

MTT RAPORTTI 5

Opas SAS Enterprise Guiden käyttäjälle

EG versio 4.2

Timo Hurme



MTI RAPORTTI 5

Opas SAS Enterprise Guiden käyttäjälle

EG versio 4.2

Timo Hurme

ISBN: painettu: 978-952-487-275-1, verkkojulkaisu: 978-952-487-276-8

ISSN: 1798-6419

www-osoite: www.mtt.fi/mttraportti/pdf/mttraportti5.pdf

Copyright: MTT

Kirjoittajat: Timo Hurme

Julkaisija ja kustantaja: MTT, 31600 Jokioinen

Julkaisuvuosi: 2010

Kannen kuva: Jii Roikonen

Painopaikka: Tampereen Yliopistopaino Oy – Juvenes Print

Opas SAS Enterprise Guiden käyttäjälle

Hurme, Timo

MTT, Kasvintuotannon tutkimus, 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 Guiden graafisessa käyttöliittymässä voidaan muokata tutkimusaineistoja, 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. Lisäksi tutustutaan Enterprise Guiden yleisiin ominaisuuksiin.

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

Avainsanat:

SAS, Enterprise Guide 4.2, tilastollinen ohjelmisto, kuvaileva tilastotiede, aineiston tarkastelu

Sisällysluettelo

1 Johdanto.....	6
2 Oppaassa käytettävät datat	7
3 Enterprise Guiden käyttöympäristö	9
4 Enterprise Guiden ohje	13
5 Aineiston käyttöönotto	14
5.1 Kirjastoviite	14
5.2 Aineiston tallennusmuoto.....	16
5.3 Aineiston syöttäminen Enterprise Guidessa.....	18
5.4 Excel-tiedoston avaaminen Enterprise Guideen.....	20
5.5 Erikoismuuttujien sisäänlukeminen ja ongelmatapaukset.....	25
5.6 Tekstimuotoisen data-tiedoston avaaminen Enterprise Guideen	27
5.7 SAS-datan avaaminen Enterprise Guideen	29
5.8 JMP-, SPSS- ja Stata-tiedostojen avaaminen Enterprise Guideen	30
5.9 Datojen tallentaminen Enterprise Guidesta.....	31
5.10 Projektin tallentaminen	31
6 Datan muokkaus.....	32
6.1 Datan muuttaminen muokkaustilaan	32
6.2 Muuttujien perusominaisuuksien määrittely	33
6.3 Muuttujan arvojen tulostusmuotojen eli formaattien käyttäminen.....	33
6.4 Datan muokkaustyökalut.....	36
6.5 Muuttujien muunnokset ja uusien muuttujien tekeminen	38
6.6 Muuttujien luokittelu ja luokkien yhdistäminen	41
6.7 Muuttujan luominen aritmeettisella operaatiolla.....	47
6.8 Muuttujien poistaminen.....	48
6.9 Muuttujien järjestyksen vaihtaminen datassa.....	49
6.10 Datan havaintorivien järjestäminen.....	50
6.11 Osa-aineiston eli tiettyjen havaintorivien valitseminen	51
6.12 Kahden tai useamman datan yhdistäminen	55
6.13 Datan kääntäminen.....	58
6.14 Muuta datojen käsittelyyn liittyvää.....	62
7 Datan tarkastelu Graph-valikon kuvatyökaluilla.....	63
7.1 Sirontakuvio (Scatter Plot).....	64
7.2 Pylväskuvio (Bar Chart).....	70
7.3 Viivakuva (Line Plot).....	74
7.4 Laatikko-janakuvio (Box Plot).....	80
7.5 Piirakkakuvi (Pie Chart).....	82
7.6 Tile Chart -kuva	83
7.7 Wizard-kuvatyökalut.....	85
7.8 Kuvien tulostuskohde ja yleisasetukset.....	92
8 Datan tarkastelu Describe- ja Analyze-valikoiden työkaluilla.....	95
8.1 Datan listaukset ja raportit.....	96
8.2 Datan luonnehdinta -työkalu (Characterize Data).....	101
8.3 Jatkuvan muuttujan jakauman tarkastelu	102
8.3.1 Distribution Analysis -työkalu	102
8.3.2 Summary Statistics -työkalu	105
8.3.3 Summary Tables -työkalu	108
8.3.4 Tunnuslukujen taulukointi wizard-työkaluilla	110

8.4	Kategorisen muuttujan jakauman tarkastelu	114
8.5	Kahden muuttujan yhteyden tarkastelu	116
8.5.1	Kahden numeerisen muuttujan välinen yhteys	116
8.5.2	Kahden kategorisen muuttujan välinen yhteys	121
9	SAS-koodin käyttäminen	126
9.1	Koodin kirjoittaminen Enterprise Guidessa tai tuominen muualta	126
9.2	Koodin generoiminen Enterprise Guidessa	126
9.3	Koodin tallentaminen SAS-kooditiedostoon	127
10	Käyttöä helpottavia toimintoja	128
10.1	Työkalussa tehtyjen määritysten tallentaminen ajopohjaksi Create Template -toiminnolla	128
10.2	Käyttäjähöteiden (Prompts) käyttäminen	129
10.3	Prosessidiagrammin haarojen ehdollistaminen (Condition)	137
11	Tulostuksen ja projektin hallintaa	141
11.1	Tulostusasetukset	141
11.1.1	Tulostuskohteen valinta	141
11.1.2	Tulostustyylin valinta	142
11.1.3	Muut tulostusasetukset	144
11.2	Yksittäisen tulosteen hallinta ja jatkokäyttö	144
11.2.1	Kohteen ja tyylin vaihtaminen yksittäiselle tulosteelle	144
11.2.2	Yksittäisen tulosteen tallentaminen tiedostoon	146
11.2.3	Tulosteen tai datan siirtäminen toiseen ohjelmaan	146
11.2.4	Valmiin kuvan muokkaaminen	147
11.2.5	Käytettävän datan vaihtaminen ja rivien valinta tietyissä työkaluissa	148
11.2.6	Selitetekstien ja formaattien lisääminen yksittäiselle tulosteelle	149
11.2.7	Otsikoiden ja alaviitteiden muokkaaminen	149
11.3	Yksittäisten tulosteiden kokoaminen yhteen	150
11.3.1	SAS-raportti	150
11.3.2	HTML-dokumentti	152
11.4	Projektin hallinta	153
11.4.1	Projektin perustoiminnot	153
11.4.2	Projektin pitäminen järjestyksessä	153
11.4.3	Tehtäväsolmujen käsittely	155
11.5	Virhetilanteet	156
	Kirjallisuus	157

1 Johdanto

SAS Enterprise Guide (EG) on graafinen 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 EG:ssä 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 4.2 yhdessä SAS 9.2:n kanssa. EG toimii ainoastaan Windows-ympäristössä, ja tässä oppaassa rajoitutaan tilanteeseen, jossa molemmat – EG ja perus-SAS – ovat asennettu samalle paikalliselle tietokoneelle (local). Oppaan kirjoittajalla on ollut käytössä Microsoft Office 2003 -versio, joka saattaa epäsuorasti vaikuttaa tiettyihin oppaassa kuvattuihin EG:n toimintoihin. Lisäksi on syytä mainita, että opasta tehtäessä käytössä ei ole ollut SASin ”Add-In for Microsoft Office” -tuotetta, joka toisi mukanaan lisätoimintoja Microsoftin ohjelmien ja SASin yhteiskäyttöön.

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

Tämä opas on suunnattu ensisijaisesti maa- ja elintarviketalouden tutkijoille ja muulle tutkimushenkilökunnalle. Lähestymistapa on suunniteltu niin, että materiaali soveltuu erityisesti aloitteleville SAS Enterprise Guiden käyttäjille. Toivon mukaan opas on kuitenkin käyttökelpoinen myös muilla aloilla toimiville EG:n käyttäjille sekä hieman kokeneemmillekin käyttäjille. Opas perustuu useiden Enterprise Guide peruskurssien kurssimateriaaleihin. Tämän oppaan rakenne noudattaa Hurmeen (2007) aiemman oppaan rakennetta, mutta sisältö on muokattu SAS Enterprise Guiden uuden 4.2 version mukaiseksi.

Tähän oppaaseen on lisätty eri työkalujen yhteyteen vaihtoehtoiset SAS-koodiratkaisut. SAS-koodin avulla voidaan tiettyjä toimintoja suorittaa helpommin ja nopeammin, varsinkin jos vaaditaan rutiininomaista työvaiheen toistoa. Kussakin koodiratkaisuissa on pyritty esittämään yksi yksinkertainen tapa tehtävän ratkaisemiseen. Koodiratkaisun tarkoituksena on toimia muistin virkistäjänä perustilanteissa ja niinpä kaikkia yksityiskohtia ei ole tyhjentävästi selitetty. Ongelmatilanteissa kannattaa avata EG:ssä ”Help” -> ”SAS Syntax Help” ja hakea avautuvasta laajasta SAS-manuaalista lisätietoa koodin yksityiskohdista. Tätä EG-opasta kattavampi esitys SAS-koodin käytöstä löytyy muun muassa Marko Laineen (2000) SAS peruskurssi -oppaan verkkopainoksesta.

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.

Haluttaessa tutustua laajemmin tilastollisten menetelmien soveltamiseen SAS Enterprise Guidella, kannattaa tutustua Helienuksen ja Vahlbergin (2008) kirjaan, jossa on esitetty tilastomenetelmien soveltamisesimerkkejä EG:n versiolla 4.1. Englanninkielisistä kirjoista Slaughterin ja Delwichen (mm. 2006) kirjat tarjoavat hyvän johdannon EG:n peruskäyttöön.

2 Oppaassa käytettävät datat

Kaikki kolme oppaassa käytettävää aineistoa löytyvät internetistä (<http://www.mtt.fi/mttraportti/pdf/egdata/EG42datat.zip>). Aineistoja tarvitaan oppaan harjoitus-tehtäviä tehtäessä, mutta harjoitusmateriaalina voi vaihtoehtoisesti käyttää myös muita vastaavan tyyppisiä aineistoja. Käytettävät aineistot kannattaa tallentaa omalle tietokoneelle esimerkiksi ”Omien tiedostojen” (tai ”My Documents”) alle alikansioon \SAS\EGKURS, jota käytetään oppaan esimerkkien tallennuspaikkana tai esimerkiksi kansioon C:\SAS\EGKURS.

Ohra-aineisto

Ohra-aineisto on simuloitu kuvitteellinen osaruutukokeen tulosaineisto. Kokeessa on tutkittu ohran satoa eri lajikkeilla ja lannoitetasoilla. Pääruututekijä on lannoitetaso, jonka luokat (kolme lannoitetasoa) on arvottu pääruutuihin satunnaistettujen täydellisten lohkojen koekaavion mukaisesti. Lohkoja kokeessa on neljä. Osaruututekijä on ohralajike, jonka luokat (neljä eri lajiketta) on täydellisesti satunnaistettu osaruutuihin erikseen jokaisen pääruudun sisällä. Aineisto on tallennettu Excelissä nimellä **OhraData.xls**. Aineistosta löytyvät seuraavat muuttujat:

- lohko:
 - 1
 - 2
 - 3
 - 4
- lajike:
 - A=”Arvo”
 - B=”Bertta”
 - C=”Cecile”
 - D=”Daavid”
- lannoite (kg/ha):
 - 280
 - 300
 - 320
- sato (kg/ha)

Viljelijä 2006 -aineisto

Viljelijä 2006 -aineisto perustuu kuvitteelliseen kyselytutkimukseen, jossa 54 tilanomistajalta on kysytty erilaisia maatalaan ja itse viljelijään liittyviä kysymyksiä. Aineisto on tallennettu Excelissä nimellä **Viljelijä2006.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)
- metsää (metsää omistuksessa, ha)
- tuotsuun (päätuotantosuurat):
 - 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"

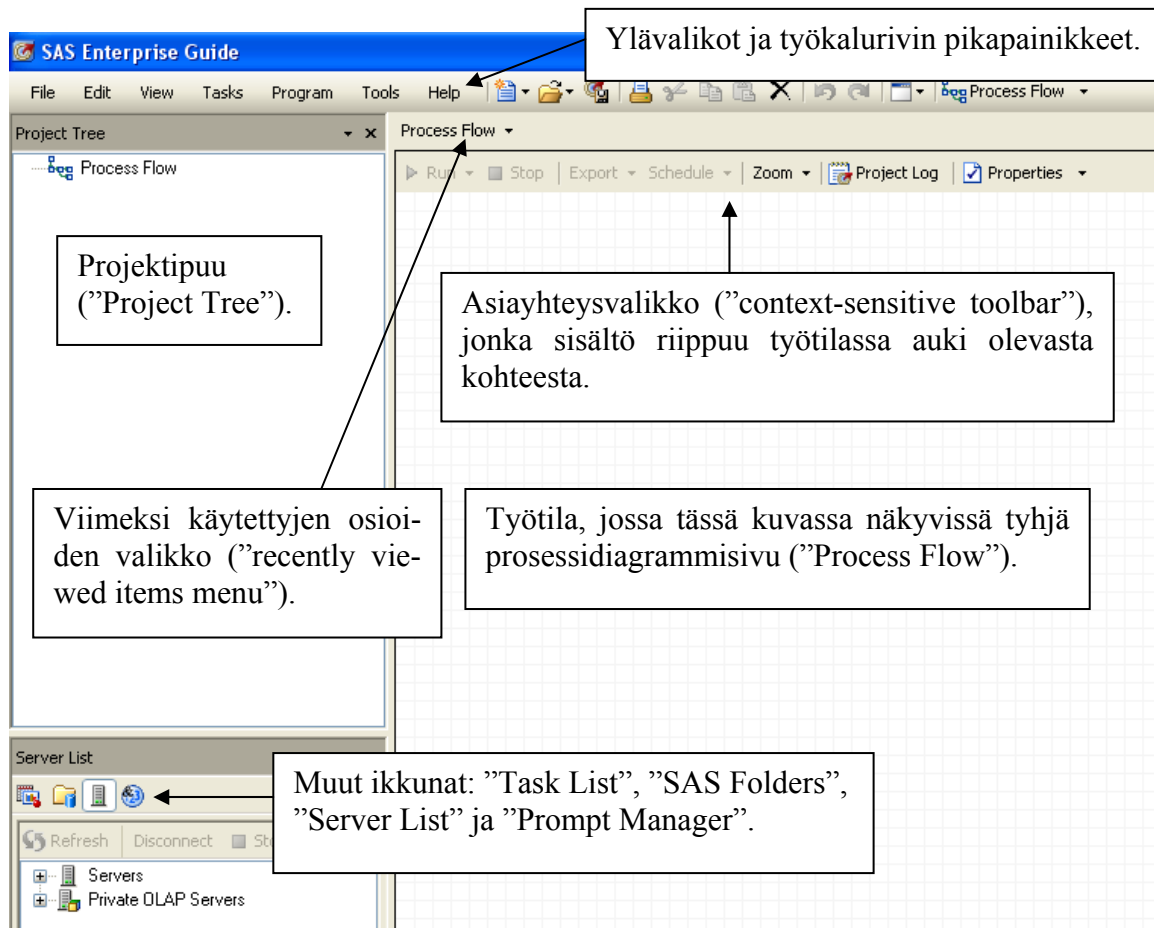
Eläinten painot -aineisto

Eläinten painot -aineisto on kuvitteellinen kokeellinen toistomittausaineisto. Kymmenen saman pentueen kettua on heinäkuussa tapahtuneen vieroituksen yhteydessä täydellisesti satunnaistettu kahteen eri käsittelyryhmään. Kettujen painon kehitystä on seurattu elokuusta joulukuuhun neljänä eri ajankohtana. Aineisto on tallennettu SASissa nimellä **KetutData.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)

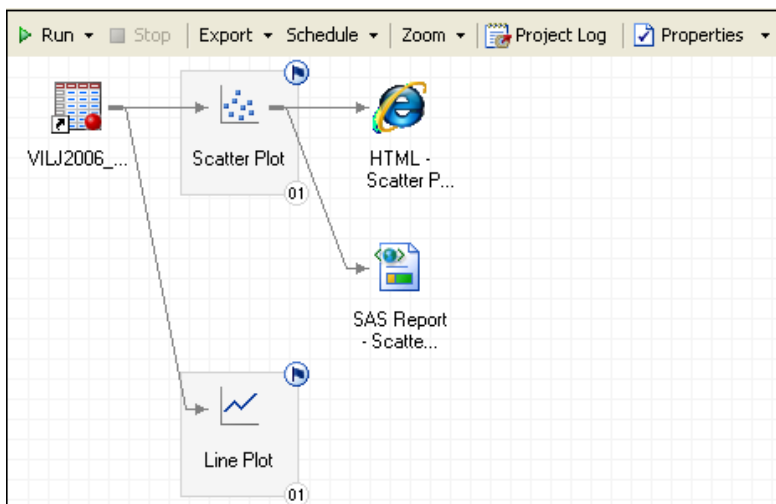
3 Enterprise Guiden käyttöympäristö

Avattaessa Enterprise Guide, perusnäky on olennaisilta osiltaan seuraava:

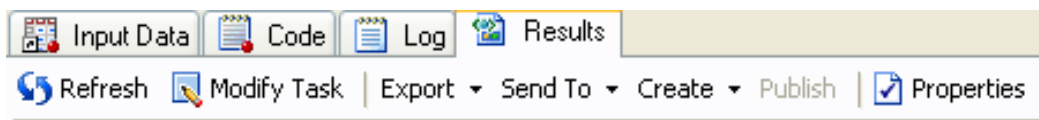


Perusnäky sisältää työtilan, johon avautuvat muun muassa tarkasteltavat aineistot ja suoritettujen tehtävien tulosteet. Työtilassa näkyy myös prosessidiagrammi (työkalurivin "Process Flow" -pikapainikkeesta), joka palvelee samaa tarkoitusta kuin projektipuu. Molempiin kertyy istunnon aikana tehdyistä tehtävistä jälki, joka voidaan tallentaa ja ottaa käyttöön uudessa istunnossa. Jälki eli EG-projektikokonaisuus sisältää erillisiä kuvakkeita eli solmuja (pääasiassa tehtäväsolmuja ja tulossolmuja) suoritetuista toimenpiteistä. Työtilassa on myös koko ajan näkyvässä niin sanottu asiayhteysvalikko, josta valittavissa olevat toimenpiteet riippuvat siitä mitä työtilassa on samalla hetkellä aktiivisena tarkastelussa.

Alla olevassa kuvassa näkyy esimerkkinä prosessidiagrammijälkeä ja siihen liittyvä asiayhteysvalikko:



Yksittäiseen tehtävään liittyvät käytetyt datat, koodit, logit ja tuloksena syntyneet datat ja/tai tulosteet on ryhmitelty esitettäväksi yhteen työtilaan avautuvaan välilehdille jaoteltuun näkymään. Tuplaklikkaamalla aiemmin määriteltyä ja suoritettua tehtäväsolmua prosessidiagrammista saa esille kyseisen välilehtinäkömän:



Välilehtien välillä selailemalla, voidaan helposti ja nopeasti tutkia käytettyä dataa ("Input Data"), tehtävän luomaa SAS-koodia ("Code"), suorituksen yhteydessä muodostettua logia ("Log") ja tehtävän tulostetta ("Results"). Aiemmin määriteltyjä tehtäväsolmuja pääsee muokkaamaan avaamalla ne yllä olevan kuvan näkymään (tuplaklikkaamalla prosessidiagrammista) ja valitsemalla asiayhteysvalikosta "Modify Task". Yksittäiseen tehtäväsolmuun liittyvistä jatkotoimenpiteistä saa valikon esiin myös klikkaamalla tehtäväsolmua hiiren oikealla painikkeella suoraan prosessidiagrammissa.

Edellä mainittujen tehtävä- ja tulossolmujen sekä niihin liittyvien SAS-koodien ja logien lisäksi Enterprise Guiden projektikokonaisuus pitää sisällään muun muassa datoihin viittaavia kuvakkeita ja muistiinpanoja toisiinsa linkitettyinä. Projektikokonaisuutta voi myös jakaa useammalle erilliselle prosessidiagrammisivulle (Process Flows). Uusia välilehtiä saa käyttöön valitsemalla "File" -> "New" -> "Process Flow". Erillisten prosessidiagrammisivujen ja muiden osioiden (muun muassa "Custom Reports") välillä voi selailta joko työkalurivin viimeisen painikkeen (edellisen sivun kuvan "Process Flow" -painike) tai viimeksi käytettyjen osioiden valikon kautta. Lisäksi osioiden välillä on kätevä liikkua käyttäen projektipuuta.

Työtilassa on oletuksena samanaikaisesti näkyvissä vain yksi tarkastelun kohde (esimerkiksi prosessidiagrammi, data tai tuloste). Useampaa kohdetta voidaan tarkastella samanaikaisesti valitsemalla työkalurivin "Workspace Layout" -pikapainikkeesta vaihtoehto "Stacked" tai "Side By Side", joiden avulla saadaan samanaikaisesti näkyviin kaksi kohdetta joko päällekkäin tai rinnakkain. Tämän avulla voidaan esimerkiksi käyttää Enterprise Guidea siten, että prosessidiagrammi on koko ajan näkyvissä toisessa ikkunassa ja valintoja tehtäessä niiden tulokset aukeavat toiseen ikkunaan.

Suoritettavien tehtävien eteneminen näkyy Enterprise Guide 4.2:ssa alapalkissa suorituksen ajan. Tällöin se voidaan myös klikata näkyviin suurempaan ikkunaan valitsemalla ”(Details)”. Aineistojen käsittelyyn ja niistä tehtävään tarkasteluun liittyvät varsinaiset SAS Enterprise Guide -työkalut löytyvät sekä asiayhteysvalikon, tehtävälistan (”Task List”) että ”Tasks”-ylävalikon kautta. Tässä oppaassa tehtävät suoritetaan pääasiassa valitsemalla työkalut käyttöön asiayhteysvalikon kautta. Ylävalikoista puolestaan 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.
- Tasks:** Tasks-valikkoon on koottu kaikki aineistojen käsittelyyn ja tarkasteluun liittyvät SAS Enterprise Guide -työkalut. Valikosta löytyvät aineistojen muokkaamiseen, aineistojen kuvailevaan tarkasteluun, kuvien laatimiseen ja tilastollisiin analyyseihin liittyvät työkalut.
- Program:** Program-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.
- 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 ”View”-valikosta. Sen avulla työtila 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 valitsemalla ”Auto Hide” valikosta, joka saadaan näkyviin klikkaamalla ×-symbolin vieressä olevaa alaspäin osoittavaa nuolta. Tällöin ikkuna menee samanlaisen pikalinkin taakse kuin ”Maximize Workspace” -toiminnossa. Saman valikon kautta voidaan tehdä myös muuta käyttöympäristön järjestelyä, muun muassa siirtää ikkunoiden sijaintia ruudulla.

Ikkunoiden tai ikkunaryppäiden kokoa ja muotoa voi muokata tarttumalla niiden reunoista kiinni ja venyttämällä 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 ”Restore Window Layout”. Yksittäisen suljetun ikkunan saa palautettua ”View”-valikosta.

Enterprise Guide tunnistaa automaattisesti erityyppisiä tiedostoja. Enterprise Guideen voidaan avata aiemmin luotuja Enterprise Guide -projektitiedostoja, joita ovat *.egp-tiedostot (EG versiot 3, 4.1 ja 4.2) ja *.seg-tiedostot (EG versiot 1 ja 2). Avattaessa EG 4.2:een EG 3.0:ssa tai 4.1:ssä tehty projektitiedosto, ohjelma kuitenkin huomauttaa, samasta tiedostotyyppistä huolimatta, että se on tehnyt projektitiedostoon tiettyjä muutoksia, jotka lopullisesti hyväksytään tallentamalla projekti EG 4.2:ssä. Muutosten tallennuksen jälkeen projektitiedostoa ei voi enää käyttää EG:n vanhemmissa versioissa, joten tallennus kannattaa yleensä tehdä uudella nimellä, jolloin alkupe-
räinen vanhemmassa EG:n versiossa luotu projektitiedosto säilyy entisellään.

Enterprise Guide tunnistaa monia erityyppisiä datatiedostoja. EG:ssä voidaan ottaa käyttöön muun muassa SAS-datat (*.sas7bdat ja *.sd2), Excel-datat (*.xls, *.xlsx, *.xslm ja *.xlsb) sekä monen muun tyyppiset datatiedostot (*.dbf, *.db, *.txt, *.csv, *.asc, *.tab). Uutena ominaisuutena Enterprise Guide 4.2:een voidaan lukea sisään myös JMP (*.jmp), SPSS (*.sav) ja Stata

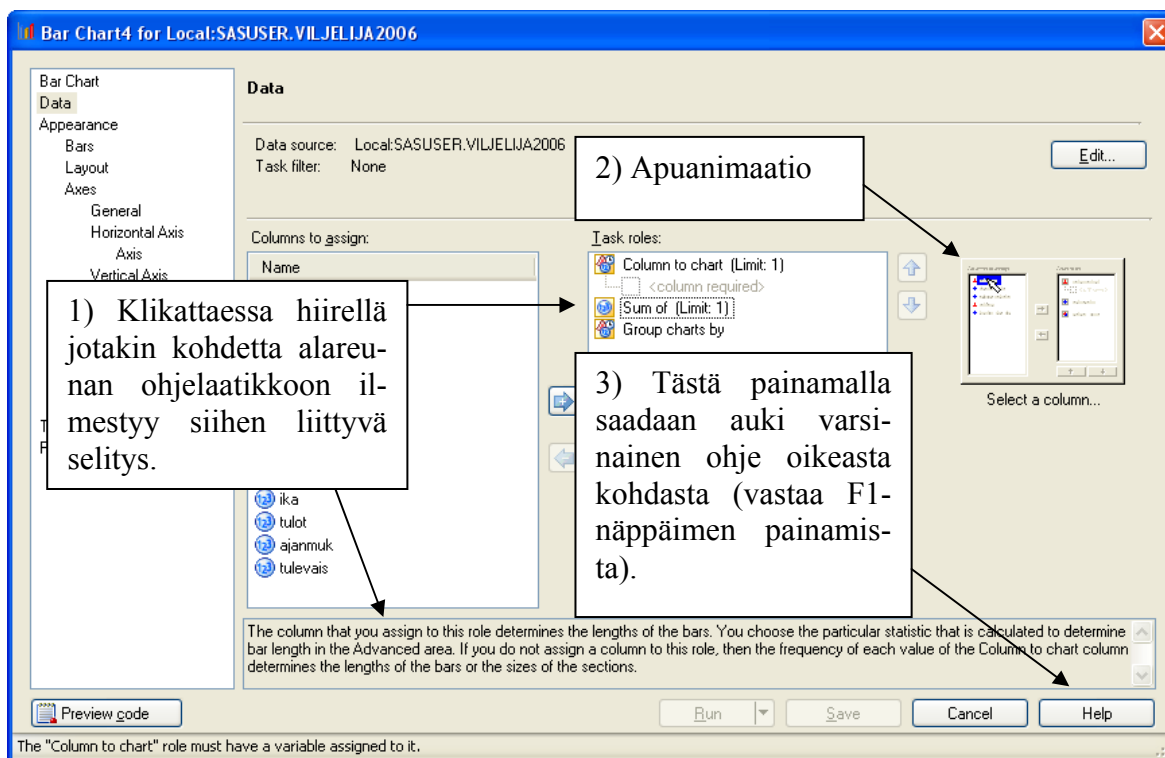
(* .dta) -ohjelmien datatiedostoja omilla erillisillä työkaluillaan. Erityyppisten data-tiedostojen käyttöönotosta kerrotaan tarkemmin luvussa **Aineiston käyttöönotto**. Lisäksi Enterprise Guiden voidaan avata muun muassa SAS-kooditiedostoja (*.sas), SAS-logitiedostoja (*.log) sekä SAS-raportti -tiedostoja (*.srx). Täydellisen listan Enterprise Guiden tunnistamista tiedostotyypeistä näkee esimerkiksi klikkaamalla: "File" -> "Open" -> "Other" ja avaamalla tiedostotyyppi-listan. Samasta "File" -> "Open"-valikosta löytyy myös omat kohdat projektitiedostojen ("Project"), datojen ("Data") ja SAS-kooditiedostojen ("Code") avaamiseen.

Tehtävä: Käynnistä Enterprise Guide pikakuvakkeesta. Selaa valikoita, ikkunoita ja välilehtiä saadaksesi alustavan kuvan Enterprise Guiden käyttöympäristöstä. Harjoittele myös ikkunoiden muokkausta. Lisäksi voit harjoitella "Restore Window Layout" -toiminnon käyttöä, jos ikkunat menevät epäjärjestykseen tai ovat edellisen käyttäjän jäljiltä sekavassa 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 järjestettynä selkeiden otsikoiden alle, joten se on varsin käyttökelpoinen ja toimiva. Esimerkiksi eri analyysityökaluihin löytyy apua klikkaamalla ohjeessa Sisällys-välilehdeltä auki kansio ”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. Enterprise Guiden ohjeen lisäksi on usein tarpeen käyttää myös tavallisen SASin ohjetta. Sen löytää kätevästi EG:n ”Help”-valikosta valitsemalla ”SAS Syntax Help”. Tämä ohje on paljon täsmällisempi ja laajempi yksityiskohtien suhteen kuin Enterprise Guiden ohje.

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



Aiemmin tavallista SASia käyttäneille EG:n hahmottamista saattaa helpottaa tieto Enterprise Guiden työkalujen taustalla käyttämissä SAS-proseduureista. Valitsemalla ”Task list” -ikkunassa ”Tasks by Name”, saadaan kunkin työkalun viereen näkyviin maininta sen käyttämissä SAS-proseduureista.

Lisäksi ”Help”-valikon kohdasta ”SAS on the Web” löytyy monta hyödyllistä linkkiä internet-sivustoille. Siellä on muun muassa linkkejä, joiden takaa löytyy lisämateriaalia EG:n käytön helpottamiseksi ja harjoittelemiseksi sekä erilaista teknistä tietoa (muun muassa linkki EG:n hotfix-korjaustiedostoihin).

5 Aineiston käyttöönnotto

5.1 Kirjastoviite

Kirjastoviite on SAS-ohjelmiston käyttämä menettelytapa, jossa avainsanalla viitataan tiettyyn hakemistoon tai kansioon tietokoneella. Sitä tarvitaan esimerkiksi, kun halutaan määrittellä mihin tietokoneen kansioon istunnon aikana luodut SAS-datat tallennetaan tai mistä kansioista aiemmin luodut SAS-datat otetaan käyttöön.

Kirjastoviitteen luominen

Tässä esitettävä työkalu tuottaa kirjastoviitteen, joka on voimassa EG-istunnon ajan. Kun Enterprise Guide tai projekti käynnistetään uudelleen, pitää kirjastoviite muistaa istunnon aluksi yhdistää uudelleen haluttuun kansioon suorittamalla kirjastoviitteen määrittelevä tehtäväsolmu uudelleen.

Kirjastoviite luodaan käyttäen siihen tarkoitettua Enterprise Guiden työkalua klikkaamalla **”Tools” -> ”Assign Project Library...”**. Aukeavassa ikkunassa määritellään vaihe kerrallaan kirjastoviite.

Ensimmäisellä sivulla annetaan kirjastoviitteelle nimi (esimerkiksi kir):

Assign Project Library

1 of 4 Specify a name and server for the library.

This wizard assigns a library for the current project and the current user. This library will be un-assigned when you exit the application. If you want to create a permanent library on a server, create a server library definition by going to menu Tools-> Enterprise Guide Explorer.

Name (enter 8 or fewer characters):
KIR

Server:
Local - The SAS server on your machine

Kirjastoviitteelle annetaan nimi kohdassa "Name".

Kohdassa "Server" valitaan SAS-palvelin, jossa kirjastoviite luodaan. Tässä oppaassa käytetään omalla tietokoneella olevaa SASia ("Local").

<Back Next> Finish Cancel Help

Toisella sivulla valitaan kansio, johon kirjastoviite viittaa, allaolevassa kuvassa esimerkki viittaamisesta ”Omien tiedostojen” alla olevaan \SAS\EGKURS -kansioon. (”Engine” puolestaan saa olla oletuksena oleva ”BASE”, kun kirjastoviitettä käytetään SAS-datojen käsittelyyn):



Loput sivut voi yleensä klikata ohi painamalla ”Next” ja viimeisellä sivulla valitaan ”Finish”.

Vaihtoehtoinen ratkaisu SAS-koodilla:

Istunnon ajan voimassa oleva kirjastoviite voidaan toteuttaa myös SAS-koodin avulla. Koodieditorin saa avattua klikkaamalla ”File” -> ”New” -> ”Program”. Tässä koodiikkunassa voi määrittää kirjastoviitteen samoin kuin perus-SASissa, esimerkiksi:

```
LIBNAME kir 'C:\tahanpolkua\sas\egkurs';
```

Siinä kir on kirjastoviitteen nimi ja yksinkertaisten lainausmerkkien sisällä on hakemistopolku, jonka osoittamaan kansioon kirjastoviite viittaa. Koodin voi suorittaa muun muassa valitsemalla ”Run” asiayhteysvalikosta tai klikkaamalla koodisolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Run Program on Local”. Huomaa, että suorittaessa sama tehtävä sekä valikkotyökalulla että koodin avulla, aiemmin luodun tehtävän tulokset saattavat korvaantua jälkimmäisen tuloksilla!

SAS-koodia hyödyntäen voidaan luoda myös kirjastoviite, joka yhdistetään automaattisesti aina EG:n käynnistyksen yhteydessä. Se onnistuu kohdassa ”Tools” -> ”Options” -> ”SAS Programs” ruksaamalla vaihtoehto ”Submit SAS code when server is connected” ja lisäämällä kohdan perässä olevan ”Edit”-painikkeen kautta aukeavaan koodi-ikkunaan kirjastoviitteen luova SAS-koodi.

Tehtävä: Tee tietokoneelle ”Omien tiedostojen” alle kansio \SAS\EGKURS ja kopioi sinne tähän oppaaseen liittyvät datatiedostot, jotka löytyvät zip-pakattuina internetistä osoitteesta: <http://www.mtt.fi/mtrraportti/pdf/egdata/EG42datat.zip>. Luo istunnon ajan voimassa oleva kirjastoviite nimeltä kir, joka viittaa kyseiseen kansioon.

Kirjastoviitteen asettaminen oletukseksi

Oletuskansio, johon eri tehtäväsolmut ehdottavat datoja 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. Uuden kirjastoviitteen saa lisättyä listaan ”Add”-valinnalla. Kunkin projektin datat 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 ensisijaiseksi kirjastoviitteeksi ja seurauksena EG tarjoaa oletusarvoisesti datoja tallennettavaksi käyttäjän projektille tarkoittamaan kansioon tietokoneella.

Tehtävä: Aseta edellisessä tehtävässä luotu kirjastoviite kir oletuskirjastoviitteeksi.

5.2 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:

LAJIKE	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 Guidessa:

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:

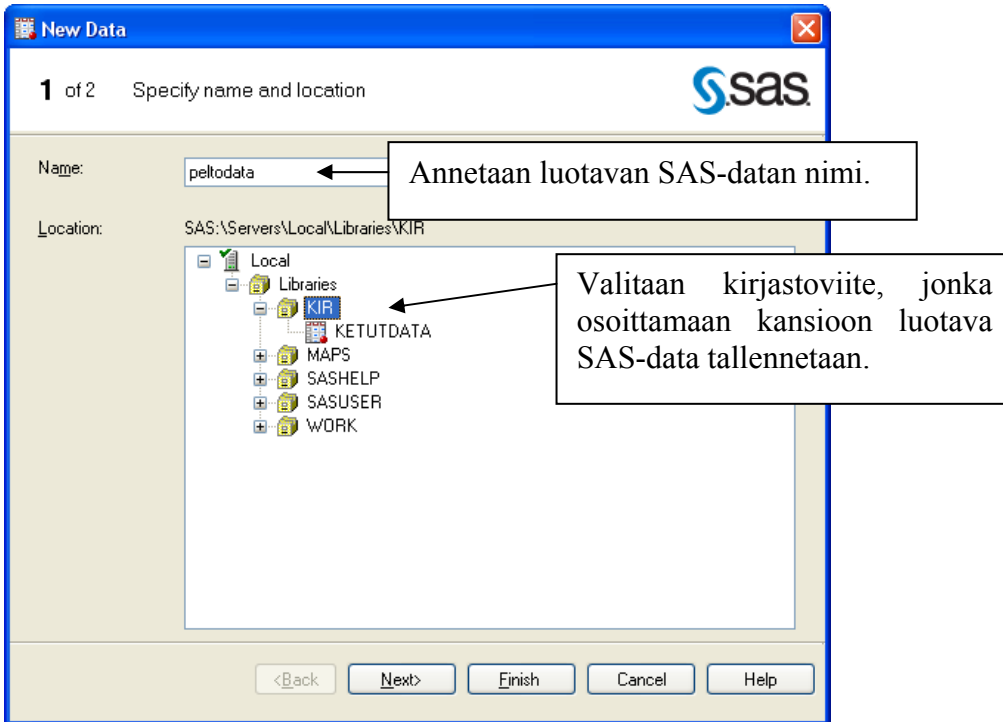
HENKILO	KAS	PAINO_1	PAINO_2	PAINO_3
Pekka	TavPuuro	87.2	87.1	87.4
Paavo	PuuroX	72.1	83.2	85.6
Matti	TavPuuro	99.1	99.8	100.2
Antti	PuuroX	78.1	83.2	86.5
Timo	PuuroX	95.1	103.2	114.2
Keijo	TavPuuro	104.2	102.1	99.2

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 alleviivausmerkkiä (_), ei siis muita erikoismerkkejä eikä välilyöntejä. Muuttujien nimiä ei myöskään kannata aloittaa numeroilla. Poikkeukset edellä mainituista käytännöistä muuttujien nimeämisessä saattavat toimia hyvin EG 4.2:ssa, mutta luotujen datojen käyttäminen esimerkiksi SAS 9.2:ssa voi jo aiheuttaa ongelmia.

