

Opas SAS Enterprise Guiden käyttäjälle

EG versio 4.1

Timo Hurme



MTT:n selvityksiä 140
115 s.

Opas SAS Enterprise Guiden käyttäjälle

EG versio 4.1

Timo Hurme

ISBN 978-952-487-102-0 (Painettu)
ISBN 978-952-487-101-3 (Verkkajulkaisu)
ISSN 1458-509X (Painettu)
ISSN 1458-5103 (Verkkajulkaisu)
www.mtt.fi/mtts/pdf/mtts140.pdf

Copyright

MTT

Timo Hurme

Julkaisija ja kustantaja

MTT, 31600 Jokioinen

Jakelu ja myynti

MTT, Tietohallinto, 31600 Jokioinen

sähköposti julkaisut@mtt.fi

Puhelin (03) 4188 2327, telekopio (03) 4188 2339

Julkaisuvuosi

2007

Kannen kuva

Jii Roikonen

Painopaikka

Tampereen Yliopistopaino Oy Juvenes Print

Opas SAS Enterprise Guiden käyttäjälle

Timo Hurme

MTT, Palveluyksikkö, Menetelmäpalvelut, M-talo, 31600 Jokioinen, timo.hurme@mtt.fi

Tiivistelmä

SAS Enterprise Guide on tilastollinen ohjelmisto, joka vaatii toimiakseen myös tavallisen SAS-ohjelmiston. Enterprise Guide on helppokäyttöinen vaihtoehto SAS-ohjelmiston monipuolisten ominaisuuksien hyödyntämiseen.

Enterprise Guidella voidaan muokata valikkopohjaisesti tutkimusaineistoa, piirtää kuvia, laskea kuvailevia tunnuslukuja ja tehdä tilastollisia analyyskejä. Tässä oppaassa tutustutaan Enterprise Guiden käyttöympäristöön, datojen käsittelyyn sekä aineiston alustavaan tarkasteluun muun muassa kuvien ja tilastollisten tunnuslukujen avulla.

Opas on suunnattu lähinnä maa- ja elintarviketalousalan tutkijoille ja muulle tutkimusväelle. Siitä on kuitenkin varmasti hyötyä kaikille SAS Enterprise Guiden käyttäjille.

Avainsanat: SAS, Enterprise Guide, tilastollinen ohjelmisto, kuvaileva tilastotiede, aineiston tarkastelu

Sisällysluettelo

1	Johdanto	7
2	Oppaassa käytettävät datat	7
3	Enterprise Guiden käyttöympäristö	9
4	Enterprise Guiden ohje.....	11
5	Tiedostojen tyypit	13
6	Aineiston käyttöönotto.....	13
6.1	Kirjastoviitteen luominen	13
6.2	Kirjastoviitteen asettaminen oletukseksi	16
6.3	Aineiston tallennusmuoto	16
6.4	Aineiston syöttäminen Enterprise Guidessa	17
6.5	Excel-tiedoston avaaminen Enterprise Guideen	18
6.6	Erikoismuuttujien sisään lukeminen ja ongelmatapaukset	22
6.7	SAS-datan avaaminen Enterprise Guideen.....	24
6.8	Datojen tallentaminen Enterprise Guidesta	24
6.9	Projektin tallentaminen	25
7	Datan muokkaus.....	26
7.1	Datan muuttaminen muokkaustilaan	26
7.2	Muuttujien perusominaisuuksien määrittely.....	27
7.3	Muuttujan arvojen tulostusmuotojen eli formaattien käyttäminen	27
7.4	Datan muokkauksen työkalut.....	30
7.5	Muuttujien muunnokset ja uusien muuttujien tekeminen.....	32
7.6	Muuttujien luokittelu ja luokkien yhdistäminen.....	35
7.7	Muuttujan luominen aritmeettisella operaatiolla	38
7.8	Muuttujien poistaminen	39
7.9	Muuttujien järjestyksen vaihtaminen datassa	39
7.10	Datan havaintorivien järjestäminen	39
7.11	Osa-aineiston eli tiettyjen havaintorivien valitseminen.....	40
7.12	Kahden tai useamman datan yhdistäminen.....	42
7.13	Datan kääntäminen	45
7.14	Muuta datojen käsittelyyn liittyvää.....	47

8	Datan tarkastelu Graph-valikon kuvatyökaluilla.....	48
8.1	Sirontakuvio (Scatter Plot).....	49
8.2	Pylväskuvio (Bar Chart).....	51
8.3	Viivakuva (Line Plot).....	54
8.4	Laatikko-janakuvio (Boxplot).....	58
8.5	Piirakkakuvio (Pie Chart).....	61
8.6	Wizard-kuvatyökalut.....	61
8.7	Kuvien tulostusmuoto ja yleisasetukset	68
9	Datan tarkastelu Describe- ja Analyze-valikoiden työkaluilla.....	70
9.1	Datan listaukset	70
9.2	Datan luonnehdinta -työkalu (Characterize Data).....	72
9.3	Jatkuvan muuttujan jakauman tarkastelu	73
9.3.1	Distribution Analysis -työkalu.....	73
9.3.2	Summary Statistics -työkalu	75
9.3.3	Summary Tables -työkalu.....	77
9.3.4	Tunnuslukujen taulukointi wizard-työkaluilla.....	79
9.4	Kategorisen muuttujan jakauman tarkastelu	83
9.5	Kahden muuttujan yhteyden tarkastelu.....	85
9.5.1	Kahden numeerisen muuttujan välinen yhteys.....	85
9.5.2	Kahden kategorisen muuttujan välinen yhteys.....	89
10	SAS-koodin käyttäminen	92
10.1	Koodin kirjoittaminen Enterprise Guidessa tai tuominen muualta.....	92
10.2	Koodin generoiminen Enterprise Guidessa.....	92
10.3	Koodin tallentaminen SAS-kooditiedostoon	93
11	Tulostuksen ja projektin hallintaa	94
11.1	Tulostusasetukset	94
11.1.1	Tulostuskohteen valinta.....	94
11.1.2	Tulostustyylin valinta	94
11.1.3	Muut tulostusasetukset	96
11.2	Yksittäisen tulosteen hallinta ja jatkokäyttö	97
11.2.1	Kohteen ja tyylin vaihtaminen yksittäiselle tulosteelle	97
11.2.2	Yksittäisen tulosteen tallentaminen tiedostoon	98
11.2.3	Tulosteen tai datan siirtäminen toiseen ohjelmaan.....	98
11.2.4	Valmiin kuvan muokkaaminen.....	99
11.2.5	Selitetekstien ja formaattien lisääminen yksittäiselle tulosteelle	100

11.2.6 Otsikoiden ja alaviitteiden muokkaaminen	100
11.3 Yksittäisten tulosten kokoaminen yhteen	101
11.3.1 HTML dokumentti	101
11.3.2 SAS Raportti	102
11.4 Projektin hallinta.....	103
11.4.1 Projektin perustoiminnot	103
11.4.2 Projektin pitäminen järjestyksessä	104
11.4.3 Tehtäväsolmujen käsittely	106
11.5 Makromuuttajat (Parameters)	107
11.5.1 Makromuuttujien määrittely	107
11.5.2 Makromuuttujien käyttäminen	110
11.6 Virhetilanteet	115
12 Kirjallisuus	115

1 Johdanto

Tämä opas on suunnattu ensisijaisesti maa- ja elintarviketalouden tutkijoille ja muulle tutkimushenkilökunnalle. Lähestymistapa on sellainen, että materiaali soveltuu erityisesti aloitteleville SAS Enterprise Guiden (EG) käyttäjille. Opas perustuu suurelta osin MTT:llä lukuisia kertoja pidettyjen Enterprise Guide peruskurssien kurssimateriaaleihin. Opas on kuitenkin varmasti käyttökelpoinen myös muilla aloilla toimiville SAS Enterprise Guiden käyttäjille sekä hieman kokeneemmille EG-käyttäjille.

SAS Enterprise Guide on valikkopohjainen hiirellä ohjattava SAS-ohjelmiston käyttöliittymä, joka madaltaa SASin käyttöönottokynnystä. EG:ssä hiirellä tehtävät valinnat generoivat taustalla SAS-koodia, joka suoritetaan automaattisesti perinteisessä SASissa. Tulokset palautuvat esitettäväksi Enterprise Guidessa ja niiden ulkoasu noudattaa viimeisintä SAS-tekniikkaa. Niinpä Enterprise Guide ei ole itsenäinen ohjelma, vaan se vaatii toimiakseen myös perus-SASin. Tässä oppaassa käytetään SAS Enterprise Guiden versiota 4.1 yhdessä SAS 9.1.3:n kanssa. Enterprise Guide toimii ainoastaan Windows-ympäristössä, ja tässä oppaassa rajoitutaan tilanteeseen, jossa molemmat – EG ja tavallinen SAS – on asennettu samalle paikalliselle tietokoneelle (local).

Enterprise Guidella voidaan muokata valikkopohjaisesti tutkimusaineistoa, piirtää kuvia, laskea kuvailevia tunnuslukuja sekä tehdä tilastollisia analyysejä. Perinteinen SAS-käyttö voidaan myös yhdistää EG:n helppouteen, sillä EG:ssa voidaan käyttää SAS-koodia. EG:n vahvuutena on myös tulosteiden, kuten kuvien ja taulukoiden, helppo siirrettävyys muihin ohjelmiin, esimerkiksi Wordiin. Oppaassa keskitytään datan käyttöönottoon ja muokkaukseen sekä datan havainnollistamiseen kuvien ja tunnuslukujen avulla. Lisäksi käydään läpi Enterprise Guiden käyttöympäristöä ja yleisiä ominaisuuksia.

Muutamia Enterprise Guidessa käytettyjä englanninkielisiä ilmauksia on syytä täsmentää. EG:ssä sanoilla ”Column” ja ”Variable” tarkoitetaan samaa asiaa. Näillä tarkoitetaan muuttujaa ja sen arvoja eli saraketta datamatriisissa. Myös sanat ”Rows” ja ”Observations” tarkoittavat keskenään samaa asiaa eli havaintoa (riviä) datamatriisissa. Ilmaukset ”SAS data table” ja ”SAS data set” puolestaan molemmat tarkoittavat SAS-dataa.

Vahlberg ym. (2006) sekä Helenius ja Katajisto (2005) ovat julkaisseet suomenkieliset kirjat Enterprise Guiden versiosta 3.0. Slaughter ja Delwiche (2006) ovat julkaisseet englanninkielisen kirjan Enterprise Guiden versiosta 4.1. Nämä kirjat ovat tämän oppaan päälähteitä.

2 Oppaassa käytettävät datat

Kaikki kolme oppaassa käytettävää aineistoa löytyvät internetistä (<http://www.mtt.fi/mtts/pdf/egdata/EGdatat.zip>). Aineistoja tarvitaan oppaan harjoitustehtäviä tehtäessä, mutta harjoitusmateriaalina voi vaihtoehtoisesti käyttää myös muita vastaavan tyyppisiä aineistoja. Käytettävät aineistot kannattaa tallentaa omalle tietokoneelle esimerkiksi kansioon C:\SAS\EGKURS, jota käytetään oppaan esimerkkien tallennuspaikkana.

Gomez-aineisto

Gomez-aineisto on Gomezin ja Gomezin (1984) kirjassa esitetty osaruutukokeen tulosaineisto. Kokeessa on tutkittu riisin satoa eri lajikkeilla ja tyyppitasoilla. Pääruututekijä on

typpitaso, jonka luokat (kuusi typpitasoa) on arvottu pääruutuihin satunnaistettujen täydellisten lohkojen koekaavion mukaisesti. Lohkoja kokeessa on kolme. Osaruututekijä on riisilajike, jonka luokat (neljä eri lajiketta) on täydellisesti satunnaistettu osaruutuihin erikseen jokaisen pääruudun sisällä. Aineisto on tallennettu Excelissä nimellä **GomezData.xls**. Aineistosta löytyvät seuraavat muuttujat:

- Variety (lajike):
IR8
IR5
C4_63
Peta
- Nitrogen (typpi kg/ha):
0
60
90
120
150
180
- Rep (lohko):
1
2
3
- Yield (sato kg/ha)

Viljelijä 2006 -aineisto

Viljelijä 2006 -aineisto perustuu keksittyyn kyselytutkimukseen, jossa 54 tilanomistajalta on kysytty erilaisia maatalaan ja itse viljelijään liittyviä kysymyksiä. Aineisto on tallennettu Excelissä nimellä **Viljeliija2006.xls**. Aineistosta löytyvät seuraavat muuttujat:

- id (havainnon tunnistenumero)
- nimi (viljelijän etunimi)
- sukup (viljelijän sukupuoli)
1="Mies"
2="Nainen"
- peltoa (peltoa hallinnassa, ha)
- nautoja (nautoja omistuksessa, kpl)
- metsaa (metsää omistuksessa, ha)
- tuotsuun (päätuotantosuunta)
1="Peltoviljely"
2="Kotieläimet"
- paino (viljelijän paino, kg)
- pituus (viljelijän pituus, cm)
- ika (viljelijän ikä vuosina)
- tulot (viljelijän vuositulot euroina)
- ajanmuk (tilan ajanmukaisuus)
1="Huippumoderni"
2="Ajan hermolla"
3="OK"
4="Vähän perässä"
5="Alkukantainen"

- tulevais (tilan tulevaisuudennäkymät)
 - 1="Hyvät"
 - 2="OK"
 - 3="Huonot"

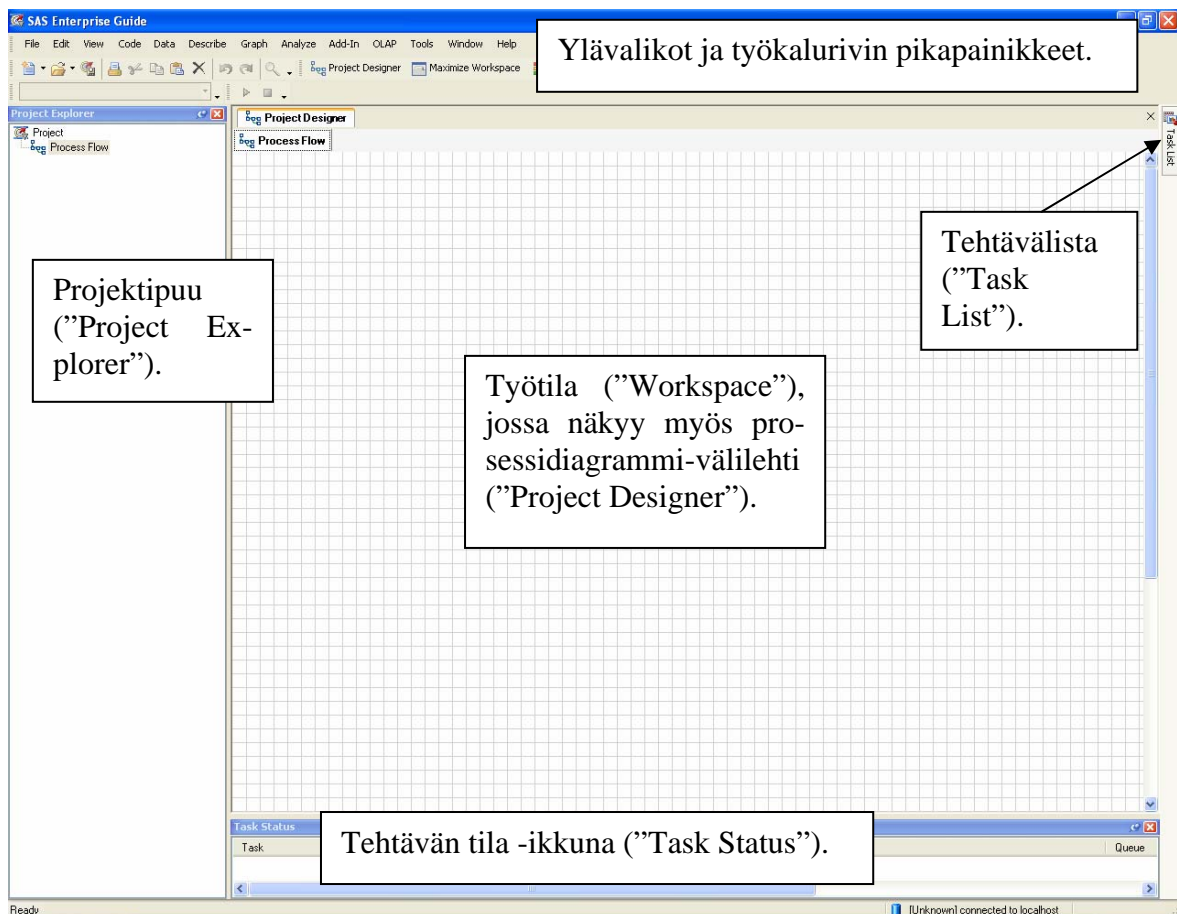
Eläinten painot -aineisto

Eläinten painot -aineisto on keksitty kokeellinen toistomittausaineisto. Kymmenen saman pentueen kettua on heinäkuussa tapahtuneen vieroituksen yhteydessä täydellisesti satunaistettu kahteen eri käsittelyryhmään. Kettujen painon kehitystä on seurattu elokuusta joulukuuhun neljänä eri ajankohtana. Aineisto on tallennettu SASissa nimellä **Ketut-Data.sas7bdat** sekä Excelissä nimellä **KetutData.xls**. Aineistosta löytyvät seuraavat muuttujat:

- ID (ketun tunnistenumero)
- ryhmä (käsittelyryhmä)
 - A
 - B
- elokuu (ketun paino elokuun mittauksessa, kg)
- syyskuu (ketun paino syyskuun mittauksessa, kg)
- lokakuu (ketun paino lokakuun mittauksessa, kg)
- joulukuu (ketun paino joulukuun mittauksessa, kg)

3 Enterprise Guiden käyttöympäristö

Avattaessa Enterprise Guide, perusnäky on seuraavanlainen:



Perusnäkyvä sisältää työtilan ("Workspace"), johon avautuvat käyttöön otettavat aineistot ja suoritettujen tehtävien tulosteet. Työtilan hallitsevana osiona on prosessidiagrammi-välilehti ("Project Designer"), joka palvelee samaa tarkoitusta kuin projektipuu. Molempiin kertyy istunnon aikana tehdyistä tehtävistä jälki, joka voidaan myös tallentaa ja ottaa käyttöön uudessa istunnossa. Jälki eli EG-projektikokonaisuus sisältää erillisiä kuvakkeita eli solmuja (pääasiassa tehtäväsolmuja ja tulossolmuja) suoritetuista toimenpiteistä, joita pääsee tarkastelemaan ja muokkaamaan tupla-klikkaamalla niitä. Hiiren oikealla painikkeella saa listan kyseiseen kuvakkeeseen liittyvistä jatkotoimenpiteistä. Oppaassa käytetään pääasiassa prosessidiagrammia projektin käsittelyyn. Edellä mainittujen tehtävä- ja tulossolmujen lisäksi Enterprise Guiden projekti pitää sisällään muun muassa dataja, koodeja, logitiedostoja ja muistiinpanoja toisiinsa linkitettyinä. Projektia voi prosessidiagrammissa myös jakaa useammalle erilliselle välilehdelle (Process Flows). Uusia välilehtiä saa valitsemalla "File" -> "New" -> "Process Flow".

Prosessidiagrammi, datat ja tulosteet ovat kukin omina välilehtinään työtilassa. Useampi välilehti saadaan samanaikaisesti näkyviin painamalla hiiren oikealla painikkeella välilehden otsakepalkkia ja valitsemalla "New Horizontal Tab Group" tai "New Vertical Tab Group". Huomaa, että jos prosessidiagrammiin on tehty useampi välilehti, niin ne ovat alivälilehtinä varsinaisen prosessidiagrammivälilehden alla. Kätevin tapa löytää prosessidiagrammi mahdollisten monien tulosteiden ja datojen joukosta, on klikata prosessidiagrammin omaa työkalurivin pikapainiketta "Project Designer".

Perusnäkyvä sisältää myös Tehtävän tila -ikkunan ("Task Status"), jossa näkyy suoritettavan tehtävän eteneminen.

Varsinaisia toimenpiteitä voidaan tehdä tehtävälistan ("Task List") tai ylävalikoiden avulla. Oletusasetuksilla tehtävälista on piilotettu oman pikalinkkinsä alle (ks. kuva edellä). Tässä oppaassa tehtävät suoritetaan pääasiassa ylävalikoiden kautta. Ylävalikoista löytyy seuraavia toimintoja:

- File:** Tyypillinen "File"/"Tiedosto"-valikko, josta löytyy muun muassa tiedostojen avaamiseen, tallentamiseen ja tulostamiseen liittyviä toimintoja.
- Edit:** Tyypillinen "Edit"/"Muokkaa"-valikko, josta löytyy muun muassa kopioi-, leikkaa-, liitä-, etsi- ja korvaa-toimintoja.
- View:** View-valikossa voidaan määritellä näkyvissä olevat ikkunat ja työkalurivit.
- Code:** Code-valikko liittyy lähinnä perinteisen SAS-koodin käyttöön. Muun muassa koodi-ikkunaan kirjoitettu SAS-koodi voidaan suorittaa tämän valikon kautta. Kuitenkin myös tavalliset tehtäväsolmut voidaan suorittaa täältä.
- Data:** Tästä valikosta löytyvät aineiston muokkaamiseen liittyvät toiminnot. Tämän valikon työkaluja käydään oppaassa läpi tarkemmin.
- Describe:** Tästä valikosta löytyy aineiston kuvailuun liittyviä toimintoja, kuten tunnuslukujen laskemista ja frekvenssitaulujen muodostamista. Tämänkin valikon työkaluihin paneudutaan oppaassa.
- Graph:** Valikosta löytyy aineiston graafiseen tarkasteluun liittyviä toimintoja, joihin tutustutaan oppaassa tarkemmin.

Analyze: Analyze-valikossa ovat toiminnot, joiden avulla voidaan suorittaa aineiston tilastollisia analyysejä.

Tools: Valikosta voidaan vaikuttaa moniin Enterprise Guiden yleisiin asetuksiin.

Help: Valikosta löytyy Enterprise Guiden Help eli ohje

Tulostusta tutkiessa saattaa tulostusikkuna tuntua turhan pieneltä. Tällöin voi käyttää ”Maximize Workspace” -toimintoa, joka löytyy omana painikkeenaan työkaluriviltä. Sen avulla tulostusikkuna suurentuu lähes koko ruudun kokoiseksi ja muut ikkunat menevät sivuun, pikalinkkien taakse. Muita ikkunoita saa kätevästi näkyviin viemällä hiiren kohdeikkunan pikalinkin päälle.

Yksittäisen ikkunan voi myös siirtää sivuun klikkaamalla ×-symbolin vieressä olevaa nastasymbolia. Tällöin ikkuna menee samanlaisen pikalinkin taakse kuin ”Maximize Workspace” -toiminnossa. Klikkaamalla nastan takaisin pystyasentoon ikkuna palautuu ennalleen.

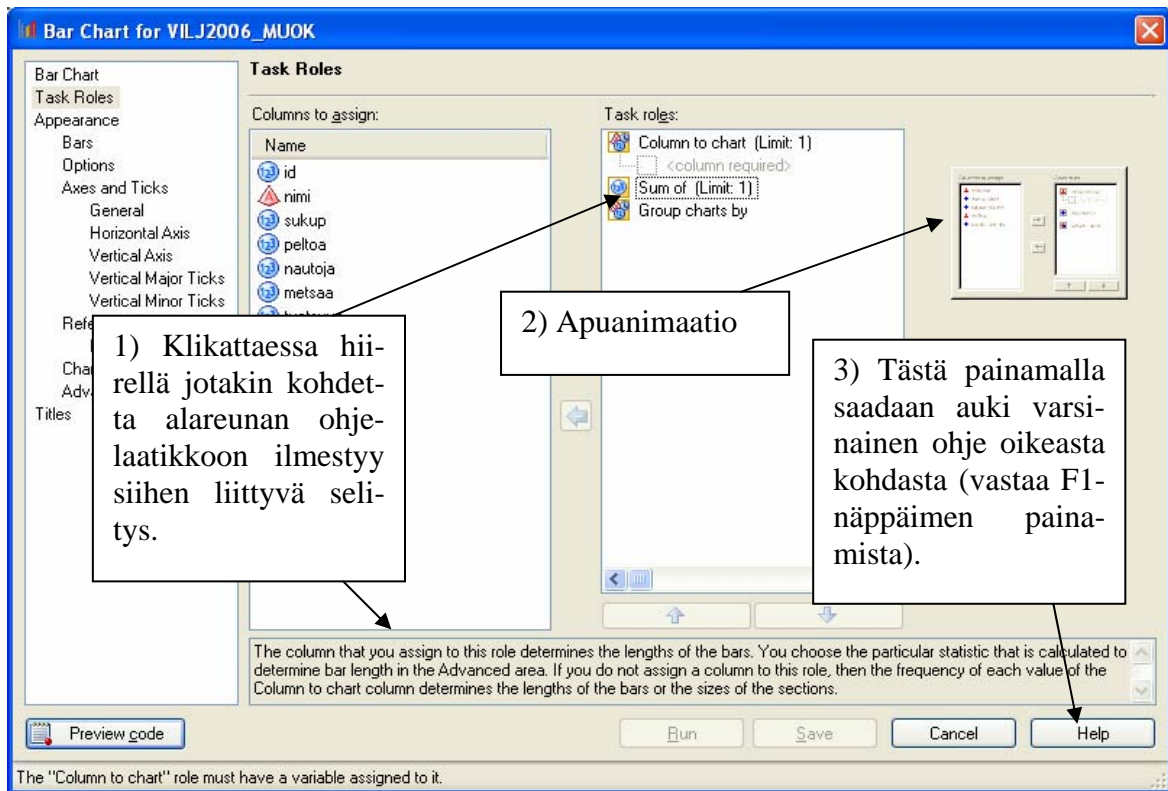
Ikkunoiden tai ikkunaryppäiden kokoa, muotoa ja paikkaa voi muokata tarttumalla niistä tai niiden reunasta kiinni ja venyttämällä tai siirtämällä niitä. Ikkunoista voi siis muokata juuri itselleen sopivat. Ohjelma tallentaa ikkunoiden järjestyksen automaattisesti EG:n lopetuksen yhteydessä, joten järjestys palautuu samanlaisena, kun EG käynnistetään seuraavan kerran. Ikkunat voi palauttaa alkuperäisille paikoilleen valitsemalla ”Tools” -> ”Options” -> ”General”-välilehdeltä ”Reset Docking Windows”. Yksittäisen suljetun ikkunan saa palautettua ”View”-valikosta. Prosessidiagrammin ja työtilan voi myös erottaa toisistaan eli muuttaa näkymän samankaltaiseksi kuin Enterprise Guide 3.0:n oletusnäkyminen näiden osalta. Tämä onnistuu klikkaamalla hiiren oikeata painiketta prosessidiagrammivälilehden otsakepalkin (”Project Designer”) päällä ja valitsemalla ”Dockable”.

Tehtävä: Avaa Enterprise Guide pikakuvakkeesta. Selaa valikoita, ikkunoita ja välilehtiä saadaksesi alustavan kuvan Enterprise Guiden käyttöympäristöstä. Harjoittele myös ikkunoiden muokkausta. Lisäksi voit harjoitella ”Reset Docking Windows” -toiminnon käyttöä, jos ikkunat menevät epäjärjestykseen tai ovat edellisen käyttäjän jäljiltä väärässä järjestyksessä.

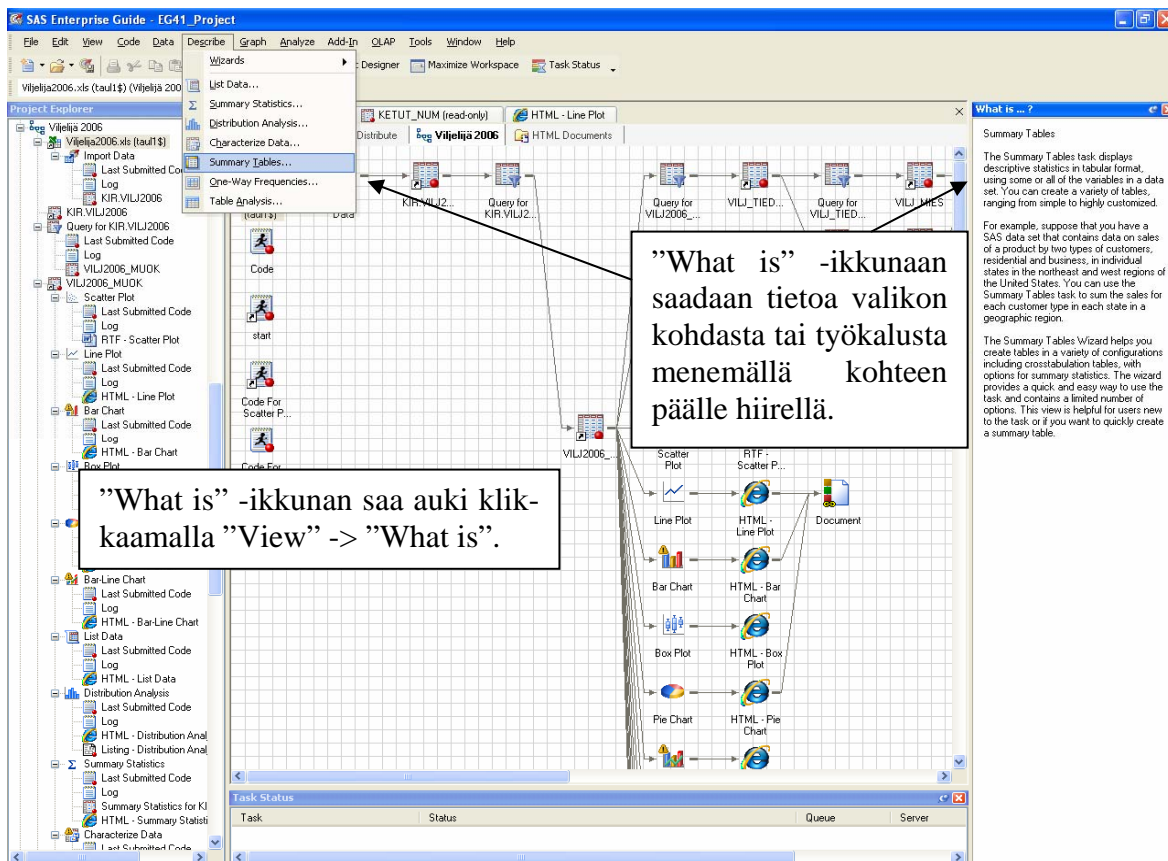
4 Enterprise Guiden ohje

Enterprise Guiden ohje löytyy klikkaamalla ”**Help**” -> ”**SAS Enterprise Guide Help**”. Samasta valikosta löytyy myös ”Getting Started Tutorial”, jossa käydään läpi EG:n perustoimintoja. Siihen kannattaa tutustua itsenäisesti. Varsinaisessa ohjeessa on eri kokonaisuuksia järjestettyinä selkeiden otsikoiden alle, joten se on varsin käyttökelpoinen ja toimiva. Esimerkiksi eri analyysityökaluihin löytyy apua klikkaamalla ohjeessa ”Sisällys”-välilehdeltä auki kansion ”Running SAS Tasks”. Lisäksi apua saa painamalla näppäimistön F1-näppäintä Enterprise Guiden käytön lomassa, jolloin EG avaa automaattisesti ohjeesta sellaisen sivun, joka liittyy sillä hetkellä käytössä olevaan osaan EG:tä.

Myös muokattaessa tehtäväsolmua (tehtäväsolmujen muokkausta käsitellään oppaassa myöhemmin) ohjeita on runsaasti tarjolla:



Kun ollaan tehtäväsolmun ulkopuolella eli varsinaisessa Enterprise Guiden päänäkymässä, saadaan tietoa eri työkaluista ja valikoista käyttämällä ”What is” -ikkunaa:



Enterprise Guiden ohjeen lisäksi on usein tarpeen käyttää myös tavallisen SASin manuaalia (löytyy internetistä: <http://support.sas.com/documentation/onlinedoc/sas9doc.html>). Varsinainen SASin manuaali on nimittäin paljon täsmällisempi ja laajempi yksityiskohtien suhteen kuin Enterprise Guiden ohje. Lisätietoa tavallisen SASin manuaalin käyttämisestä löytyy myös luvusta **SAS-koodin käyttäminen**.

Aiemmin tavallista SASia käyttäneille EG:n hahmottamista saattaa helpottaa tieto Enterprise Guiden työkalujen käyttämisestä SAS-proseduureista. ”Task list” -ikkunan toisella välilehdellä on kunkin työkalun vieressä mainittu sen käyttämät SAS-proseduurit.

5 Tiedostojen tyypit

Enterprise Guide tunnistaa automaattisesti erityyppisiä tiedostoja. Enterprise Guideen voidaan avata aiemmin luotuja Enterprise Guide -projektitiedostoja, joita ovat *.epg-tiedostot (EG versio 3 ja 4) ja *.seg-tiedostot (EG versiot 1 ja 2). Avattaessa EG 4.1:een EG 3.0:ssa tehty projektitiedosto, ohjelma kuitenkin huomauttaa, samasta tiedostotyypistä huolimatta, että se on tehnyt projektitiedostoon tiettyjä muutoksia, jotka lopullisesti hyväksytään tallentamalla projekti EG 4.1:ssä. Jos muutokset tallennetaan, projektitiedostoa ei voi enää käyttää EG 3.0:ssa.

Enterprise Guideen voidaan avata erityyppisiä datatiedostoja. EG:ssä aukeavat muun muassa SAS-datat (*.sas7bdat, *.sd7 ja *.sd2), Excel-datat (*.xls) sekä monen muun tyyppiset datatiedostot (*.dbf, *.txt, *.csv, *.asc, *.tab). Lisäksi Enterprise Guideen voidaan avata muun muassa SAS-kooditiedostoja (*.sas) sekä SAS-logitiedostoja (*.log). Täydellisen listan Enterprise Guiden tunnistamista tiedostotyypeistä näkee esimerkiksi klikkaamalla: ”File” -> ”Open” -> ”Other” -> ”Local Computer” ja avaamalla tiedostotyyppilistan. Samasta ”File” -> ”Open”-valikosta löytyy myös omat kohdat projektitiedostojen (”Project”), datojen (”Data”) ja SAS-kooditiedostojen (”Code”) avaamiseen.

6 Aineiston käyttöönotto

6.1 Kirjastoviitteen luominen

Kirjastoviite on SAS-ohjelmiston käyttämä menettelytapa, jossa avainsanalla viitataan tiettyyn hakemistoon tai kansioon tietokoneella. Sitä tarvitaan esimerkiksi, jos halutaan määrittellä, mihin tietokoneen kansioon istunnon aikana luodut SAS-datat tallennetaan tai mistä kansioista aiemmin luodut SAS-datat otetaan käyttöön. SAS Enterprise Guiden versiossa 4.1 kirjastoviitteen luomiseen on olemassa kolme käyttökelpoista tapaa.

Istunnon ajan voimassa olevan kirjastoviitteen luominen

Kaksi ensimmäistä tapaa tuottavat kirjastoviitteen, joka on voimassa EG-istunnon ajan. Kun Enterprise Guide tai projekti käynnistetään uudelleen, pitää kirjastoviite muistaa istunnon aluksi yhdistää uudelleen haluttuun hakemistoon suorittamalla kirjastoviitteen määrittelevä solmu uudelleen.

Ensimmäinen tapa on luoda kirjastoviite käyttäen siihen tarkoitettua Enterprise Guiden työkalua. Kirjastoviite luodaan tällöin klikkaamalla ”Tools” -> ”Assign Library”. Au-
keavassa ikkunassa määritellään vaihe kerrallaan kirjastoviite.

Ensimmäisellä sivulla annetaan kirjastoviitteelle nimi (esimerkiksi kir):

Library Wizard

1 of 6 Enter a name and an optional description for the library.

Name: KIR

Description:

Note:
This library definition will not be permanently stored with the SAS server. To use this library again in a future session, you will need to rerun this task within your project. To create a permanent library definition, use the New Library wizard in SAS Enterprise Guide Explorer.

Enter the name of the library and a description of its contents.

< Back Next > Finish Cancel Help

Toisella sivulla valitaan palvelin (usein ainoa vaihtoehto on "Local"):

Library Wizard

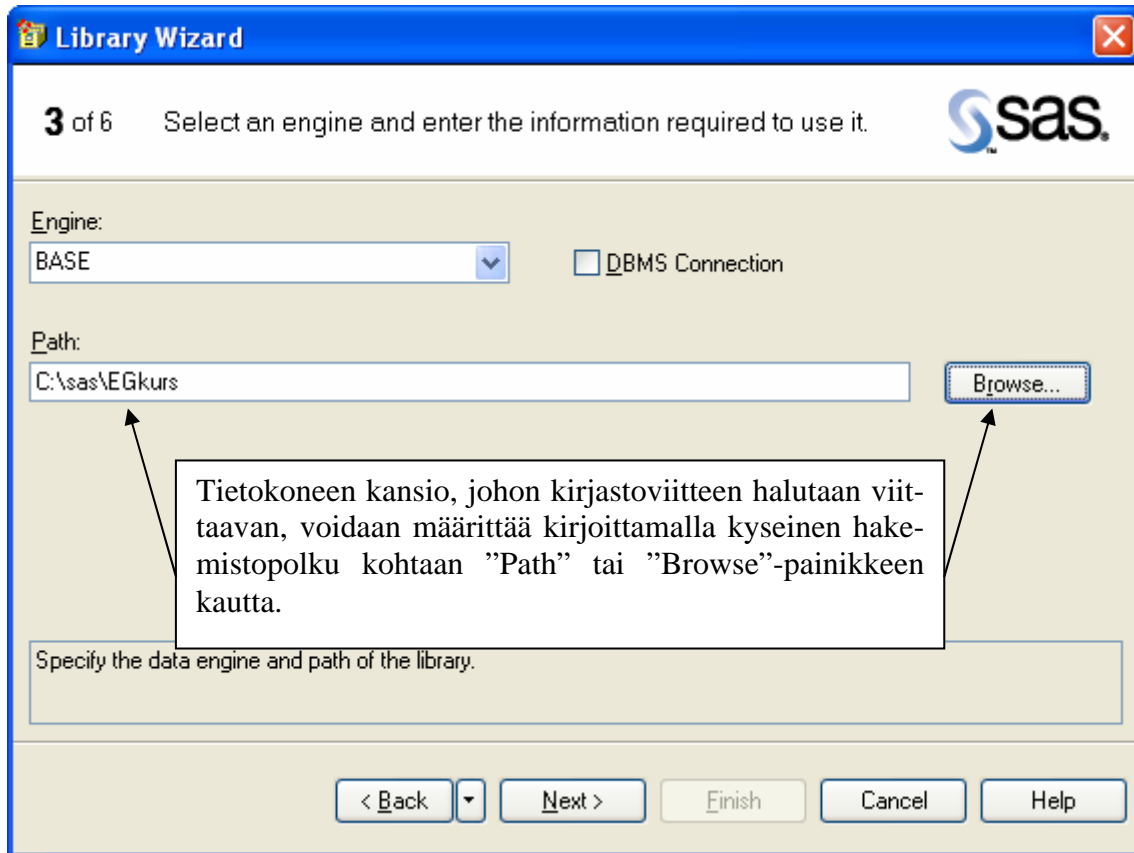
2 of 6 Specify the server for the library.

Server	Description
Local	User's local machine

Klikataan "Local" siniseksi.

< Back Next > Finish Cancel Help

Kolmannella sivulla valitaan kansio, johon kirjastoviite viittaa, esimerkiksi C:\SAS\EGKURS ("Engine" puolestaan saa pääsääntöisesti olla oletuksena oleva "Base"):



Loput sivut voi yleensä klikata ohii painamalla "Next" ja viimeisellä sivulla valitaan "Finish".

Toinen tapa luoda kirjastoviite, on tehdä se SAS-koodin avulla. Koodieditorin saa avattua klikkaamalla "File" -> "New" -> "Code". Tässä koodi-ikkunassa voi määrittää kirjastoviitteen samoin kuin perus-SASissa, esimerkiksi:

```
LIBNAME kir 'c:\sas\egkurs';
```

Siinä kir on kirjastoviitteen nimi ja yksinkertaisten lainausmerkkien sisällä on hakemistopolku, jonka osoittamaan kansioon kirjastoviite viittaa. Koodin voi suorittaa muun muassa klikkaamalla koodisolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla "Run Code On Local".

Pysyvän kirjastoviitteen luominen

Pysyvä kirjastoviite luodaan klikkaamalla ensin "Tools" -> "SAS Enterprise Guide Explorer". Aukeavassa ikkunassa valitaan "File" -> "New" -> "Library". Sitten määritellään kirjastoviite samoin kuin Assign Library -työkalulla, mutta toisella sivulla laitetaan merkki kohtaan "Assign at Server Startup".

Tehtävä: Tee tietokoneelle kansio C:\SAS\EGKURS ja kopioi sinne tähän oppaaseen liittyvät datatiedostot (<http://www.mtt.fi/mtts/pdf/egdata/EGdatat.zip>). Luo istunnon ajan voitavissa oleva kirjastoviite nimeltä kir, joka viittaa kansioon C:\SAS\EGKURS.

6.2 Kirjastoviitteen asettaminen oletukseksi

Oletuskansio, johon eri tehtäväsolmut ehdottavat dataa ja tulosteita tallennettavaksi, on mahdollista vaihtaa. Se onnistuu klikkaamalla **“Tools”** -> **“Options”** -> **“Tasks”** -> **“Output Library”** ja vaihtamalla siellä näkyvien kirjastoviitteiden järjestystä. Järjestystä voi vaihtaa valitsemalla jonkin kirjastoviitteen aktiiviseksi ja siirtämällä sitä ”Up”- tai ”Down”-painikkeilla. Kunkin projektin dataa kannattaa tallentaa omaan kansioonsa. Yksi hyvä menettelytapa on käyttää kaikissa projekteissa samannimistä kirjastoviitettä (esimerkiksi kir), mutta määrittää aina projektin alussa kir-niminen viite osoittamaan kyseisen projektin data-kansioon tietokoneella. Tällöin kir-kirjastoviite voidaan määrittää asetussissa ensisijaiseksi kirjastoviitteeksi ja seurauksena EG tarjoaa oletusarvoisesti dataa tallennettavaksi käyttäjän projektille tarkoitamaan kansioon tietokoneella.

Tehtävä: Aseta edellisessä tehtävässä luotu kirjastoviite kir oletuskirjastoviitteeksi.

6.3 Aineiston tallennusmuoto

Aineiston pitää olla tallennettuna sellaiseen muotoon, että siinä eri tilastoyksiköiden havainnot ovat eri riveillä ja jokaisen muuttujan arvot omissa sarakkeissaan. Seuraavassa muutama esimerkkiaineisto:

Esimerkki. Lehmädata:

LEHNIMI	KAS	TUOTOS
Mansikki	Perus	8600
Rebekka	Perus	8450
Pokemon	Super	9100
Pikachu	Perus	8700
MelC	Perus	8750
MelB	Super	8950
NylonBeat	Super	9150
Rulu	Super	8700
Viira	Perus	8750
Ylva	Super	8950

Esimerkki. Peltodata:

LAIKE	LOHKO	SATO
A	1	6500
A	2	5850
A	3	6250
B	1	5025
B	2	4520
B	3	5260
C	1	4520
C	2	3950
C	3	4310

Toistomittausaineistoksi kutsutaan dataa, jossa samaa ominaisuutta on mitattu useamman kerran kokeen tai seurannan aikana samoilta tilastoyksiköiltä. Toistomittausaineisto voidaan tallentaa kahdella eri tavalla ja tarvittaessa kääntää muodosta toiseen Enterprise Guiddessa:

Esimerkki. Elintarvikedata, toistomittaukset yhdessä sarakkeessa:

HENKILO	KAS	KK	PAINO
Pekka	TavPuuro	1	87.2
Pekka	TavPuuro	2	87.1
Pekka	TavPuuro	3	87.4
Paavo	PuuroX	1	72.1
Paavo	PuuroX	2	83.2
Paavo	PuuroX	3	85.6
Matti	TavPuuro	1	99.1
Matti	TavPuuro	2	99.8
Matti	TavPuuro	3	100.2
Antti	PuuroX	1	78.1
Antti	PuuroX	2	83.2
Antti	PuuroX	3	86.5
Timo	PuuroX	1	95.1
Timo	PuuroX	2	103.2
Timo	PuuroX	3	114.2
Keijo	TavPuuro	1	104.2
Keijo	TavPuuro	2	102.1
Keijo	TavPuuro	3	99.2

Esimerkki. Elintarvikedata, toistomittaukset omissa sarakkeissaan:

HENKILO	KAS	PAINO_1	PAINO_2	PAINO_3
Pekka	TavPuuro	87.2	87.1	87.4
Paavo	PuuroX	72.1	83.2	85.6
Matti	TavPuuro	99.1	99.8	100.2
Antti	PuuroX	78.1	83.2	86.5
Timo	PuuroX	95.1	103.2	114.2
Keijo	TavPuuro	104.2	102.1	99.2

Aineistoa tallennettaessa pitää huomioida, että muuttujien nimet saavat olla EG:ssä korkeintaan 32 merkin pituisia, mutta suositeltavaa olisi pitää nimet vieläkin lyhyempinä. Monissa ohjelmissa, mukaan lukien SASin vanhemmat versiot, rajana nimen pituudelle on kahdeksan merkkiä ja pidemmät nimet saattavat aiheuttaa yhteensopivuusongelmia. Muuttujien nimissä kannattaa käyttää ainoastaan kirjaimia (ei skandinaavisia), numeroita ja alleviihausmerkkiä (_), ei siis muita erikoismerkkejä eikä välilyöntejä. Muuttujien nimiä ei myöskään kannata aloittaa numeroilla.

6.4 Aineiston syöttäminen Enterprise Guidessa

Uuden datamatriisin avaaminen ja tallentaminen

Aineisto voidaan muodostaa syöttämällä se data-matriisiin Enterprise Guidessa. Tämä tapa on hyvin samanlainen kuin datan syöttäminen käsin Excelissä. Uusi data-matriisi saadaan ruudulle valitsemalla yläreunan valikoista **"File"** -> **"New"** -> **"Data"**. Tällöin aukeaa kaksivaiheinen työkalu, jonka ensimmäisessä vaiheessa annetaan muodostettavalle datalle nimi "Name"-kohdassa ja valitaan tallennuksen kohdekansio kirjastoviitteen avulla "Location"-kohdassa. Klikkaamalla "Next" päästään seuraavalle sivulle, jossa voidaan antaa datamatriisin muuttujien nimet ja ominaisuudet. Lopuksi klikataan "Finish".

Datamatriisin muokkaaminen

Datamatriisin muokkaaminen selitetään tarkemmin luvussa **Datan muokkaus**. Tässä käydään lyhyesti läpi muutamat perusasiat, jotka liittyvät uuden datan luomiseen.

Jotta dataa pääsee muokkaamaan, on datasta ensin poistettava kirjoitussuojaus. Tämä tehdään datan ollessa päällimmäisenä työtilassa klikkaamalla ”Data” ja poistamalla mahdollinen valinta kohdasta ”Read-only”.

Sarakkeiden nimiä, tyyppejä yms. pääsee tässä vaiheessa muokkaamaan klikkaamalla sarakkeen otsikkoa hiiren oikealla painikkeella ja valitsemalla ”Properties”. Uusia sarakkeita voidaan luoda klikkaamalla hiiren oikealla painikkeella jotain sarakeotsikkoa ja valitsemalla ”Insert Column”. Riveihin kohdistuviin toimenpiteisiin päästään käsiksi klikkaamalla hiiren oikealla painikkeella tietyn rivin numeron kohdalla. Uusia rivejä saa lisätyksi myös painamalla Enter-näppäintä, kun kohdistin on muokattavan datamatriisin viimeisellä rivillä. Muuttujien järjestystä voi vaihtaa siten, että valitaan jokin sarake aktiiviseksi, otetaan muuttujan nimestä kiinni painamalla hiiren vasen painike pohjaan ja raahataan sarake eri kohtaan.

Datamatriisiin voi käsin syöttämisen lisäksi myös tuoda dataa esimerkiksi Excelistä leikepöydän kautta (”Copy”/”Kopioi” ja ”Paste”/”Liitä”).

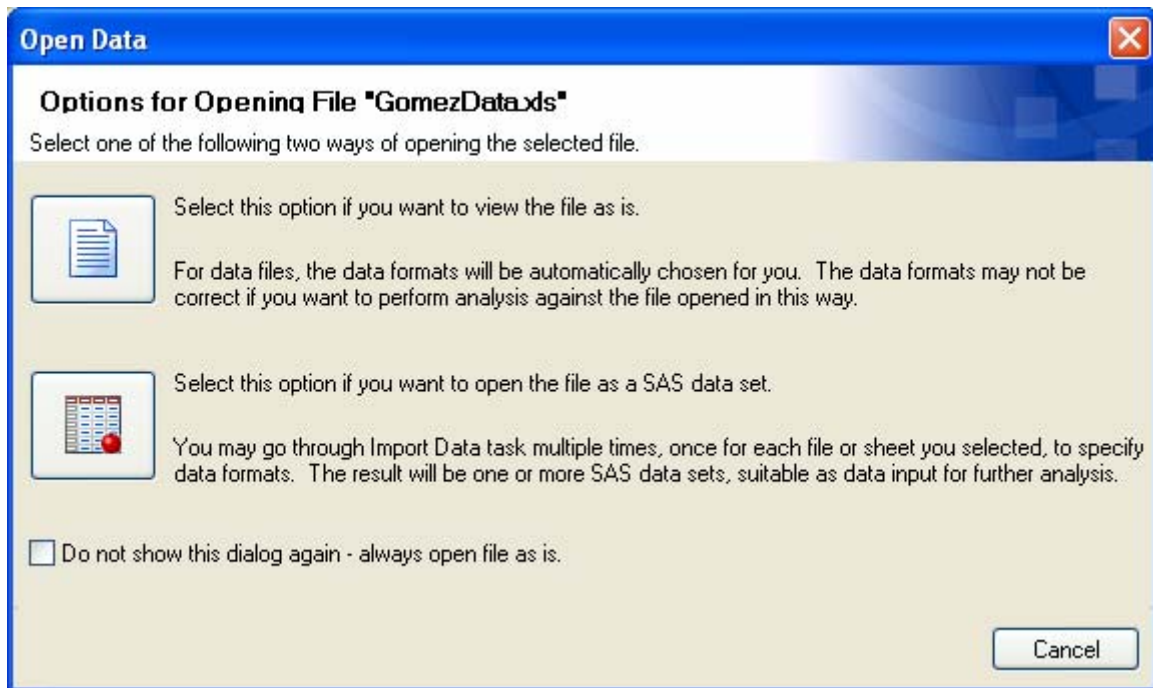
Tehtävä: Avaa Enterprise Guidessa uusi tyhjä data-matriisi ja tallenna siihen seuraava aineisto:

LAIJKE	KERR	SATO
A	1	6500
A	2	5850
A	3	6250
B	1	5025
B	2	4520
B	3	5260
C	1	4520
C	2	3950
C	3	4310

Tallenna aineisto pysyväksi SAS-datatiedostoksi kirjastoviitteesi osoittamaan kansioon nimellä **peltodata**.

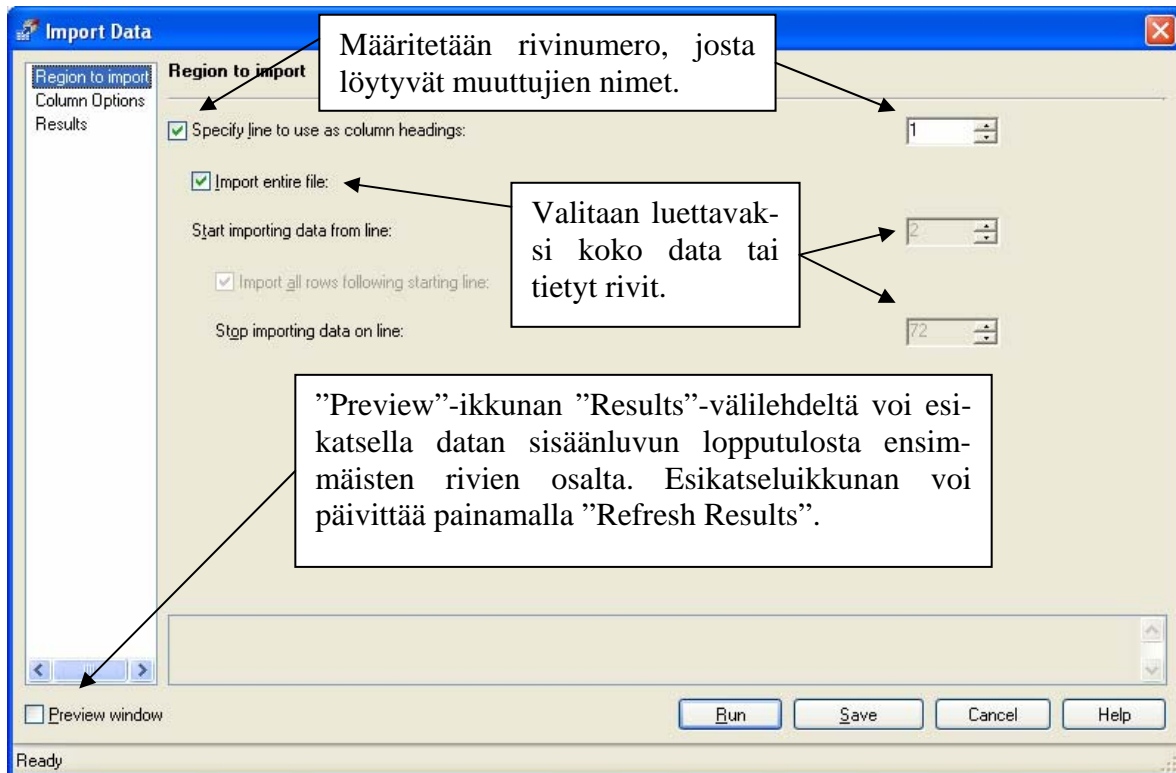
6.5 Excel-tiedoston avaaminen Enterprise Guideen

Enterprise Guideen voidaan avata Excel-tiedostoja. Tämä tapahtuu klikkaamalla ”File” -> ”Open” -> ”Data” ja ”Local Computer” (tai ”File” -> ”Import Data”) ja valitsemalla haluttu tiedosto. Vaihtoehtoisesti Excel-tiedosto voidaan yksinkertaisesti raahata omalta tietokoneelta EG:n prosessidiagrammi-välilehdelle. Tämän jälkeen valitaan yksi tai useampi Excel-datan välilehti, jolta tai joilta dataa halutaan tuoda Enterprise Guideen. Tämän jälkeen näkyviin tulee seuraava valintaruutu:

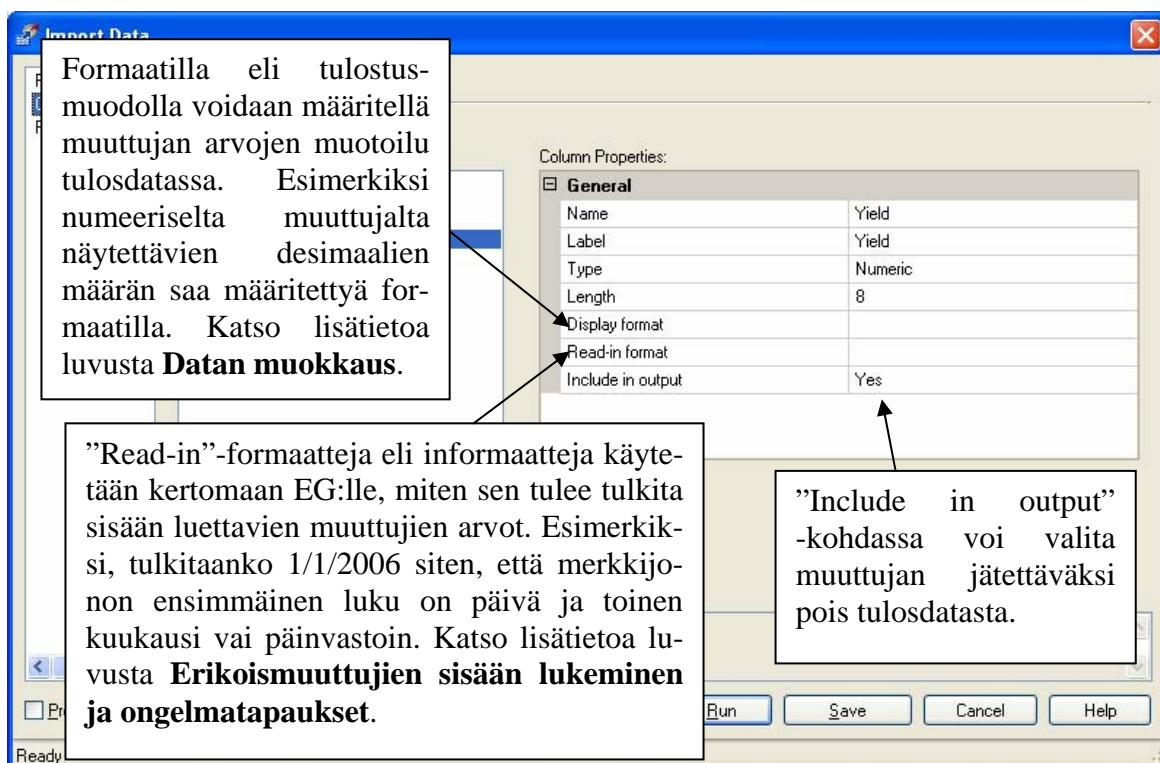


Ensimmäinen vaihtoehto avaa ”näkyvän” Excel-dataan. Toinen vaihtoehto luo Excel-datasta SAS-datan ja ottaa sen käyttöön EG:ssä. Erona vaihtoehtojen välillä on, että EG pystyy käyttämään SAS-dataksi muunnettua dataa monipuolisemmin. Jos alkuperäistä Excel-dataa muutetaan (Excelissä), muutos päivittyy EG:ssä olevaan Excel-näkymään automaattisesti, kun dataa seuraavan kerran käytetään (kannattaa sulkea mahdollisesti auki oleva datamatriisi EG:n työtilasta ja klikata se uudelleen auki prosessidiagrammista). Jos taas luodaan SAS-data, projektiin syntyy tuloksena sekä näkymä Excel-dataan että näkymä luotuun SAS-dataan. SAS-dataa voidaan sekä muokata että käyttää EG:ssä, mutta samalla alkuperäinen Excel-data säilyy muuttumattomana. Tässäkin tapauksessa mahdolliset muualla tehdyt muutokset Excel-dataan päivittyvät samoin kuin edellä Excel-datanäkymään, mutta jos halutaan tehdä muutokset myös SAS-dataan, on ajettava uudelleen prosessidiagrammissa Excel-datanäkymän ja SAS-datanäkymän välissä oleva ”Import Data” -tehtäväsolmu. Jälkimmäinen vaihtoehto eli datan lukeminen SAS-dataksi on suositeltavampi, koska se mahdollistaa aineiston monipuolisemman käytön EG:ssä.

