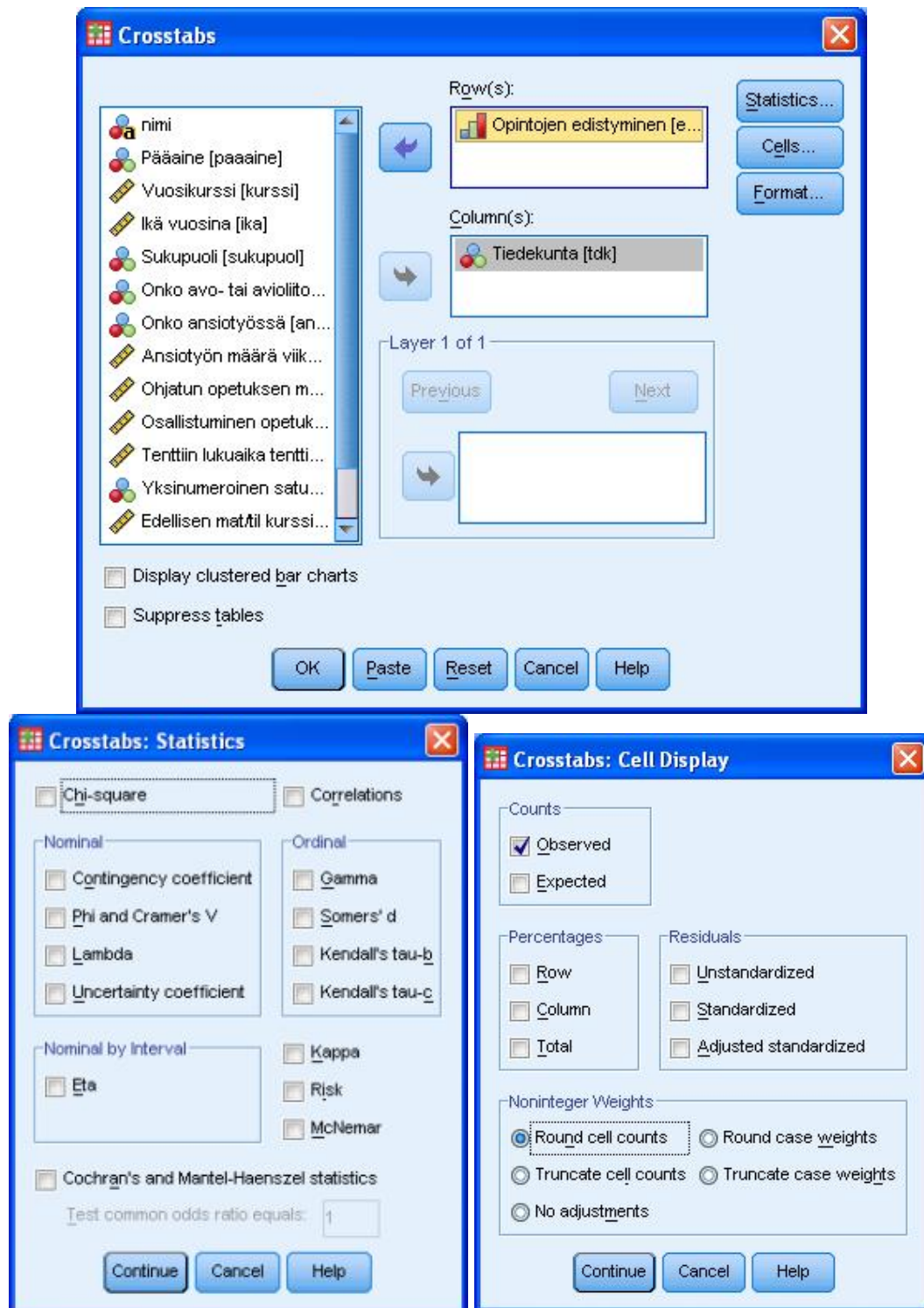
Avautuvan päämäärittelyikkunan (kuvio 53) **Rows**-ruutuun valitaan rivimuuttuja(t) (esimerkissä *Opintojen edistyminen*) ja **Columns**-ruutuun sarakemuuttuja(t) (esimerkissä *Tiedekunta*). **Layer**-ruutuun voidaan valita ryhmitteleviä muuttujia (, jolloin saadaan aikaiseksi kolmi- tai moniulotteisia ristiintaulukoita). Proseduurin avulla voidaan myös tuottaa ryhmiteltyjen pylväiden kuvio valinnalla **Display clustered bar charts** ja ristiintaulukoiden tulostus saadaan estettyä valinnalla **Suppress tables**.

Statistics-valinnalla avautuu lisämäärittelyikkuna, josta voi valita, mitä tunnuslukuja ristiintaulukosta lasketaan. **Chi-square** -valinnalla saadaan muodostettua χ^2 -testisuureita (esim. Pearsonin χ^2 -riippumattomuudesta, uskottavuussuhdetesti ja nelikentälle lisäksi esim. Fisherin testi), joilla voidaan tutkia riippuvuuden voimakkuutta. **Correlations**-valinnalla saadaan muodostettua Spearmanin järjestyskorrelaatio, joka sopii riippuvuusluvuksi, kun muuttujat ovat vähintään järjestysasteikollisia sekä Pearsonin lineaarinen korrelaatio, joka sopii riippuvuusluvuksi määrällisille muuttujille.

Nominal-ruudussa on sellaisia riippuvuustunnuslukuja, jotka sopivat kaikille mitta-asteikoille nominaaliasteikosta lähtien. **Ordinal**-ruudussa on sellaisia riippuvuuslukuja, jotka sopivat järjestysasteikollisille tai määrällisille muuttujille. **Nominal by Interval**-valinnalla muodostuu **Eta**-kerroin, joka sopii riippuvuusluvuksi, kun toinen muuttuja on nominaaliasteikolla mitattu ja toinen määrällinen.

Jos ristiintaulukon rivi- ja sarakeluokat ovat samat, valitaan riippuvuusluvuksi **Kappa**. Nelikenttätaulukossa voidaan valita **Risk**, jolla saadaan muodostettua suhteellisen riskin estimaatit. **McNemar**-testi sopii nelikenttään, jossa ristiintaulukko on muodostettu ennen-jälkeen-luokittelun perusteella. **Cochran's and Mantel-Haenzel** -testejä käytetään, kun tutkitaan dikotomisen ryhmittelymuuttujan ja dikotomisen vastemuuttujan välistä riippuvuutta.