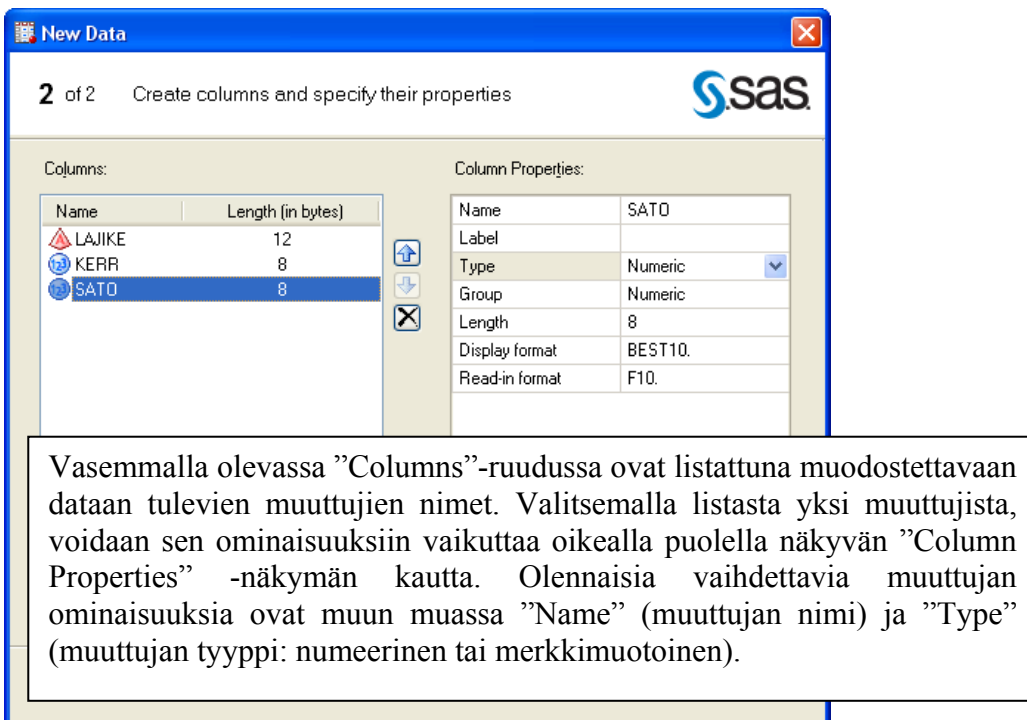
5.3 Aineiston syöttäminen Enterprise Guidessa

Uuden datamatriisin avaaminen ja tallentaminen

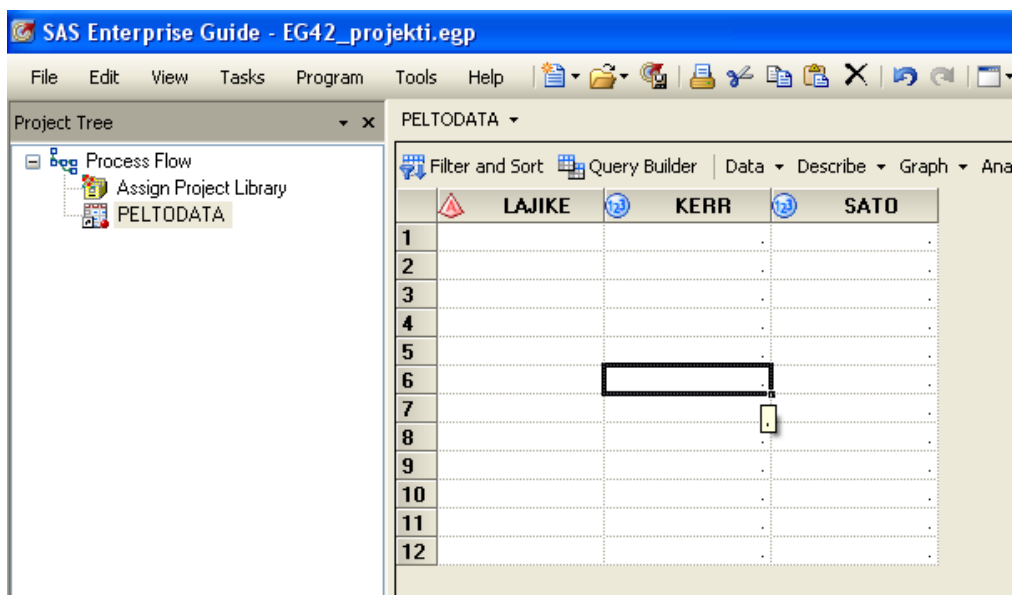
Aineisto voidaan muodostaa syöttämällä se data-matriisiin Enterprise Guidessa. Tämä tapa on hyvin samanlainen kuin datan syöttäminen käsin Excelissä. Uusi data-matriisi saadaan ruudulle valitsemalla yläreunan valikoista ”File” -> ”New” -> ”Data”. Tällöin aukeaa kaksivaiheinen työkalu, jonka ensimmäisessä vaiheessa annetaan muodostettavalle datalle nimi ja valitaan tallennuksen kohdekansio kirjastoviitteen avulla:



Klikkaamalla ”Next” päästään seuraavalle sivulle, jossa voidaan antaa datamatriisin muuttujien nimet ja ominaisuudet:



Lopuksi klikataan ”Finish”, jolloin työtilaan aukeaa tyhjä datamatriisi, johon voidaan syöttää muuttujien arvot:



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 kirjoitus suojaus. Tämä tehdään, datamatriisin ollessa näkyvässä työtilassa, valitsemalla ylävalikko ”Edit” ja poistamalla mahdollinen valinta kohdasta ”Protect Data”. Jos muokattava data on juuri luotu uusi datamatriisi, se on yleensä jo valmiiksi muokkaustilassa.

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 jonkin rivinumeron 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”).

Uusi datamatriisiin lisätty sisältö tallentuu reaaliaikaisesti ja automaattisesti datamatriisin luonnin yhteydessä määritettyyn erilliseen SAS-datatiedostoon, joka näkyy ja on käytettävissä EG:ssä.

Vaihtoehtoinen ratkaisu SAS-koodilla:

Uusi data voidaan syöttää myös käyttäen SAS-koodia. Tämä voidaan tehdä esimerkiksi seuraavalla tavalla DATA-vaiheessa:

```
DATA kir.peltodata;  
  INPUT lajike $ lohko sato;  
  CARDS;  
  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  
  ;  
RUN;
```

DATA-lauseessa annetaan uuden muodostettavan SAS-datan nimi ja tallennuskohde. Tässä esimerkissä uudelle SAS-datalle on annettu nimeksi peltodata ja se tallennetaan kir-nimisen kirjastoviitteen osoittamaan kansioon. INPUT-lauseessa puolestaan luetellaan tallennettavien muuttujien nimet ja \$-merkillä osoitetaan sitä edeltävän muuttujan olevan merkkimuotoinen (oletuksena olevan numeerisen sijaan). Muuttujien arvot annetaan esimerkin mukaisesti CARDS;-komennon perään. DATA-vaihe päättyy RUN;-komentoon.

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

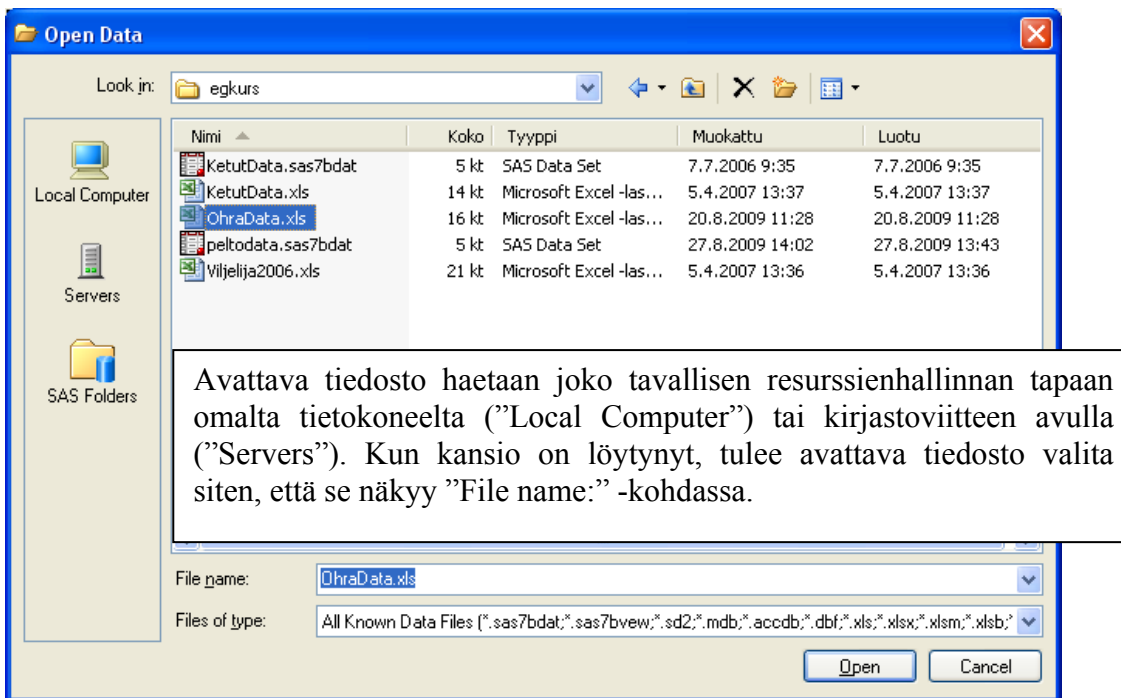
LAJIKE	KERR	SATO
A	1	6500
A	2	5850
A	3	6250
B	1	5025
B	2	4520
B	3	5260
C	1	4520
C	2	3950
C	3	4310

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

5.4 Excel-tiedoston avaaminen Enterprise Guideen

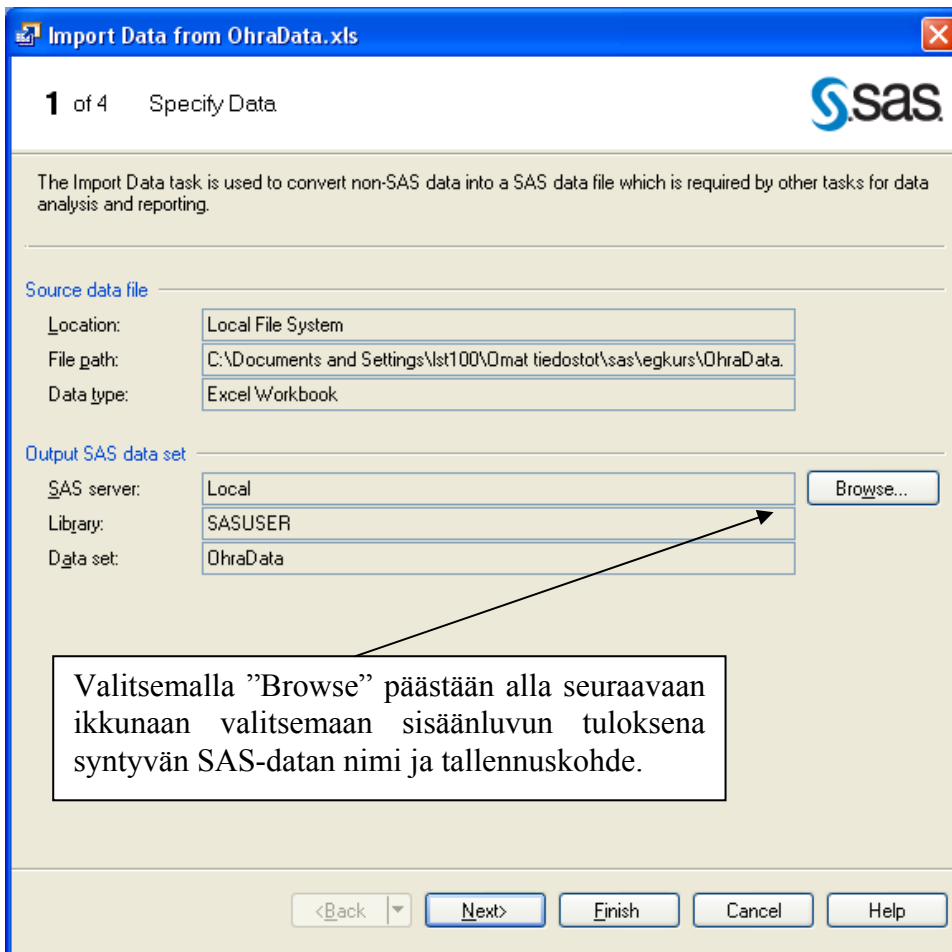
Enterprise Guideen voidaan avata monen eri Excel-version tiedostoja. Tässä oppaassa Excel-tiedostolla tarkoitetaan pääasiassa *.xls -muotoista Excel 2003:n tiedostoa, sillä opas on kirjoitettu käyttäen Microsoft Officen 2003 -versiota yhdessä EG 4.2:n kanssa. Kun Excel-tiedosto avataan Enterprise Guide -projektiin, taustalla Excel-tiedostosta tehdään SAS-datatiedosto, joka tallennetaan erillisenä tiedostona jonkin kirjastoviitteen osoittamaan kansioon. EG-projektissa näkyvä SAS-datakuvake on siis todellisuudessa vain viittaus ulkoiseen SAS-datatiedostoon. Näin ollen varsinainen data ei tallennu projektin mukana, mutta niin kauan kuin data on saatavilla viittauksen osoittamasta kohteesta, dataa voi tarkastella ja käyttää EG:ssä.

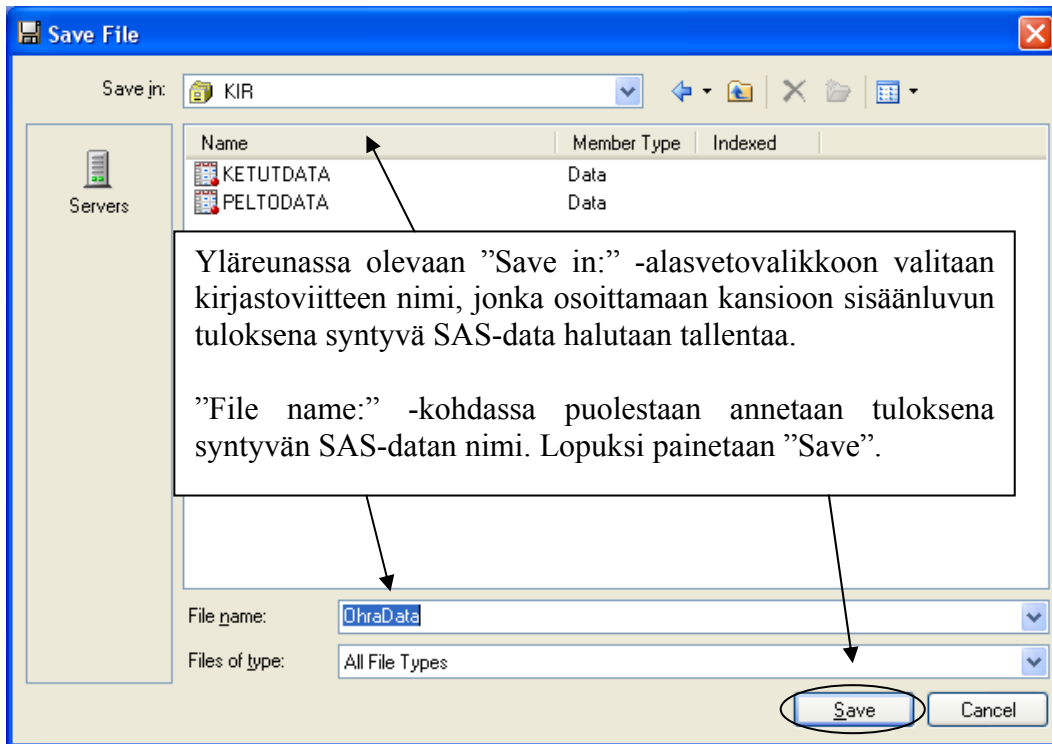
Excel-tiedoston käyttöönotto tapahtuu klikkaamalla ”File” -> ”Open” -> ”Data” (tai ”File” -> ”Import Data”), jolloin aukeaa seuraava ruutu:



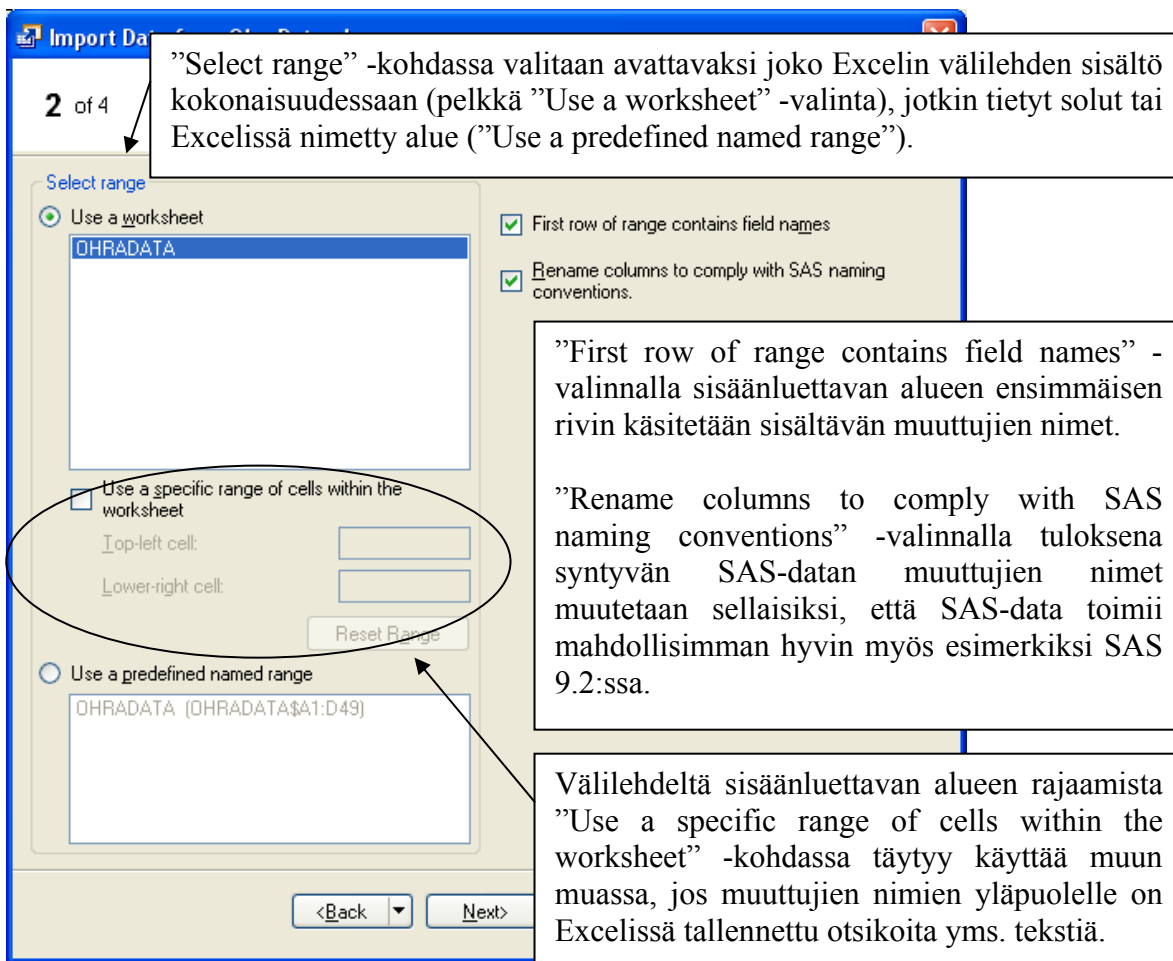
Vaihtoehtoisesti Excel-tiedoston kuvake voidaan yksinkertaisesti raahata Windowsin resurssienhallinnasta EG:n prosessidiagrammi-välilehdelle.

Tämän jälkeen tulee näkyviin seuraava wizard-tyyppinen käyttöliittymä Excel-tiedoston muuttamiseksi SAS-dataksi:





Tämän jälkeen siirrytään Wizard-käyttöliittymässä seuraavalle sivulle valitsemalla "Next", jossa valitaan Excelistä sisäänluettava alue:



Seuraavalla sivulla ("Next") puolestaan tarkastellaan ja muokataan sisäänluettavia sarakkeita (muun muassa muuttujien nimiä, tyyppisiä ja selitetekstejä voidaan muokata):

3 of 4 Define Field Attributes

”Inc” -kohdassa voi valita muuttujan jätettäväksi pois tulosdatasta.

Select columns and define attributes:

Inc	Source Name	Name	Label	Type	Source Informat	Len.	Output Format	Output Informat
<input checked="" type="checkbox"/>	lohko	lohko	lohko	Number	BEST12.	8	F12.	BEST12.
<input checked="" type="checkbox"/>	lajike	lajike	lajike	String	\$CHAR1.	1	\$CHAR1.	\$CHAR1.
<input checked="" type="checkbox"/>	lannoite	lannoite	lannoite	Number	BEST12.	8	F12.	BEST12.
<input checked="" type="checkbox"/>	sato	sato	sato	Number	BEST12.	8	F12.	BEST12.

Informaatteja ("Source Informat") käytetään kertomaan EG:lle, miten sen tulee tulkita sisään luettavien muuttujien arvot. Esimerkiksi, tulkitaanko 1/1/2009 siten, että merkkijonon ensimmäinen luku on päivä ja toinen kuukausi vai päinvastoin. Katso lisätietoa luvusta **Erikoismuuttujien sisänlukeminen ja ongelmatapaukset**.

Formaatilla eli tulostusmuodolla ("Output Format") voidaan määrittellä muuttujan arvojen muotoilu tulosdatassa. Esimerkiksi numeerisen muuttujan näytettävien desimaalien määrän saa määritettyä formaatilla. Katso lisätietoa luvusta **Muuttujan arvojen tulostusmuotojen eli formaattien käyttäminen**.

Select All Clear All Modify... Cancel Help

Modify-valinnalla muokkaukset voidaan tehdä erillisessä editorissa.

Informaatit ja formaatit ovat olennaisimpia erilaisilla erikoismuotoisilla muuttujien arvoilla, kuten päivämäärämuuttujien arvoilla. Niinpä erikoisemmilta muuttujilta kannattaa tarkistaa Enterprise Guiden automaattisesti valitsemat informaatit ja formaatit, jotta tieto tulee varmasti luettua sisään oikein ja esitettyä datamatriisissa ja sen myötä tulosteissa mielekkäällä tavalla. Tavallisimmille päivämäärämuotoisille muuttujille Enterprise Guide 4.2 osaa kuitenkin valita sisäänlukformaatit yleensä automaattisesti oikein.

Viimeisellä sivulla ("Next") on vielä muutama sisäänlukuun liittyvä enemmän tekninen valinta, jotka tavallinen EG-käyttäjä voi yleensä ohittaa:

4 of 4 Advanced Options sas

Embed the data within the generated SAS code.

Import the data using SAS/ACCESS Interface to PC Files whenever possible.

Remove characters that can cause transmission errors from text-based data files.

"Embed the data within the generated SAS code" -valinnalla saadaan sisäänluettavan datan arvot sisällytettyä taustalla suoritettavaan SAS-koodiin mahdollista jatkokäyttöä varten.

"Import the data using SAS/ACCESS Interface to PC Files whenever possible" -valinta tekemällä otetaan käyttöön vaihtoehtoinen tekniikka datan sisään lukemiseksi (PROC IMPORT).

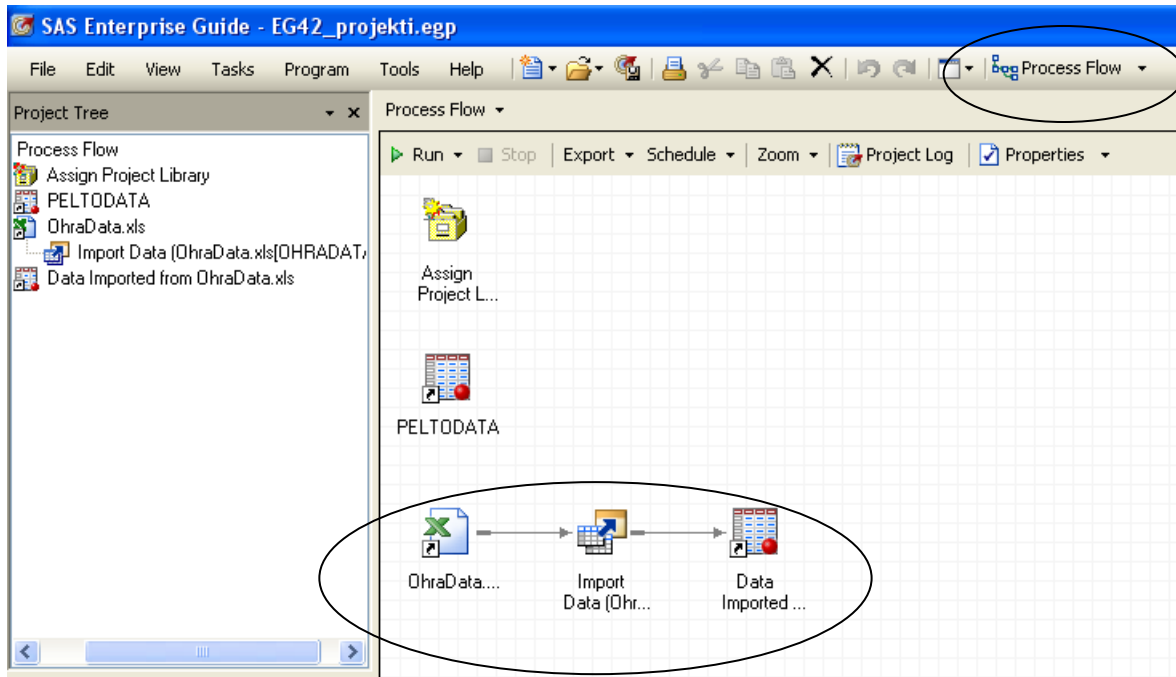
Lisätietoa näistä valinnoista saa muun muassa klikkaamalla "Help".

<Back Next> Finish Cancel Help

Aineiston sisäänluku on nyt määritetty, joten voidaan valita "Finish", jolloin sisäänluettu aineisto tulee näkyviin työtilaan:

	lohko	lajike	lannoite	sato
1	1	A	280	3371
2	1	B	280	3088
3	1	C	280	3259
4	1	D	280	2838
5	1	A	300	3536
6	1	B	300	3106
7	1	C	300	3477
8	1	D	300	2987
9	1	A	320	3211
10	1	B	320	3305
11	1	C	320	3589
12	1	D	320	2962
13	2	A	280	3303
14	2	B	280	3079
15	2	C	280	3188
16	2	D	280	2981
17	2	A	300	3351
18	2	B	300	3255
19	2	C	300	3298
20	2	D	300	3165
21	2	A	320	3596
22	2	B	320	3614

Samalla prosessidiagrammiin muodostuut datasolmut, jotka viittaavat sisäänluettuun Excel-dataan ja tuloksena syntyneeseen SAS-dataan. Lisäksi muodostuu tehtäväsolmu, joka viittaa edellä tehtyihin SAS-datan muodostamiseen liittyviin määrittelyihin ("Import Data" -tehtäväsolmu). Prosessidiagrammi saadaan näkyviin klikkaamalla "Process Flow" -painiketta työkaluriviltä:



Vaihtoehtoinen ratkaisu SAS-koodilla:

Jos käytössä on SAS/ACCESS Interface to PC Files -tuote, voidaan Excel-tiedoston sisäänluku suorittaa seuraavalla SAS-koodilla (koodin yksityiskohtiin kannattaa tutustua tarkemmin valikosta "Help" -> "SAS Syntax Help"):

```
PROC IMPORT DATAFILE='C:\tahanpolku\OhraData.xls' OUT=kir.OhraData
      REPLACE DBMS=EXCEL;
      SHEET="OHRADATA";
      GETNAMES=YES;
RUN;
```

Näin luotu SAS-data pitää avata Enterprise Guidessa vielä erikseen ("File" -> "Open" -> "Data"), jos data halutaan prosessidiagrammiin näkyviin. Koodilla tehtävän jatkokäytön kannalta avaaminen ei kuitenkaan ole välttämätöntä.

Tehtävä: Avaa \SAS\EGKURS -kansioista Excel-data nimeltä **OhraData.xls** ja muodosta siitä pysyvä SAS-data nimeltä **OhraData** kirjastoviitteesi osoittamaan kansioon.

5.5 Erikoismuuttujien sisäänlukeminen ja ongelmatapaukset

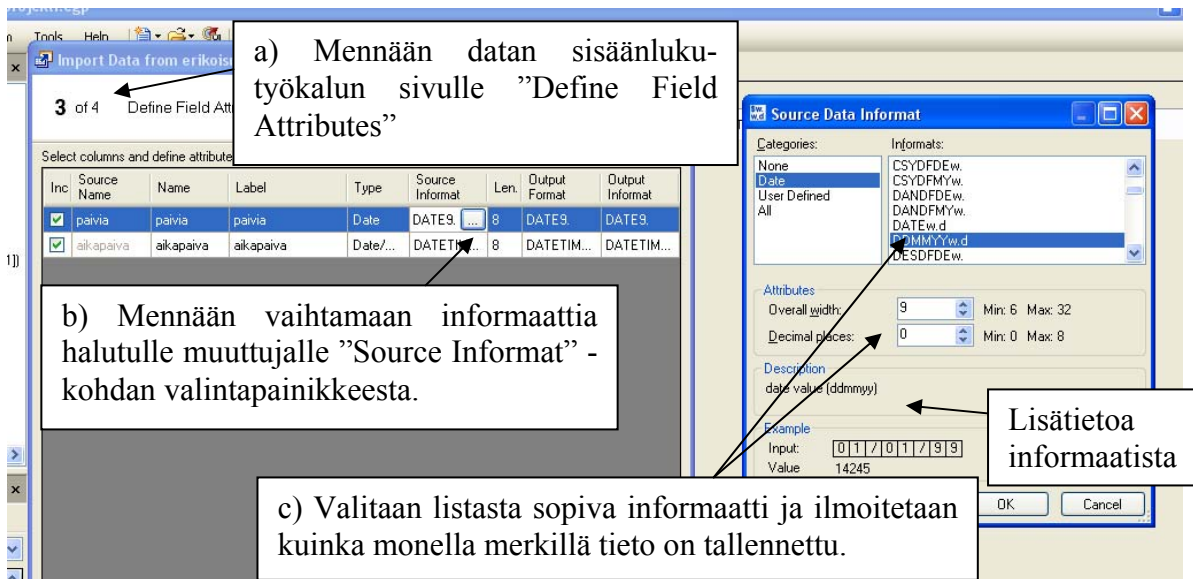
Sisäänlukuformaattit eli informaattit

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 ja aikojen sisäänlukuun. Ongelmien syynä on se, että päivämäärät ja kellonajat voidaan ilmoittaa hyvin monella tavalla ja siten ohjelmakin voi tulkita niitä eri tavoin. Informaattien toimintaperiaatteena on, että niiden avulla kerrotaan ohjelmalle, miten sen tulee tulkita sisäänluettavan datan muuttujien arvot. Samalla voidaan käyttää myös tavallisia formaatteja eli tulostusmuotoja, joiden

avulla puolestaan kerrotaan ohjelmalle, miten sen tulee esittää sisäänlukemiensa muuttujien arvot.

Datan sisäänluvun yhteydessä informaattit ja formaatit vaihdetaan yhdelle muuttujalle kerrallaan alla olevassa kuvassa esitetyllä tavalla:



Päivämäärämuuttujien tapauksessa yleisesti käytetty informaatti/formaatti on DDMMYYw.d. Informaatin nimi kertoo että arvoissa tieto esiintyy järjestyksessä päivä (dd), kuukausi (mm), vuosi (yy) ja kaiken kaikkiaan niin monella merkillä kuin "width"-kohdassa määritellään (esim. 10 merkkiä: dd.mm.yyyy). Formaattia eli muuttujan arvojen tulostusmuotoa voidaan vaihtaa vastaavasti kuin informaattia. Sisäänluvun yhteydessä se tapahtuu kohdassa "Output format", jonka kautta päästään selaamaan vastaavaa formaattilistaa kuin informaattien määrittelyn yhteydessä. Yleisimpiä informaatteja ovat edellä mainitun DDMMYYw.d:n lisäksi muun muassa ANYDTEw. (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 lisätietoa informaattien description-kohdista). Erityyppisille muuttujille on valittavissa lista tilanteeseen soveltuvia informaatteja/formaatteja, joten kaikki edellä luetellut informaattit/formaatit eivät ole näkyvissä kaikkien muuttujien yhteydessä. Lisätietoa informaateista löytyy SASin ohjeesta ("Help") ja muun muassa Slaughterin ja Delwichen (2006) kirjasta, josta löytyy kattavampi lista yleisimmistä informaateista.

Erilaiset ongelmatilanteet Excel-datan sisäänluvussa

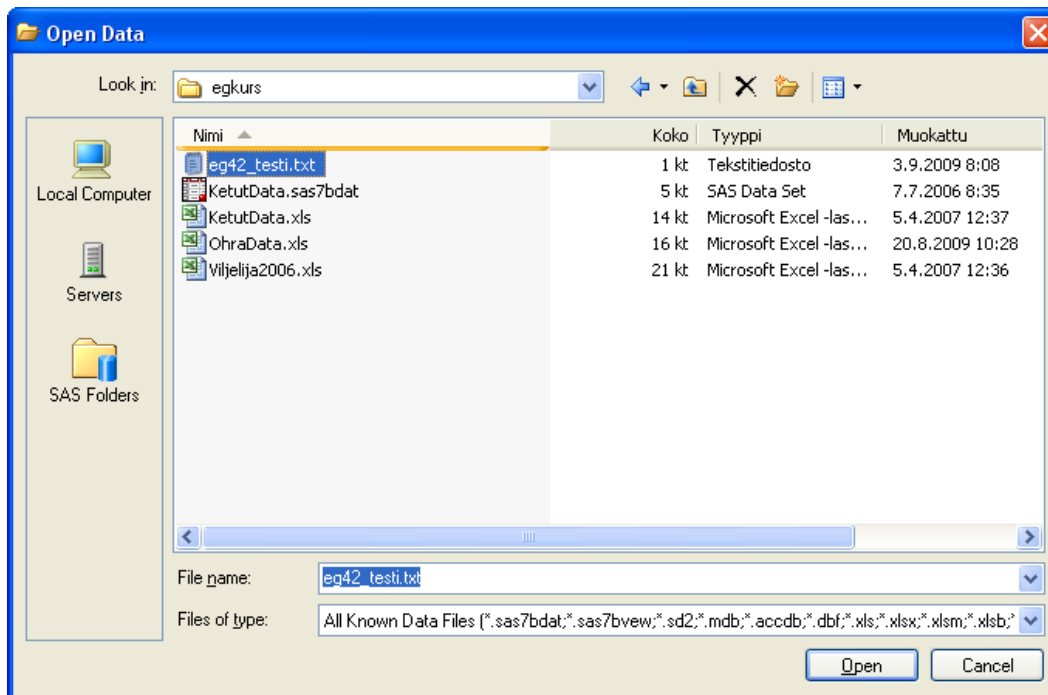
Datan sisäänlukemiseen liittyvissä mahdollisissa ongelmatilanteissa saattaa auttaa jokin Excelissa tehtävä toimenpide. Tällaisia toimenpiteitä ovat sarakkeiden tyyppien määrittely niiden arvoja vastaaviksi, puuttuvien havaintojen jättäminen tyhjiksi (esimerkiksi pisteiden sijaan), muuttujan tyyppin kannalta 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 sisäänlukemisen yhteydessä voidaan määrittää, että havainnot otetaan mukaan tulosdataan vasta toiselta riviltä alkaen eli poistetaan apurivi.