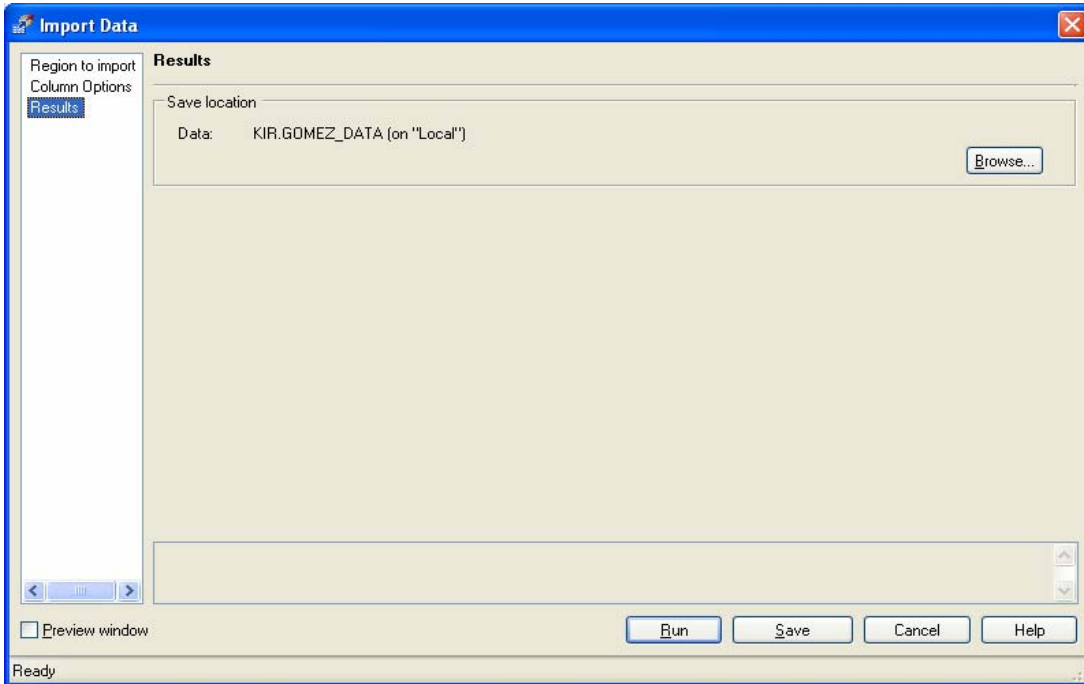
Jos valitaan jälkimmäinen vaihtoehto, päädytään seuraavaan näkymään (välilehtiä voi vaihtaa ruudun vasemmassa reunassa olevasta listasta), jossa ”Region to import” -välilehdellä määritetään datan lukeminen sisään rivien osalta:



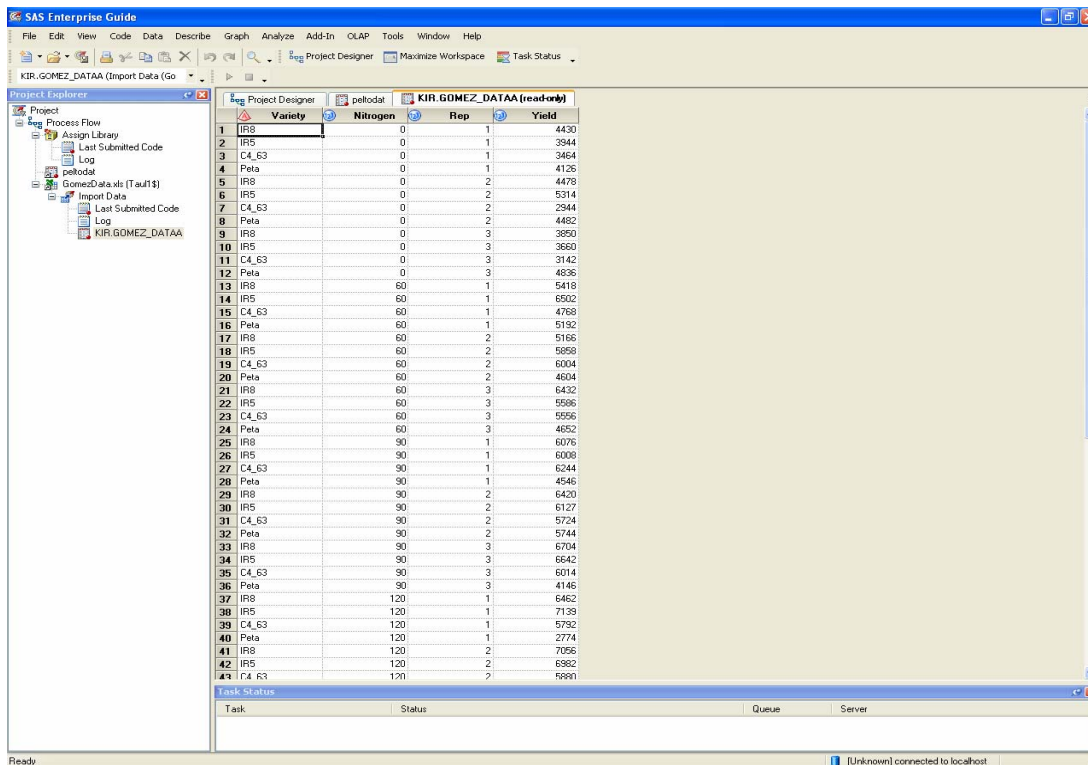
Huomaa, että EG jättää Excel-datasta huomioimatta otsikkorivit ja tyhjät rivit yms., joten jos annat yllä rivien numeroita, aloita rivien laskeminen vasta siitä kohtaa, missä varsinainen data sarakeotsikoineen alkaa. ”Column Options” -välilehdellä puolestaan tarkastellaan ja muokataan sisään luettavia sarakkeita:



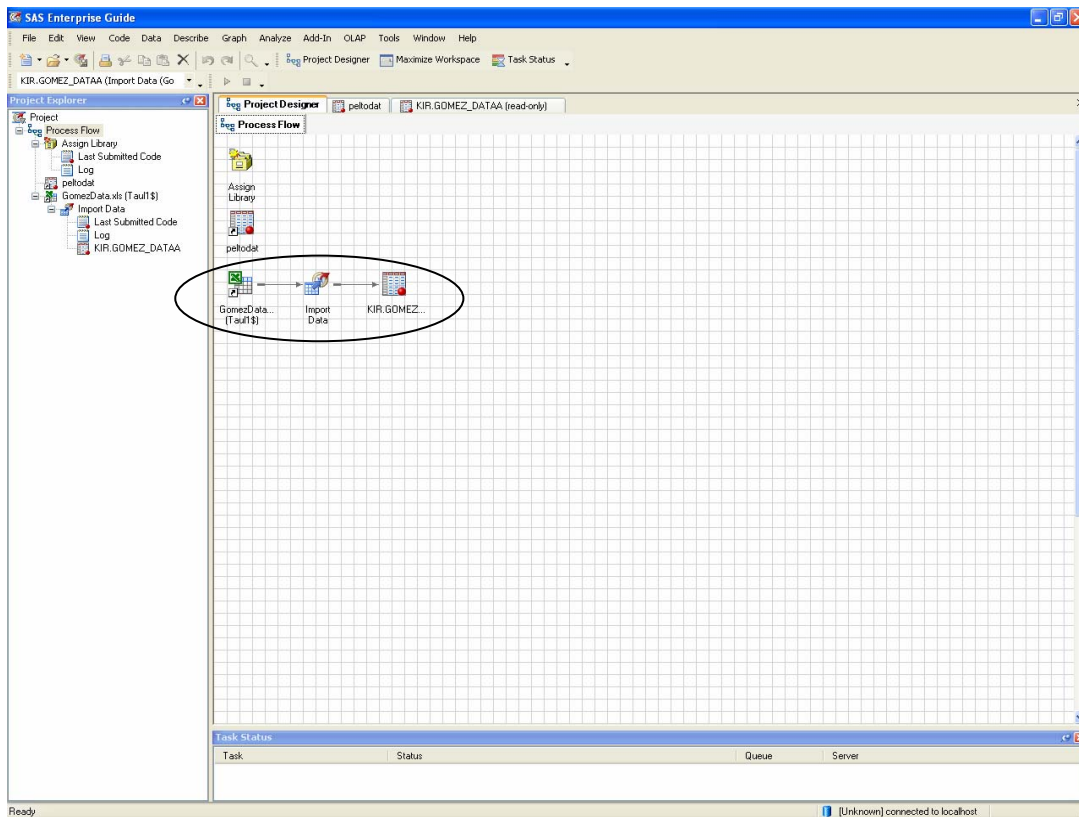
Päivämäärämuuttujista ja muista erikoisemmista muuttujista kannattaa yleensä tarkistaa Enterprise Guiden automaattisesti valitsemat informaattit ja formaattit, jotta tieto tulee varmasti luettua sisään oikein ja esitettyä datamatriisissa ja sen myötä tulosteissa mielekkäällä tavalla. Tavallisimmille päivämäärämuotoisille muuttujille Enterprise Guide 4.1 osaa valita sisäänlukuformaattit oikein, toisin kuin Enterprise Guiden edellinen versio. ”Results”-välilehdellä voidaan vaihtaa tuloksena syntyvän SAS-datan nimeä ja sijaintia. Alla on muutettu datan nimeä ja määritetty, että se tallennetaan kirjastoviitteen kir osoittamaan kansioon:



Lopuksi painetaan ”Run”, jolloin uusi data näkyy ruudussa:



Samalla prosessidiagrammiin muodostuvat datanäkökulmat, jotka viittaavat Excel-datanäkymään ja SAS-dataan. Samoin muodostuu tehtäväsolmu, joka sisältää SAS-datan muodostamiseen liittyvät määrittelyt ("Import Data" -tehtäväsolmu). Prosessidiagrammi saadaan näkyviin klikkaamalla auki "Project Designer" -välilehti:



Excel-datanäkymässä otetaan automaattisesti alkuperäisen datan ensimmäinen rivi sarakeotsikoiksi. Jos tätä oletusta halutaan vaihtaa, se onnistuu "Tools" -> "Options" -> "Data" -> "Performance" -välilehdellä.

Tekstimuotoisen datan sisään lukemiseen kuuluvat hyvin pitkälti samat välivaiheet kuin Excel-datan sisään lukemiseen. Lisänä sisään luku -työkalussa on "Text Format" -välilehti, jossa tehdään tekstimuotoiseen dataan liittyviä erityismäärittelyjä.

Tehtävä: Avaa kansio C:\SAS\EGKURS Excel-data nimeltä **GomezData.xls** ja muodosta siitä pysyvä SAS-data nimeltä **gomez_data** kirjastoviitteesi osoittamaan kansioon.

6.6 Erikoismuuttujien sisään lukeminen ja ongelmatapaukset

Enterprise Guide pystyy yleensä tulkitsemaan erikoisiakin muuttujia automaattisesti oikein lukiessaan aineistoa SAS-dataksi. Jotkin muuttujat aiheuttavat kuitenkin ongelmia, jotka pystytään kiertämään käyttämällä sisäänlukuformaatteja eli informaatteja.

Tyypilliset ongelmatilanteet, jotka voidaan ratkaista informaateilla, liittyvät päivämäärien tai aikojen sisään lukuun. Ongelmien syynä on se, että päivämäärät ja kellonajat voidaan ilmoittaa hyvin monella tavalla ja siten ohjelmakin voi tulkita niitä eri tavoin. Informaation toimintaperiaatteena on, että niiden avulla kerrotaan ohjelmalle, miten sen tulee tulkita sisään luettavan datan muuttujien arvot. Samalla voidaan käyttää myös tavallisia formaatteja eli tulostusmuotoja, joiden avulla puolestaan kerrotaan ohjelmalle, miten sen tulee esittää sisään lukemiensa muuttujien arvot.

Datan sisään luvun yhteydessä informaatiit ja formaatit vaihdetaan yhdelle muuttujalle kerrollaan alla olevassa kuvassa esitetyllä tavalla:

"Import Data" -solmu, joka saadaan, kun luetaan data SAS-dataksi.

a) Valitaan "Column Options" -välilehdellä muuttuja, jolle määritetään informaatti.

b) Mennään vaihtamaan informaattia.

c) Valitaan listasta sopiva informaatti ja ilmoitetaan kuinka monella merkillä tieto on tallennettu.

Lisätieto informaattista.

Tässä tapauksessa Enterprise Guide oli valinnut jo valmiiksi tilanteeseen soveltuvan informaatin DATETIME18 ja vastaavan tulostusformaatin. DATETIME kuitenkin pitää sisällään tiedon sekä päivämäärästä että ajasta, kun taas alkuperäisessä datassa oli ainoastaan päivämäärätieto. Olennainen tieto tulee luettua sisään näinkin, mutta esimerkissä vaihdetaan informaattiksi tilanteeseen paremmin sopiva DDMMYY10. Formaatin nimi kertoo että arvoissa tieto on annettu järjestyksessä päivä (dd), kuukausi (mm), vuosi (yy) ja kaiken kaikkiaan kymmenellä merkillä (dd.mm.yyyy). Vaihdettaessa informaattia vaihtuu myös tulostusformaatti automaattisesti. Sitä voidaan kuitenkin myös vaihtaa informaatin vaihtoa vastaavalla tavalla kohdassa "Display format". Yleisimpiä informaatteja ovat edellä mainitun DDMMYYw.:n lisäksi muun muassa ANYDTDEw. (lukee monia erilaisia päivämäärä- ja päivämäärä-aika -muotoisia muuttujia päivämäärinä), DATETIMEw. (päivämäärä-aika informaatti), FIN-alkuiset (suomalaisten omia informaatteja muutama päivämäärä- ja päivämäärä-aika-tilanteeseen, katso lisätieto informaattien description-kohdista). Lisätieto informaateista löytyy SASin manuaalista ja Slaughterin ja Delwichen (2006) kirjasta, josta löytyy muun muassa kattavampi lista yleisimmistä informaateista.

Datan sisään lukemiseen liittyvissä mahdollisissa ongelmatilanteissa saattaa auttaa jokin Excelissa tehtävä toimenpide. Tällaisia toimenpiteitä ovat sarakkeiden tyyppien määrittely niiden arvoja vastaaviksi, puuttuvien havaintojen jättäminen tyhjiksi (esimerkiksi pisteiden sijaan), muuttujan tyyppin kannalta virheellisten merkkien poistaminen niiden arvoista (esimerkiksi numeerisilla arvoilla välilyöntien, erikoismerkkien (esimerkiksi >, <, =, %) ja kirjaimien poistaminen) sekä muuttujien nimien määrittely oikein eli lähinnä ilman erikoismerkkejä ja välilyöntejä. Joskus ongelmia saattaa aiheuttaa myös tilanne, jossa muuttujalla on paljon peräkkäisiä puuttuvia havaintoja tiedoston alusta lukien. Tällöin saattaa

auttaa, kun tiedoston alkuun tehdään ylimääräinen rivi, jossa annetaan muuttujan tyyppin mukainen arvo jokaiseen soluun. Kun data on luettu sisään, tämä ylimääräinen rivi poistetaan. Vaihtoehtoisesti jo sisään lukemisen yhteydessä voidaan määrittää, että havainnot otetaan mukaan tulosdataan vasta toiselta riviltä alkaen eli poistetaan apurivi.

Sisään lukemisen ongelmatilanteissa saattaa auttaa myös se, että data tallennetaan Excellissä esimerkiksi DBF 4 (dBASE IV) eli *.dbf -muotoon. Tällöin Excelin taustalle tekemät määrittelyt poistuvat ja data tallentuu yksinkertaisesti sellaisessa muodossa kuin se näkyy ruudullakin. Kun data sitten luetaan sisään Enterprise Guideen tässä dBASE-muodossa, onnistutaan kiertämään monia ongelmia ylimääräisten ja mahdollisesti sisään lukua sekoittavien muotoilujen hävityä.

6.7 SAS-datan avaaminen Enterprise Guideen

SAS-data avataan Enterprise Guideen valitsemalla ”File” -> ”Open” -> ”Data” -> ”Local Computer”, valitsemalla haluttu SAS-data ja klikkaamalla ”Avaa”. Vaihtoehtoisesti data voidaan hakea tietyistä kirjastoviitteen osoittamasta kansioista, jolloin valitaan ”File” -> ”Open” -> ”Data” -> ”SAS Servers”.

Tehtävä: Avaa Enterprise Guideen SAS-data nimeltä **KetutData.sas7bdat** kansioista C:\SAS\EGKURS.

6.8 Datojen tallentaminen Enterprise Guidesta

Enterprise Guide -projektissa oleva data voidaan tallentaa tiedostoksi. Tämä tapahtuu klikkaamalla dataa prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Export” -> ”Export *datan_nimi*” ja ”Local Computer”. Sen jälkeen valitaan haluttu kansio tietokoneelta, tiedostotyyppi alasvetovalikosta ja annetaan tiedostolle nimi. Valittavia tiedostotyyppejä ovat muun muassa *.sas7bdat (SAS-data), *.xls (Excel-tiedosto), *.dbf (dBase-tiedosto) ja *.txt (tekstitiedosto). Tavallisen ”Export”-vaihtoehdon lisäksi valittavana on myös vaihtoehto ”Export *datannimi* As A Step In Project”. Tämän vaihtoehdon avulla datan tallentaminen tiedostoksi saadaan pysyväksi vaiheeksi prosessidiagrammiin, jolloin se voidaan toteuttaa rutiininomaisena vaiheena projektikokonaisuuden suorittamisen yhteydessä.

Datat voidaan tallentamisen lisäksi myös lähettää suoraan toiseen ohjelmaan. Tämä onnistuu klikkaamalla prosessidiagrammin datakuvaketta hiiren oikealla painikkeella ja valitsemalla ”Send To” -> ”E-mail Recipient” tai ”Microsoft Word” tai ”Microsoft Excel”. ”E-mail Recipient” -vaihtoehto toimii ainakin, jos sähköpostiohjelmana on Outlook Express. Sitä klikattaessa sähköpostiohjelma avaa uuden viestin ja liittää datan viestiin liitetiedostona valitussa tiedostomuodossa. ”Microsoft Word” -vaihtoehto avaa Wordin ja sijoittaa datan word-dokumenttiin taulukkona. ”Microsoft Excel” -vaihtoehto avaa datan Excelliin. Vaihtoehtona on myös ”E-mail Recipient as a Step in Project”, jonka avulla sähköpostin lähetys saadaan pysyväksi osaksi projektikokonaisuutta. Ennen tämän vaihtoehdon käyttöä pitää määrittää sähköpostiasetukset Enterprise Guideen.

Tehtävä: Tallenna jokin projektissa olevista SAS-datoista Excel-tiedostoksi.

6.9 Projektin tallentaminen

Enterprise Guidessa kaikki työ tallennetaan projekteina. Projekti koostuu toisiinsa liittyvistä datoista, toimenpiteistä, tulosteista, koodeista, logi-tiedostoista ja muistiinpanoista. Tässä kohtaa on syytä mainita, että projektin sisältämät data-kuvakkeet eivät varsinaisesti sisällä dataa, vaan ne ovat pikakuvakkeita varsinaisiin datoihin. Niinpä datat on tallennettava kansioon tietokoneelle, jos ne halutaan säästää EG-istunnon jälkeen. Ne eivät siis tallennu projektin sisälle. SAS-datojen tallentaminen tapahtuu määrittelemällä niiden nimeämisen yhteydessä tietokoneella olevaan kansioon osoittava kirjastoviite (pysyvään kansioon osoittava kirjastoviite eli jokin muu kuin "Work"), jolloin EG tallentaa SAS-datan pysyvästi kyseiseen kansioon. Jos data tallennetaan väliaikaisena SAS-datana ("Work"), Enterprise Guide varoittaa ohjelman sulkemisen yhteydessä siitä, että projekti sisältää väliaikaisia dataa.

Prosessidiagrammiin muodostunut projektikokonaisuus voidaan tallentaa valitsemalla **"File" -> "Save Project As"**. Tämän jälkeen tallentaminen samalla nimellä onnistuu valitsemalla "File" -> "Save *projektin_nimi*" tai painamalla näppäimistöä CTRL+S.

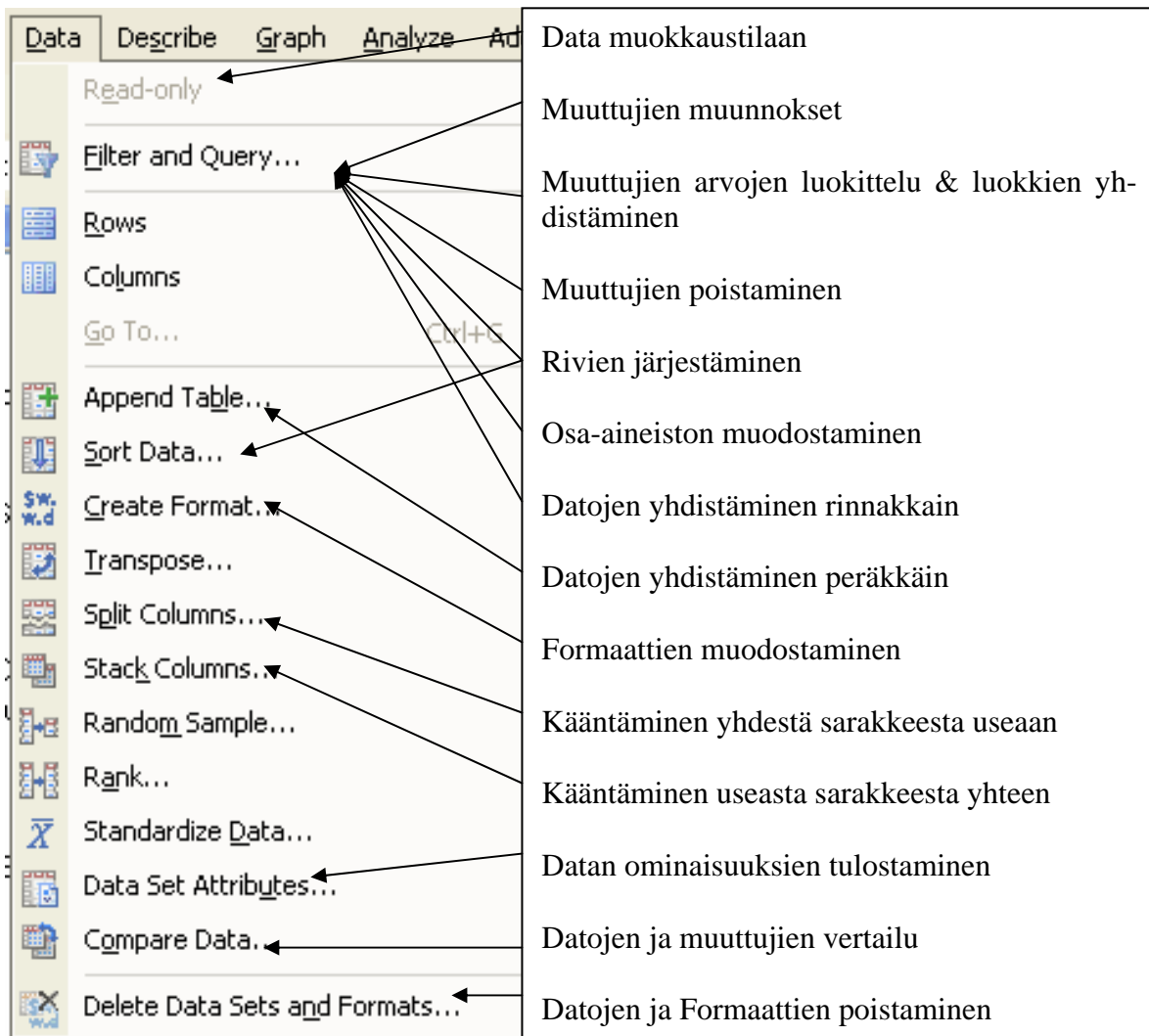
Jos tallennettu projekti avataan toisella tietokoneella tai tiedostot siirretään eri kansioon, kirjastoviite pitää muuttaa viittaamaan uuteen kansioon. Uudessa tilanteessa viittaukset datojen sijainteihin pitää muuttaa myös niille datoilta, joihin ei viitata kirjastoviitteellä vaan jotka on avattu suoraan kansioista projektiin ("File" -> "Open" -> "Data", esimerkiksi Excel-datat). Viittaus muutetaan klikkaamalla prosessidiagrammin datakuvaketta hiiren oikealla painikkeella ja valitsemalla "Properties" -> "Change".

Tehtävä: Tallenna projektisi nimellä **EGkurssi_projekti**.

7 Datan muokkaus

Data pitää yleensä olla valittuna aktiiviseksi, jotta sitä voidaan muokata tai muuten käsitellä eri työkaluilla. Datan saa aktiiviseksi klikkaamalla data-kuvaketta prosessidiagrammissa, jolloin se muuttuu siniseksi. Vaihtoehtoinen tapa tehdä data aktiiviseksi, on valita se työkalurivin ”Active Data” -valikosta.

Tässä luvussa käydään läpi seuraavat Data-valikon työkalut:



7.1 Datan muuttaminen muokkaustilaan

Ennen kuin dataa voidaan muokata suoraan datamatriisissa, sen tulee olla muokkaustilassa. Datan voi muuttaa muokkaustilaan poistamalla mahdollisen valinnan ”Data”-valikon ”Read-only” -kohdasta. Datan pitää olla näkyvässä työtilassa, jotta ”Read-only”-valintaa voi muuttaa.

Tehtävä: Ota työtilaan näkyviin SAS-data **gomez_data** ja poista siitä kirjoitussuojaus.

7.2 Muuttujien perusominaisuuksien määrittely

Muuttujien nimiä, nimien selitystekstejä ja tyyppejä pääsee muokkaamaan klikkaamalla Enterprise Guiden datamatriisissa sarakkeen otsikkoa hiiren oikealla painikkeella ja valitsemalla ”Properties”. Tällöin aukeaa ”General”-välilehti. Kohdassa ”Name” voidaan muuttaa muuttujan nimeä ja kohdassa ”Label” antaa muuttujalle pidempi selitysteksti. Selitysteksti ei tule näkyviin data-matriisiin, mutta se näkyy monissa tulosteissa. Kohdassa ”Type” voidaan puolestaan määrittää muuttuja joko numeeriseksi tai merkkimuotoiseksi. Numeeriset muuttujat voidaan jakaa vielä kohdassa ”Group” tavallisiin numeerisiin (”Numeric”), päivämääriin (”Date”), aikoihin (”Time”) ja rahayksiköihin (”Currency”). Näillä kaikilla muuttujatyypeillä on omia erityisiä työkaluja, joita muuttujiin voidaan soveltaa (informaatteja, formaatteja ja funktioita). Lisäksi kullakin muuttujatyypillä on datamatriisissa erilainen tunnistesymboli, jonka avulla näkee nopeasti minkä tyyppinen muuttuja on kyseessä. Yleensä muuttujilla käytetään joko merkkimuotoa (symboli punainen pyramidi) tai tavallista numeerista muotoa (symboli sininen pallo).

Tehtävä: Määritä **gomez_dataan** suomenkieliset muuttujien nimet ja selitystekstit seuraavasti:

Muuttujan nimi (Name):

- Variety -> lajike
- Nitrogen -> typpi
- Rep -> lohko
- Yield -> sato

Selitysteksti (Label):

- lajike -> Riisilajike
- typpi -> Typpitaso (kg/ha)
- lohko -> Lohko
- sato -> Hehtaarisato (kg/ha)

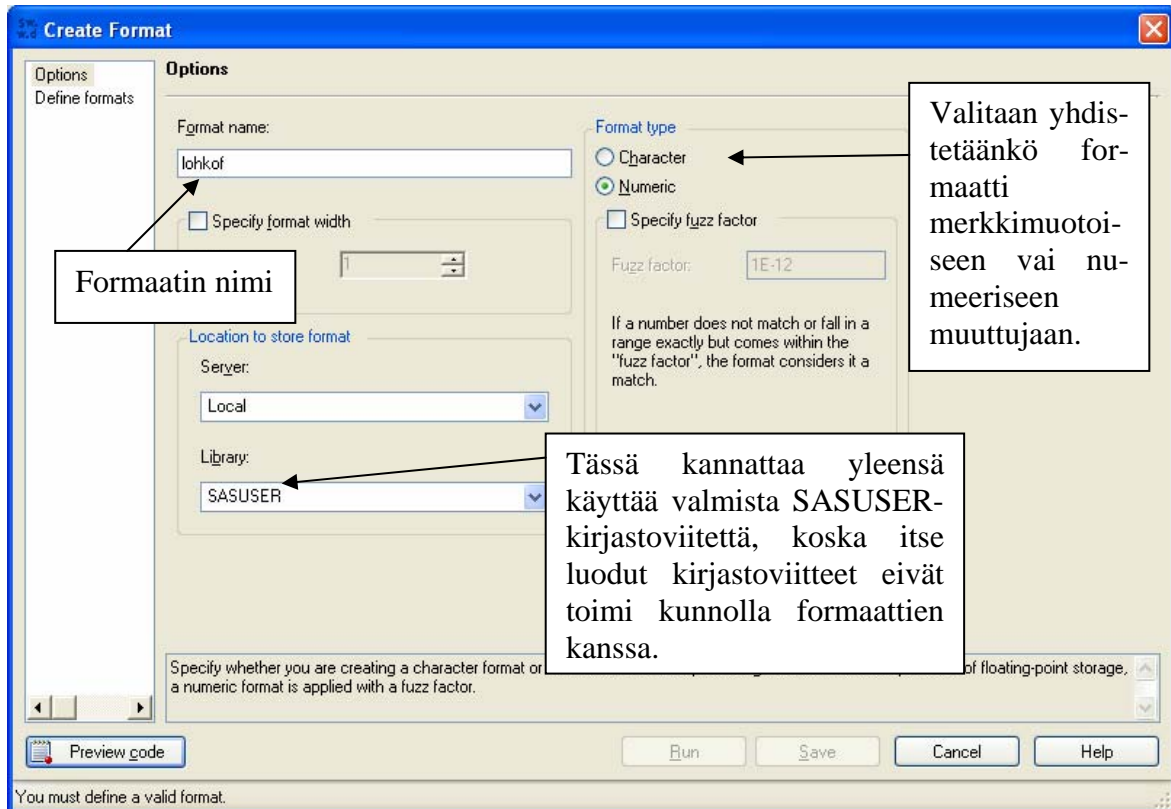
7.3 Muuttujan arvojen tulostusmuotojen eli formaattien käyttäminen

Formaattien käytössä on tilanteesta riippuen kaksi tapaa. Joillakin muuttujilla voidaan käyttää Enterprise Guiden valmiita formaatteja. Niitä löytyy esimerkiksi useille aika- ja päivämäärämuuttujille. Monille muuttujille käyttäjän täytyy kuitenkin itse määrittellä eri arvoihin liittyvät selitteet. Näin on, jos esimerkiksi halutaan suomenkieliset selitystekstit arvoille 1 = ’kissa’ ja 2 = ’koira’.

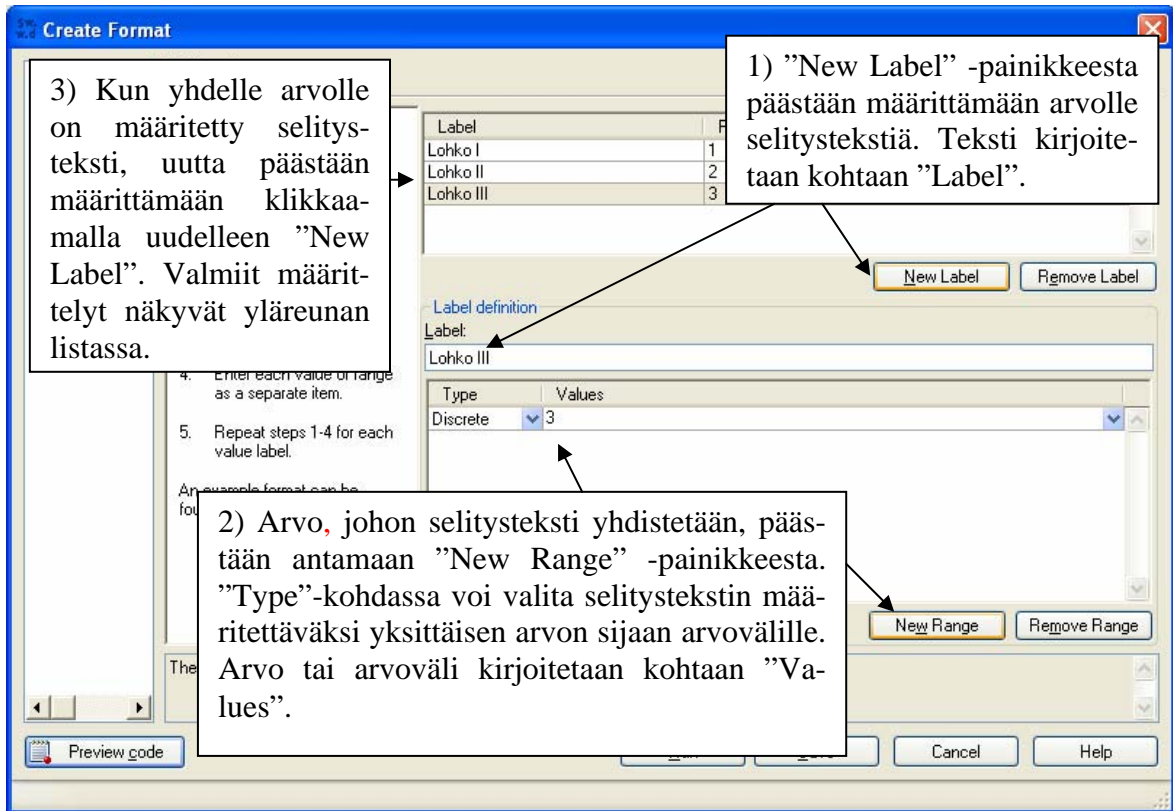
Seuraavaksi käydään läpi, miten formaatti voidaan muodostaa ja miten itse muodostetun tai valmiin formaatin voi yhdistää johonkin muuttujaan.

Formaatin luominen

Formaatteja pääsee tekemään klikkaamalla ”Data” -> ”Create Format”, minkä jälkeen ”Options”-välilehdellä määritellään formaatin nimi, tyyppi (ovatko arvot datassa merkkimuotoisia vai numeerisia) ja tallennuskohde:



Kun formaatti tallennetaan SASUSER-kirjastoviitteen avulla pysyvään kansioon, se tallentuu formaattitiedostoksi ja on aina käytettävissä, koska SASUSER kirjastoviite yhdistyy automaattisesti kun EG käynnistetään. Tiettyihin muuttujan arvoihin liittyvät selitetekstit määritellään ”Define Formats” -välilehdellä:

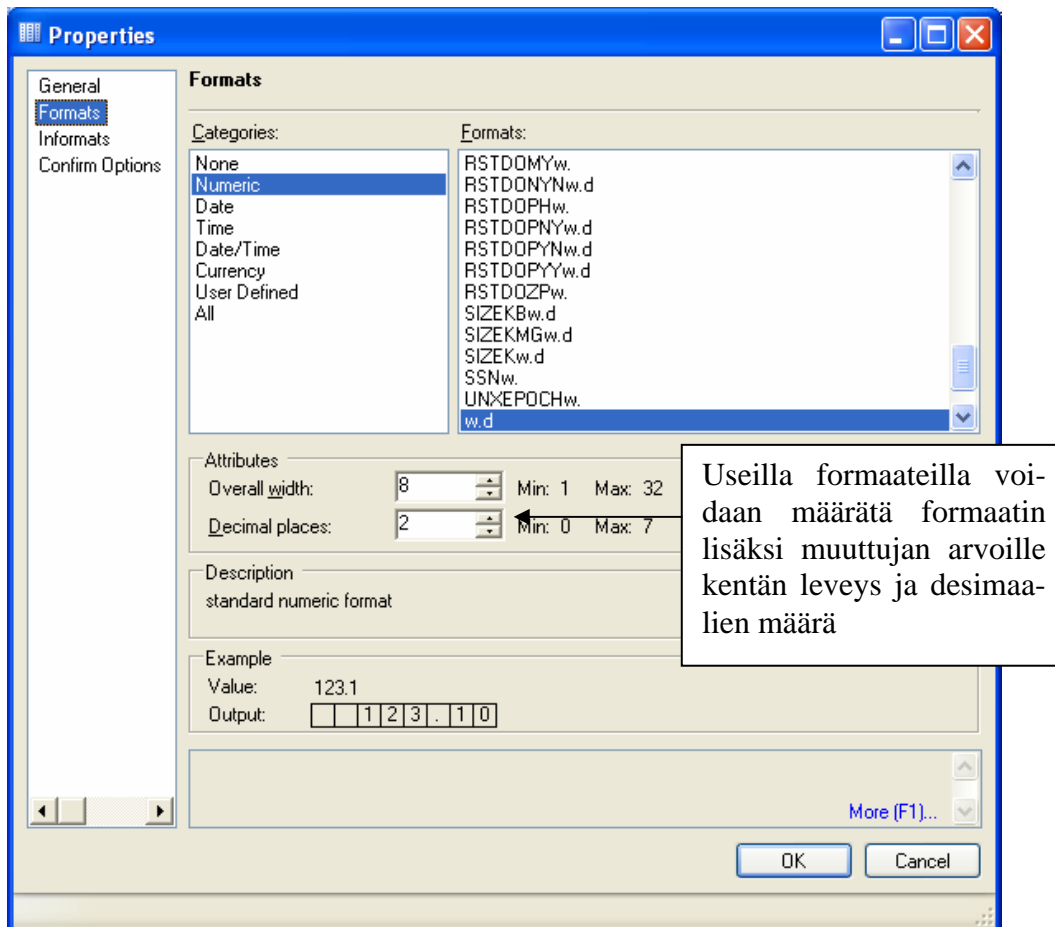


Lopuksi formaatti luodaan valitsemalla "Run". "Save"-valinnan avulla voi tallentaa valinnat kuitenkin vielä suorittamatta tehtävää.

Formaatin yhdistäminen muuttajaan

Ensin tulee varmistaa että data ei ole "Read-only" -tilassa. Formaattit yhdistetään muuttajaan klikkaamalla hiiren oikealla painikkeella muuttujan nimeä ja valitsemalla "Properties". "Formats"-välilehden kategoriasta "User Defined" löytyvät käyttäjän määrittelemät formaatit. Valmiit formaatit löytyvät muista kategorioista. Kun formaatti on valittu, painetaan "OK" ja muuttujien arvot muuttuvat data-matriisissa määritetyn formaatin mukaisiksi. Tyypillisiä muuttujia, joille määritetään valmis formaatti, ovat erilaiset aika- ja päivämäärämuuttujat. Käyttökelpoisia lähinnä päivämääriin liittyviä formaatteja ovat muun muassa DDMMYYw. (päivämäärä järjestyksessä päivä (dd), kuukausi (mm) ja vuosi (yy)) ja erilaiset FIN-alkuiset formaatit (suomalaisille räätälöityjä formaatteja).

Välilehdellä ”Formats” voidaan myös määrittää näytettävien desimaalien määrä numeerisen muuttujan arvoille käyttäen *w.d*-formaattia:



Tehtävä: Muodosta **gomez_datan** lohko-muuttujan arvoille 1, 2 ja 3 selitystekstit:

- 1="Lohko I"
- 2="Lohko II"
- 3="Lohko III"

Luo formaatti ja yhdistä se muuttujaan lohko. Tarkastele datamatriisista, onnistuiko selitystekstien lisääminen muuttujan arvoille.

7.4 Datan muokkauksen työkalut

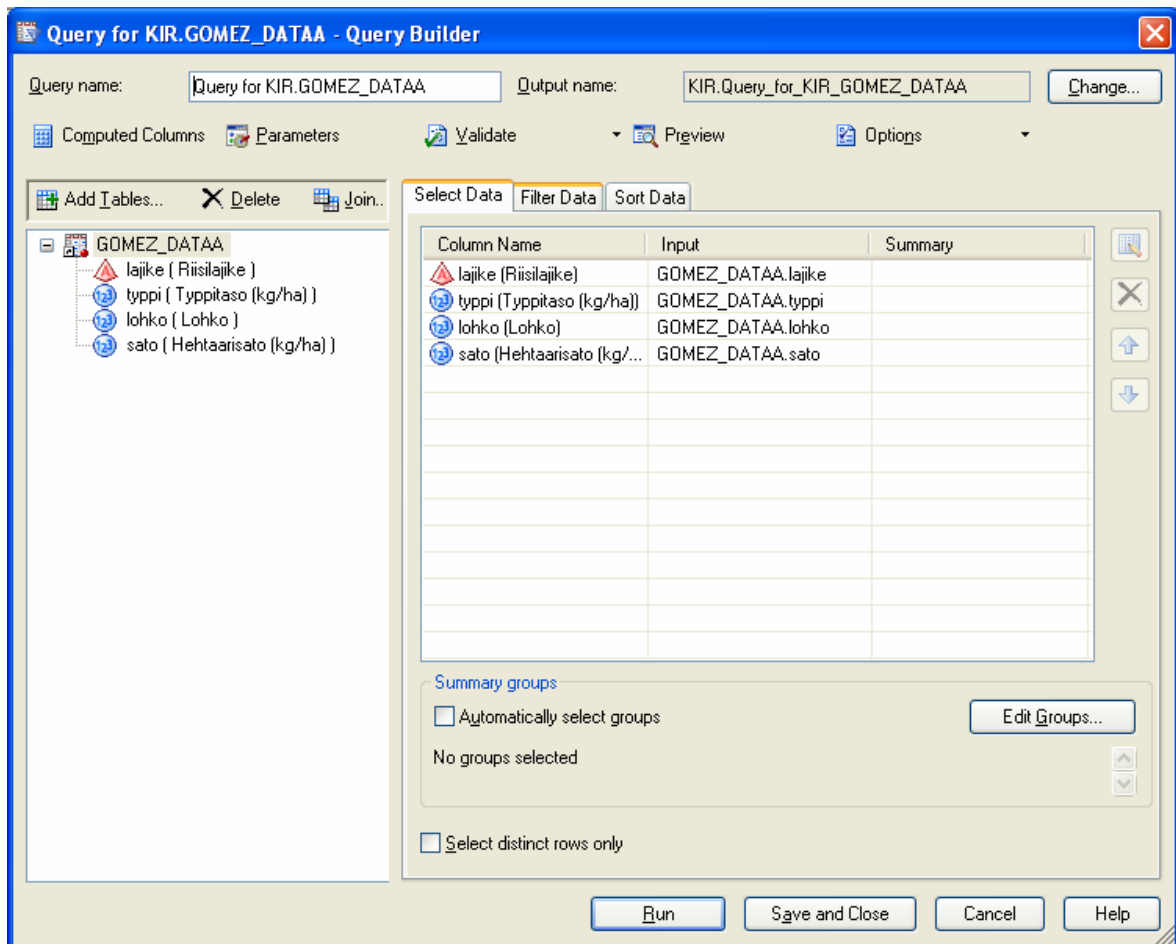
Enterprise Guidessa ensisijainen aineiston muokkaustyökalu on Query Builder, joka on eräänlainen kyselytyökalu ("kysely"-sana tulee SQL-ohjelmointikielestä, koska Query Builder käyttää SASin PROC SQL:ää). Kysely-sanalla viitataan siihen, että työkalun avulla tehdään dataan kysely ja tuloksena saadaan uusi muokattu data. Näiden kyselyjen avulla voidaan muodostaa osa-aineistoja, luoda uusia muuttujia sekä järjestää ja yhdistää aineistoja. Query Builderilla tehdään siis tavallista datan muokkausta. Kaikkea datan muokkausta ei kuitenkaan pystytä tekemään Query Builderilla, vaan tarvitaan myös muita Data-valikosta löytyviä työkaluja. Monet asiat voidaan kuitenkin tehdä sekä Query Builderilla että jollakin vaihtoehdoisella tavalla.

Kun muokataan olemassa olevaa dataa (uusien muuttujien tekeminen, muuttujien poistaminen, havaintorivien lisääminen ja poistaminen sekä muuttujien uudelleennimeäminen), kannattaa yleensä käyttää Query Builderia. Query Builderin vahvuutena on, että alkuperäi-

nen data säilyy koskemattomana ja tuloksena syntyy uusi muokattu data (samoin voi tehdä tavallisessa SASissa). Tällöin datan muokkauksesta jää myös dokumentti EG-projektiin ja muokkauksia voi tarkastella Query-tehtäväsolmusta jälkikäteen.

Enterprise Guidessa voidaan käyttää myös niin sanottuja datamatriisin perustoimintoja: sarakkeita ja rivejä voidaan lisätä ja poistaa samaan tapaan kuin Excelissä, eli kohdistamalla toimenpiteet suoraan datamatriisiin jonkin työkalun sijasta. Nämä toiminnot ovat suoraviivaisempia ja helppokäyttöisempiä kuin Query Builder, mutta niiden kanssa pitää olla varovainen, koska muutokset kohdistuvat suoraan alkuperäiseen dataan ja muutoksista ei jää projektiin dokumenttia. Niinpä näiden käytössä on sama vaara kuin Excelissä, eli dataa saattaa vahingossa hävitä ("Edit" -> "Undo" -toimintokaan ei auta). Myös täysin oikein tehdyissä datan muunnoksissa dokumentointi jää perustoimintoja käytettäessä käyttäjän itsensä varaan. Näitä toimintoja kannattaa käyttää pääasiassa, kun ollaan tallentamassa uutta dataa Enterprise Guidessa tai luomassa tallennuspohjaa. Jos kuitenkin halutaan käyttää niin sanottuja perustoimintoja datan muokkaukseen, kannattaa ehdottomasti ottaa datasta varmuuskopio ennen muokkausta. Varmuuskopiointi onnistuu esimerkiksi klikkaamalla datakuvaketta hiiren oikealla napilla ja valitsemalla "Export".

Dataa päästään muokkaamaan Query Builderilla valitsemalla ensin muokattava data aktiiviseksi prosessidiagrammista (klikataan hiirellä siniseksi) ja klikkaamalla "Data" -> "Filter and Query", jolloin aukeaa seuraava näkymä:



Query Builderista löytyy muun muassa seuraavaa:

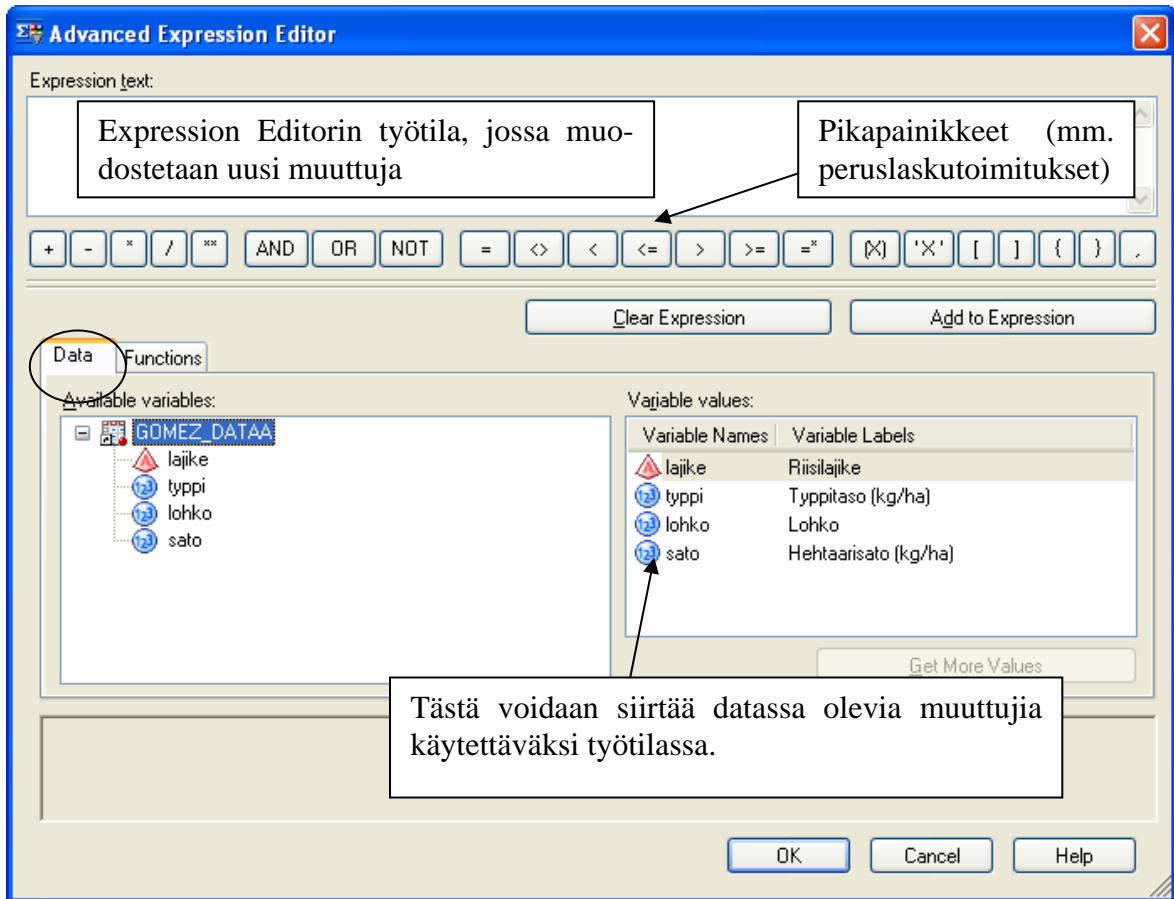
Select Data -välilehti	Lopputuloksena syntyvään dataan mukaan tulevat muuttujat. Tässä voidaan myös tehdä tiettyjä muuttujiin liittyviä toimenpiteitä. Useimmiten ”Tools” -> ”Options” -> ”Query” -valikosta kannattaa tehdä valinta ”Automatically add columns from input tables to result set of query”, jotta tulodataan saadaan alkupe- räisestä datasta lähtökohtaisesti kaikki muuttujat mukaan.
Filter Data -välilehti	Muodostetaan osa-aineistoja eli valitaan jonkin tai joidenkin ehtojen mukaan tietyt rivit mukaan dataan.
Sort Data -välilehti	Aineiston rivien järjestäminen jonkin tai joidenkin muuttujien arvojen mukaiseen järjestykseen.
Computed Columns	Uusien muuttujien tekeminen vanhoista muuttujista luokitte- malla tai funktioiden tai aritmeettisten operaatioiden avulla.
Join	Kahden tai useamman aineiston yhdistäminen rinnakkain siten, että samoille riveille eli tilastoyksiköille tulee muuttujia useista aineistoista.

Query Builderissa luotu kysely dataan suoritetaan valitsemalla ”Run”, mutta sitä ennen kannattaa yleensä antaa tuloksena syntyvälle SAS-datalle nimi klikkaamalla oikeassa ylä-
kulmassa olevaa ”Change”-valintaa. Valitsemalla ”Save and Close” voi tallentaa kyselyn suorittamatta sitä. Tulodataa voi myös esikatsella ennen kyselyn suorittamista painamalla ”Preview”.

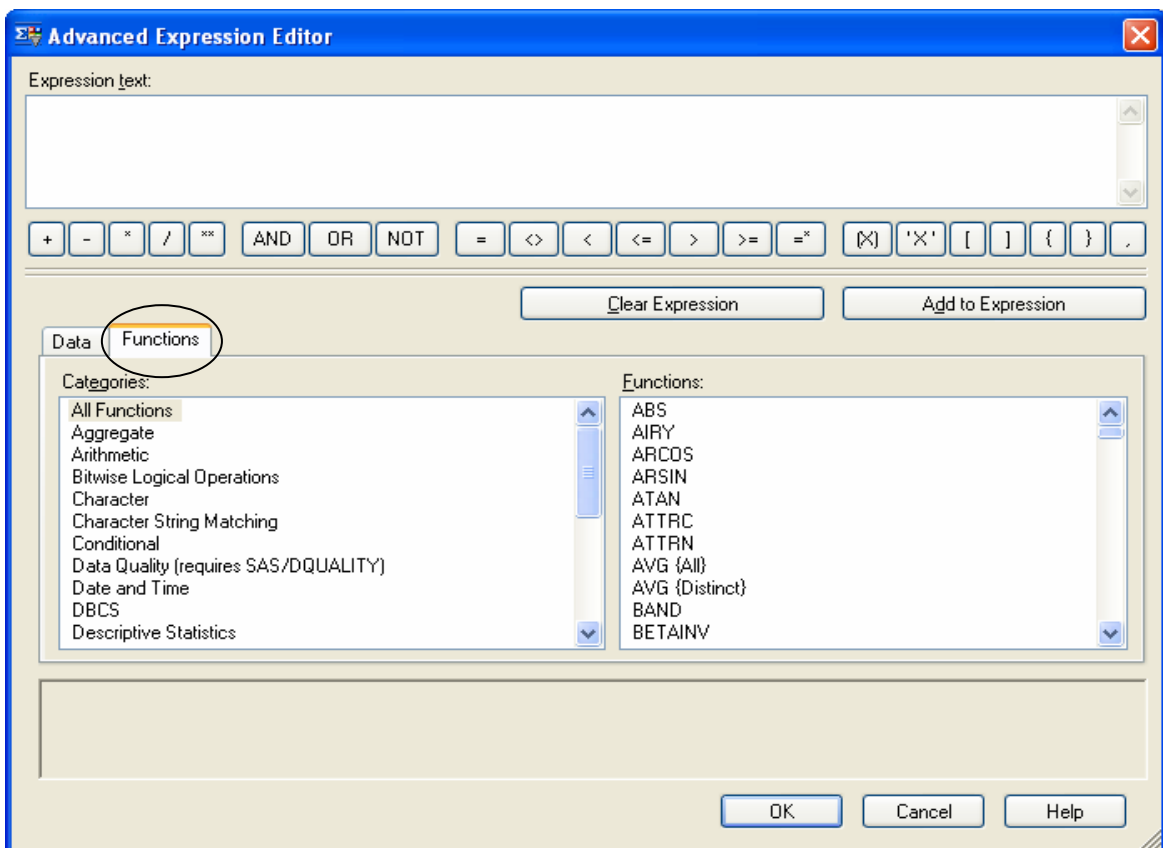
Tehtävä: Tutki data-valikkoa ja avaa Query Builder (**gomez_data** aktiivisena). Tarkastele Query Builderia.

7.5 Muuttujien muunnokset ja uusien muuttujien tekeminen

Muuttujien muunnoksia voi tehdä Expression Editorin avulla. Tähän työkaluun päästään käsiksi Query Builderin kautta siten, että valitaan muokattava data aktiiviseksi ja klikataan ”Data” -> ”Filter and Query” ja valitaan Query Builderista ”Computed Columns” -> ”New” -> ”Build Expression”. Expression Editorissa voidaan tehdä uusia muuttujia sekä tavallisten laskutoimitusten avulla että käyttämällä valmiita funktioita. Näkymä Expression Editorissa on seuraava:



Klikkaamalla välilehteä "Functions" saadaan oikeaan reunaan näkyviin käytettävissä olevat valmiit funktiot, joita voidaan siirtää työtilaan:



Muuttujien nimiä, valmiita funktioita sekä välilehtien yläpuolella olevia pikapainikkeita käyttämällä voidaan vanhoja muuttujia muuntamalla tehdä uusia muuttujia. Muunnoksen voi myös kirjoittaa suoraan yläreunan työtilaan, jos tiedetään käytettävien funktioiden ja muuttujien nimet. Kun muuttujan nimi haetaan listasta, se tulee työtilaan muodossa, jossa muuttujan nimen edessä on myös datan nimi. Datan nimi ei kuitenkaan ole välttämätön, jos muuttujan nimi kirjoitetaan työtilaan käsin listasta hakemisen sijaan. Alla on esimerkkejä muutamista muunnoksista:

Logaritmimuunnos: LOG(sato)

Muuttujan arvot muunnetaan (luonnolliselle) logaritmiasteikolle siten, että haetaan ensin LOG-funktio funktiolistasta ja lisätään se työtilaan ("Add to Expression"). Seuraavaksi poistetaan työtilassa LOG-funktiosta kaikki ylimääräinen (eli "<numValue>") sulkujen sisältä ja lisätään sen paikalle sato-muuttuja (valitaan "Data"-välilehdeltä ja klikataan "Add to Expression" tai kirjoitetaan suoraan sulkujen sisälle "sato").

Neliöjuurimuunnos: SQRT(sato)

Funktio ja muuttujan nimi haetaan listoilta samoin kuin edellisessä esimerkissä.