Cells-valinnalla avautuu lisämäärittelyikkuna, jossa ristiintaulukon havaittujen solufrekvenssien (**Counts: Observed**) lisäksi voidaan valita tulostuuko soluihin odotetut eli teoreettiset frekvenssit (**Expected**), prosenttijakaumat (**Percentages**) rivisummista (**Row**), sarakesummista (**Column**) tai kokonaismäärästä (**Total**) sekä jäännökset (**Residuals**) sellaisenaan (**Unstandardized**) tai standardoituina (**Standardized/Adj. standardized**). Tässä ikkunassa valitaan myös se, miten toimitaan sellaisessa tilanteessa, jossa käytetään sellaista painomuuttujaa, jonka arvot eivät ole kokonaislukuja. Tällöin solufrekvenssit pyöristetään (**Round cell counts**) tai katkaistaan (**Truncate cell counts**), painomuuttujan arvot pyöristetään (**Round case weights**) tai katkaistaan (**Truncate case weights**) tai mitään muutoksia ei tehdä (**No adjustments**).



Kuvio 53. Ristiintaulukoinnin määrittelyikkunoita

Crosstabs-proseduurin tulostusta on esitetty kuviossa 54. χ^2 -riippumattomuustestin (= **Pearson Chi Square**) luotettavuuden takia ristiintaulukon odotettujen frekvenssien toteuttaa kaksi ehtoa: kaikkien pitäisi olla ykköstä suurempia, ja korkeintaan 20 % saa olla pienempiä kuin viisi. **Chi-Square Tests** -taulukon alapuolella on tietoja odotetuista frekvensseistä. Jos em. ehdot eivät ole voimassa, pitäisi ristiintaulukon kokoa pienentää luokkia yhdistämällä (tai käytetään χ^2 -riippumattomuustestin sijasta esimerkiksi Uskottavuussuhde-testiä (= **Likelihood Ratio**)).

χ^2 -riippumattomuustestin hypoteesit ovat

H_0 : populaatiossa ko. muuttujat ovat riippumattomia

H_1 : populaatiossa ko. muuttujan jakauma ei ole normaalijakauma

Miten opintoni edistyvät muihin saman alan opiskelijoihin nähden? * Tiedekunta Crosstabulation

			Tiedekunta		Total
			humanistinen	yhteiskuntatieteellinen	
Miten opintoni edistyvät muihin saman alan opiskelijoihin nähden?	keskimääräistä hitaammin	Count	47	16	63
		Expected Count	42,7	20,3	63,0
		% within Tiedekunta	21,9%	15,7%	19,9%
		Residual	4,3	-4,3	
	keskimääräisesti	Count	142	65	207
		Expected Count	140,4	66,6	207,0
		% within Tiedekunta	66,0%	63,7%	65,3%
		Residual	1,6	-1,6	
	keskimääräistä nopeammin	Count	26	21	47
		Expected Count	31,9	15,1	47,0
		% within Tiedekunta	12,1%	20,6%	14,8%
		Residual	-5,9	5,9	
Total	Count	215	102	317	
	Expected Count	215,0	102,0	317,0	
	% within Tiedekunta	100,0%	100,0%	100,0%	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4,751 ^a	2	,093
Likelihood Ratio	4,633	2	,099
Linear-by-Linear Association	4,308	1	,038
N of Valid Cases	317		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 15,12.

Symmetric Measures

	Value	Approx. Sig.
Nominal by Nominal Contingency Coefficient	,122	,093
N of Valid Cases	317	

a. Not assuming the null hypothesis.

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

Kuvio 54. Crosstabs -proseduurin tulostusta

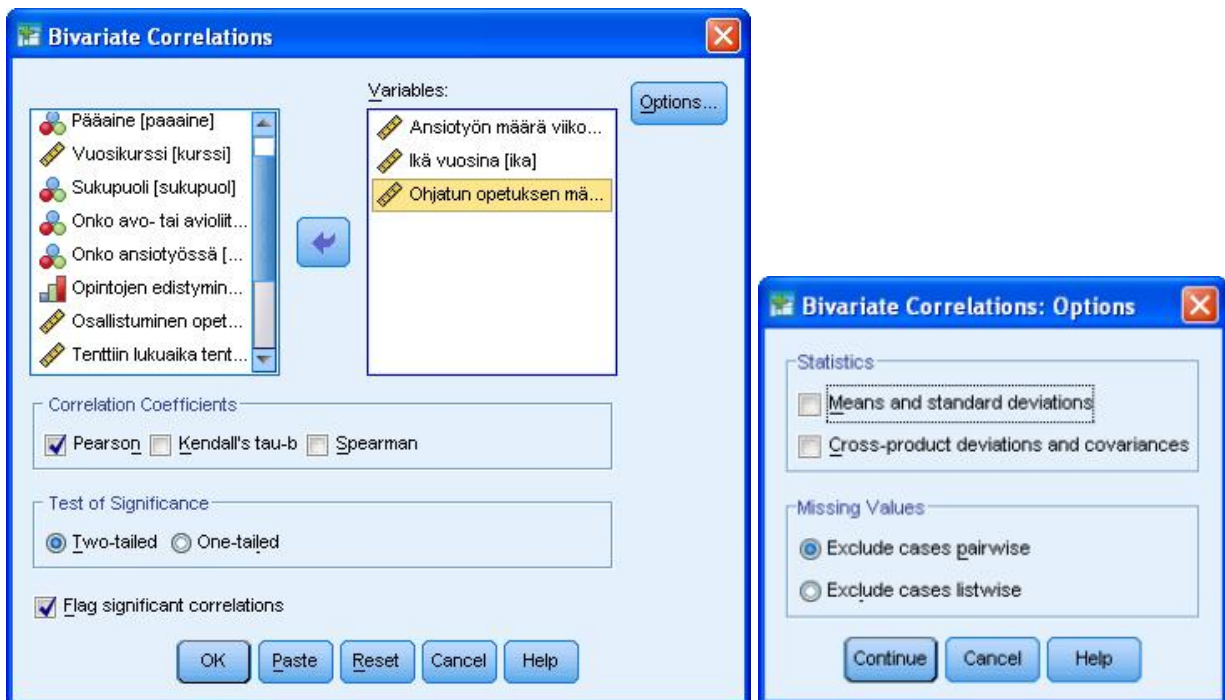
Tulostuksesta nähdään mm. seuraavaa. **Kontingenssikertoimen** arvo on 0.122, joten 317 opiskelijan joukossa tiedekunnan (nominaaliasteikko) ja edistymisen (järjestysasteikko) välillä on heikkoa riippuvuutta. Pienin odotettu frekvenssi on 15.12, joten χ^2 -riippumattomuustestin testisuure antaa luotettavan kuvan riippuvuuden voimakkuudesta. Pearsonin χ^2 -testin eli χ^2 -riippumattomuustestin testisuureen arvo on 4.751 ja sen havaittu merkitsevyytaso 0.093, joten opiskelijapopulaatiossa tiedekunnan ja opintojen edistymisen välinen riippuvuus ei ole tilastollisesti merkitsevää.

7. KORRELAATIOISTA

Muuttujien välisiä korrelaatiokertoimia saadaan muodostettua **Analyze**-valikon valinnan **Correlate**-proseduureilla. Korrelaatioita tutkittaessa kannattaa piirtää pisteparvikuvio, koska riippuvuussuhde voi vallita myös silloin, kun korrelaation arvo on lähellä nollaa.