Sisäänlukemisen ongelmatilanteissa saattaa auttaa myös se, että data tallennetaan Excelissä esimerkiksi DBF 4 (dBASE IV) eli *.dbf -muotoon. Tällöin Excelin taustalle tekemät määrittelyt

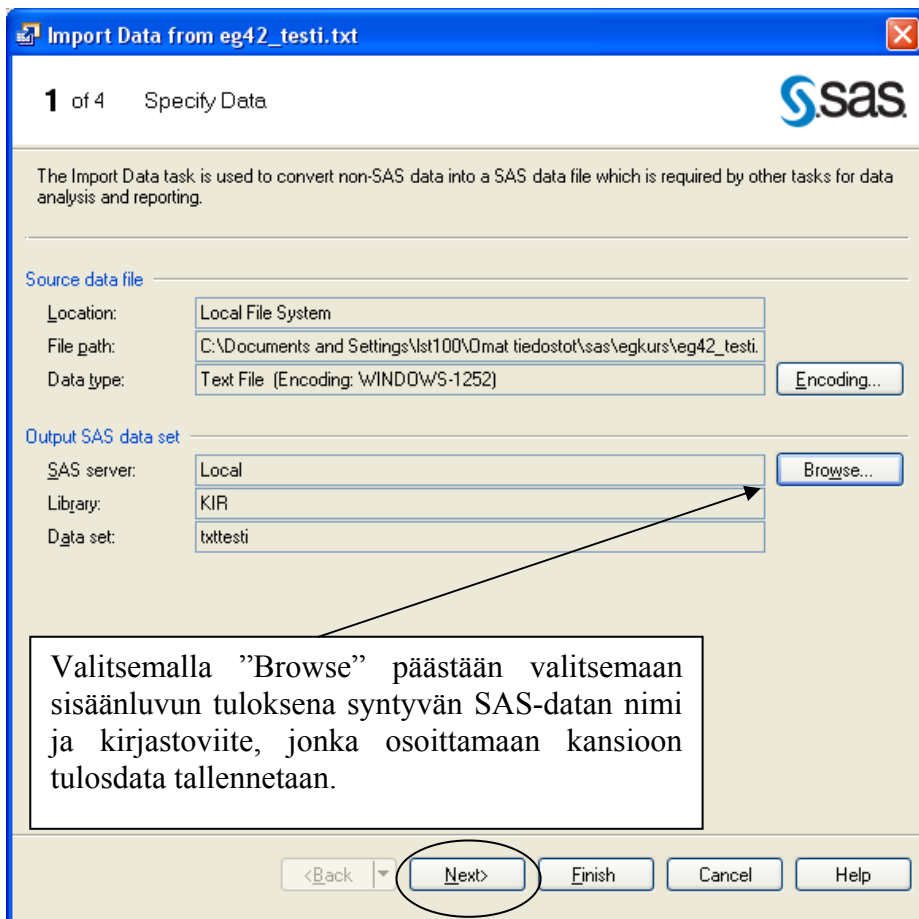
poistuvat ja data tallentuu yksinkertaisesti sellaisessa muodossa kuin se näkyy ruudullakin. Kun data sitten luetaan sisään Enterprise Guideen tässä dBASE-muodossa, onnistutaan kiertämään monia ongelmia ylimääräisten ja mahdollisesti sisäänlukua sekoittavien muotoilujen hävittyä. Vaihtoehtoisesti taustalla olevista laskentakaavoista ja muista määrittelyistä voi päästä eroon siten, että Excel-aineisto kopioidaan ja ”liitetään määräten” kokonaisuudessaan uuteen Excelin laskentataulukkaan siten, että liitetään vain pelkät solujen arvot.

5.6 Tekstimuotoisen data-tiedoston avaaminen Enterprise Guideen

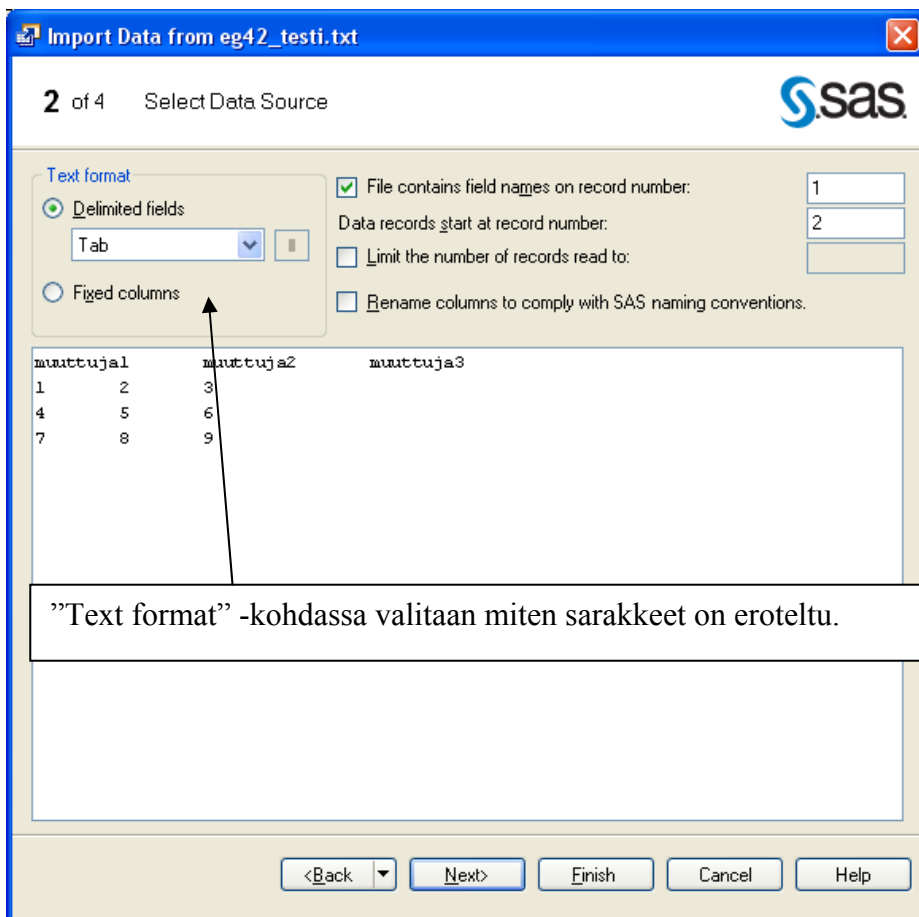
Tekstimuotoisen datan sisäänlukemiseen voidaan käyttää työkalua, johon kuuluvat hyvin pitkälti samat välivaiheet kuin Excel-datan sisäänlukemiseen. Tekstimuotoisen data-tiedoston sisäänlu-
keminen pitää aloittaa valitsemalla ”File” -> ”Import Data”. Tällöin aukeaa vastaava wizard-
tyyppinen työkalu kuin Excel-tiedostoa avattaessa. Ensin valitaan sisäänluettava tiedosto:



Valitsemalla ”Open” päästään eteenpäin:



Toisella sivulla ("Next") tehdään tekstitiedoston sisäänlukemiseen liittyviä erityisvalintoja:



Kolmannelta ja neljänneltä sivulta ("Next") löytyvät samat valinnat kuin Excel-aineiston sisäänluvun vastaavilta sivuilta. Lopuksi painetaan "Finish" ja sisäänluku on valmis.

Vaihtoehtoinen ratkaisu SAS-koodilla:

Alla olevalla koodilla voidaan lukea sisään eg42_testi-niminen tekstitiedosto, jossa on havaintoja kolmesta muuttujasta (muuttuja1, muuttuja2 ja muuttuja3) eroteltuna tabulaattorilla. Huomaa, että koodia käytettäessä tekstitiedostosta luetaan vain varsinaista dataa. Tässä tapauksessa FIRSTOBS=2 -optiolla ilmoitetaan, että datan lukeminen aloitetaan vasta toiselta riviltä, koska ensimmäisellä rivillä ovat muuttujien nimet. Muuttujien nimet annetaan INPUT-lauseessa.

```
DATA kir.txttesti;  
  INFILE 'C:\tahanpolku\eg42_testi.txt' EXPANDTABS MISSEVER FIRSTOBS=2;  
  INPUT muuttuja1 muuttuja2 muuttuja3;  
RUN;
```

5.7 SAS-datan avaaminen Enterprise Guideen

SAS-data avataan Enterprise Guideen valitsemalla "File" -> "Open" -> "Data", etsimällä haluttu SAS-data ja klikkaamalla "Open". SAS-data voidaan paikallistaa omalta tietokoneelta joko selailemalla kansioita resurssienhallintatyypillisesti (valitsemalla "Local Computer" ikkunan vasemmasta reunasta) tai kirjastoviitteen avulla (valitsemalla "Servers" ikkunan vasemmasta reunasta).

Vaihtoehtoinen ratkaisu SAS-koodilla:

Haluttaessa käyttää olemassa olevaa SAS-dataa, pitää yhdistää kirjastoviite, joka osoittaa kansioon, jossa data sijaitsee. Data-vaiheessa data otetaan käyttöön SET-lauseessa seuraavasti:

```
DATA valiaikainentestidata;  
  SET kir.ketutdata;  
RUN;
```

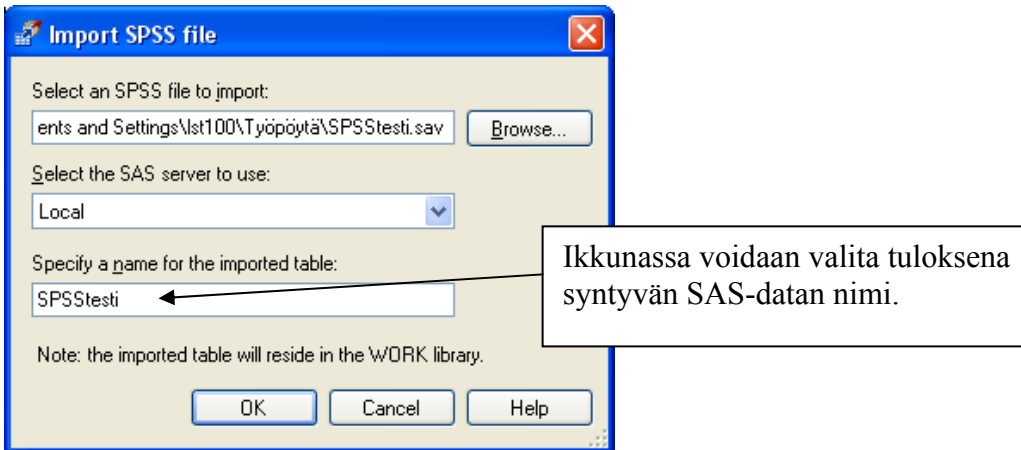
Proseduureissa data otetaan käyttöön vastaavalla tavalla, yleensä DATA= -kohdassa. Esimerkkinä MEANS-proseduuri:

```
PROC MEANS DATA=kir.ketutdata;  
  VAR elokuu;  
RUN;
```

Tehtävä: Avaa esimerkkidatat sisällään pitävästä kansioista (\SAS\EGKURS -kansioista) Enterprise Guideen SAS-data nimeltä **KetutData.sas7bdat**.

5.8 JMP-, SPSS- ja Stata-tiedostojen avaaminen Enterprise Guideen

Enterprise Guide 4.2:een on tullut uutena ominaisuutena erilliset työkalut JMP-, SPSS- ja Stata-ohjelmien tuottamien datatiedostojen käyttöönottamiseksi. Nämä työkalut löytyvät valikosta ”Tasks” -> ”Data” nimillä ”Import JMP file...”, ”Import SPSS file...” ja ”Import Stata file...”. Näiden kolmen työkalun rakenne on hyvin samanlainen ja käyttö yksinkertaista, joten seuraavassa esitellään esimerkin vuoksi vain yksi niistä. Luetaan sisään *.sav -muotoinen SPSS-tiedosto valitsemalla ”Tasks”-> ”Data” -> ”Import SPSS file...”, haetaan SPSS-tiedosto omalta tietokoneelta ja valitaan ”Avaa”, jolloin aukeaa seuraava ikkuna:



Kun valitaan ”OK”, EG lukee SPSS-datan sisään ja muuttaa sen SAS-dataksi. Huomaa, että tuloksena syntynyt SAS-data tallentuu WORK kirjastoviitteen osoittamaan kansioon eli väliaikaisena SAS-datana, joka häviää kun Enterprise Guide istunto lopetetaan. Jos datatiedosto halutaan säilyttää SAS-muotoisena, se tulee muuttaa pysyväksi tiedostoksi esimerkiksi Query Builder -työkalulla (Lisätietoa luvussa **Datan muokkaus**).

Vaihtoehtoinen ratkaisu SAS-koodilla:

Alla esitetään koodiratkaisu JMP-, SPSS- ja Stata-datatiedostojen lukemiseksi SAS-muotoon. Huomaa, että koodilla datat voi tehdä suoraan pysyviksi SAS-datoiksi kirjastoviitteen avulla.

JMP:

```
PROC IMPORT OUT=kir.JMPtesti
  DATAFILE = 'C:\tahanpolku\JMPtesti.jmp'
  DBMS = JMP REPLACE;
RUN;
```

SPSS:

```
PROC IMPORT OUT=kir.SPSSstesti
  DATAFILE = 'C:\tahanpolku\SPSSstesti.sav'
  DBMS = SAV REPLACE;
RUN;
```

Stata:

```
PROC IMPORT OUT=kir.Statatesti
  DATAFILE = 'C:\tahanpolku\Statatesti.dta'
  DBMS = DTA REPLACE;
RUN;
```

5.9 Datojen tallentaminen Enterprise Guidesta

Enterprise Guide -projektissa oleva data voidaan tallentaa tiedostoksi. Tämä tapahtuu klikkaamalla datamatriisi näkyviin ja valitsemalla asiayhteysvalikosta ”**Export**” -> ”Export Data...”. Sen jälkeen valitaan haluttu kansio tietokoneelta, tiedostotyyppi alasetusvalikosta ja annetaan tiedostolle nimi. Valittavia tiedostotyyppisiä ovat muun muassa *.sas7bdat (SAS-data), *.xls (Excel-tiedosto), *.dbf (dBase-tiedosto) ja *.txt (tekstitiedosto). Tavallisen ”Export”-vaihtoehdon lisäksi valittavana on myös vaihtoehto ”Export Data ... 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 datamatriisi näkyviin ja valitsemalla asiayhteysvalikosta ”**Send To**” -> ”E-mail Recipient” tai ”Microsoft Word” tai ”Microsoft Excel”. ”E-mail Recipient”-vaihtoehto toimii yleisimpien sähköpostiohjelmien kanssa (testattu Mozilla Thunderbird -ohjelmalla). 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.

Sekä ”Export” että ”Send To” -toiminnot löytyvät asiayhteysvalikon lisäksi EG 4.2:sta myös valikosta, joka aukeaa kun klikataan datakuvaketta prosessidiagrammissa hiiren oikealla painikkeella.

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

5.10 Projektin tallentaminen

Enterprise Guidessa kaikki työ tallennetaan projekteina. Projekti koostuu toisiinsa liittyvistä datoina, tehtävistä, tulosteista, koodeista, log-tiedostoista ja muistiinpanoista. Tässä kohtaa on syytä mainita, että projektin sisältämät data-kuvat eivät varsinaisesti sisällä dataa, vaan ne ovat pikakuvia varsinaisiin datoihin. Niinpä datat on tallennettava kansioon tietokoneelle, jos ne halutaan säästää EG-istunnon jälkeen. Datat eivät siis tallennu projektitiedoston 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.

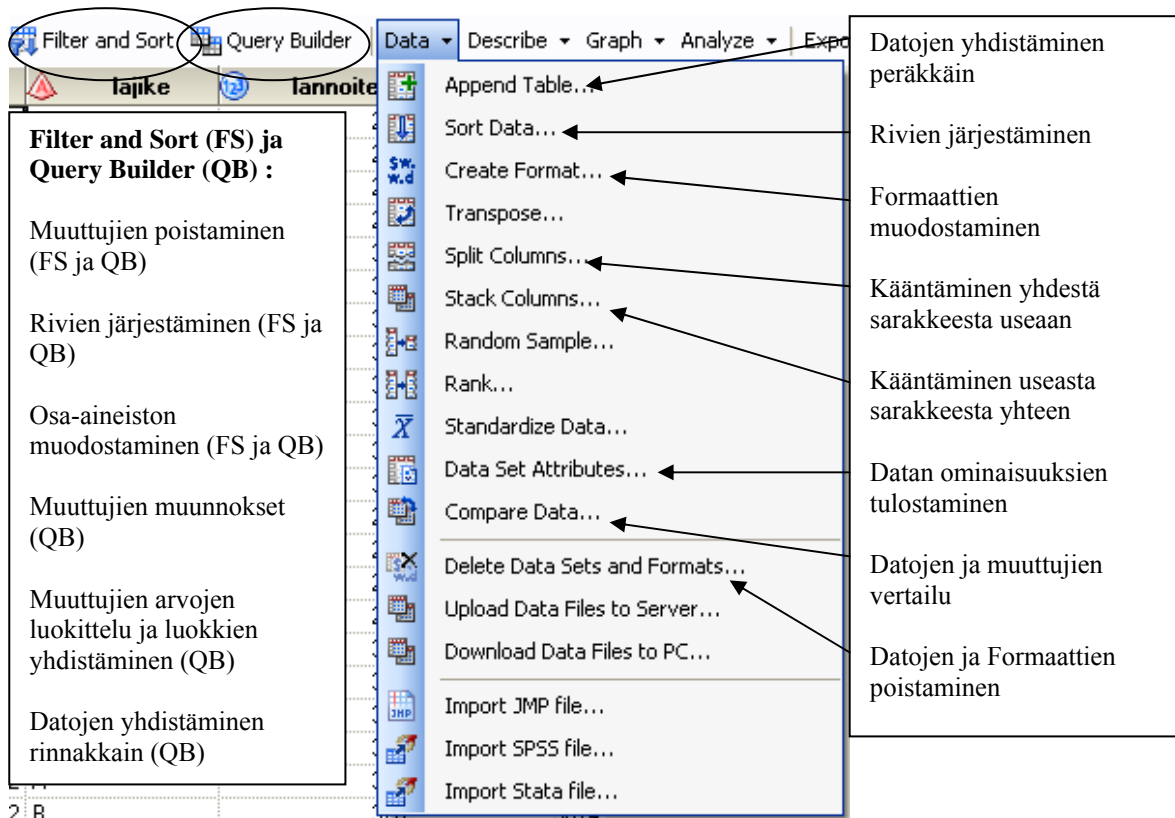
Jos tallennettu projekti avataan toisella tietokoneella tai tiedostot siirretään eri kansioon, kirjastoviite pitää muuttaa viittaamaan uuteen kansioon. Uudessa tilanteessa viittaukset datojen sijainteihin pitää muuttaa myös niille datoilta, joihin ei viitata kirjastoviitteellä vaan jotka on avattu suoraan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

Tehtävä: Tallenna projektisi nimellä **EGkurssi_projekti**.

6 Datan muokkaus

Tässä oppaassa erilaisia datan muokkaustyökaluja käytetään pääasiassaniin sanotun asiayhteysvalikon kautta (katso alla oleva kuva asiayhteysvalikosta). Asiayhteysvalikko saadaan näkyviin tuplaklikkaamalla datasolmua prosessidiagrammissa ja siten avaamalla datamatriisi työtilaan. Datan muokkaustyökalut löytyvät myös ylävalikosta ”Tasks” -> ”Data”. Käytettäessä työkaluja ylävalikon kautta, data pitää olla valittuna aktiiviseksi, jotta sitä voidaan muokata tai muuten käsitellä eri työkaluilla. Datan saa aktiiviseksi klikkaamalla kertaalleen data-kuvaketta prosessidiagrammissa, jolloin se muuttuu siniseksi.

Tässä luvussa käydään läpi seuraavia datan muokkaustyökaluja:



6.1 Datan muuttaminen muokkaustilaan

Ennen kuin dataa voidaan muokata suoraan datamatriisissa, sen tulee olla muokkaustilassa. Datan voi muuttaa muokkaustilaan poistamalla mahdollisen valinnan ”**Edit**”-valikon ”**Protect Data**” -kohdasta tai suoraan kohdistamalla muutoksen johonkin datan soluun, jolloin ruudulle tulee varmistuskysymys muokkaustilaan siirtymisestä. Datan pitää olla näkyvissä työtilassa, jotta ”Edit”-valikon valintaa voidaan muuttaa. Huomaa, että useimpia tämän luvun datan muokkaustyökaluja käytettäessä dataa ei tarvitse muuttaa muokkaustilaan. Muokkaustilaa tarvitaan vain kun halutaan tehdä muutoksia suoraan datamatriisiin työtilassa.

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

6.2 Muuttujien perusominaisuuksien määrittely

Muuttujien nimiä, nimien selitystekstejä ja tyyppejä pääsee muokkaamaan klikkaamalla Enterprise Guiden työtilassa auki olevassa 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 vielä jakaa 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 **OhraDataan** täsmällisemmät nimet ja selitystekstit muuttujille seuraavasti:

Muuttujan nimi (Name):

- lohko -> lohko
- lajike -> ohralajike
- lannoite -> lannoitetaso
- sato -> sato

Selitysteksti (Label):

- lohko -> Lohko
- lajike -> Ohralajike
- lannoite -> Lannoitetaso (kg/ha)
- sato -> Hehtaarisato (kg/ha)

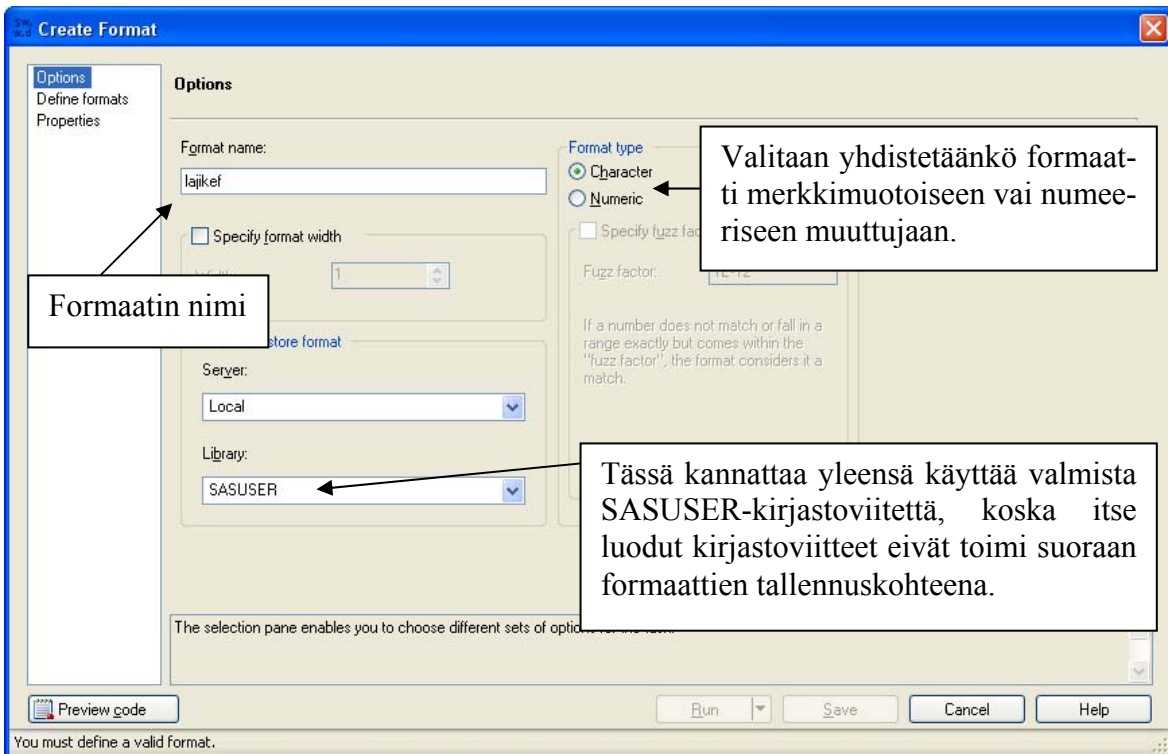
6.3 Muuttujan arvojen tulostusmuotojen eli formaattien käyttäminen

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. Esimerkkinä suomenkielisten selitystekstien määrittäminen arvoille 1 = ’kissa’ ja 2 = ’koira’.

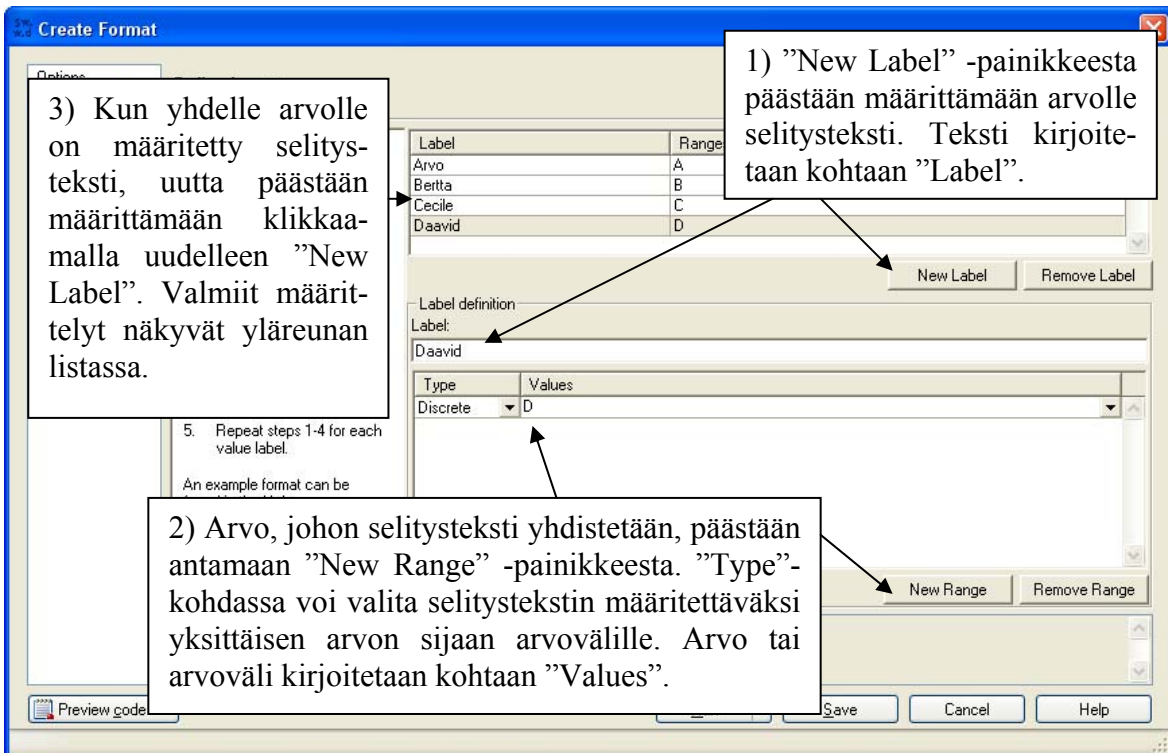
Seuraavaksi käydään läpi, miten formaatti muodostetaan ja miten itse muodostetun tai valmiin formaatin voi yhdistää johonkin muuttujaan.

Formaatin luominen

Formaatteja pääsee määrittämään klikkaamalla asiayhteysvalikosta **"Data"** -> **"Create Format"**. Tämän 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. Kuhunkin muuttujan arvoon liittyvät selitetekstit määritellään **"Define Formats"** -välilehdellä:



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

Vaihtoehtoinen ratkaisu SAS-koodilla:

```
PROC FORMAT;  
  VALUE $ lajikef  
    'A'='Arvo'  
    'B'='Bertta'  
    'C'='Cecile'  
    'D'='Daavid'  
  ;  
RUN;
```

Yllä olevalla SAS-koodilla luodaan istunnon ajan voimassaoleva lajikef-niminen formaatti. Jos sen sijaan halutaan muodostaa pysyvä formaatti, niin se onnistuu seuraavasti:

```
LIBNAME library 'c:\tahankansiopolku';
```

```
PROC FORMAT LIBRARY=library;  
  VALUE $ lajikef  
    'A'='Arvo'  
    'B'='Bertta'  
    'C'='Cecile'  
    'D'='Daavid'  
  ;  
RUN;
```

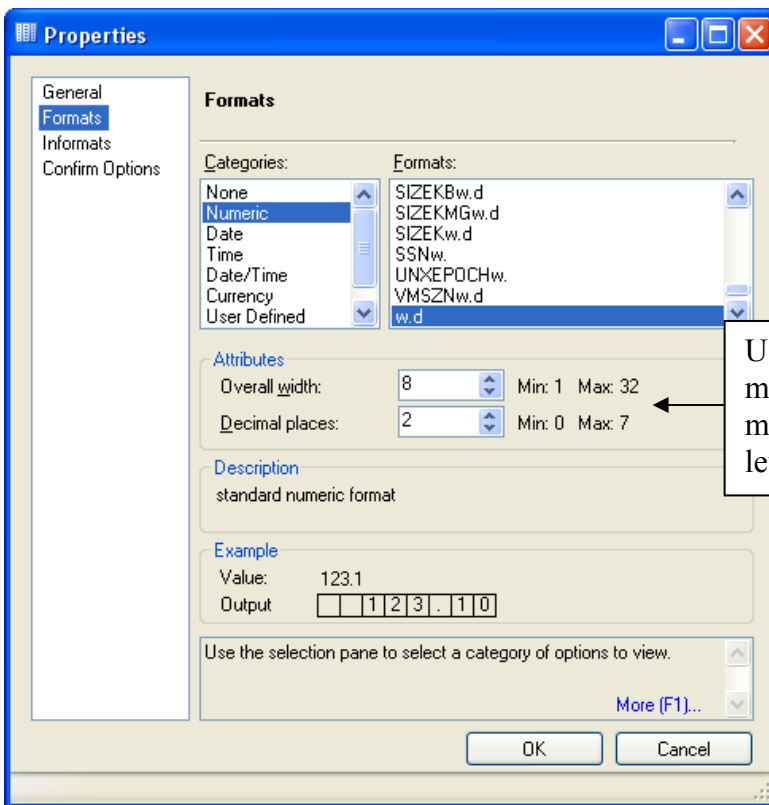
Tässä jälkimmäisessä FORMAT-proseduurissa formaatti tallentuu pysyvästi kirjastoviitteen library osoittamaan kansioon.

Numeerisilla muuttujilla formaatin luominen eroaa siten, että VALUE-lauseesta jätetään \$-merkki pois ja muuttujien arvoja (eli yllä A, B, C ja D) ei laiteta lainausmerkkien väliin. Selitystekstit sen sijaan laitetaan lainausmerkkeihin molemmissa tapauksissa.

Formaatin yhdistäminen muuttujaan

Ensin data pitää muuttaa muokkaustilaan (katso luku Datan muuttaminen muokkaustilaan). Tämän jälkeen formaatit yhdistetään muuttujaan 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 datamatriisissa 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:



Vaihtoehtoinen ratkaisu SAS-koodilla:

Formaatti voidaan yhdistää muuttujaan DATA-vaiheen FORMAT-lauseessa:

```
DATA ohra_data2;  
  SET kir.ohradata;  
  FORMAT lajike lajikef.;  
RUN;
```

Huomaa, että tässä esimerkkikoodissa muodostetaan tuloksena väliaikainen SAS-data nimeltä ohra_data2. Jos halutaan muodostaa pysyvä SAS-data, se ohjataan kirjastoviitteen avulla tallennettavaksi kansioon (esimerkiksi kir.ohra_data2).

Tehtävä: Muodosta **OhraDataan** lajike-muuttujan arvoille A, B, C ja D selitystekstit:

```
A="Arvo"  
B="Bertta"  
C="Cecile"  
D="Daavid"
```

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

6.4 Datan muokkaustyökalut

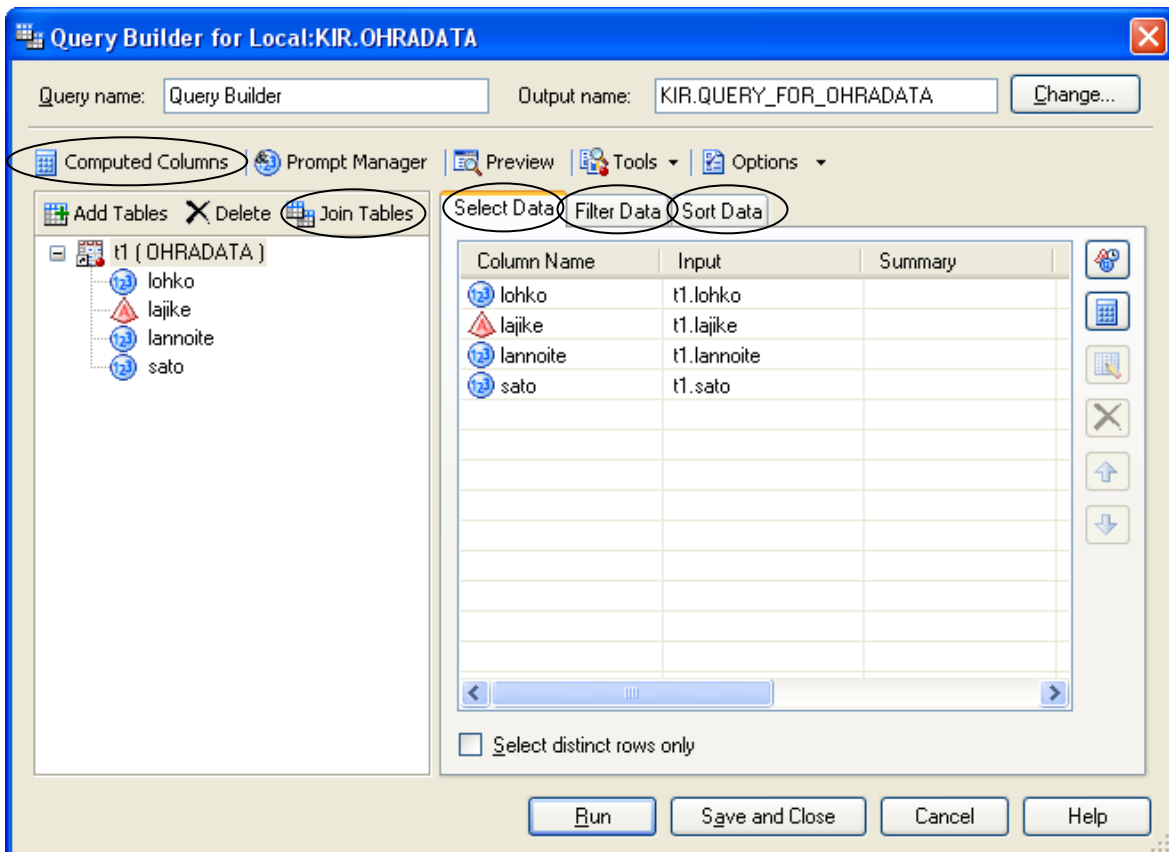
Enterprise Guidessa ensisijainen datan muokkaustyökalu on Query Builder, joka on eräänlainen kyselytyökalu (”kysely”-sana tulee SQL-ohjelmointikielestä, koska Query Builder käyttää SASin PROC SQL:ää). Kysely-sanalla viitataan siihen, että työkalun avulla tehdään dataan kysely ja tuloksena saadaan uusi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

daan kuitenkin tehdä sekä Query Builderillä että jollakin vaihtoehdoisella tavalla. Näitä vaihtoehtoisia datan muokkaustyökaluja ovat dataan liittyvässä asiayhteysvalikossa ”Filter and Sort” sekä ”Data”-valikon työkalut.

Kun muokataan olemassa olevaa dataa (uusien muuttujien tekeminen, muuttujien poistaminen, havaintorivien lisääminen ja poistaminen sekä muuttujien uudelleennimeäminen), kannattaa yleensä käyttää Query Builderia tai muita asiayhteysvalikosta löytyviä datan muokkaustyökaluja. Näiden työkalujen vahvuutena on, että alkuperäinen 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 jälkikäteen tuplaklikkaamalla prosessidiagrammissa kyseistä datan muokkaustyökalua ja valitsemalla asiayhteysvalikosta ”Modify Task”.