Aritmeettinen muunnos (esim. sato tonneissa): sato/1000

Muuttujan nimi voidaan taas hakea listalta, jakolaskuoperaattori löytyy pikapainikkeista ja luku 1000 syötetään käsin.

Lopuksi painetaan OK, jolloin palataan Computed Columns -ikkunaan, jossa voidaan antaa tuloksena syntyvälle uudelle muuttujalle nimi valitsemalla "Calculation1" (tai vastaava listalla oleva uuden muuttujan nimi) ja klikkaamalla "Rename". Tämän jälkeen painetaan "Close". Lopuksi painetaan "Run". Query Builderista jää myös oma tehtäväkuvake prosessidiagrammiin, josta muunnoksia pääsee halutessaan myöhemmin muokkaamaan. Query Builder luo tuloksena aina uuden datan, johon muunnosten avulla luodut muuttujat on lisätty. Jos halutaan itse valita tuloksena syntyvän datan nimi ja kohde, johon se tallennetaan (kirjastoviite), sen voi tehdä Query Builderissä klikkaamalla oikeassa yläkulmassa olevaa "Change"-valintaa, jossa "Save in:" -kohdassa voi valita kirjastoviitteen ja "File name:" -kohdassa tuloksena syntyvän SAS-datan nimen.

Jos myöhemmin halutaan palata muokkaamaan muuttujan määrittelyä, se onnistuu klikkaamalla prosessidiagrammista auki tehtäväsolmu, jossa määrittelyt on tehty, klikkaamalla "Computed Columns", valitsemalla listasta muuttuja, jonka määrittelyä halutaan muokata, ja valitsemalla "Edit". Kun muokkaus on tehty, pitää tehtäväsolmu suorittaa uudelleen valitsemalla Query Builderin päänäkymässä "Run".

Tehtävä: Tee **gomez_dataan** uudet muuttujat tsato (sato tonneina) ja sqrt_sato (saton neliöjuuri). Käytä tähän Query Builderia. Anna tulosdatalle nimeksi **gomez_muok**.

7.6 Muuttujien luokittelu ja luokkien yhdistäminen

Jatkuvien muuttujien arvojen luokittelu ja kategoristen muuttujien luokkien yhdistäminen voidaan myös tehdä Query Builderin avulla. Molemmat tehdään käyttäen samanlaista tekniikkaa. Seuraavassa esitetään, kuinka jatkuvan muuttujan arvot saa luokitetuksi.

Ensin avataan Query Builder ("Data" -> "Filter and Query") aktiiviselle datalle. Liikelle lähdetään aivan samoin kuin muunnoksissa eli ensin valitaan "Computed Columns" -> "New", mutta sitten valitaankin "Recode a Column" ja aukeavasta listasta muuttuja, joka luokitellaan. Seurauksena aukeaa "Recode Column" -ikkuna, jossa muuttujan luokittelu määritellään. Klikkaamalla "Add" saadaan lisättyä yhden luokan määrittely kokonaisuuteen:

Valitaan "Replace a range".

Nuoli-painikkeista saadaan näkyviin lista alkuperäisen muuttujan arvoista.

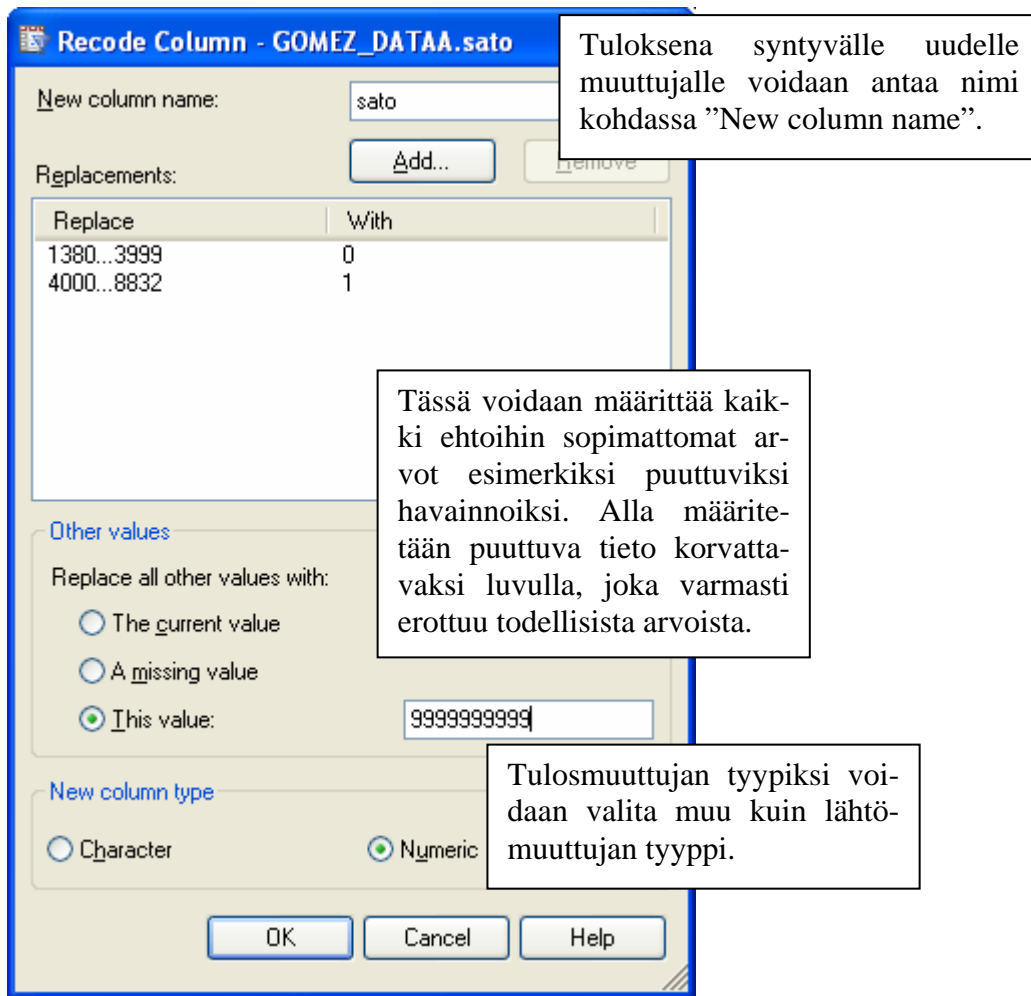
Voidaan määrittää kuhunkin luokkaan kuuluvien arvojen yläraja, alaraja tai molemmat.

Tässä määritetään uusi arvo.

Huomaa, että numeerisen muuttujan puuttuvat arvot koodataan pisteellä. Sen avulla voi asettaa esimerkiksi virheellisen mittauksen puuttuvaksi. Merkkimuotoisella muuttujalla puuttuvat jätetään tyhjiksi.

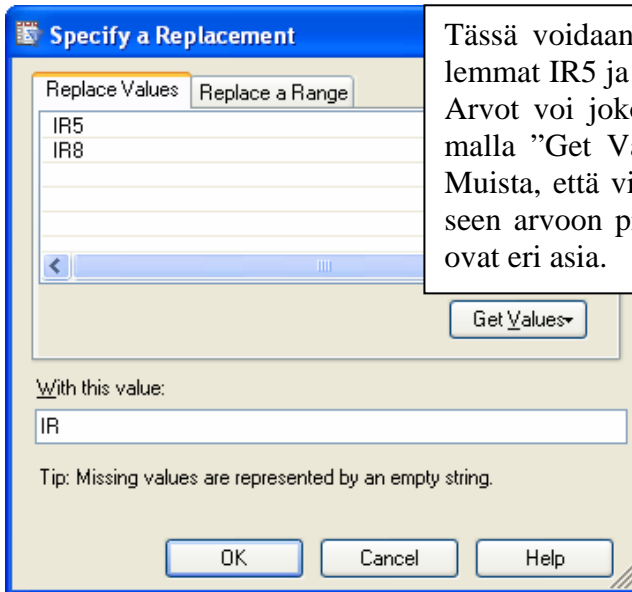
Tip: Missing values are represented by a period.

Kun kaikki luokat on määritelty, näkyy kokonaisuus ”Recode Column” -ikkunassa:



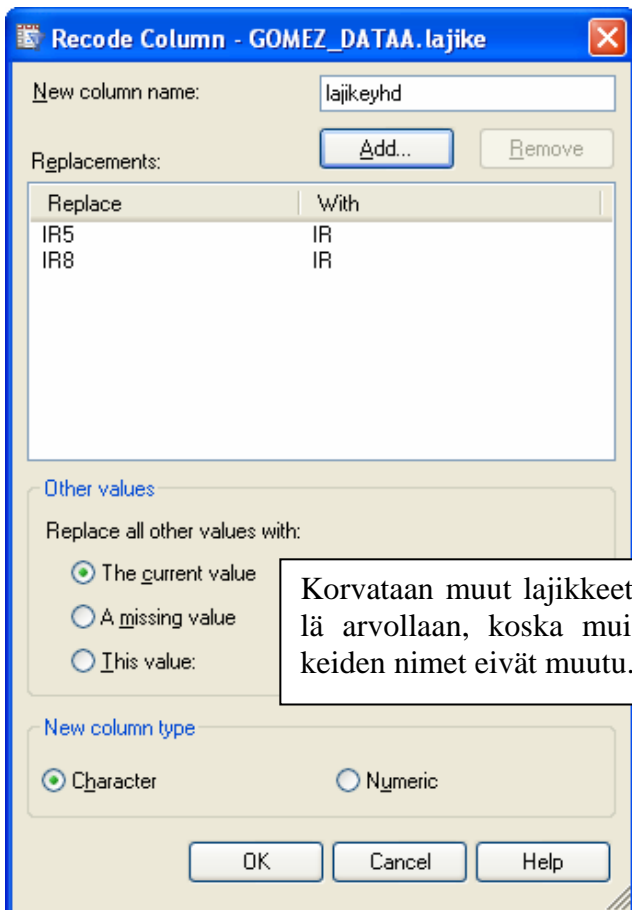
Lopuksi painetaan ”OK”, ”Close” ja ”Run Query”.

Luokittelumuuttujan luokkien yhdistäminen voidaan aivan samalla tavalla kuin jatkuvan muuttujan arvojen luokittaminen, mutta ylä- ja alarajoiksi tulee vain alkuperäisen luokkamuuttujan arvoja. Vaihtoehtoisesti voidaan käyttää toimintoa ”Replace”, jolla voidaan korvata yksi arvo kerrallaan toisella arvolla. Pelkkää ”Replace”-toimintoa käyttäen gomez_datassa voidaan yhdistää lajikkeet IR5 ja IR8 yhdeksi lajikkeeksi IR (”Replace Values” -kohtaan on päästy samoin kuin edellä jatkuvan muuttujan arvoja luokiteltaessa):



Tässä voidaan samalla määrittellä molemmat IR5 ja IR8 korvattavaksi IR:llä. Arvot voi joko hakea listasta klikkaamalla "Get Values" tai kirjoittaa itse. Muista, että viitattaessa merkkimuotoiseen arvoon pieni kirjain ja iso kirjain ovat eri asia.

Lopuksi "Recode Column" -ikkuna näyttää seuraavalta:



Korvataan muut lajikkeet nykyisellä arvolla, koska muiden lajikkeiden nimet eivät muutu.

Lopuksi painetaan "OK", "Close" ja "Run Query".

Luokkien yhdistämistä voi tehdä myös "Expression Editorin" ("Computed Columns" -> "New" -> "Build Expression") avulla käyttäen Case-When-rakennetta (funktiolistassa CASE {short}). Rakenteen käyttöä voi harjoitella esimerkiksi seuraavien tehtävien yhteydessä. Lisätietoa aiheesta löytyy muun muassa kirjasta Vahlberg ym. (2006).

Jos muuttujan määrittelyä halutaan palata muokkaamaan myöhemmin, se onnistuu klikkaamalla prosessidiagrammista auki tehtäväsolmu, jossa määrittely on tehty, klikkaamalla

”Computed Columns”, valitsemalla listasta muuttuja, jonka määrittelyä halutaan muokata, ja valitsemalla ”Edit”. Kun muokkaus on tehty, tehtäväsolmu pitää suorittaa uudelleen valitsemalla Query Builderin päänäkyessä ”Run”.

Tehtävä: Tee **gomez_dataan** uusi sato-muuttuja, joka saa arvot 1 ja 2 seuraavasti:

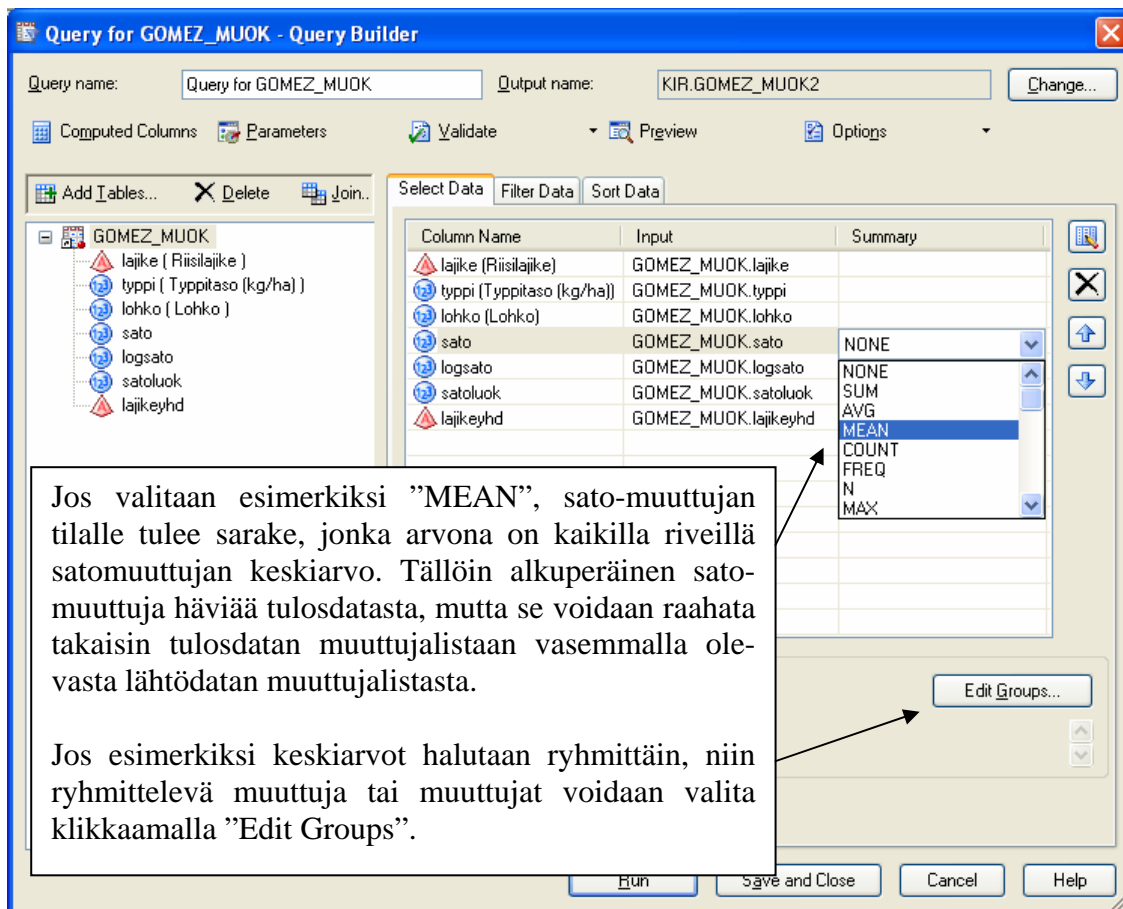
- 1 = ”alle 4000 kg/ha”
- 2 = ”4000 kg/ha tai enemmän”

Tehtävä: Tee **gomez_dataan** lajike-muuttujasta uusi muuttuja laj_tyyppi, jossa lajikkeet saavat seuraavat arvot:

- IR5 -> IR
- IR8 -> IR
- C4_63 -> C4_63
- Peta -> Peta

7.7 Muuttujan luominen aritmeettisella operaatiolla

Toisinaan on tarpeen luoda muuttuja, jonka arvo on osalle riveistä sama tai kaikille riveille yhteinen vakio. Esimerkkinä tilanne, jossa kaikilla riveillä on sato-muuttujan keskiarvo. Tällaisen vakion luominen onnistuu Query Builderin (”Data” -> ”Filter and Query”) ”Select Data” -välilehdellä. Klikkaamalla Query Builderissa ”Select Data” -välilehdeltä solua, joka on ”Summary”-sarakkeen ja satomuuttujan rivin risteyskohdassa, saadaan näkyviin seuraava valikko:



Query for GOMEZ_MUOK - Query Builder

Query name: Query for GOMEZ_MUOK Output name: KIR.GOMEZ_MUOK2 Change...

Computed Columns Parameters Validate Preview Options

Add Tables... Delete Join...

Column Name	Input	Summary
lajike (Riisilajike)	GOMEZ_MUOK.lajike	
typpi (Tyyppitaso (kg/ha))	GOMEZ_MUOK.typpi	
lohko (Lohko)	GOMEZ_MUOK.lohko	
sato	GOMEZ_MUOK.sato	NONE
logsato	GOMEZ_MUOK.logsato	NONE
satoluok	GOMEZ_MUOK.satoluok	SUM
lajikeyhhd	GOMEZ_MUOK.lajikeyhhd	AVG

MEAN

COUNT

FREQ

N

MAX

Edit Groups...

Run Save and Close Cancel Help

Jos valitaan esimerkiksi ”MEAN”, sato-muuttujan tilalle tulee sarake, jonka arvona on kaikilla riveillä satomuuttujan keskiarvo. Tällöin alkuperäinen sato-muuttuja häviää tulosdatasta, mutta se voidaan raahata takaisin tulosdatan muuttujalistaan vasemmalla olevasta lähtödatan muuttujalistasta.

Jos esimerkiksi keskiarvot halutaan ryhmittäin, niin ryhmittelevä muuttuja tai muuttujat voidaan valita klikkaamalla ”Edit Groups”.

Tehtävä: Tee **gomez_dataan** uusi muuttuja, jonka arvot ovat lohkokohtaisia satokeskiarvoja. Ota myös alkuperäinen satomuuttuja mukaan tulosdataan.

7.8 Muuttujien poistaminen

Muuttujien poistaminen kannattaa tehdä Query Builderin (**"Data"** -> **"Filter and Query"**) avulla seuraavasti: valitaan **"Select Data"** -välilehdellä muuttuja ensin aktiiviseksi ja sitten painetaan oikeassa reunassa olevaa X-painiketta.

Tehtävä: Poista edellisissä tehtävissä tehdyt luokka-muuttujat ja satokeskiarvomuuttuja.

7.9 Muuttujien järjestyksen vaihtaminen datassa

Muuttujien järjestyksen vaihtaminen onnistuu myös Query Builderin (**"Data"** -> **"Filter and Query"**) **"Select Data"** -välilehdellä. Muuttujan paikkaa voi vaihtaa listassa ja siten myös tulosdatassa klikkaamalla muuttuja aktiiviseksi ja painamalla oikeassa reunassa olevia valintapainikkeita, joissa nuolet osoittavat alas ja ylös.

Tehtävä: Harjoittele muuttujien järjestyksen vaihtamista käyttäen edellisen tehtävän tuloksena saatua dataa.

7.10 Datan havaintorivien järjestäminen

Datamatriisin selaamista ja asioiden hahmottamista helpottaa usein, jos havaintorivit järjestetään jonkin muuttujan arvojen mukaan esimerkiksi suuruusjärjestykseen tai aakkosjärjestykseen. Havaintojen järjestäminen ei ole Enterprise Guidessa kuitenkaan yhtä tärkeää kuin perus-SASissa, koska EG osaa pääasiassa järjestää datan automaattisesti, jos jokin tehtävä sitä vaatii. Kuitenkin esimerkiksi **"Line Plot"** -kuvissa eli viivakuviissa täytyy itse järjestää aineisto vaakakselille tulevan muuttujan mukaan, jotta viivat kulkisivat jatkumona vasemmalta oikealle (asiasta lisää kappaleessa **Viivakuva (Line Plot)**).

Järjestäminen voidaan tehdä EG:ssä joko Query Builderin avulla tai käyttäen työkalua **"Data"** -> **"Sort Data"**. Tässä käytetään kuitenkin Query Builderia, koska silloin voidaan yhdistää yhteen kyselyyn useita datan muokkaustoimenpiteitä. **"Sort Data"** -työkalussa on kuitenkin ominaisuuksia, joita Query Builderin vastaavasta ei löydy, joten siihenkin kannattaa tutustua.

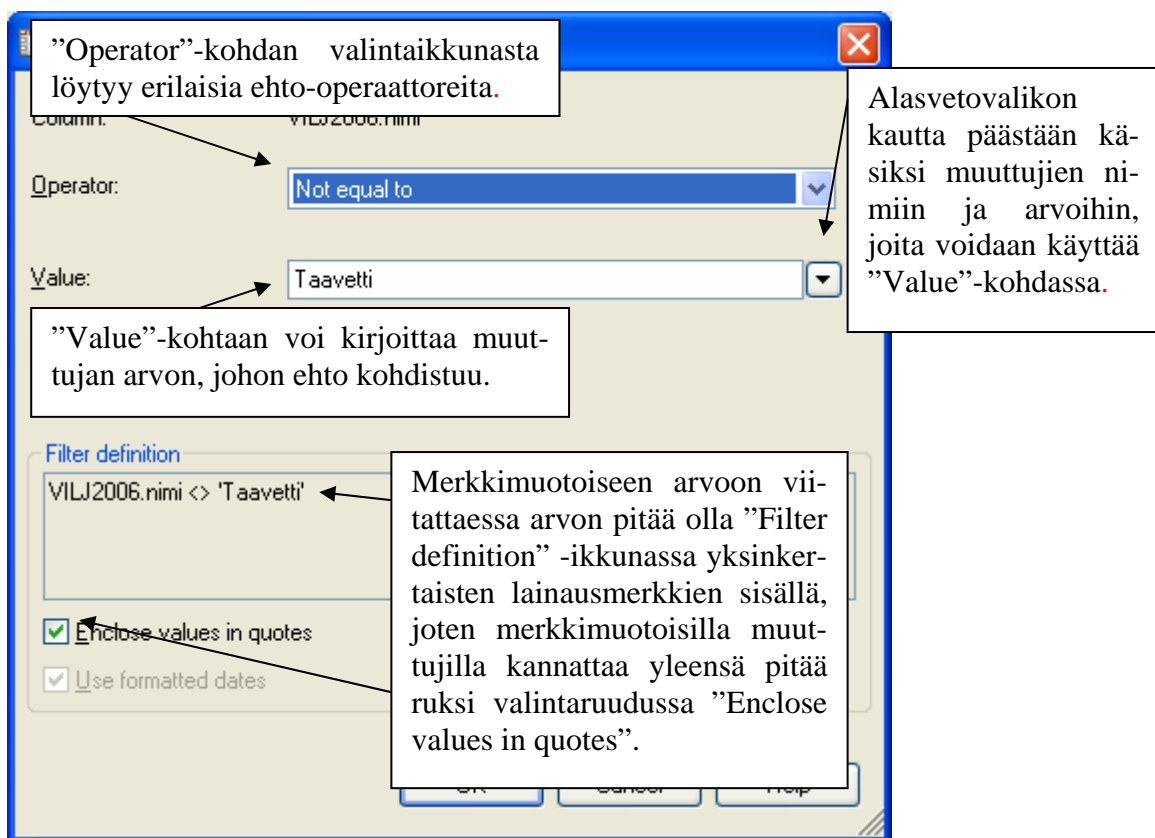
Query Builderissa uusi kysely saadaan auki valitsemalla **"Data"** -> **"Filter and Query"**. Jos taas halutaan muokata jo olemassa olevaa kyselyä, se onnistuu tupla-klikkaamalla prosessidiagrammissa olevaa **"Query"**-tehtäväsolmua. Järjestäminen tehdään siten, että vasemmalla olevasta muuttujalistasta raahataan **"Sort Data"** -välilehdelle se muuttuja tai ne muuttujat, joiden mukaan järjestäminen tehdään. Tämän jälkeen valitaan **"Sort Direction"** -kohdasta joko **"Ascending"** tai **"Descending"**, riippuen halutaanko nouseva vai laskeva rivien järjestäminen. Muuttujien järjestys **"Sort Data"** -välilehdellä määrittelee muuttujien priorisoinnin järjestämisen yhteydessä. Rivit laitetaan ensin järjestykseen ensimmäisen muuttujan arvojen mukaan ja sitten ensimmäisen muuttujan kunkin arvon sisällä toisen muuttujan arvojen mukaan.

Tehtävä: Järjestä edellisessä tehtävässä käytetyn aineiston rivit suuruusjärjestykseen suurimmasta pienimpään sadon mukaan. Tee järjestäminen muokkaamalla edellisessä tehtävässä luotua tehtäväsolmua.

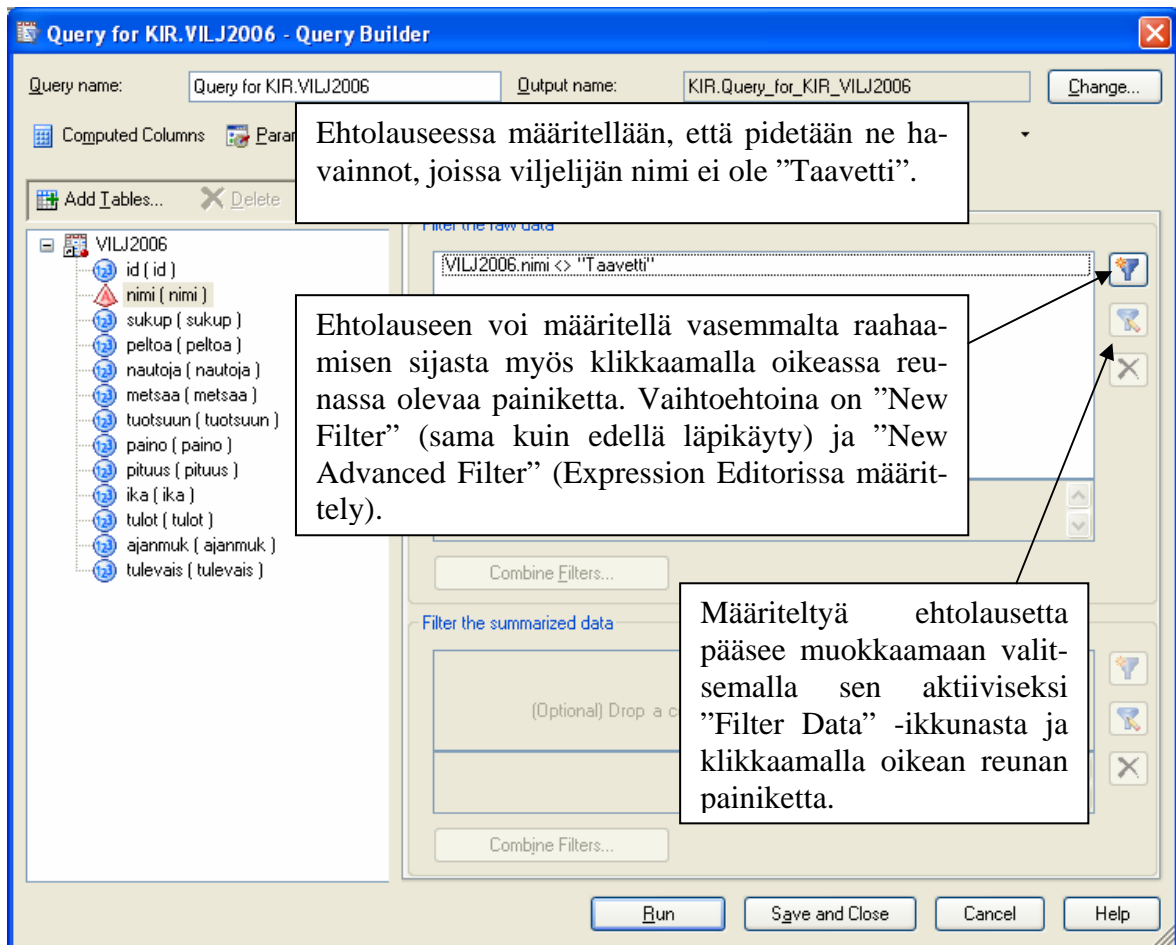
7.11 Osa-aineiston eli tiettyjen havaintorivien valitseminen

Data täytyy usein jakaa osiin, kuten esimerkiksi silloin, kun halutaan valita mukaan analyysiin vain tietty osaryhmä koko datasta. Osittaminen voidaan toteuttaa määrittämällä tietty havaintojen valintaehto Query Builderissa. Myös useissa myöhemmin läpikäytävissä aineistoa kuvien tai tunnuslukujen avulla havainnollistavissa EG:n työkaluissa voidaan määrittellä ”Group analysis by” -muuttuja tai vastaava, jonka avulla saadaan tehdyksi tarkastelut erikseen muuttujan kunkin luokan määrittelemälle osaryhmälle.

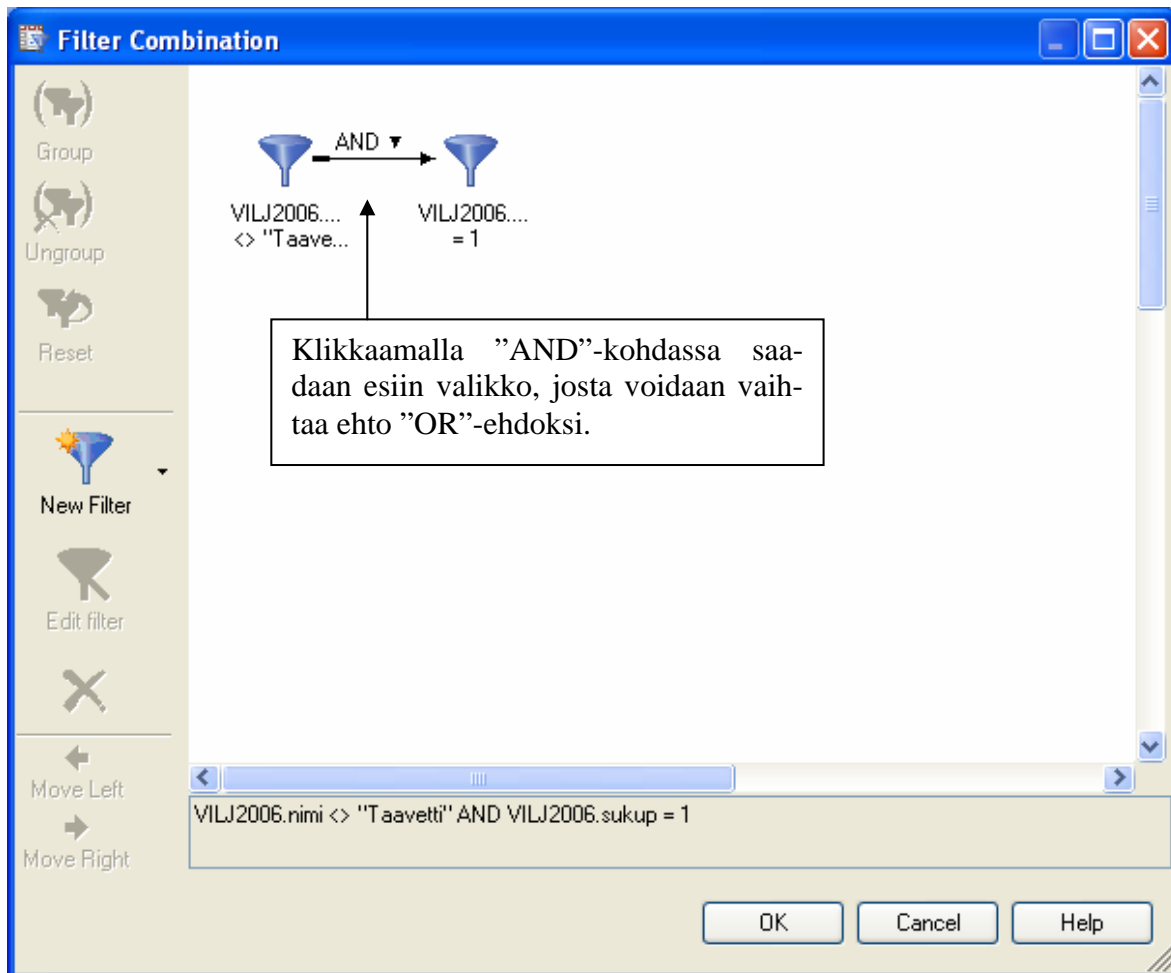
Havaintorivien poistaminen kannattaa Enterprise Guidessa tehdä käyttäen Query Builderia (”Data” -> ”Filter and Query”), koska tällöin muutoksista jää jälki, josta voidaan seurata, mitä muutoksia on tehty, ja tarvittaessa palata alkuperäiseen dataan. Dataa karsitaan Query Builderin ”Filter Data” -välilehdellä. Tämä tapahtuu raahaamalla muuttuja, johon karsintaehto halutaan kohdistaa, vasemman reunan laatikosta ”Filter Data” -ikkunaan. Tällöin aukeaa seuraava ehdon asettamisikkuna:



Kun klikataan ”OK”, ehtolause tulee näkyviin ”Filter Data” -välilehdelle:



Havaintojen suodatuksessa voidaan myös käyttää useampaa kuin yhtä muuttujaa. Tällöin tarvitaan AND (ja) tai OR (tai) -ehtoja tai molempia. Voidaan esimerkiksi määrittellä, että poistetaan sellaiset henkilöt datasta, jotka ovat Jokioisista ja työskentelevät MTT:ssä tai vaihtoehtoisesti voidaan poistaa henkilöt, jotka ovat Jokioisista tai työskentelevät MTT:ssä. Jos yllä olevaan ikkunaan määritetään useampi ehto, ohjelma yhdistää ehdot automaattisesti, mutta käyttäjän täytyy usein käydä vaihtamassa käyttöön oikea yhdistämisoperaattori (AND tai OR). Tämä onnistuu klikkaamalla ”Combine Filters” -painiketta, jolloin aukeaa seuraava ikkuna:



Tehtävä: Ota käyttöön aineisto **Viljelija2006.xls** omalle välilehdelleen prosessidiagrammiin. Tee siitä ensin kirjastoviitteen kir osoittamaan kansioon pysyvä SAS-data nimeltä **vilj2006**. Muokkaa dataa poistamalla poikkeava havainto nimeltä Taavetti ja tallenna tulodata pysyvänä SAS-datana nimellä **vilj2006_muok**.

Tehtävä: Tehdään **vilj2006_muok**-aineistosta kolme erillistä osa-aineistoa, jotka tallennetaan myöhempää käyttöä varten erillisinä SAS-datatiedostoina. Tee ensin kysely, jossa otat **vilj2006_muok**-aineistosta mukaan muuttujat id, nimi, sukup, paino, pituus ja ika, ja tallenna data kirjastoviitteen kir osoittamaan kansioon nimellä **vilj_tiedot** (ks. luku **Muuttujien poistaminen**). Tee uusi kysely **vilj2006_muok** dataan ja ota tuloksena syntyvään uuteen dataan mukaan muuttujat id, peltoa, nautoja, metsaa, tuotsuun, tulot, ajanmuk ja tulevais, ja tallenna data kir-kirjastoviitteen osoittamaan kansioon nimellä **tila_tiedot**. Valitse nyt aktiiviseksi data **vilj_tiedot** ja jaa se kahdeksi erilliseksi dataksi siten, että ensimmäiseen tulevat naiset (sukup = 2) ja toiseen miehet (sukup = 1). Tee datoina pysyviä ja anna niille nimiksi **vilj_nainen** ja **vilj_mies**.

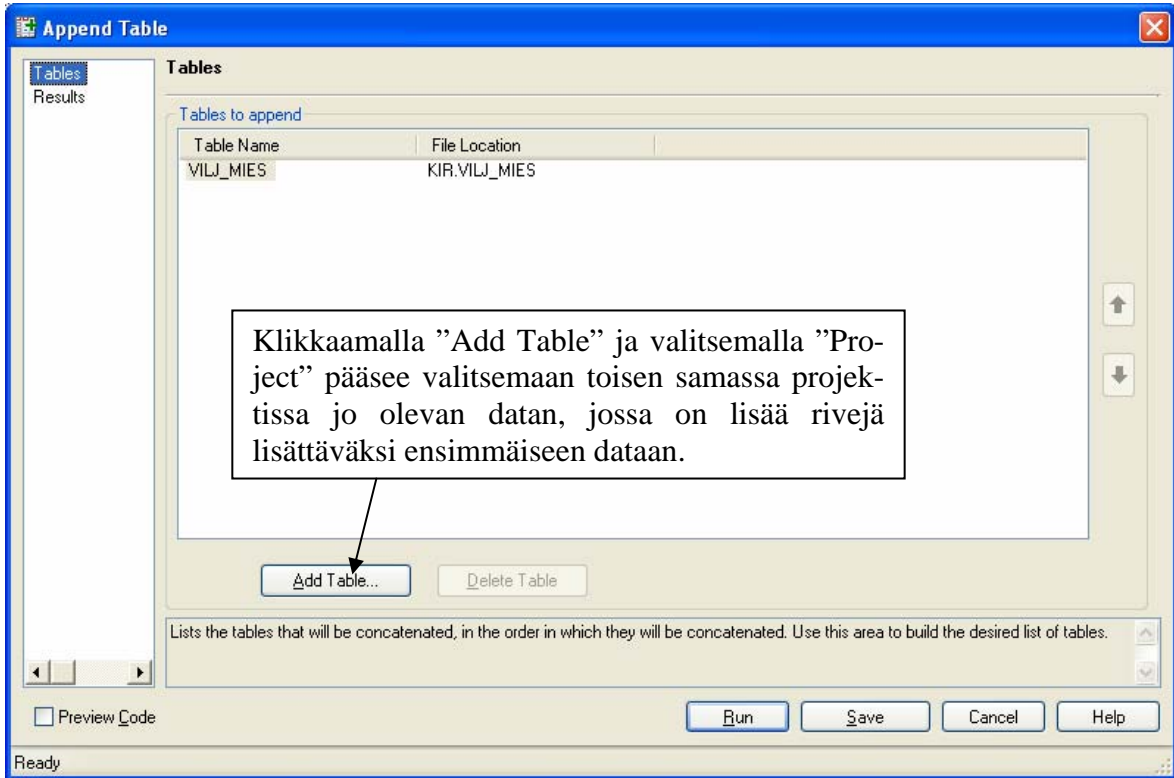
7.12 Kahden tai useamman datan yhdistäminen

Usein tutkimusaineistot on syystä tai toisesta tallennettu monena erillisenä datana, mutta esimerkiksi tilastollista käsittelyä varten ne pitää saada yhdistettyä. Ensimmäinen mahdollinen tilanne on sellainen, jossa yhdistettävissä datoina on samat muuttujat, mutta havainnot ovat jakautuneet erillisiin tiedostoihin. Tällöin datat yhdistetään peräkkäin. Rinnakkain sen sijaan yhdistetään, kun useammassa eri datassa on havaintoja samoilta tilastoyksiköiltä,

mutta eri muuttujista. Tällöin yhdistetään siten, että eri datojen saman tilastoyksikön havainnot tulevat yhdistetyssä datassa samalle riville.

Datojen yhdistäminen peräkkäin

Datoja voidaan Enterprise Guidessa yhdistää peräkkäin valitsemalla ensin ensimmäinen yhdistettävä data aktiiviseksi prosessidiagrammissa ja klikkaamalla ”Data” -> ”Append Table”, jolloin aukeaa seuraava valintaruutu:

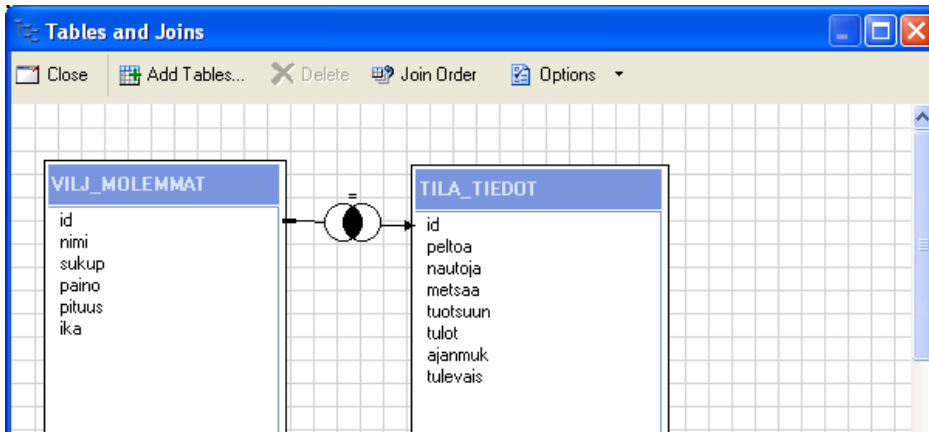


”Results”-välilehdellä voi puolestaan klikkaamalla ”Browse” määritellä, minkä nimisenä ja mihin kansioon lopullinen tulosdata tallennetaan. Lopuksi painetaan ”Run”, kun yhdistettävät datat on valittu yllä olevaan ruutuun.

Tehtävä: Yhdistä aiemmin muodostetut datat **vilj_nainen** ja **vilj_mies** peräkkäin (takaisin yhteen) ja tallenna yhdistetty data nimellä **vilj_molemmat**.

Datojen yhdistäminen rinnakkain

Datojen yhdistäminen rinnakkain voidaan tehdä Query Builderilla. Valitaan ensimmäinen data aktiiviseksi ja klikataan ”Data” -> ”Filter and Query”. Tämän jälkeen painetaan ”Join”-valintapainiketta, jolloin aukeaa ”Tables and Joins” -ikkuna. Ikkunassa voidaan lisätä toinen data painamalla ”Add Tables”, jolloin EG kysyy datan sijaintia. Jos toinenkin data on jo mukana projektissa, voidaan klikata ”Project” ja valita kyseinen data. Tällöin ohjelma yhdistää datat ja näkymä on seuraava:



Jos molemmissa datoissa on samanniminen sarake, niin EG yhdistää datat automaattisesti riveittäin sen mukaan (ID yllä). Jos taas samaa tarkoittava sarake löytyy, mutta se on datoissa eri nimillä, käyttäjä voi itse vetää yhdysviivan sarakkeiden välille. Se onnistuu ottamalla ensimmäisessä datassa muuttujan nimestä hiirellä kiinni ja vetämällä sen toisen päälle. Jos datat yhdistetään useamman kuin yhden muuttujan perusteella (esimerkiksi vuosi- ja kuukausitieto), ainoastaan ensimmäinen muuttujapari yhdistyy automaattisesti ja muut joutuu yhdistämään käsityönä.

Oletuksena yhdistämisessä syntyvään dataan otetaan mukaan vain sellaiset rivit, jotka löytyvät molemmista alkuperäisistä datoista. Näitä yhdistysehtoja pääsee muokkaamaan klikkaamalla hiiren oikealla painikkeella edellisessä kuvassa näkyvän yhdistämissiivan päällä ja valitsemalla ”Modify Join”. Tällöin aukeaa seuraava ikkuna:

- 1) Vain molemmista datoista löytyvät yhdistämisessä käytetyn muuttujan tasot otetaan mukaan tulosdataan.
- 2) Vasemmanpuoleisesta datasta otetaan kaikki rivit ja oikeanpuoleisesta vain siihen yhteensopivat.
- 3) Oikeanpuoleisesta datasta otetaan kaikki rivit ja vasemmanpuoleisesta vain siihen yhteensopivat.
- 4) Molemmista datoista otetaan mukaan kaikki rivit.

Lopuksi valitaan ”OK”, ”Close” ja ”Run”, jolloin datat yhdistyvät. Vaihtoehtoisesti yhdistämisessä voidaan lähteä liikkeelle klikkaamalla ensin ”Add Tables”, jolloin haetaan ensin toinen data, jota ollaan yhdistämässä, ja vasta tämän jälkeen klikataan ”Join”, josta päästään tarkastelemaan ja muokkaamaan datojen automaattista yhdistämistä.

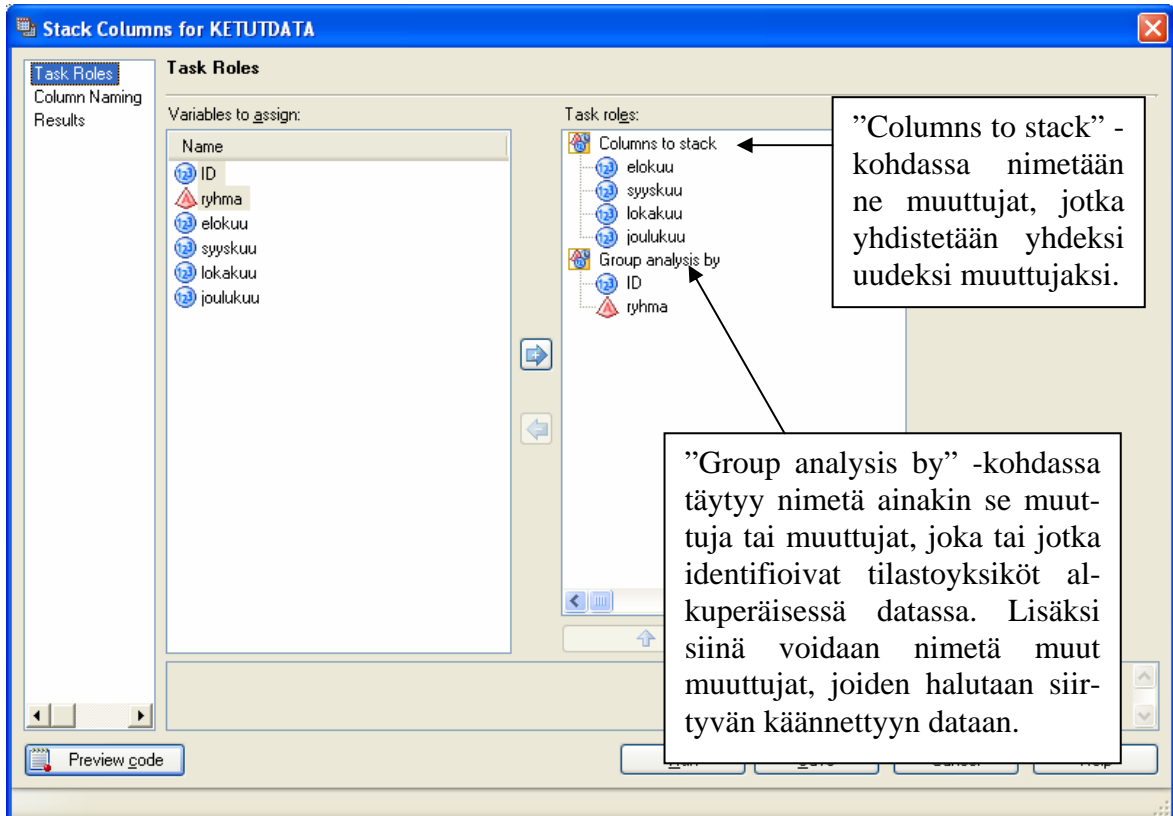
Tehtävä: Yhdistä rinnakkain edellisen tehtävän tuloksena syntynyt **vilj_molemmat** sekä aikaisemmassa tehtävässä luotu **tila_tiedot**. Tee lopputuloksena pysyvä SAS-data nimeltä **vilj06_takaisin**.

7.13 Datat käänttäminen

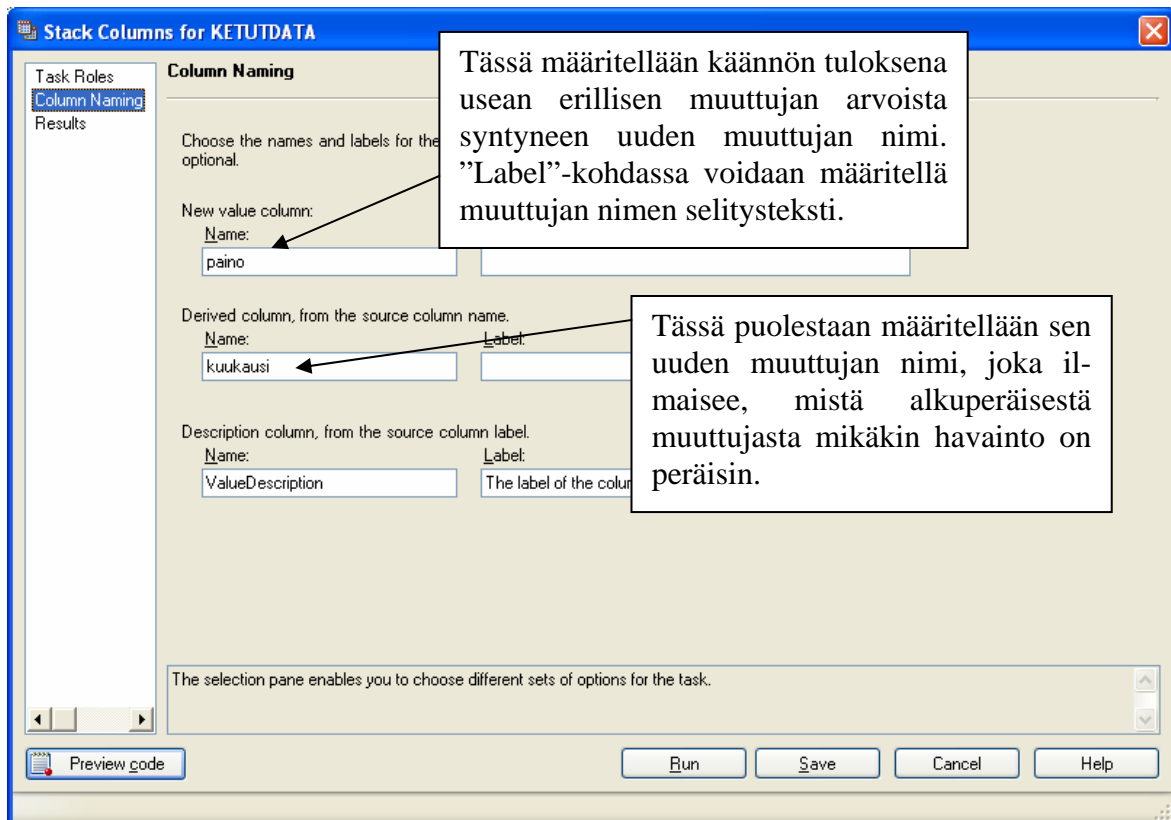
Dataa voidaan kääntää kahteen suuntaan. Ensinnäkin voidaan yhdistää usean sarakkeen tiedot yhteen sarakkeeseen, jolloin yleensä myös luodaan uusi muuttuja, joka ilmaisee missä sarakkeessa tieto oli alun perin. Toisaalta voidaan mennä vastakkaiseen suuntaan eli jakaa yhden sarakkeen tieto useaan sarakkeeseen. Tällaisia aineiston kääntö tilanteita tulee usein vastaan muun muassa silloin, kun samalta tilastoyksiköltä (esimerkiksi eläimeltä) on mitattu sama ominaisuus monena eri ajankohtana.

Useasta sarakkeesta yhteen sarakkeeseen

Usean eri sarakkeen tiedot voidaan yhdistää samaan sarakkeeseen valitsemalla aktiiviseksi muokattava data ja klikkaamalla ”Data” -> ”Stack Columns”. Käännössä käytettävät ja mukaan otettavat muuttujat nimetään ”Task roles” -välilehdellä:



Välilehdellä ”Column naming” puolestaan määritellään käynnön tuloksena syntyvien uusien muuttujien nimet:

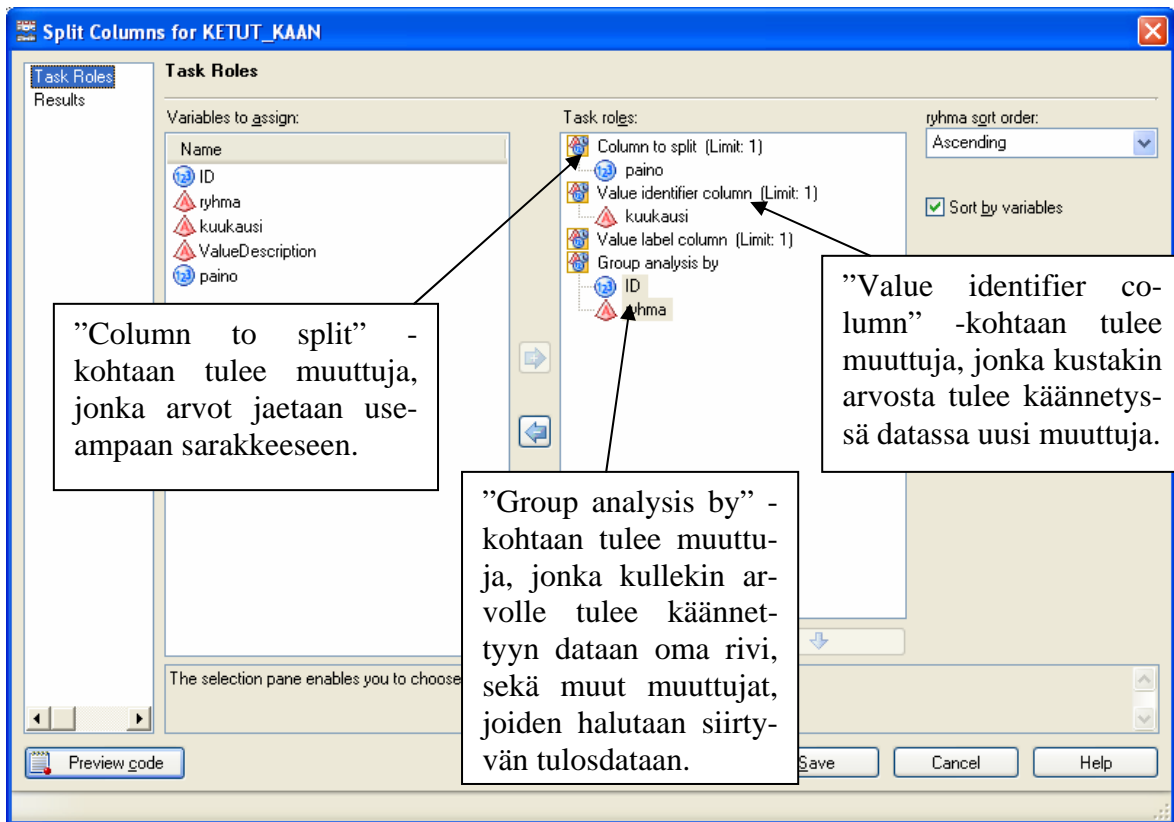


Lisäksi ”Results”-välilehdellä voidaan määrittellä, minkä nimisenä ja mihin kansioon uusi käännetty data tallennetaan.

Tehtävä: Avaa SAS-data nimeltä **KetutData.sas7bdat** ja käännä se siten, että muuttujien elokuu, syyskuu, lokakuu ja joulukuu tiedot siirtyvät samaan sarakkeeseen. Nimeä uudet muuttujat siten, että kaikkien kuukausien painot sisältävä sarake on nimeltään ”paino” ja alkuperäisten kuukausimuuttujien nimet sisältävä sarake on nimeltään ”kuukausi”. Muodosta pysyvä SAS-data nimeltä **ketut_kaan** kansioon C:\SAS\EGKURS.

Yhdestä sarakkeesta useaan sarakkeeseen

Yhdestä sarakkeesta useaan aineisto voidaan kääntää valitsemalla se aktiiviseksi ja klikkaamalla ”Data” -> ”Split Columns”, jolloin aukeaa seuraava välilehti:



”Results”-välilehdellä voi valita tuloksena syntyvän käännetyn datan nimen ja sijainnin. Painamalla lopuksi ”Run” suoritetaan datan kääntö.

Tehtävä: Käännä edellisessä tehtävässä muodostettu **ketut_kaan** takaisin alkuperäiseen muotoonsa siten, että kunkin ketun havainnot ovat yhdellä rivillä ja eri kuukausien painohavainnot ovat omina sarakkeinaan. Tallenna syntyvä SAS-data nimellä **ketut_takaisin**.