Kahden muuttujan välisiä korrelaatioita saadaan proseduurilla **Bivariate**. Määrittelyikkunassa (kuvio 55) voidaan korrelaatioista (**Correlation Coefficients**) valita *lineaarinen korrelaatiokerroin (Pearson)*, joka sopii vain *määrällisille* muuttujille ja joka on käytetyin korrelaatiokerroin. Määrittelyikkunassa voidaan valita *järjestyskorrelaatioita (Kendall's tau-b ja Spearman)*, joita voi käyttää kaikilla muilla mitta-asteikoilla, mutta ei nominaliasteikolla. Korrelaatioiden merkitsevyyden (**Test of Significance**) voi testata joko kaksi- tai yksisuuntaisesti (**Two-tailed/One-tailed**). Tilastollisesti merkitsevät korrelaatiot voidaan merkitä (**Flag significant correlations**).

Options-valinnalla voidaan määritellä, mitä tunnuslukuja (**Statistics**) muuttujista halutaan, valittavana ovat keskiarvot ja keskihajonnat (**Means and standard deviations**) sekä tulo- ja neliösummat ja kovarianssit (**Cross-product deviations and covariances**). Puuttuvien havaintojen (**Missing Values**) käsittelytapa voidaan valita. **Exclude cases pairwise**-valinnalla laskennasta jätetään pois ne tilastoyksiköt, joilla on puuttuva arvo toisessa muuttujaparin muuttujista. **Exclude cases listwise**-valinnalla laskennasta jäävät pois ne tilastoyksiköt, joilla on puuttuva arvo jossakin **Variables**-listan muuttujista.



Kuvio 55. Korrelaatiokertoimien määrittelyikkunat

Pearsonin korrelaatiokerroin (kuvio 56) mittaa lineaarisen riippuvuuden voimakkuutta.. Korrelaatiokerroin on arvoltaan välillä $[-1, 1]$. Mitä suurempi korrelaation itseisarvo on, sitä voimakkaammasta lineaarisesta riippuvuudesta on kyse. Positiivisen korrelaation tapauksessa kyseessä on positiivinen lineaarinen riippuvuus, jolloin x :n arvojen kasvaessa y :n arvot kasvavat tasaisesti. Negatiivinen korrelaatiokertoimen arvo taas osoittaa negatiivista lineaarista riippuvuutta, jolloin x :n arvojen kasvaessa y :n arvot pienenevät tasaisesti.

Korrelaatiotestin kaksisuuntaisessa versiossa (**Test of Significance: Two-tailed**) hypoteesit ovat

H_0 : populaatiossa ko. muuttujat ovat lineaarisesti riippumattomia eli $\rho = 0$

H_1 : populaatiossa ko. muuttujat eivät ole lineaarisesti riippumattomia eli $\rho \neq 0$

Merkitsevyystestin tulos on luotettava, jos muuttujien yhteisjakauma on kaksiulotteinen normaalijakauma (eli muuttujista piirrettyssä pisteparvessa pisteet sijoittuvat pääsääntöisesti ellipsin sisään.)

Tulostuksesta nähdään mm, että ansiotyömäärän ja iän välinen korrelaatio on 0.484 ja sen havaittu merkitsevyystaso on 0.000, joten ko. muuttujien välillä on tilastollisesti erittäin merkitsevää lineaarista riippuvuutta. Ohjatun opetuksen määrän ja ansiotyömäärän välinen korrelaatio on -0.222 ja havaittu merkitsevyystaso 0.061, joten ko. muuttujien välillä ei ole tilastollisesti merkitsevää lineaarista riippuvuutta.

		Correlations		
		Ansiotyön määrä viikossa tunteina	Ikä vuosina	Ohjatun opetuksen määrä viikossa tunteina
Ansiotyön määrä viikossa tunteina	Pearson Correlation	1,000	,484**	-,222
	Sig. (2-tailed)	,	,000	,061
	N	72	72	72
Ikä vuosina	Pearson Correlation	,484**	1,000	-,093
	Sig. (2-tailed)	,000	,	,096
	N	72	322	320
Ohjatun opetuksen määrä viikossa tunteina	Pearson Correlation	-,222	-,093	1,000
	Sig. (2-tailed)	,061	,096	,
	N	72	320	320

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

Kuvio 56. Pearsonin korrelaatioon liittyvää tulostusta

Järjestyskorrelaatiot mittaavat monotonisen riippuvuuden voimakkuutta. Järjestyskorrelaatioita voidaan laskea ja tulkita *jo järjestysasteikkolisista* muuttujista. Muuttujien ei tarvitse olla normaalijakautuneita, kun merkitsevyystestausta tehdään. Järjestyskorrelaatiot ovat arvoltaan välillä $[-1, 1]$. Jos järjestyskorrelaation arvo on $+1$, ovat tilastoyksiköiden järjestykset samanlaiset eri muuttujilla. Jos järjestyskorrelaatio on -1 , ovat tilastoyksiköiden järjestykset päinvastaiset.

Järjestyskorrelaatioiden merkitsevyystestien kaksisuuntaisessa versiossa (**Test of Significance: Two-tailed**) hypoteesit ovat

H_0 : populaatiossa ko. muuttujat ovat monotonisesti riippumattomia

H_1 : populaatiossa ko. muuttujat eivät ole monotonisesti riippumattomia

Correlate-valinnan proseduurilla **Partial** voidaan muodostaa osittaiskorrelaatioita, joiden avulla voidaan eliminoida yhden tai usean muuttujan vaikutus kahden muuttujan lineaariseen korrelaatioon. Määrittelyikkuna on samantyyppinen kuin **Bivariate**-proseduurissa, mutta ruutuun **Controlling for** valitaan se muuttuja, jonka vaikutus halutaan eliminoida.