Enterprise Guidessa voidaan käyttää varsinaisten datan muokkaustyökalujen lisäksi 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 esimerkiksi Query Builder, mutta niiden kanssa pitää olla varovainen, koska muutokset kohdistuvat suoraan alkupe- räiseen dataan ja muutoksista ei jää dokumentointia projektiin. Niinpä näiden käytössä on sama vaara kuin Excelissä, eli dataa saattaa vahingossa hävitä (”Edit” -> ”Undo” -toimintokaan ei auta). Myös täysin oikein tehdyissä datan muokkauksissa 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 perustoimintoja datan muokkaukseen, kannattaa ehdottomasti ottaa datasta varmuuskopio ennen muokkausta. Varmuuskopiointi onnistuu esimerkiksi valitsemalla dataan liit- tyvästä asiayhteysvalikosta ”Export” -> ”Export datan nimi”.

Dataa päästään muokkaamaan Query Builderilla avaamalla ensin data työtilaan ja klikkaamalla asiayhteysvalikosta ”**Query Builder**”, jolloin aukeaa seuraava näkymä:



Query Builderista löytyy muun muassa seuraavaa:

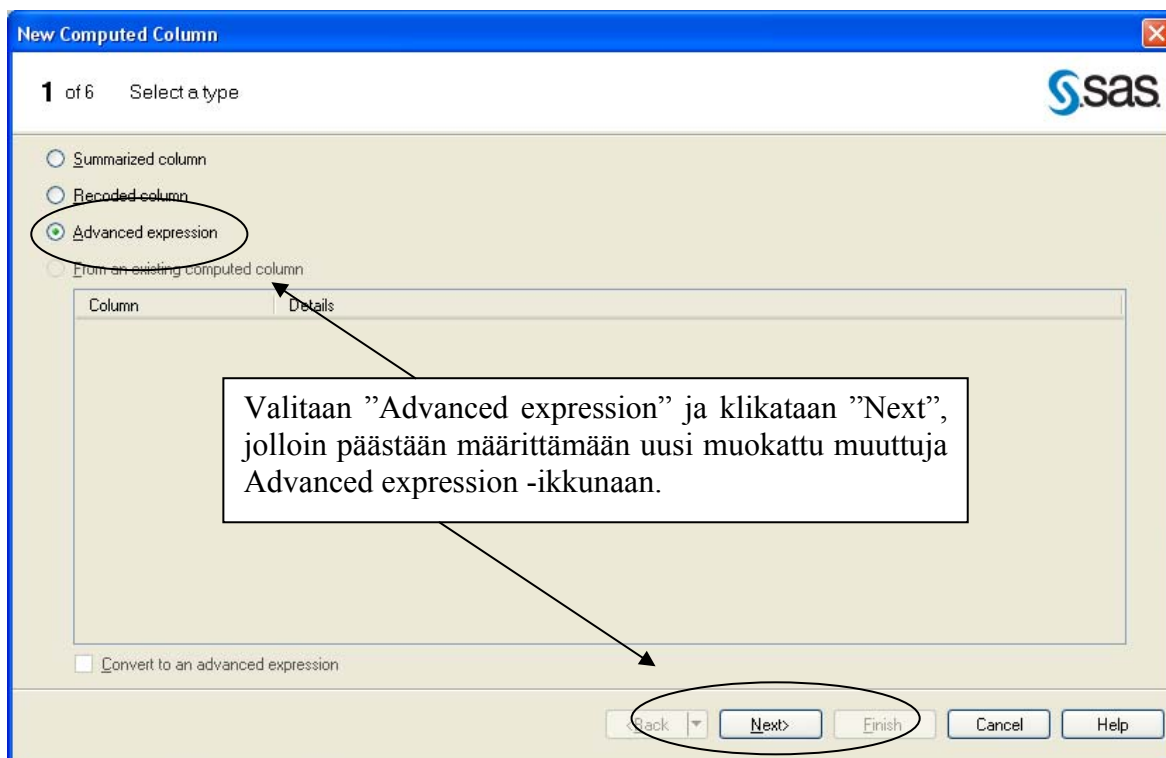
- Select Data -välilehti** Valitaan lopputuloksena syntyvään dataan mukaan tulevat muuttujat. Tässä voidaan myös tehdä tiettyjä muuttujiin liittyviä toimenpiteitä. Useimmiten ”Tools” -> ”Options” -> ”Query” -valikosta kannattaa tehdä valinta ”Automatically add columns from input tables to result set 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 tulodataan.
- Sort Data -välilehti** Aineiston rivien järjestäminen jonkin tai joidenkin muuttujien arvojen mukaiseen järjestykseen.
- Computed Columns** Uusien muuttujien tekeminen vanhoista muuttujista luokittelemalla tai funktioiden tai aritmeettisten operaatioiden avulla.
- Join Tables** Kahden tai useamman aineiston yhdistäminen rinnakkain siten, että samoille riveille eli tilastoyksiköille tulee muuttujia useammasta aineistosta.

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

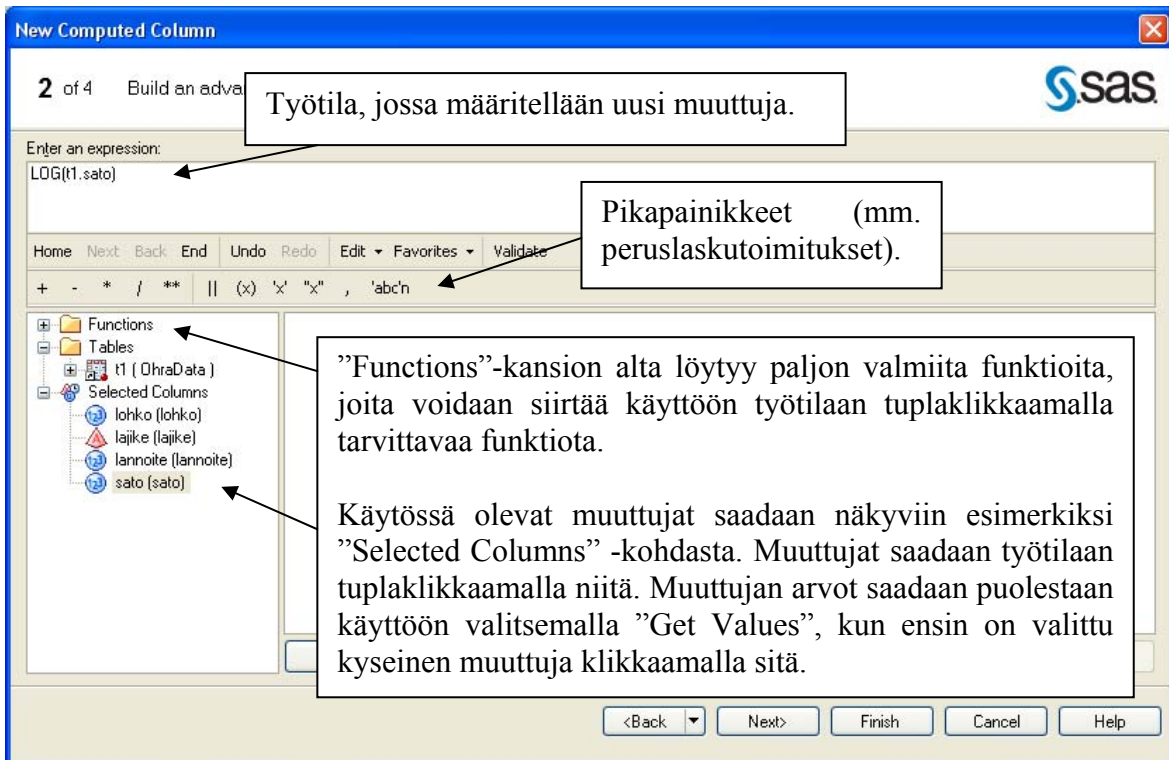
Tehtävä: Tutki datan muokkaus -työkaluja ja avaa Query Builder (**OhraData** lähtödatana). Tarkastele Query Builderia.

6.5 Muuttujien muunnokset ja uusien muuttujien tekeminen

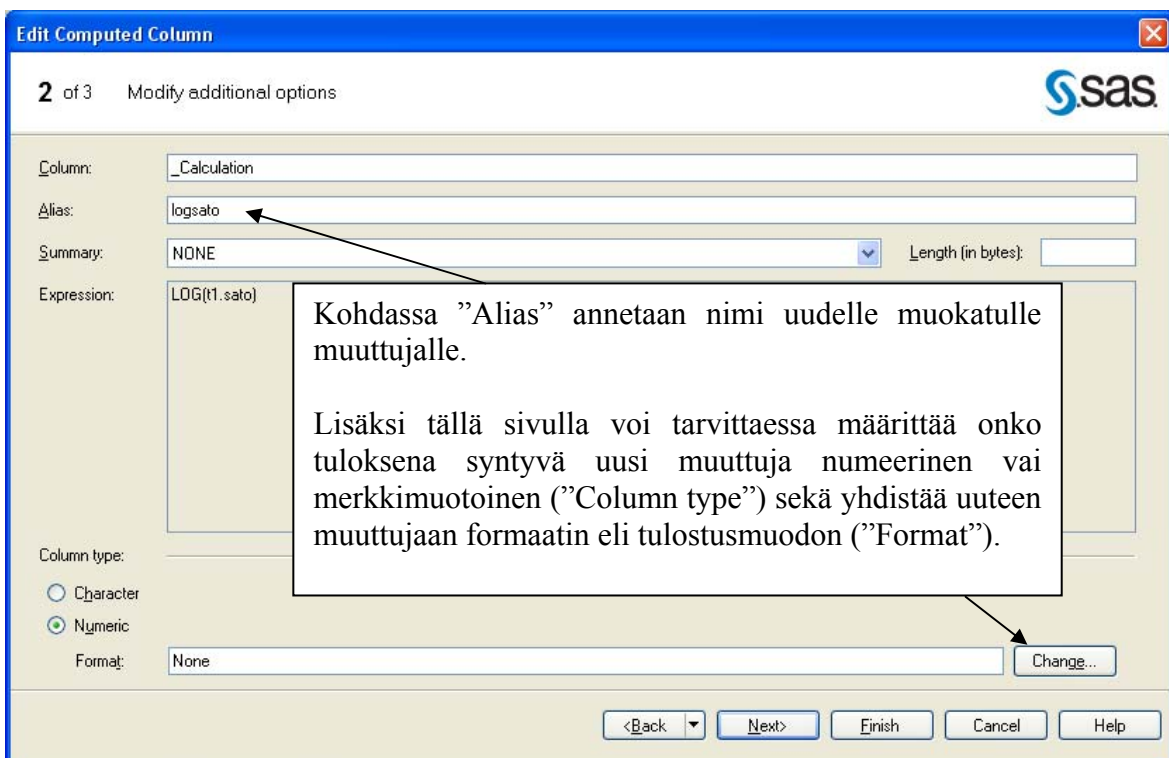
Muuttujien muunnoksia voi tehdä Expression Editorin avulla. Tähän työkaluun päästään käsiksi Query Builderin kautta siten, että avataan muokattava data työtilaan ja klikataan asiayhteysvalikosta ”**Query Builder**” ja valitaan aukeavasta Query Builderin pääikkunasta ”Computed Columns” -> ”New”. Tällöin aukeaa Query Builderin New Computed Column -wizardtyökalu:



Advanced expression -ikkunassa voidaan tehdä uusia muuttujia sekä tavallisten laskutoimitusten avulla että käyttämällä valmiita funktioita, hyödyntäen tarvittaessa aineistossa olevia muuttujia sekä niiden arvoja. Näkymä ikkunassa on seuraava:



Muuttujan 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 dataan viittaava nimi. Tämä datan nimi ei kuitenkaan ole välttämätön, ellei sitten Query Builderissa ole samanaikaisesti otettu käyttöön useampaa lähtödataa joissa saattaa olla samannimisiä muuttujia. Klikkaamalla ”Next” päästään uuden muokatun muuttujan määrittelyssä seuraavalle sivulle:



Painamalla ”Next” päästään wizardin kolmannelle sivulle, josta voi tarkistaa edellä tehdyn määrittelyn yksityiskohdat. Wizard-työkalu lopetetaan painamalla ”Finish”, jolloin palataan Computed Columns -ikkunaan, jossa painetaan ”Close”. Lopuksi painetaan ”Run”. Query Builderissä tehdyt määrittelyt tallentuvat projektin mukana, joten 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 tallennuskohde (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 tuplaklikkaamalla tulosdatan luovaa Query Builder tehtäväsolmua prosessidiagrammissa, valitsemalla asiayhteysvalikosta ”Modify Task”, 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”.

Alla on vielä koottuna muutamia esimerkkejä muuttujien muunnoksista, joita Query Builderissa voi tehdä:

Logaritmuunnos: LOG(sato)

Muuttujan arvot muunnetaan (luonnolliselle) logaritmiasteikolle siten, että haetaan ensin LOG-funktio funktiolistasta ja lisätään se tuplaklikkaamalla työtilaan. Seuraavaksi lisätään funktion sulkujen väliin sato-muuttuja (tuplaklikataan listasta Selected Columns -kohdan alta tai kirjoitetaan suoraan sulkujen sisälle ”sato”). Huomaa, että funktion sulkujen välistä tulee tässä yhteydessä poistaa kaikki ylimääräinen sisältö eli toisin sanottuna ”<numValue>” -teksti, joka on siellä ainoastaan ilmaisemassa että sen paikalle tarvitaan numeerisen muuttujan nimi tai numeerinen arvo.

Neliöjuurimuunnos: SQRT(sato)

Funktio ja muuttujan nimi lisätään työtilaan samoin kuin edellisessä esimerkissä.

Muuttujan arvojen jakaminen vakiolla (esim. sato tonneissa): sato/1000

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

Kahden muuttujan erotus (esim. erotus jälkeen ja ennen käsittelyä): jalkeen-ennen

Jos aineistosta löytyy muuttujat nimeltä ”ennen” ja ”jalkeen”, jotka pitävät sisällään mittauksia kultakin tilastoyksiköltä ennen ja jälkeen jotakin kiinnostuksen kohteena olevaa käsittelyä, erotusmuuttuja voidaan muodostaa siten, että muuttujien nimet lisätään työtilaan listalta samoin kuin edellisissä kohdissa ja väliin lisätään erotusoperaattori pikapainikkeista tai näppäimistöltä. Vastaavalla tavalla erotus voidaan laskea kahdelle päivämäärämuotoiselle muuttujalle, jolloin erotusmuuttujasta tulee numeerinen muuttuja, joka ilmaisee kunkin tilastoyksikön kohdalla kahden päivämäärän erotuksen päivinä.

Päivämäärämuuttujan ja vakiopäivämäärän erotus: pvmmtja-MDY(1,1,2006)

Jos aineistossa olevasta muuttujasta, jonka arvot ovat päivämääriä (tässä ”pvmmtja”), halutaan laskea erotus johonkin tiettyyn vakiopäivämäärään, esimerkiksi 1.1.2006, se voidaan tehdä käyttäen edellä esitettyä menettelyä. Päivämäärämuuttuja (”pvmmtja”), erotusoperaattori sekä MDY-funktio lisätään työtilaan kuten muuttujat, operaattorit ja funktiot on edellisissäkin kohdissa lisätty. Tässä kohdassa uutta on MDY-funktion käyttö, jonka parametrit annetaan numeeroina pilkulla eroteltuna järjestyksessä kuukausi, päivä, vuosi. Kun halutaan laskea päivämäärämuodossa olevan muuttujan ja vakiopäivämäärän 1.1.2006 erotus, pitää vakiopäivämäärä ensin muuttaa SASin ymmärtämään päivämäärämuotoon. MDY-funktio muodostaa sille parametreinä annetuista kuukaudesta (”M”), päivästä (”D”) ja vuodesta (”Y”) päivämäärän SASin päivämää-

rämuotoon, jolloin erotus voidaan laskea ja tuloksena saadaan muuttuja, jonka arvot ovat päiviä. MDY-funktio on päivämääriä käsiteltäessä hyvin käyttökelpoinen, koska sille voidaan antaa vakioarvojen lisäksi parametreinä myös numeeriset kuukausi-, päivä- ja vuosimuuttujat. MDY-funktion lisäksi vakioarvoinen päivämäärä on mahdollista antaa niin sanotussa ”SAS Date Constant” -muodossa, jolloin MDY-funktion sijaan erotus voitaisiin antaa muodossa ”pvmmtja-’1JAN2006’d”. Vastaavat vakiot löytyvät myös kellonajoille (esimerkiksi ’00:00’t) ja päivämäärä-kellonaikatiedoille (esimerkiksi ’1JAN2006 00:00’dt). Muita käyttökelpoisia päivämääriin liittyviä funktioita ovat muun muassa MONTH-, DAY- ja YEAR-funktiot, joiden avulla saadaan SASin päivämäärämuotoisesta muuttujasta eroteltua pelkkä kuukausi-, päivä- tai vuositieto. Hyödyllisiä ovat myös DATEPART ja TIMEPART-funktiot, joiden avulla saadaan sekä päivämäärä- että aikatiedon arvoinaan sisällään pitävästä muuttujasta eroteltua joko pelkkä päivämäärätieto (DATEPART-funktiolla) tai aikatieto (TIMEPART-funktiolla). Katso lisätietoa funktioista ja päivämäärien käsittelystä SAS Enterprise Guiden ohjeesta eli Helpistä.

Vaihtoehtoinen ratkaisu SAS-koodilla:

Muuttujien muunnoksia ja uusia muuttujia voi tehdä SAS-koodilla muun muassa DATA-vaiheessa:

```
DATA ohra_data3;  
  SET ohra_data2;  
  log_sato=LOG(sato);  
  sqrt_sato=SQRT(sato);  
  tonnisato=sato/1000;  
RUN;
```

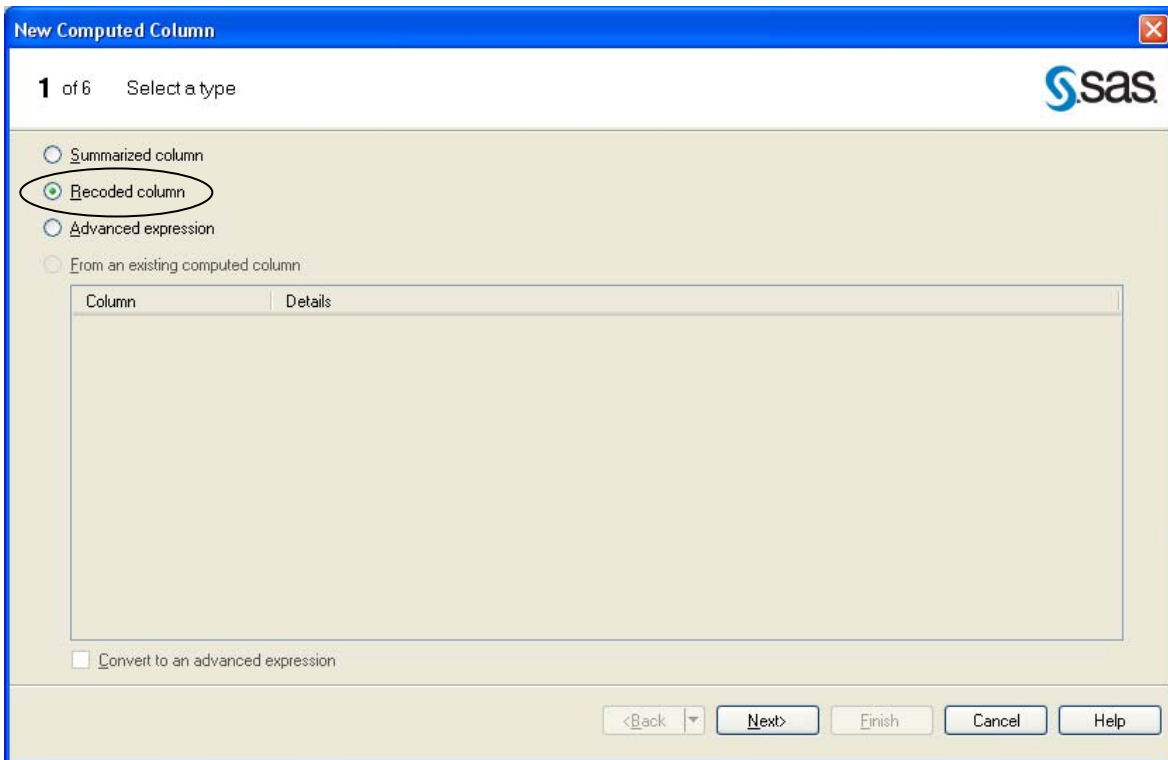
Tehtävä: Tee **OhraDataan** uudet muuttujat tsato (sato tonneina) ja sqrt_sato (saton neliöjuuri). Käytä tähän Query Builderia. Anna tulosdatalle nimeksi **ohra_muok**.

6.6 Muuttujien luokittelu ja luokkien yhdistäminen

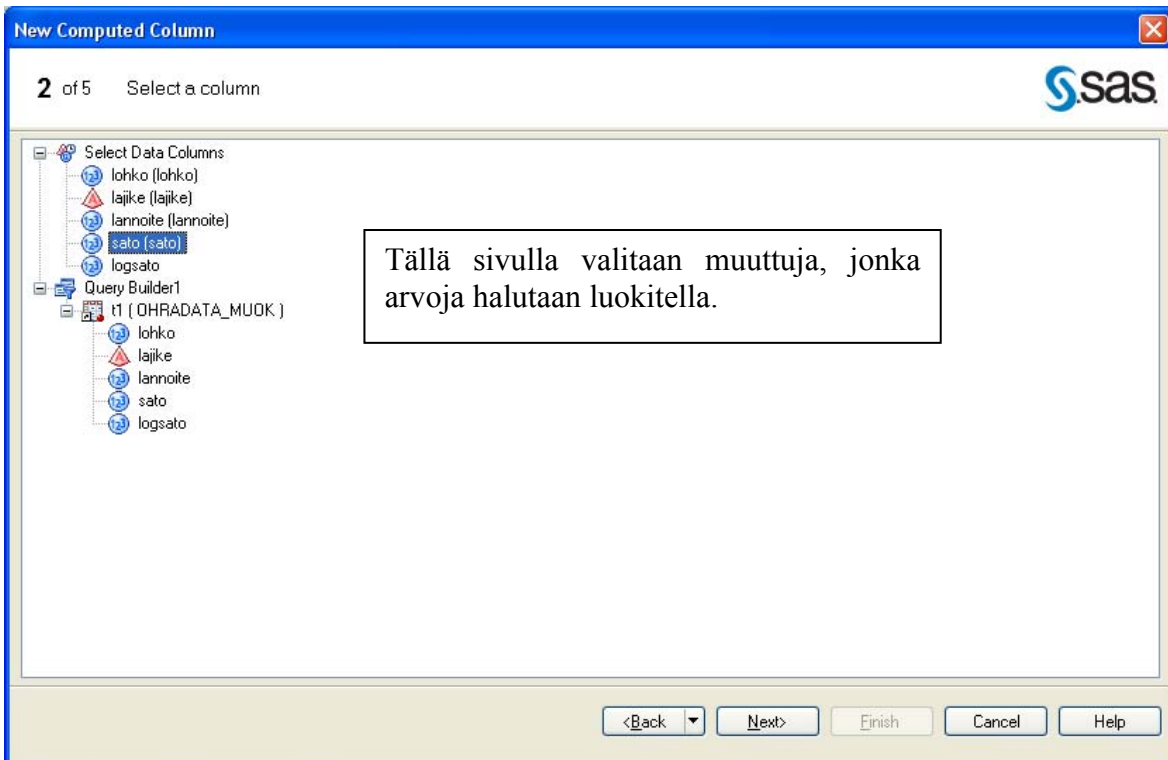
Jatkuvien muuttujien arvojen luokittelu ja kategoristen muuttujien luokkien yhdistäminen voidaan myös tehdä Query Builderin avulla. Molemmat tehdään käyttäen samanlaista tekniikkaa. Seuraavassa esitetään, kuinka jatkuvan muuttujan arvot saa luokitetuksi.

Jatkuvan muuttujan arvojen luokittelu

Ensin avataan Query Builder (valitaan asiayhteysvalikosta ”Query Builder” käsiteltävän aineiston ollessa avattuna työtilassa). Liikkeelle lähdetään aivan samoin kuin muunnoksissa eli ensin valitaan ”Computed Columns” -> ”New”, mutta sitten valitaankin ”Recorded Column”:



Painamalla ”Next” päästään seuraavalle sivulle:



Kolmannella sivulla päästään määrittämään miten jatkuvan muuttujan arvot halutaan luokitella. Klikkaamalla ”Add” aukeaa ikkuna, jossa voidaan lisätä yhden luokan määrittely kokonaisuuteen:

Valitaan ”Replace a range”.

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

Nuoli-painikkeista saadaan näkyviin lista alkuperäisen muuttujan arvoista.

Tässä määritetään uusi arvo.

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

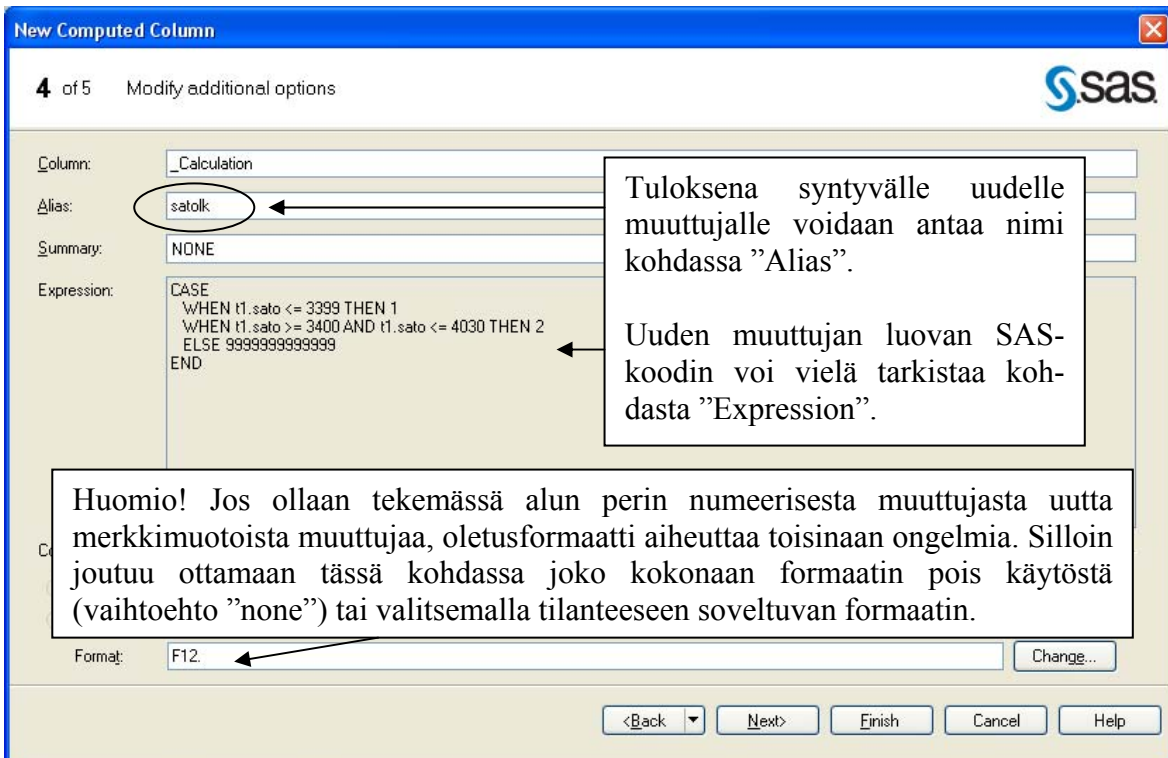
Kun kaikki luokat on määritelty, näkyy kokonaisuus wizard-työkalun kolmannella sivulla:

Tässä voidaan määrittää kaikki ehtoihin sisältymättömät arvot esimerkiksi puuttuviksi havainnoiksi. Tässä määritetään puuttuva tieto korvattavaksi luvulla, joka varmasti erottuu todellisista arvoista.

Tulosmuuttujan tyyppiä voidaan valita muu kuin lähtömuuttujan tyyppi.

”Add”-valinnalla päästäisiin määrittämään uusi luokittelu edellisessä kuvassa esitettyyn ikkunaan.

Lisää määrittelyjä voidaan tehdä neljännellä sivulla, jonne päästään valitsemalla ”Next”:



Painamalla "Next" pääsee vielä yhteenvetosivulla ja wizard lopetetaan painamalla "Finish". Tämän jälkeen painetaan "Computed Columns" -ikkunassa "Close". Query Builderin pääikkunassa voi tässäkin yhteydessä tarvittaessa vaihtaa tuloksena syntyvän SAS-datan nimen ja kirjastoviitteen osoittaman tallennuskohteen valitsemalla oikeasta ylänurkasta "Change". Query Builder suoritetaan lopuksi valitsemalla "Run".

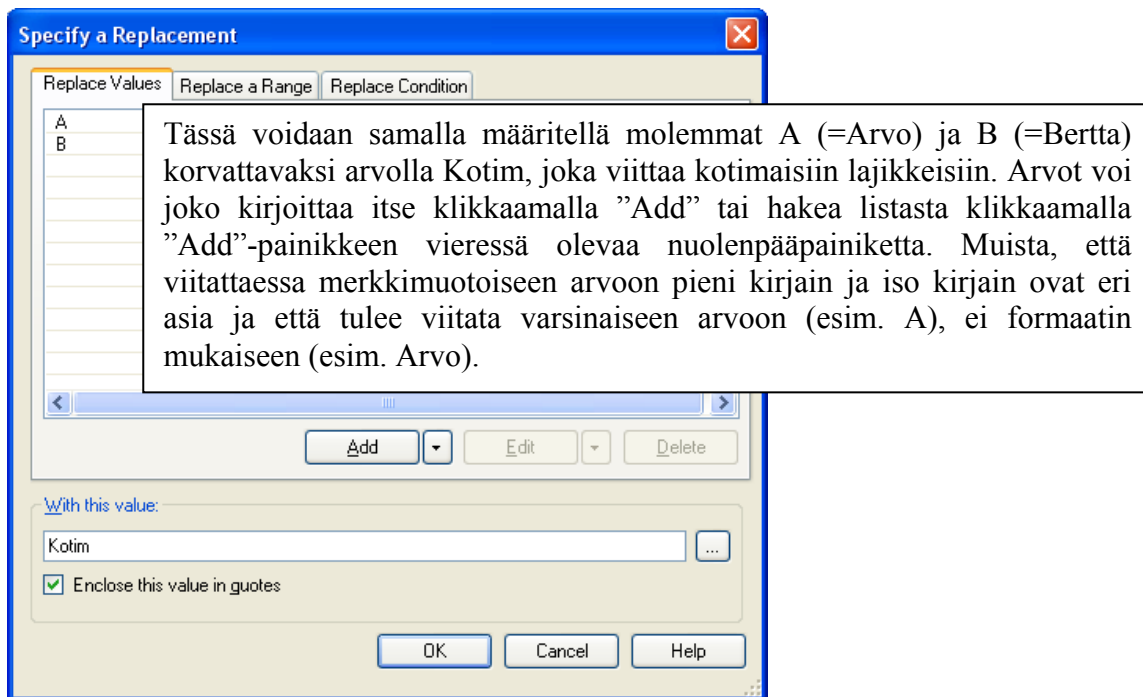
Vaihtoehtoinen ratkaisu SAS-koodilla:

Jatkuvan muuttujan arvojen luokittelua voi tehdä SAS-koodilla muun muassa DATA-vaiheessa:

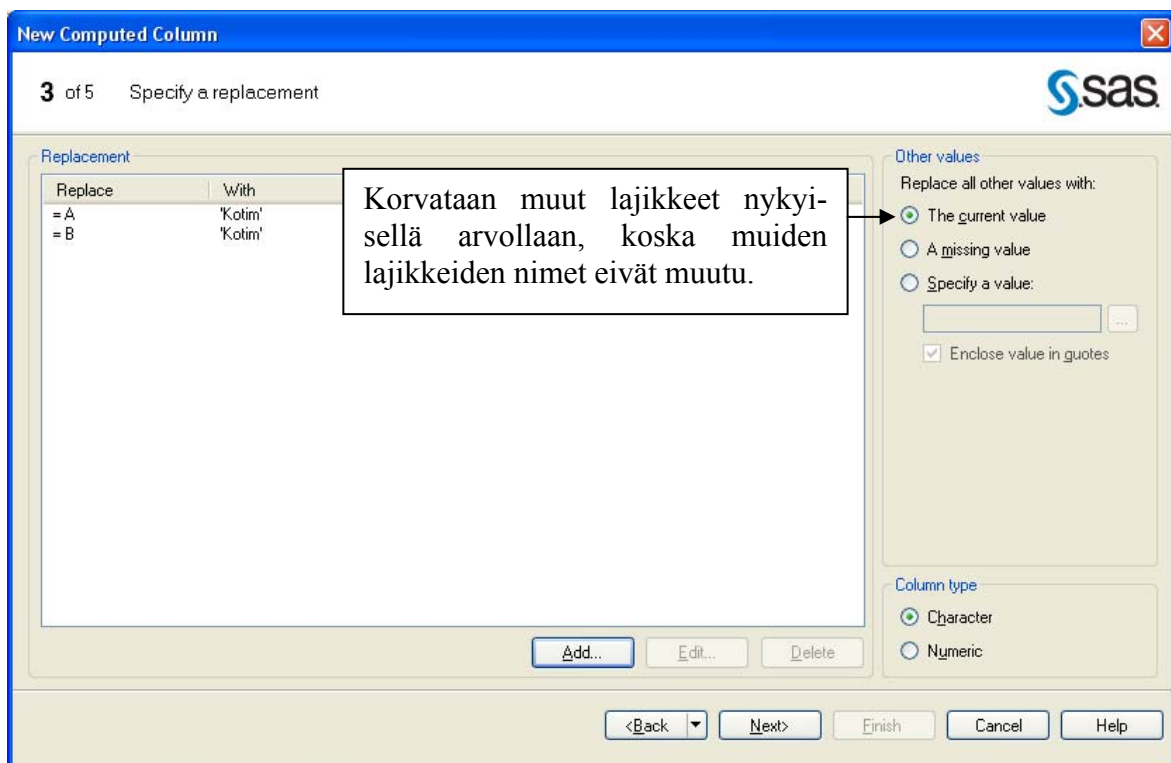
```
DATA ohra_data4;
  SET ohra_data3;
  IF sato < 3400 THEN satolk=1;
  IF sato >= 3400 THEN satolk=2;
RUN;
```

Kategorisen muuttujan luokkien yhdistäminen

Luokkien yhdistämisessä voidaan käyttää toimintoa "Replace", jolla voidaan korvata yksi arvo kerrallaan toisella arvolla. Pelkkää "Replace"-toimintoa käyttäen Ohra-datassa voidaan yhdistää lajikkeet "Arvo" ja "Berta" yhdeksi ryhmäksi "Kotimaiset" ("Replace Values" -kohtaan päästään samoin kuin edellä jatkuvan muuttujan arvoja luokiteltaessa eli valitsemalla Query Builderissa "Computed Columns" -> "New", ensimmäiseltä sivulta "Recoded column", toiselta sivulta käsiteltävä muuttuja (tässä lajike) ja kolmannelta sivulta "Add"):



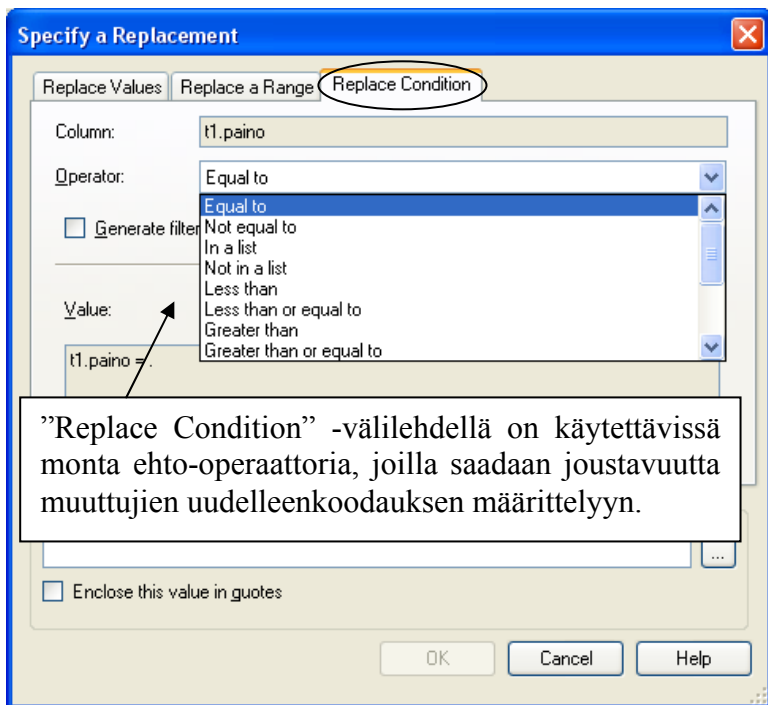
Painettaessa "OK" wizardin kolmas sivu eli "Specify a replacement" -sivu näyttää seuraavalta:



Wizardin neljännellä sivulla voidaan tässäkin yhteydessä antaa tuloksena syntyvän uuden muuttujan nimi kohdassa "Alias" ja viidennellä sivulla voidaan vielä tarkistaa tehdyt valinnat yhteenvedosta. Lopuksi painetaan "Finish" ja sen jälkeen "Close" ja "Run".