7.14 Muuta datojen käsittelyyn liittyvää

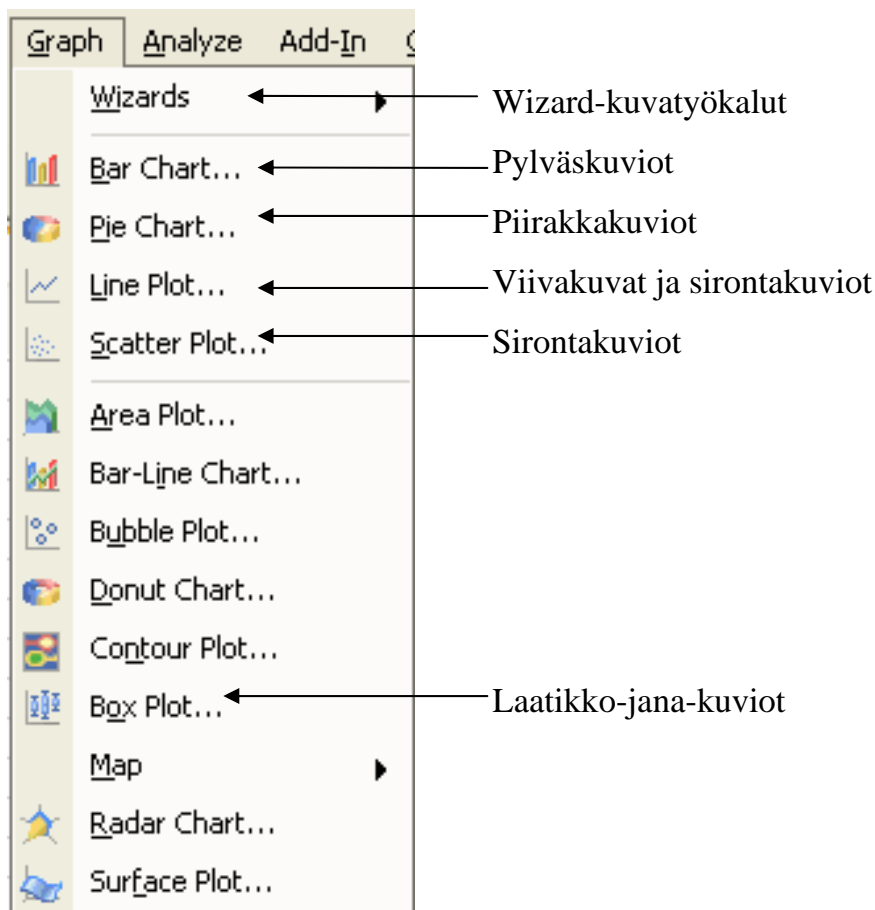
”Data”-valikossa on myös työkalu ”Data Set Attributes”. Tämän työkalun avulla saadaan erilaista tietoa aktiivisesta datasta. Esimerkiksi suuren datan muuttujien ominaisuuksista saa kätevästi yhteenvedon. Kahta dataa voidaan puolestaan vertailla Data-valikon työkalulla ”Compare Data”. Tässä työkalussa valitaan kaksi dataa ja ohjelma vertaa, ovatko datat täsmälleen samat. Työkalulle voi olla käyttöä esimerkiksi silloin, kun projektissa on paljon eri versioita samoista datoista ja käyttäjältä on päässyt unohtumaan, miten datat eroavat toisistaan. Toisaalta työkalun avulla voidaan vertailla myös kahta yksittäistä muuttujaa, jotka ovat peräisin joko samasta datasta tai kahdesta eri datasta. Jos tietokoneelle on kertynyt paljon pysyviä SAS-datoja, niitä pääsee kätevästi poistamaan Data-valikon työkalulla ”Delete Data Sets and Formats”. Työkalulla voi valita kirjastoviitteen osoittaman kansion ja sitten sieltä poistettavaksi datoja sekä formaattitiedostoja. Esimerkiksi SASUSER-kirjastoviitteen osoittamaan kansioon saattaa helposti kertyä turhia datoja. Kannattaa kuitenkin olla varovainen, ettei vahingossa poista tarpeellisia datoja.

8 Datan tarkastelu Graph-valikon kuvatyökaluilla

Yksi SAS Enterprise Guiden vahvuuksista perinteiseen SASiin verrattuna on kuvien piirtämisen helppous. Kuvat ovat käteviä apuvälineitä, kun halutaan hahmottaa muuttujien jakaumien ominaispiirteitä sekä tarkastella muuttujien välisiä yhteyksiä. Kuvia piirrettäessä muuttujien tyyppi on olennainen. Muuttujat voidaan jakaa kategorisiin ja numeerisiin muuttujiin. Kategoristen muuttujien arvot ovat luokkia. Kategoriset muuttujat voidaan edelleen jakaa laatueroasteikollisiin ja järjestysasteikollisiin muuttujiin. Laatueroasteikollisen muuttujan luokkia ei voida laittaa järjestykseen (esim. värit), kun taas järjestysasteikollisilla muuttujilla luokkien järjestyksen määräävä sääntö on olemassa (esim. likertasteikko). Numeerisilla muuttujilla puolestaan kukin tilastoyksikkö saa määrää kuvaavan numeerisen arvon (esim. pituus ja paino). Numeeriset muuttujat voidaan myös edelleen jakaa välimatka- ja suhdeasteikon sekä absoluuttisen asteikon muuttujiin.

Seuraavassa käydään läpi muutamia usein tarvittavia kuvia. Kuvien piirtäminen on Enterprise Guidessa varsin helppoa, kunhan ensin oppii perusasiat. ”Graph”-valikosta löytyviä kuvatyökaluja kannattaa siis käydä myös itsenäisesti läpi. Myöhemmin esiteltävissä datan kuvailu -työkaluissa on myös mahdollista tuottaa varsinaisen tulostuksen lisäksi sivutuotteenä osittain samoja kuvia.

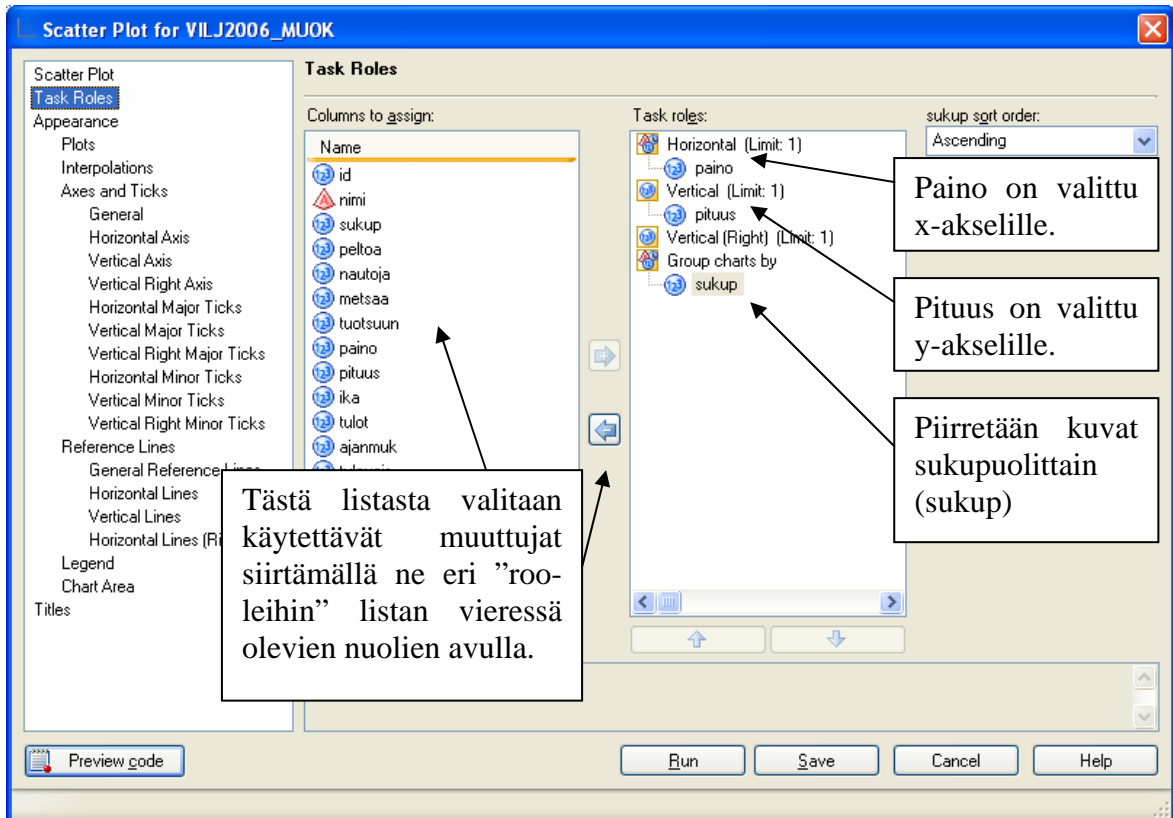
Tässä luvussa käydään läpi seuraavat ”Graph”-valikosta löytyvät työkalut:



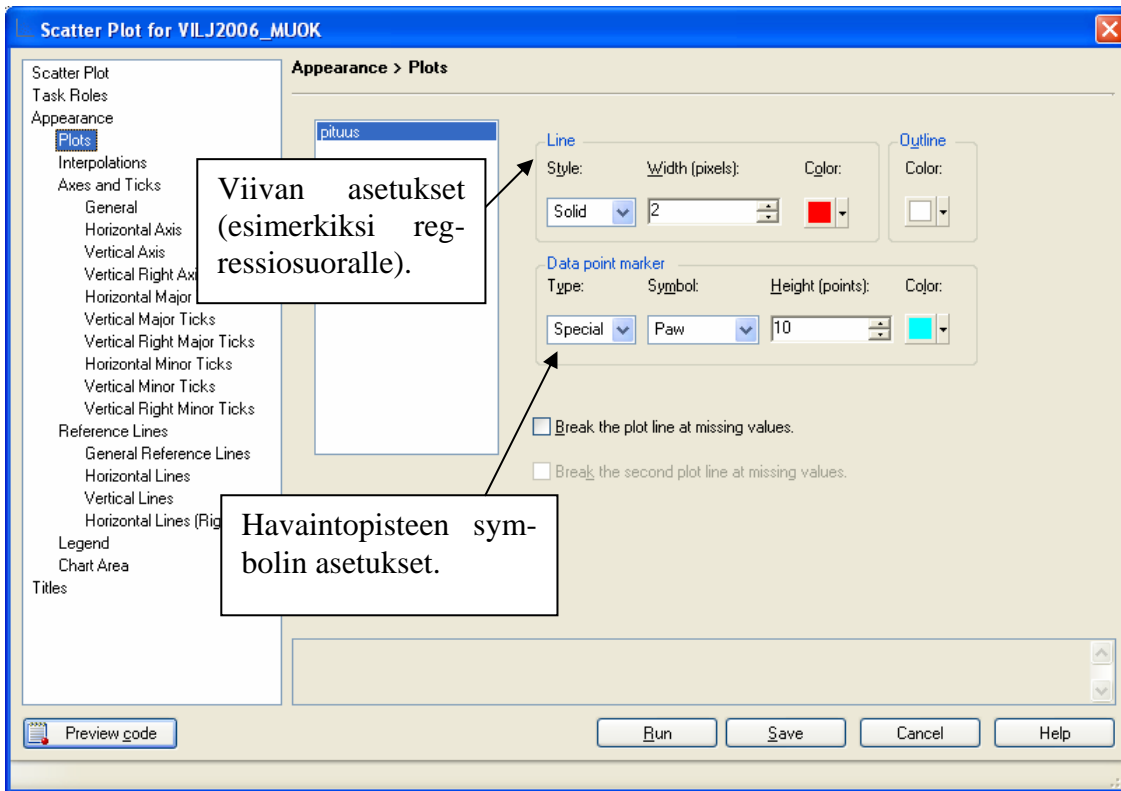
8.1 Sirontakuviio (Scatter Plot)

Sirontakuvioksi kutsutaan kuvaa, jolla tarkastellaan kahden, yleensä numeerisen, muuttujan yhteyttä. Kuvio muodostetaan piirtämällä koordinaatistoon havaintoparit (x_i, y_i) , jossa x_i on vaaka-akselimuuttujan arvo ja y_i pystyakselimuuttujan arvo tilastoyksiköllä i . Sirontakuviio piirtämiseen löytyy työkalu "Graph"-valikosta.

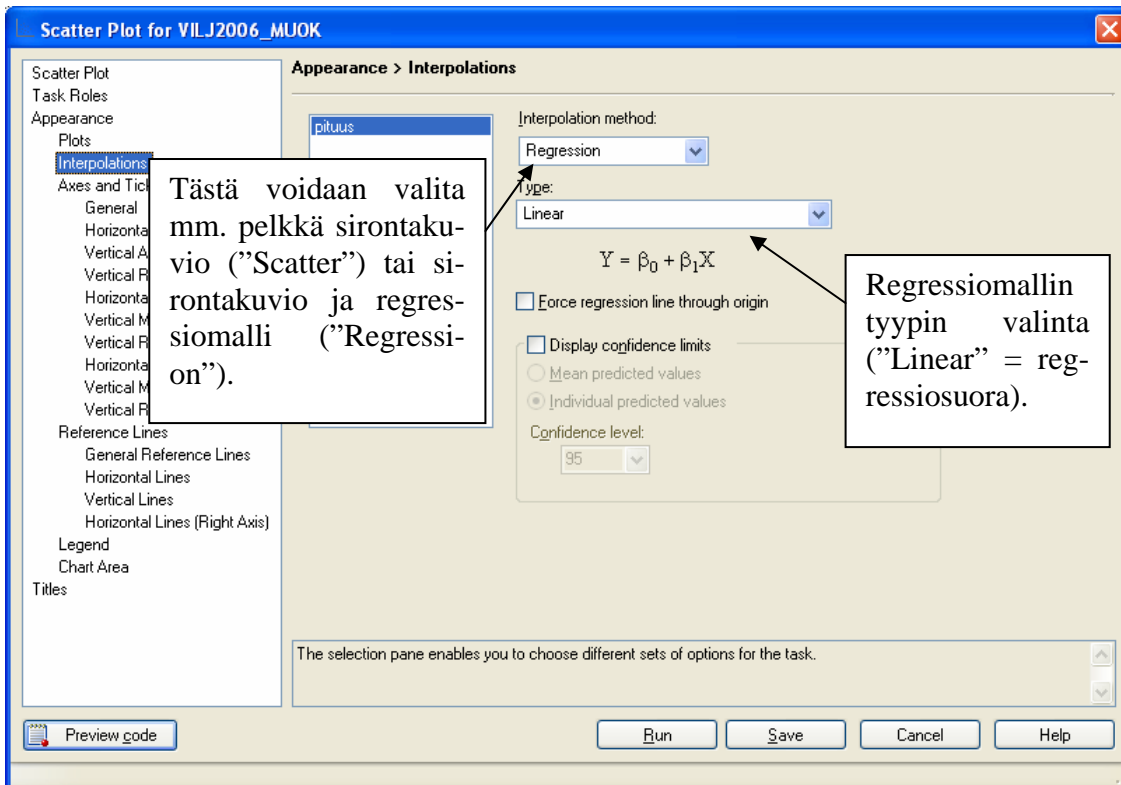
Käytettävä data valitaan ensin aktiiviseksi klikkaamalla se siniseksi prosessidiagrammissa ja sitten valitaan "Graph" -> "Scatter Plot". "Scatter Plot" -välilehdeltä valitaan tavallisen sirontakuviio piirtämiseksi "2D Scatter Plot". "Task Roles" -välilehdellä valitaan muuttujat, joiden arvoista sirontakuviio piirretään:



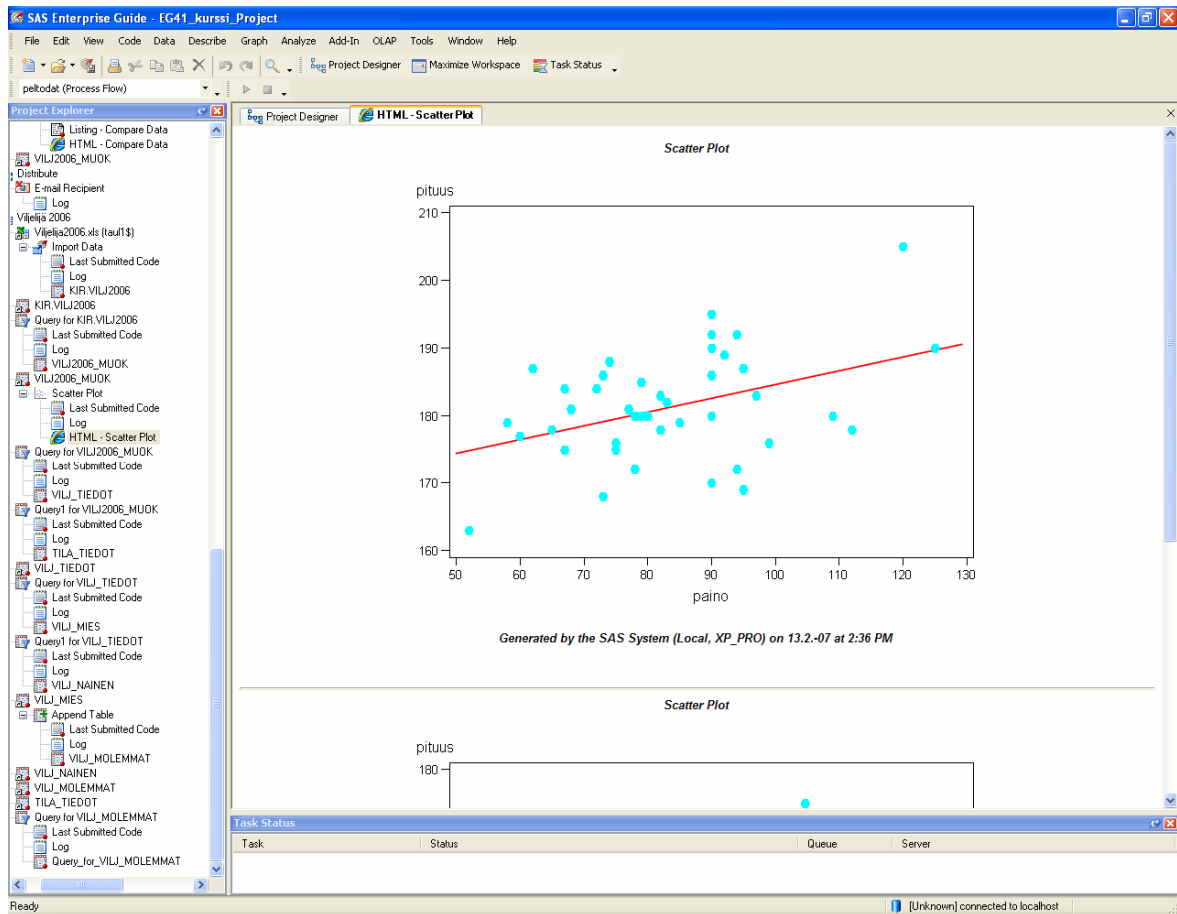
”Plots”-välilehdellä voidaan muokata kuvan ulkoasua:



Kuvan asetuksia voi vaihtaa RTF- ja HTML-tulosteissa myös valmiista kuvasta siten, että painetaan hiiren oikeata painiketta kuvan päällä ja valitaan ”Graph Properties”. Aukeavasta valikosta päästään editoimaan kuvaa. ”Interpolations”-välilehdellä voidaan valita esimerkiksi regressiosuora sovitettavaksi pisteparveen:



Akselien pituudet voi määrittellä oletusarvoista poikkeaviksi ”Horizontal ja Vertical Major Ticks” -välilehdillä valitsemalla oletuksena olevan ”Automatic” vaihtoehdon sijasta ”Specify”. Lisäksi eri välilehdillä voidaan tehdä muun muassa akseleihin, referenssiiviivoihin sekä kuvan otsikoihin ja kokoon liittyviä määrittelyjä. Lopuksi valitaan ”Run”, jolloin sirontakuvio tulostuu oletustulostusmuodossa:



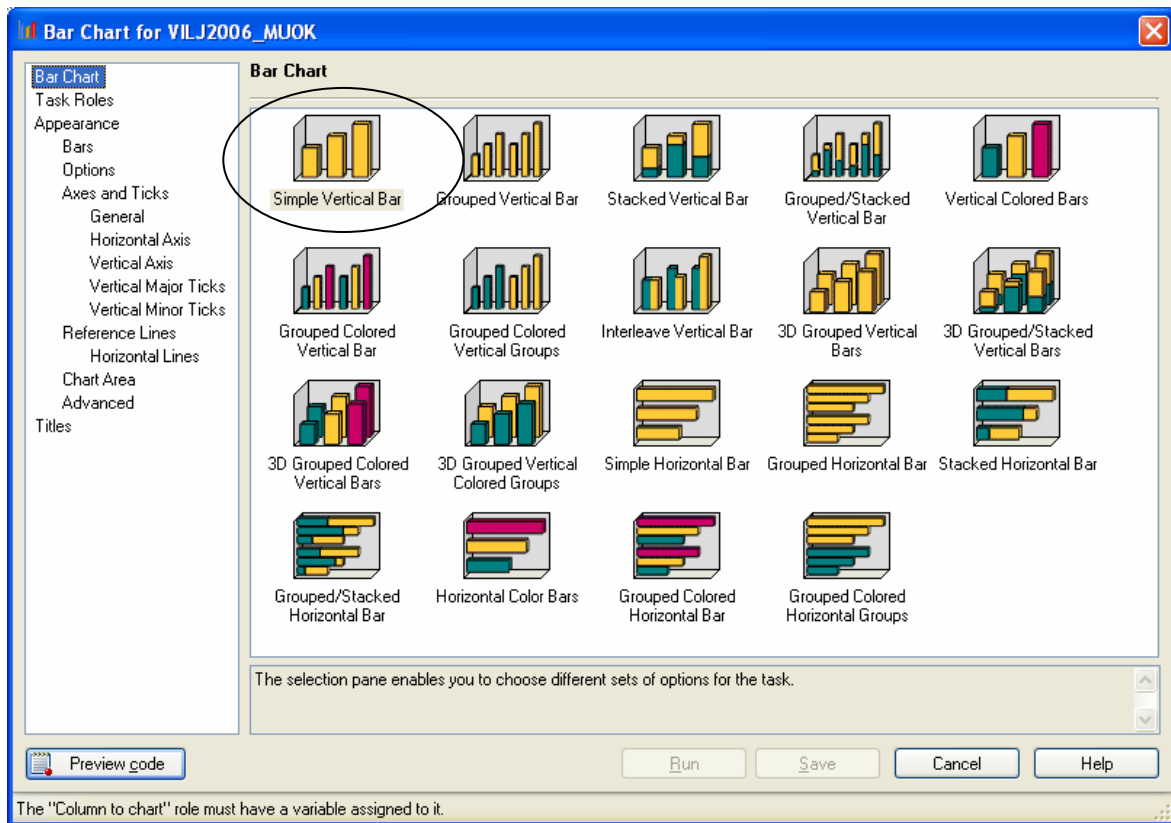
Jos halutaan piirtää sirontakuvio, jossa pisteet on merkitty eri värillä tai symbolilla jonkin ryhmittelevän muuttujan luokkien mukaisesti, se voidaan tehdä ”Line Plot” -työkalulla. Valitaan ”Graph” -> ”Line Plot” ja avautuvalla ”Line Plot”-välilehdellä ”Multiple line plots by group column”. ”Task Roles” -välilehdellä ryhmän määräävä muuttuja laitetaan kohtaan ”Group” ja ”Interpolations”-välilehdellä kohdassa ”Interpolation method” valitaan joko ”Scatter” tai ”Regression” (kummallekin ryhmälle erikseen).

Tehtävä: Käytä projektissa olevaa muokattua Viljelijä 2006 -dataa nimeltä **vilj2006_muok**. Piirrä sirontakuvio, jossa on x-akselilla viljelijän paino ja y-akselilla viljelijän pituus. Piirrä kuvat sukupuolittain ja lisää vielä painon ja pituuden yhteyttä havainnollistavat regressiosuorat sekä naisten että miesten kuviin. Muokkaa kuvista mieleisesi jälkikäsittelemällä valmiita kuvia.

8.2 Pylväskuvio (Bar Chart)

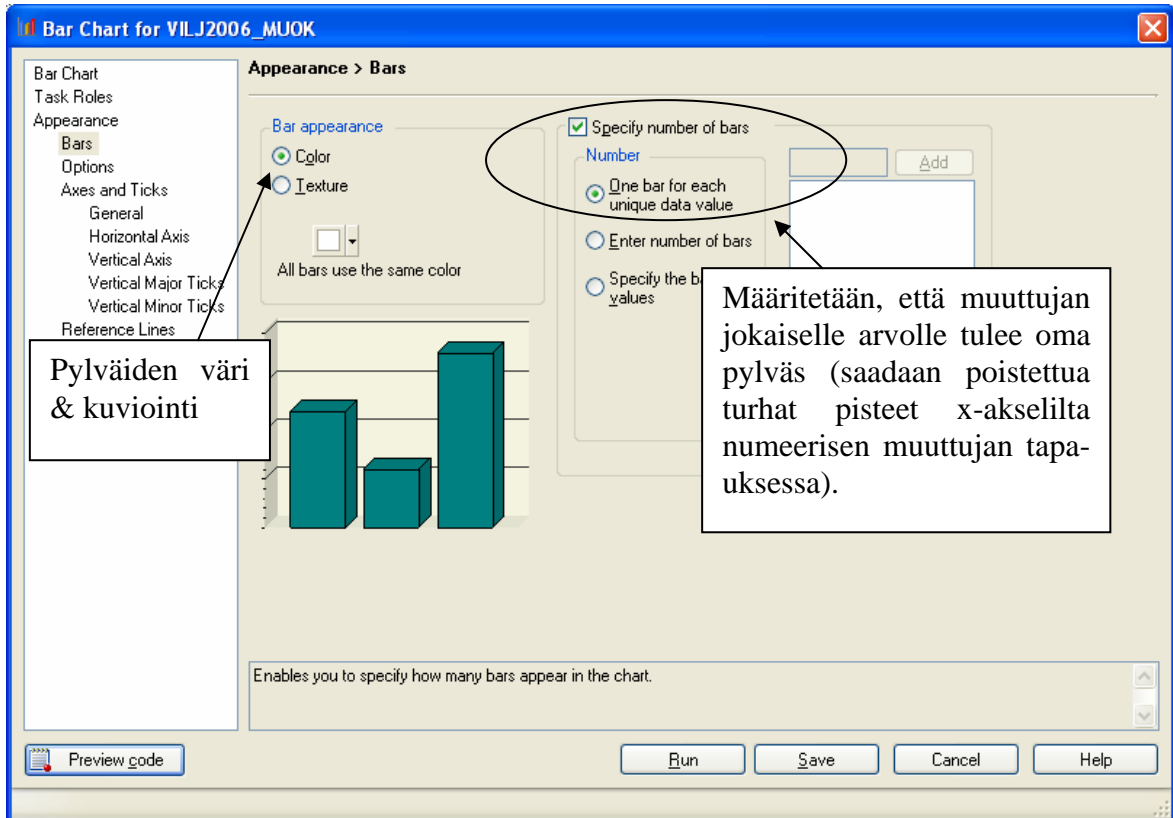
Pylväskuvio on yleinen kategorisen muuttujan frekvenssijakauman havainnollistamistapa. Muuttujan kullekin luokalle (esimerkiksi sukupuolimuuttujalla miehille ja naisille) piirretään oma pylväs, jonka pituus kuvaa luokan esiintymisten määrää aineistossa tai prosenttiosuutta koko aineistosta. Toisaalta voidaan piirtää pylväitä, joiden pituudet edustavat esimerkiksi jonkin muuttujan keskiarvoa toisen muuttujan eri luokissa. Pylväskuvio voidaan piirtää valitsemalla käytettävä data aktiiviseksi eli klikkaamalla datakuvake siniseksi

prosessidiagrammissa ja valitsemalla ”Graph” -> ”Bar Chart”. Tällöin aukeaa päällimmäiseksi ”Bar Chart” -ruutu, josta voidaan valita kuvan tyyppi:



Samoin kuin sirontakuvion yhteydessä muuttujille annetaan ”roolit” välilehdellä ”Task Roles”:

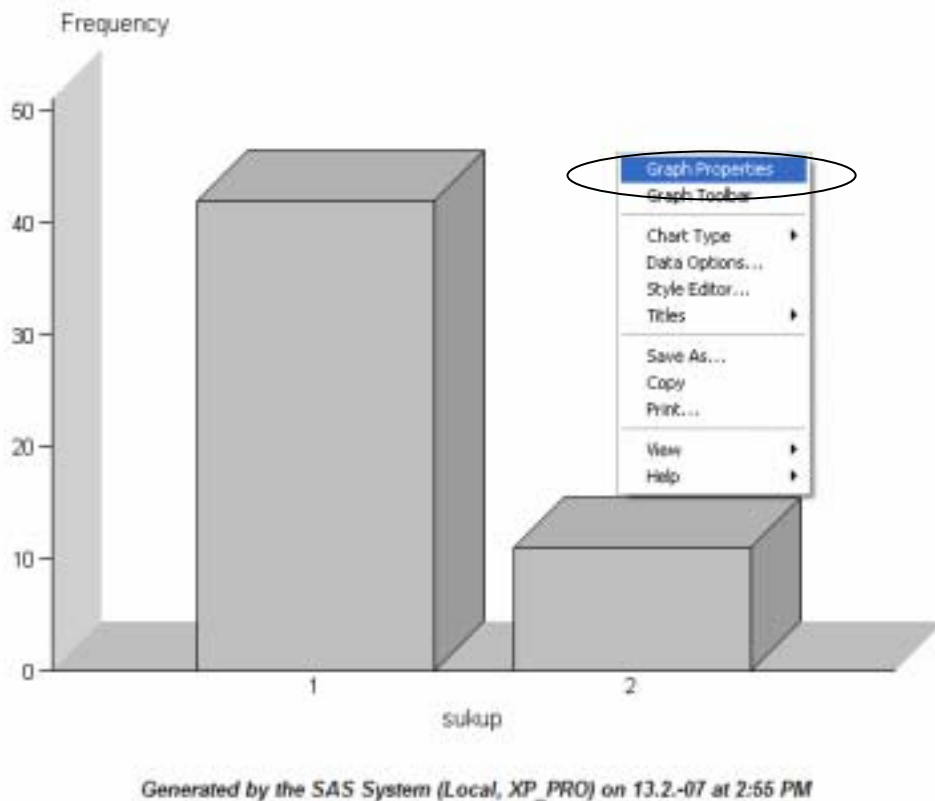
”Bars”-välilehdellä voidaan tehdä seuraavia määrittäviä:



”Options”-välilehden kohdassa ”2D” voidaan määrittää kuva kaksiulotteiseksi oletuksena olevan kolmiulotteisuuden sijaan. ”Advanced”-välilehdellä puolestaan voidaan valita pylvään pituuden laskentatapa (”Statistic used to calculate bar”). Jos ”Sum of” -muuttujaa ei ole määritetty, pylväät voivat edustaa muun muassa kunkin luokan esiintymisten määrää tai prosenttiosuutta aineistossa. Jos taas käytetään ”Sum of” -muuttujaa, pylväät voidaan määrittellä joko keskiarvopylväiksi tai summapylväiksi. ”Display error bars” -kohdassa puolestaan voidaan lisätä prosentti- tai keskiarvopylväille luottamusvälipalkit. Luottamusvälien määrittäminen riippuu tutkimusasetelmasta. Tästä saadaan keskiarvojen luottamusvälit asetelmalle, jossa kaikki havainnot ovat toisistaan riippumattomia ja peräisin normaalijakaumasta. Prosenttiosuuksien luottamusvälejä puolestaan voi käyttää suurten otosten tilanteessa, jossa kaikkien havaintojen riippumattomuusoletus on voimassa. Kannattaa siis varmistaa, että aineisto perustuu siihen tutkimusasetelmaan, johon EG:n esittämät luottamusvälit soveltuvat käytettäviksi.

Muilta välilehdiltä löytyy erilaisia kuvan ulkoasuun vaikuttavia asetuksia (muun muassa otsikoihin, akseleihin, referenssiiviivoihin ja kuvan kokoon liittyviä määrittäviä). Esimerkiksi prosenttiluvut pylväiden päälle saadaan ”Appearance” -> ”Advanced”-välilehdellä tekemällä valinta ”Additional statistical value to show next to bar”.

Valmista kuvaa pääsee tässäkin tapauksessa muokkaamaan RTF- ja HTML-tulostusten yhteydessä klikkaamalla kuvaa hiiren oikealla painikkeella ja valitsemalla ”Graph properties”:



Kannattaa huomioida, että muokattaessa yllämainitulla tavalla valmista kuvaa, muutokset eivät tallennu projektin mukana. Muokattu kuva kannattaa siis joko tallentaa omaksi kuvatiedostokseen (valitsemalla yllä olevassa kuvassa näkyvästä valikosta ”Save As”) tai kopioida leikepöydän kautta talteen esimerkiksi Wordiin (valitsemalla yllä olevassa kuvassa näkyvästä valikosta ”Copy”).

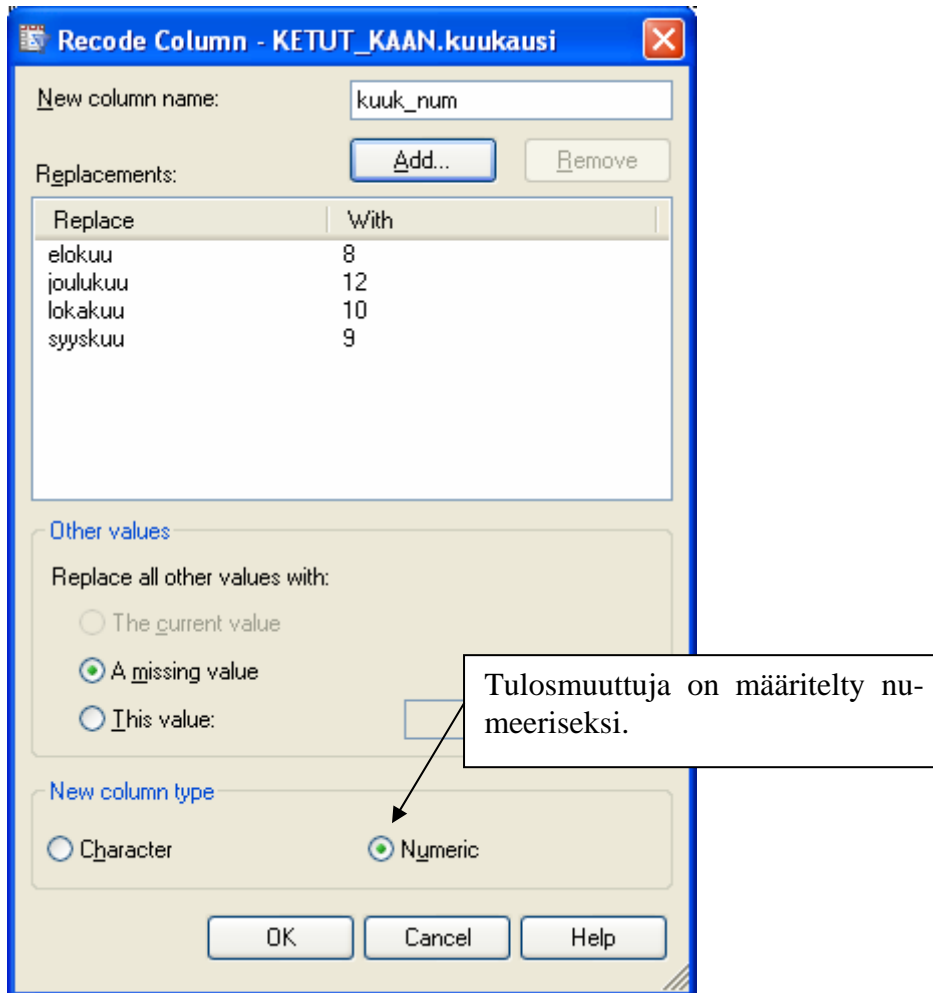
Tehtävä: Piirrä kummallekin sukupuolelle pylväskuvio Viljelijä 2006 -aineistosta (**vilj2006_muok**) siten, että pylväiden pituudet kuvaavat miesten ja naisten keskipituutta aineistossa.

8.3 Viivakuva (Line Plot)

Viivakuva on koordinaatistoon piirrettävä kuva, jossa osa havainnoista yhdistetään toisiinsa viivalla. Pienillä kokeellisilla aineistoilla viivakuva on usein kätevä esitystapa muuttujien arvojen tarkasteluun kaikilla koeyksiköillä. Kuva voidaan piirtää esimerkiksi siten, että x-akselilla ovat toistotekijän (esimerkiksi vuoden) arvot ja saman tilastoyksikön havainnot yli toistotekijän tasojen yhdistetään viivalla. Tilastoyksiköt tai niiden muodostamat ryhmät voidaan erottaa toisistaan erilaisella viivatyyppillä, symbolilla tai piirtämällä ryhmille omat kuvat.

Jotta viivat kulkisivat pisteiden läpi x-akselin arvojen mukaisessa järjestyksessä vasemmalta oikealle, aineisto täytyy järjestää ennen kuvan piirtämistä x-akselin muuttujan mukaan suuruusjärjestykseen, ellei se ole valmiiksi järjestyksessä. X-akselin arvojen tulee myös olla sellaisia, että ohjelma ymmärtää niiden oikean järjestyksen (jonkin säännön mu-

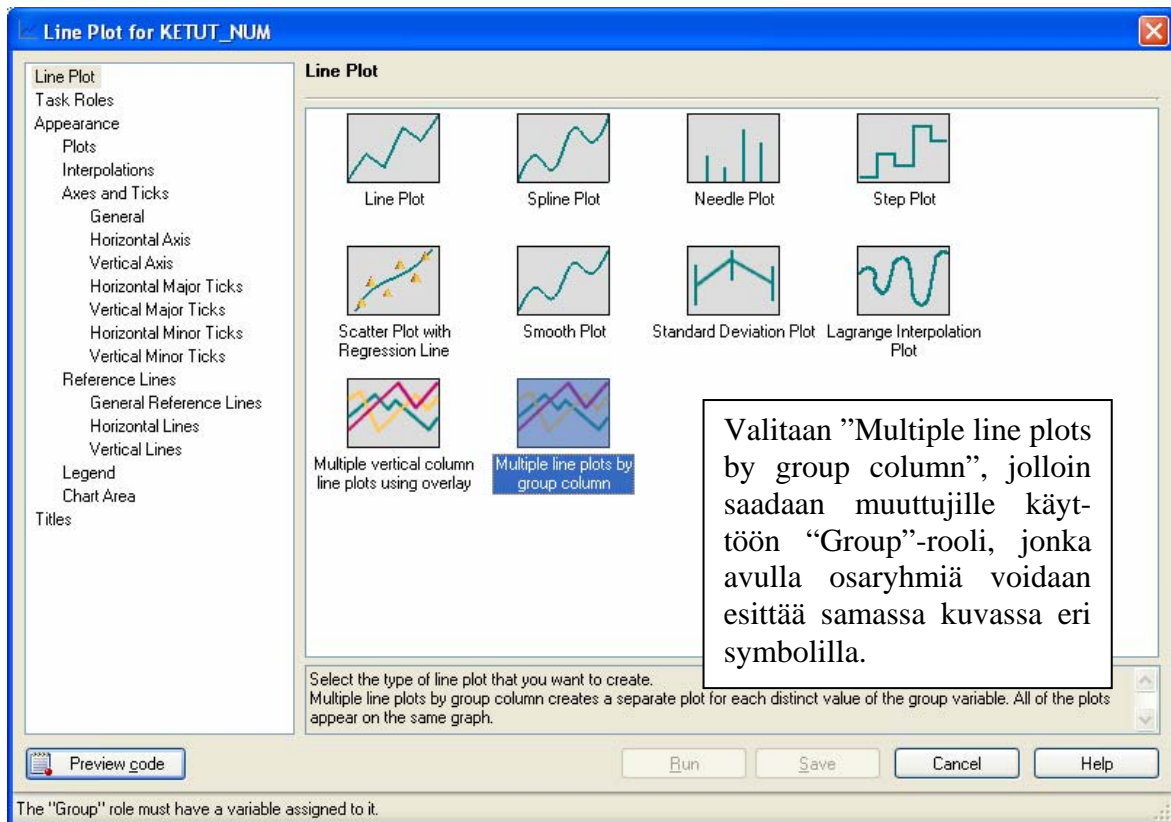
kaisesti), muuten ohjelma sijoittaa arvot akselille epäloogiseen järjestykseen. Merkkimuotoiset arvot ovat usein ongelmallisia, koska ohjelma ei osaa järjestää niitä muuten kuin aakkosjärjestykseen. Tästä syystä arvot on usein syytä koodata uudelleen numeerisiksi, kun niitä käytetään viivakuvan x-akselilla. Koodaaminen voidaan tehdä samalla tavalla Query Builderissa kuin muuttujien luokittelu ja luokkien yhdistäminen. Query Builderissa valitaan siis "Computed Columns" -> "New" -> "Recode a Column", jossa voidaan tehdä seuraavan ikkunan mukainen määrittely:



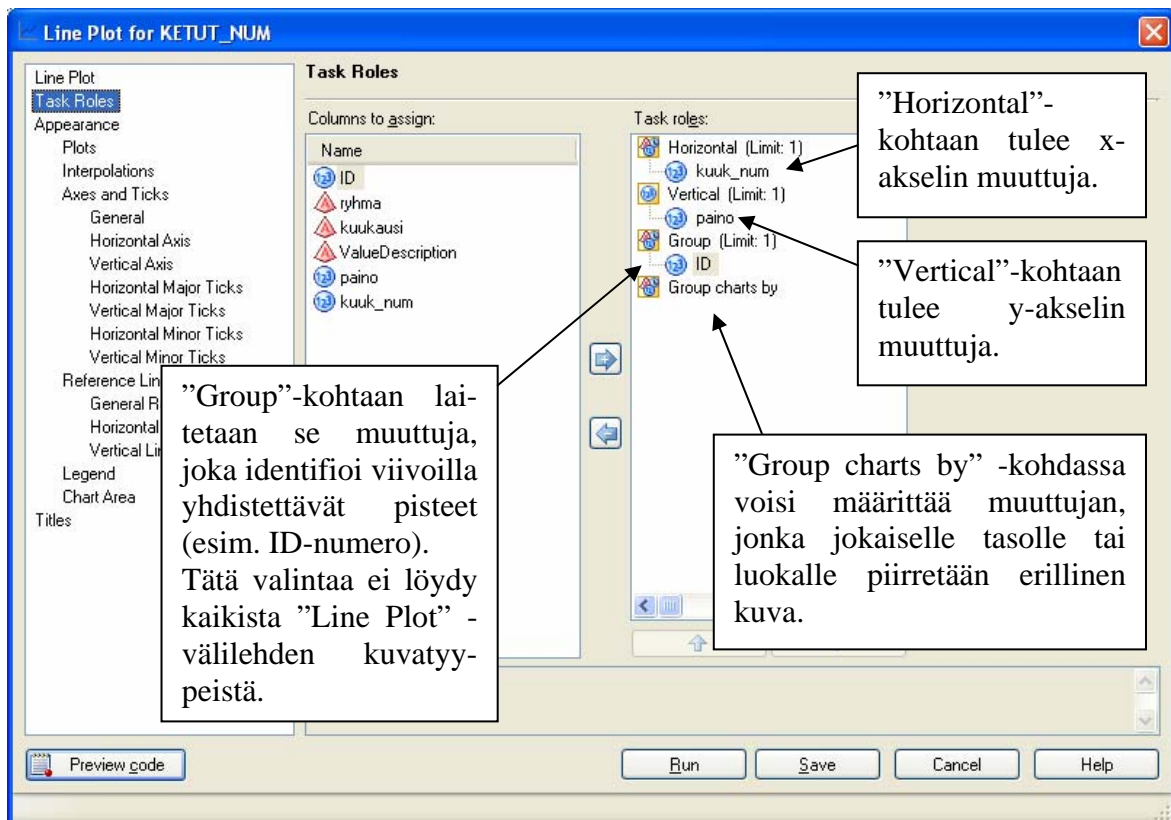
Query Builderin tuloksena saatavan datan kuuk_num-muuttujaa voidaan nyt käyttää viivakuvan x-akselin muuttujana.

Viivakuva voidaan piirtää valitsemalla käytettävä data aktiiviseksi ja klikkaamalla "Graph" -> "Line Plot". Käytettävä kuvatyyppejä valitaan "Line Plot" -välilehdellä. Nämä kuvatyypit antavat pohjan kuvan muokkaukselle, mutta erilaisten asetusvalintojen avulla kahden eri pohjan avulla saatetaan päätyä samaan lopputulokseen.

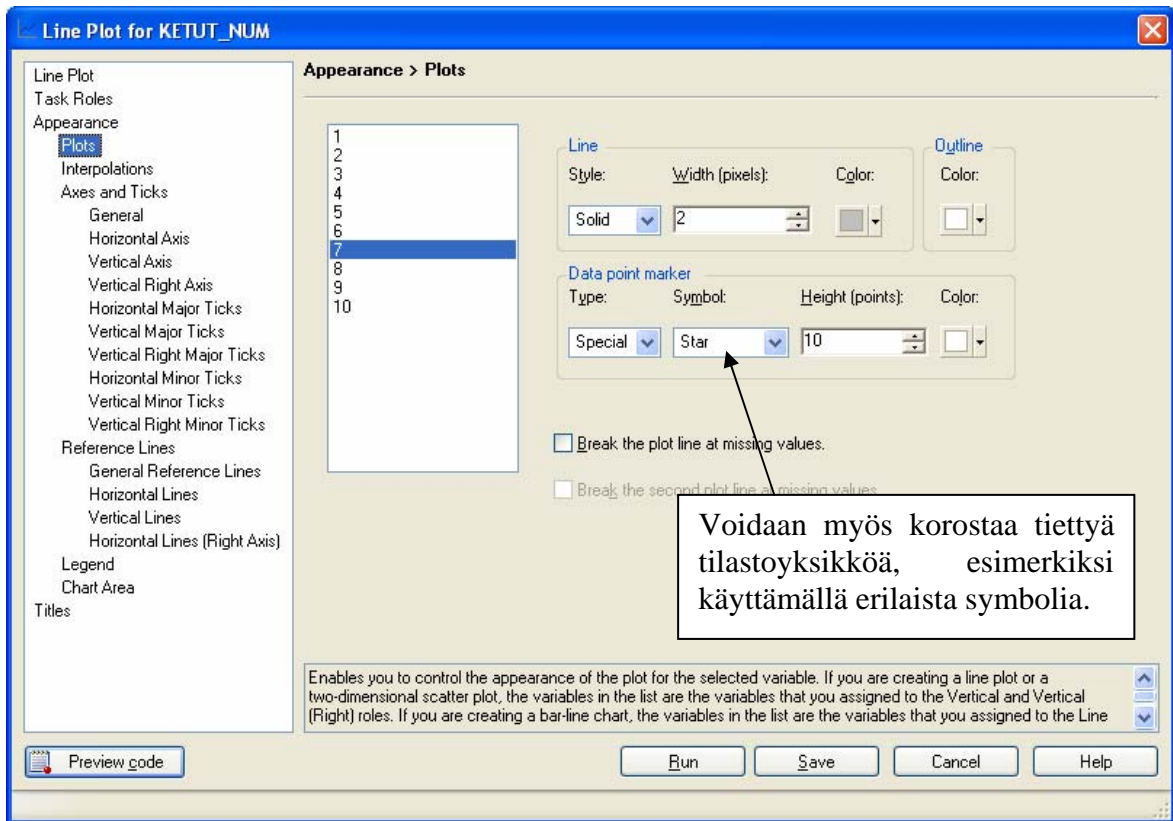
Esimerkiksi seuraavassa valittavalla kuvatyypillä voidaan tehdä myös ”Standard Deviation Plot” -tyyppinen kuva:



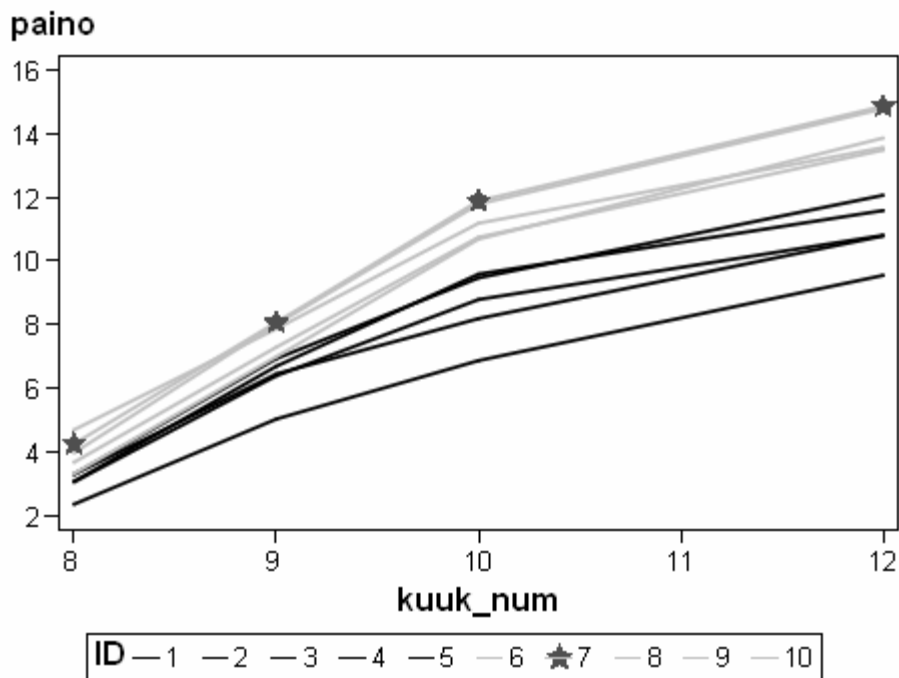
”Task Roles” -välilehdellä määritellään muuttujien tehtävät:



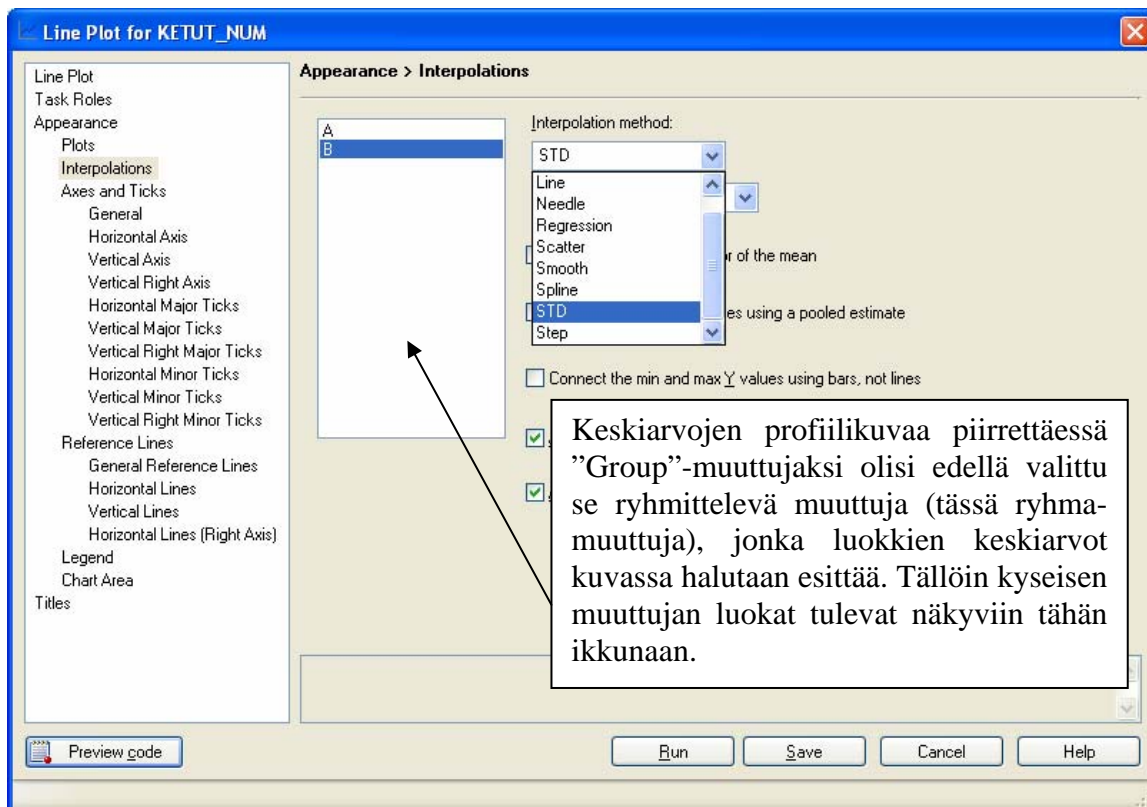
”Plots”-välilehdellä voi määrittellä viivatyypin, viivan paksuuden ja värin erikseen kullekin ”Group”-muuttujan tasolle. Siinä voi siis esimerkiksi määrätä samaa käsittelyä saaneiden tilastoyksiköiden profiilit piirrettäviksi samanlaisella viivalla:



Muut valinnat ovat samankaltaisia kuin muidenkin kuvien yhteydessä, eli niillä voidaan muun muassa vaihtaa tai muokata kuvien otsikoita, referenssi viivoja, akseleiden asetuksia sekä viivojen selitteiden ja kuvan koon asetuksia. Tuloksena syntyy näillä asetuksilla seuraava kuva:



”Interpolations”-välilehdellä olisi voitu yksittäisiä havaintoja esittävän viivakuvasijaan vaihtaa piirrettäväksi esimerkiksi keskiarvojen profiilikuvan (eli ”Standard Deviation Plot”-tyyppisen kuvan) valitsemalla ”STD”:



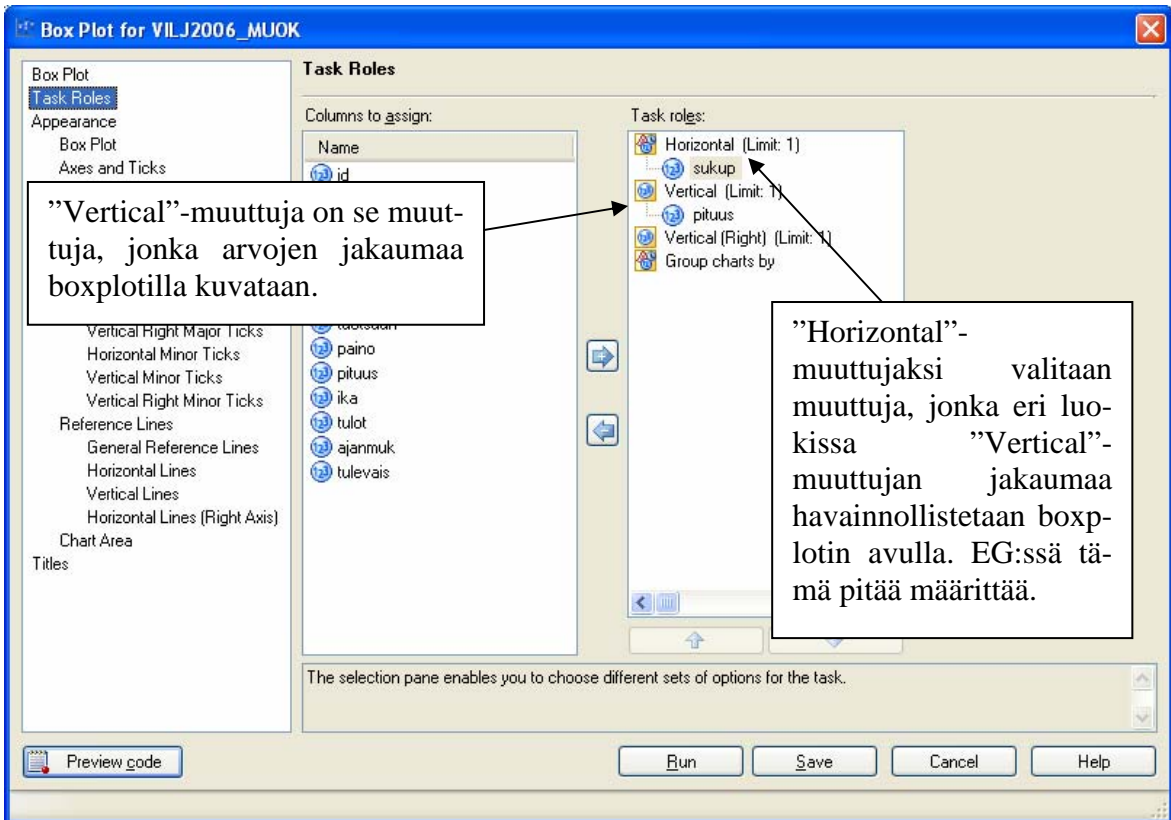
Tehtävä: Ota projektista käyttöön käännetty kettudata **ketut_kaan**. Tee ensin Query Builderissa uusi muuttuja, joka sisältää kuukausitiedon numeerisena. Tallenna data nimellä **ketut_num**. Piirrä kuva, jossa x-akselilla on kuukausi ja y-akselilla ketun paino. Yhdistä kunkin ketun toistomittaukset yli kuukausien viivalla ja erota eri käsittelyä saaneet ketut toisistaan erilaisella viivalla.

Tehtävä: Käytä samaa dataa kuin edellisessä tehtävässä ja piirrä kuva, jossa on viivalla yhdistetty samaa käsittelyä saaneiden kettujen keskimääräiset painot eri ajankohtina (yksi viiva kummallekin käsittelylle). Piirrä kuvaan myös keskihajonnan pituiset janat kuhunkin aikapisteeseen kummallekin käsittelylle.

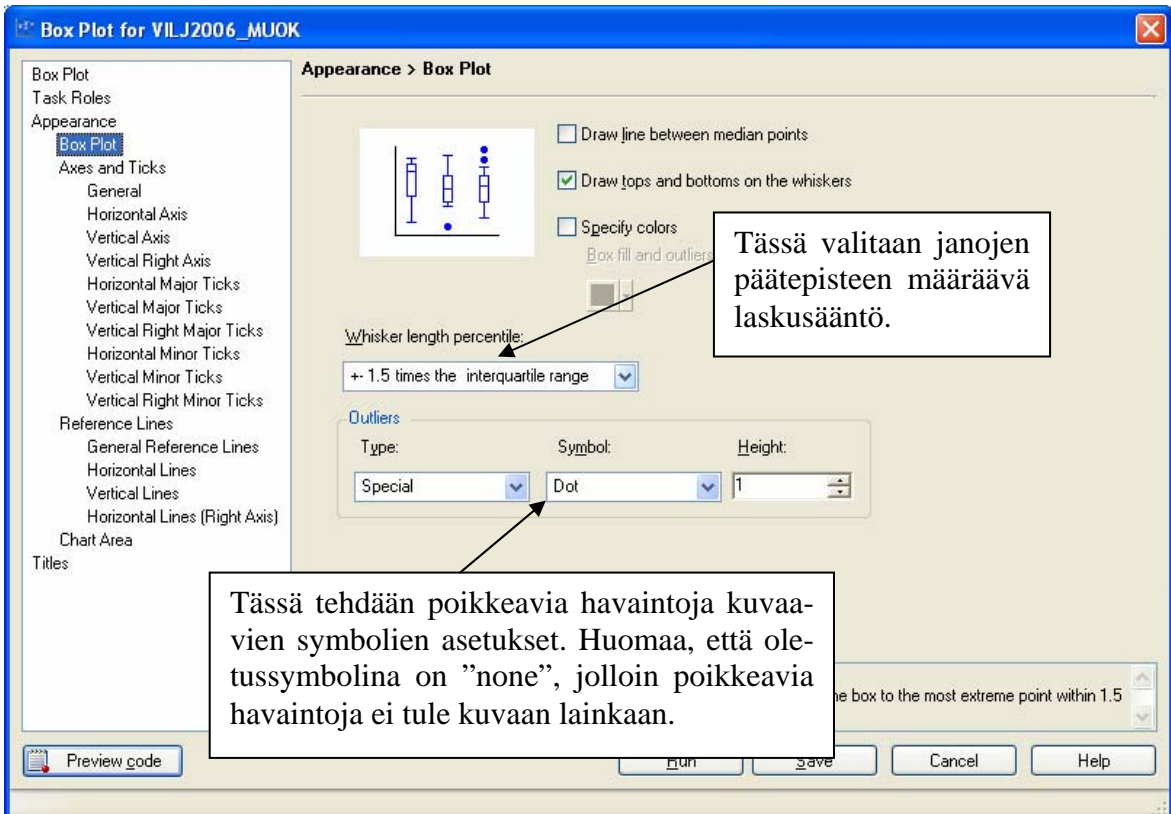
8.4 Laatikko-janakuvio (Boxplot)

Laatikko-janakuvio soveltuu numeerisen muuttujan jakauman havainnollistamiseen. Kuvioista saa nopealla silmäyksellä käsityksen monista jakauman piirteistä.