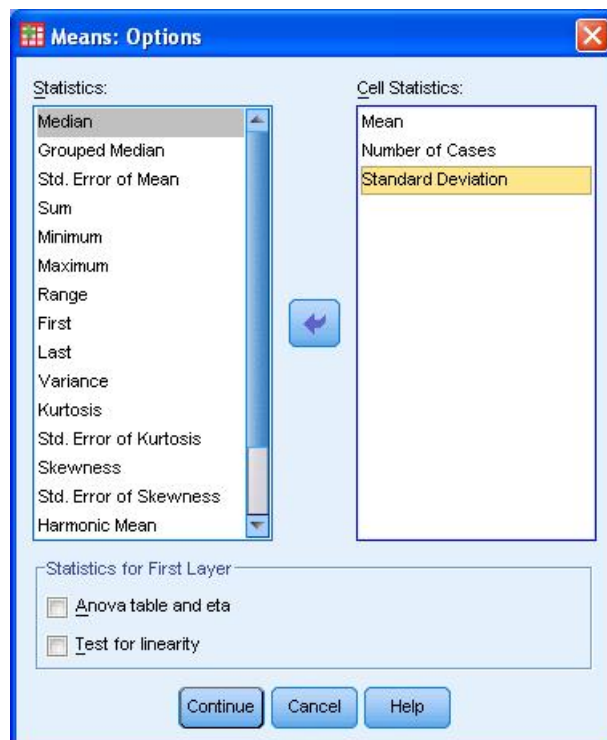
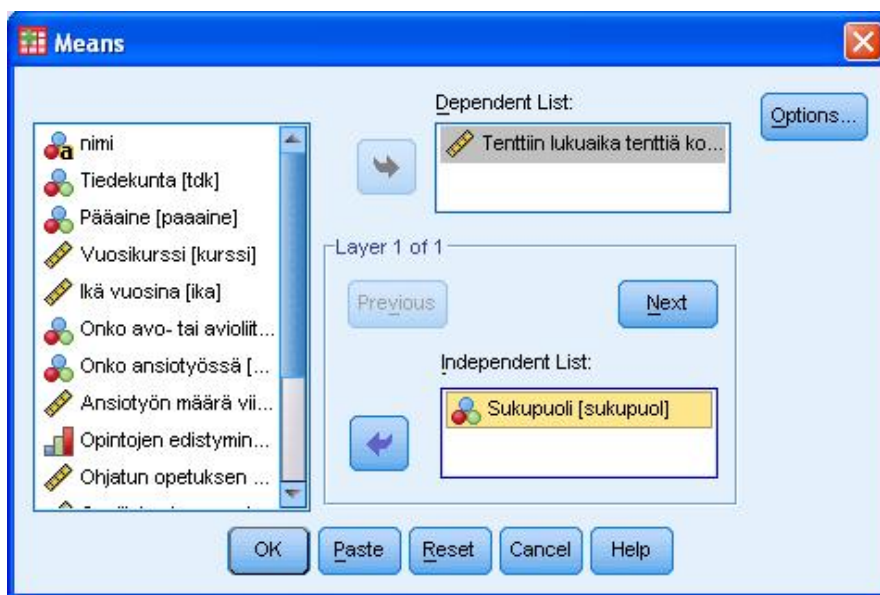
Correlate-valinnan **Distances**-proseduurin avulla voidaan muodostaa useita erilaisia tilastollisia tunnuslukuja, joiden avulla voidaan tutkia muuttujien tai tilastoyksiköiden samanlaisuutta (esim. tilastoyksiköiden etäisyyksiä), Näitä tunnuslukuja voidaan hyödyntää muissa prosedureissa (esim. faktorianalyysi, ryhmittelyanalyysi), kun halutaan analysoida isoja havainto- ja muuttujajoukkoja.

8. KESKIARVOJEN VERTAILUA JA KESKIARVOTESTEJÄ

Keskiarvojen laskemisen ja vertailun vaihtoehtoja sisältyy useita Analyze-valikon valintaan **Compare Means**.

8.1. Keskiarvojen ja muiden tunnuslukujen laskentaa ryhmille

Proseduurin **Means** määrittelyikkunan (kuvio 57) ruutuun **Independent List** valitaan ne muuttujat, jonka mukaan havaintoaineisto ryhmitellään (esimerkissä *Sukupuoli*) ja ruutuun **Dependent List** valitaan ne **määrälliset** muuttujat, joista lasketaan keskiarvot ja muut **Options**-valinnalla valitut tunnusluvut ryhmittäin (esimerkissä *Tenttiin ..*).



Kuvio 57. Means-proseduurin määrittelyikkunat

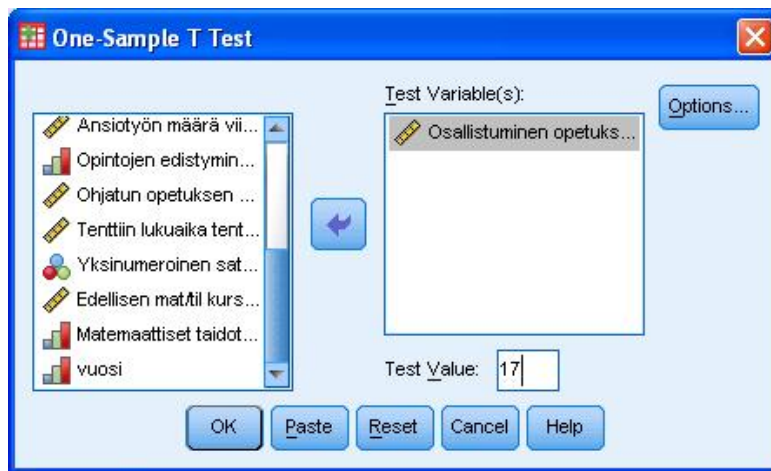
8.2. Yhden otoksen keskiarvotesti

Proseduurilla **One-Sample T Test** (kuvio 58) voidaan muodostaa *yhden otoksen keskiarvotesti*, jossa populaation varianssi oletetaan tuntemattomaksi ja valittujen **määrällisten** muuttujien (**Test Variables**, esimerkissä *Osallistuminen...*) keskiarvoja verrataan itse valittuun arvoon μ_0 (**Test Value**, esimerkissä 17). **Options**-valinnalla voidaan mm. määrittää halutun tasoinen luottamusväli (**Confidence Interval**).

Yhden otoksen keskiarvotestin oletuksena on se, että testattavan muuttujan jakauma on normaalijakauma (tai että jakauma on ainakin symmetrinen, jos otoskoko on iso). Otokseen olisi suotavaa olla ainakin 30. Testin hypoteesit ovat

H_0 : populaatiossa ko. muuttujan keskiarvo on μ_0 eli $\mu = \mu_0$

H_1 : populaatiossa ko. muuttujan keskiarvo ei ole μ_0 eli $\mu \neq \mu_0$



Kuvio 58. One-Sample T Test -proseduurin päämäärityikkuna

Kuviossa 59 on esitetty em. proseduurin tulostusta. Tulostuksen ensimmäinen taulukko sisältää testattavan muuttujan otostunnuslukuja (eli havaintomäärä 50, keskiarvo 15.2 tuntia, keskihajonta 4.994 tuntia, keskiarvon keskivirhe 0.706 tuntia). **One-Sample Test** -taulukossa on keskiarvotestin t-testisuuren arvo -2.549 ja testin havaittu merkitsevyytaso 0.014, joten populaatiokeskiarvon ei voida katsoa olevan yhtä suuri kuin väitetty 17 tuntia.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Osallistuminen opetukseen (tuntia viikossa)	50	15,20	4,994	,706

One-Sample Test

	Test Value = 17					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Osallistuminen opetukseen (tuntia viikossa)	-2,549	49	,014	-1,800	-3,22	-,38