Jos muuttujan määrittelyä halutaan palata muokkaamaan myöhemmin, se onnistuu tuplaklikkaamalla tulokset luovaa Query Builder tehtäväsolmua prosessidiagrammissa ja valitsemalla asiayhteysvalikosta "Modify Task". Sieltä jatketaan klikkaamalla "Computed Columns", valitsemalla listasta muuttuja, jonka määrittelyä halutaan muokata, ja valitsemalla "Edit". Tällöin aukeaa suppeampi kolmisivuinen versio wizard-käyttöliittymästä, jossa voidaan muokata aiemmin tehtyjä valintoja. Kun muokkaus on tehty, tehtäväsolmu pitää suorittaa uudelleen valitsemalla Query Builderin päänäkymässä "Run".

Enterprise Guiden 4.2 version myötä muuttujien arvojen uudelleenmäärittelyyn on tarjolla myös uusia joustavia mahdollisuuksia tarjoava toiminto ”Replace Condition”, joka löytyy edellä esitetyn työkalun kolmantena vaihtoehtona ”Replace Values” ja ”Replace a Range” -välilehtien rinnalta:



Vaihtoehtoinen ratkaisu SAS-koodilla:

```
DATA ohra_data5;  
  SET ohra_data4;  
  IF lajike IN ('A','B') THEN lajikeyhhd='Kotim';  
  ELSE lajikeyhhd=lajike;  
RUN;
```

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

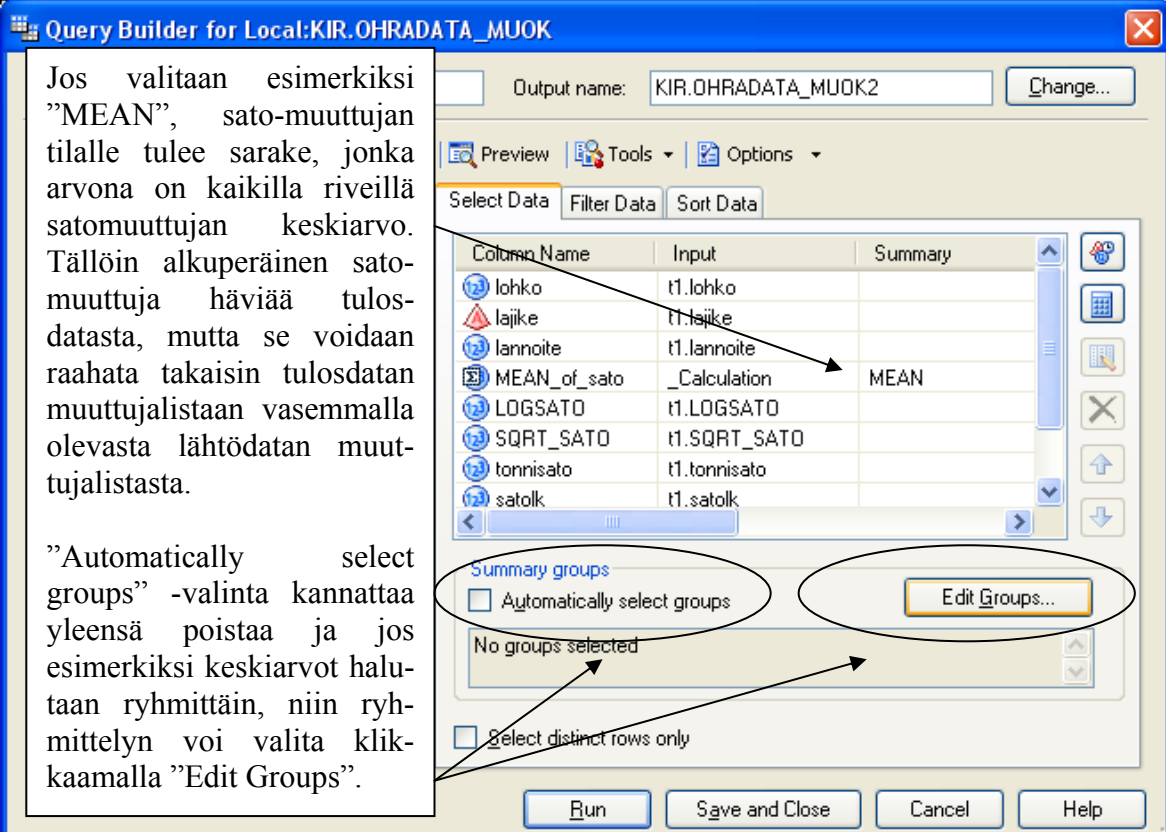
- 1 = ”alle 3400 kg/ha”
- 2 = ”3400 kg/ha tai enemmän”

Tehtävä: Tee **OhraDataan** lajike-muuttujasta uusi muuttuja lajikeyhhd, jossa lajikkeet saavat seuraavat arvot:

- A -> Kotim
- B -> Kotim
- C -> C
- D -> D

6.7 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** (löytyy asiayhteysvalikosta muokattavan datan ollessa auki työtilassa) ”Select Data” -välilehdellä. Klikkaamalla Query Builderissa ”Select Data” -välilehdeltä solua, joka on ”Summary”-sarakkeen ja satomuuttujan rivin risteyskohdassa, saadaan näkyviin seuraava valikko, josta alla olevassa kuvassa on valittu laskettavaksi keskiarvo (mean):



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.

”Automatically select groups” -valinta kannattaa yleensä poistaa ja jos esimerkiksi keskiarvot halutaan ryhmittäin, niin ryhmittelyn voi valita klikkaamalla ”Edit Groups”.

Column Name	Input	Summary
lohko	t1.lohko	
lajike	t1.lajike	
lannoite	t1.lannoite	
MEAN_of_sato	_Calculation	MEAN
LOGSATO	t1.LOGSATO	
SQRT_SATO	t1.SQRT_SATO	
tonnisato	t1.tonnisato	
satolk	t1.satolk	

Summary groups

Automatically select groups

No groups selected

Select distinct rows only

Edit Groups...

Run Save and Close Cancel Help

Vaihtoehtoinen ratkaisu SAS-koodilla:

Muuttujan luominen artimeettisellä operaatiolla onnistuu koodissa helpoiten PROC SQL:n avulla:

```
PROC SQL;
  CREATE TABLE work.ohra_data6 AS
  SELECT ohra.lohko,
         ohra.lajike,
         ohra.lannoite,
         ohra.sato,
         ohra.log_sato,
         ohra.sqrt_sato,
         ohra.tonnisato,
         ohra.satolk,
         ohra.lajikkeyhd,
         (MEAN(ohra.sato)) AS MEAN_of_sato
  FROM work.ohra_data5 AS ohra
  GROUP BY ohra.lohko;
```

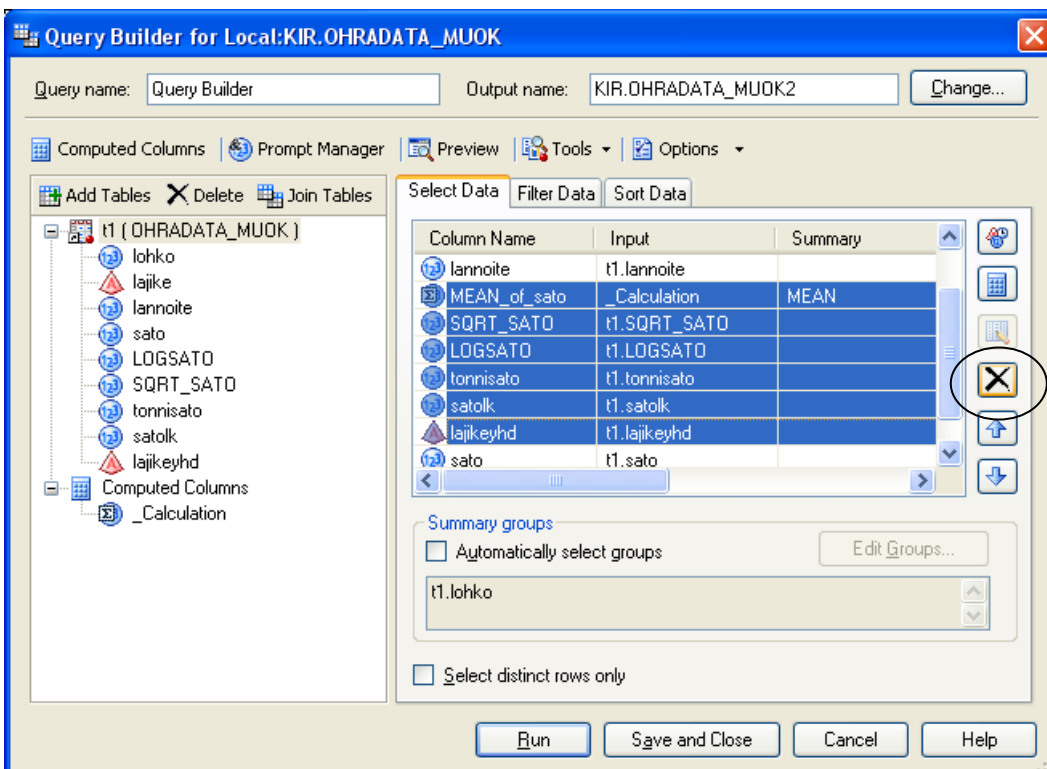
QUIT;

SQL-proseduurin rakenne poikkeaa selvästi DATA-vaiheesta. Jos haluat tutustua tarkemmin SQL-proseduuriin, hyviä lähteitä ovat Enterprise Guiden Helpin ”SAS Syntax Help” sekä Prairien (2005) kirja ”The essential PROC SQL handbook for SAS users”.

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

6.8 Muuttujien poistaminen

Muuttujien poistaminen kannattaa tehdä **Query Builderin** (löytyy asiayhteysvalikosta muokattavan datan ollessa auki työtilassa) avulla seuraavasti: valitaan ”Select Data” -välilehdellä muuttuja tai muuttujat ensin aktiiviseksi ja sitten painetaan oikeassa reunassa olevaa X-painiketta:



Vaihtoehtoinen ratkaisu SAS-koodilla:

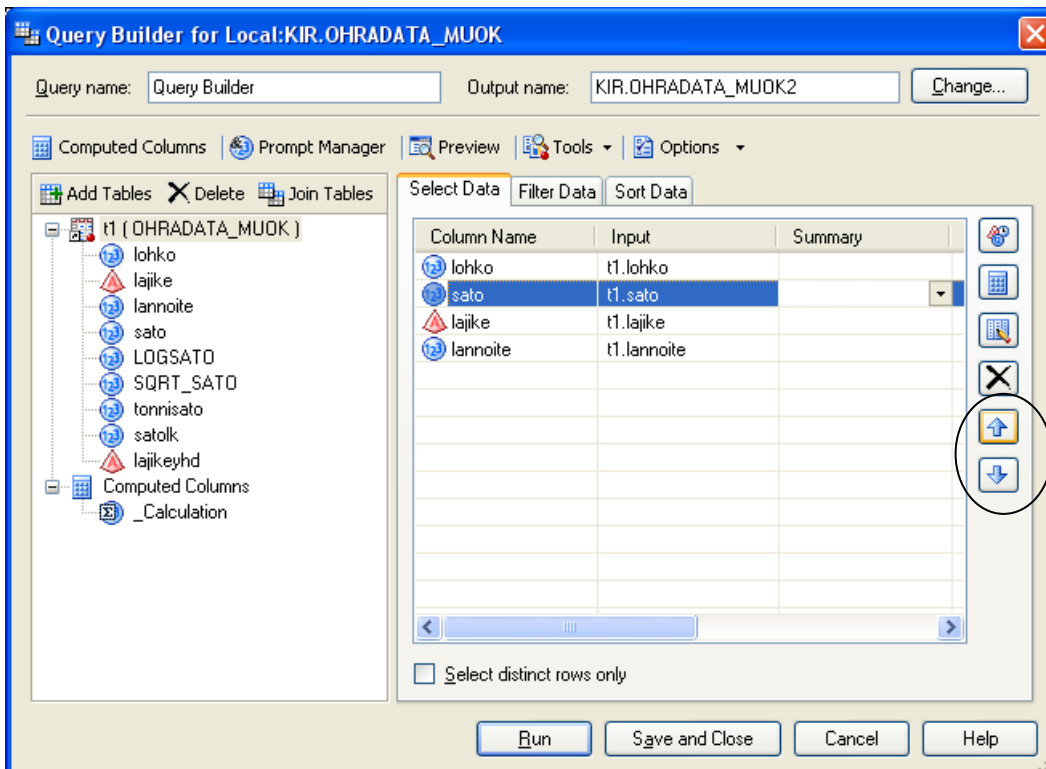
DATA-vaiheessa muuttujia voidaan jättää pois tulosdatasta listaamalla ne DROP-lauseessa. Vaihtoehtoisesti mukaan otettavat muuttujat voitaisiin listata KEEP-lauseessa.

```
DATA ohra_data7;
  SET ohra_data6;
  DROP log_sato sqrt_sato tonnisato satolk lajikeyhd mean_of_sato;
RUN;
```

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

6.9 Muuttujien järjestyksen vaihtaminen datassa

Muuttujien järjestyksen vaihtaminen onnistuu myös **Query Builderin** (löytyy asiayhteysvalikosta muokattavan datan ollessa auki työtilassa) ”Select Data” -välilehdellä. Muuttujan paikkaa voi vaihtaa listassa ja siten myös tulosdatassa klikkaamalla muuttujan aktiiviseksi ja painamalla oikeassa reunassa olevia valintapainikkeita, joissa nuolet osoittavat alas ja ylös:



Vaihtoehtoinen ratkaisu SAS-koodilla:

Yksi tapa valita muuttujien järjestys tulosdatassa on luetella ne halutussa järjestyksessä SQL-proseduurin SELECT -komennon jälkeen:

```
PROC SQL;
  CREATE TABLE ohra_data8 AS SELECT
    ohra.lohko,
    ohra.sato,
    ohra.lajike,
    ohra.lannoite
  FROM ohra_data7 AS ohra;
QUIT;
```

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

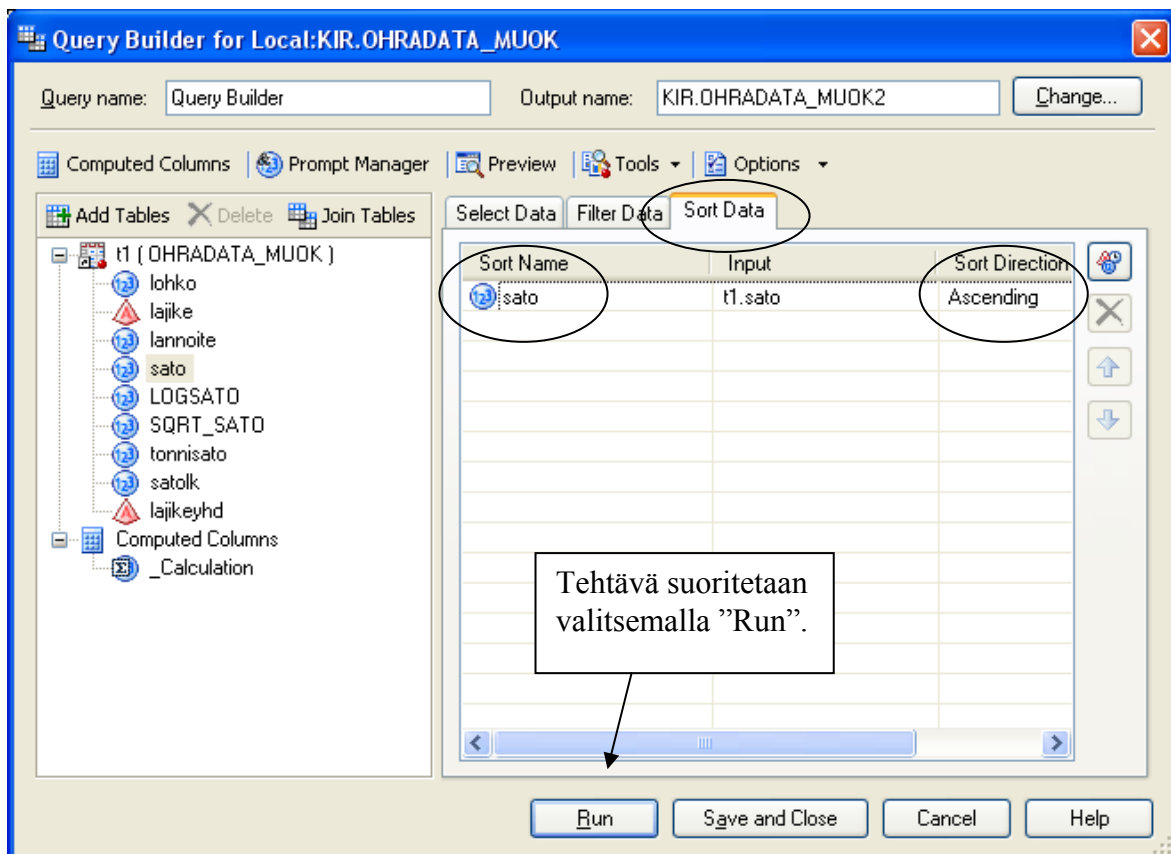
6.10 Datan havaintorivien järjestäminen

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 perinteisessä SAS-käyttöliittymässä, koska EG osaa pääasiassa järjestää datan automaattisesti, jos jokin tehtävä sitä vaatii.

Järjestäminen voidaan tehdä EG:ssä joko Query Builderillä, Filter and Sort -työkalulla tai Sort Data -työkalulla. Tässä esitellään kuitenkin vain monikäyttöisiä Query Builder ja Filter and Sort -työkaluja, koska silloin voidaan yhdistää yhteen työvaiheeseen useita datan muokkaustoimenpiteitä. Pelkästään datan havaintorivien järjestämiseen tarkoitettussa Sort Data -erikoistyökalussa on kuitenkin ominaisuuksia, joita Query Builderistä ja Filter and Sortista ei löydy, joten siihenkin kannattaa tutustua.

Havaintorivien järjestäminen Query Builderilla

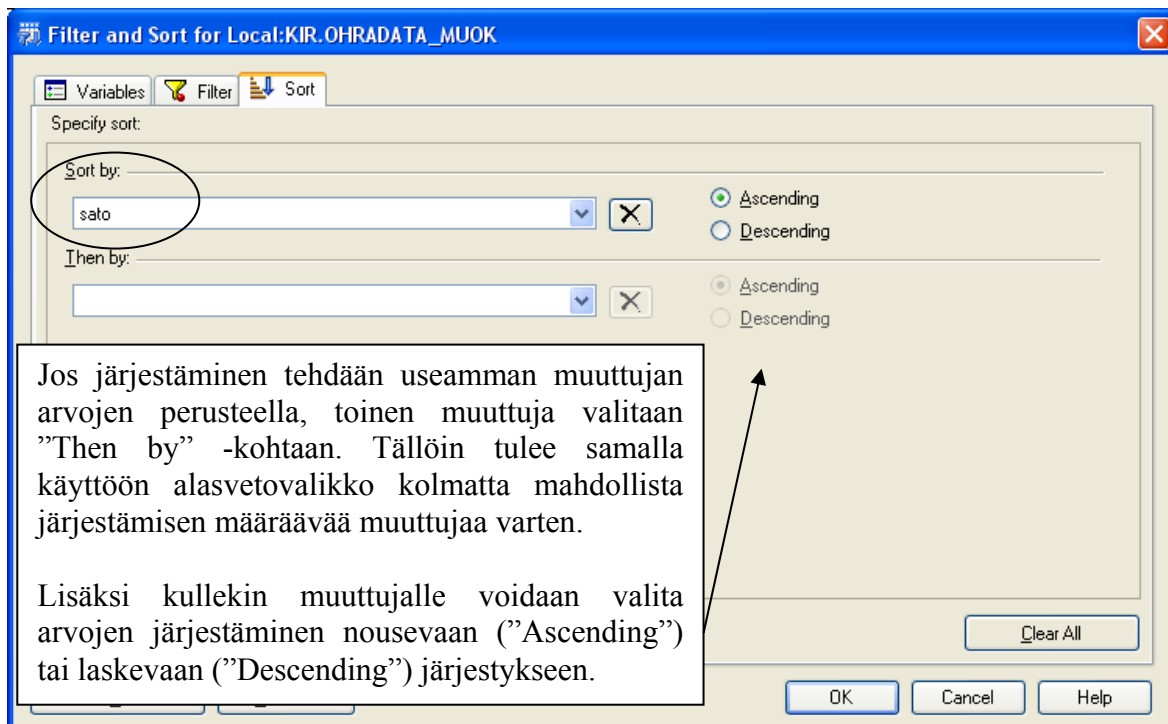
Query Builderissa (löytyy asiayhteysvalikosta muokattavan datan ollessa auki työtilassa) järjestäminen tehdään siten, että vasemmalla olevasta muuttujalistasta raahataan ”Sort Data” -välilehdelle muuttuja tai muuttujat, joiden arvojen mukaan järjestäminen tehdään. Tämän jälkeen valitaan ”Sort Direction” -kohdasta joko ”Ascending” tai ”Descending”, riippuen halutaanko rivit nousevaan vai laskevaan järjestykseen:



Useamman muuttujan arvojen mukaan järjestettäessä muuttujien järjestys ”Sort Data” -välilehdellä määrittelee muuttujien priorisoinnin. Rivit laitetaan ensin järjestykseen ensimmäisen muuttujan arvojen mukaan ja sitten ensimmäisen muuttujan kunkin arvon sisällä toisen muuttujan arvojen mukaan.

Havaintorivien järjestäminen Filter and Sortilla

Filter and Sort -työkalussa (löytyy asiayhteysvalikosta muokattavan datan ollessa aktiivisena) järjestäminen tehdään siten, että ”Sort”-välilehdellä valitaan ”Sort by” -kohtaan se muuttuja, jonka arvojen mukaan järjestäminen tehdään:



Filter and Sort -työkalu suoritetaan painamalla ”OK”. Tämän työkalun yhtenä heikkoutena on, että tulosdatan nimeä ei voi valita ja se tehdään automaattisesti oletuskirjastoviitteen osoittamaan kansioon. Tämä asettaa rajoituksia työkalun käytettävyydelle.

Vaihtoehtoinen ratkaisu SAS-koodilla:

Aineisto voidaan järjestää seuraavalla SAS-koodilla:

```
PROC SORT DATA=ohra_data8 OUT=ohra_data9;  
  BY DESCENDING sato;  
RUN;
```

Yllä oleva koodi järjestää aineiston rivit satomuuttujan arvojen mukaan laskevaan järjestykseen. Nousevan järjestyksen saa valittua jättämällä avainsanan DESCENDING pois koodista.

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

6.11 Osa-aineiston eli tiettyjen havaintorivien valitseminen

Data täytyy usein jakaa osiin, kuten esimerkiksi silloin, kun halutaan valita mukaan analyysiin vain tietty osaryhmä koko datasta. Osa-aineiston valinta voidaan toteuttaa määrittämällä tietty havaintojen valintaehto Query Builderissa tai Filter and Sort -työkalussa. Enterprise Guide 4.2:n myötä havaintorivien valitseminen voidaan useissa EG-työkaluissa tehdä myös ”Data”-välilehden ”Edit”-valinnan kautta, jolloin osa-aineiston valinta kohdistuu vain kyseiseen yksittäiseen tarkasteluun (tarkempi läpikäynti luvussa **Käytettävän datan vaihtaminen ja rivien valinta tietyissä työkaluissa**). Tämä ominaisuus on muun muassa useimmissa myöhemmin läpikäytävissä aineistoa kuvien tai tunnuslukujen avulla havainnollistavissa EG-työkaluissa. Samoissa EG:n työkaluissa voidaan määritellä myös ”Group analysis by” -muuttuja tai vastaava, jonka

avulla saadaan tehdyksi tarkastelut erikseen muuttujan kunkin luokan määrittelemälle osaryhmälle. Tässä yhteydessä tutustutaan tarkemmin Query Builderin ja Filter and Sort -työkalun mahdollisuuksiin, joiden avulla osa-aineiston valinnan tuloksena saadaan uusi muokattu SAS-data.

Havaintorivien valitseminen valintaehdon avulla Query Builderissa

Havaintorivien poistaminen tai tiettyjen rivien valitseminen kannattaa tehdä Enterprise Guidessa käyttäen esimerkiksi **Query Builderiä** (löytyy asiayhteysvalikosta muokattavan datan ollessa auki työtilassa), 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, jonka arvojen perusteella karsintaehto halutaan määrittää, vasemman reunan muuttujaikkunasta ”Filter Data” -ikkunaan. Tällöin aukeaa seuraava ehdon asettamisikkuna:

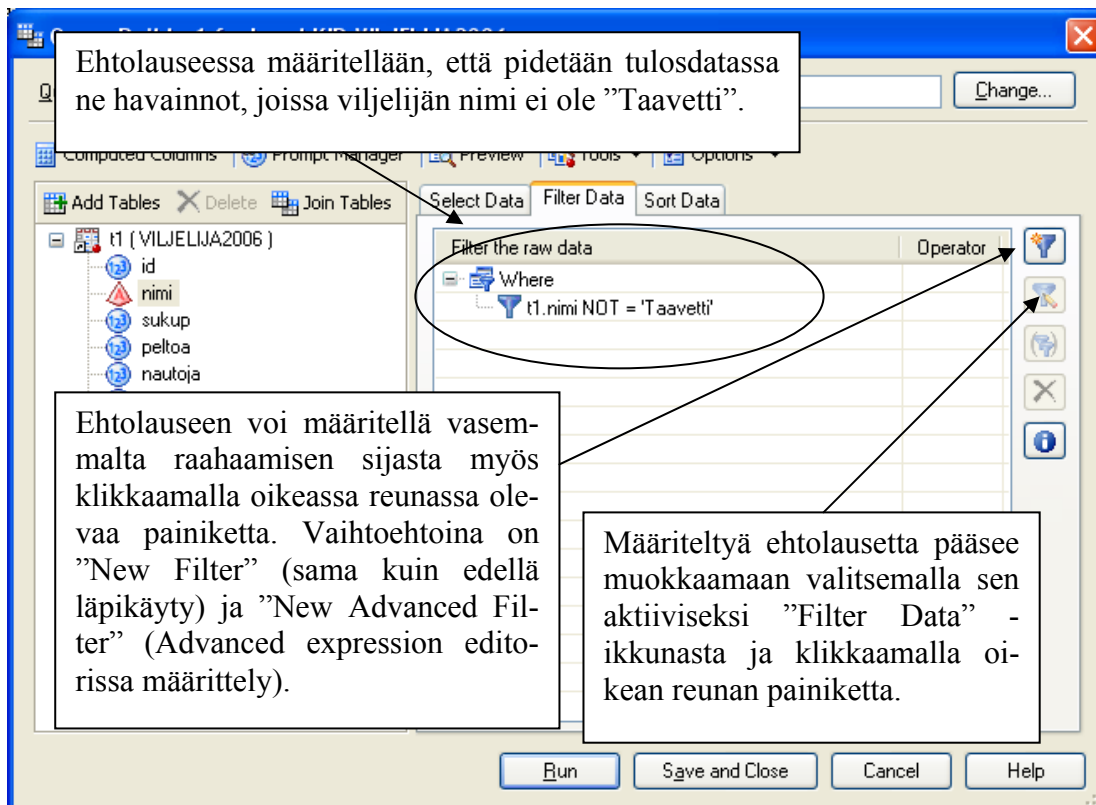
”Operator”-kohdan valintaikkunasta löytyy erilaisia ehto-operaattoreita.

Alasvetovalikon kautta päästään käsiksi muuttujien nimiin ja arvoihin, joita voidaan käyttää ”Value”-kohdassa.

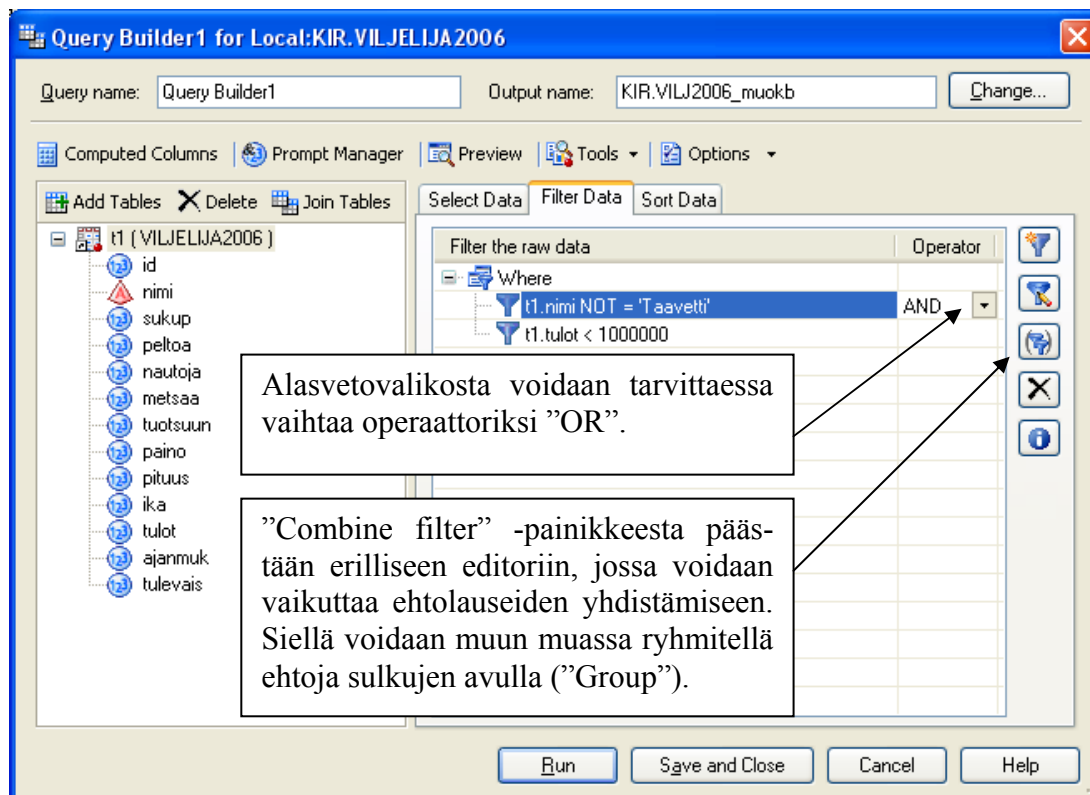
”Value”-kohtaan voi myös suoraan kirjoittaa muuttujan arvon, johon ehto kohdistuu.

Merkkimuotoiseen arvoon viitattaessa arvon pitää olla muodostuvassa SAS-koodissa yksinkertaisten lainausmerkkien sisällä, joten merkkimuotoisilla muuttujilla kannattaa yleensä pitää ruksi valintaruudussa ”Enclose values in quotes”.

”Next”-valinnalla päästään vielä yhteenvetosivulle ja painamalla ”Finish” takaisin Query Builderiin, johon tulee näkyviin uusi ehtolause ”Filter Data”-välilehdelle:



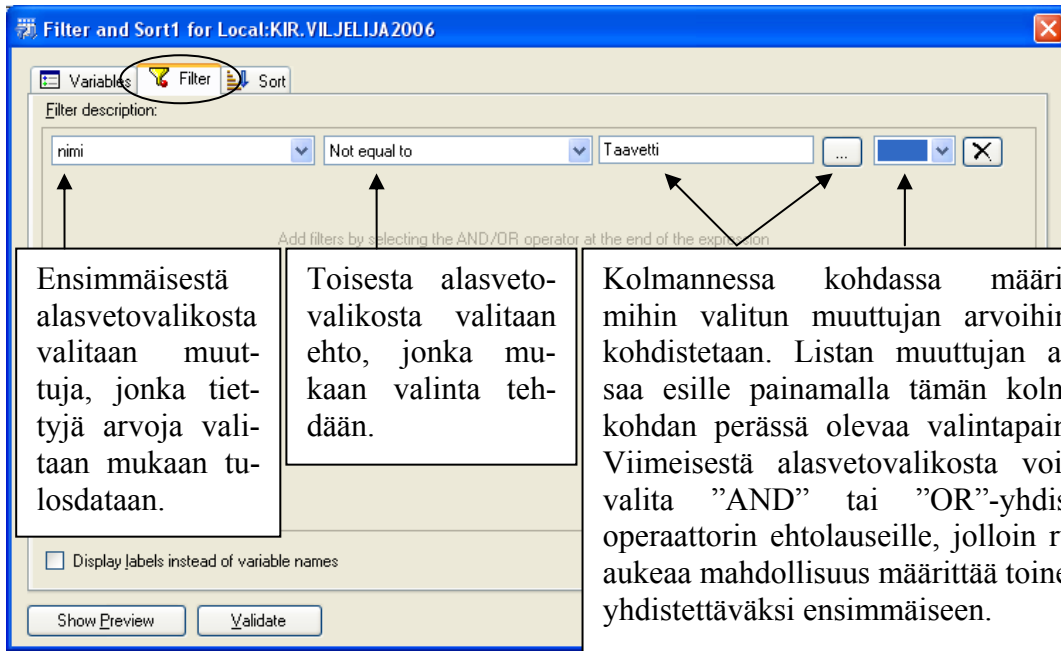
Havaintojen valinnassa voidaan käyttää myös useampaa kuin yhtä ehtolauseita. Tällöin tarvitaan AND (ja) tai OR (tai) ehtojen yhdistämisoperaattoria tai molempia. Voidaan esimerkiksi määrittellä, että poistetaan sellaiset henkilöt datasta, jotka ovat Jokioisista ja työskentelevät MTT:ssä tai vaihtoehtoisesti 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 "Operator"-kohdassa:



Kun ehto on määritelty valitaan "Run".

Havaintorivien valitseminen valintaehdon avulla Filter and Sort -työkalussa

SAS Enterprise Guiden version 4.2 myötä havaintorivien valintaa voidaan tehdä myös uuden Filter and Sort -työkalun avulla. Tämä työkalu saadaan käyttöön valitsemalla asiayhteysvalikosta **Filter and Sort**, muokattavan aineiston ollessa auki työtilassa. Tulodataan mukaan tulevien havaintorivien valinta tehdään määrittelemällä ehtolause "Filter"-välilehdellä:



Ensimmäisestä alavetovalikosta valitaan muuttuja, jonka tiedot valitaan mukaan tulodataan.

Toisesta alavetovalikosta valitaan ehto, jonka mukaan valinta tehdään.

Kolmannessa kohdassa määritellään mihin valitun muuttujan arvoihin ehto kohdistetaan. Listan muuttujan arvoista saa esille painamalla tämän kolmannen kohdan perässä olevaa valintapainiketta. Viimeisestä alavetovalikosta voi vielä valita "AND" tai "OR"-yhdistämisoperaattorin ehtolauseille, jolloin ruutuun aukeaa mahdollisuus määrittää toinen ehto yhdistettäväksi ensimmäiseen.

Filter and Sort -työkalu suoritetaan painamalla "OK". Tämän työkalun yhtenä heikkoutena on, että tulodataan nimeä ei voi valita ja se tehdään automaattisesti oletuskirjastoviitteen osoittamaan kansioon. Tämä asettaa rajoituksia työkalun käytettävyydelle.

Vaihtoehtoinen ratkaisu SAS-koodilla:

Tietyt havaintorivit voidaan SAS-koodissa valita esimerkiksi DATA-vaiheen WHERE-lauseessa:

```
DATA kir.vilj2006_muok;  
  SET kir.vilj2006;  
  WHERE nimi NE 'Taavetti';  
RUN;
```

Ennen tämän koodin suorittamista pitää muodostaa SAS-data nimeltä vilj2006 kir-kirjastoviitteen osoittamaan kurssikansioon, kuten seuraavassa tehtävässä kehoitetaan.

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 tulodata 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 tulodata 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 tulodataista pysyviä ja anna niille nimiksi **vilj_nainen** ja **vilj_mies**.

6.12 Kahden tai useamman datan yhdistäminen

Usein tutkimusaineistot ovat syystä tai toisesta tallennettu monena erillisenä datana, mutta esimerkiksi tilastollista käsittelyä varten ne pitää saada yhdistettyä. Ensimmäinen tilanne on sellainen, jossa yhdistettävissä datoissa 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 datat yhdistetään siten, että eri datojen saman tilastoyksikön havainnot tulevat yhdistetyssä datassa samalle riville.