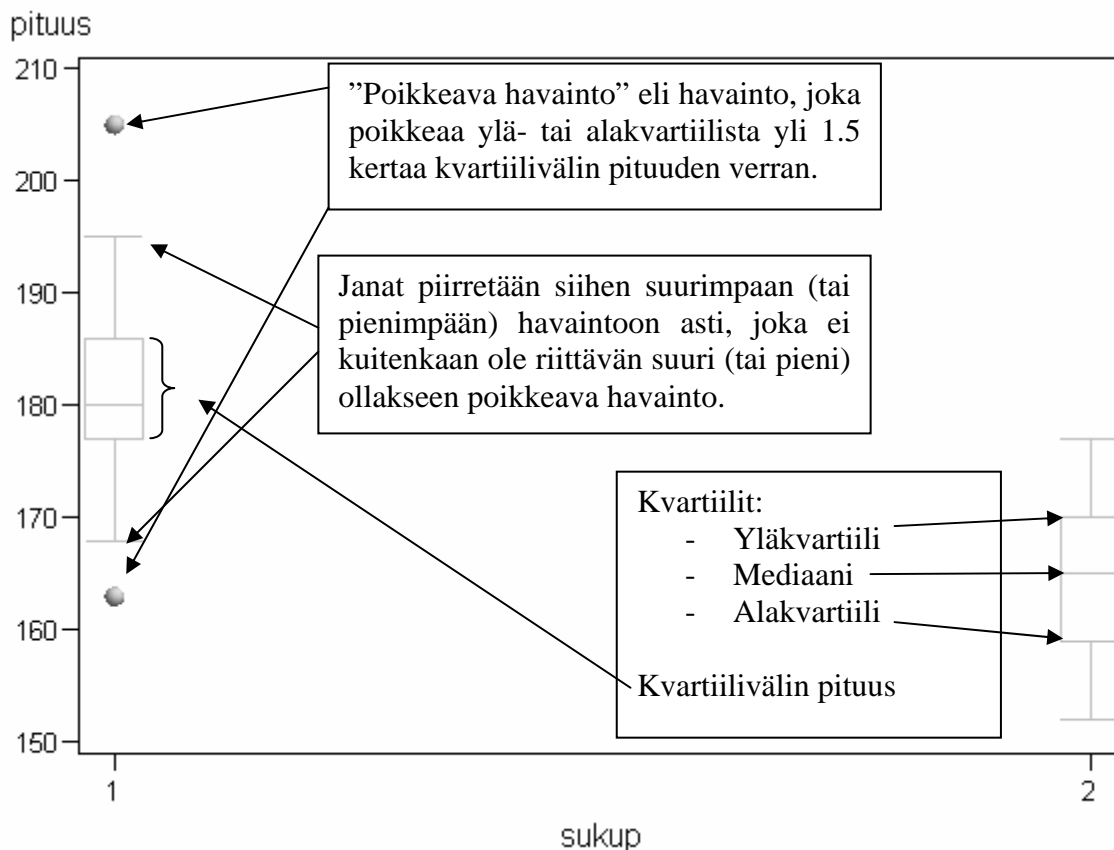
Enterprise Guidessa piirretään boxplot klikkaamalla ”Graph” -> ”Boxplot” ja valitsemalla ”Box Plot” -välilehdeltä ”Box Plot” ja sitten ”Task Roles”:



Boxplotista on erilaisia versioita. Käyttäjä voi ”Appearance” -> ”Box Plot” -välilehdellä valita, millaisen version haluaa tulostaa:



Muilla välilehdillä tehtävät valinnat koskevat muun muassa kuvan akseleiden, referenssi-
viivojen, otsikoiden ja koon asetuksia. Tulostus näyttää valituilla asetuksilla seuraavalta:

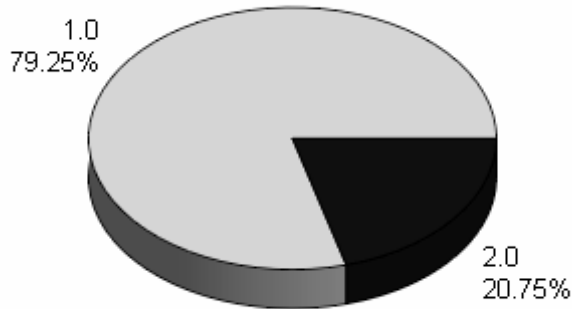


EG:ssä myös jotkut muut työkalut tuottavat boxplot-kuvia. Kuten tässäkin nähtiin, boxplot-kuvien piirtämisessä voidaan käyttää erilaisia sääntöjä. Kannattaa siis aina varmistaa, millä periaatteilla käyttämäsi tai tulkitsemäsi boxplot-kuva on piirretty.

Tehtävä: Piirrä päätuotantosuunnittain tulojen jakaumia havainnollistavat boxplot-kuvat Viljelijä 2006 -aineistosta (**vilj2006_muok**). Voit myös tarkastella muiden numeeristen muuttujien jakaumia.

8.5 Piirakkakuvi (Pie Chart)

Myös piirakkakuvion avulla voidaan havainnollistaa kategorisen muuttujan jakaumaa. Niinpä piirakkakuvion tekeminen toimii hyvin pitkälle samoilla periaatteilla kuin pylväs-kuvion tekeminen. Ensin valitaan ”Graph” -> ”Pie Chart” ja sitten valitaan ”Task Roles” -välilehdellä muuttujien roolit: esimerkiksi ”Column to chart” voi olla sukupuoli ja ”Advanced”-välilehdellä voidaan valita esitetäänkö kuvassa frekvenssejä vai prosenttiosuuksia. Tulostus näyttää kolmiulotteisena seuraavalta:



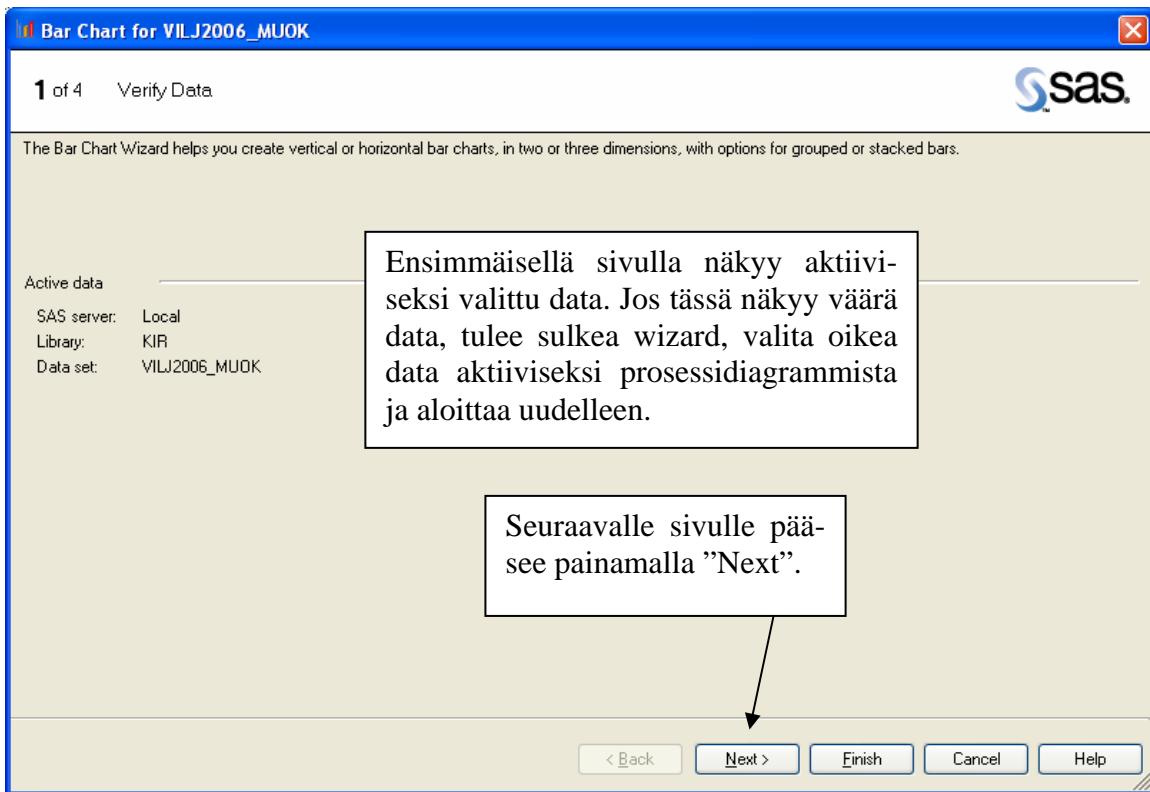
8.6 Wizard-kuvatyoikalut

Wizardien tarkoituksena on tarjota vaihtoehtoinen tapa tiettyjen työkalujen käyttämiseksi. Tässä esiteltävissä wizardeissa on yhteistä se, että niillä pystyy tekemään tärkeimmät tavallisilla työkaluillakin tehtävät määrittelyt, mutta osa asioista jää wizardien ulkopuolelle. Niinpä tavalliset kuvatyoikalut ovat usein käyttökelpoisempia, kun halutaan käyttää kuvatyoikalun ominaisuuksia mahdollisimman monipuolisesti. Wizard-kuvatyoikalut voivat puolestaan olla käyttökelpoisia esimerkiksi sellaisissa tilanteissa, joissa kuvia käytetään aineiston alustavaan tarkasteluun ja käyttäjä kokee wizard-kuvatyoikalujen logiikan tavallisten kuvatyoikalujen logiikkaa selkeämpänä. Lisäksi wizardista on mahdollista siirtyä tavalliseen vastaavaan työkaluun. Tämä onnistuu klikkaamalla määriteltä wizardeista tehtäväsolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Open In Advanced View”. Tällöin wizardissa tehdyt määrittelyt näytetään tavallisessa työkalussa ja käyttäjä voi jatkaa määrittelyä siellä.

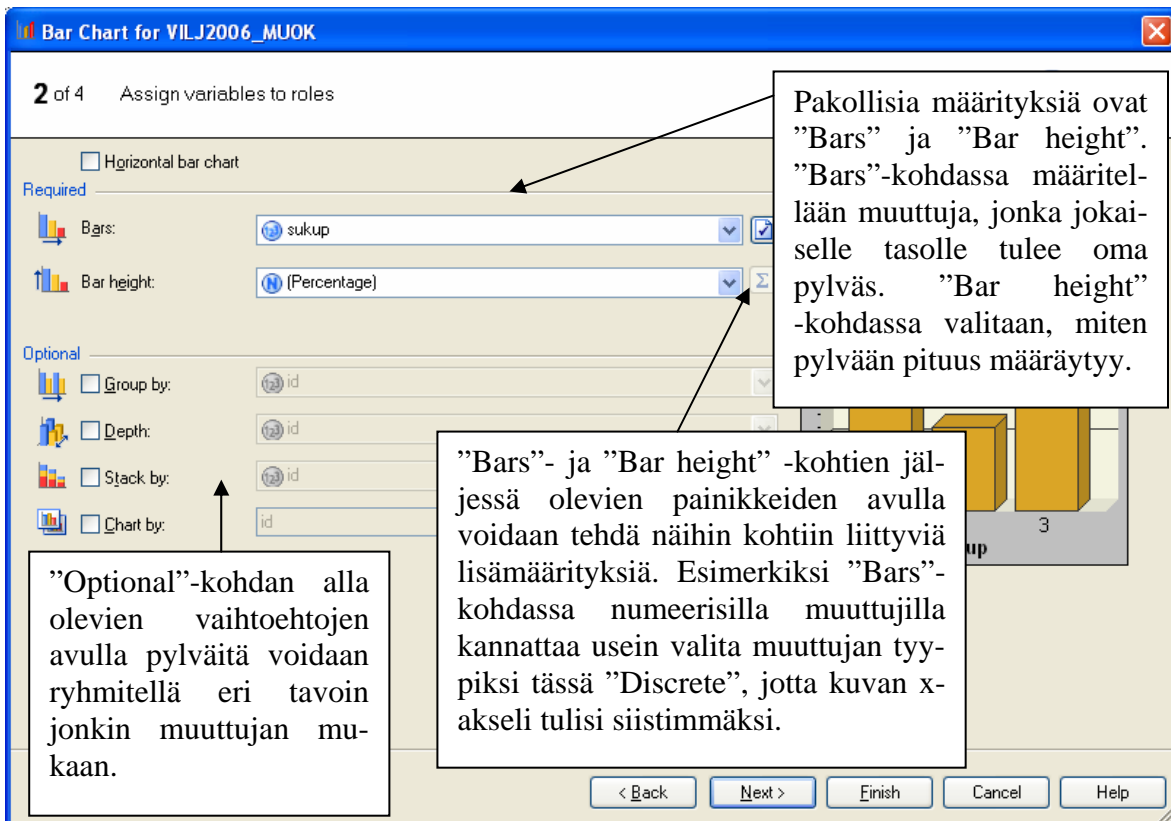
Toimintatapa wizardeissa on sellainen, että ensin valitaan prosessidiagrammista käytettävä data aktiiviseksi ja sitten käynnistetään wizard-työkalu. Wizardissa määrittelyt on jaettu useammalle sivulle, joiden läpi edetään järjestyksessä, ja lopuksi valitaan ”Finish”. Seuraavassa esitellään lyhyesti EG:stä löytyvät wizard-kuvatyoikalut.

Wizard-työkalu pylväskuville

Pylväskuvien tekoon suunniteltu Wizard saadaan avattua klikkaamalla ”Graph” -> ”Wizards” -> ”Bar Chart”. Tällöin aukeaa wizardin ensimmäinen sivu:



Toisella sivulla määritellään muuttujien roolit:



Kolmannella sivulla määritellään ulkoasuun liittyviä asioita:

3 of 4 Specify appearance

Kolmiulotteisuus

3D chart

Colors

Color bars by: [All bars the same]

Bar color: **Pylväiden värit**

Labels

Legend: South

Data labels: Percentage

Axis Labels

Lines and ticks

Use reference lines

Tick marks

Selitykset pylväiden päihin (esimerkiksi prosentit)

Sample chart: (Percentage)

sukupu	Percentage
1	33.33%
2	16.67%
3	50.00%

< Back Next > Finish Cancel Help

Viimeisellä sivulla voi vielä määrittellä kuvien otsikoita ja alaviitteitä:

4 of 4 Provide a title and footnote

Graph: Bar Chart

Footnote: Generated by the SAS System (&_SASSERVERNAME, &SYSSCP) on %SYSFUNC(DATE(), EURDFDE9.) at %SYSFUNC(TIME(), TIMEAMP8.)

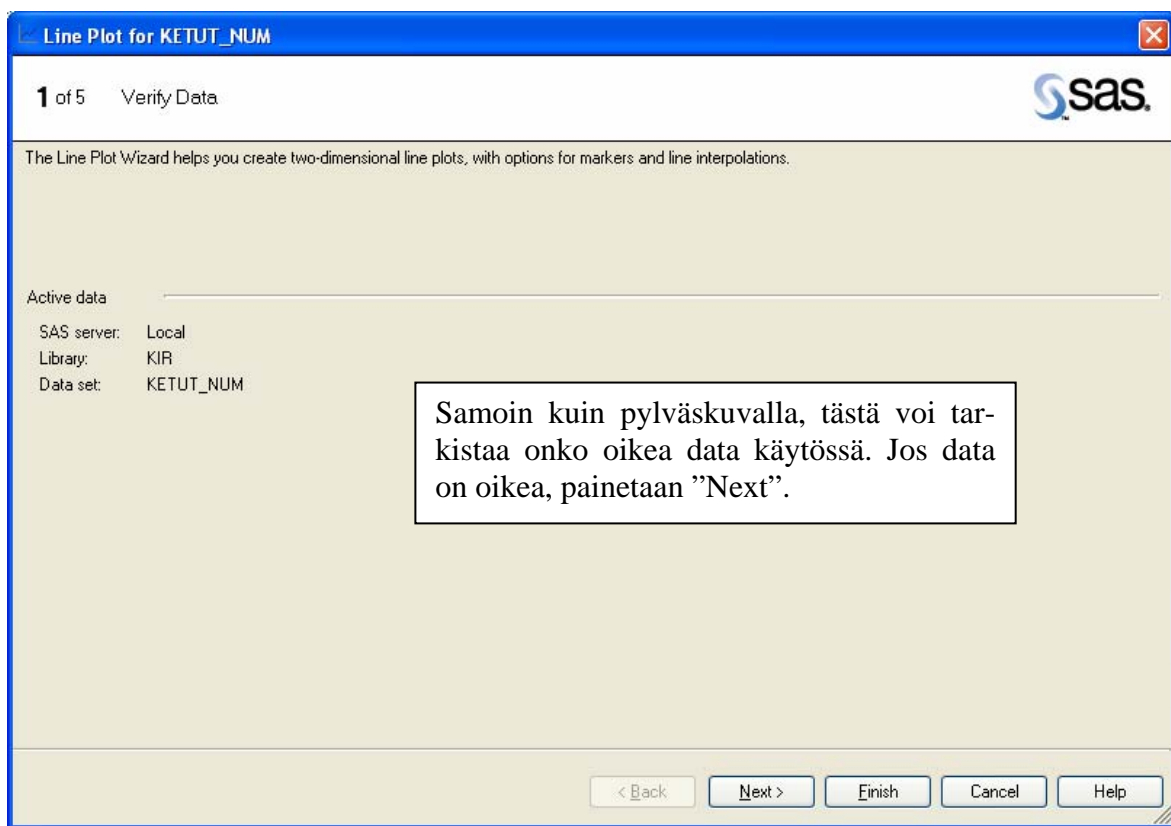
< Back Next > Finish Cancel Help

Lopuksi painetaan "Finish", jolloin valmis kuva tulee samalla tavalla osaksi projektia kuin tavallisilla kuvatyökaluilla.

Tehtävä: Piirrä wizard-työkalua käyttäen pylväskuvio, jonka avulla havainnollistetaan sukupuolimuuttujan frekvensijakaumaa.

Wizard-työkalu viivakuville

Viivakuvissa wizard noudattaa samoja toimintaperiaatteita kuin pylväskuvissa. Viivakuvan wizardin heikkoutena verrattuna tavalliseen viivakuvatyökaluun on, että sen avulla kuvaan ei pysty määrittelemään ryhmittelevää muuttujaa, jonka avulla samaan kuvaan saataisiin eri ryhmiä kuvaavia erilaisia viivoja. Wizard-työkalu viivakuville avataan valitsemalla ensin data aktiiviseksi prosessidiagrammista ja klikkaamalla "Graph" -> "Wizards" -> "Line Plot". Tällöin aukeaa ensimmäinen sivu:



Seuraavalla sivulla määritellään muuttujien roolit:

The screenshot shows the 'Line Plot for KETUT_NUM' dialog box at step 2 of 5, 'Assign variables to roles'. The 'Required' section has 'Horizontal' set to 'kuuk_num' and 'Vertical' set to 'paino'. The 'Optional' section has 'Chart by' checked and set to 'ID', and 'Summarize for each distinct horizontal value' unchecked with 'Function' set to 'Average'. A preview window on the right shows a line plot with two data points. A text box explains that 'Horizontal' and 'Vertical' are mandatory. Another text box explains that 'Chart by' allows for separate plots for each horizontal value, and 'Summarize for each distinct horizontal value' allows for averaging values across the y-axis.

”Horizontal” eli x-akselin ja ”Vertical” eli y-akselin muuttujien määritykset ovat pakollisia.

Lisäksi kohdassa ”Chart by” voidaan määrittellä, että jonkin muuttujan tasoille piirretään erilliset kuvat. Kohdassa ”Summarize for each distinct horizontal value” voidaan määrätä, että kuvaan piirretään y-muuttujan arvojen sijasta listasta valittavan tunnusluvun arvot, jotka on laskettu erikseen kullekin x-muuttujan arvolle.

Kolmannella sivulla määritetään viivakuvan asetuksia:

The screenshot shows the 'Line Plot for KETUT_NUM' dialog box at step 3 of 5, 'Line Plot options'. The 'Interpolation' section has 'Type' set to 'Line'. The 'Line' section has 'Style' set to 'Solid', 'Width (pixels)' set to 2, and 'Color' set to 'Default color'. The 'Data point marker' section has 'Symbol' set to 'Circle', 'Height (points)' set to 10, and 'Color' set to 'Default color'. A text box explains that 'Interpolation' allows for different plot types like 'Scatter', 'Regression', and 'STD'. Another text box explains that 'Line' and 'Data point marker' settings can be customized. A preview window on the right shows a line plot with a grid and x-axis labeled 'kuuk_num'.

”Interpolation”-kohdassa määritetään kuvan tyyppi. Viivakuvan (”Line”) lisäksi vaihtoehtoina ovat muun muassa sirontakuviio (”Scatter”), sirontakuviio ja regressiomalli (”Regression”) ja keskihajontakuviio (”STD”). Kuhunkin kuvatyyppiin liittyvät lisämääritykset saa valittua kohdasta ”Detail”.

Myös viivojen (”Line”) ja symbolien (”Data point marker”) ulkoasuun voi vaikuttaa.

Neljännellä sivulla on vielä muutama kuvan ulkoasua koskeva lisämäärittäminen:

Line Plot for KETUT_NUM

4 of 5 Specify appearance

Appearance options

- Use reference lines
- Tick marks

Axis Labels

Tällä sivulla voidaan valita käytetäänkö referenssi viivoja ("Use reference lines") ja asteikon jakoviivoja ("Tick marks"). Akseleiden selitystekstit voi myös vaihtaa ("Axis Labels").

Sample chart:

paino

kuuk_num	paino
1	12
2	6
3	18
4	15
5	3
6	9
7	12

< Back Next > Finish Cancel Help

Viimeisellä sivulla voi vielä vaihtaa kuvien otsikoita ja alaviitteitä. Sivun on täsmälleen samanlainen kuin pylväskuvion wizard-työkalun yhteydessä. Kuva saadaan valmiiksi painamalla lopuksi "Finish".

Tehtävä: Ota projektista käyttöön käännetty kettudata nimeltä **ketut_num**, johon on tehty numeerinen kuukausimuuttuja (ks. luku **Viivakuva (Line Plot)**). Piirrä wizard-työkalulla viivakuva, jossa x-akselilla on kuukausi ja y-akselilla ketun paino. Yhdistä kunkin ketun toistomittaukset yli kuukausien viivalla.

Wizard-työkalu piirakkakuviolle

Piirakkakuvioiden wizard-työkalussa lähdetään liikkeelle samalla tavalla kuin muissa wizard-työkaluissa eli valitaan data aktiiviseksi prosessidiagrammissa ja klikataan **"Graph" - > "Wizards" -> "Pie Chart"**. Ensimmäinen ikkuna on sama kuin muissa wizard-työkaluissa, eli siitä voi tarkistaa, onko käytössä oikea data. Toisella sivulla määritetään muuttujien roolit:

Pie Chart for VILJ2006_MUOK

2 of 4 Assign variables to roles

Required

Slice: sukup

Slice size: (Percentage)

Optional

Group by: id

Stack by: id

Chart by: id

Lisäksi piirakkakuvioita voidaan ryhmitellä jonkin muun muuttujan arvojen mukaan erillisiksi kuvioiksi samaan kuvaan, eri kuviin tai eri kerroksiin. Tämä voidaan tehdä "Optional"-otsikon alla olevista kohdista.

Välttämättömät asetukset tehdään kohdissa "Slice" ja "Slice size". "Slice"-kohdassa määritellään muuttuja, jonka kullekin luokalle tulee oma sektori piirakkaan. "Slice size" -kohdassa määritellään, miten sektorin suuruus määräytyy. Lisäasetuksiin pääsee tässäkin vaikuttamaan kohtien jäljessä olevista painikkeista.

< Back Next > Finish Cancel Help

Kolmannella sivulla voidaan vaikuttaa piirakkakuvion ulkoasuun:

Pie Chart for VILJ2006_MUOK

3 of 4 Specify appearance

3D chart

Colors

Schemes: Spring

Labels

Legend: South

Slice name: Outside

Percentage: Outside

Data value: Outside

Sample chart:

Tällä sivulla voidaan vaikuttaa kolmiulotteisuuteen, värimaailmaan sekä tekstien ja numeroiden sijoitteluun kuviossa.

< Back Next > Finish Cancel Help

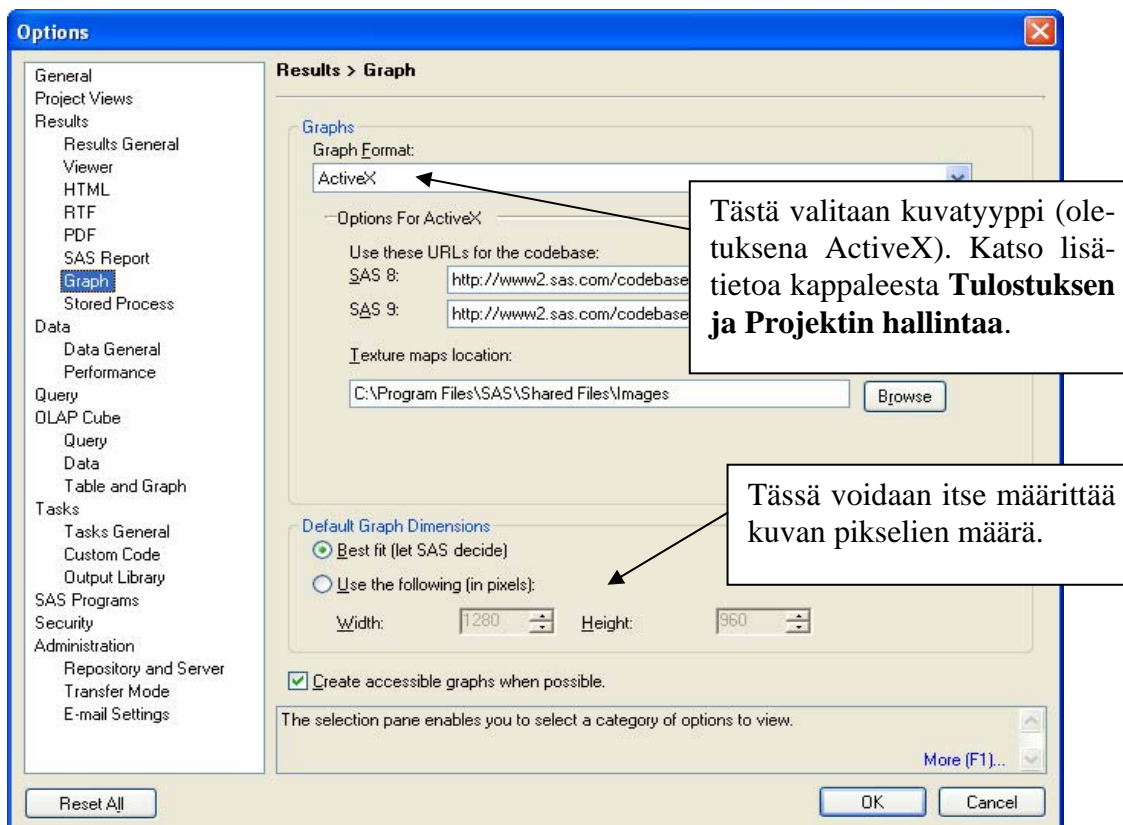
Neljännellä sivulla voidaan vaikuttaa kuvan otsikkoon ja alaviitteeseen samoin kuin muissa wizard-työkaluissa. Lopuksi painetaan "Finish", ja kuva on valmis.

Tehtävä: Piirrä wizard-työkalua käyttäen piirakkakuvi, jonka avulla havainnollistetaan sukupuolimuuuttujan frekvensijakaumaa.

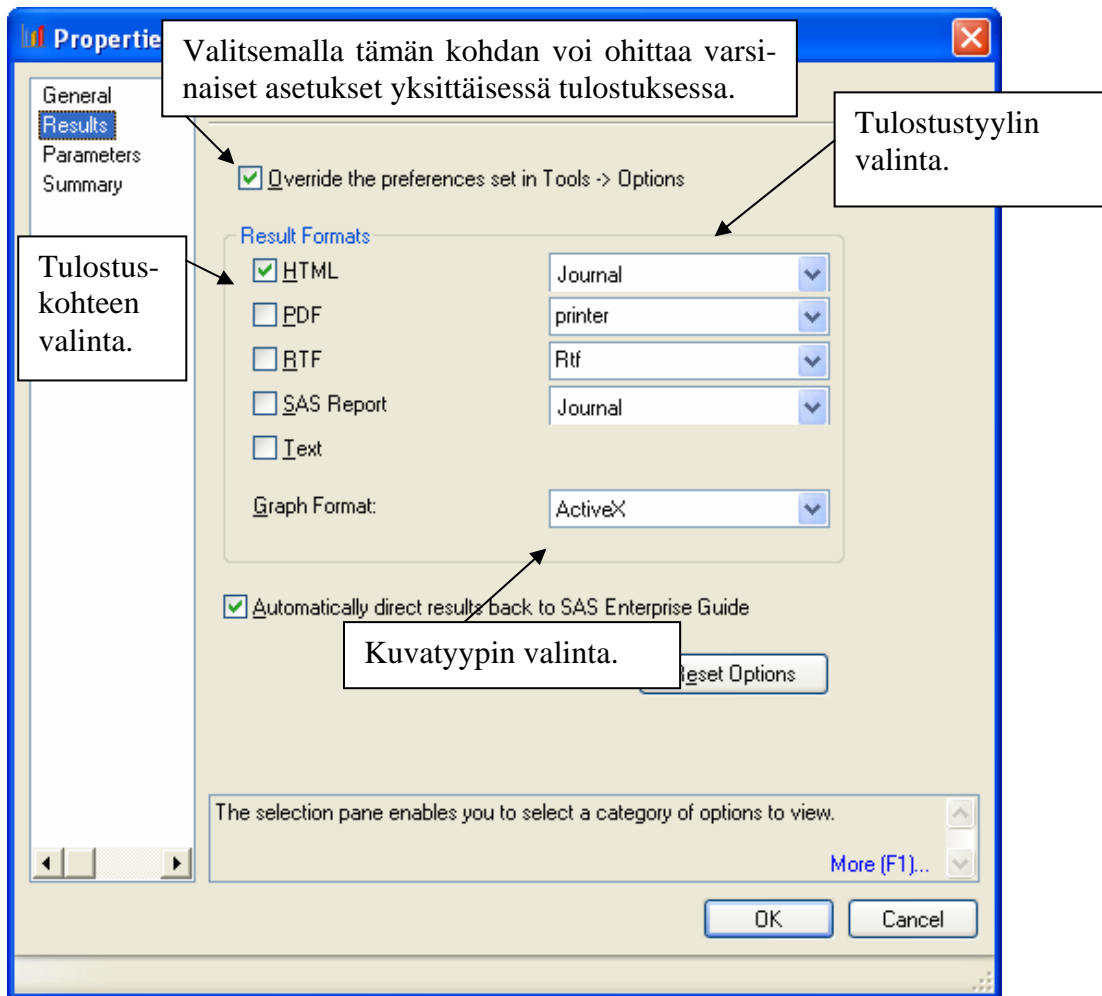
8.7 Kuvien tulostusmuoto ja yleisasetukset

Kuvien asetukset muokataan pääasiassa kuvien määrittelyn yhteydessä. Kuvia voidaan kuitenkin muokata myös jälkikäteen, jolloin päästään osittain vaikuttamaan sellaisiin asioihin, joita ei kuvaa määriteltäessä saa muokattua. Jos kuva on tulostettu HTML- tai RTF-kohteeseen, sitä voidaan muokata klikkaamalla kuvan päällä hiiren oikealla painikkeella ja valitsemalla ”Graph Properties”. Tällöin aukeaa valikko, josta päästään vaihtamaan kuvan asetuksia. Kannattaa kuitenkin huomioida, että muokattaessa valmista kuvaa tällä tavalla, muutokset eivät tallennu projektin mukana. Muokattu kuva kannattaa siis joko tallentaa omaksi kuvatiedostokseen (hiiren oikealla painikkeella aukeavasta valikosta ”Save As”) tai kopioida leikepöydän kautta talteen esimerkiksi Wordiin (hiiren oikealla painikkeella aukeavasta valikosta ”Copy”). Valmiin kuvan muokkaaminen onnistuu vain tavallisilla ActiveX-kuvilla EG:ssä (HTML ja RTF) ja Internet Explorerissa (HTML). Huomaa, että muokkaaminen ei onnistu muiden ohella ActiveX image (SASV9) -kuvatyyppillä. ActiveX image (SASV9) -kuvatyyppi on kuitenkin usein käyttökelpoinen, koska sen avulla saadaan hyvälaatuisena tulostuvia kuvia. Tulostuslaatuun voi vaikuttaa myös suurentamalla kuvien pikselimäärää (”Tools” -> ”Options” -> ”Graph”).

Tulostuskohteen pääsee valitsemaan klikkaamalla ”Tools” -> ”Options” -> ”Results General”. Valittavana on HTML (oletus), PDF, RTF, SAS Report ja Text Output (= perinteinen SAS-tulostus, jota ei voida käyttää kuville). ”Viewer”-välilehdellä voi valita ohjelman, johon tulostus automaattisesti aukeaa, ja kunkin tulostuskohteen omalla välilehdellä voi tehdä niihin liittyviä tarkempia määrittämiä (muun muassa tulostustyylin määrittäminen). ”Graph”-välilehdellä ovat varsinaiset kuvan asetusten määrittäykset:



Asetusten määrittäminen ennen kuvan määrittelyä ei ole kuitenkaan ainoa tapa. Asetuksia voidaan lisäksi muokata yksittäiselle jo määritellylle tehtäväsolmulle seuraavasti: klikataan tehtäväsolmua prosessidiagrammissa hiiren oikealla painikkeella, valitaan ”Properties” ja välilehti ”Results”:



Näiden valintojen jälkeen kyseinen tehtäväsolmu pitää vielä suorittaa uudelleen. Se onnistuu klikkaamalla tehtäväsolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Run työkalun nimi”.

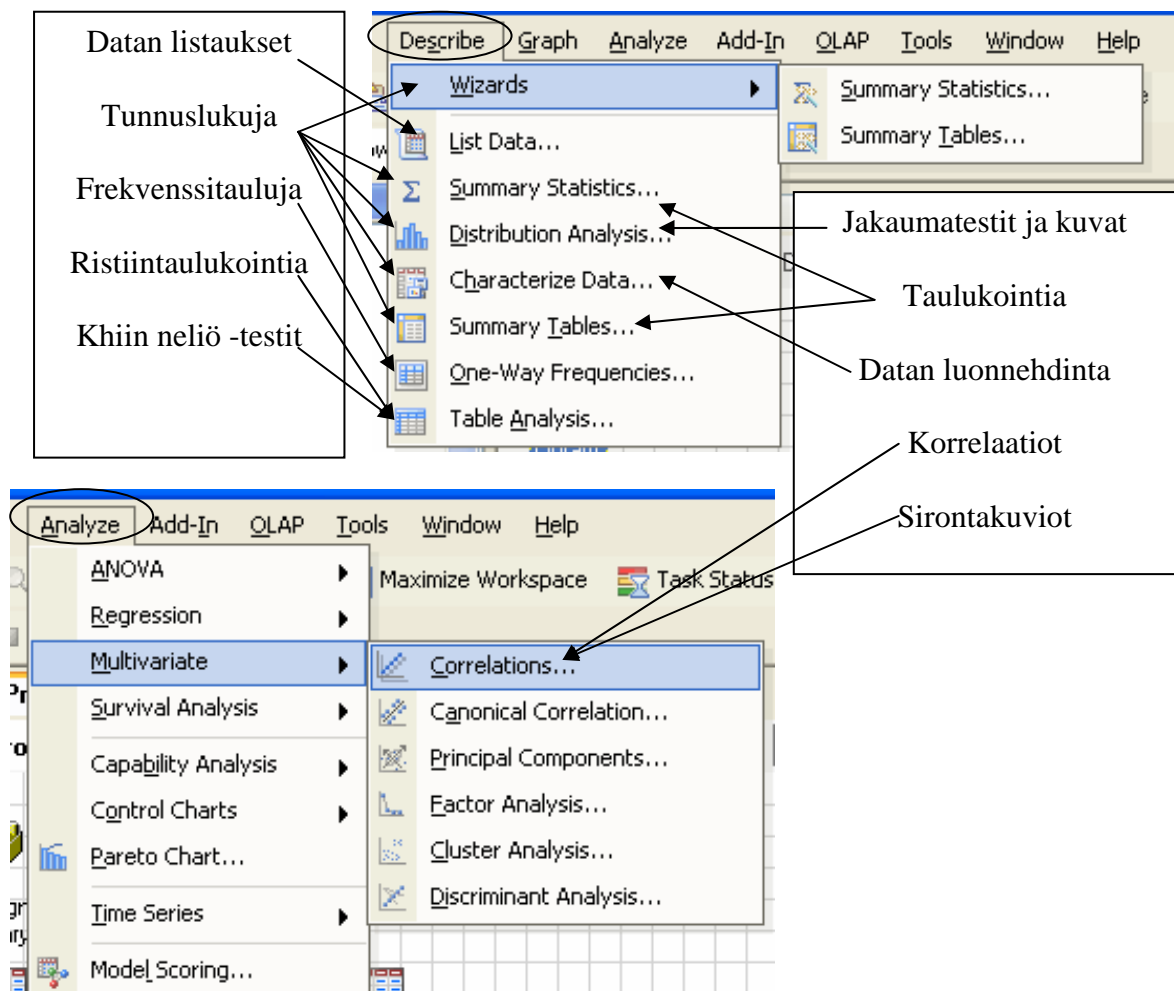
Yksittäisen tulostusikkunaan tulostetun kuvan voi myös tallentaa kuvatiedostoon tai siirtää sen leikepöydän kautta (”copy-paste”) esimerkiksi wordiin. Kuvan tallentaminen *.jpeg-tiedostoksi onnistuu klikkaamalla valmista kuvaa tulostusikkunassa hiiren oikealla painikkeella ja valitsemalla ”Save as”. Tällöin aukeaa ikkuna, jossa voidaan valita kansio ja tiedoston nimi tallennettavalle kuvalle. Leikepöydän kautta kopioiminen puolestaan onnistuu valitsemalla samasta valikosta ”Copy” ja esimerkiksi Wordissa ”Liitä” tai ”Paste”.

Valmiin tulostussolmun voi myös avata toisessa ohjelmassa. Se onnistuu klikkaamalla solmua hiiren oikealla painikkeella ja valitsemalla ”Open With ohjelman nimi”. Vaihtoehdot ovat HTML:lle Internet Explorer, RTF:lle Word ja PDF:lle Adobe Reader.

9 Datan tarkastelu Describe- ja Analyze-valikoiden työkaluilla

Edellä on käyty läpi erilaisia kuvatyyppejä, joiden avulla voidaan tutkia ja vertailla aineistossa olevien muuttujien jakaumia ja havainnollistaa muuttujien välisiä yhteyksiä. Kuvat ovat erinomaisia tulosten esittämisessä ja haluttaessa saada alustava käsitys aineistosta. Tietoa aineistossa olevien muuttujien jakaumista ja muuttujien välisistä yhteyksistä saadaan myös erilaisten kuvailevien tunnuslukujen, taulukoiden ja datalistausten avulla. Tässä esiteltävien datan kuvailuun käytettävien SAS Enterprise Guide -työkalujen avulla pystytään tunnuslukujen tai taulukoiden lisäksi tuottamaan myös joitain graafisia havainnollistuksia muuttujien jakaumista ja yhteyksistä.

Datan tarkastelusta käydään läpi seuraavat Describe- ja Analyze-valikoiden työkalut:

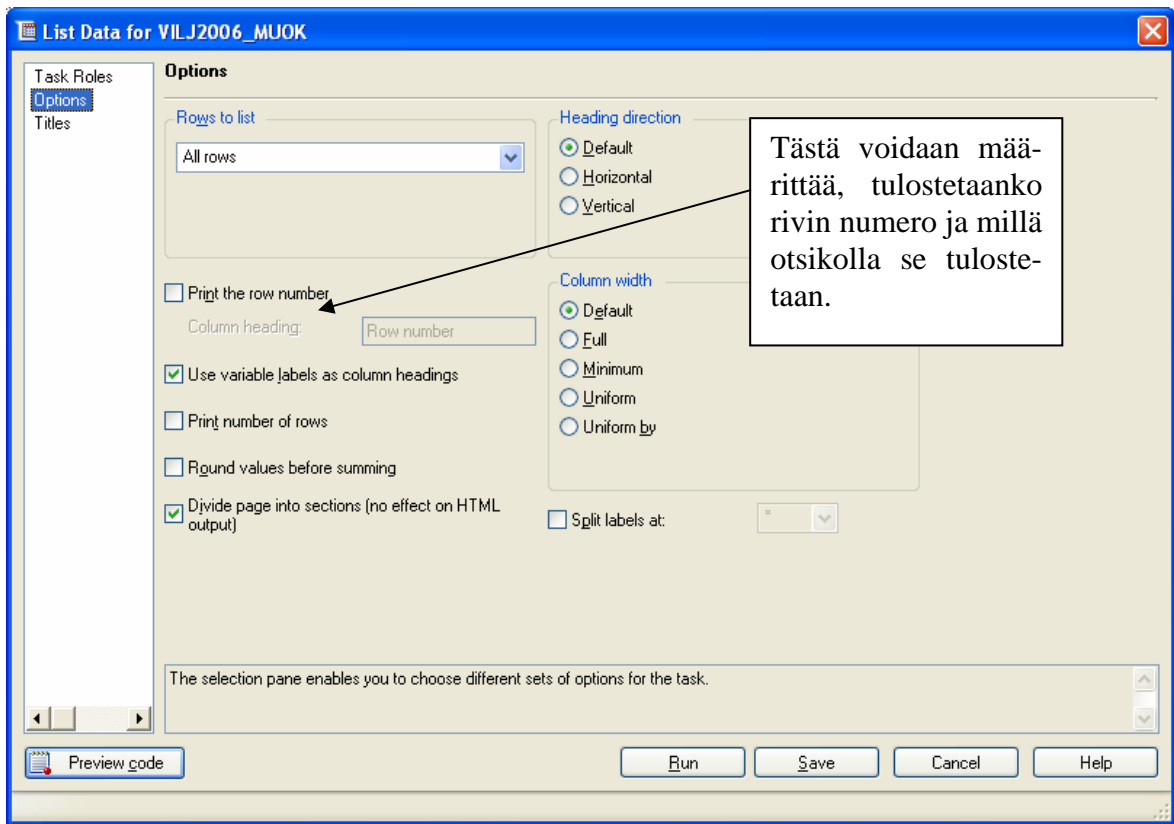


9.1 Datan listaukset

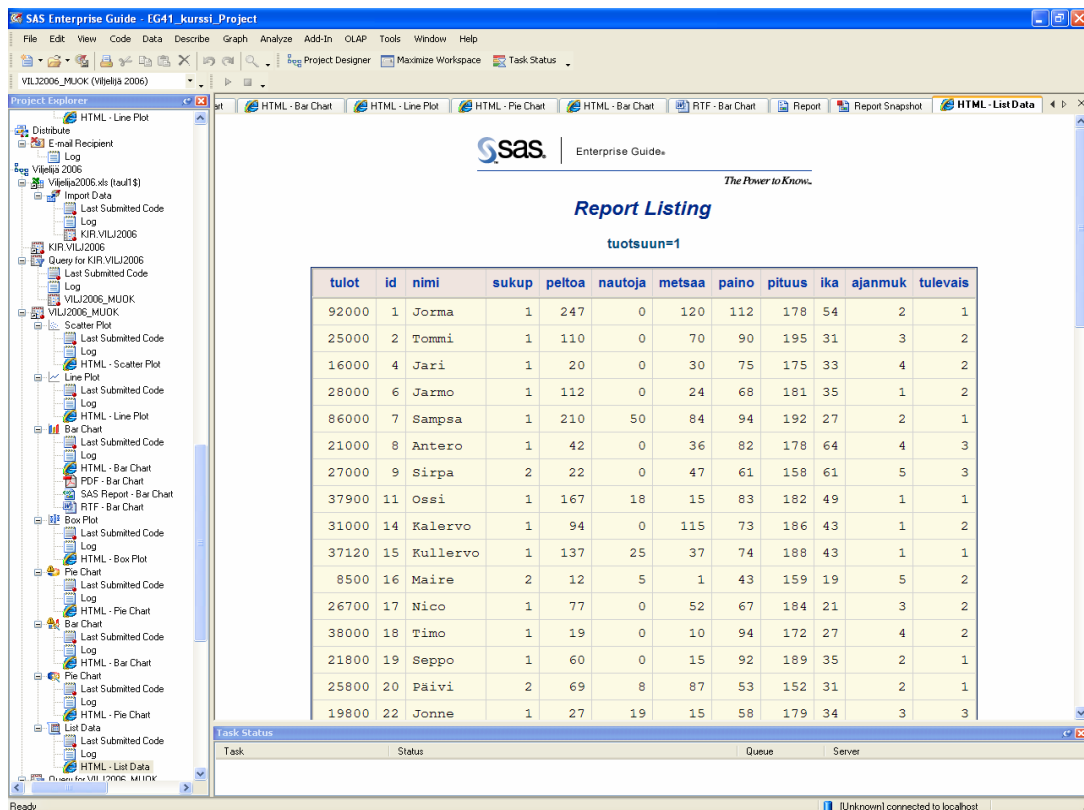
Yksinkertaisin tapa tarkastella dataa on tulostaa koko data tai osia siitä. Listausta voidaan myös muokata lukijaystävällisemmäksi. Listaukseen voidaan valita vain osa muuttujista ja suorittaa erilaista datan järjestämistä ja ryhmittelyä. Listausta tuotetaan valitsemalla ”Describe” -> ”List Data”, jolloin aukeaa ”Task Roles” -välilehti. ”List variables” -kohdassa luetellaan muuttujat, joiden arvot halutaan tulostaa. ”Group analysis by” -kohtaan puolestaan laitetaan sellainen muuttuja tai sellaiset muuttujat, joiden mukaan tulostus halutaan järjestää osaryhmiin. Tällöin kukin osaryhmä tulostuu omana taulukkonaan. Jos halutaan käyttää osaryhmiin jakoa, mutta tulostaa kaikki rivit silti samaan yhteen taulukkoon, ryh-

mittelevä muuttuja tai muuttujat laitetaan ”Grouping analysis by” -kohdan lisäksi ”Identifying label” -kohtaan.

Klikkaamalla ”Options” päästään määrittelemään tulostusta tarkemmin:



Painamalla ”Run” saadaan seuraavanlainen tulostus:

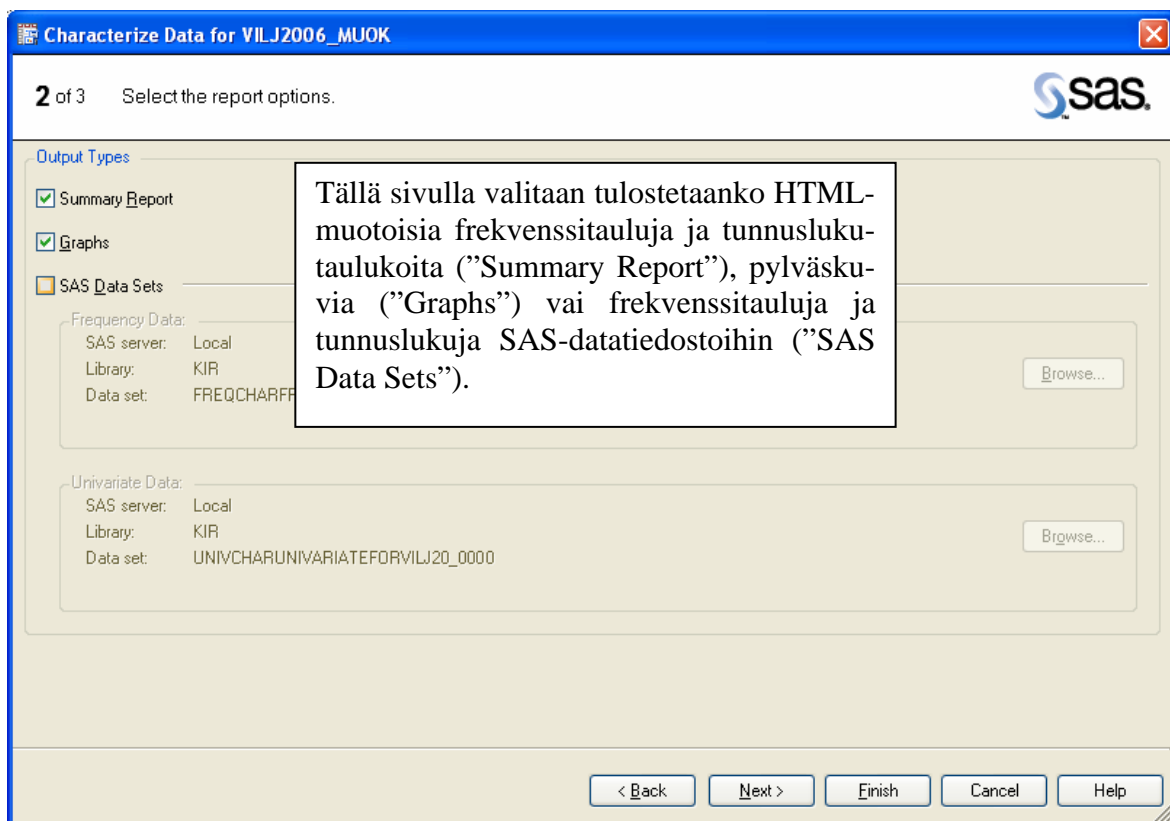


Tulostusta voitaisiin vielä hioa esimerkiksi määrittelemällä dataan selitetekstejä muuttujille ja formaatteja muuttujien arvoille käyttäen samaa tekniikkaa kuin **Datan muokkaus** -luvussa.

Tehtävä: Tee Viljelijä 2006 -aineistosta (**vilj2006_muok**) listaus, jossa tulostetaan viljelijän nimi sekä tiedot kunkin viljelijän peltohehtaareista, metsähehtaareista, nautojen lukumäärästä ja tuloista.

9.2 Datan luonnehdinta -työkalu (Characterize Data)

Datan luonnehdinta -työkalun avulla saadaan helposti yleiskuva aineiston muuttujista. Työkalu tulostaa kategorisille muuttujille frekvenssitauluja sekä pylväskuvioita. Numeerisille muuttujille puolestaan voidaan tulostaa tunnuslukuja sekä histogrammeja. Kaikki tulostus ei ole välttämättä mielekästä, mutta yleiskuvan muuttujien jakaumista tällä työkalulla saa nopeasti. Työkalu saadaan käyttöön valitsemalla data aktiiviseksi prosessidiagrammista ja valitsemalla ”Describe” -> ”Characterize Data”. Ensimmäisellä sivulla voi vaikuttaa käytettävään dataan. Toisella sivulla tehdään tärkeimmät valinnat:



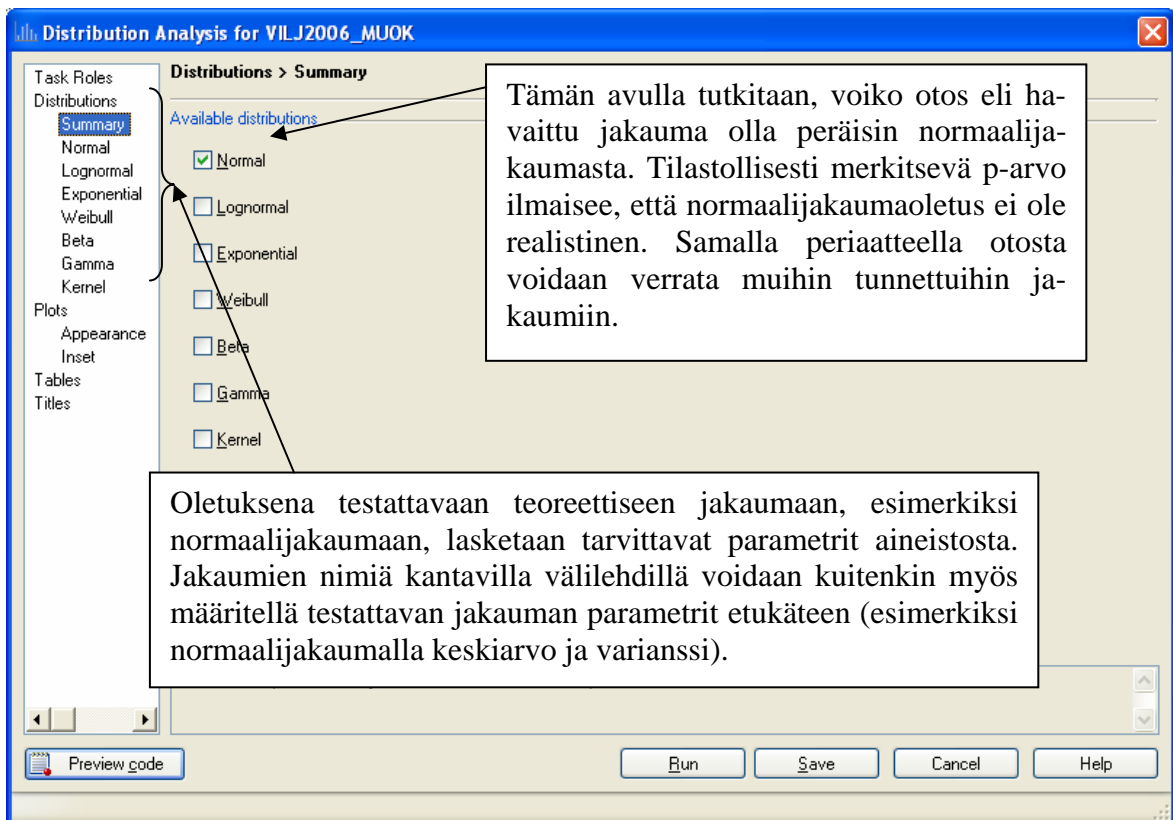
Viimeisellä sivulla voi vielä valita, montako kategorisen muuttujan luokkaa huomioidaan maksimissaan erillisinä luokkina. Lopuksi painetaan "Finish".

9.3 Jatkuvan muuttujan jakauman tarkastelu

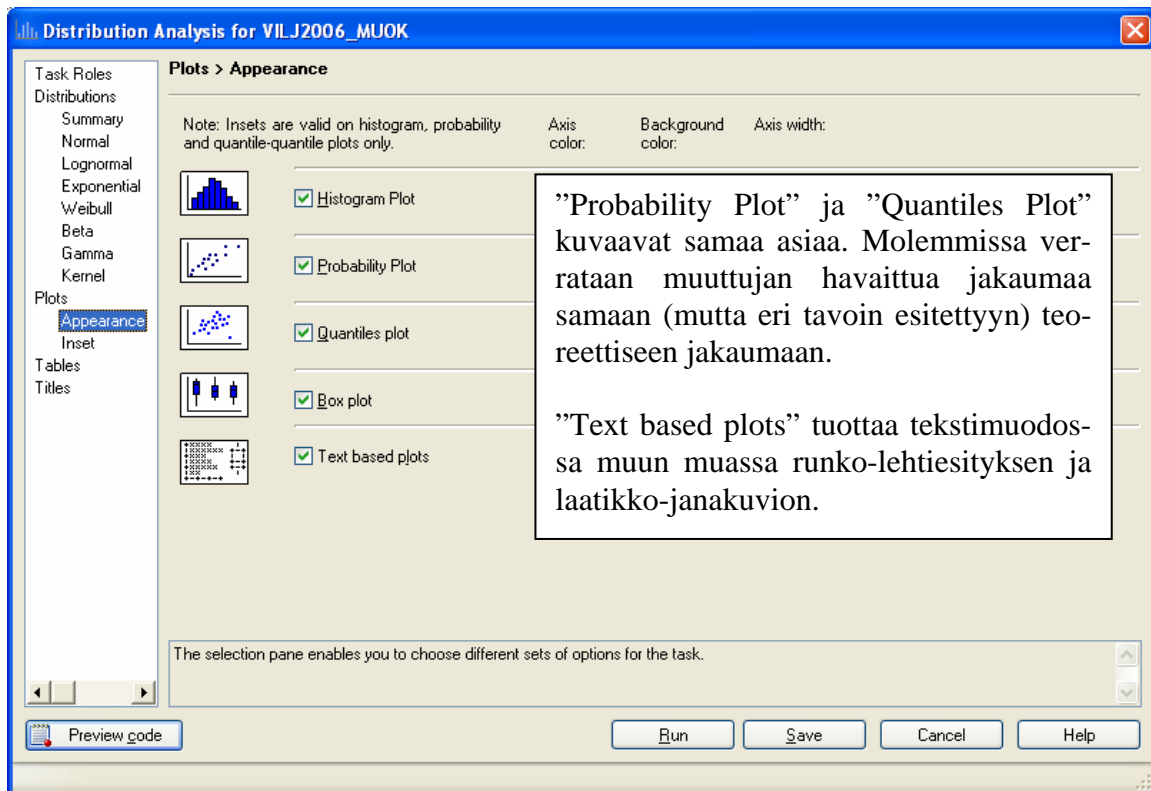
Datassa olevan jatkuvan muuttujan jakauman tarkastelua varten esitellään kolme varsinaista SAS Enterprise Guiden työkalua sekä tunnuslukujen taulukointiin käytettävissä olevat wizard-työkalut.