Kuvio 59. One-Sample T Test -proseduurin tulostusta

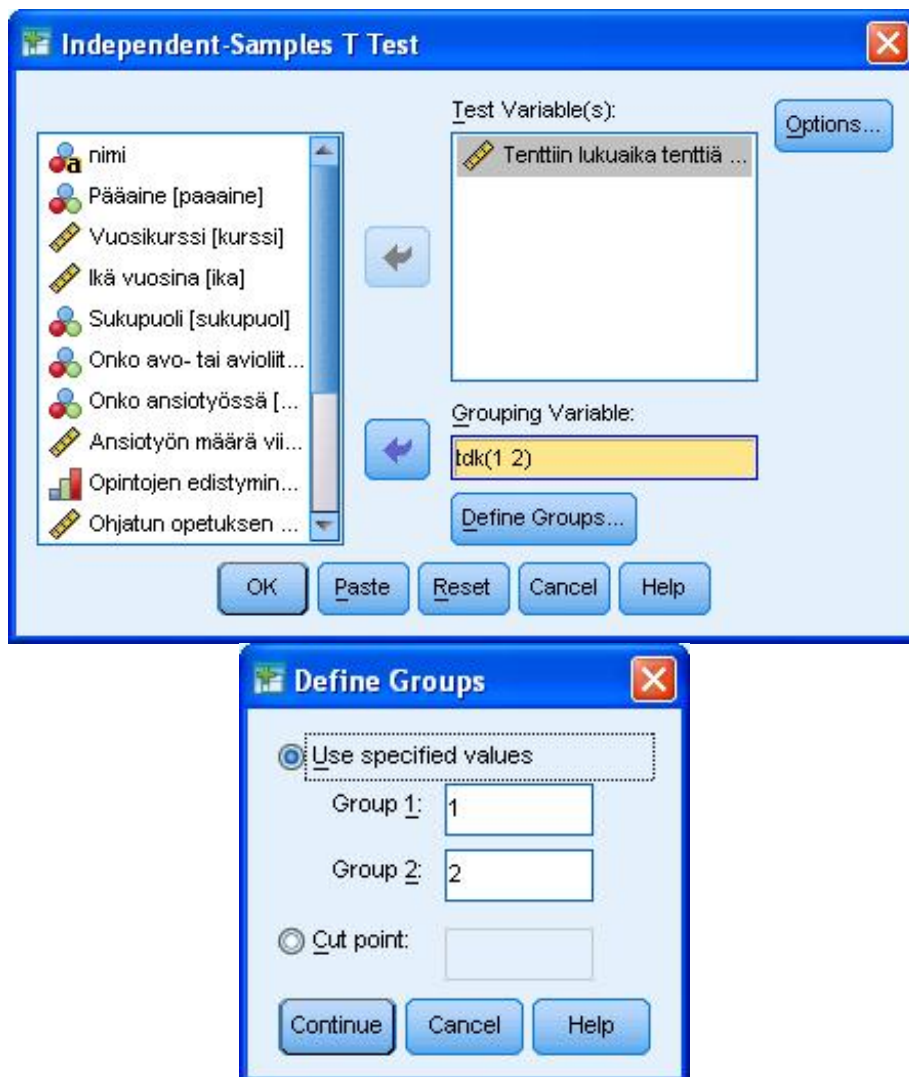
8.3. Kahden riippumattoman otoksen keskiarvotesti

Proseduurilla **Independent-Samples T Test** muodostetaan *kahden riippumattoman otoksen keskiarvotesti*, jossa populaatioiden varianssit oletetaan tuntemattomiksi ja valitun *määrällisen* muuttujan keskiarvot lasketaan eri otoksissa. Itse asiassa näitä keskiarvotestejä on kaksi erilaista, ja analysoija tekee tulostulkinnassa valinnan siitä, kumpaa testiä tulkitsee varianssistien tulosten perusteella. Määrittelyikkunassa (kuvio 60) ruutuun **Test Variables** valitaan testattavat määrälliset muuttujat (esimerkissä *Tenttiin ...*) ja kohtaan **Grouping Variable** se muuttuja, jonka perusteella havaintoaineisto jaetaan kahteen ryhmään (esimerkissä *tdk*). **Define Groups** -valinnalla määritetään, miten ryhmät muodostuvat. **Use specified values** -valinnalla ryhmittelevän muuttujan arvoista valita kaksi arvoa, joiden perusteella muodostuu kaksi ryhmää. **Cut point** -valinnalla valitaan ryhmittelevän muuttujan se arvo, jota pienemmistä arvoista muodostuu 1. ryhmä ja jota suuremmista (tai samankokoisista) arvoista muodostuu 2. ryhmä.

Kahden riippumattoman otoksen keskiarvotesti oletuksena on, että määrällisen muuttujan jakauma on eri ryhmissä normaalijakauma (tai symmetrinen, jos otokset ovat isoja). Kummankin otoksen koon olisi suotavaa olla ainakin 30. Testin hypoteesit ovat

H_0 : populaatioissa ko. muuttujan keskiarvot ovat yhtä suuret eli $\mu_1 = \mu_2$

H_1 : populaatioissa ko. muuttujan keskiarvot eivät ole yhtä suuret eli $\mu_1 \neq \mu_2$



Kuvio 60. Independent -Samples T Test -määrittelyikkunat

Group Statistics

Tiedekunta		N	Mean	Std. Deviation	Std. Error Mean
Tenttilukuaika	humanistinen	212	13,18	8,34	,57
	yhteiskuntatieteellinen	102	16,71	11,80	1,17

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	95% Confidence Interval of the Difference	
									Lower	Upper
Tenttilukuaika	Equal variances assumed	7,292	,007	-3,045	312	,003	-3,52	1,16	-5,80	-1,25
	Equal variances not assumed			-2,706	151	,008	-3,52	1,30	-6,09	-,95

Kuvio 61. Independent –Samples T Test -proseduurin tulostusta

Kuviossa 61 on em. proseduurin tulostusta. Tulostuksen ensimmäinen taulukko sisältää otoksittaista kuvailua tenttiinlukuaajasta.

Independent Samples Test -taulukossa on varsinaisen keskiarvotestin tulos ja myös *varianssitestin* tulos. Levenen varianssitestin (=varianssien yhtäsuuruustestin) hypoteesit ovat

H_0 : populaatioissa ko. muuttujan varianssit ovat yhtä suuret eli $\sigma_1^2 = \sigma_2^2$

H_1 : populaatioissa ko. muuttujan varianssit eivät ole yhtä suuret eli $\sigma_1^2 \neq \sigma_2^2$.

Levenen testisuureen arvo on 7.292 ja sen havaittu merkitsevyystaso on 0.007, joten populaatioiden (= humanistisen tdk:n kaikki opiskelijat ja yhteiskuntatieteellisen tdk:n kaikki opiskelijat) tenttilukuaajan varianssit eivät ole yhtäsuuret. Keskiarvotestituloksista valitaan siten rivi **Equal variances not assumed**, josta nähdään keskiarvotestin t-testisuureen arvon olevan -2.706 ja testin havaittu merkitsevyystaso on 0.008, joten populaatioiden keskiarvoilla on tilastollisesti merkitsevää eroa. Otoksissa Humanistisen tdk:n opiskelijoiden keskimääräinen tenttilukuaika on n. 13 tuntia ja yhteiskuntatieteellisen tdk:n opiskelijoilla se on siis selvästi enemmän: lähes 17 tuntia.