Datojen yhdistäminen peräkkäin

Datoja voidaan Enterprise Guidessa yhdistää peräkkäin avaamalla ensin ensimmäinen yhdistettävä datamatriisi työtilaan tupla-klikkaamalla sen kuvaketta prosessidiagrammissa ja sen jälkeen valitsemalla asiayhteysvalikosta ”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.

Vaihtoehtoinen ratkaisu SAS-koodilla:

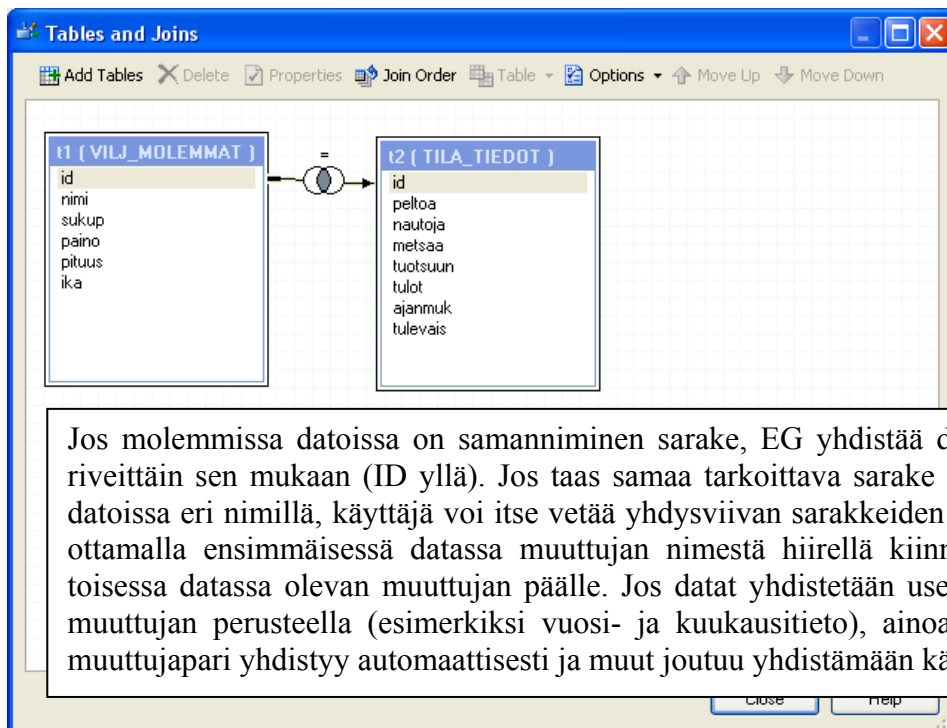
DATA-vaiheessa peräkkäin yhdistäminen tehdään luettelemalla yhdistettävät SAS-datat SET-lauseessa:

```
DATA kir.vilj_molemmat;  
  SET kir.vilj_nainen kir.vilj_mies;  
RUN;
```

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. Avataan ensimmäinen yhdistettävä datamatriisi työtilaan tupla-klikkaamalla sen kuvaketta prosessidiagrammissa ja sen jälkeen valitsemalla asiayhteysvalikosta ”Query Builder”. Tämän jälkeen painetaan ”Join Tables”-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”, valita kyseinen data ja painaa ”Open”. Tällöin ohjelma yhdistää datat ja näkymä on seuraava:



Oletuksena yhdistämisessä syntyvään tulosdataan otetaan mukaan vain sellaiset rivit, joihin löytyy tietoa molemmista alkuperäisistä datoista. Yhdistämisasetuksia pääsee muokkaamaan klikkaamalla hiiren oikealla painikkeella edellisessä kuvassa näkyvän yhdistämissiivan päällä ja valitsemalla ”Properties”, jolloin aukeaa seuraava ikkuna:

Inner Join = Tulodataan otetaan mukaan vain sellaiset rivit, joille löytyvät vastinparit yhdistettävistä datoista (oletustapa)

Left Join = Vasemmanpuoleisesta datasta otetaan kaikki rivit ja oikeanpuoleisesta vain siihen yhteensopivat.

Right Join = Oikeanpuoleisesta datasta otetaan kaikki rivit ja vasemmanpuoleisesta vain siihen yhteensopivat.

Full Outer Join = Molemmista datoista otetaan mukaan kaikki rivit.

Lisäksi listasta löytyy useita muita yhdistämisehtoja, joihin voi tutustua tarkemmin painamalla "Help"-painiketta.

Lopuksi valitaan "OK", "Close" ja "Run", jolloin datat yhdistyvät. Huomaa, että ennen "Run"-valintaa, kannattaa kuitenkin antaa tuloksena syntyvälle uudelle datalle nimi valitsemalla "Change" Query Builderin oikeassa yläkulmassa.

Datojen yhdistämisessä Query Builderissa voidaan vaihtoehtoisesti lähteä liikkeelle klikkaamalla ensin "Add Tables", jolloin haetaan ensin toinen data, jota ollaan yhdistämässä, ja vasta tämän jälkeen klikataan "Join Tables", josta päästään tarkastelemaan ja muokkaamaan datojen automaattista yhdistämistä.

Vaihtoehtoinen ratkaisu SAS-koodilla:

DATA-vaiheessa rinnakkain yhdistäminen tehdään luettelemalla yhdistettävät SAS-datatiedostot MERGE-lauseessa. Lisäksi muuttuja (tai muuttujat), jonka arvojen mukaan rivit yhdistetään, pitää mainita BY-lauseessa ja kumpikin data pitää ensin järjestää kyseisen yhdistämissuuttujan arvojen mukaan:

```
PROC SORT DATA=kir.vilj_molemmat;
  BY id;
RUN;

PROC SORT DATA=kir.tila_tiedot;
  BY id;
RUN;

DATA kir.vilj06_takaisin;
  MERGE kir.vilj_molemmat kir.tila_tiedot;
  BY id;
RUN;
```

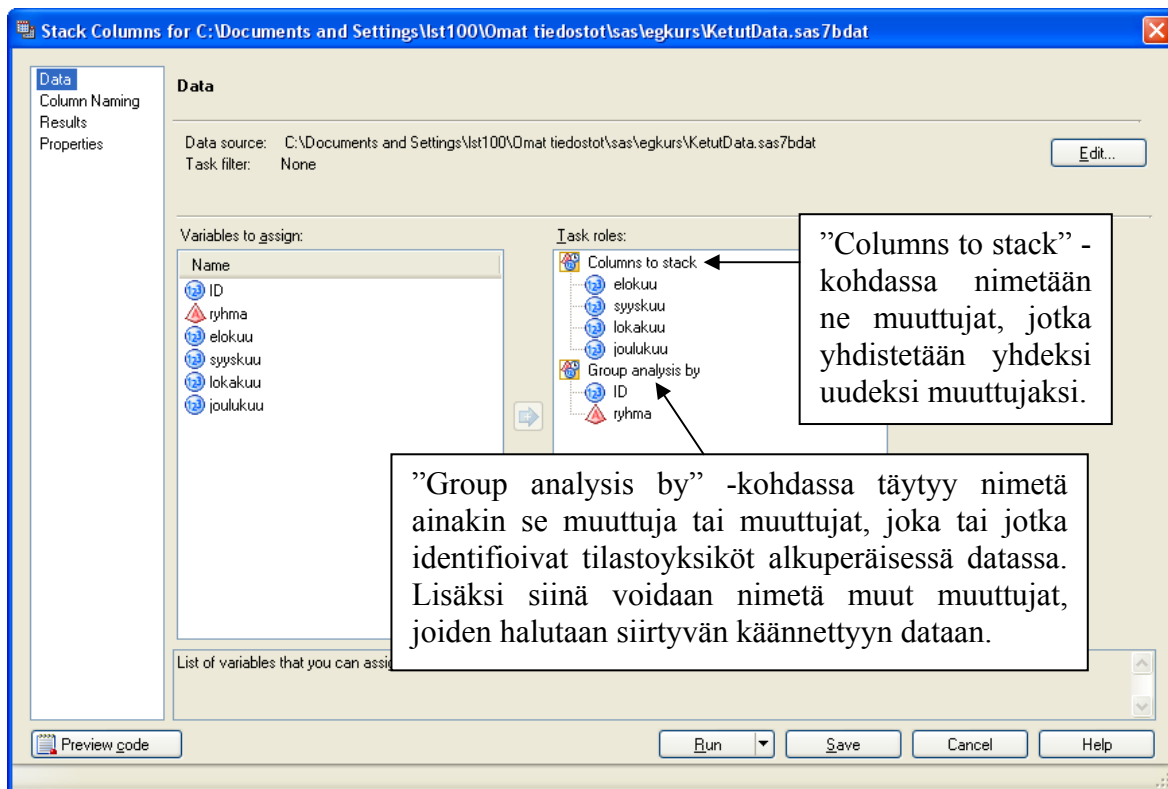
Tehtävä: Yhdistä rinnakkain edellisen tehtävän tuloksena syntynyt **vilj_molemmat** sekä aikaisemmassa tehtävässä luotu **tila_tiedot**. Tee lopputuloksena pysyvä SAS-data nimeltä **vilj06_takaisin**.

6.13 Datan kääntäminen

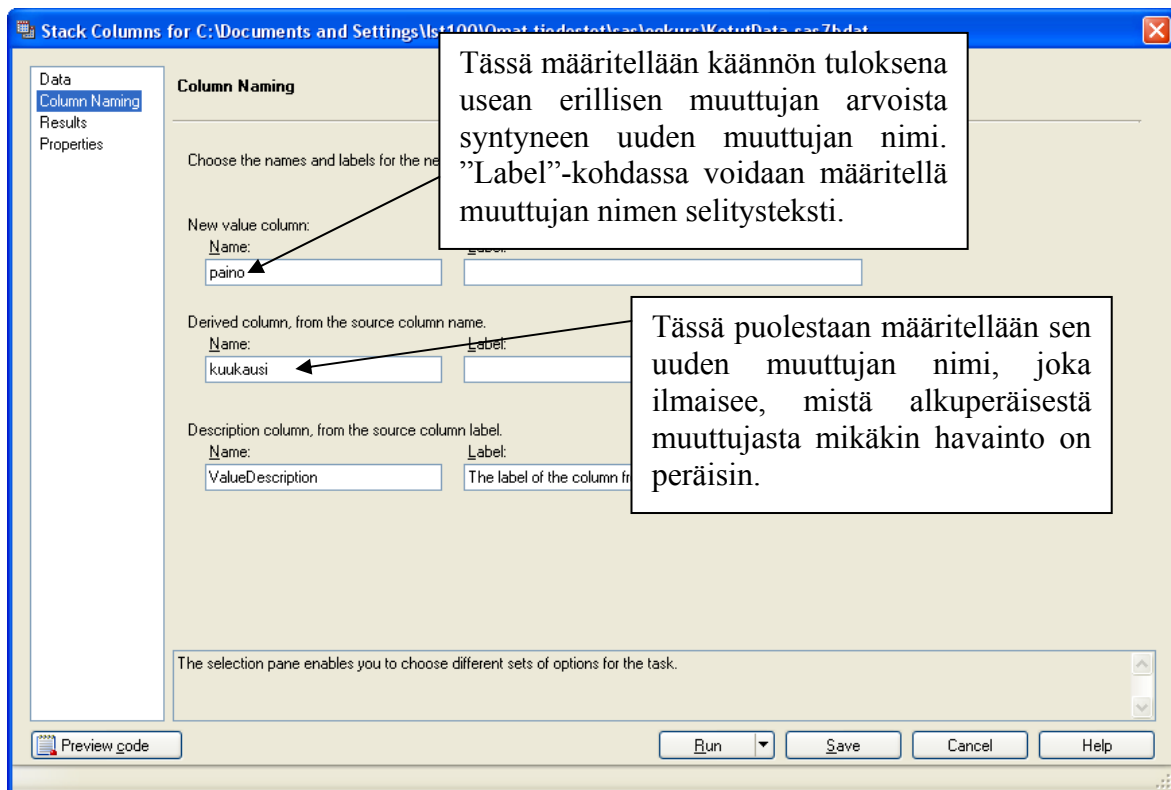
Dataa voidaan kääntää kahteen suuntaan. Ensimmäkin voidaan yhdistää usean sarakkeen tiedot yhteen sarakkeeseen, jolloin yleensä myös luodaan uusi muuttuja, joka ilmaisee missä sarakkeessa tieto oli lähtödatassa. Toisaalta data voidaan kääntää vastakkaiseen suuntaan eli jakaa yhden sarakkeen tieto useaan sarakkeeseen. Tällaisia aineiston kääntö tilanteita tulee 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 avaamalla käännettävä data-matriisi työtilaan tupla-klikkaamalla sen kuvaketta prosessidiagrammissa ja sen jälkeen valitsemalla asiayhteysvalikosta ”Data” -> ”Stack Columns”. Käännössä käytettävät ja mukaan otettavat muuttujat nimetään ”Data” -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. Lopuksi painetaan ”Run”, jolloin datan käyntö suoritetaan.

Vaihtoehtoinen ratkaisu SAS-koodilla:

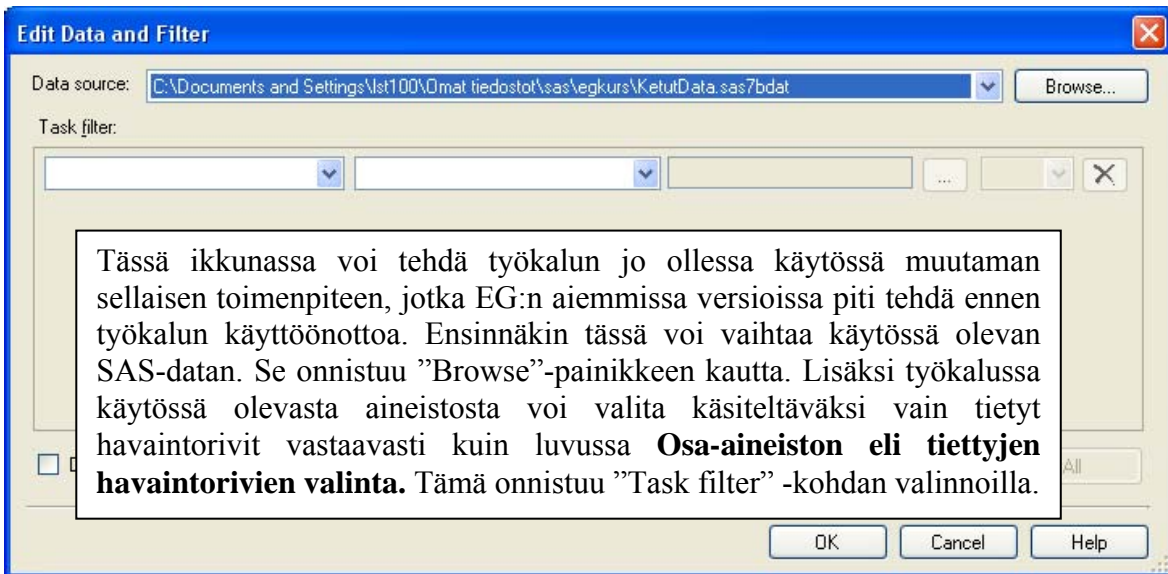
SAS-koodilla kääntämisen voi tehdä muun muassa käyttäen TRANSPOSE-proseduuria:

```
PROC TRANSPOSE DATA=kir.ketutdata OUT=kir.ketut_kaan NAME=kuukausi
                PREFIX=paino;
  VAR elokuu syyskuu lokakuu joulukuu;
  BY ID ryhma;
RUN;
```

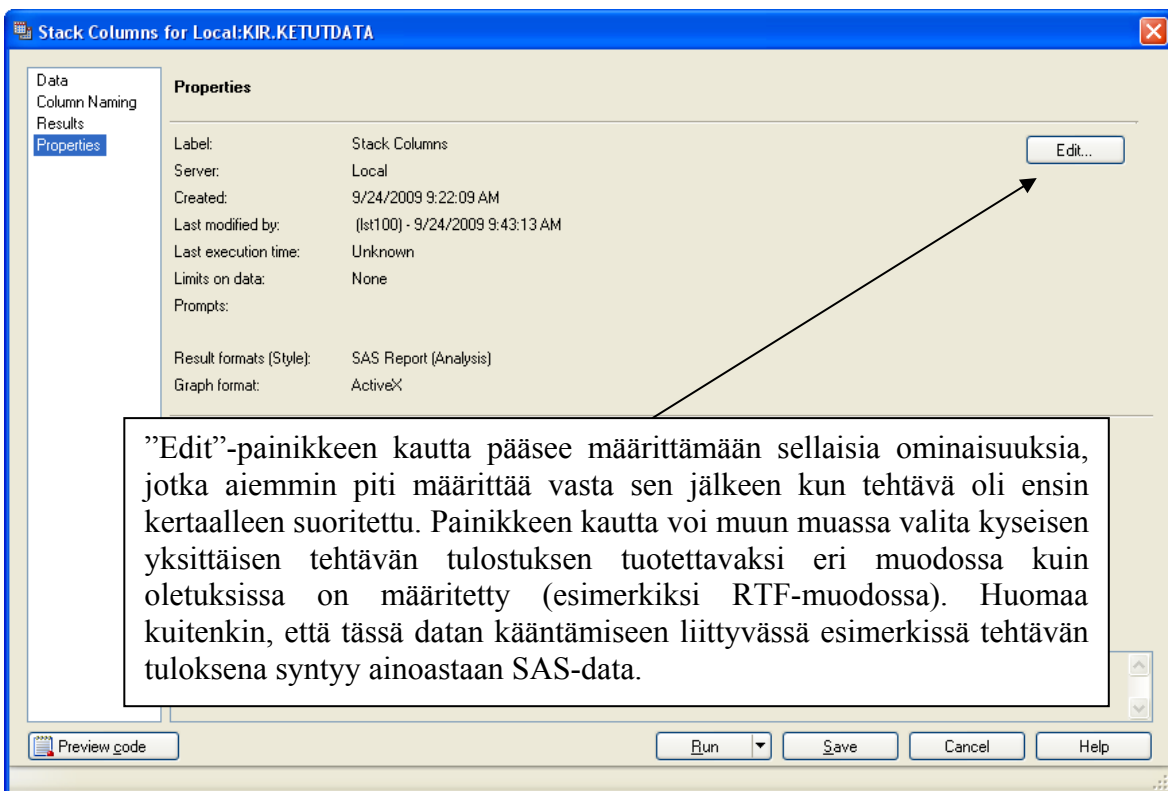
PROC TRANSPOSE-lauseen NAME-optio tuottaa kuukausi-nimisen muuttujan, joka saa arvoikseen VAR-lauseessa luetellut muuttujien nimet. PREFIX=paino valinnan seurauksena puolestaan uuden, kaikki kettujen painot sisältään pitävän, muuttujan nimeksi tulee ”paino1”.

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** tietokoneellesi.

Tämän työkalun yhteydessä on käytettävissä muutamia moniin työkaluihin liitettyjä Enterprise Guide 4.2:n uusia ominaisuuksia. Ensinnäkin data-välilehdellä on näkyvissä valintapainike ”Edit”, jota painamalla aukeaa seuraava ikkuna:



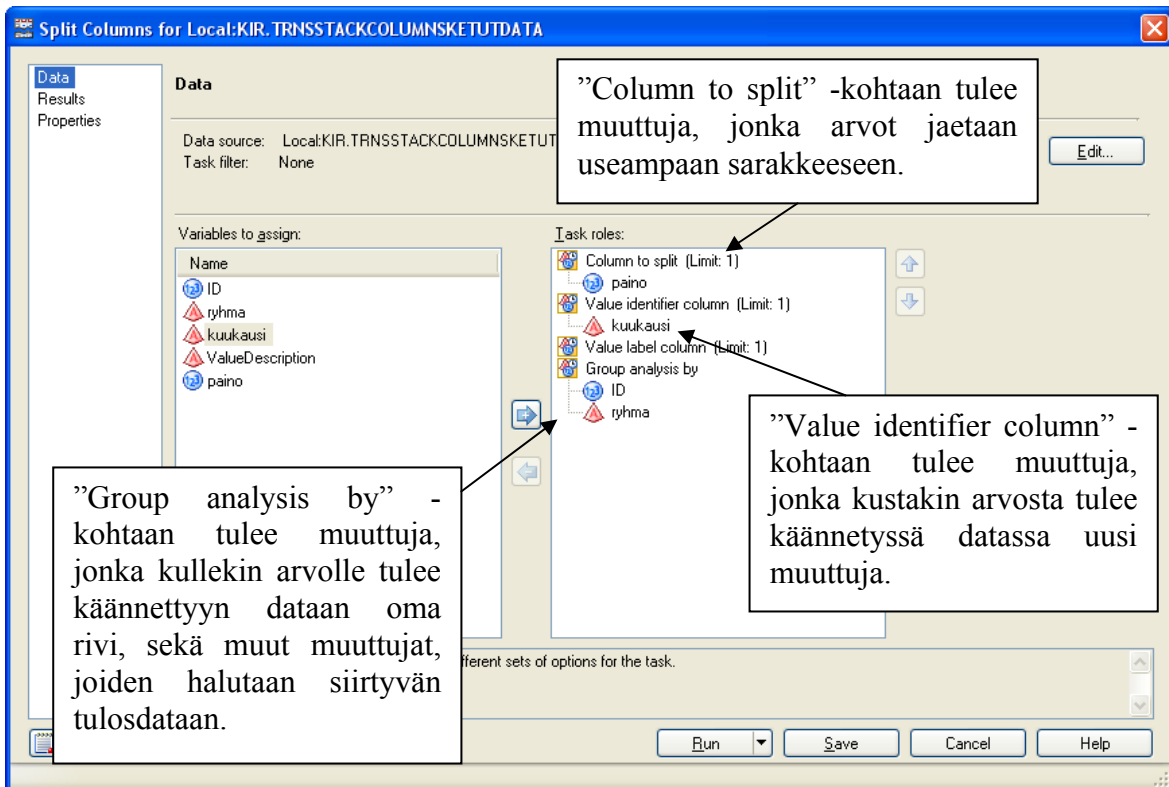
Toisena uutuuksena on ”Properties”-välilehti:



Näistä uusista ominaisuuksista löytyy yksityiskohtaisempaa tietoa oppaan loppupuolelta kappaleesta **Yksittäisen tulosteen hallinta ja jatkokäyttö**.

Yhdestä sarakkeesta useaan sarakkeeseen

Yhdestä sarakkeesta useaan aineisto voidaan kääntää avaamalla datamatriisi työtilaan tupla-klikkaamalla sen kuvaketta prosessidiagrammissa ja sen jälkeen valitsemalla asiayhteysvalikosta ”Data” -> ”Split Columns”, jolloin aukeaa seuraava välilehti:



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

Vaihtoehtoinen ratkaisu SAS-koodilla:

SAS-koodilla kääntämisen voi tehdä muun muassa käyttäen TRANSPOSE-proseduuria:

```
PROC TRANSPOSE DATA=kir.ketut_kaan OUT=ketut_takaisin;  
  BY id ryhma;  
  ID kuukausi;  
  VAR paino1;  
RUN;
```

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**.

6.14 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 sieltä poistettavaksi datoja sekä formaattitiedostoja. Esimerkiksi SASUSER-kirjastoviitteen osoittamaan kansioon saattaa helposti kertyä turhia datoja. Kannattaa kuitenkin olla varovainen, ettei vahingossa poista tarpeellisia datoja.

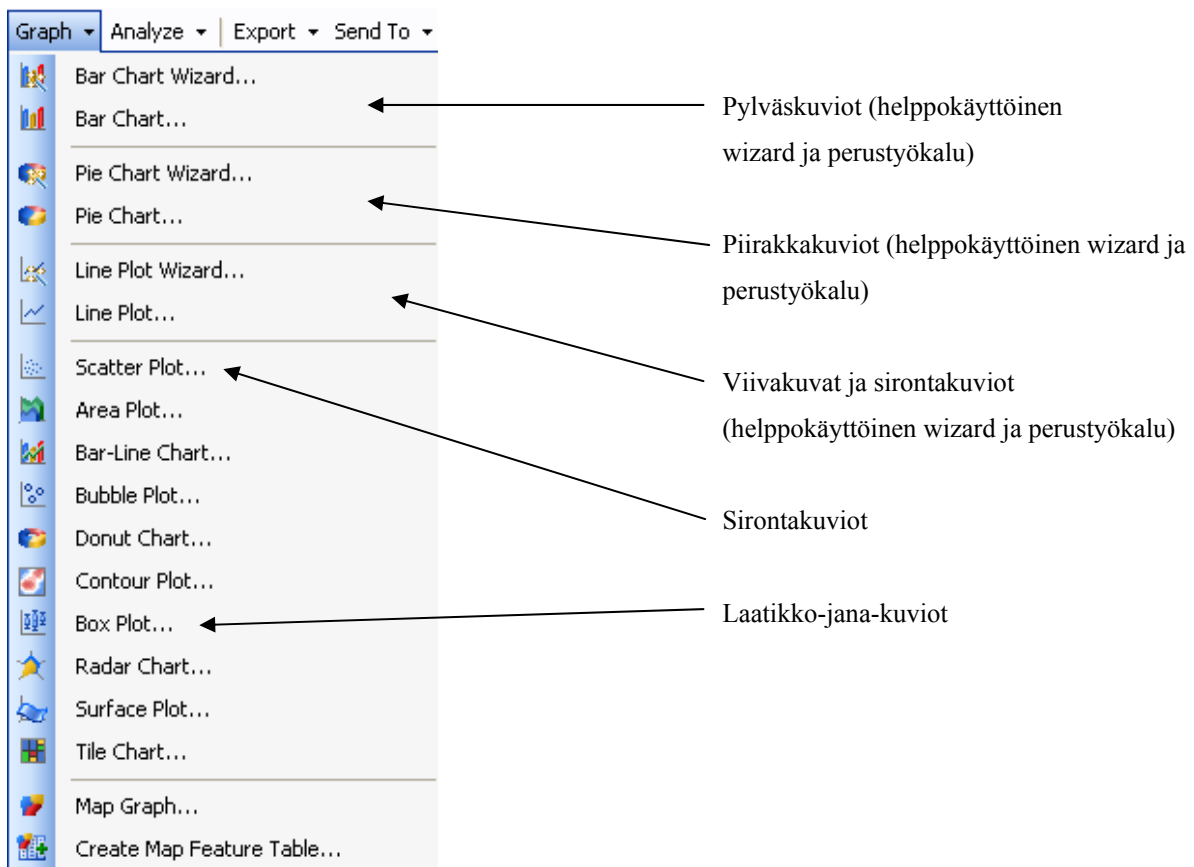
Lisäksi **Data**-valikosta löytyy työkalut ”**Upload Data Files to Server**” ja ”**Download Data Files to PC**”. Nämä liittyvät nimiensä mukaisesti data-tiedostojen siirtoon palvelimen ja paikallisen tietokoneen välillä.

7 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 kategoriisiin ja numeerisiin muuttujiin. Kategoristen muuttujien arvot ovat luokkia. Kategoriset muuttujat voidaan edelleen jakaa laatueroasteikollisiin ja järjestysasteikollisiin muuttujiin. Laatueroasteikollisen muuttujan luokkia ei voida laittaa järjestykseen (esimerkiksi värit), kun taas järjestysasteikollisilla muuttujilla luokkien järjestyksen määräävä sääntö on olemassa (esimerkiksi arvosteluasteikko: ”Huono”, ”Tyydyttävä”, ”Hyvä”, ”Erinomainen”). Numeerisilla muuttujilla puolestaan kukin tilastoyksikkö saa määrää kuvaavan numeerisen arvon (esimerkiksi pituus- ja painohavainnoilla). Numeeriset muuttujat voidaan myös edelleen jakaa välimatka- ja suhdeasteikon sekä absoluuttisen asteikon muuttujiin.

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

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



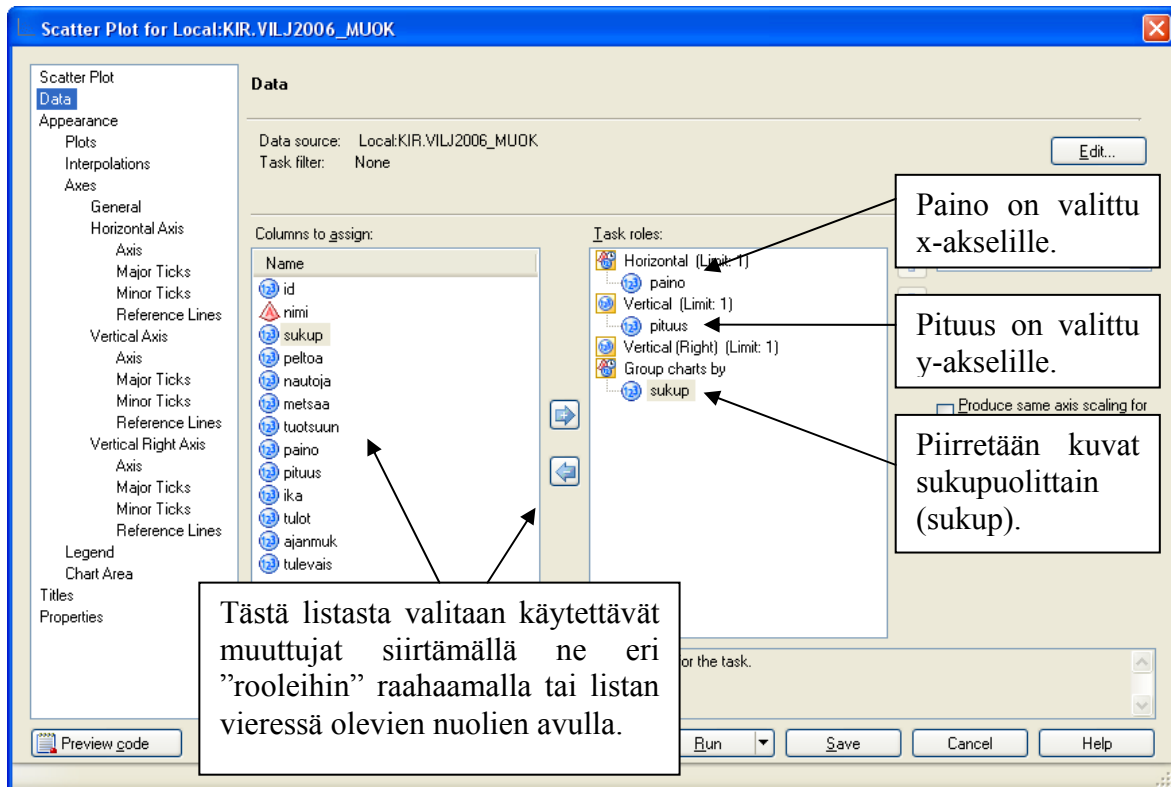
Graph-valikkoa voidaan käyttää joko asiayhteysvalikon tai ylävalikon ”Tasks”-valikon kautta.

7.1 Sirontakuviio (Scatter Plot)

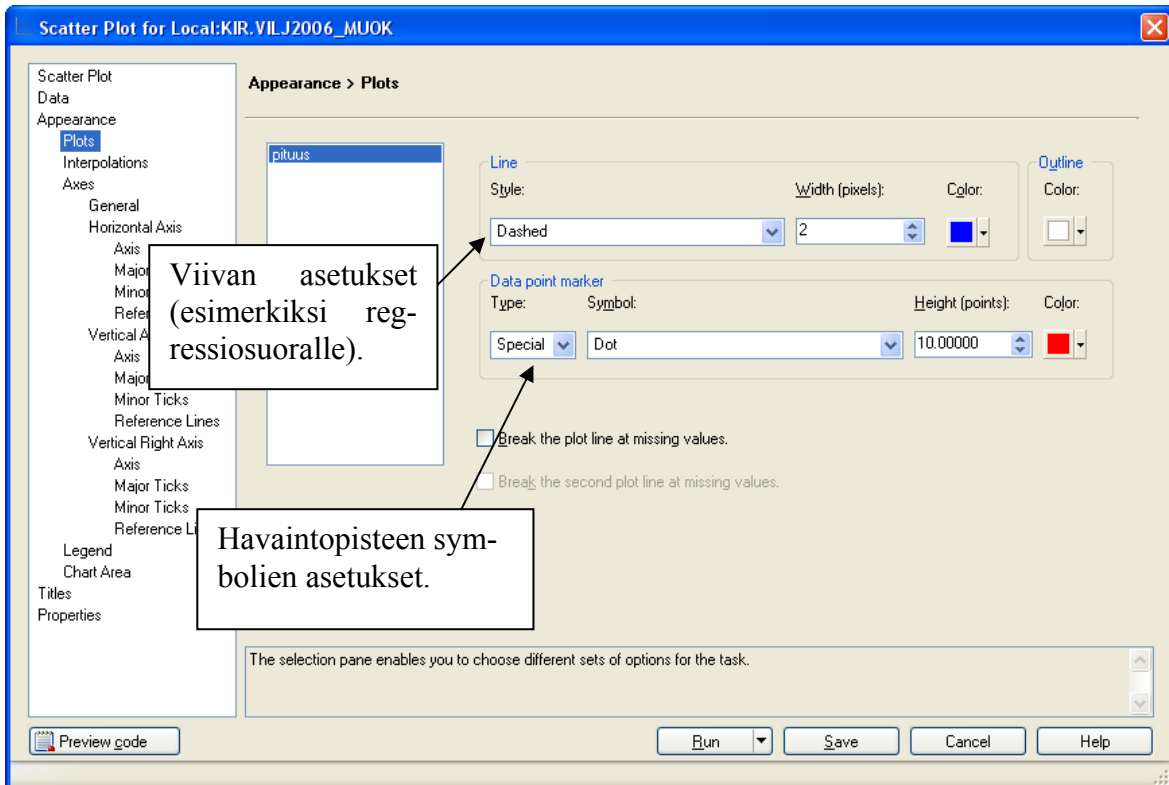
Sirontakuvioksi kutsutaan kuvaa, jolla tarkastellaan kahden, yleensä numeerisen, muuttujan yhteyttä. Kuvio muodostetaan piirtämällä koordinaatistoon havaintoparit (x_i, y_i) , jossa x_i on vaakakselimuuttujan arvo ja y_i pystyakselimuuttujan arvo tilastoyksiköllä i . Sirontakuviio piirtämiseen löytyy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

Tavallinen sirontakuviio

Käytettävä data avataan ensin työtilaan tupla-klikkaamalla datakuvaketta prosessidiagrammissa ja tämän jälkeen valitaan asiayhteysvalikosta ”Graph” -> ”Scatter Plot”. ”Scatter Plot” -välilehdeltä valitaan tavallisen sirontakuviio piirtämiseksi ”2D Scatter Plot”. ”Data” -välilehdellä valitaan muuttujat, joiden arvoista sirontakuviio piirretään:

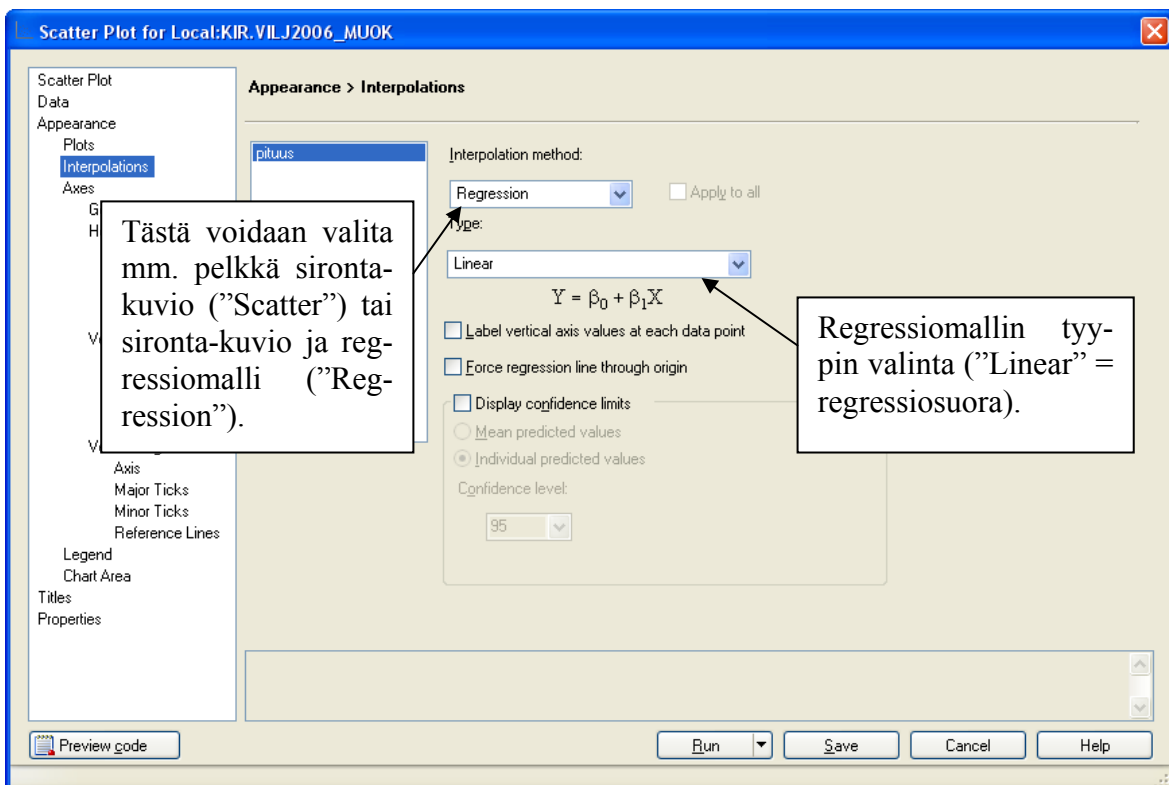


”Plots”-välilehdellä voidaan muokata kuvan ulkoasua:

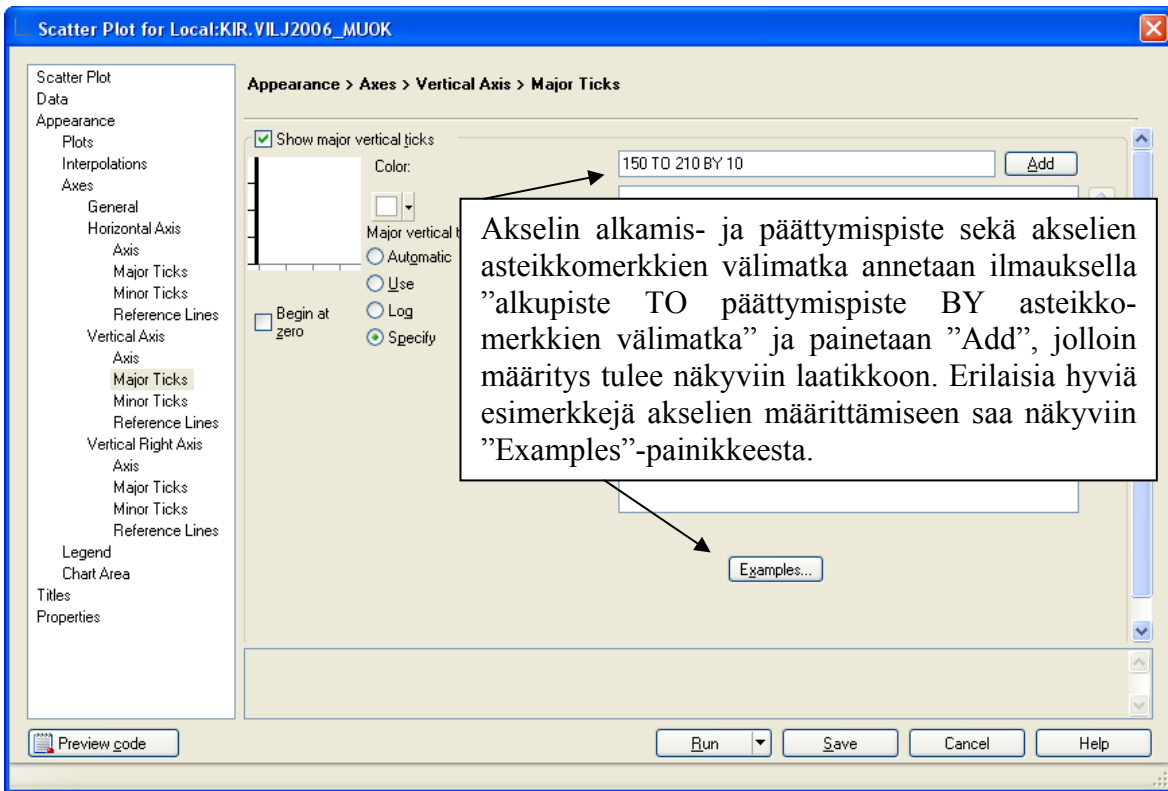


Kuvan asetuksia voi vaihtaa HTML-tulosteissa myös valmiista kuvasta siten, että painetaan hiiren oikeata painiketta kuvan päällä ja valitaan ”Graph Properties”. Vastaava kuvien jälkimuokaus onnistuu myös SAS Report -oletustulostuskohteeseen tehdyillä kuvilla, jos tulosteista on ensin tehty niin sanottu SAS-raportti (katso lisätietoa kappaleesta **Kuvien tulostuskohde ja yleisasetukset**).