9.3.1 Distribution Analysis -työkalu

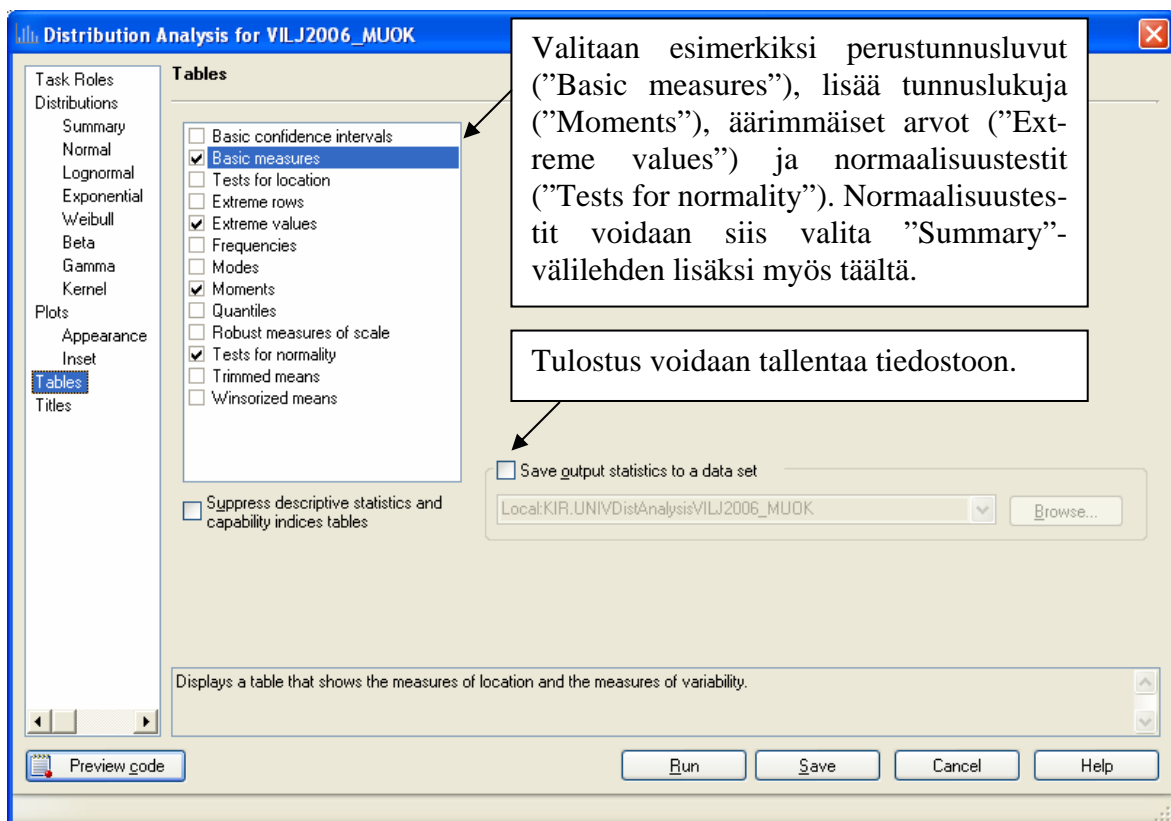
Yhden jatkuvan muuttujan jakauman tarkastelua on kätevää tehdä ”Distribution Analysis” työkalulla, jolla saadaan tulostetuksi listauksia jakaumaa kuvaavista tunnusluvuista. Lisäksi voidaan testata, onko havaittu otos peräisin jostakin teoreettisesta jakaumasta, esimerkiksi normaalijakaumasta. Myös erilaisia jakaumia havainnollistavia kuvia saa tulostettua kätevästi. Alkuun päästään valitsemalla käytettävä data aktiiviseksi ja klikkaamalla ”Describe” -> ”Distribution Analysis”. Tällöin aukeaa ”Task Roles” -välilehti, jossa tarkasteltava jatkuva muuttuja tai jatkuvat muuttujat sijoitetaan kohtaan ”Analysis variables”. Jos haluttaisiin tarkastella yhden jatkuvan muuttujan jakaumaa jonkin kategorisen muuttujan eri luokissa, sijoitettaisiin kategorinen muuttuja kohtaan ”Group analysis by”. Välilehdellä ”Summary” puolestaan voidaan määritellä testit, joilla tutkitaan ovatko havaitut muuttujan arvot otos jostakin tunnetusta jakaumasta:



”Plots” -> ”Appearance”-välilehdellä voidaan puolestaan määrittellä muuttujan jakaamaa havainnollistavat kuvat:



”Tables”-välilehdellä puolestaan määritetään tulostettavat tunnuslukutaulut:

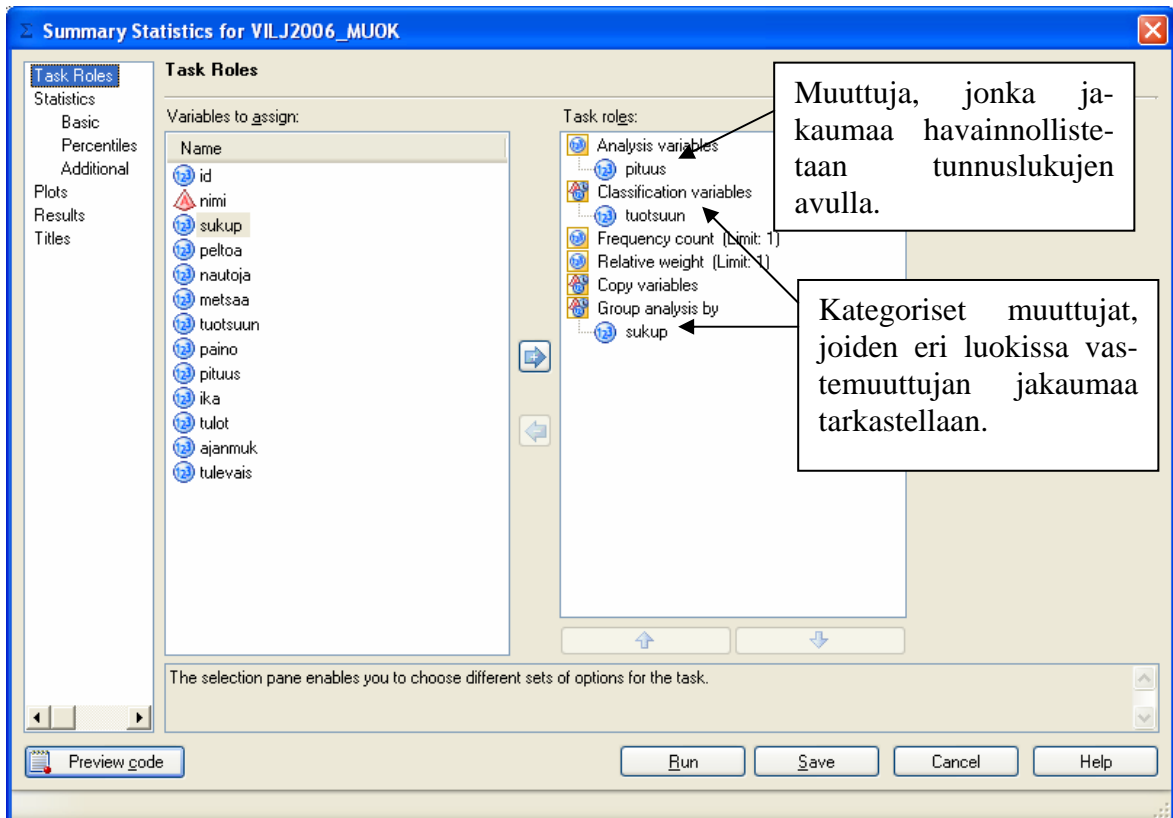


Lopuksi valitaan ”Run”.

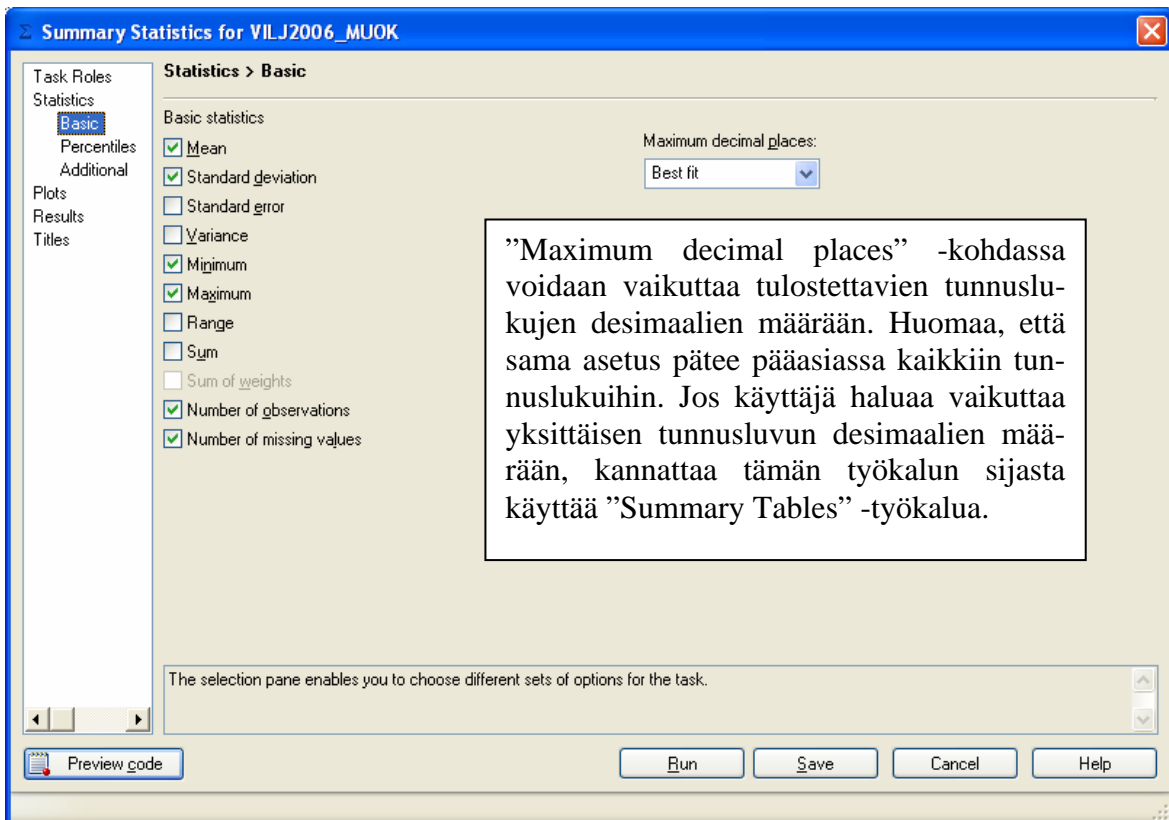
Tehtävä: Tutki Viljelijä 2006 -aineistosta (**vilj2006_muok**) ikä-muuttujan (ika) jakaumaa. Tarkastele sijaintia, hajontaa, muotoa ja poikkeavia havaintoja.

9.3.2 Summary Statistics -työkalu

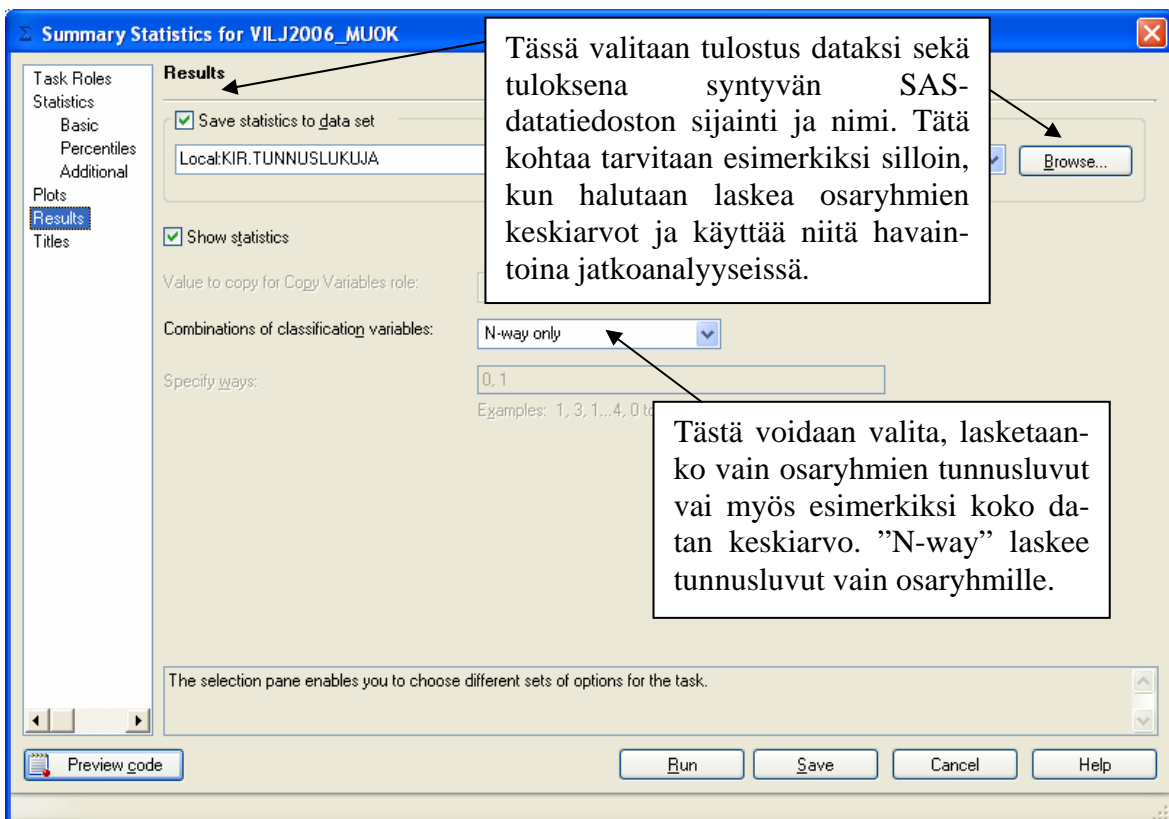
Enterprise Guiden Summary Statistics -työkalulla on mahdollista tulostaa jatkuvan muuttujan jakaumaa kuvaavia tunnuslukuja taulukkomuodossa. Tämä on erityisen kätevä työkalu, jos halutaan laskea tunnuslukuja erikseen yhden tai useamman kategorisen muuttujan luokissa. Työkalu saadaan käyttöön valitsemalla käytettävä data aktiiviseksi prosessidiagrammista ja klikkaamalla **"Describe"** -> **"Summary statistics"**, jolloin aukeaa "Task Roles" -ikkuna:



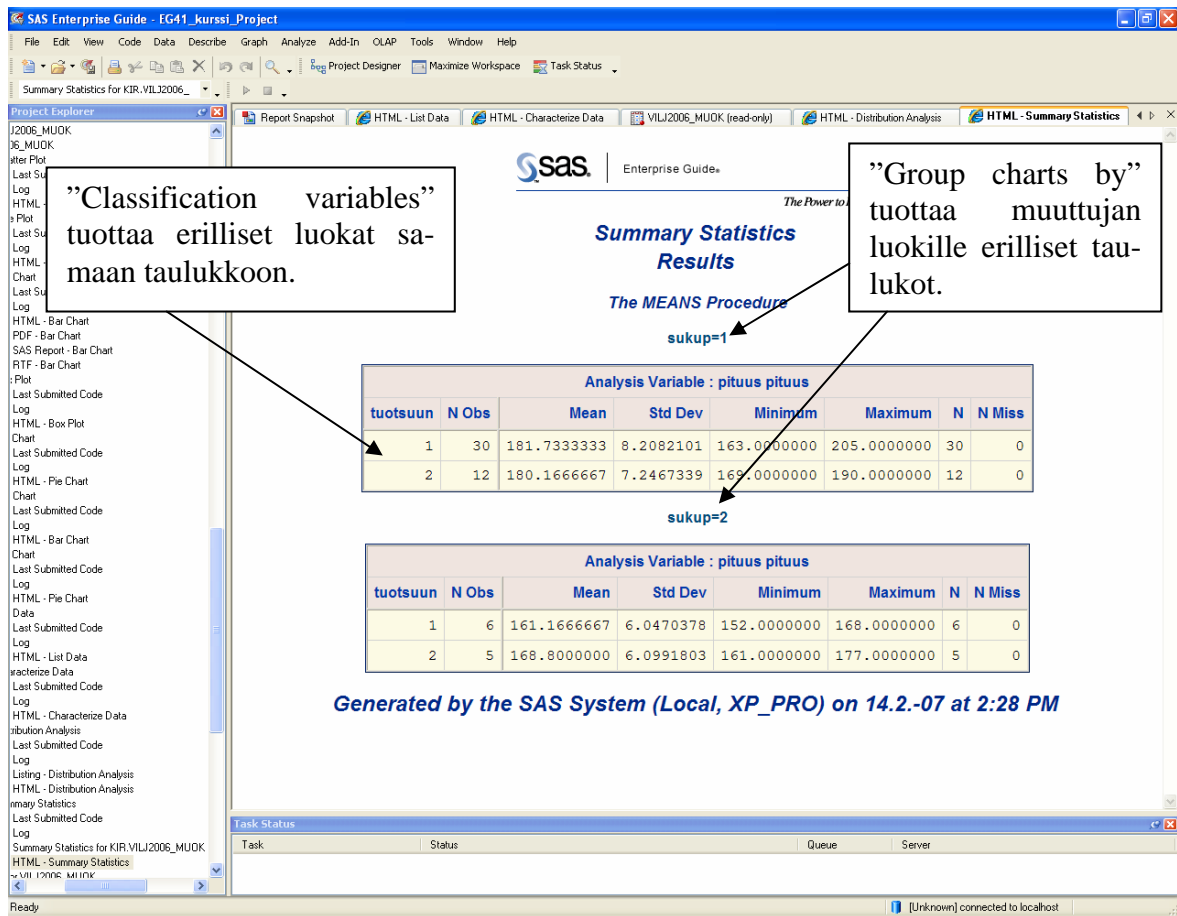
Laskettavat tilastolliset tunnusluvut ja testit valitaan ”Basic”-, ”Percentiles”- ja ”Additional”-välilehdillä. Ohessa ”Basic”-välilehti:



”Plots”-välilehdellä voidaan määrätä piirrettäväksi histogrammi ja boxplot-kuva. Tulostuksen ajaminen SAS-dataksi voidaan määritellä ”Results”-välilehdellä:



Tulostuksena saadaan seuraava taulukko:

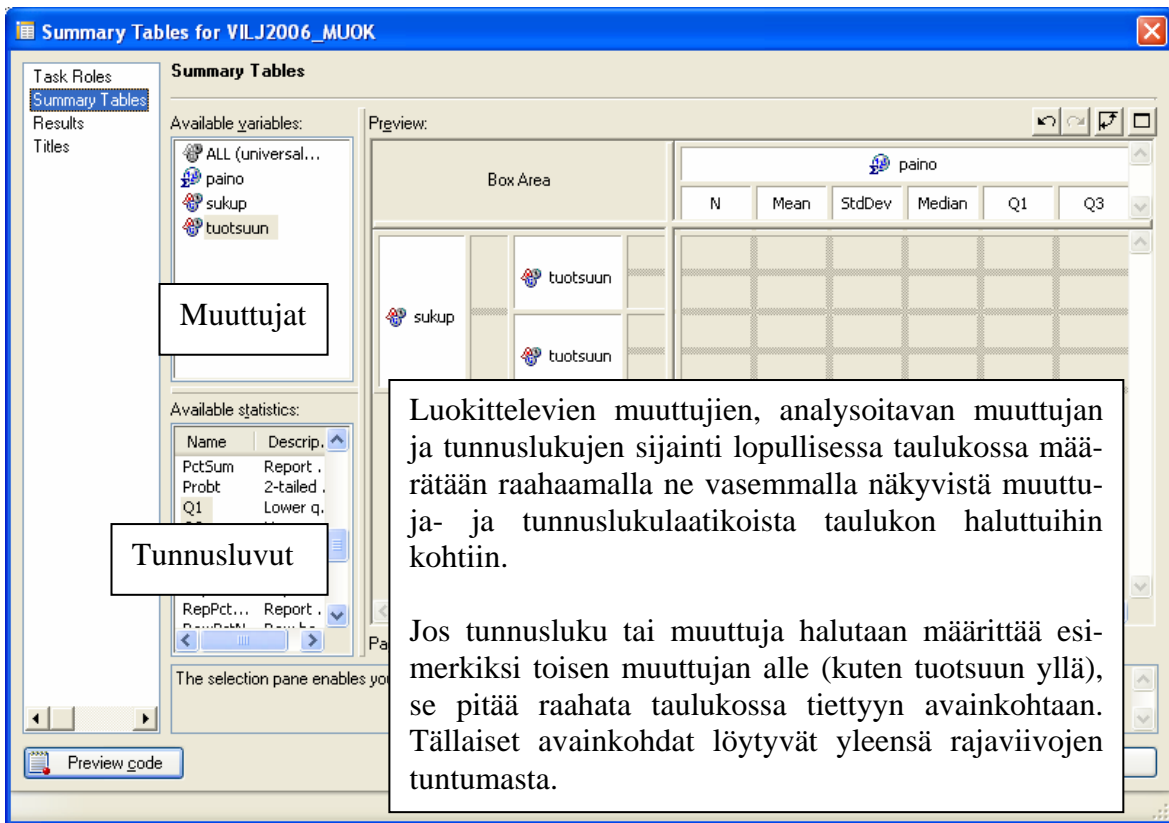


Tehtävä: Tutki Viljelijä 2006 -aineistosta (**vilj2006_muok**) pituus-muuttujan jakaumaa sukupuolen ja tuotantosuunnan eri luokkakombinaatioissa. Käytä tähän tarkasteluun tilastollisia tunnuslukuja ja mahdollisesti myös kuvia.

9.3.3 Summary Tables -työkalu

Kolmas tapa laskea tunnuslukuja on käyttää taulukointityökalua Summary Tables. Tämä työkalu saadaan käyttöön valitsemalla käytettävä data aktiiviseksi ja klikkaamalla ”Describe” -> ”Summary Tables”. Sitten annetaan ”Task Roles” -välilehdellä muuttujat, joiden jakaumaa tarkastellaan (”Analysis variables”), sekä ryhmittelevät muuttujat (”Classification variables”).

Varsinaiseen taulukon määrittelyyn päästään ”Summary Tables” -välilehdellä:



Tunnuslukuihin voi tehdä lisämäärittäyksiä valikosta, joka aukeaa klikkaamalla hiiren oikealla näppäimellä taulukkoon lisätyn tunnusluvun päällä. Tällöin voidaan valita esimerkiksi ”Data Value Properties”, jossa voidaan ”Format”-välilehdellä yhdistää tunnuslukuun formaatti w.d, jonka avulla saadaan vaihdettua desimaalien määrää tulostuksessa.

Tuloksena syntyy seuraava taulukko (RTF-kohteeseen ohjattuna):

		paino					
		N	Mean	StdDev	Median	Q1	Q3
sukupuoli	tuotsuun						
1	1	30	83.07	15.58	82.50	73.00	92.00
	2	12	84.08	17.15	80.50	76.50	92.50
2	1	6	54.33	9.95	57.00	43.00	63.00
	2	5	73.80	15.75	80.00	64.00	81.00

RTF-taulukko on siirretty Wordiin EG:n tulostusikkunasta valitsemalla se aktiiviseksi ja kopioimalla ja liittämällä leikepöydän kautta. Samalla tekniikalla voidaan siirtää muitakin EG:n RTF-tulostusobjekteja Wordiin. Tällöin taulukkoa päästään vielä muokkaamaan Wordissa muun muassa siirtämällä sarakkeiden rajoja ja muuttamalla fonttia. Taulukosta saadaan Wordissa muokattua esimerkiksi seuraavanlainen:

		Paino (kg)					
		N	Keskiarvo	Keski-hajonta	Mediaani	Ala-kvartiili	Ylä-kvartiili
Sukupuoli	Tuotantosuunta						
Mies	Peltoviljely	30	83.1	15.58	82.5	73	92
	Kotieläimet	12	84.1	17.15	80.5	76.5	92
Nainen	Peltoviljely	6	54.3	9.95	57	43	63
	Kotieläimet	5	73.8	15.75	80	64	81

Taustavärit saa poistettua Wordissa esimerkiksi valitsemalla koko taulukon aktiiviseksi (klikkaamalla vasemmassa yläkulmassa toisinaan näkyvää laatikkoa), klikkaamalla hiiren oikeaa näppäintä ja valitsemalla ”Reunat ja sävytys”. Aukeavassa ikkunassa mennään ”Sävytys”-välilehdelle ja valitaan ”Täyttökuvio”-kohdassa ”Ei täyttöä”.

Tehtävä: Tee Viljelijä 2006 -aineistosta (**vilj2006_muok**) taulukko, jossa ika-muuttujalle on laskettu soveltuvia tunnuslukuja sukupuolen ja tuotantosuunnan eri luokkakombinaatioissa. Vie taulukko myös Wordiin tai muuhun tekstinkäsittelyohjelmaan ja harjoittele taulukon muokkaamista siellä.

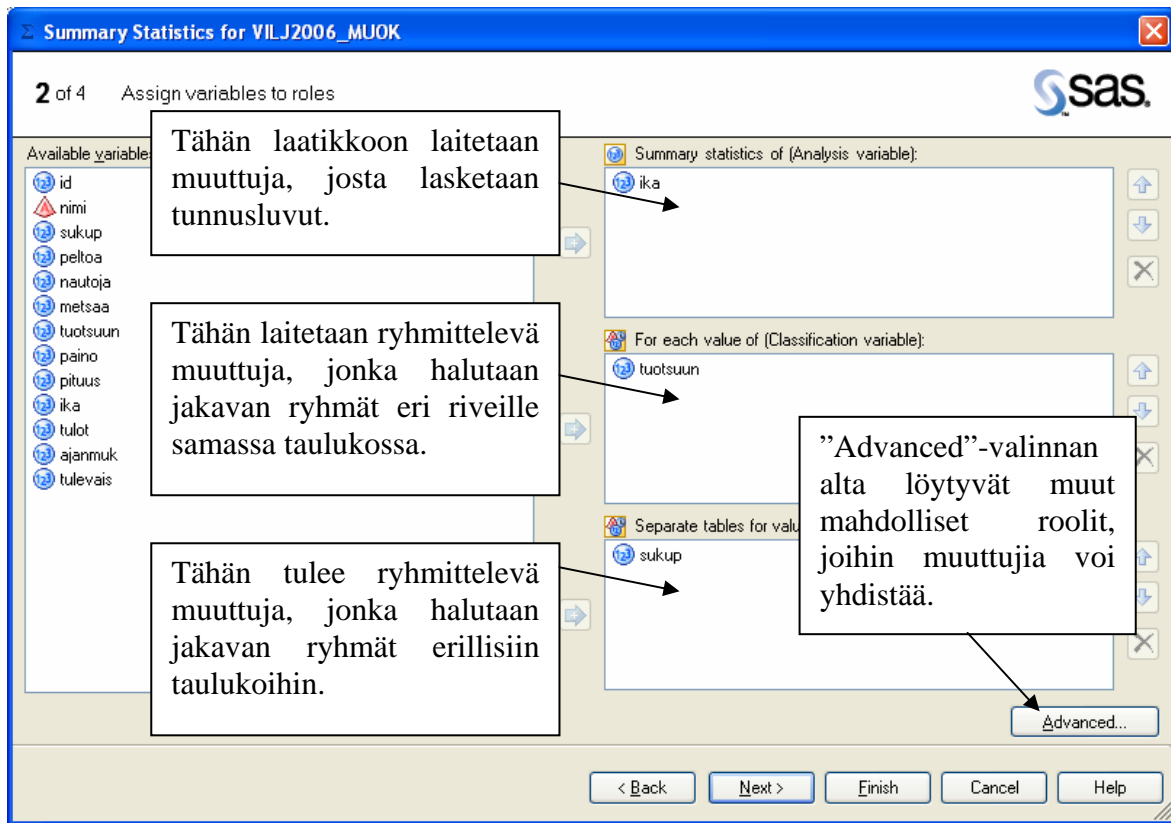
9.3.4 Tunnuslukujen taulukointi wizard-työkaluilla

Samoin kuin kuvien myös jatkuvan muuttujan jakauman tarkasteluun löytyy Enterprise Guiden versiosta 4.1 muutama wizard-työkalu. Näiden työkalujen tarkoituksena on tarjota entistä helppokäyttöisempi, vaihtoehtoinen tapa tunnuslukujen taulukointiin. Wizardissa määrittelyt on jaettu useammalle sivulle, joiden läpi edetään järjestyksessä, ja lopuksi valitaan ”Finish”. Seuraavassa esitellään lyhyesti EG:stä löytyvät ”Summary Statistics”- ja ”Summary Tables” -wizard-työkalut. Nämä vastaavat samannimisiä tavallisia EG-työkaluja.

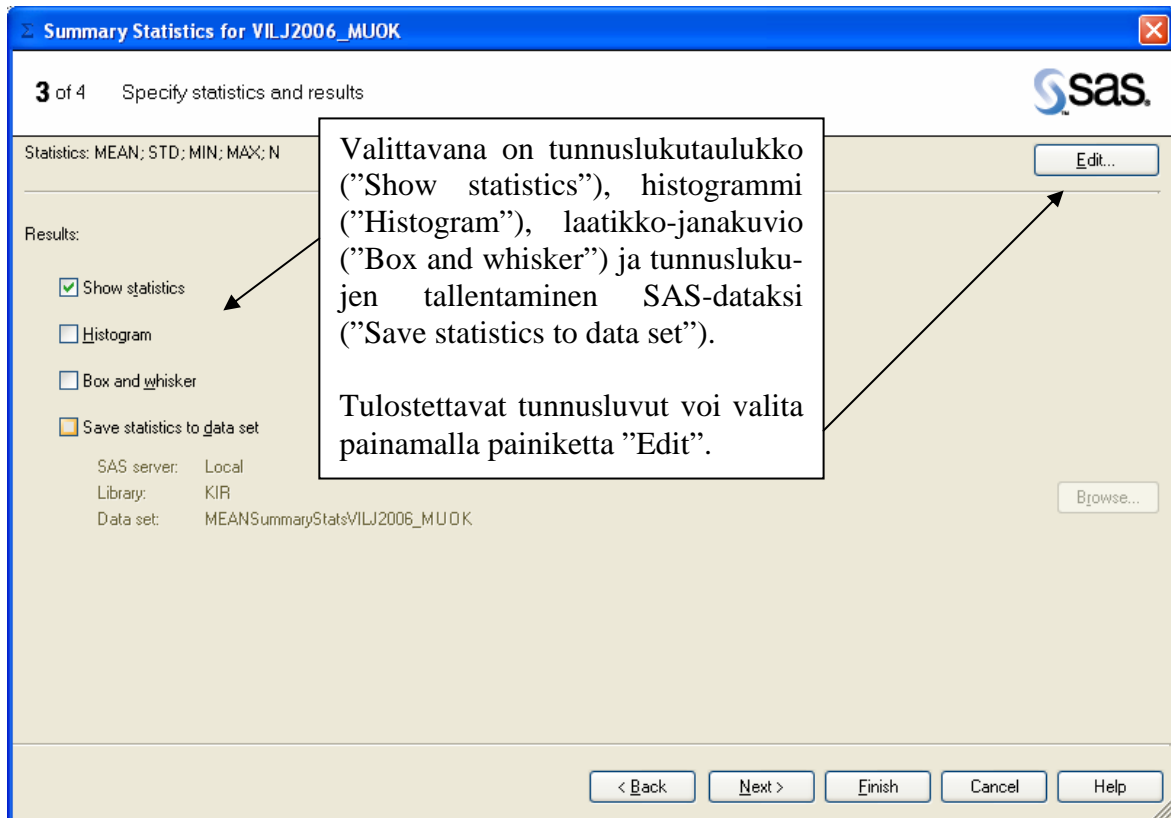
Summary Statistics wizard

Tarkastellaan ensin ”Summary Statistics” -työkalua, joka saadaan käyttöön, kun valitaan aktiiviseksi käytettävä data prosessidiagrammista ja klikataan ”Describe” -> ”Wizards” -> ”Summary Statistics”. Ensimmäisellä sivulla voidaan tarkistaa, että aktiivisena oli varmasti oikea data. Toiselle välilehdelle päästään klikkaamalla ”Next”.

Toisella välilehdellä määritellään muuttujien roolit:



Kolmannella sivulla valitaan mitä tulostusta halutaan:



Viimeisellä sivulla voi vaihtaa kuvaan liittyvät otsikot ja alaviitteet. Lopuksi painetaan ”Finish”, jolloin saadaan samanlainen tulostus kuin vastaavilla valinnoilla tavallista ”Summary Statistics” -työkalua käyttäen:

The screenshot displays the SAS Enterprise Guide interface with the 'Summary Statistics Results' window open. The window shows the results of a Summary Statistics procedure for the variable 'ika' (age), grouped by 'sukupuoli' (sex). The results are presented in two tables, one for 'sukupuoli=1' and one for 'sukupuoli=2'. Each table shows the distribution of 'ika' values across different 'tuotsuun' (production) categories.

Summary Statistics Results
The MEANS Procedure
sukupuoli=1

Analysis Variable : ika ika						
tuotsuun	N Obs	Mean	Std Dev	Minimum	Maximum	N
1	30	39.6333333	11.3577661	16.0000000	64.0000000	30
2	12	44.3333333	12.8511289	26.0000000	60.0000000	12

sukupuoli=2

Analysis Variable : ika ika						
tuotsuun	N Obs	Mean	Std Dev	Minimum	Maximum	N
1	6	34.5000000	14.4741148	19.0000000	61.0000000	6
2	5	38.6000000	16.9351705	23.0000000	67.0000000	5

Generated by the SAS System (Local, XP_PRO) on 14.2.-07 at 3:29 PM

Tehtävä: Tutki Viljelijä 2006 -aineistosta (vilj2006_muok) ika-muuttujan jakaumaa sukupuolen ja tuotantosuunnan eri luokkakombinaatioissa. Laske tarvittavat tilastolliset tunnusluvut Summary Statistics wizard -työkalulla.

Summary Tables wizard

Toinen taulukointia tekevä wizard-työkalu saadaan käyttöön valitsemalla käytettävä data aktiiviseksi ja klikkaamalla ”Describe” -> ”Wizards” -> ”Summary Tables”. Ensimmäisellä sivulla voidaan taas vahvistaa, että käytössä on haluttu data.

Toisella sivulla puolestaan määritellään muuttuja tai muuttujat, joiden arvoista lasketaan tunnuslukuja, sekä laskettavat tunnusluvut:

2 of 6 Select analysis variables and statistics

Analysis variables:

Variable	Statistic
paino	Average
paino	Standard Deviation

Preview:

	paino	paino
	Mean	StdDev
	999	999

Analysis variable labels: in columns

Statistics labels: in columns

Buttons: < Back, Next >, Finish, Cancel, Help

Tehtäessä valintoja työkalu näyttää esikatselun muodostettavasta taulukosta.

Muuttujan saa lisättyä listaan klikkaamalla "Add".

Laskettavan tunnusluvun saa puolestaan valittua alaspäin-valikosta muuttujan kohdalta. Tässä työkalussa on vähemmän tunnuslukuja valittavana kuin varsinaisessa "Summary Tables" -työkalussa. Lisäksi kullekin muuttujalle saa listassa valittua vain yhden tunnusluvun, joten sama muuttuja pitää lisätä listaan useamman kerran, jos halutaan laskea useampi tunnusluku.

Kolmannella sivulla määritellään ryhmittelevät muuttujat:

3 of 6 Select classification variables

Columns: (Optional) Click Add to insert a column variable.

Rows:

sukupu
tuotsuun

Pages: (Optional) Click Add to insert a page variable.

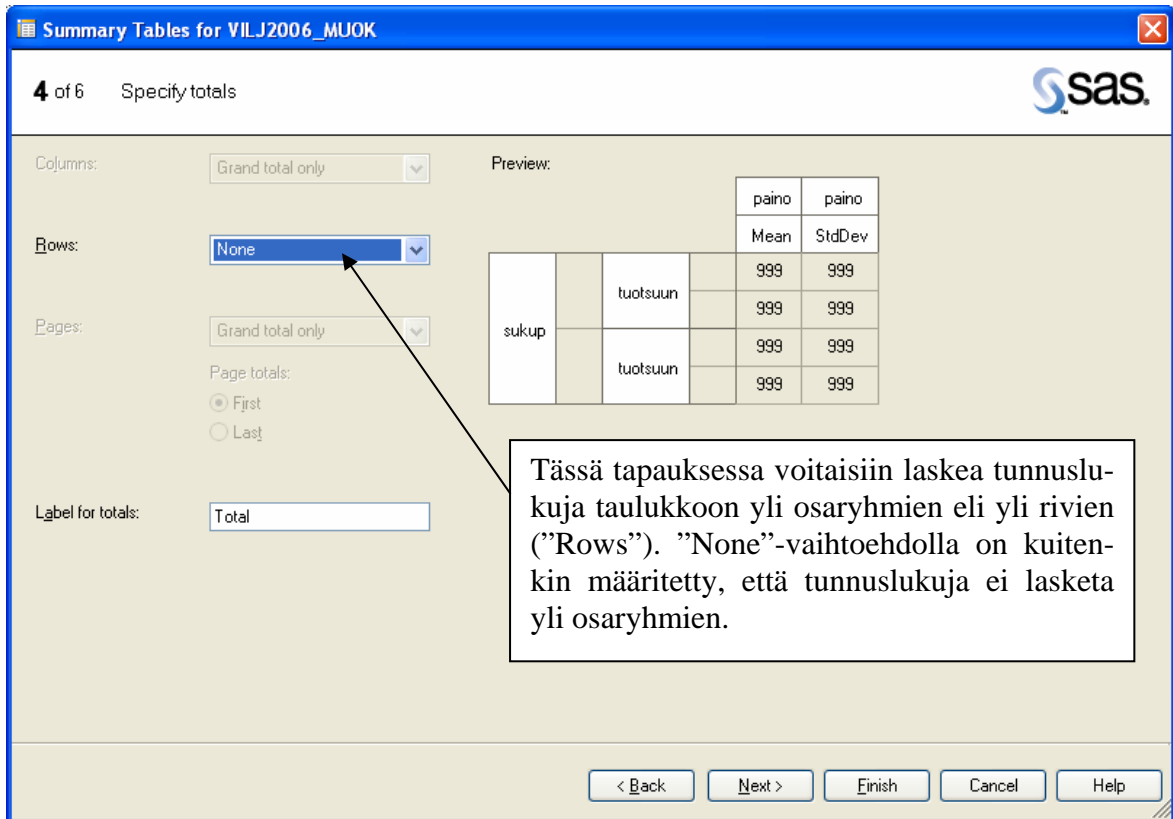
Preview:

		paino	paino
		Mean	StdDev
sukupu	tuotsuun	999	999
	tuotsuun	999	999
Total		999	999

Buttons: < Back, Next >, Finish, Cancel, Help

Ryhmittelevät muuttujat voi sijoittaa sarakemuuttujiksi ("Columns"), rivimuuttujiksi ("Rows") tai sivumuuttujiksi ("Pages"), jolloin muuttujien tasoista tulee valinnan mukaan joko erillisiä rivejä, sarakkeita tai taulukoita. Tässä ryhmittelevät muuttujat on sijoitettu rivimuuttujiksi, koska analysoitava muuttuja on oletusasetuksilla sarakemuuttujana.

Neljännellä sivulla voi määrittellä, lasketaanko tunnuslukuja taulukkoon yli osaryhmien:



Summary Tables for VILJ2006_MUOK

4 of 6 Specify totals

Columns: Grand total only

Rows: None

Pages: Grand total only

Page totals:
 First
 Last

Label for totals: Total

Preview:

		paino	paino
		Mean	StdDev
sukupu	tuotsuun	999	999
	tuotsuun	999	999
sukupu	tuotsuun	999	999
	tuotsuun	999	999

Tässä tapauksessa voitaisiin laskea tunnuslukuja taulukkoon yli osaryhmien eli yli rivien ("Rows"). "None"-vaihtoehdolla on kuitenkin määritetty, että tunnuslukuja ei lasketa yli osaryhmien.

< Back Next > Finish Cancel Help

Viidennellä sivulla voitaisiin vielä haluttaessa tallentaa tunnusluvut SAS-datatiedostoon, ja kuudennella sivulla määrittää tulostuksen otsikoita ja alaviitteitä. Lopuksi painetaan "Finish", jolloin saadaan tulostus, joka vastaa tavallisen "Summary Tables" -työkalun tulostusta samoilla asetuksilla.

Tehtävä: Tutki Viljelijä 2006 -aineistosta (**vilj2006_muok**) paino-muuttujan jakaumaa sukupuolen ja tuotantosuunnan eri luokkakombinaatioissa. Laske tarvittavat tilastolliset tunnusluvut Summary Tables wizard -työkalulla.

9.4 Kategorisen muuttujan jakauman tarkastelu

Kategorisen muuttujan jakaumaa voidaan tarkastella Enterprise Guidessa valitsemalla käytettävä data aktiiviseksi prosessidiagrammissa ja klikkaamalla "Describe" -> "One-Way Frequencies". Tarkasteltava kategorinen muuttuja tai kategoriset muuttujat laitetaan "Task Roles" -välilehdellä kohtaan "Analysis variables". Jos halutaan erilliset frekvenssitaulut jonkin toisen muuttujan tai toisten muuttujien eri luokissa, kyseinen muuttuja tai muuttujat laitetaan "Group analysis by" -kohtaan.

”Statistics”-välilehdellä määritellään frekvenssijakauman tarkasteluun liittyviä asetuksia:

Statistics

Frequency table options

Include:

- Frequencies and percentages with cumulatives
- Frequencies and cumulative frequencies
- Frequencies and percentages
- Frequencies only

Binomial proportions

- Asymptotic test
- Exact p-values

Test proportion: 0.5

Confidence level: 95%

Use Monte Carlo estimation

Missing values

- Show frequencies
- Include in calculations

Chi-square goodness of fit

- Asymptotic test
- Exact p-values

Valitaan, mitä tulostetaan kuvaamaan kategorisen muuttujan jakaamaa. Vaihtoehtona on erilaisia yhdistelmiä seuraavista: frekvenssit eli luokkien esiintymisten lukumäärät, kumulatiiviset frekvenssit, prosenttiosuudet ja kumulatiiviset prosenttiosuudet.

Puuttuvien arvojen käsittely. Jos valintoja ei tehdä, frekvenssitaulun alle tulee maininta puuttuvien havaintojen määrästä. ”Show frequencies” -valinnalla puuttuvista havainnoista tulee oma rivinsä varsinaiseen frekvenssitauluun. ”Include in calculations” -valinnalla puolestaan luokkien prosenttiosuuksien laskemisessa puuttuvia havaintoja pidetään omana luokkana ja puuttuvien osuudeksi lasketaan osa sadasta prosentista.

”Plots”-välilehdellä voidaan määrittellä tulostettavaksi pylväskuvio ja ”Results”-välilehdellä määrittellä, että frekvenssijakauma tulostetaan SAS-datatiedostoon. Samalla voidaan määrittellä, että ruudulle ei tulosteta mitään. Kun painetaan ”Run”, saadaan tulostus:

Tulevaisuudennäkymät:

- 1 = ”Hyvät”
- 2 = ”OK”
- 3 = ”Huonot”

tulevais	Frequency	Percent
1	19	35.85
2	22	41.51
3	12	22.64

System (Local, XP_PRO) on 14.2

One-Way Frequencies Plots

Tulostukseen oli valittu:

- 1) frekvenssijakauma
- 2) pylväskuvio

Tehtävä: Käytä Viljelijä 2006 -aineistoa (**vilj2006_muok**) ja tutki muuttujan ”tilan ajanmukaisuus” (=ajanmuk) jakaumaa.

9.5 Kahden muuttujan yhteyden tarkastelu

Kahden muuttujan yhteyttä on jo tarkasteltu aiemmin sirontakuvioiden avulla, mutta nyt otetaan käyttöön kuvien lisäksi tilastolliset tunnusluvut.

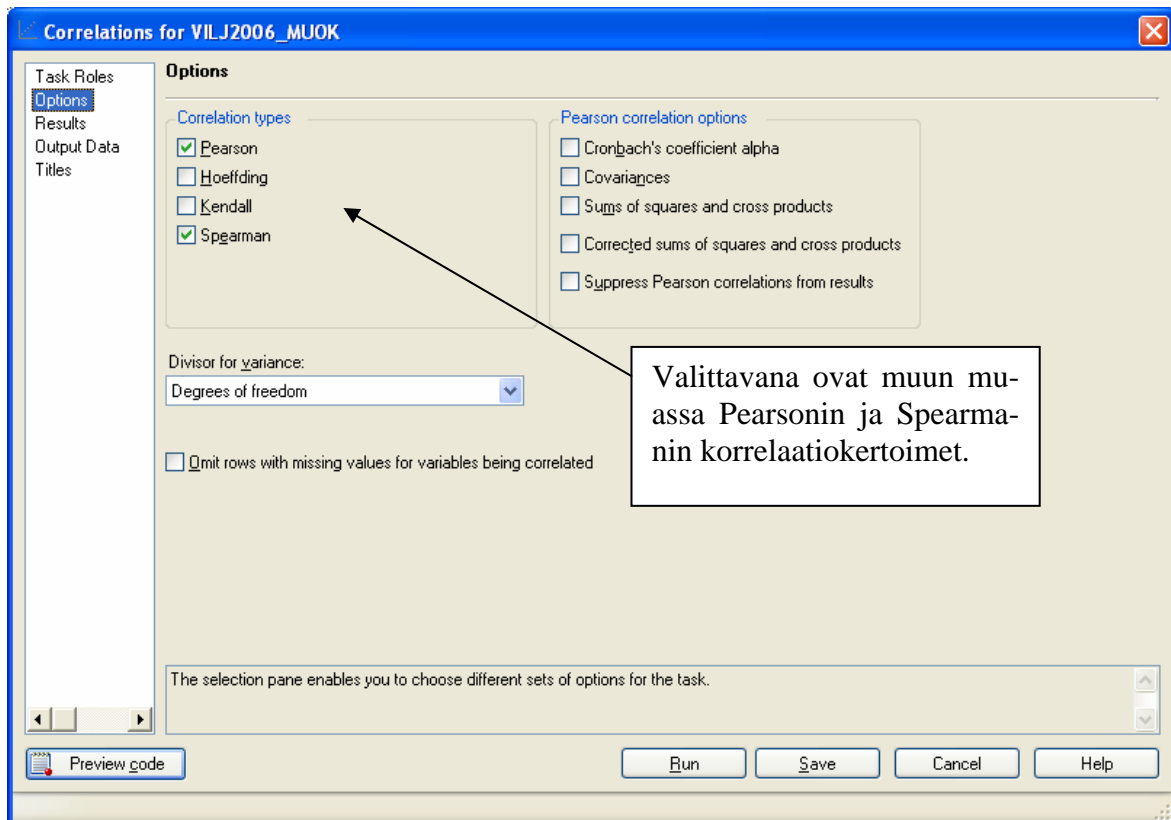
Tarkasteltaessa muuttujien välisiä yhteyksiä seuraavassa esitettävien menetelmien avulla, on hyvin oleellista että kaikkia havaintoja voidaan pitää riippumattomina. Jos havaintojen välillä esiintyy riippuvuuksia, muuttujien yhteyksien tarkastelut tulee tehdä osaryhmittäin siten, että havainnot ovat osaryhmän sisällä riippumattomia. Esimerkkinä toisistaan riippuvista havainnoista on tilanne, jossa kymmeneltä lehmältä on mitattu sama ominaisuus ensin tammikuussa ja sitten uudelleen helmikuussa. Näiden kahdenkymmenen havainnon käsittely riippumattomina olisi virheellistä, koska molempina ajankohtina mittaukset on tehty samoista eläimistä. Niinpä yhteyden tarkastelu muuttujien ”X” ja ”Y” välillä tulee tehdä erikseen kummallekin kuukaudelle, koska kuukauden sisällä havainnot ovat peräisin eri eläimistä ja sen vuoksi niitä yleensä voidaan pitää riippumattomina. On olemassa tilastollisia menetelmiä, joiden avulla voidaan ottaa huomioon riippuvuudet ja siten analysoida koko aineisto yhdessä. Kyseisiä menetelmiä ei kuitenkaan käsitellä tässä oppaassa.

9.5.1 Kahden numeerisen muuttujan välinen yhteys

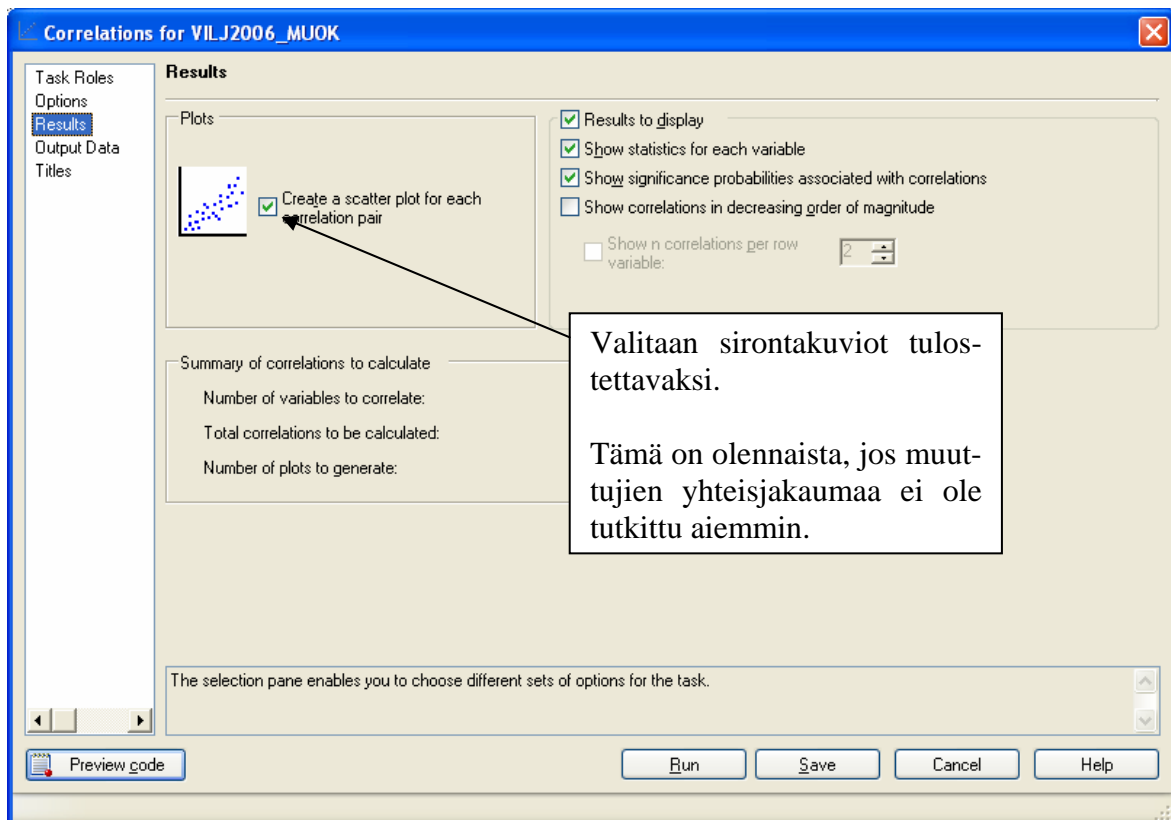
Kahden numeerisen muuttujan välistä yhteyttä voidaan luonnehtia korrelaatiokertoimen avulla. Tavallinen eli Pearsonin korrelaatiokerroin on kahden muuttujan lineaarisen yhteyden voimakkuuden ja suunnan mitta. Oletuksena Pearsonin korrelaatiokertoimen käytölle on se, että tarkasteltavien muuttujien yhteisjakauma on kaksiulotteinen normaalijakauma eli käytännössä yhteyden lineaarisuus. Oletuksen voimassaoloa on kätevä tarkastella sirontakuvioiden avulla. Sirontakuviossa pisteparven tulee asettua ympyrän tai litistyvän ellipsin muotoon, jotta kaksiulotteinen normaalijakaumaoletus olisi voimassa. Pitää muistaa myös havaintojen riippumattomuusoletus eli sirontakuviot tulee piirtää ja korrelaatiokertoimet laskea erikseen riippumattomien havaintojen muodostamille osaryhmille. Jos kaksiulotteinen normaalijakaumaoletus ei ole realistinen, voidaan käyttää epäparametrista Spearmanin järjestyskorrelaatiokerrointa, joka soveltuu myös järjestysasteikollisten muuttujien yhteyden tutkimiseen. Riippumattomuusoletuksen lisäksi korrelaatiokertoimia käytettäessä on kuvasta syytä tarkastella poikkeavia havaintoja, jotka saattavat vaikuttaa merkittävästi korrelaatiokertoimen suuruuteen.

Korrelaatiokertoimet saadaan Enterprise Guidesta valitsemalla käytettävä data aktiiviseksi prosessidiagrammissa ja klikkaamalla ”**Analyze**” -> ”**Multivariate**” -> ”**Correlations**”, jossa ”Task Roles” -välilehdellä määritellään muuttujien tehtävät. ”Analysis variables” -kohtaan laitetaan muuttujat, joiden välistä yhteyttä tutkitaan. Vaihtoehtoisesti voidaan asettaa joku muuttuja ”Analysis variables” -kohtaan ja toiset ”Correlate with” -kohtaan. Muuttujat, jotka määrittelevät riippumattomia havaintoja sisältävät osaryhmät, laitetaan kohtaan ”Group analysis by”.