8.4. Kahden riippuvan otoksen keskiarvotesti

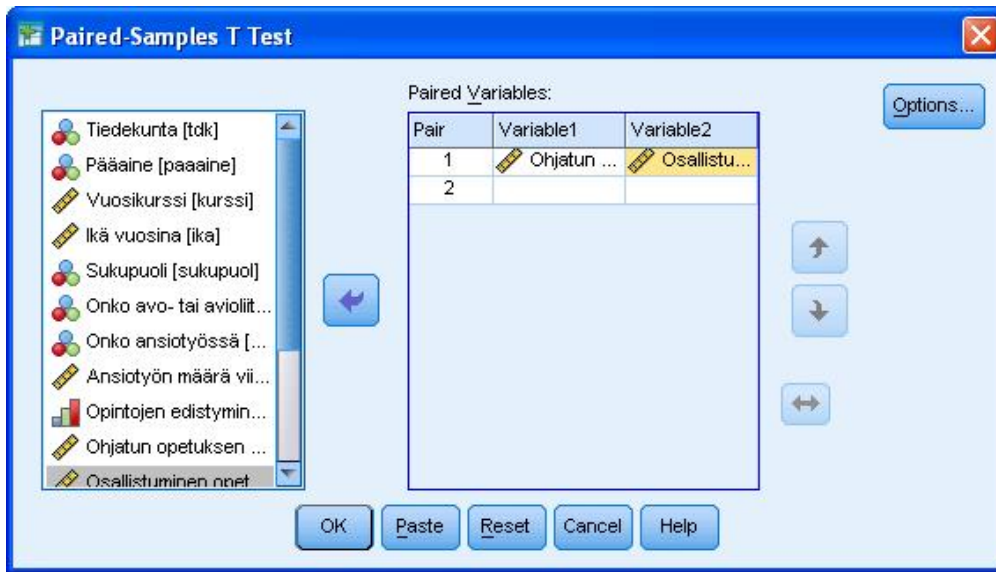
Paired-Samples T Test –proseduurilla voidaan muodostetaan *kahden riippuvan otoksen* (= ennen-jälkeen tai koe-vertailuryhmä) *keskiarvotesti*, jossa tutkitaan kahden *määrällisen* muuttujan muuttujarvoista muodostettuja erotuksia. Oletuksena on testissä se, että muuttujarvojen erotuksen jakauma on normaalijakauma (tai isossa otoksessa ainakin symmetrinen). Suotavaa olisi, että pareja olisi ainakin 30.

Ko. testin hypoteesit ovat

H_0 : populaatiossa ko. muuttujien keskiarvot ovat yhtä suuret eli $\mu_1 = \mu_2$

H_1 : populaatiossa ko. muuttujien keskiarvot eivät ole yhtä suuret eli $\mu_1 \neq \mu_2$

Määrittelyikkunassa (kuvio 62) ruutuun **Paired Variables** valitaan ne muuttujat pareina, joiden keskiarvoja halutaan vertailla.



Kuvio 62. Paired-Samples T Test -määrittelyikkuna

Kuviossa 63 on em. proseduurin tulostusta. Ensimmäinen taulukko sisältää tilastollisia tunnuslukuja kummastakin muuttujaparin muuttujasta. Toisessa taulukossa on tuloksia Pearsonin korrelaatiosta. Kolmannesta taulukosta nähdään keskiarvotestin t-testisuureen arvo olevan 13.442, ja sen havaittu merkitsevyytaso on 0.000, joten muuttujien Ohjatun opetuksen määrä ja Osallistuminen opetukseen keskiarvoilla on tilastollisesti erittäin merkitsevä ero siten, että Ohjatun opetuksen määrän keskiarvo on lähes 2 tuntia suurempi kuin muuttujalla Osallistuminen opetukseen.

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 Ohjatun opetuksen määrä	18,82	316	5,531	,311
Osallistuminen opetukseen	16,92	316	5,232	,294

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 Ohjatun opetuksen määrä & Osallistuminen opetukseen	316	,893	,000

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 Ohjatun opetuksen määrä - Osallistuminen opetukseen	1,896	2,507	,141	1,618	2,173	13,442	315	,000

Kuvio 63. Paired Samples T Test -proseduurin tulostusta

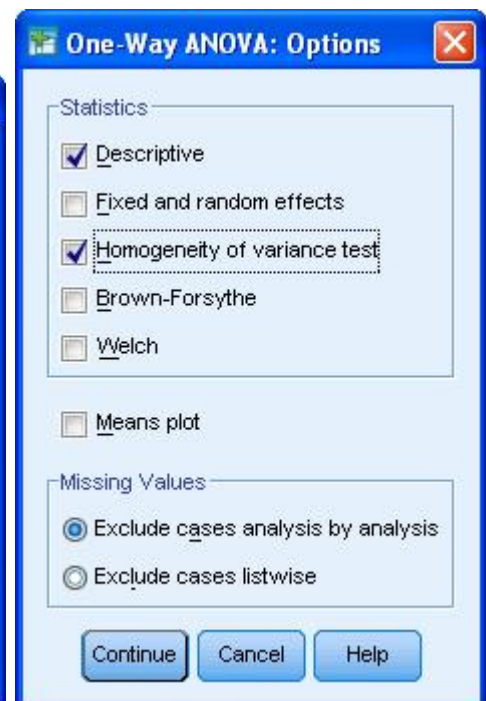
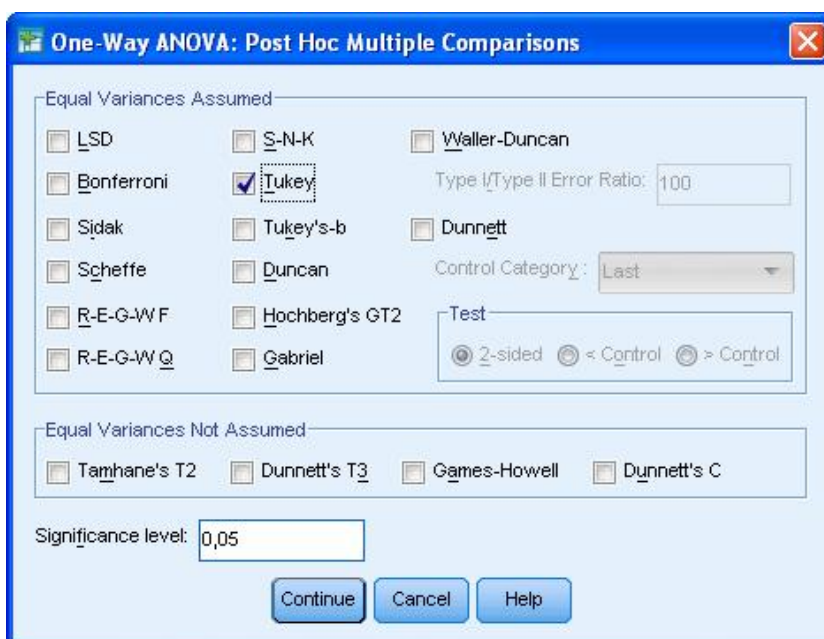
8.5. Varianssianalyysi