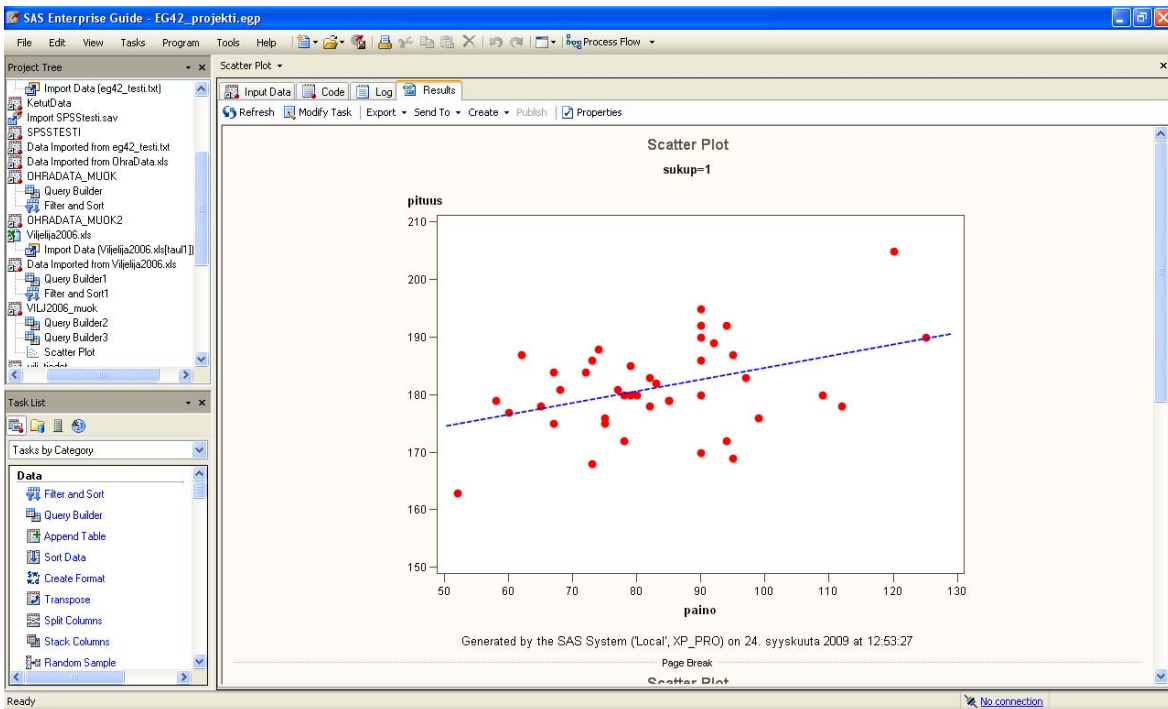
”Interpolations”-välilehdellä voidaan valita esimerkiksi regressiosuora sovitettavaksi pisteparveen:



Akselien pituudet voi määrittellä oletusarvoista poikkeaviksi eri akseleihin ("Horizontal Axis", "Vertical Axis" ja "Vertical Right Axis") liittyvillä "Major Ticks" -välilehdillä valitsemalla oletuksena olevan "Automatic" vaihtoehdon sijasta "Specify", jolloin aukeaa seuraava näkymä:



Lisäksi eri välilehdillä voidaan tehdä muun muassa akseleihin, referenssiivivoihin sekä kuvan otsikoihin ja kokoon liittyviä määrittelyjä. Lopuksi valitaan "Run", jolloin sirontakuviot tulostuu tehdyillä valinnoilla:



Vaihtoehtoinen ratkaisu SAS-koodilla:

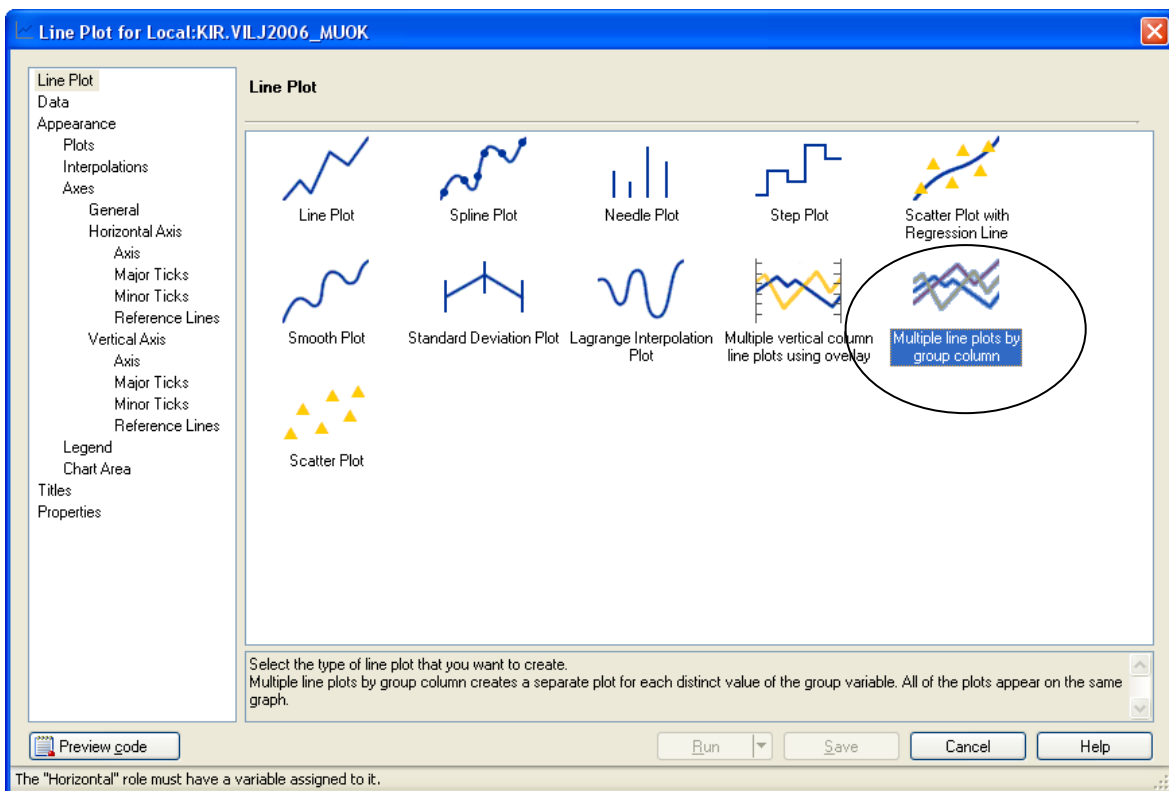
Sirontakuvion voi piirtää esimerkiksi käyttäen GPLOT-proseduuria:

```
PROC SORT DATA=kir.vilj2006_muok;  
  BY sukupuoli;  
RUN;  
  
GOPTIONS RESET=all;  
PROC GGPLOT DATA=kir.vilj2006_muok;  
  BY sukupuoli;  
  PLOT paino*pituus;  
  SYMBOL COLOR=blue VALUE=dot INTERPOL=rl;  
RUN;
```

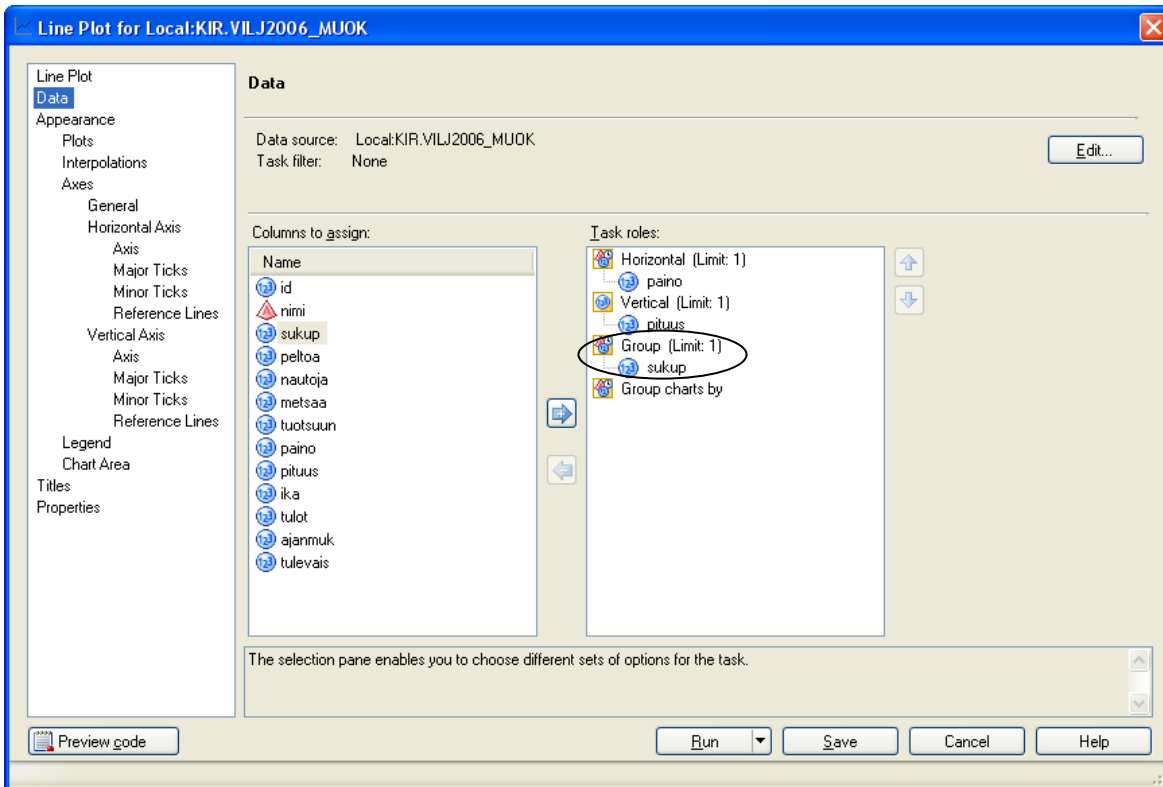
Huomaa, että kun koodissa käytetään BY-lausetta, pitää aineisto aina ensin järjestää BY-muuttujan arvojen mukaan.

Sirontakuviassa ryhmät eri symbolilla samassa kuvassa

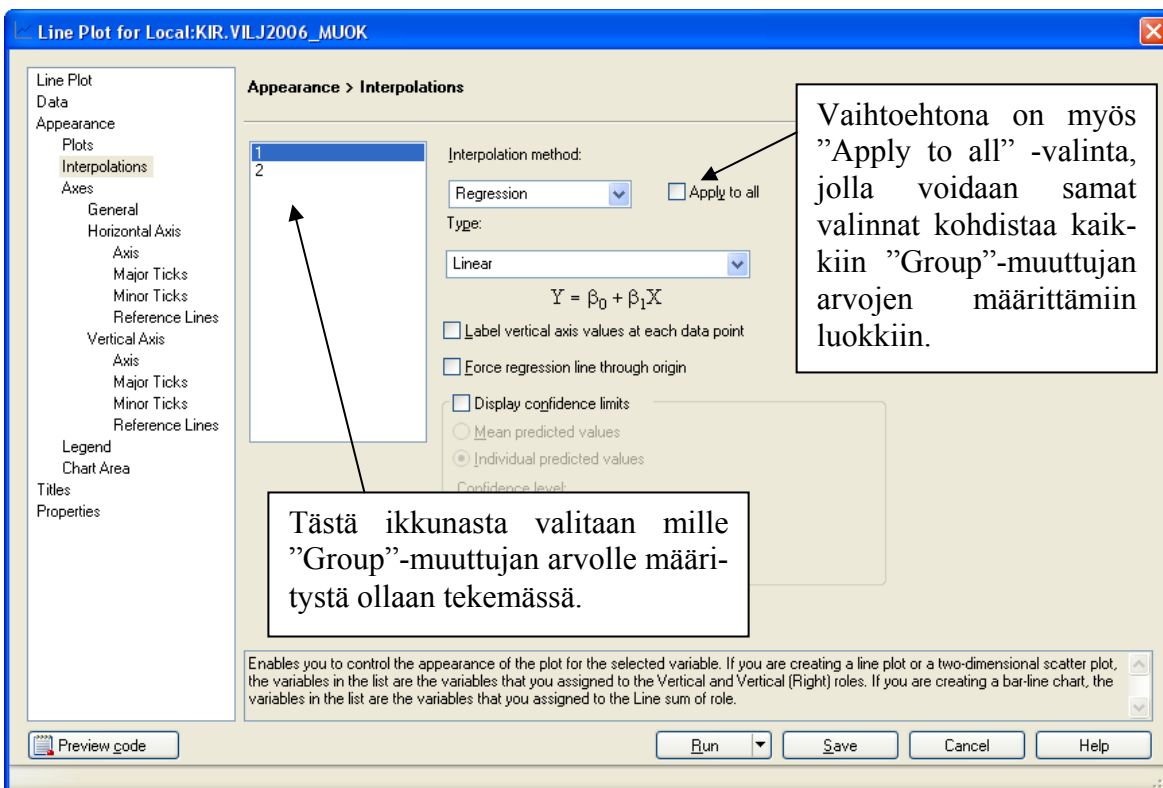
Jos halutaan piirtää sirontakuviota, jossa pisteet on merkitty eri värillä tai symbolilla jonkin ryhmittelevän muuttujan luokkien mukaisesti, se voidaan tehdä "Line Plot" -työkalulla. Avataan käytettävä datamatriisi työtilaan ja valitaan asiayhteysvalikosta "Graph" -> "Line Plot" ja avautuvalla "Line Plot"-välilehdellä "Multiple line plots by group column":



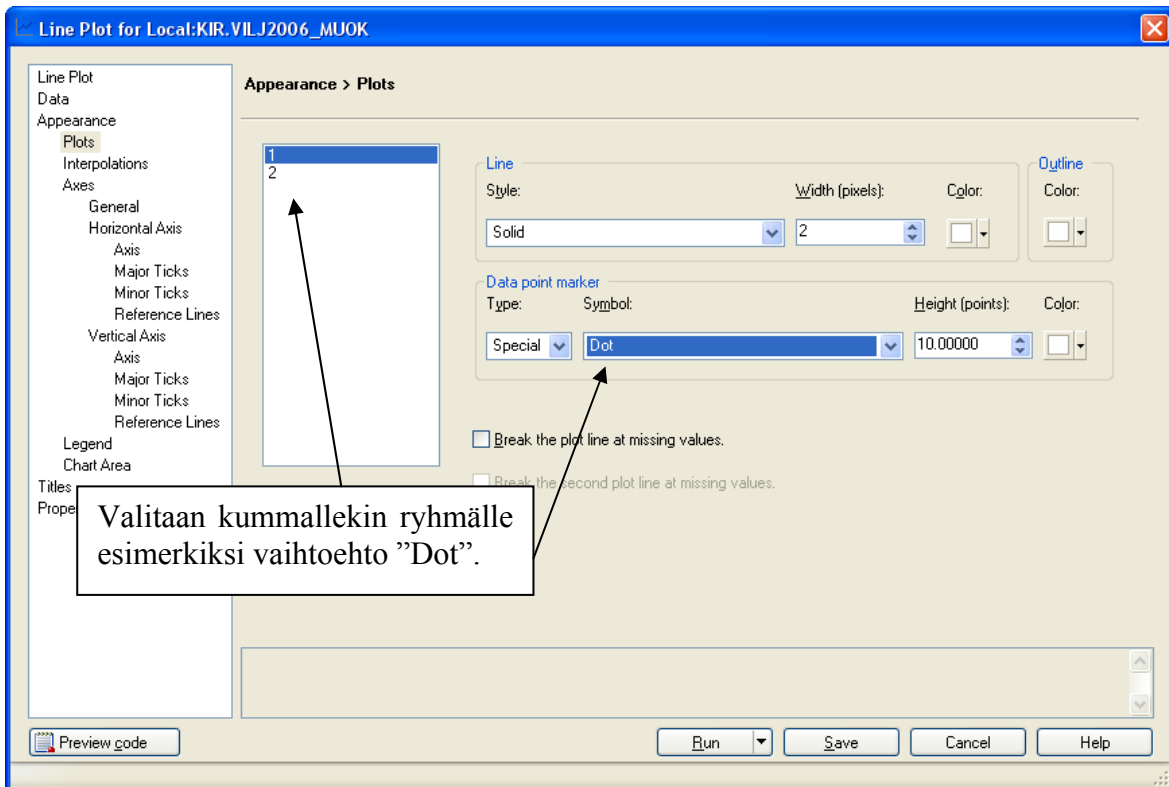
”Data” -välilehdellä ryhmän määräävä muuttuja laitetaan kohtaan ”Group”:



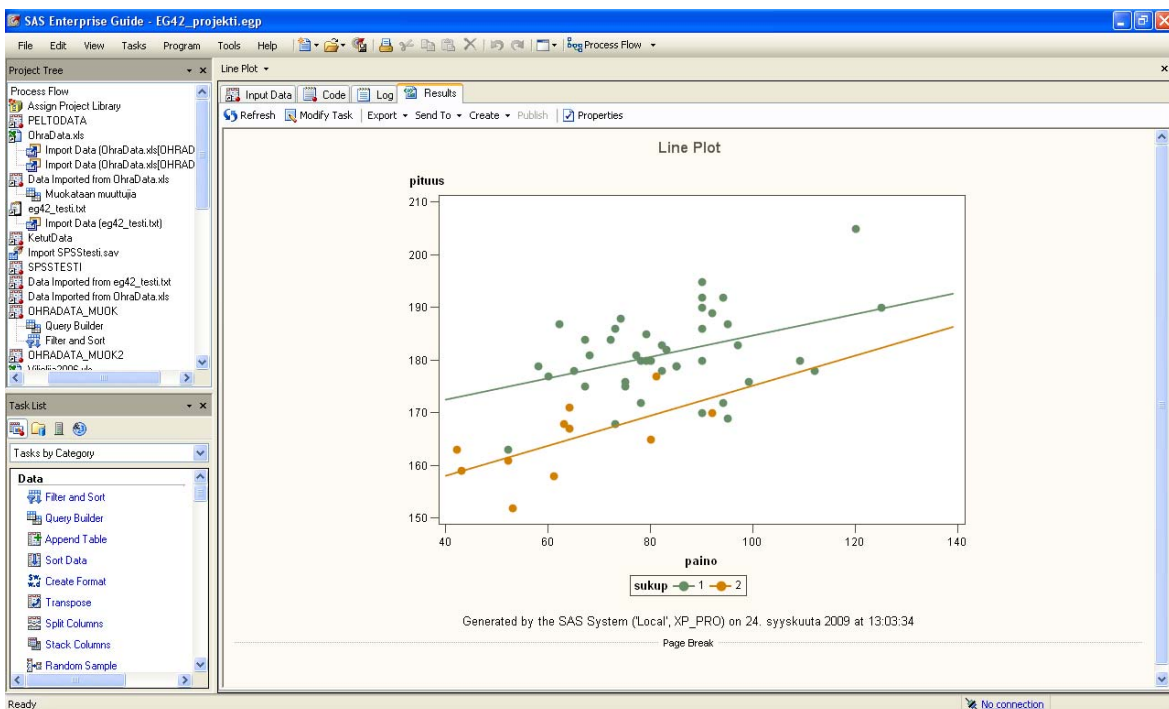
”Interpolations”-välilehdellä kohdassa ”Interpolation method” valitaan joko ”Scatter” tai ”Regression” jokaiselle ryhmälle erikseen:



Jos ”Interpolations”-välilehdellä valitaan regressioyhtälön sovittaminen pelkän sirontakuvion sijaan, ”Plots”-välilehdeltä pitää vielä valita pisteille jokin symboli ”none”-vaihtoehdon sijaan, haluttaessa piirtää sirontakuvion myös yksittäiset pisteet:



Lopuksi valittaessa "Run" piirtyy seuraava kuva työtilaan:



Vaihtoehtoinen ratkaisu SAS-koodilla:

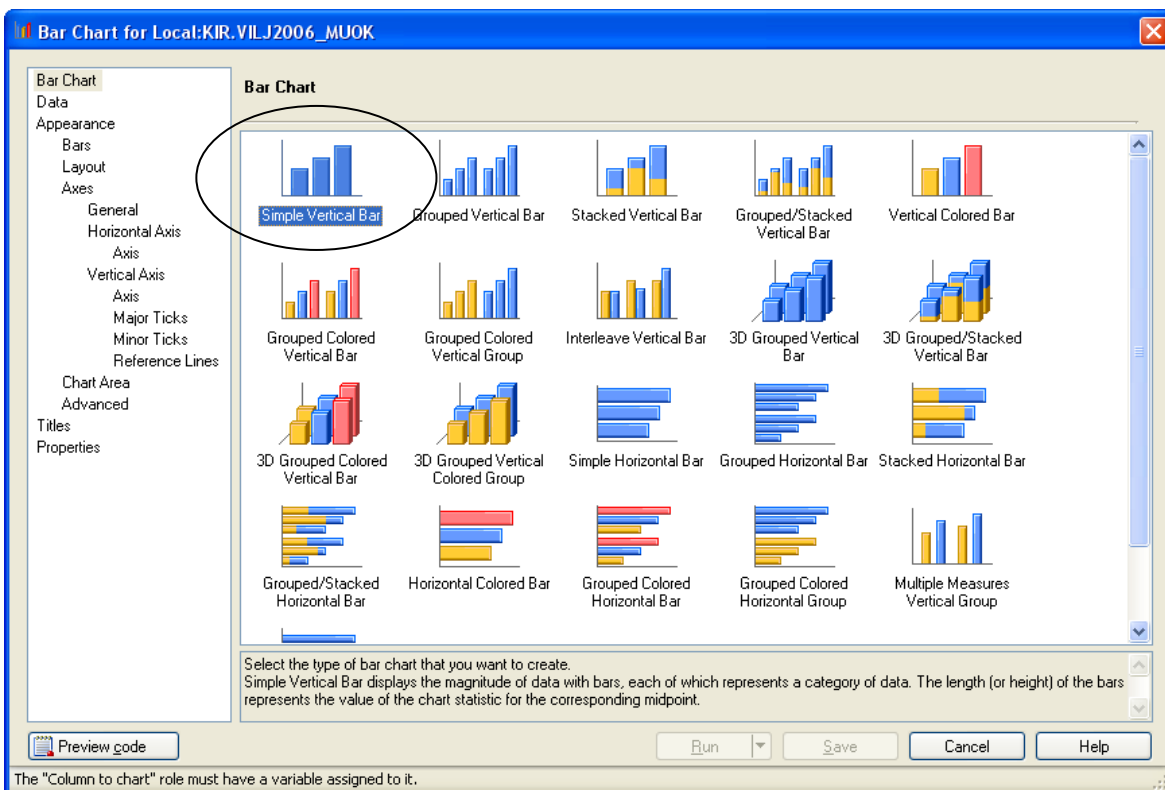
Sirontakuvion voi piirtää esimerkiksi käyttäen GPLOT-proseduuria:

```
GOPTIONS RESET=all;
PROC GPLOT DATA=kir.vilj2006_muok;
  PLOT paino*pituus=sukupu;
  SYMBOL1 COLOR=blue VALUE=dot INTERPOL=r1;
  SYMBOL2 COLOR=red VALUE=circle INTERPOL=r1;
RUN;
```

Tehtävä: Käytä projektissa olevaa muokattua Viljelijä 2006 -dataa nimeltä **vilj2006_muok**. Piirrä sirontakuviio, 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.

7.2 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 avaamalla käytettävä datamatriisi työtilaan tupla-klikkaamalla sitä prosessidiagrammissa ja valitsemalla asiayhteysvalikosta **"Graph" -> "Bar Chart"**. Tällöin aukeaa "Bar Chart" -ruutu, josta voidaan valita kuvan tyyppi:



Samoin kuin sirontakuvioiden yhteydessä muuttujille annetaan ”roolit” välilehdellä ”Data”:

”Column to chart” -kohtaan siirretään muuttuja, jonka eri arvoille piirretään erilliset pylväät.

Jos halutaan kokonaan erilliset pylväskuviot joillekin osaryhmille, ryhmittelevä muuttuja määrätään kohdassa ”Group charts by”.

Jos jokin muu kuin ”Column to chart” -muuttuja pitää sisällään tiedon pylväiden pituuksista, kyseisen muuttujan nimi tulee kohtaan ”Sum of”. Jos taas ”Column to chart” -muuttuja pitää sisällään kaiken tarvittavan tiedon, tämä kohta jätetään tyhjäksi.

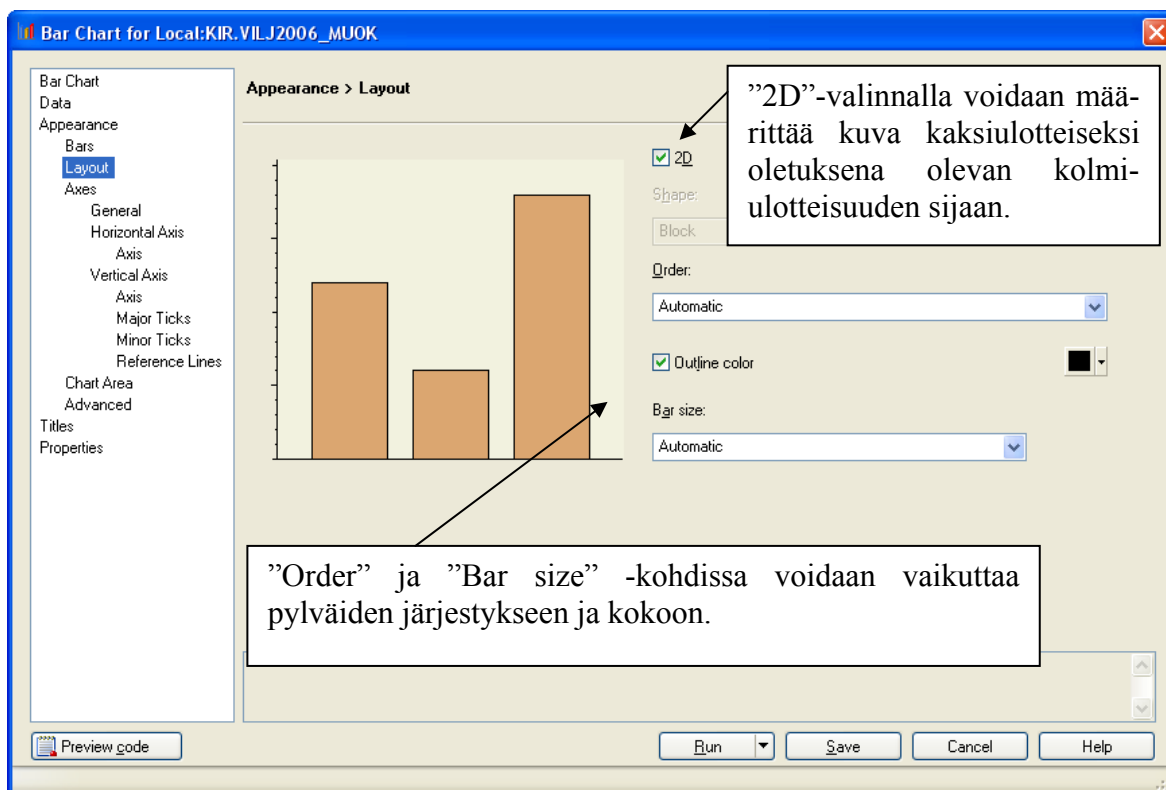
”Sum of” -kohtaan voidaan sijoittaa muuttuja, jonka arvojen summa tai keskiarvo määrää pylvään pituuden.

”Bars”-välilehdellä voidaan tehdä seuraavia määrytyksiä:

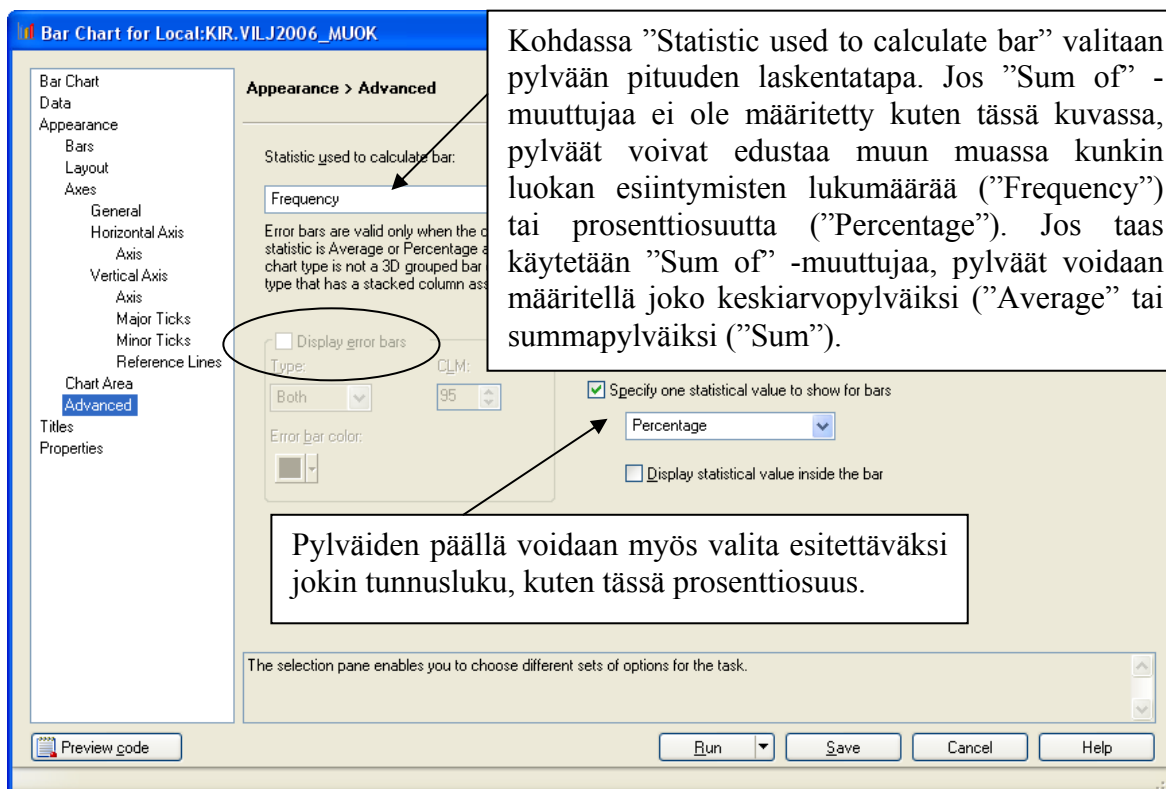
Pylväiden väreiksi voidaan määrittää joko oletusvärit tai käyttäjä voi antaa värit ”Custom colors” -kohdassa.

Määritetään, että ”Column to chart” -muuttujan jokaiselle havaitulle arvolle tulee oma pylväs (saadaan poistettua turhat pisteet x-akselilta numeerisen muuttujan tapauksessa).