”Options”-välilehdellä puolestaan määritellään käytettävät korrelaatiokertoimet ja tehdään niihin liittyviä lisämäärittäyksiä:

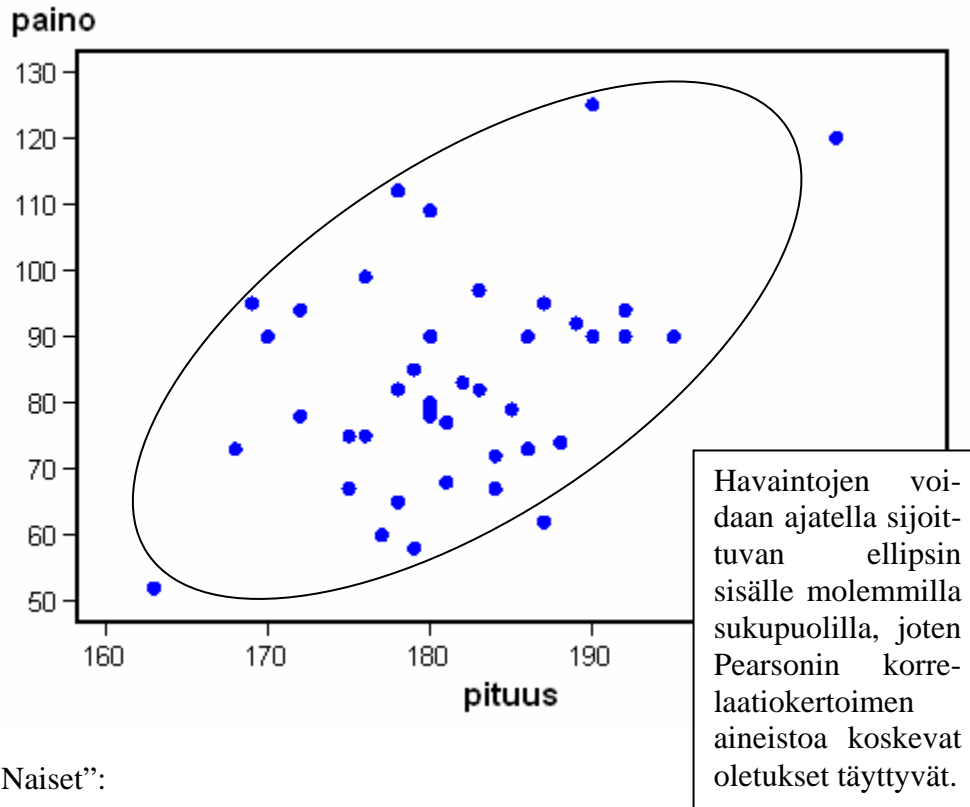


”Results”-välilehdellä voidaan valita muun muassa sirontakuviot mukaan tulostukseen:

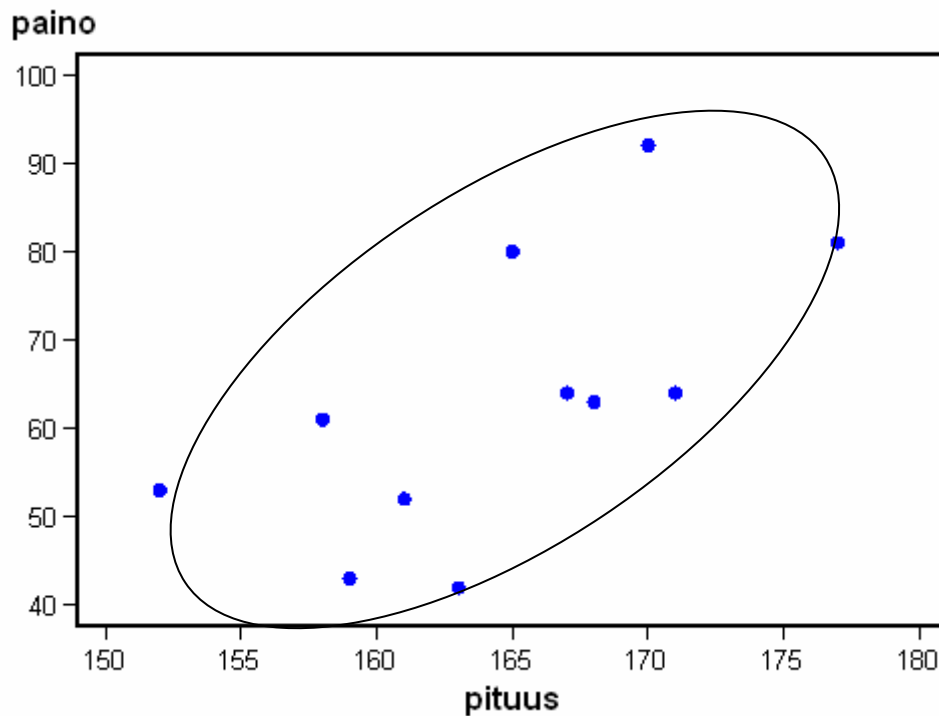


”Output Data” -välilehdellä voidaan valita tulokset tulostettavaksi SAS-dataksi. Kun painetaan lopuksi ”Run”, saadaan tuloksena sukupuolittain seuraavat sirontakuviot (kopioitu leikepöydän kautta Wordiin):

”Miehet”:



”Naiset”:



Korrelaatiokerroin miehille:

Pearson Correlation Coefficients, N = 42 Prob > r under H0: Rho=0		
	paino	pituus
paino paino	1.00000	0.40783 0.0073
pituus pituus	0.40783 0.0073	1.00000

Korrelaatiokerroin naisille:

Pearson Correlation Coefficients, N = 11 Prob > r under H0: Rho=0		
	paino	pituus
paino paino	1.00000	0.64712 0.0314
pituus pituus	0.64712 0.0314	1.00000

Painon ja pituuden välillä näyttää odotetusti olevan positiivinen lineaarinen yhteys kummallakin sukupuolella. Yhteyden suunta (positiivinen tai negatiivinen) nähdään korrelaatiokertoimen etumerkistä. Luku korrelaatiokertoimen alla taulukossa on p-arvo testille, jonka nollahypoteesi on, että korrelaatiokerroin on nolla. Pieni p-arvo tarkoittaa, että korrelaatiokerroin poikkeaa tilastollisesti merkitsevästi nolasta. Yllä olevista tuloksista näkee, miten paljon havaintomäärä vaikuttaa p-arvoon. Miesten ryhmällä korrelaatiokerroin on pienempi kuin naisilla, mutta silti p-arvo on selvästi pienempi. Tämä johtuu siitä, että miehillä havaintoja on selvästi enemmän. Korrelaatiokerroin vaihtelee miinus yhdestä plus yhteen. Korrelaatiokertoimen ollessa -1 kaikki havainnot ovat laskevalla suoralla ja kertoimen ollessa 1 havainnot ovat nousevalla suoralla. Korrelaatiokertoimen itseisarvo pienenee sitä mukaa, kun havaintojen poikkeama suoralta kasvaa. Kertoimen suuruuden käytännön merkittävyyden tulkintaan ei ole olemassa selkeätä sääntöä. Tulkinta vaihtelee ilmiön mukaan, ja eri asioissa käytetäänkin erisuuruksia raja-arvoja merkittävälle korrelaatiolle.

Jos haluttaisiin esimerkiksi tietää, kuinka paljon kymmenen sentin lisäys pituudessa keskimäärin nostaa painoa, vastaus saataisiin sovittamalla aineistoon suora ja määräämällä sen kulmakerroin. Tähän asiaan ei siis saada vastausta korrelaatiokertoimen avulla. Myös suoran ja yleisemmin regressiomallien sovitus onnistuu Enterprise Guidella. Lisätietoa asiasta löytyy muun muassa kirjasta Helenius & Katajisto (2005).

Tehtävä: Käytä Viljelijä 2006 -aineistoa (**vilj2006_muok**) tutkiaksesi painon ja pituuden yhteyttä sukupuolittain.

9.5.2 Kahden kategorisen muuttujan välinen yhteys

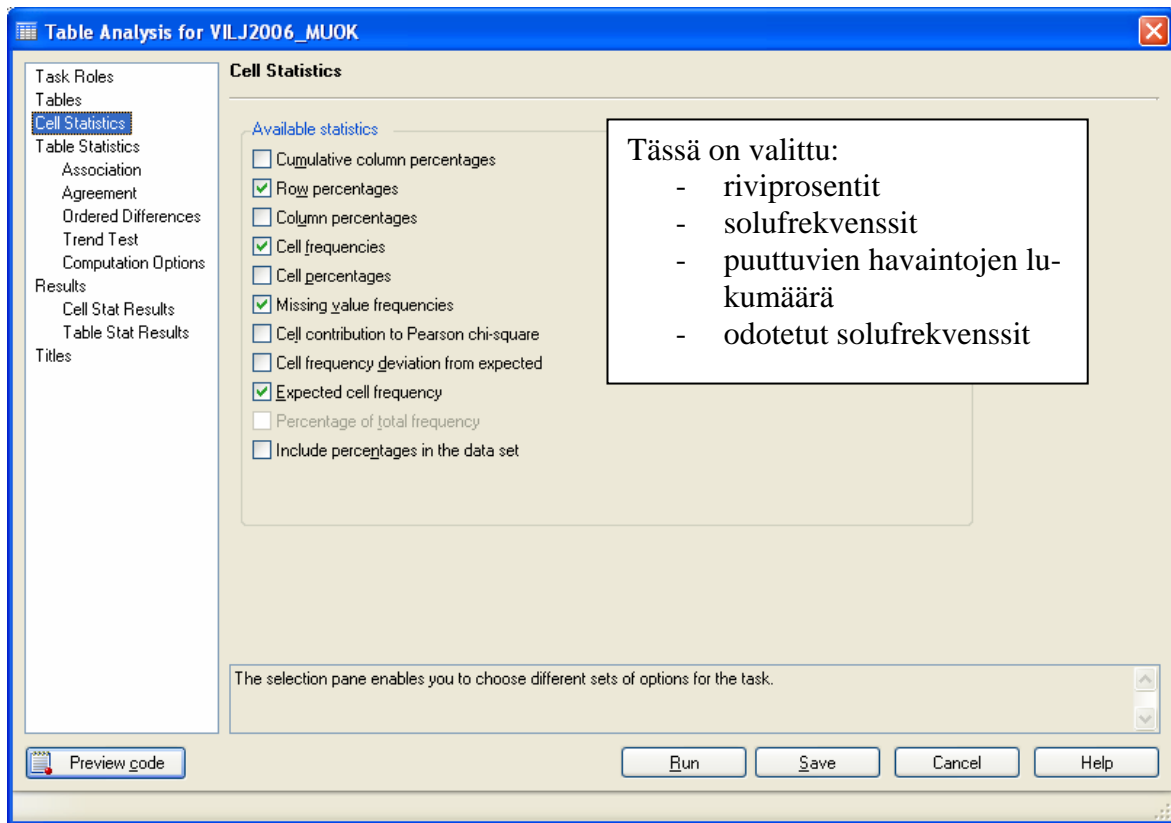
Kahden kategorisen muuttujan välistä yhteyttä voidaan tarkastella muun muassa ristiintaulukoinnin ja Khiin neliö -testin avulla. Samoin kuin korrelaatiokertoimien määrittäminen Khiin neliö -testi edellyttää kaikkien havaintojen riippumattomuutta. Lisäksi Khiin neliö -testin vaatimukset otoskoon suhteen ovat seuraavat: 1) korkeintaan 20 % odotetuista solufrekvensseistä saa olla pienempiä kuin viisi ja 2) kaikkien odotettujen solufrekvenssien pitää olla suurempia kuin yksi. Jos oletukset otoskoon suhteen eivät ole aineistossa voimassa, EG tulostaa automaattisesti varoituksen tulostusikkunaan. Tällaisessa tilanteessa voidaan yhdistää luokkia tai korvata suurten otosten Khiin neliö -testi niin sanotulla tarkalla (exact) testillä.

SAS Enterprise Guiden työkalu ristiintaulukointiin ja Khiin neliö -testin tekemiseen löytyy klikkaamalla ”Describe” -> ”Table Analysis”, jolloin ensin aukeaa ”Task Roles” -välilehti. Kategoriset muuttujat, joiden yhteyttä ollaan tutkimassa, sijoitetaan kohtaan ”Table variables” ja mahdollinen aineiston osaryhmiin jakava muuttuja sijoitetaan kohtaan ”Group analysis by”. ”Tables”-välilehdellä tehdään tarkempi taulukon määrittely:

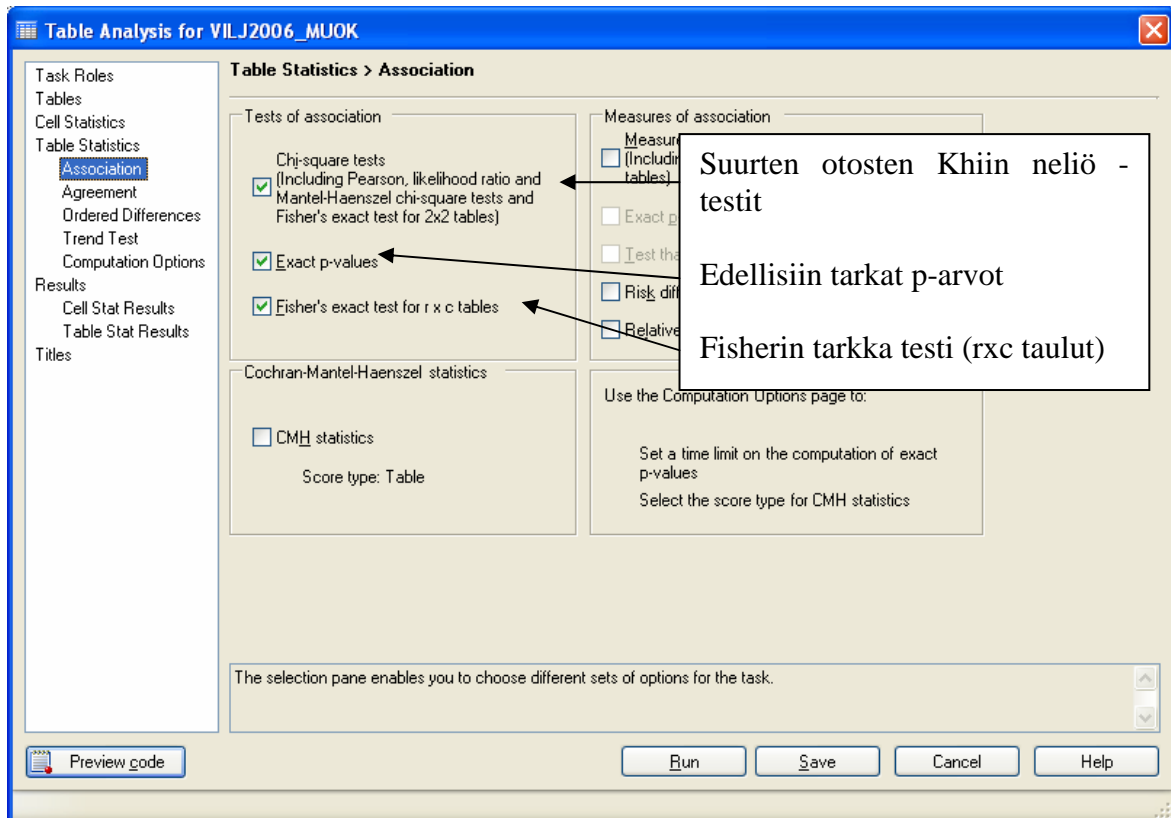
Tästä klikkaamalla voit määrittää uuden taulun.

Muuttujat raahataan hiirellä paikalleen vasemmalla olevasta laatikosta tähän taulukkoon. ”Table variables” -muuttujia voi olla kolmekin. Tällöin kolmannen muuttujan tasolle tulevat erilliset taulut, samanlaiset kuin käytämällä yhtä muuttujista kohdassa ”Group analysis by”. Jos kuitenkin halutaan tutkia kolmen muuttujan yhteyttä tilastollisella testillä, kaikkien muuttujien pitää olla ”Table variables” -kohdassa.

Ristiintaulukon soluihin tulostettavat luvut valitaan ”Cell Statistics” -välilehdellä:



Erilaisia tilastollisia testejä voidaan valita ”Association”-välilehdellä:



”Resultsin” alla olevilla välilehdillä puolestaan voidaan valita ristiintaulukko ja tilastollisten testien tulokset ajettavaksi SAS-dataksi. Painamalla ”Run” saadaan tulostus:

Ristiintaulukko

Table of ajanmuk by tulevais				
ajanmuk(ajanmuk)	tulevais(tulevais)			Total
Frequency Expected Row Pct	1	2	3	
1	6 3.2264 66.67	3 3.7358 33.33	0 2.0377 0.00	9
2	10 4.6604 76.92	3 5.3962 23.08	0 2.9434 0.00	13
3	1 4.6604 7.69	9 5.3962 69.23	3 2.9434 23.08	13
4	1 5.0189 7.14	6 5.8113 42.86	7 3.1698 50.00	14
5	1 1.434 25.00	1 1.6604 25.00	2 0.9057 50.00	4
Total	19	22	12	53

Suurten otosten
Khiin neliö -testit.

Statistics for Table of ajanmuk by tulevais

	DF	Value	Prob
Chi-Square	8	29.5438	0.0003
Likelihood Ratio Chi-Square	8	33.8016	<.0001
Mantel-Haenszel Chi-Square	1	18.5257	<.0001
Phi Coefficient		0.7466	
Contingency Coefficient		0.5983	
Cramer's V		0.5279	
WARNING: 73% of the cells have expected counts less than 5. (Asymptotic) Chi-Square may not be a valid test.			

Suurten otosten testien otoskoko-oletukset eivät ole tässä voimassa.

Pearson Chi-Square Test	
Chi-Square	29.5438
DF	8
Asymptotic Pr > ChiSq	0.0003
Exact Pr >= ChiSq	1.190E-04

Fisher's Exact Test	
Table Probability (P)	3.035E-10
Pr <= P	6.012E-05

Tarkat testit

Tehtävä: Tutki Viljelijä 2006 -aineistosta (**vilj2006_muok**) tilan ajanmukaisuuden (ajanmuk) ja tulevaisuudennäkymien (tulevais) yhteyttä ristiintaulukoinnin ja testien avulla.

10 SAS-koodin käyttäminen

Erityisesti kokeneet SAS-käyttäjät saattavat kokea hyödylliseksi mahdollisuuden käyttää SAS-koodia Enterprise Guidessa valikkotoimintojen lisäksi. Myös uusille käyttäjille joidenkin toimintojen suorittaminen voi olla kätevintä koodin avulla. Enterprise Guiden valikkokäytössä on myös omat rajoituksensa, joten tavallisen SASin monipuolisempien ominaisuuksien hyödyntäminen Enterprise Guidessa voi edellyttää koodin käyttöä.

Koodin käyttöön liittyviä asetuksia voidaan muokata valitsemalla ”Tools” -> ”Options” -> ”SAS Programs”.

10.1 Koodin kirjoittaminen Enterprise Guidessa tai tuominen muualta

Koodin kirjoittaminen onnistuu EG:ssä koodi-ikkunassa. Uuden koodi-ikkunan saa avattua valitsemalla ”File” -> ”New” -> ”Code”. Koodi-ikkunat tulevat projektiin irrallisiksi ikkunoiksi, joissa esimerkiksi datoihin viitataan kirjastoviitteen avulla samoin kuin tavallisessa SASissa. Koodi-ikkuna tallentuu automaattisesti projektin sisälle, mutta jos se halutaan tallentaa erillisenä tiedostona, tallennus onnistuu klikkaamalla hiiren oikealla painikkeella koodisolmua prosessidiagrammissa ja valitsemalla ”Properties”. Avautuvassa ikkunassa valitaan ”Save As”.

Valmis, esimerkiksi tavallisessa SASissa tehty kooditiedosto (*.sas) voidaan myös avata Enterprise Guideen. Tämä onnistuu valitsemalla ”File” -> ”Open” -> ”Code” -> ”Local Computer” ja valitsemalla SAS-kooditiedosto tietokoneen jostakin hakemistosta. Kooditiedosto ei kuitenkaan automaattisesti tallennu projektin sisälle, vaan tietokoneen kansiossa olevaan kooditiedostoon avataan näkymä (pikakuvake). Jos koodi halutaan sisällyttää projektiin, voidaan klikata koodisolmua hiiren oikealla painikkeella ja valita ”Properties”. Avautuvassa ikkunassa valitaan ”Embed”.

SAS-koodia voi myös kätevästi liikutella leikepöydän kautta esimerkiksi tavallisen SASin, EG:n ja Wordin välillä. Tämä onnistuu maalaamalla ensin kopioitava koodinpätkä hiirellä ja valitsemalla ”Edit”/”Muokkaa” -> ”Copy”/”Kopioi”, minkä jälkeen mennään kohdeohjelmaan ja valitaan siellä ”Edit”/”Muokkaa” -> ”Paste”/”Liitä”.

Jos Enterprise Guidessa tehdään kuvia koodin avulla, saattaa vastaan tulla tilanne, jossa koodissa tehtävät asetukset eivät välttämättä vaikuta valmiiseen kuvaan. Tähän saattaa auttaa kuvatyyppin vaihtaminen joksikin muuksi kuin oletuksena oleva ActiveX. Esimerkiksi symbol-lausetta käytettäessä ActiveX ohittaa koodissa annetut määrittelyt.

10.2 Koodin generoiminen Enterprise Guidessa

Kaikki Enterprise Guiden työkalut luovat koodia samalla, kun hiirellä tehdään valintoja. Enterprise Guide lähettää automaattisesti koodit SASin suoritettavaksi, ja tulokset näytetään Enterprise Guidessa. Kaikki toiminnot EG:ssä perustuvat siis pohjimmiltaan SAS-koodeihin, joita voidaan tarkastella ja muokata omiin tarkoituksiin.

Valmiista tehtäväsolmusta saa sen luoman SAS-koodin tarkasteltavaksi klikkaamalla solmua hiiren oikealla painikkeella ja valitsemalla ”Open Last Submitted Code”. Tällöin ruudulle aukeaa SASista tuttu koodieditori, josta löytyy tehtävän käyttämä SAS-koodi. Jos koodia yrittää muokata, EG ilmoittaa, että koodi on suojattu ja editointia varten pitää tehdä

kopio koodista. ”Kyllä”-valinnalla koodi aukeaa uuteen editori-ikkunaan, jossa sitä voidaan muokata. Vaihtoehtoisesti voidaan valita ”Open Last Submitted Code” -vaihtoehdon sijaan ”Add as Code Template”, jolloin EG tekee koodista suoraan muokattavan kopion. Tämän ominaisuuden avulla voidaan esimerkiksi luoda koodipohjia ja käyttää niitä tavallisessa SASissa.

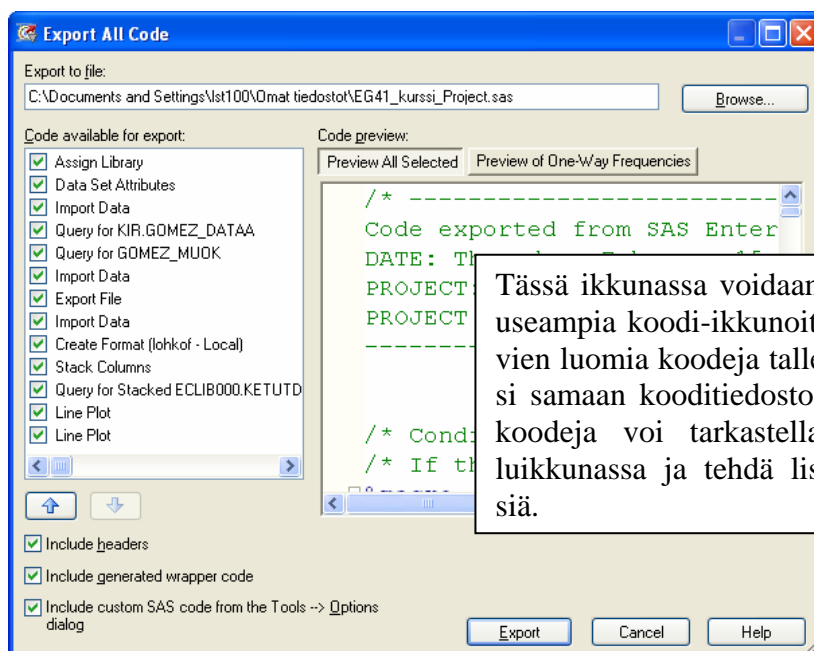
Koodia voidaan myös tarkastella tehtävän määrittelyn ollessa vielä kesken. Useimmissa tehtävämäärittelyikkunoissa löytyy vasemmasta alakulmasta ruutu ”Preview code”. Tätä klikkaamalla saadaan siihen asti tehdyillä valinnoilla generoitu koodi esikatseluikkunaan. Klikkaamalla ”Insert code” pääsee koodin muokkaustilaan. EG tarjoaa mahdollisuuden lisätä tiettyihin kohtiin omia koodinpätkiä. Koko koodia ei kuitenkaan pääse tällä keinolla muokkaamaan.

Koodin tarkastelu jälkikäteen tai esikatselun avulla on käyttökelpoista myös silloin, jos halutaan varmistaa, millä menetelmillä EG toteuttaa jonkin toimenpiteen. Koodista voi tarkistaa EG:n käyttämän menetelmän tai valinnan (eli avainsanan koodissa) ja hakea ”tavallisen” SASin manuaalista tarkemman selityksen kyseiselle asialle. Tähän manuaaliin pääsee käsiksi muun muassa Enterprise Guiden ohjeen haku-toiminnolla.

10.3 Koodin tallentaminen SAS-kooditiedostoon

SAS-kooditiedosto voidaan tallentaa tietokoneelle omana tiedostonaan klikkaamalla koodisolmua hiiren oikealla painikkeella ja valitsemalla ”Export” -> ”Export Code For *tehtävän_nimi*” tai ”Save Code For *tehtävän_nimi* As”. Erona näiden tallennustapojen välillä on, että ”Export”-tavalla projektiin sisällytetty koodisolmu (embedded) säilyy sisällytettynä tallennuksen jälkeenkin, kun taas ”Save As” -tavalla koodisolmu muuttuu viittaukseksi tietokoneen kansioon tallennettuun kooditiedostoon. Jos EG:ssä luotua koodia ei tallenneta erikseen, se tallentuu kuitenkin projektin mukana. Silloin siitä ei kuitenkaan tule erillistä SAS-kooditiedostoa.

Jos halutaan tallentaa samanaikaisesti useamman koodi-ikkunan sisältämä tai tehtävien luoma koodi samaan SAS-kooditiedostoon, se onnistuu valitsemalla ”File” -> ”Export” -> ”Export All Code”. Tällöin aukeaa seuraava ikkuna:



Tässä ikkunassa voidaan siis valita useampia koodi-ikkunoita tai tehtävien luomia koodeja tallennettavaksi samaan kooditiedostoon. Lisäksi koodeja voi tarkastella esikatseluikkunassa ja tehdä lisämäärittäyksiä.

11 Tulostuksen ja projektin hallintaa

Tässä kappaleessa on esitetty yleisiä asioita, jotka pätevät Enterprise Guideen ja sen monein eri työkaluihin. Esitystapa on hieman erilainen kuin aiemmissa kappaleissa. Asiat on pyritty ilmaisemaan hieman tiiviimmin ja listamaisemmin kuin aiemmin, jotta kappaleesta olisi helppo hakea ratkaisuja Enterprise Guiden yleiskäytössä ilmeneviin ongelmiin.

11.1 Tulostusasetukset

Monia asioita, mukaan lukien erilaisia tulostukseen liittyviä määrittämiä, voidaan EG:ssä säädellä asetusvalikosta. Asetuksiin pääsee käsiksi valitsemalla **"Tools"** -> **"Options"**. Asetukset voidaan palauttaa oletusasetuksiksi klikkaamalla **"Tools"** -> **"Options"** ja **"Reset All"**.

11.1.1 Tulostuskohteen valinta

Tulostuskohteella tarkoitetaan tulostusdokumentin tiedostomuotoa. Tulostus voidaan tehdä HTML-, RTF-, PDF-, SAS Report- ja Text-muodossa. HTML on tiedosto, joka voidaan avata internet-selaimella. RTF aukeaa Wordissa, PDF Acrobat Readerilla ja Text on tekstimuotoista tulostusta, jonka voi avata vaikka Notepadilla (Muistio). Yksittäisistä SAS Report -tulosteista puolestaan voi koota oman SAS raportin ("File" -> "New" -> "Report"). RTF:llä tehtyä tulostetta on melko kätevä muokata Wordissa. PDF-tulostetta sen sijaan ei pääse kovin helposti muokkaamaan, mikä saattaa olla etu joissakin tilanteissa. Kaikki tulostuskohteet aukeavat tietysti Enterprise Guidessa. Tulostuskohdetta pääsee vaihtamaan klikkaamalla **"Tools"** -> **"Options"** -> **"Results General"**. Tulostus voidaan ohjata yhteen tai useampaan kohteeseen. Ohjelma, jolla tulostus avataan, voidaan puolestaan valita "Viewer"-välilehdeltä.

Kuvien tulostus ohjautuu samalla tavalla johonkin edellä mainituista tulostuskohteista (paitsi "Text", joka ei käy kuville), mutta lisäksi voidaan valita kuvatyyppi. Kuva siis tallentuu jollain kuvatyypillä jonkin edellä mainitun tallennuskohteen tyyppiseen tiedostoon. Kuvatyyppejä pääsee vaihtamaan valitsemalla **"Tools"** -> **"Options"** -> **"Graph"**, jolloin **"Graph Format"** -valikosta voidaan valita joku seuraavista tyypeistä:

- ActiveX (oletus)
- Java
- GIF
- JPEG
- ActiveX image (SAS V9)
- Java image (SAS V9)
- SAS EMF (SAS V9)

Yleensä kannattaa käyttää oletuksena olevaa ActiveX-kuvatyyppeä, jolloin käytössä ovat monipuoliset kuvan muokkausominaisuudet. Esimerkiksi paperille tulostettaessa myös ActiveX image (SASV9) -kuvatyyppi on kuitenkin käyttökelpoinen.

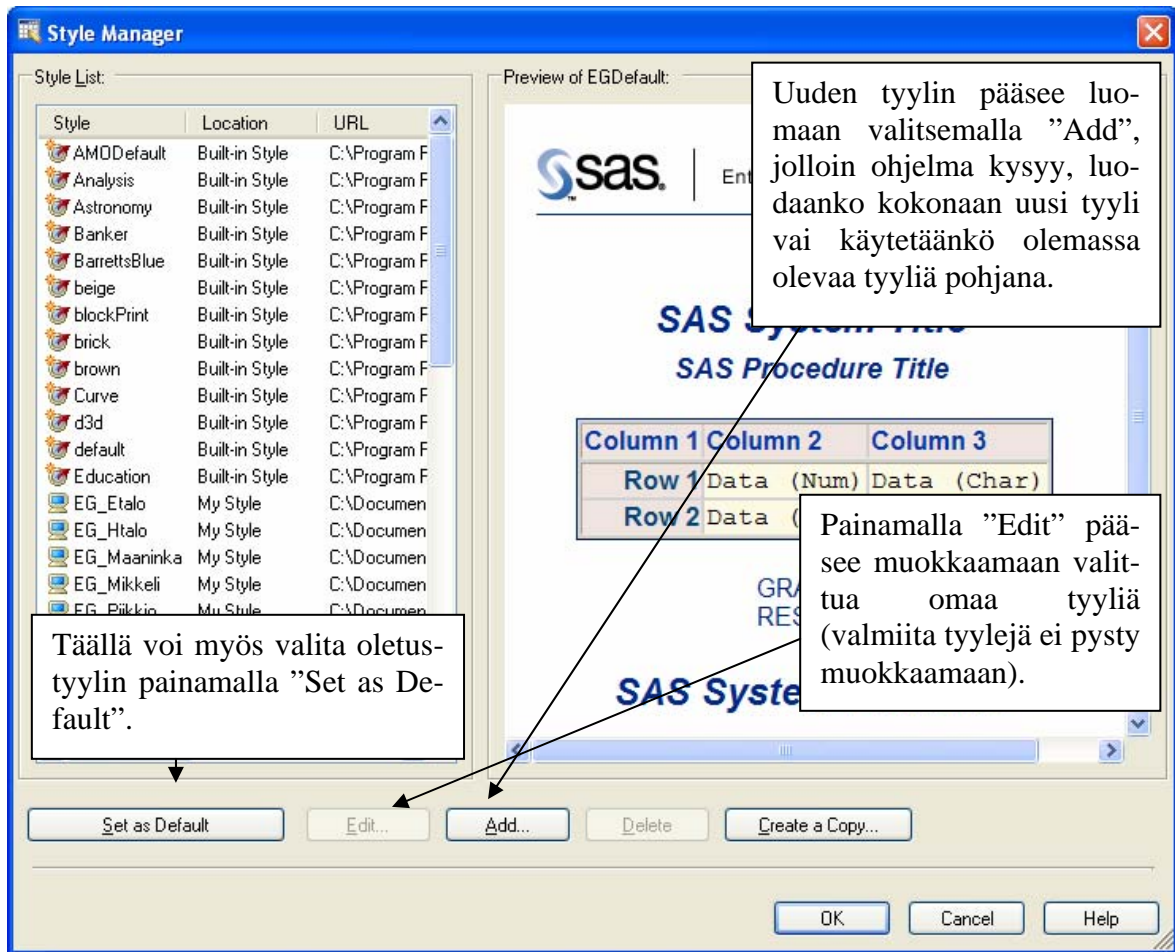
11.1.2 Tulostustyylin valinta

Tulostustyyllillä tarkoitetaan tulosteen taulujen ja taustan graafista ulkoasua. Tulostustyyllillä voidaan vaikuttaa tulosteen väreihin, fontteihin ja aseteluun. Tulostustyyli perustuu aina johonkin tyyliin, joita EG:stä löytyy useita. Oman tyyliin voi tehdä helpoiten muokkaamalla jotakin EG:n valmista pohjaa. Muokatulla tyyliin pohjalla voi luoda kuvia

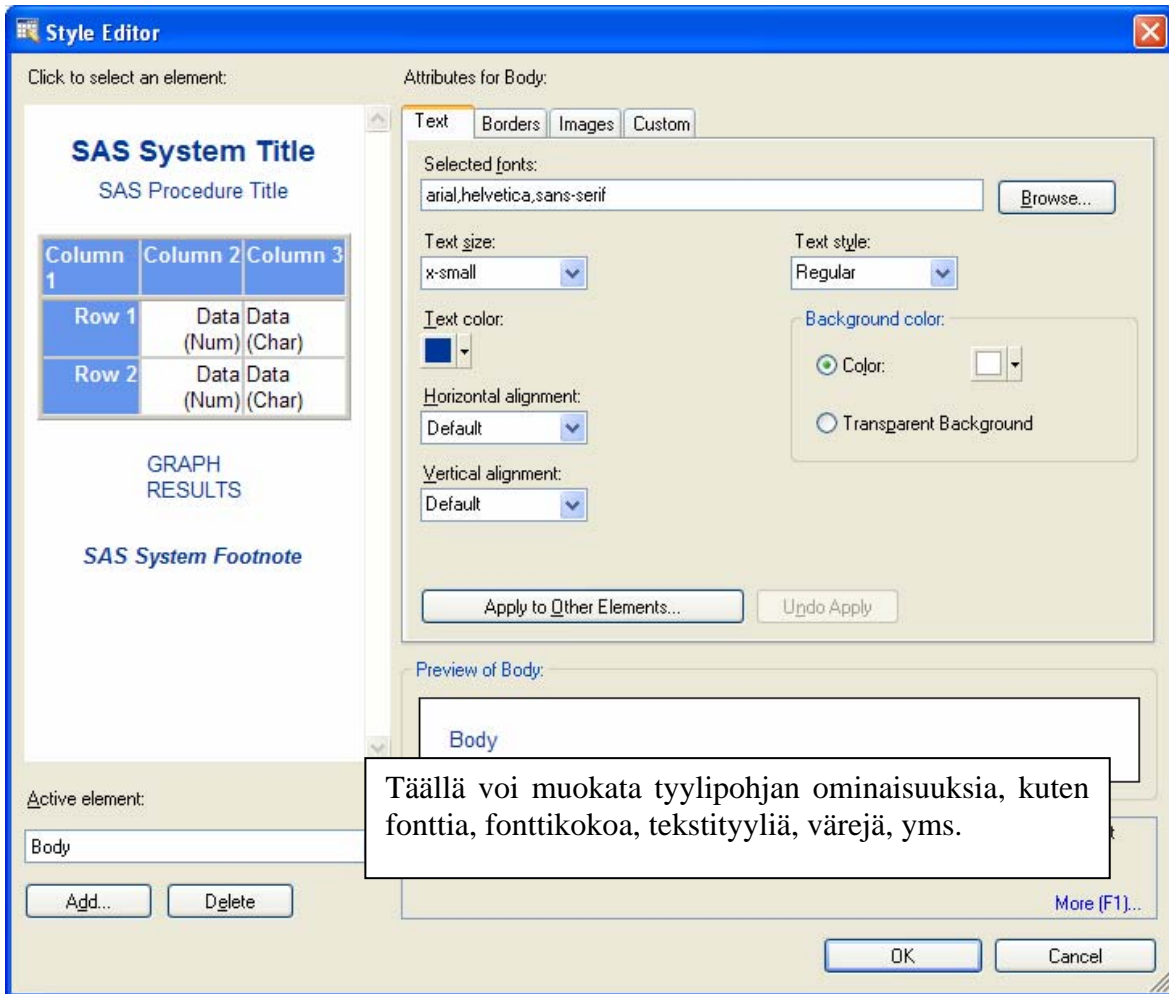
tai taulukoita, jotka ovat ulkoasun puolesta sellaisenaan kelvollisia esimerkiksi julkaisuihin.

Oletuksena olevan tulostustyylin voi vaihtaa toiseksi valitsemalla ”Tools” -> ”Options” ja käytössä olevan tulostuskohteen mukaan ”HTML”-, ”RTF”-, ”PDF”- tai ”SAS Report”-välilehden. ”Text”-tulostuskohteen tyyliä sen sijaan ei pysty muokkaamaan. Kaikista näistä välilehdistä löytyy kohta ”Style”, josta tulostustyylin voi vaihtaa. Hyviä vaihtoehtoisia tulostustyyliä (mustavalkoisia) ovat ainakin ”printer”, ”journal” ja ”minimal”.

Oman tyylin pääsee luomaan valitsemalla ”Tools” -> ”Style Manager”, jolloin aukeaa seuraava ikkuna:



Varsinainen tyyli pohjan muokkaustyökalu (joka aukeaa ”Edit”-valinnalla) näyttää seuraavalta:



11.1.3 Muut tulostusasetukset

Tulostukseen tulee oletusasetuksilla aina alaviite: ”Generated by the SAS System...”. Tämän alaviitteen voi vaihtaa toiseksi tai poistaa kokonaan valitsemalla ”Tools” -> ”Options” -> ”Tasks” -> ”Tasks General” ja muuttamalla sieltä ”Default footnote text for task output” -kohtaa. Samalta välilehdeltä löytyy myös kohta, josta voi lisätä pysyvän otsikon tulostukseen.

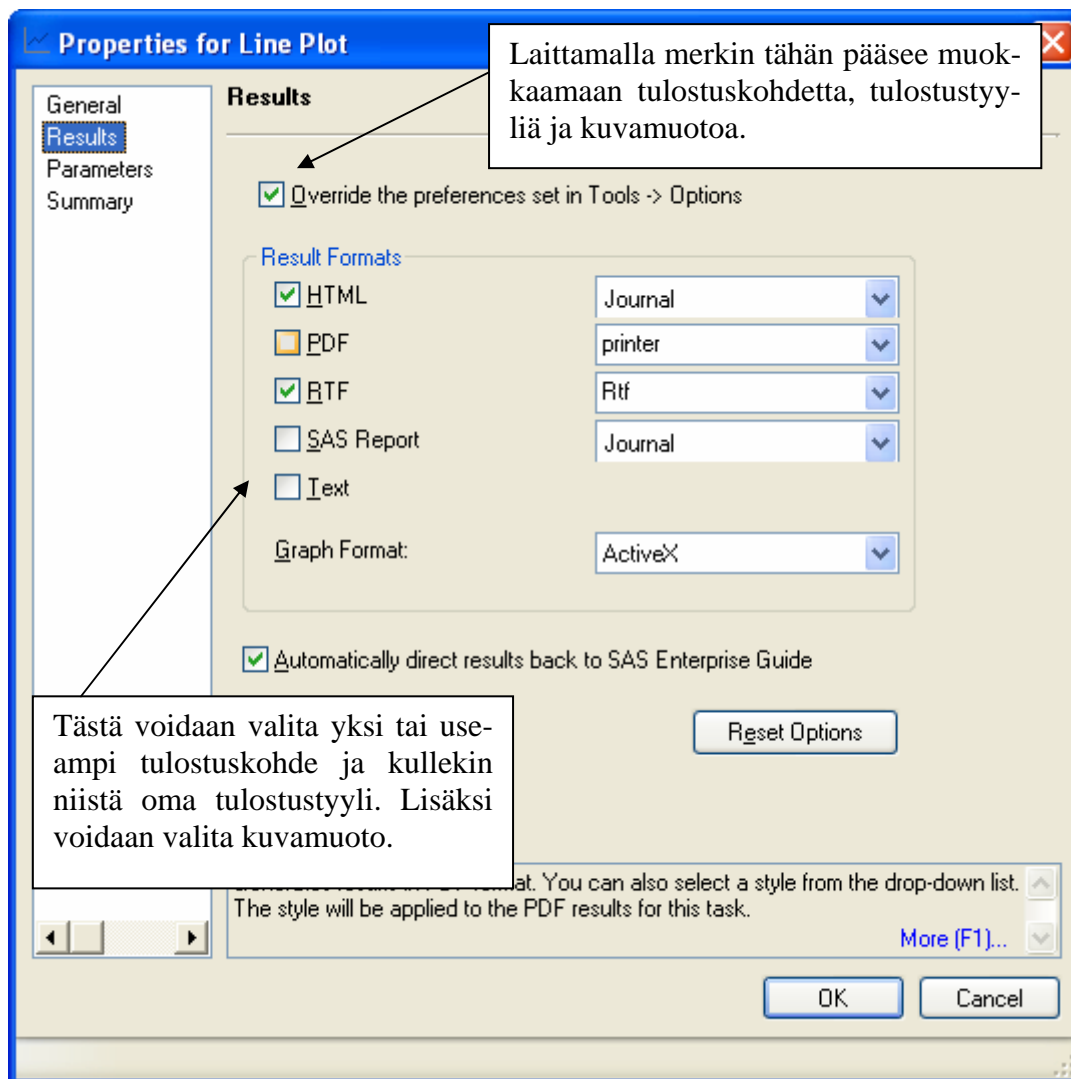
Työkalujen käyttämien SAS-proseduurien nimet saa poistettua tulostuksesta. Tämä onnistuu myös ”Tools” -> ”Options” -> ”Tasks” -> ”Tasks General” -välilehdellä, kun poistetaan merkintä kohdasta ”Include SAS procedure titles in results”.

11.2 Yksittäisen tulosteen hallinta ja jatkokäyttö

Enterprise Guidessa voidaan muuttaa yleisasetusten lisäksi monia yksittäisen tulosteen asetuksia. Tulosteita voidaan helposti myös muokata ja siirtää toisiin ohjelmiin. Seuraavassa tarkastellaan näitä asioita lähemmin.

11.2.1 Kohteen ja tyylin vaihtaminen yksittäiselle tulosteelle

Yksittäiselle tulosteelle voidaan vaihtaa tulostuskohdetta ja tulostustyyliä, kun tulostuksen tekävä tehtäväsolmu on jo olemassa prosessidiagrammissa. Asetuksia päästään muokkaamaan painamalla tulostuksen tuottavaa tehtäväsolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Properties” ja sieltä ”Results”:



Tämän jälkeen painetaan ”OK”. Tehtäväsolmu pitää suorittaa vielä uudelleen, jotta muutokset menevät tulosteeseen asti. Tämä onnistuu klikkaamalla tehtäväsolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Run *tehtävän nimi*”.

11.2.2 Yksittäisen tulosteen tallentaminen tiedostoon

Yksittäinen tuloste voidaan tallentaa tiedostoon klikkaamalla tulosteen kuvaketta hiiren oikealla painikkeella prosessidiagrammissa ja valitsemalla ”Export” -> ”Export *tulostusmuoto – työkalun nimi*” -> ”Local Computer”. Tällöin aukeaa ikkuna, josta voidaan valita tallennettavalle tulostetiedostolle kansio ja nimi. Eri tulostuskohteet tallentuvat pääasiassa seuraavan tyyppisinä tiedostoina:

HTML ->	HTML-tiedosto (*.html)
RTF ->	Rich Text Format -tiedosto (*.rtf)
PDF ->	Portable Document Format -tiedosto (*.pdf)
SAS Report ->	SAS Report -tiedosto (*.srx)
Text ->	List-tiedosto (*.lst) tai tekstitiedosto (*.txt)

Lisäksi on mahdollista tallentaa tuloste käyttäen muita tiedostotyyppisiä kuin yllämainitut. Esimerkiksi HTML-tulostus voidaan tallentaa Excel-tiedostoksi kirjoittamalla tiedoston nimeksi ”*nimi.xls*”. Tällöin tulostus on avattavissa Excelissä. Tämä on kuitenkin epävirallinen tapa, ja saattaa aiheuttaa ongelmia. Tulostus voi näyttää Exceliin avattuna erilaiselta kuin EG:ssä, ja myös desimaaliluvut voivat aiheuttaa ongelmia.

Klikattaessa tulostussolmua hiiren oikealla painikkeella ”Export”-kohdasta löytyy myös vaihtoehto ”Export *tulostusmuoto – työkalun nimi* As A Step In Project”. Tämän vaihtoehdon avulla tulosteen tallentaminen tiedostoksi saadaan pysyväksi vaiheeksi prosessidiagrammiin, jolloin se voidaan milloin tahansa toteuttaa helposti suorittamalla kyseinen ”Export File” -tehtäväsolmu.

Kuvan voi tallentaa joko edellä mainitulla tavalla koko tulosteen mukana tai yksittäisenä kuvatiedostona. Jälkimmäinen tapa onnistuu siten, että ensin avataan työtilaan tuloste, jossa kuva sijaitsee ja klikataan kuvan päällä hiiren oikealla painikkeella. Sen jälkeen valitaan ”Save as” ja aukeavasta ikkunasta tallennuskansio ja annetaan kuvatiedostolle nimi. Käytettävissä olevat tiedostomuodot määräytyvät sen mukaan, mikä on valittu kuvatyypiksi ”Tools” -> ”Options” -> ”Graph” -välilehdellä.

11.2.3 Tulosteen tai datan siirtäminen toiseen ohjelmaan

Ensimmäinen tapa siirtää tulostus toiseen ohjelmaan (esim. Wordiin), on avata EG:ssä tallennettu tiedosto (esim. *.rtf) kyseisessä ohjelmassa. Kuvatiedostoina tallennetut kuvat puolestaan voidaan lisätä Wordiin käyttäen Wordin toimintoa ”Lisää” -> ”Kuva” -> ”Tiedostosta”.

Toisaalta Enterprise Guiden tulostus voidaan joko ohjata lähtökohtaisesti toiseen ohjelmaan (”Tools” -> ”Options” -> ”Results” -> ”Viewer”) tai avata yksittäinen EG:ssä oleva tulostussolmu muulla ohjelmalla. Yksittäinen tulostussolmu voidaan avata toiseen ohjelmaan klikkaamalla tulostussolmua hiiren oikealla painikkeella ja valitsemalla ”Open With *ohjelman nimi*”. RTF-solmuissa tässä lukee yleensä (tietokoneen oletusohjelmista riippuen) ”Microsoft Office Word”, PDF-solmuissa ”Acrobat Reader” ja HTML-solmuissa ”Internet Explorer”.

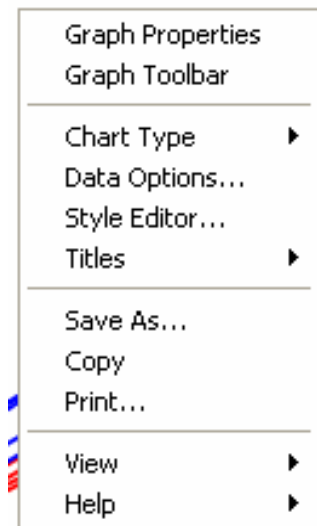
Nopein tapa siirtää dataa tai tulosteita toiseen ohjelmaan on klikata kyseistä solmua hiiren oikealla painikkeella ja valita ”Send To”, minkä jälkeen valitaan kohde. Yksi vaihtoehto on ”E-mail Recipient”, jolloin tuloste tai data menee suoraan uuden sähköpostiviestin liitetiedostoksi. Tämä toimii ainakin Outlook Expressin kanssa. ”Microsoft Word” ja ”Microsoft Excel” -vaihtoehdot puolestaan avaavat kohteen Wordissa tai Excelissa. Tämä ”Send To” -

toiminto toimii parhaiten datoilla. Tulosteilla kaikki kolme edellä mainittua vaihtoehtoa eivät ole aina valittavissa. Neljäntenä vaihtoehtona on ”E-mail Recipient as a Step in Project”, jonka avulla sähköpostin lähetys saadaan pysyväksi osaksi projektikokonaisuutta. Tätä vaihtoehtoa käytettäessä pitää sähköpostiasetusten olla määritettynä Enterprise Guidessa.

Yksittäisiä kuvia tai tulostustauluja voidaan myös siirtää toiseen ohjelmaan leikepöydän kautta (”copy-paste”). Tämä onnistuu joko klikkaamalla kohdetta hiiren oikealla painikkeella EG:ssä ja valitsemalla ”Copy” tai valitsemalla kohde aktiiviseksi ja klikkaamalla ”Edit -> ”Copy”. Sen jälkeen mennään toiseen ohjelmaan, jossa laitetaan kohdistin paikkaan, johon kopioitava objekti halutaan liittää, ja valitaan ”Muokkaa” -> ”Liitä” (tai ”Edit” -> ”Paste”). Erittäin käyttökelpoinen tapa siirtää tulostusta Exceliin onkin yksinkertaisesti kopioida RTF-muotoon tulostettu yksittäinen taulukko leikepöydän kautta Exceliin.

11.2.4 Valmiin kuvan muokkaaminen

EG:llä piirrettyä kuvaa voidaan muokata monin tavoin piirtämisen jälkeen. Valikko muokkaustoimenpiteistä saadaan, kun klikataan kuvaa hiiren oikealla painikkeella:



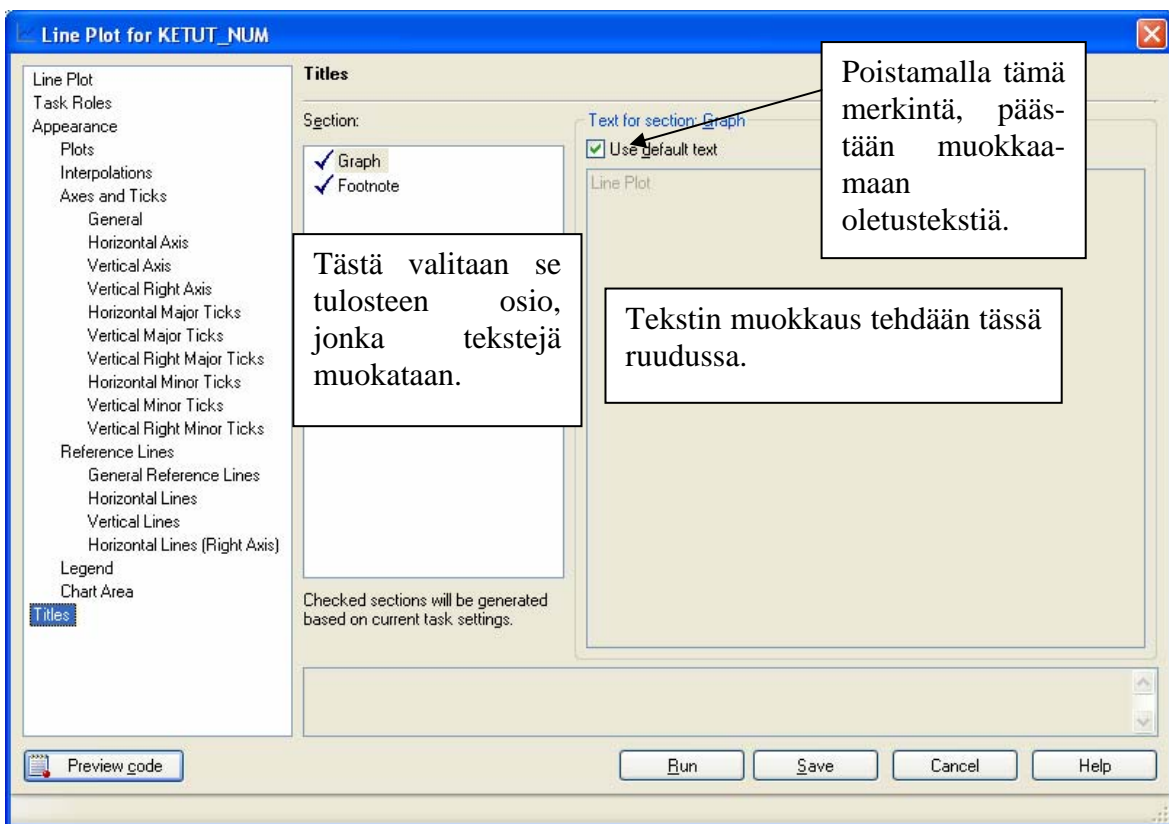
Suurin osa toimenpiteistä löytyy ”Graph Properties” -kohdan alta, mutta muualtakin saatavaa löytyä tutustumisen arvoisia toimenpiteitä. ”Graph Properties”-kohdasta voidaan muun muassa vaihtaa kuvan taustan tyyliä kohdassa ”Style”. Tällä voi tehdä esimerkiksi esityksiin näyttäviä kuvia. Oikealla painikkeella aukeava valikko voi hieman vaihdella kuvan mukaan. Kannattaa kuitenkin huomioida, että muokattaessa yllämainitun valikon kautta valmista kuvaa, muutokset eivät tallennu projektin mukana. Muokattu kuva kannattaa siis joko tallentaa omaksi kuvatiedostokseen (valitsemalla yllä olevassa kuvassa näkyvästä valikosta ”Save As”) tai kopioida leikepöydän kautta talteen esimerkiksi Wordiin (valitsemalla yllä olevassa kuvassa näkyvästä valikosta ”Copy”). Jos kuvatyypinä on ActiveX, yllä näkyvä valikko aukeaa myös Internet Explorerissa klikattaessa kuvaa hiiren oikealla painikkeella. ActiveX ja Java ovat kuvatyyppejä, jotka antavat käyttäjälle mahdollisuuden muokata kuvaa jälkikäteen (tarkastella sitä interaktiivisesti). Tällöin käyttäjä ei tarvitse edes EG:tä tai SASia. ActiveX-kuvaa voidaan editoida Internet Explorerissa koneella, jossa on käyttöjärjestelmänä Windows. Java-kuvaa voidaan puolestaan muokata kaikissa internet-selaimissa, jotka tukevat Javaa.

11.2.5 Selitetekstien ja formaattien lisääminen yksittäiselle tulosteelle

Tulosteessa esiintyville muuttujien nimille voidaan määrittää selitetekstejä ja muuttujien arvoille formaatteja lisäämättä niitä varsinaiseen käytettävään dataan. Tämä onnistuu tulosteen tuottavan tehtävän määrittelyssä. ”Task Roles” -välilehdellä klikataan käsiteltävän muuttujan nimeä hiiren oikealla painikkeella ja valitaan ”Properties”. Tällöin aukeaa ikkuna, jossa voidaan määrittää kyseiselle muuttujalle seliteteksti (”Label”) tai formaatti (”Format” -> ”Change”). Näin voidaan esimerkiksi formaatin avulla säädellä desimaalien määrää tulosteessa.

11.2.6 Otsikoiden ja alaviitteiden muokkaaminen

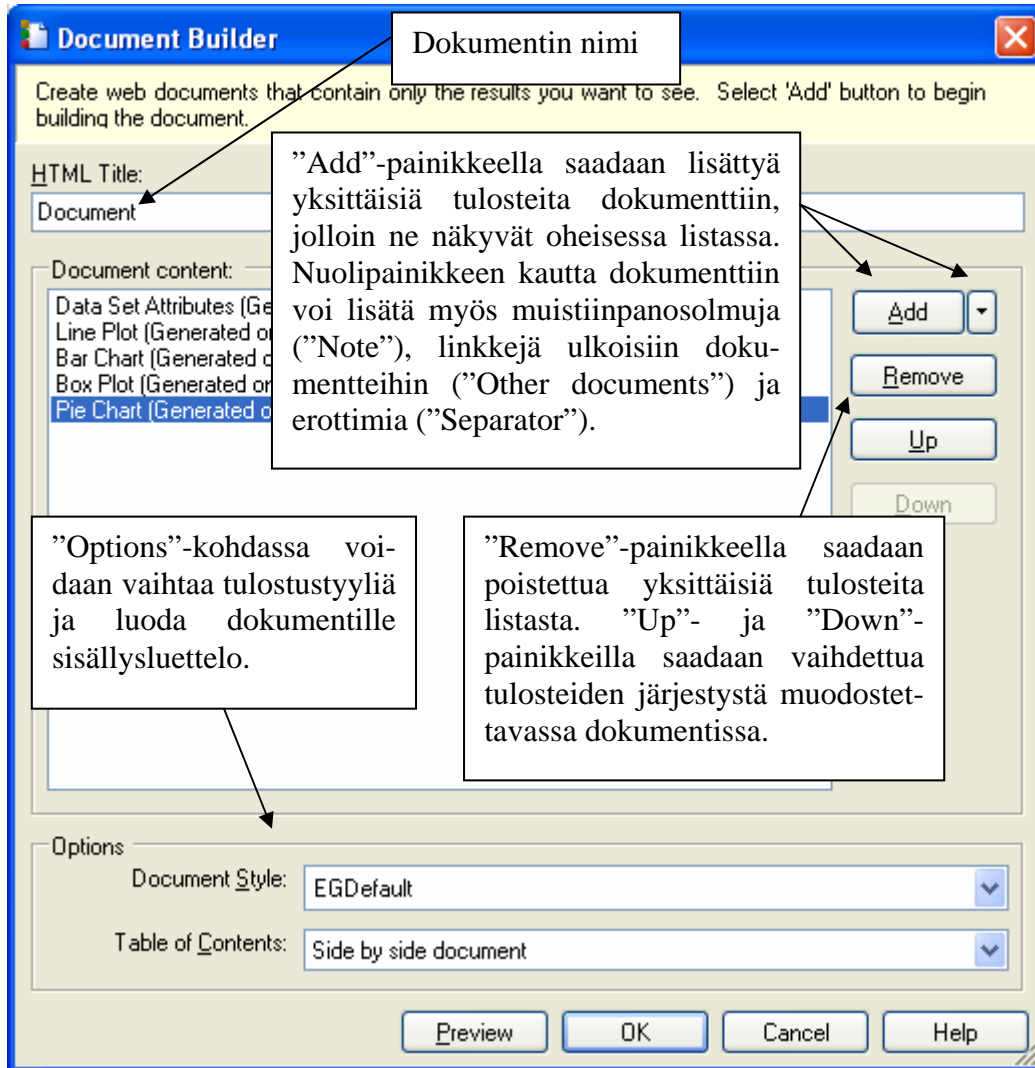
Otsikoiden ja alaviitteiden muokkaaminen toimii samalla tavalla kaikissa EG:n tehtäväsolmuissa, jotka tuottavat tulosteita. Tehtäviä muokatessa on mahdollista valita ”Titles”-välilehti, joka näyttää seuraavalta:



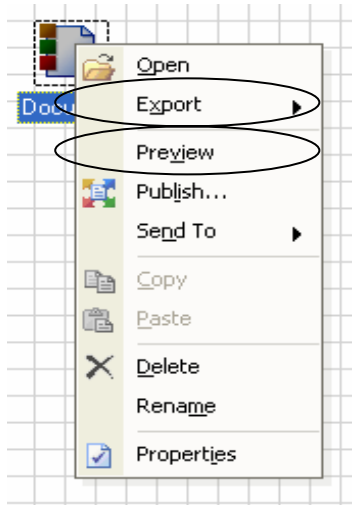
11.3 Yksittäisten tulosten kokoaminen yhteen

11.3.1 HTML dokumentti

Erillisistä tulostussolmuista on helppoa koota yksi yhteinen www-dokumentti käyttämällä Document Builderia. Document Builderin saa auki klikkaamalla ”Tools” -> ”Create HTML Document”, jolloin aukeaa seuraava ikkuna:

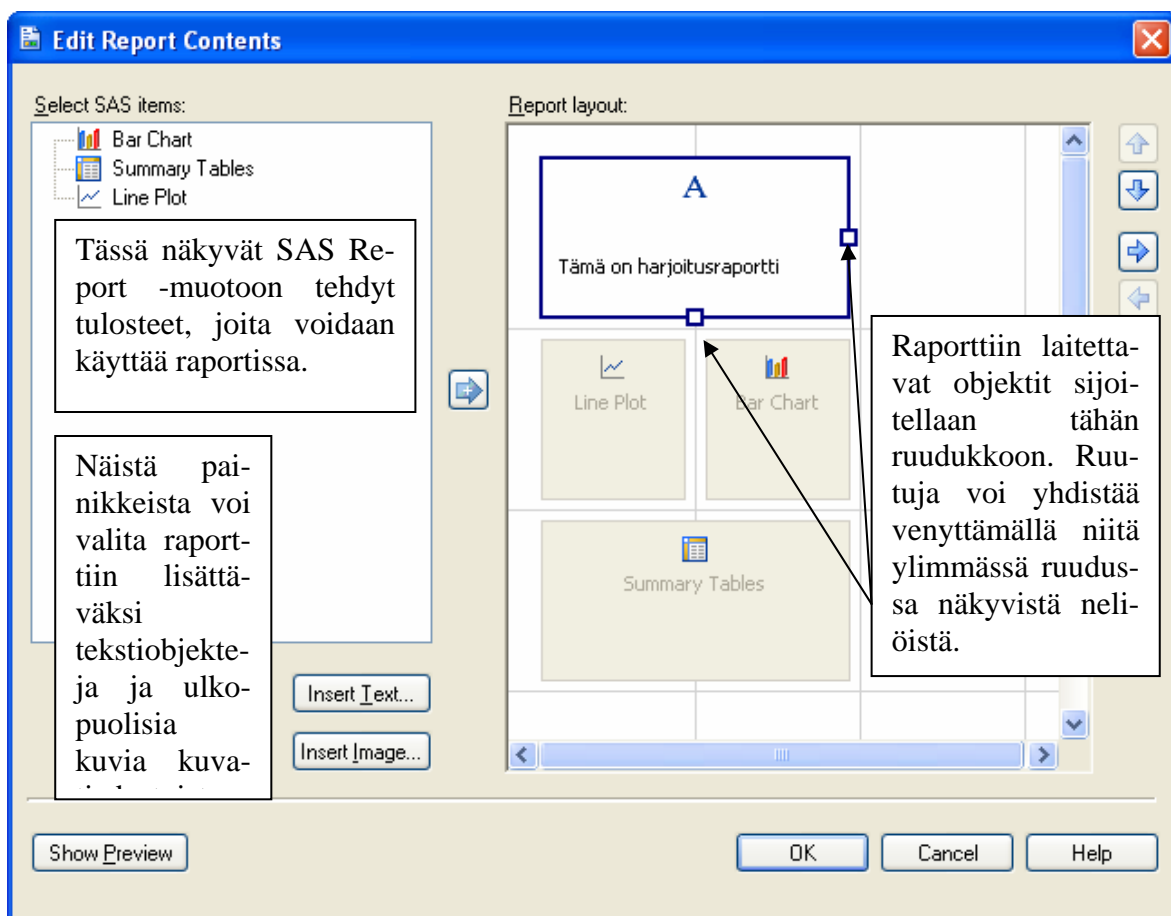


Painettaessa "OK" dokumentista tulee projektiin oma kuvake (prosessidiagrammin omalle "HTML Documents" -välilehdelle), jota voidaan joko esikatsella ("Preview") tai joka voidaan tallentaa HTML-tiedostoksi ("Export"):



11.3.2 SAS Raportti

Käyttäjä voi koota SAS Report -tulostusmuotoon tehdyistä tulosteista yhdistetyn SAS-raportin, jossa voidaan yhdistellä tulosteita, tekstiä, ulkopuolelta tuotavia kuvia ja otsikoita haluttuun muotoon. Uusi raporttipohja saadaan auki klikkaamalla "File" -> "New" -> "Report":



Koottua raporttia pääsee tarkastelemaan painamalla ”OK”:

Tästä valikosta voi poistaa sivunvaihdot raportista painamalla ”Remove All Page Breaks”. Valikon saa esille klikkaamalla ”Edit Report” -painikkeen vieressä olevaa nuolenpäätä.