One-Way ANOVA -proseduurilla (kuvio 64) muodostetaan *yksisuuntainen varianssianalyysi* yhden selittävän muuttujan varianssianalyysi eli yhden *luokittelevan* muuttujan (**Factor**, esimerkissä *pääaine*) perusteella muodostetaan ainakin kolme otosta eli ryhmää, joiden välillä halutaan tutkia *määrällisten* muuttujien (**Dependent List**, esimerkissä *Osallistuminen opetuks...*) keskiarvojen yhtäsuuruutta.

Oletuksena perusvarienssianalyysissä on, että määrällisen muuttujan jakauma on kussakin ryhmässä on normaalijakauma (tai ainakin symmetrinen, jos otokset ovat isoja) vielä siten, että määrällisen muuttujan varianssit ovat eri ryhmissä yhtä suuret. Suotavaa olisi, että kunkin ryhmän koko on ainakin 30. Yksisuuntaisessa varianssianalyysissä hypoteesit ovat

H_0 : populaatioissa ko. muuttujan keskiarvot ovat yhtä suuret eli $\mu_1 = \mu_2 = \dots = \mu_k$

H_1 : ainakin yhdessä populaatioissa ko. muuttujan keskiarvo eroaa muista



Kuvio 64. One-Way ANOVA -proseduurin määrittelyikkunat

Options-valinnalla voidaan valita laskettavaksi määrällisen muuttujan tilastollisia tunnuslukuja (**Statistics: Descriptive**) selittävän muuttujan eri luokissa ja tutkia ryhmien varianssien yhtäsuuruutta (**Homogeneity of variance test**). **Means Plot** -valinnalla voidaan muodostaa viivakuvio ryhmäkeskiarvoista. (Ryhmäkeskiarvojen vertailussa kannattaa F-testin sijasta käyttää **Welch-** tai **Brown-Forsythe** -testiä, jos oletus ryhmävarienssien yhtäsuuruudesta ei toteudu.)

Jos varianssianalyysin tuloksena on se, että ryhmäkeskiarvoissa on eroa, voidaan vielä muodostaa ryhmille kontrasteja (**Contrasts**) tai **Post Hoc**-testejä, joiden avulla voidaan tutkia tarkemmin esimerkiksi sitä, mitkä ryhmät ovat keskenään samankaltaisia keskiarvojen osalta. Post Hoc -testejä voidaan muodostaa joko tapauksissa, joissa ryhmien varianssit ovat yhtä suuret (**Equal Variances Assumed**) tai erisuuret (**Equal Variances Not Assumed**).

Kuviossa 65 esitetään **One-Way ANOVA** -proseduurin perustulostusta. Tulostuksesta nähdään, että havaintoaineisto on jaettu yhdeksään ryhmään pääaineen mukaan. Kolmessa ryhmässä on tilastoyksiköiden määrä alle 30, joten varianssianalyysin suositukset eivät aivan toteudu. Kaikista ryhmistä on esitetty muuttujan Osallistuminen opetukseen tunnuslukuja, joita esitetään **Descriptives**-taulukossa. Esimerkiksi havaintoaineistossa on 13 opiskelijaa, joiden pääaine on englantia, ja näiden joukossa opetukseen osallistutaan keskimäärin 19.23 tuntia/viikko.

Descriptives

Osallistuminen opetukseen (tuntia viikossa)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
englanti	13	19,23	5,732	1,590	15,77	22,69	8	28
nykysuomi	31	16,42	4,856	,872	14,64	18,20	7	28
ruotsi	68	18,00	4,785	,580	16,84	19,16	10	32
saksa	48	19,21	4,053	,585	18,03	20,39	10	31
viestintätieteet	53	17,00	5,317	,730	15,53	18,47	8	33
hallintotiede	58	15,95	5,817	,764	14,42	17,48	2	32
julkisoikeus	32	13,63	4,164	,736	12,12	15,13	5	20
Sos. ja terveyshall.	8	13,00	5,318	1,880	8,55	17,45	6	20
aluetiede	4	14,00	5,888	2,944	4,63	23,37	8	20
Total	315	16,91	5,238	,295	16,33	17,49	2	33

Test of Homogeneity of Variances

Osallistuminen opetukseen (tuntia viikossa)

Levene Statistic	df1	df2	Sig.
,936	8	306	,487

ANOVA

Osallistuminen opetukseen (tuntia viikossa)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	967,394	8	120,924	4,839	,000
Within Groups	7646,118	306	24,987		
Total	8613,511	314			

Kuvio 65. One-Way ANOVA -proseduurin tulostusta

Test of Homogeneity of Variances -taulukossa tutkitaan ryhmien varianssien yhtäsuuruutta. Levenen testin hypoteesit ovat

H_0 : populaatioissa ko. muuttujan varianssit ovat yhtä suuret eli $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$

H_1 : ainakin yhdessä populaatioissa ko. muuttujan varianssi eroaa muista

Levenen testin havaittu merkitsevyystaso on 0.487, joten populaatiovariansseja voidaan pitää yhtä suurina. ANOVA-taulukossa on esitetty ryhmien sisäisiä ja välisiä vaihtelu- ja varianssilukuja, joiden perusteella tulee F-testisuureen arvoksi 4.839 ja sen havaituksi merkitsevyystasoksi 0.000, joka tarkoittaa sitä, että populaatioiden keskiarvoilla on tilastollisesti erittäin merkitsevä ero.

Kuviossa 66 esitetään **One-Way ANOVA** -proseduurin Post Hoc -testituloksista pieni osuus. Tukeyn testin perusteella voidaan esimerkiksi todeta, että englanninkielen pääaineopiskelijat osallistuvat opetukseen keskimäärin 2.811 tuntia enemmän kuin nykysuomen opiskelijat, mutta keskiarvojen ero ei ole tilastollisesti merkitsevä, koska testin havaittu merkitsevyystaso on 0.745. Sen sijaan englanninkielen ja julkisoikeuden osallistumiskeskiarvojen ero on jo 5.606 tuntia, joka on tilastollisesti melkein merkitsevä, koska havaittu merkitsevyystaso on 0.021.