Edelliseen ikkunaan pääsee takaisin painamalla ”Edit Report Contents” tai suoraan ”Edit Report” -painiketta.

”Page View” -painikkeella pääsee tarkastelemaan raporttia sivutettuna.

”Header & Footer” -painikkeella pääsee muokkaamaan raportin otsikkoa ja alaviitettä sekä valitsemaan, käytetäänkö yksittäisten tulostusobjektien omia otsikoita ja alaviitteitä raportissa.

Kuvia voi pienentää ja suurentaa ottamalla kuvasta kiinni ja vetämällä.

		paino					
		N	Mean	StdDev	Median	Q1	Q3
sukupu	tuotsuun						
1	1	30	83.07	15.58	82.50	73.00	92.00

Valmiin raportin voi tulostaa paperille tai lähettää toiselle Enterprise Guiden käyttäjälle tiedostona. Wordiin ja Exceliin näitä raportteja saa siirrettyä ainoastaan kopioimalla leikepöydän kautta. Prosessidiagrammiin raportit tulevat kuvakkeina omalle SAS Reports -välilehdelle.

11.4 Projektin hallinta

Enterprise Guidessa kaikki työ tallennetaan projekteina. Projekti koostuu toisiinsa liittyvistä datoista, toimenpiteistä, tulosteista, koodeista, log-tiedostoista ja muistiinpanoista. Projektin avulla kokonaisuus pysyy järjestyksessä, vaikka käytettäisiin useita eri datoja eri tehtävissä. Niinpä jälkikäteen pystyy kätevästi esimerkiksi selvittämään tarkalleen, mitä dataa on käytetty mihinkin analyysiin. Projekteja voi olla auki kerrallaan vain yksi, mutta samaan projektiin voidaan tehdä useita prosessidiagrammin välilehtiä.

11.4.1 Projektin perustoiminnot

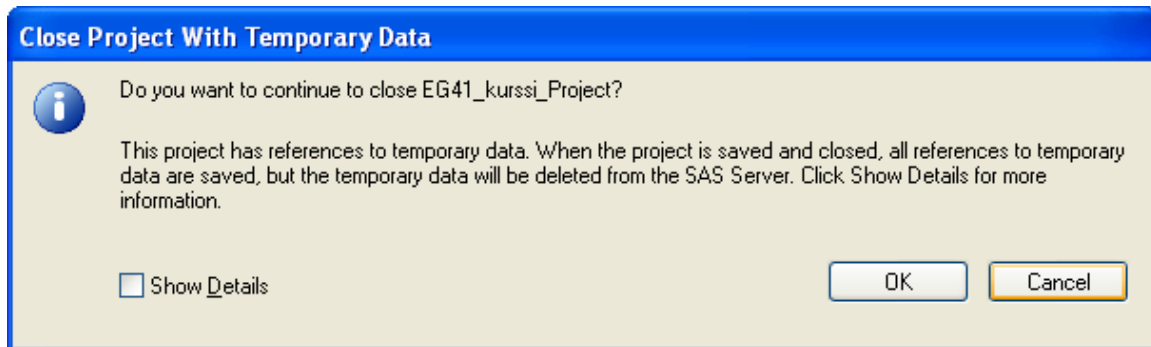
Uuden projektin luominen onnistuu valitsemalla ”File” -> ”New” -> ”Project”.

Vanhan projektin avaaminen puolestaan onnistuu valitsemalla ”File” -> ”Open” -> ”Project” -> ”Local Computer” ja hakemalla projektitiedoston tietokoneelta.

Projekti tallennetaan valitsemalla ”File” -> ”Save Project As” -> ”Local Computer” tai ”File” -> ”Save Project”.

Projektipuuhun ("Project Explorer -ikkuna") tulostuu automaattisesti kaikista suoritetuista tehtävistä log-tiedostot ja kooditiedostot. Niinpä koodeja ja log-tiedostoja on kätevä tarkastella projektipuusta, vaikka muuten käyttäisi pääasiassa prosessidiagrammia ("Project Designer"). Myös prosessidiagrammiin on mahdollista valita kooditiedostojen ja log-tiedostojen tallennus muokkaamalla asetuksia "Tools" -> "Options" -> "Project Views" -välilehdellä.

Enterprise Guide -istunnon lopettaminen tehdään normaalisti klikkaamalla oikean yläkulman ruksipainiketta tai valitsemalla "File" -> "Exit". Jos projektia ei ole tallennettu ennen sulkemista, EG kysyy tallentamisesta suljettaessa. Jos taas projektissa on tehty väliaikaisia SAS-datoja, EG varoittaa asiasta seuraavalla ilmoituksella:



Väliaikaiset SAS-datat häviävät tietokoneelta, kun EG suljetaan, ja edellä kuvattu varmistus mahdollistaa palaamisen projektiin ja datojen tallentamisen pysyviksi datoina. Yleensä väliaikaisia datoja ei välttämättä tarvitse tallentaa pysyvinä datoina, jos ne on helppo luoda takaisin käyttöön seuraavassa EG istunnossa suorittamalla tietyt tehtäväsolmut uudelleen.

11.4.2 Projektin pitäminen järjestyksessä

Prosessidiagrammissa on muutamia käytännöllisiä ominaisuuksia projektin järjestelmälliseen hallintaan. Ensinnäkin prosessidiagrammissa voidaan solmuja jakaa useammalla eri välilehdelle. Uuden välilehden saa avatuksi valitsemalla "File" -> "New" -> "Process Flow". Erillisille välilehdille on kätevä ryhmitellä esimerkiksi samaan projektiin kuuluvien erillisten datojen analyysit. Lisätyn välilehden nimeä saa vaihdetuksi klikkaamalla sen oletusnimen kohdalla hiiren oikealla painikkeella ja valitsemalla "Rename".

Järjestyksen ylläpitoa prosessidiagrammissa auttaa ennen kaikkea sinne tulostuvien kuvakkeiden eli solmujen organisoiminen loogiseen järjestykseen. Oletuksena Enterprise Guidessa on käytössä toiminto "Autoarrange", joka järjestää solmuja käyttäjän puolesta prosessidiagrammissa. "Autoarrange"-toiminnon saa pois käytöstä klikkaamalla prosessidiagrammin tyhjässä kohdassa hiiren oikealla painikkeella ja klikkaamalla "Autoarrange", jos sen kohdalla näkyy valintamerkki. Tämän jälkeen solmuja saa siirrettyä ottamalla niistä hiirellä kiinni ja raahaamalla ne uuteen paikkaan. Solmujen välillä olevat yhdysviivat venyvät siirrettäessä.

Prosessidiagrammin selkeyden kannalta on myös tärkeätä, että luotaessa SAS-datoja tehtäväsolmuissa, kiinnitetään huomiota nimien valintaan ja tallennuskohteeseen. Kannattaa siis tarkistaa, onko oletuskirjastoviitteen osoittama kansio hyvä kohde tallennukselle, ja tarvittaessa vaihtaa kirjastoviitettä. Samoin kannattaa aina valita luotavalle datalle looginen nimi.

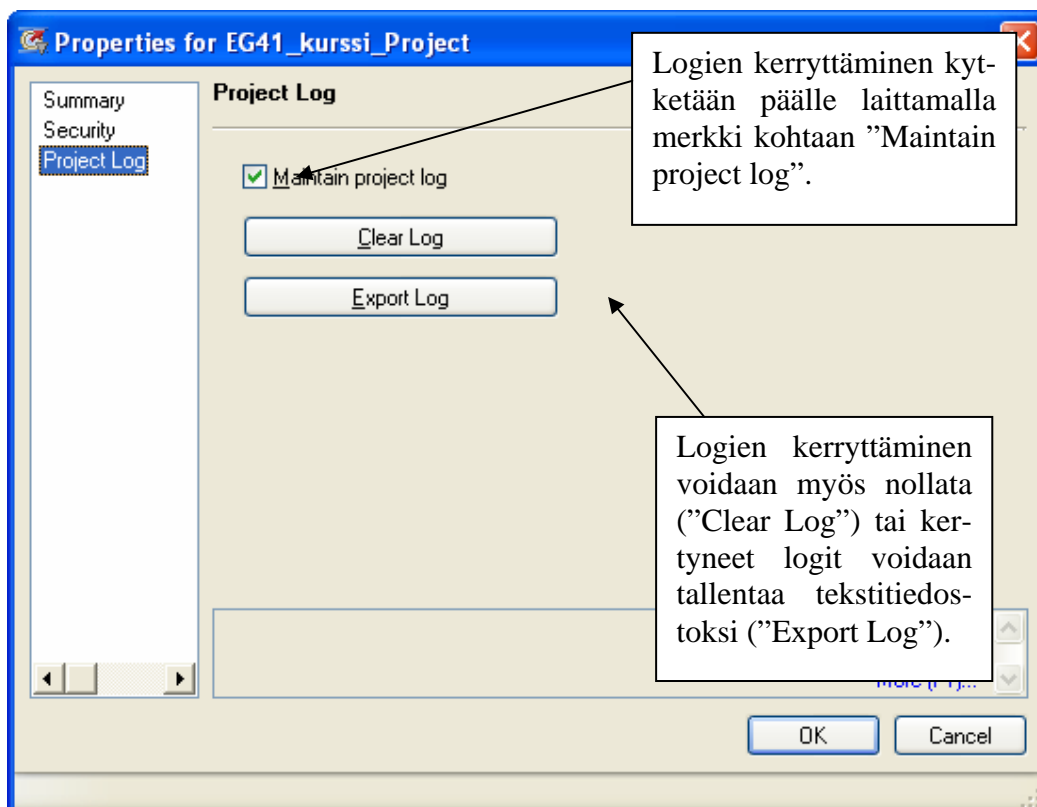
Joskus prosessidiagrammin jäsentämistä helpottaa ylimääräisten yhdysviivojen vetäminen sellaisen diagrammissa olevien solmujen välille, joiden välillä yhdysviivaa ei muuten ole.

Tällainen linkitys onnistuu klikkaamalla ensimmäistä linkitettävää solmua hiiren oikealla painikkeella ja valitsemalla "Link solmun_nimi to". Tällöin aukeaa ikkuna, josta voidaan valita toinen solmu, johon yhdysviiva vedetään. Linkittäminen on käyttökelpoista esimerkiksi SAS-koodisolmuja käytettäessä, jolloin voidaan linkittää koodit niiden käyttämiin SAS-datoihin (tällaisia linkkejä ei automaattisesti synny). Toisaalta yhdysviivojen avulla voidaan myös pakottaa solmut suoritettavaksi tietyssä järjestyksessä.

Prosessidiagrammin näkymää voi myös "zoomata" eli lähentää tai loitontaa. Tämä onnistuu klikkaamalla prosessidiagrammin tyhjää kohtaa hiiren oikealla painikkeella ja valitsemalla "Zoom"-kohdasta pienennys- tai suurennusprosentin.

Projektin dokumentoinnissa käyttökelpoinen työkalu on muistiinpano-solmu. Muistiinpanosolmun saa lisättyä klikkaamalla "File" -> "New" -> "Note". Jos samalla on valittuna esimerkiksi joku tehtäväsolmu, muistiinpanosolmu yhdistyy siihen yhdysviivalla. Muistiinpanosolmut voivat kuitenkin olla myös irrallisina projektissa. Muistiinpanosolmuun voi kirjoittaa mitä vain, ja solmut tallentuvat projektin mukana pysyviksi muistiinpanoiksi projektista. Tätä ominaisuutta käyttämällä projektiin palaaminen pitkän ajan kuluttua helpottuu.

Enterprise Guiden versiossa 4.1 on uutena ominaisuutena mahdollisuus kerätä kaikki projektin logit yhdeksi kokonaisuudeksi. Tämä ominaisuus vastaa tavallisen SASin log-ikkunaa, johon kertyy tieto kaikista suoritetuista koodeista sekä niiden suoritukseen liittyvistä ilmoituksista mukaan lukien virheilmoitukset. Enterprise Guidessa ikkunaan kertyy tieto kaikkien suoritettujen tehtäväsolmujen taustalla olevista SAS-koodeista ja niihin liittyvistä ilmoituksista sekä tietysti vastaavat tiedot tavallisen SAS-koodin suorittamisesta. Enterprise Guidessa koodin kerryttäminen log-ikkunaan pitää kuitenkin ensin kytkeä toimintaan. Tämä päästään tekemään klikkaamalla hiiren oikealla painikkeella projektipuussa ("Project Explorer") ylimmäisenä olevaa projektikuvaketta ja aukeavasta valikosta "Properties" -> "Project Log". Tällöin aukeaa seuraava ikkuna:



Logien tallentamisen lisäksi niitä voidaan tarkastella myös log-ikkunasta, joka saadaan auki klikkaamalla "View" -> "Project Log". Tällöin työtilaan aukeaa oma Project Log -välilehti, johon kertyy tietoa tehtäväsolmujen tai koodien suorittamisesta.

11.4.3 Tehtäväsolmujen käsittely

Projektissa olevan tehtäväsolmun voi suorittaa klikkaamalla sitä hiiren oikealla painikkeella ja valitsemalla "Run This Task". Tehtävän suoritus onnistuu myös valitsemalla kyseinen solmu aktiiviseksi prosessidiagrammissa ja klikkaamalla työkalurivissä olevaa vihreää "Play-nappulaa". Kolmas vaihtoehto on tuplaklikata tehtäväsolmu muokkaustilaan ja suorittaa se kuten uudet tehtäväsolmut. Tehtävän suorituksen pysäyttäminen puolestaan onnistuu klikkaamalla työkalurivissä olevaa punaista "Stop-nappulaa". Edellä mainittujen tapojen lisäksi tehtävien suorittaminen ja suorituksen pysäyttäminen onnistuu myös "Code"-valikon kautta. Virhe tehtäväsolmun suorittamisessa näkyy siten, että prosessidiagrammissa olevaan tehtäväkuvakkeeseen tulee pieni punainen ruksi ja tehtävään liittyvä log-ikkuna aukeaa. Log-ikkunasta löytyy tällöin punaisella merkitty virheilmoitus, josta yleensä löytyy virheen syy. Pienemmistä rikkeistä tehtäväsolmun kuvakkeeseen tulee keltaisella pohjalla oleva huutomerkki. Nämä huutomerkkit eivät yleensä viittaa mihinkään merkittävään virheeseen, mutta ainakin jos tehtävän tulos ei ole looginen, asia kannattaa tarkistaa logista (esim. projektipuun kautta).

Yksittäisten tehtäväsolmujen suorittamisen lisäksi voidaan suorittaa myös useampi tehtäväsolmu kerrallaan. Käyttäjä voi suorittaa kaikki samalla prosessidiagrammin välilehdellä olevat tehtäväsolmut klikkaamalla hiiren oikealla painikkeella jossakin tyhjässä prosessidiagrammin kohdassa ja valitsemalla "Run Process Flow". Toisaalta käyttäjä voi suorittaa tietyn tehtäväsolmun ja kaikki sen perään yhdysviivoilla yhdistetyt tehtäväsolmut klikkaamalla ensimmäistä solmua hiiren oikealla painikkeella ja valitsemalla "Run Branch From tehtäväsolmun nimi".

Kuten edellä jo mainittiin, projektissa jo olevia valmiita tehtäväsolmuja pääsee muokkaamaan jälkikäteen tuplaklikkaamalla solmua (tehtävän suoritettavaa solmua, ei tulostussolmua) prosessidiagrammissa tai klikkaamalla solmua hiiren oikealla painikkeella ja valitsemalla "Open". Muokkausten jälkeen solmun voi suorittaa uudelleen valitsemalla "Run". Tällöin EG kysyy, halutaanko korvata edellisen suorituksen tulokset. Jos vastataan "Kyllä", tehdyt muutokset jäävät tehtäväsolmuun ja uudet tulokset ajetaan vanhojen päälle. Jos taas vastataan "Ei", EG tekee uuden tehtäväsolmun ja uuden tulostussolmun, jolloin vanhat tehtävä- ja tulostussolmut jäävät koskemattomiksi.

Myös tehtäväsolmun käyttämää dataa voidaan vaihtaa. Tämä on tarpeellista esimerkiksi silloin, kun tuodaan kokonaan uusi data projektiin ja sille halutaan tehdä monia toiselle datalle aiemmin määriteltyjä toimenpiteitä. Tällöin data ikään kuin asetetaan "ketjun alkuun". Datan vaihtaminen onnistuu klikkaamalla hiiren oikealla painikkeella tehtäväsolmua, jolle halutaan uusi data käyttöön, ja valitsemalla "Select Input Data" ja siitä haluttu, projektissa mukana oleva data. Query Builder -tehtäväsolmuille dataa ei voi vaihtaa tällä keinolla.

Tehtäväsolmun nimeä voi myös muuttaa. Tästä voi olla apua projektin dokumentoinnissa tulevaa käyttöä varten. Solmun nimeä pääsee muokkaamaan klikkaamalla solmua hitaasti kahdesti tai klikkaamalla solmua hiiren oikealla painikkeella ja valitsemalla "Rename".

Turhan tehtäväsolmun voi poistaa. Tämä onnistuu valitsemalla solmu aktiiviseksi ja painamalla "Delete"-näppäintä tai klikkaamalla solmua hiiren oikealla painikkeella ja valit-

semalla ”Delete”. Kun solmu poistetaan, samalla poistuu sen perässä olevia, siihen liittyviä muita solmuja. Kun esimerkiksi poistetaan sirontakuvion tuottava tehtäväsolmu, samalla poistuu siihen liittyvät tulostussolmu, koodisolmu ja log-solmu.

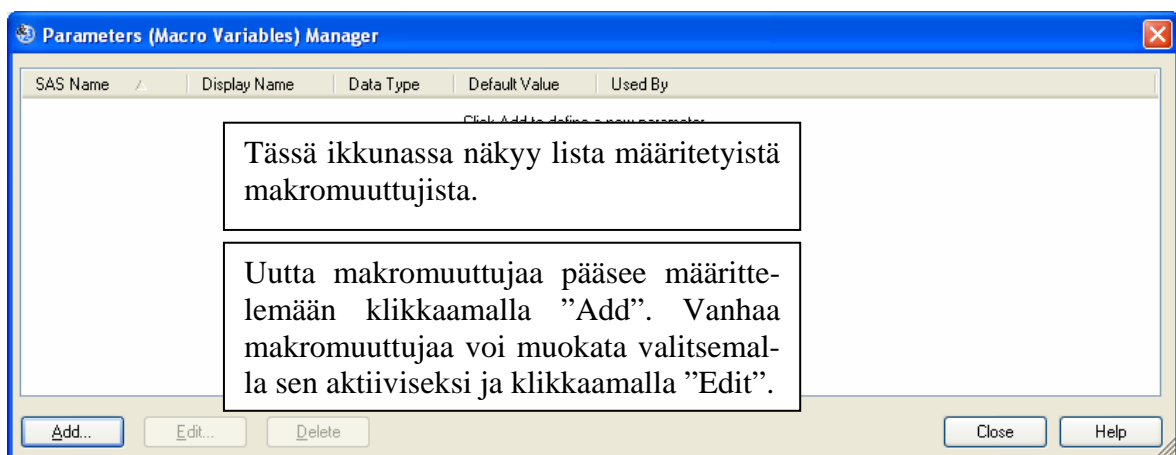
Koko projekti voidaan myös määrätä suoritettavaksi ajastuksen avulla. Ajastustyökalu löytyy klikkaamalla ”Tools” -> ”Schedule Project”. ”Create Ordered List” -työkalulla voi puolestaan suorittaa useamman projektissa olevan tehtäväsolmun halutussa järjestyksessä. Työkalussa valitaan projektissa olevia tehtäviä listaan ja tuloksena syntyy ”Ordered List” -tehtäväsolmu. Listaan valitut tehtävät voidaan tämän jälkeen suorittaa helposti sarjassa suorittamalla kyseinen ”Ordered List” -tehtäväsolmu. Tämän työkalun saa käyttöön klikkaamalla ”Tools” -> ”Create Ordered List”.

11.5 Makromuuttujat (Parameters)

Makromuuttujien avulla pystytään lisäämään uusia ulottuvuuksia Enterprise Guiden tehtäväsolmujen käyttämiseen. Esimerkiksi osaryhmittäin keskiarvoja laskevassa tehtäväsolmussa voitaisiin korvata sekä analysoitava muuttuja että ryhmittelevä muuttuja makromuuttujilla, jolloin käyttäjä saisi valita todelliset käytettävät muuttujat vasta tehtäväsolmun suorituksen yhteydessä. Tällöin kerran määritellyllä tehtäväsolmulla saataisiin laskettua keskiarvoja käyttäen monia eri muuttujia. Suorittaessa makromuuttujia sisältävä tehtäväsolmu, Enterprise Guideen aukeaa ikkuna, jossa kysytään käyttäjän makromuuttujan määrittelyn yhteydessä antama kysymys ja annetaan tietyt vastausvaihtoehdot, jotka ovat makromuuttujan mahdollisia arvoja. Arvot voivat olla esimerkiksi muuttujien nimiä, lukuarvoja tai merkkijonoja. Enterprise Guide suorittaa tehtävän käyttäjän eri vaihtoehdoista valitsemalla arvolla. Makromuuttujien avulla voidaan siis rakentaa eräänlaisia aliohjelmiä, jolloin yhdellä tehtäväsolmulla saadaan suoritetuksi useita tehtäviä. Makromuuttujia voidaan käyttää muun muassa jonkin tehtäväsolmun käyttämien muuttujien määrittelyyn tai Query Builderissä aineiston osa-aineistoiksi jakavan ehdon määrittelyyn.

11.5.1 Makromuuttujien määrittely

Makromuuttujia päästään määrittelemään valitsemalla ”Tools” -> ”Parameter (Macro Variable) Manager”. Tällöin aukeaa seuraava ikkuna:



Klikkaamalla ”Add” päästään seuraavaan ikkunaan:

”Display name” -kohtaan laitetaan kysymys tai kehoitus, joka tulee ruudulle, kun käyttäjältä kysytään makromuuttujan arvoa.

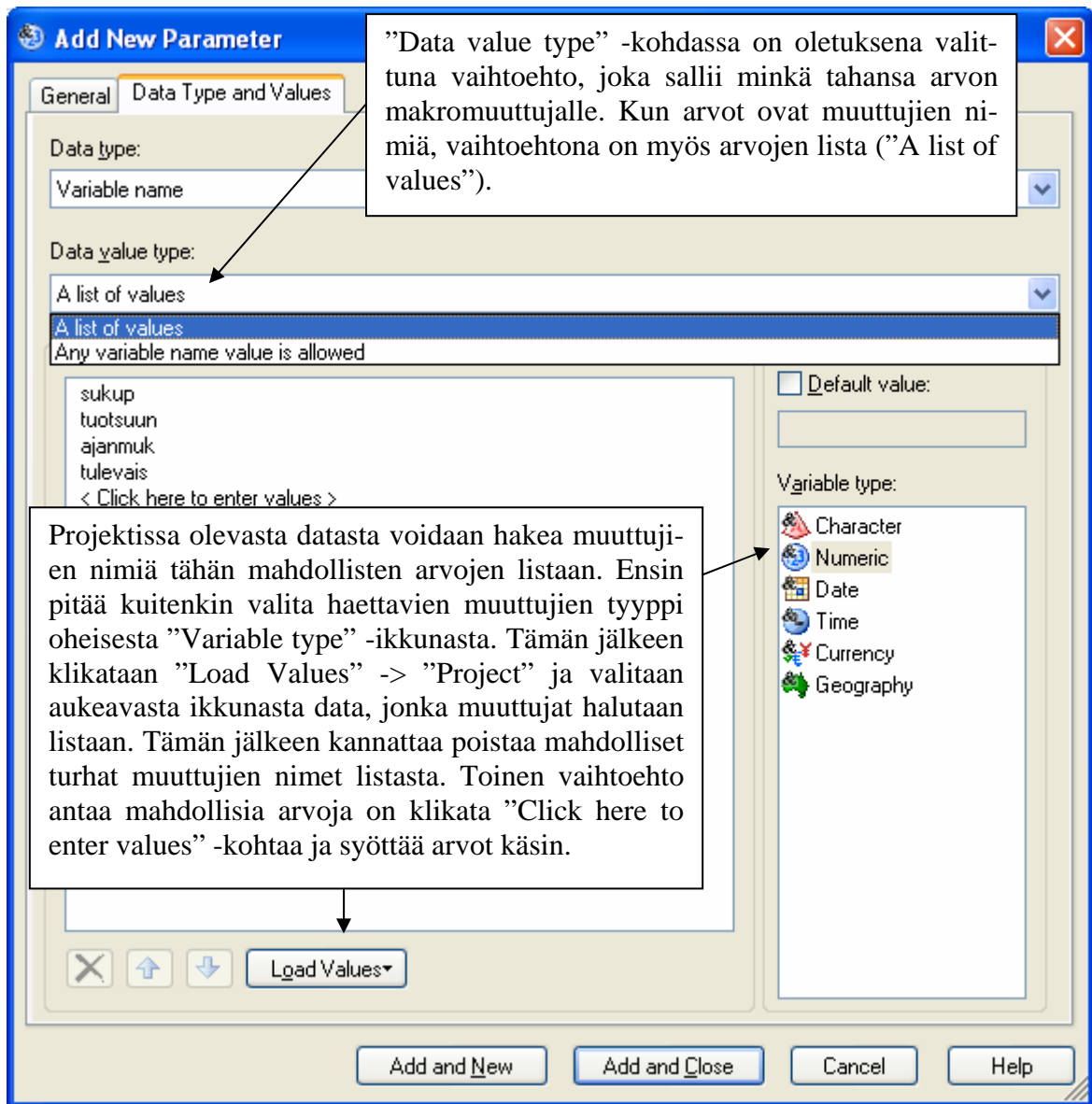
”SAS code name” -kohtaan puolestaan laitetaan makromuuttujan nimi. Nimeen pätevät samat rajoitukset kuin tavallisten muuttujien nimiin.

”Description”-kohdassa voi antaa lisäselitystä makromuuttujasta.

”Data type” -kohdassa määritellään makromuuttujan mahdollisten arvojen tyyppi. Valittavana ovat merkkijonot (”String”), kokonaisluvut (”Integer”), desimaaliluvut (”Float”), erilaiset aika- ja päivämääräarvot sekä muuttujien nimet (”Variable name”).

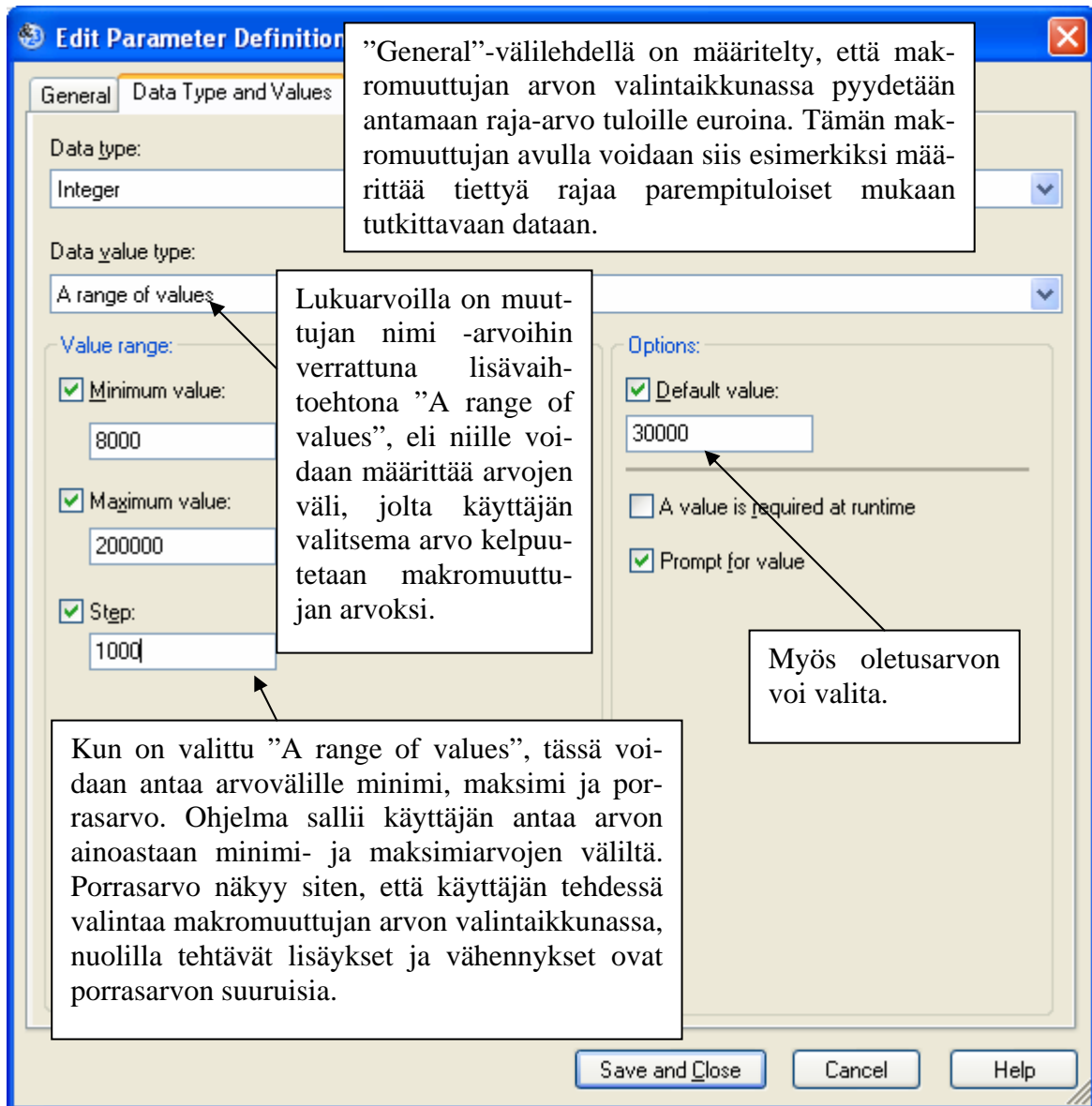
Edellä tehtyjen määritysten lisäksi ei ole pakko tehdä muita määrittämiä, mutta jos makromuuttujalle halutaan määrittää lista mahdollisia arvoja tai jokin mahdollisten arvojen väli, se voidaan tehdä välilehdellä ”Data Type and Values”. Arvojen lista tai arvoväli kannattaa useimmiten määrittää, koska se helpottaa makromuuttujan myöhempää käyttöä.

”Data Type and Values” -välilehti näyttää seuraavalta:



Lopuksi painetaan ”Add and Close” ja ”Close”. Jos halutaan määrittää heti uusi makromuuttuja, yllä olevassa ikkunassa voidaan valita myös ”Add and New”.

Kun makromuuttuja on määritelty, se on valmis käytettäväksi jossakin tehtäväsolmussa tai SAS-koodissa. Katsotaan kuitenkin ensin, miten numeroarvojen määrittely makromuuttujalle eroaa edellä läpikäydystä muuttujan nimi -arvojen määrittelystä. Liikkeelle lähdetään ihan samalla tavalla eli avataan Parameters Manager ja aletaan määrittellä uutta makromuuttujaa. Uuden makromuuttujan määrittelyssä ”General”-välilehdellä toimitaan samoin kuin edellä, mutta valitaan vain ”Data type” -kohtaan esimerkiksi ”Integer” eli kokonaisluvut makromuuttujan mahdollisiksi arvoiksi. Tällöin ”Data Type and Values” -välilehti näyttää seuraavalta:

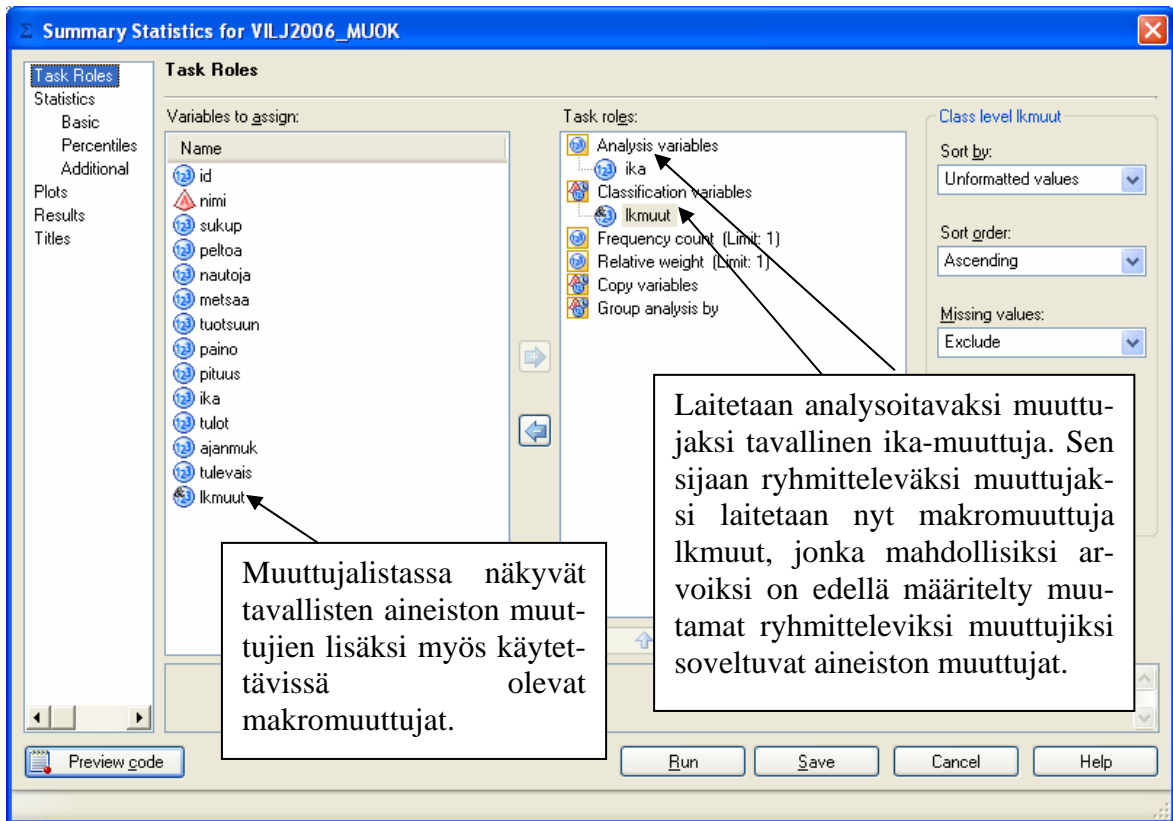


Lopuksi painetaan ”Save and Close” ja ”Close”.

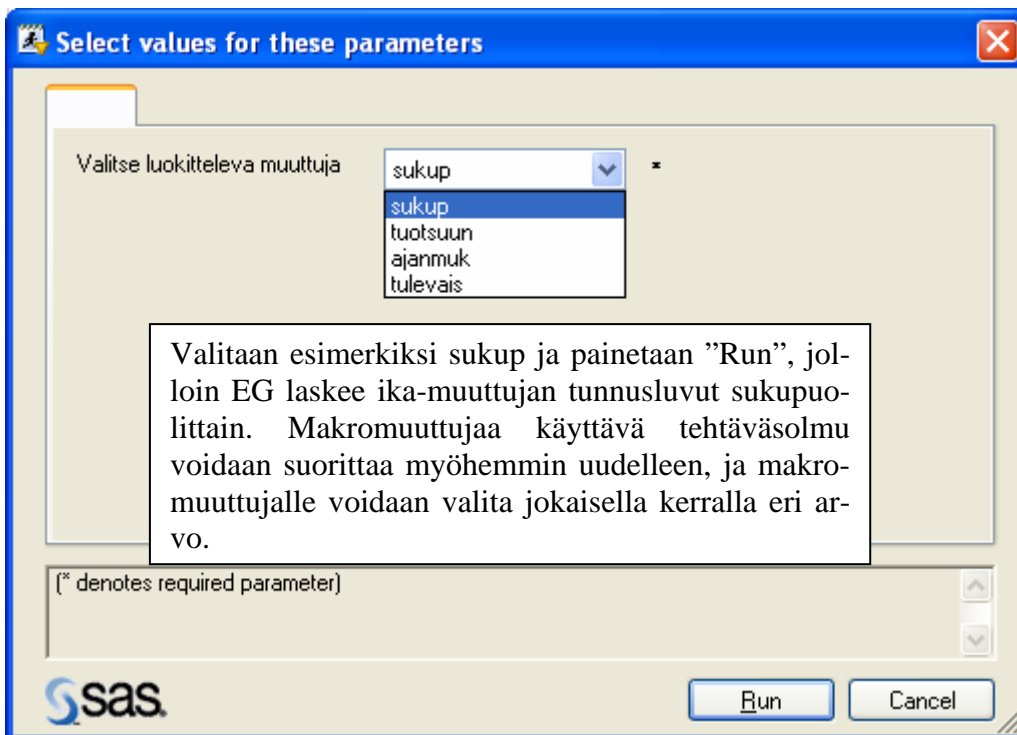
11.5.2 Makromuuttujien käyttäminen

Edellä on käyty läpi muutaman erilaisen makromuuttujan muodostaminen. Valmiita makromuuttujia voidaan käyttää muun muassa erilaisissa tulostusta tuottavissa tehtäväsolmuissa muuttujien valinnassa sekä Query Builderin ”Filter Data” -välilehdellä osaineiston valintaan liittyvissä ehtolauseissa. Myös SAS-koodissa voidaan käyttää makromuuttujia.

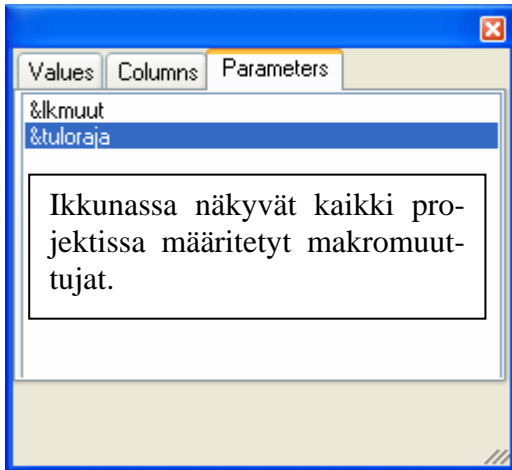
Ensin käydään läpi tilanne, jossa valitaan ryhmittelevä muuttuja makromuuttujan avulla. Tätä varten avataan ensin ”Summary Statistics” -työkalu siten, että Viljelijä 2006 -aineisto on aktiivisena:



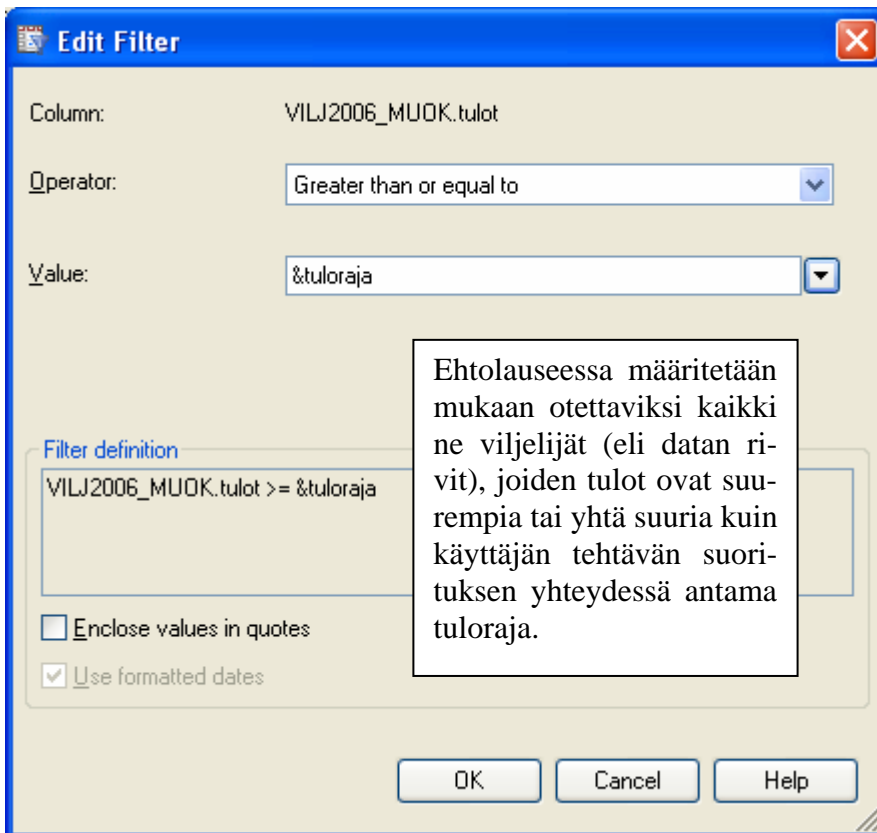
Kun painetaan ”Run”, aukeaa valintaikkuna, jossa pitää valita mitä makromuuttujan mahdollisista arvoista tässä tilanteessa käytetään:



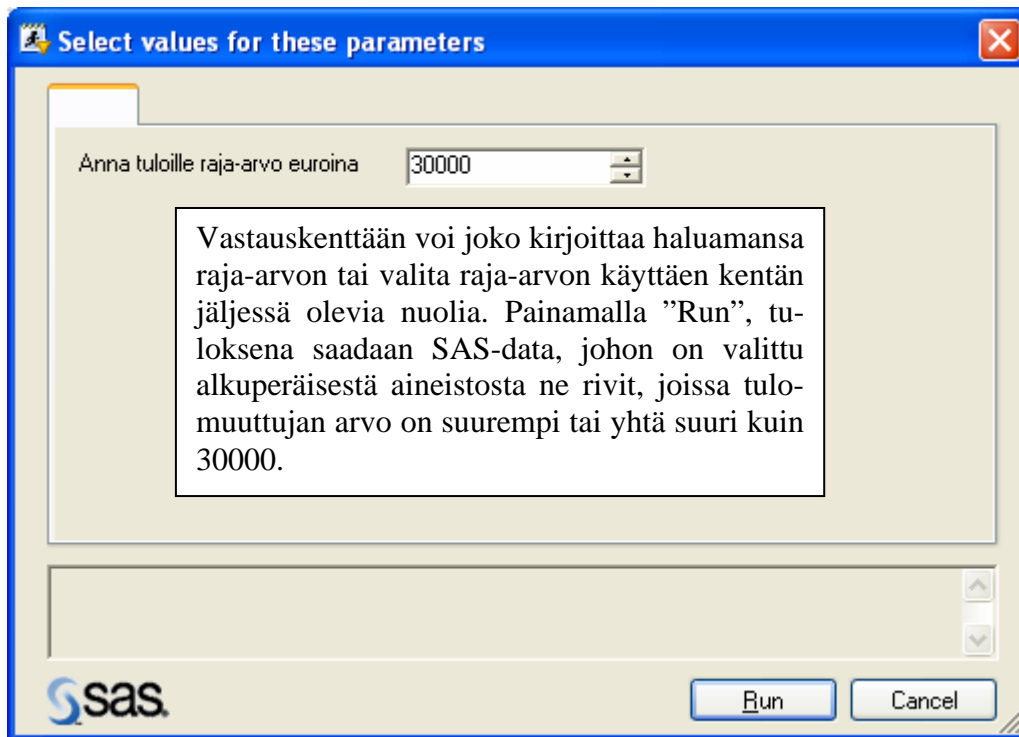
Toinen tilanne, jossa makromuuttujat ovat käyttökelpoisia, liittyy osa-aineiston valintaan. Avataan ensin Query Builder siten, että käytettävä data on aktiivisena ja mennään "Filter Data" -välilehdelle. Raahataan "Filter Data" -välilehdelle muuttuja, jonka arvoihin liittyvä makromuuttuja on aiemmin määritelty. Aukeavassa ikkunassa valitaan ensin "Operator"-kohdassa ehto eli esimerkiksi "Greater than or equal to" ja sitten klikataan "Value"-kohdan päässä olevaa nuolta. Tällöin aukeaa seuraava ikkuna, jonka "Parameters"-välilehdellä voidaan määrittää makromuuttuja käytettäväksi ehtolauseessa:



Kun listasta valitaan makromuuttuja, EG palaa edelliseen ikkunaan, jossa näkyy makromuuttujan avulla määritetty ehtolause:



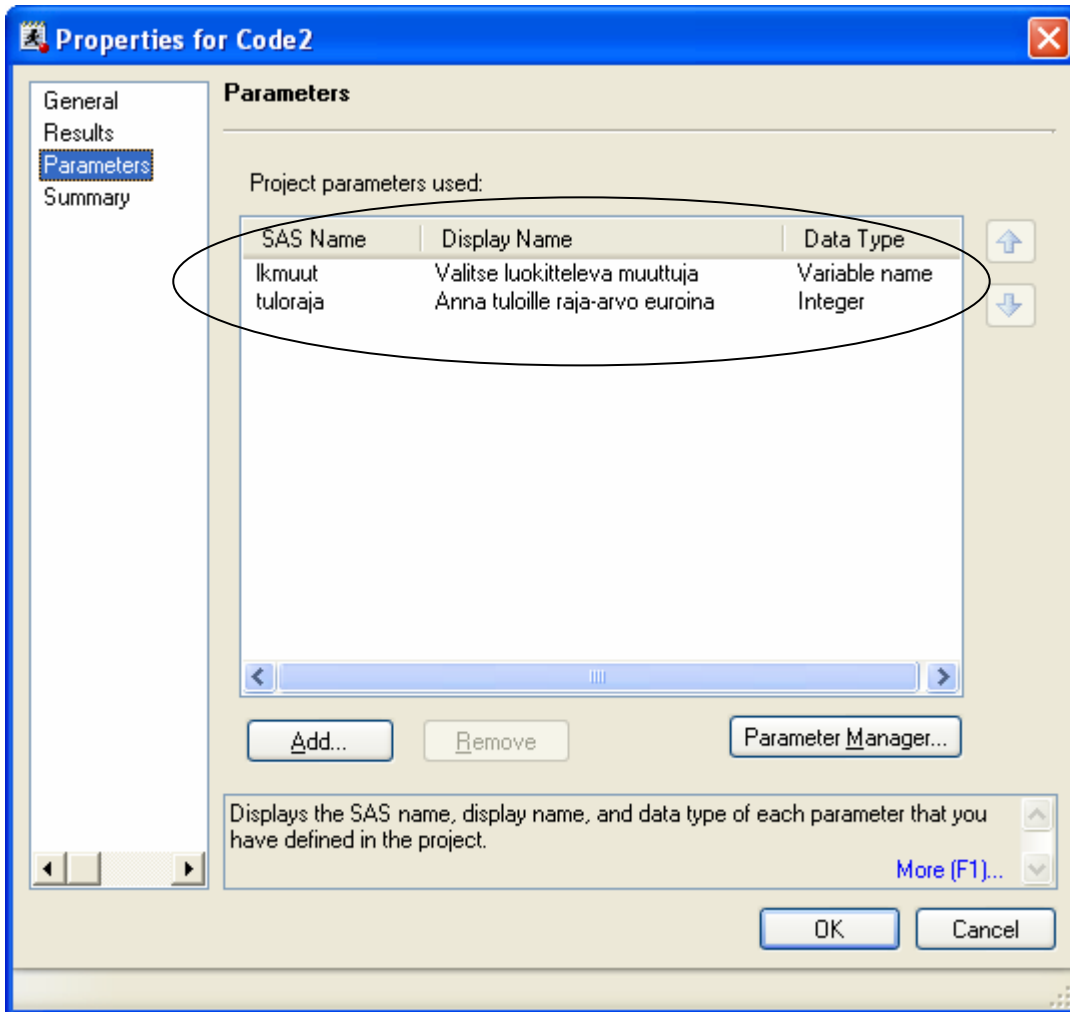
Lopuksi painetaan ”OK” ja suoritetaan Query Builder normaalisti. Suorituksen yhteydessä aukeaa seuraava ikkuna:



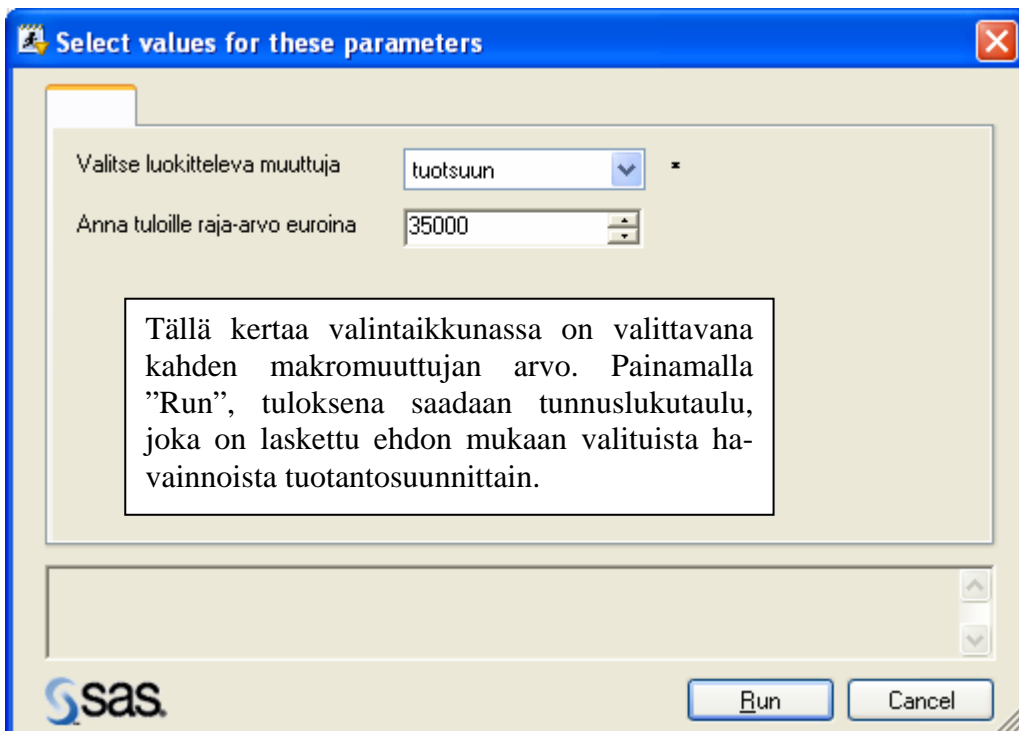
SAS-koodia käytettäessä makromuuttujiin viitataan niiden nimellä sillä lisäyksellä, että nimen eteen tulee makromuuttujiin viittaava &-merkki. Samassa tehtävässä tai SAS-koodissa voi myös käyttää useampia makromuuttujia. Alla on esimerkki SAS-koodista, jossa on käytetty molempia edellä määriteltyjä makromuuttujia:

```
Project Designer | VILJ2006_MUOK (read-only) | Code2*  
PROC MEANS DATA=kir.vilj2006_muok;  
  WHERE tulot>=&tuloraja;  
  CLASS &lkmuut;  
  VAR ika;  
RUN;
```

Ennen koodin suorittamista makromuuttujat pitää vielä erikseen yhdistää tähän koodiin, koska ne on määritelty muualla kuin tässä koodissa. Tämä onnistuu klikkaamalla koodisolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Properties”. Tämän jälkeen mennään ”Parameters”-välilehdelle, klikataan ”Add” ja valitaan käytettävät makromuuttujat. Kun muuttujat on valittu, ne näkyvät seuraavassa ikkunassa:



Sitten painetaan "OK" ja suoritetaan koodi, jolloin aukeaa makromuuttujien arvojen valintaikkuna:



11.6 Virhetilanteet

- 1) Enterprise Guiden selittämättömissä virhetilanteissa auttaa useimmiten ohjelman sulkeminen ja uudelleenkäynnistäminen. Joskus täytyy käynnistää uudelleen koko tietokone. Ennen sulkemista kannattaa tallentaa projektiin tehdyt muutokset, ja uudelleenkäynnistyksen jälkeen pitää yhdistää tarvittavat kirjastoviitteet uudelleen.
- 2) Apua Enterprise Guide -ongelmiin löytyy myös SASin teknisestä tuesta. Kannattaa hakea apua <http://support.sas.com> -sivujen hakutoiminnolla.
- 3) Asennuksen jälkeen kannattaa asentaa Enterprise Guiden uusin hotfix eli korjauspaketti, jonka saa SASin teknisen tuen internet-sivuilta (http://ftp.sas.com/techsup/download/hotfix/ent_guide.html). Tämä paketti korjaa monia perustoimenpiteissäkin esiintyviä ohjelmavirheitä.
- 4) Tässä oppaassa mahdollisesti havaituista virheistä voi lähettää palautetta kirjoittajalle esimerkiksi sähköpostilla osoitteeseen timo.hurme@mtt.fi.

12 Kirjallisuus

Documentation for SAS 9 Products (Verkkodokumentti) Viitattu 4.4.2007. Saatavissa internetistä: <http://support.sas.com/documentation/onlinedoc/sas9doc.html>

Gomez, K. A. ja Gomez, A. A. 1984. Statistical procedures for agricultural research. Second edition. New York, NY: John Wiley & Sons. 680 s.

Helenius, H. ja Katajisto, J. 2005. Tilastomenetelmien soveltamisesimerkkejä SAS® Enterprise Guide:lla. Espoo: Numos Oy. 305 s.

Slaughter, S. J. ja Delwiche, L.D. 2005. The Little SAS book for Enterprise Guide 3.0. Cary, NC: SAS Publishing. 364 s.

Slaughter, S. J. ja Delwiche, L. D. 2006. The Little SAS book for Enterprise Guide 4.1. Cary, NC: SAS Institute Inc. 388 s.

Vahlberg, T., Helenius, H. ja Katajisto, J. 2006. SAS® Enterprise Guide® 3.0:n käyttöympäristö ja aineistojen muokkaus. Espoo: Numos Oy. 139 s.

MTT:n selvityksiä 140