Multiple Comparisons

Osallistuminen opetukseen (tuntia viikossa)

Tukey HSD

(I) Pääaine	(J) Pääaine	Mean Difference (I-J)	Std. Error	Sig.
englanti	nykysuomi	2,811	1,652	,745
	ruotsi	1,231	1,513	,996
	saksa	,022	1,563	1,000
	viestintätieteet	2,231	1,547	,881
	hallintotiede	3,282	1,534	,448
	julkisoikeus	5,606	1,644	,021
	sosiaali- ja terveyshallinto	6,231	2,246	,128
	aluetiede	5,231	2,858	,662

Kuvio 66. One-Way ANOVA -proseduurin Post Hoc -testitulostusta

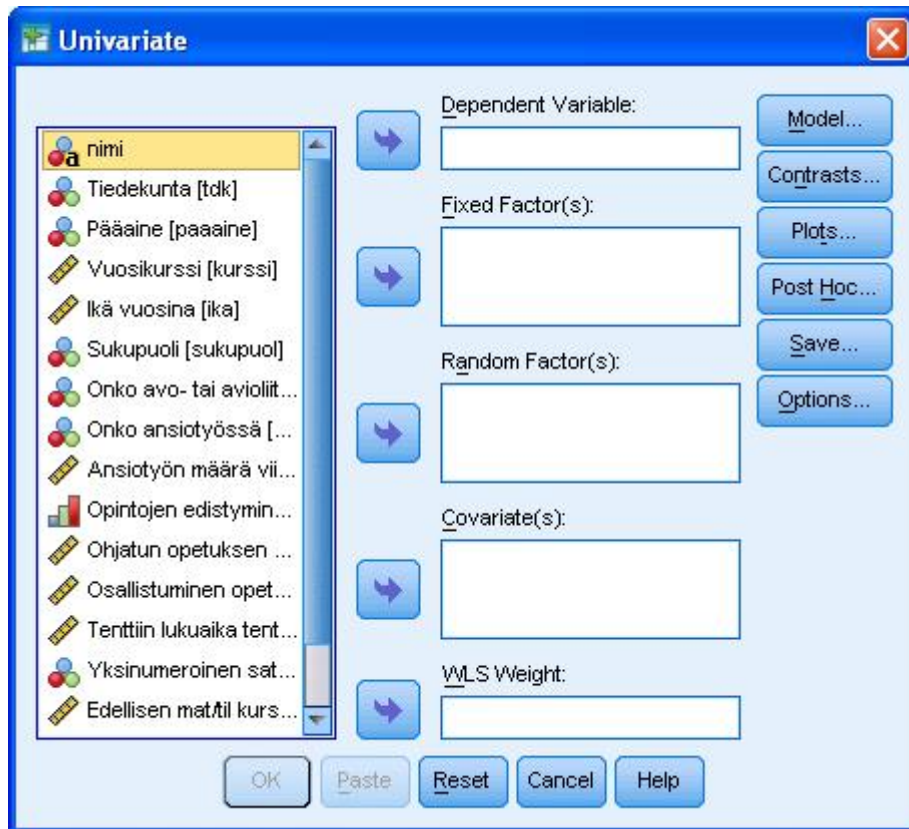
Analyze-valikon **General Linear Model** -valinnan proseduurilla **Univariate** voidaan suorittaa *kaksi-* tai *useampi-* **useampi-** **varianssianalyysi**, jossa selvitetään yksittäisten *laadullisten* muuttujien omavaikutusta ja muuttujaparien ja -ryhmien yhdysvaikutusta *määrälliseen* muuttujaan. Oletuksena tässäkin varianssianalyysissä on se, että määrällisen muuttujan jakauma on normaalijakauma (tai ainakin symmetrinen, jos otokset ovat isoja) jokaisessa tarkasteltavassa ryhmässä siten, että varianssit eri ryhmissä ovat kuitenkin yhtä suuret. Suositeltavaa on, että ryhmissä on ainakin 30 havaintoa.

Määrittelyikkunan (kuviokuva 67) kohtaan **Dependent Variable** valitaan se määrällinen muuttuja, jonka keskiarvoja halutaan vertailla eri ryhmissä. Ryhmittelevä muuttuja valitaan kohtaan **Fixed Factors**, jos oletetaan, että se ei voi saada muita arvoja kuin ko. havaintoaineistossa on. Jos ryhmittelevä muuttuja valitaan kohtaan **Random Factors**, niin oletetaan, että sen arvot ovat otos isommasta arvojoukosta. Jos selittävä muuttuja on määrällinen, valitaan se kohtaan **Covariates**, jolloin tehdään ns. kovarianssianalyysi.

Lisämäärittelyllä **Model** päästään tekemään malliin liittyviä määrittelyksiä. Valinnalla **Full factorial** otetaan malliin mukaan kaikki oma- ja yhdysvaikutukset. **Custom**-valinnalla valitaan halutut vaikutukset

siirtämällä **Factors & Covariates** –ruudusta nuolipainikkeen avulla **Model**-ruutuun muuttujia yksitellen ja kerralla useita. **Sum of Squares** -kohdassa voidaan valita yhdys- ja omavaikutusten testausmenetelmä.

Malliin liittyviä kuvia saadaan muodostettua **Plots**-lisämäärityillä. **Options**-lisämäärityillä saadaan tulostettua mm. ryhmien tunnuslukuja ja voidaan testata varianssianalyysin oletusten voimassaoloa.



Kuvio 67. Univariate-proseduurin päämäärityikkuna

Esimerkiksi jo kaksisuuntaisessa varianssianalyysissä testattavia hypoteeseja voidaan asettaa kolmella eri tavalla.

Voidaan tutkia mm. onko yksittäisellä laadullisen muuttujan A (tai B) eri ryhmissä määrällisen muuttujan keskiarvot yhtä suuret, jolloin hypoteesit ovat

H_0 : A:n (B:n) perusteella muodostuvissa populaatioissa määrällisen muuttujan keskiarvot ovat yhtä suuret eli muuttujalla A (B) ei ole omaa vaikutusta (=omavaikutusta) määrälliseen muuttuajaan

H_1 : ainakin yhdessä A:n (B:n) perusteella muodostuvassa populaatiossa määrällisen muuttujan keskiarvo eroaa muista eli muuttujalla A (B) on omavaikutusta määrälliseen muuttuajaan.

Voidaan tutkia myös, onko laadullisilla muuttujilla A ja B yhdessä vaikutusta määrälliseen muuttuajaan, jolloin hypoteesit ovat

H_0 : A:n ja B:n perusteella muodostuvissa populaatioissa määrällisen muuttujan keskiarvot ovat yhtä suuret eli muuttujilla A ja B ei ole yhteistä vaikutusta (=yhdysvaikutusta = yhteisvaikutusta) määrälliseen muuttuajaan

H_1 : ainakin yhdessä A:n ja B:n perusteella muodostuvassa populaatiossa määrällisen muuttujan keskiarvo eroaa muista eli muuttujilla A ja B on yhdysvaikutusta määrälliseen muuttuajaan.