”Layout”-välilehdellä voidaan tehdä erilaista kuvan ulkoasuun liittyvää määrittelyä:



”Advanced”-välilehdellä puolestaan voidaan muun muassa valita pylvään pituuden laskentatapa:

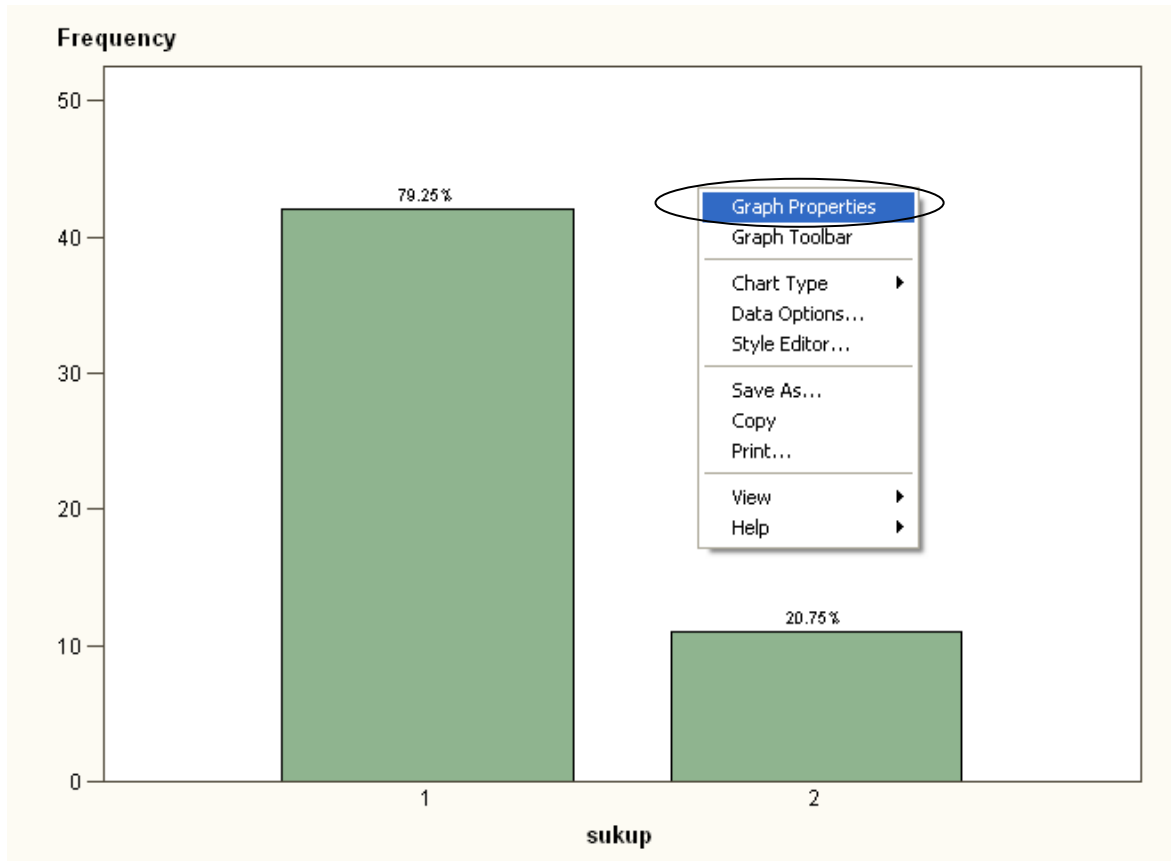


Edellisen kuvan ”Display error bars” -kohdassa puolestaan voidaan lisätä prosentti- tai keskiarvopylväille luottamusvälipalkit. Luottamusvälien määrääminen riippuu tutkimusasetelmasta. Valinnalla tuotettavat keskiarvojen luottamusvälit soveltuvat asetelmalle, jossa kaikki havainnot ovat toisistaan riippumattomia ja peräisin normaalijakaumasta. Prosenttiosuuksien luottamusvälillä puolestaan voi käyttää suurten otosten tilanteessa, jossa kaikkien havaintojen riippumatto-

muus oletus on voimassa. Kannattaa siis varmistaa, että aineisto perustuu siihen tutkimusasetelmaan, johon EG:n esittämät luottamusvälit soveltuvat käytettäväksi.

Muilta välilehdiltä löytyy erilaisia kuvan ulkoasuun vaikuttavia asetuksia (muun muassa otsikoihin, akseleihin, referenssiivivoihin ja kuvan kokoon liittyviä määrittäjäasetyksiä).

Valmista kuvaa pääsee tässäkin tapauksessa muokkaamaan 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”). Vastaava kuvien jälkimuokkaus onnistuu myös SAS Report -oletustulostuskohteeseen tehdyillä kuvilla, jos tulosteista on ensin tehty niin sanottu SAS-raportti (katso lisätietoa kappaleesta **Kuvien tulostuskohde ja yleisasetukset**).

Vaihtoehtoinen ratkaisu SAS-koodilla:

Pylväskuvion voi piirtää esimerkiksi käyttäen GCHART-proseduuria:

```
GOPTIONS RESET=all;  
PROC GCHART DATA=kir.vilj2006_muok;  
  VBAR sukup / DISCRETE TYPE=FREQ;  
RUN;
```

Jos pylväiden pituuden halutaan kuvastavan esimerkiksi paino-muuttujan keskiarvoa, voidaan käyttää seuraavaa koodiratkaisua:

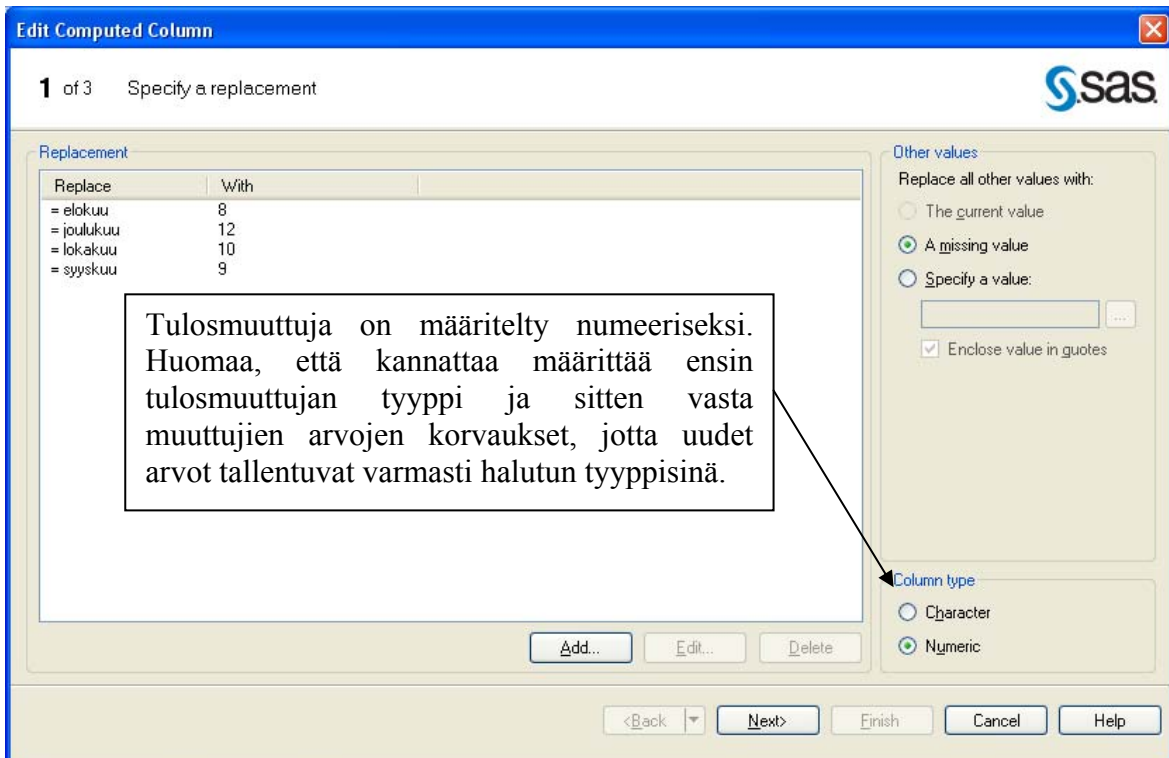
```
GOPTIONS RESET=all;  
PROC GCHART DATA=kir.vilj2006_muok;  
  VBAR sukup / SUMVAR=paino DISCRETE TYPE=MEAN;  
RUN;
```

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

7.3 Viivakuva (Line Plot)

Viivakuva on muuten kuin sirontakuvi, mutta 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, jolloin saadaan aikaan yksilökohtaiset profiilikuvat. Tilastoyksiköt tai niiden muodostamat ryhmät voidaan erottaa toisistaan erilaisella viivatyyppillä, symbolilla tai piirtämällä ryhmille omat kuvat.

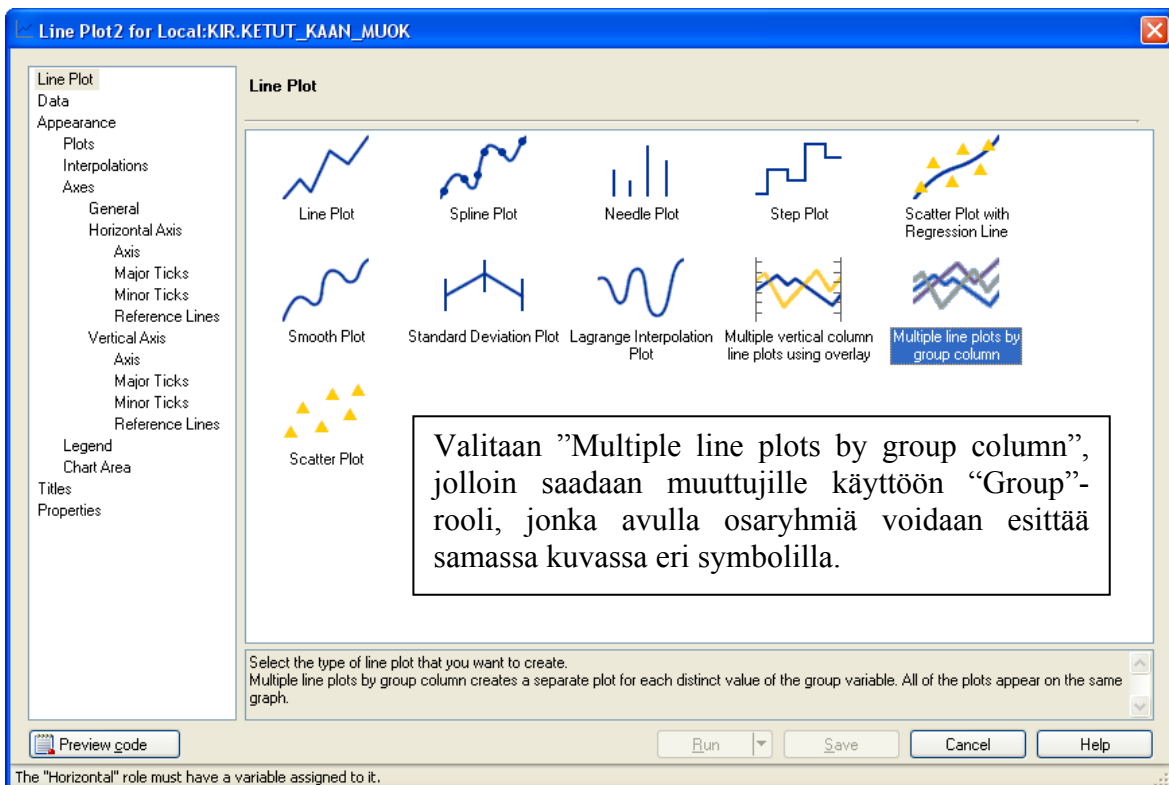
Enterprise Guide 4.2 järjestää x-akselille tulevan muuttujan arvot automaattisesti järjestykseen akselille, joten datan manuaalista järjestämistä ei tässä EG:n versiossa enää aiemmista versioista poiketen tarvitse tehdä. Merkkimuotoiset muuttujan arvot ovat kuitenkin usein ongelmallisia, koska ohjelma 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 ”Computed Columns” -> ”New” -> ”Recoded column”, jossa voidaan tehdä seuraavan ikkunan mukainen määrittely:



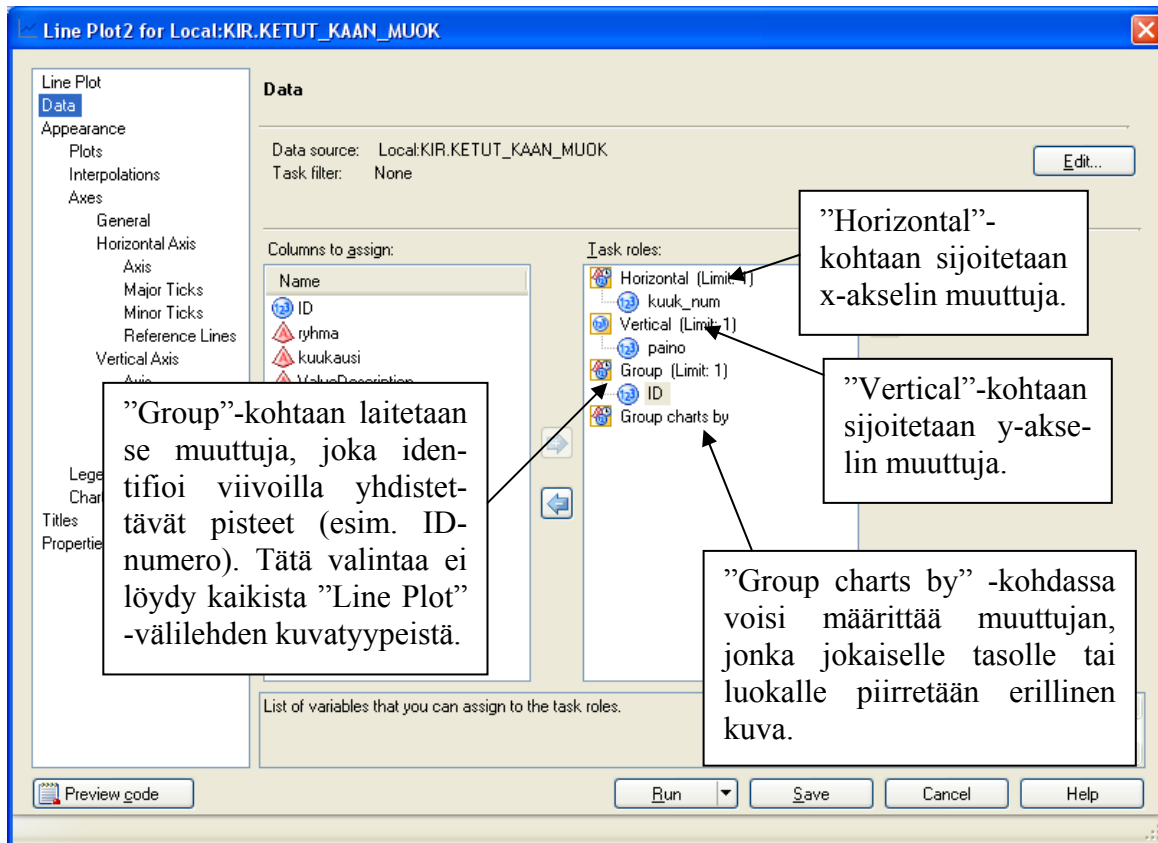
Query Builderin tuloksena saatavan datan numeerista kuukausi-muuttujaa voidaan nyt käyttää viivakuvan x-akselin muuttujana.

Viivakuva – yksilökäyrät

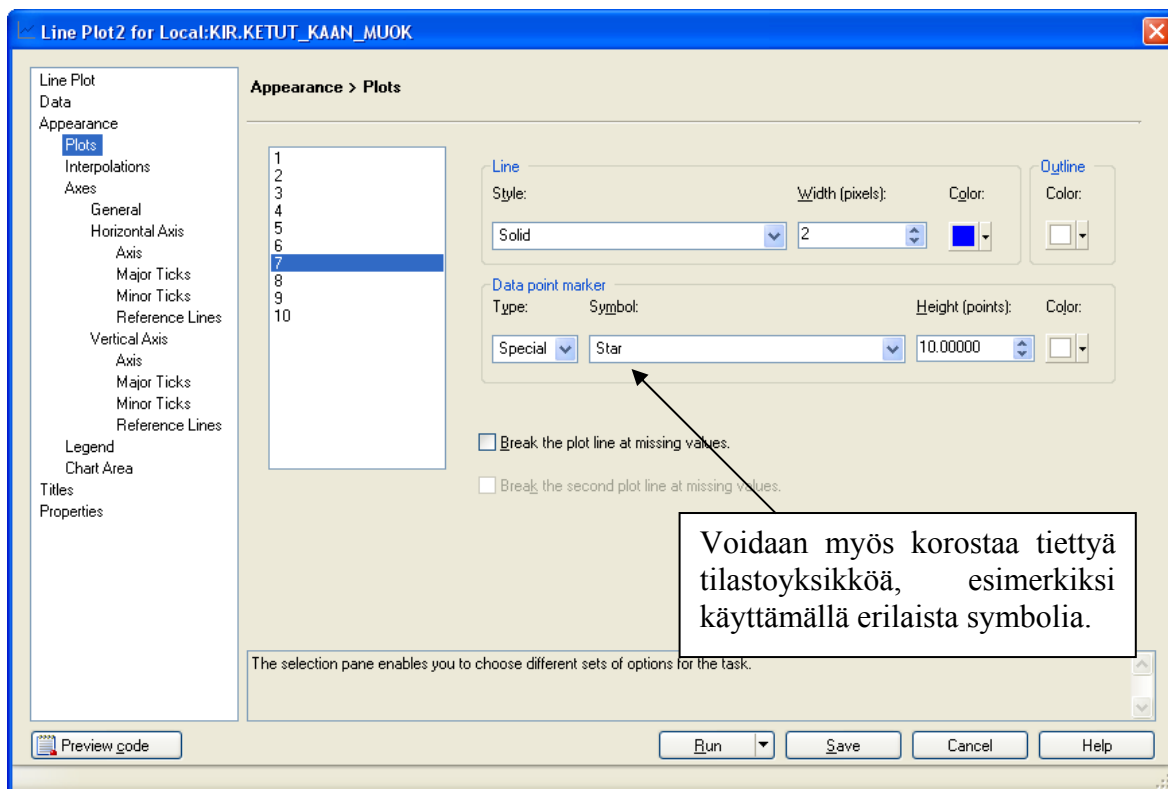
Viivakuva voidaan tuottaa valitsemalla käytettävään dataan liittyvästä asiayhteysvalikosta ”Graph” -> ”Line Plot”. Käytettävä kuvatyyppi valitaan ”Line Plot” -välilehdellä. Nämä kuvatyypit antavat pohjan kuvan muokkaukselle, mutta erilaisten asetus-valintojen avulla kahden eri pohjan avulla saatetaan päätyä samaan lopputulokseen. Esimerkiksi seuraavassa valittavalla kuvatyypillä voidaan tehdä myös ”Standard Deviation Plot” -tyyppinen kuva:



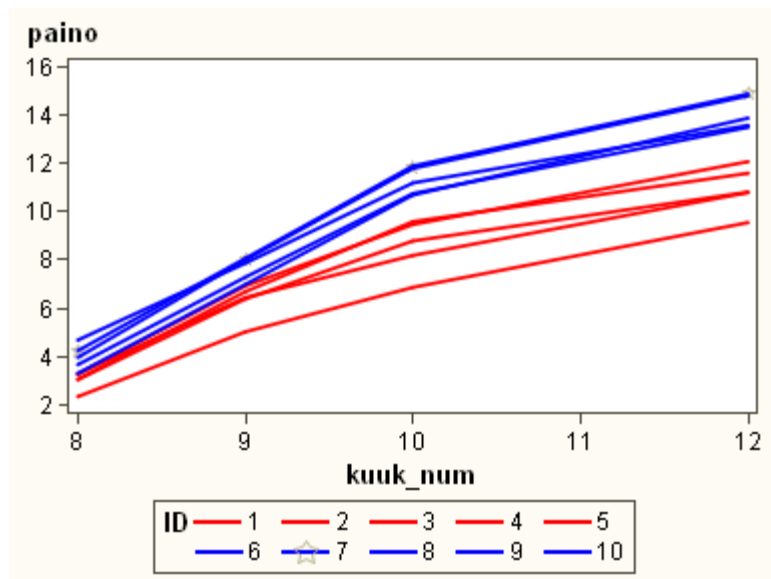
”Data” -välilehdellä määritellään muuttujien tehtävät:



”Plots”-välilehdellä voi määrittellä viivatyyppin, 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 (kopioitu leikepöydän kautta):



Vaihtoehtoinen ratkaisu SAS-koodilla:

Ennen viivakuvan piirtämistä piti muokata merkkimuotoisesta kuukausi-muuttujasta uusi numeerinen muuttuja:

```
DATA kir.ketutdata_muok2;
  SET kir.ketut_kaan;
  IF kuukausi='elokuu' THEN kuuk_num=8;
  IF kuukausi='syyskuu' THEN kuuk_num=9;
  IF kuukausi='lokakuu' THEN kuuk_num=10;
  IF kuukausi='joulukuu' THEN kuuk_num=12;
RUN;
```

Aineisto pitää järjestää x-akselin muuttujan mukaan:

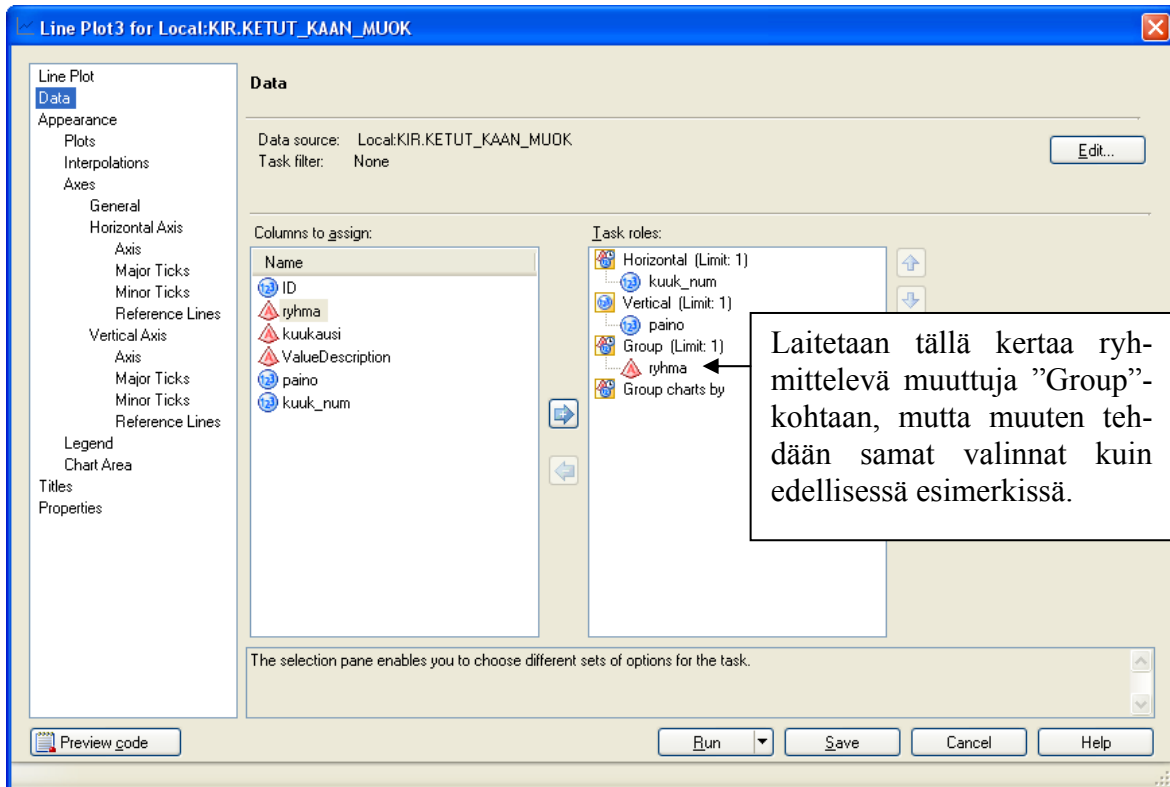
```
PROC SORT DATA=kir.ketutdata_muok2;
  BY kuuk_num;
RUN;
```

Kuva voidaan piirtää GPLOT-proseduurilla, jossa kullekin viivalle annetaan omat määrittämisensä SYMBOL-lauseissa:

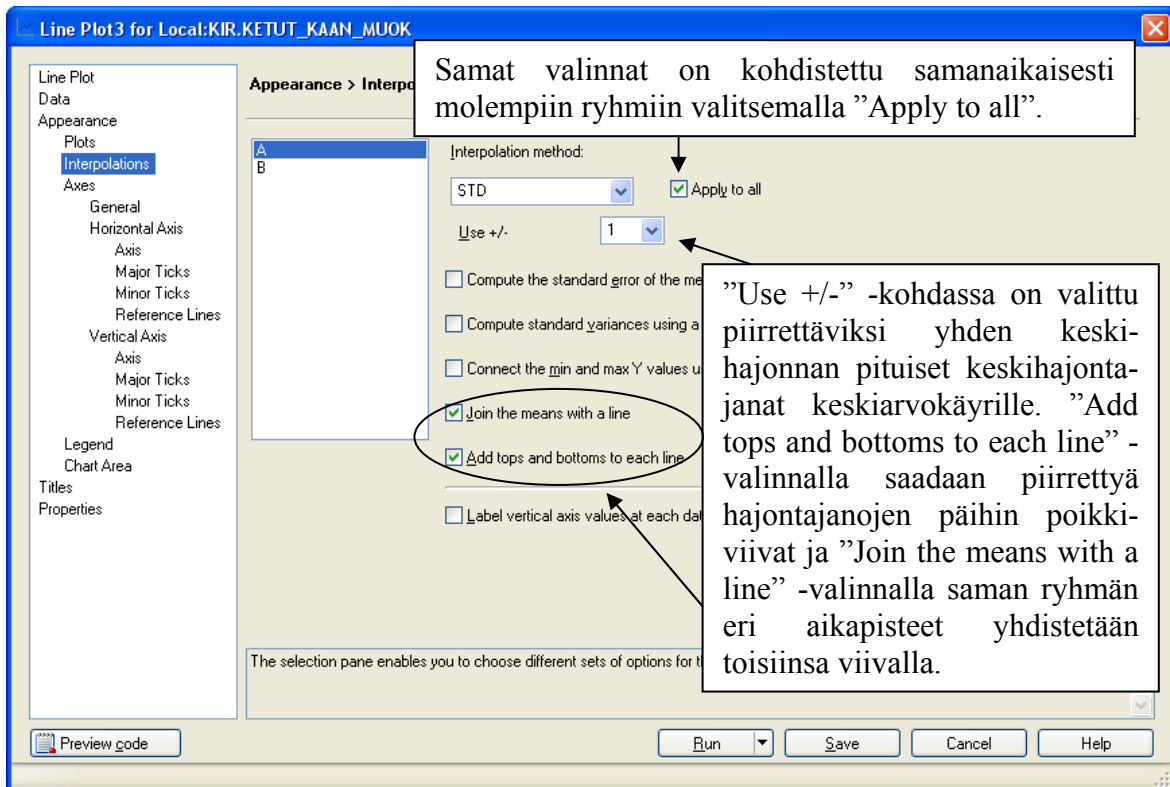
```
GOPTIONS RESET=all;
PROC GPLOT DATA=kir.ketutdata_muok2;
  PLOT paino1*kuuk_num=id;
  SYMBOL1 INTERPOL=join COLOR=red VALUE=dot;
  SYMBOL2 INTERPOL=join COLOR=red VALUE=dot;
  SYMBOL3 INTERPOL=join COLOR=red VALUE=dot;
  SYMBOL4 INTERPOL=join COLOR=red VALUE=dot;
  SYMBOL5 INTERPOL=join COLOR=red VALUE=dot;
  SYMBOL6 INTERPOL=join COLOR=blue VALUE=dot;
  SYMBOL7 INTERPOL=join COLOR=blue VALUE=star;
  SYMBOL8 INTERPOL=join COLOR=blue VALUE=dot;
  SYMBOL9 INTERPOL=join COLOR=blue VALUE=dot;
  SYMBOL10 INTERPOL=join COLOR=blue VALUE=dot;
RUN;
```

Viivakuva – keskiarvokäyrät

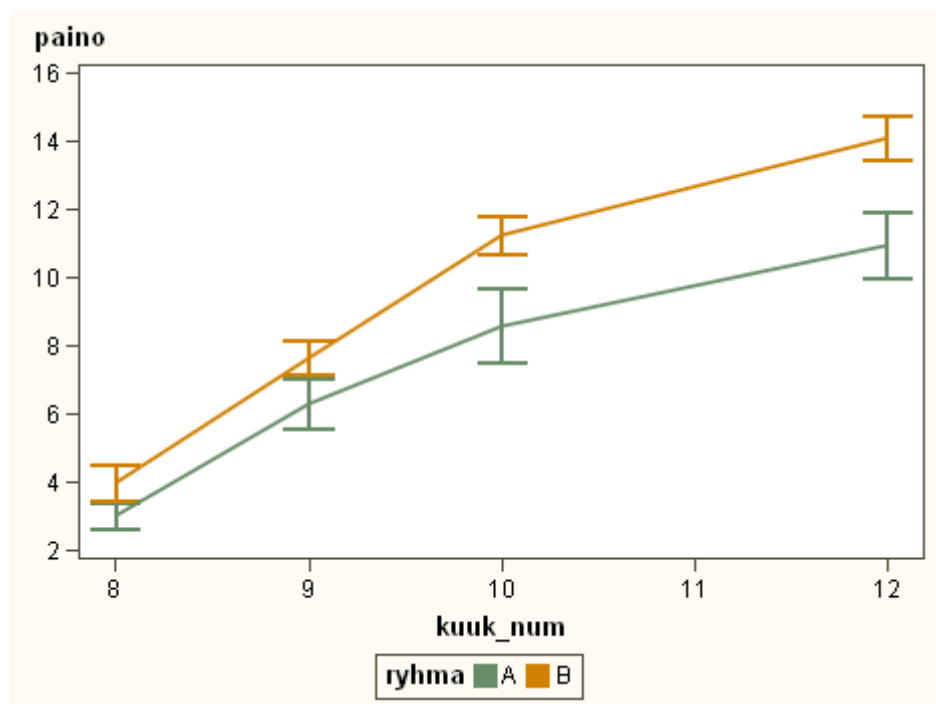
Edellisen esimerkkitalanteen ryhmien keskiarvoja eri aikapisteissä havainnollistava kuva saadaan aikaiseksi tekemällä hieman eri valinnat ”Data”-välilehdellä:



Nyt kun piirretään yksittäisiä havaintoja esittävän viivakuvan sijaan ryhmäkeskiarvoja havainnollistava kuva (eli ”Standard Deviation Plot” -tyyppinen kuva), valitaan ”Interpolations”-sivulla kullekin ryhmälle ”Interpolation method” -kohtaan vaihtoehto ”STD”:



Muilta osin tässä tilanteessa voidaan soveltaa samankaltaisia valintoja kuin edellisessä esimerkkitilanteessa ja painamalla ”Run” saadaan tulokseksi seuraava kuva:



Vaihtoehtoinen ratkaisu SAS-koodilla:

Myös keskiarvokuva voidaan piirtää käyttäen GPLOT-proseduuria. Olennaiset määrittelyt tehdään INTERPOL-optiossa (Katso lisätietoa ”Help” -> ”SAS Syntax Help”):

```
GOPTIONS RESET=all;
PROC GPLOT DATA=kir.ketutdata_muok2;
  PLOT paino1*kuuk_num=ryhma;
  SYMBOL1 INTERPOL=std1jt COLOR=red VALUE=none;
  SYMBOL2 INTERPOL=std1jt COLOR=blue VALUE=none;
RUN;
```

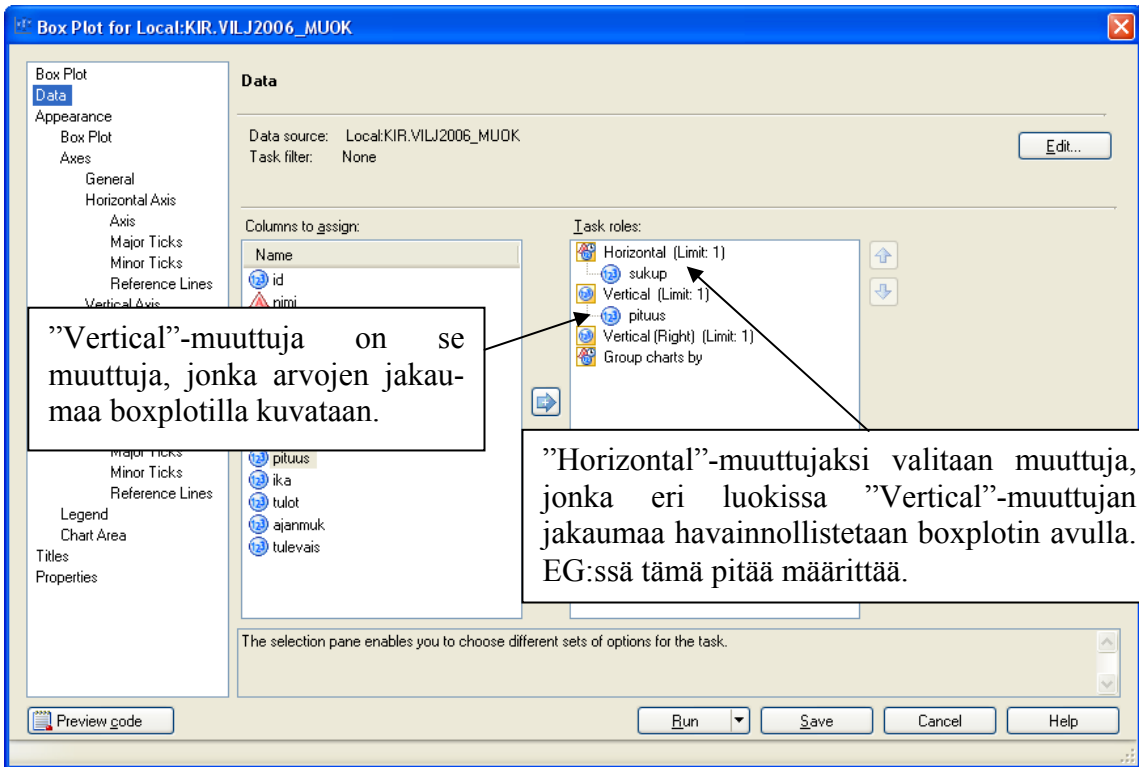
Tehtävä: Ota projektista käyttöön käännetty kettudata **ketut_kaan**. Tee ensin Query Builderissa uusi muuttuja, joka sisältää kuukausitiedon numeerisena. Tallenna data nimellä **ketut_num**. Piirrä kuva, jossa x-akselilla on kuukausi ja y-akselilla ketun paino. Yhdistä kunkin ketun toistomittaukset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

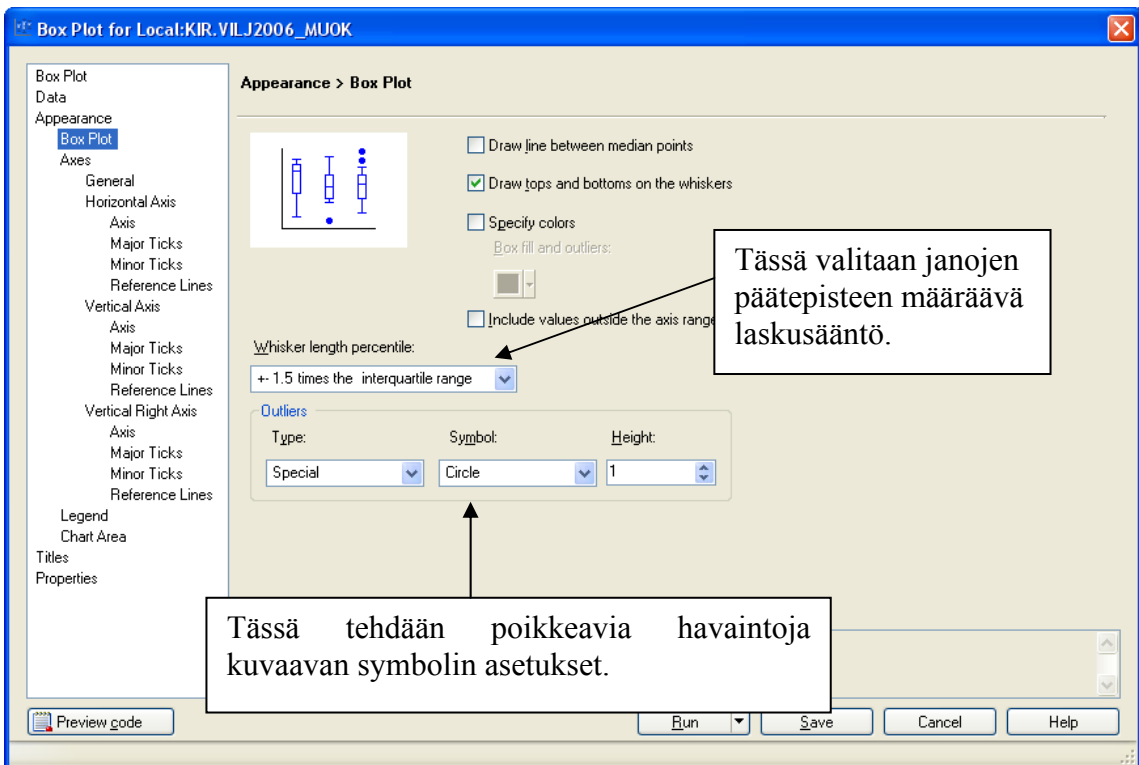
7.4 Laatikko-janakuvio (Box Plot)

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

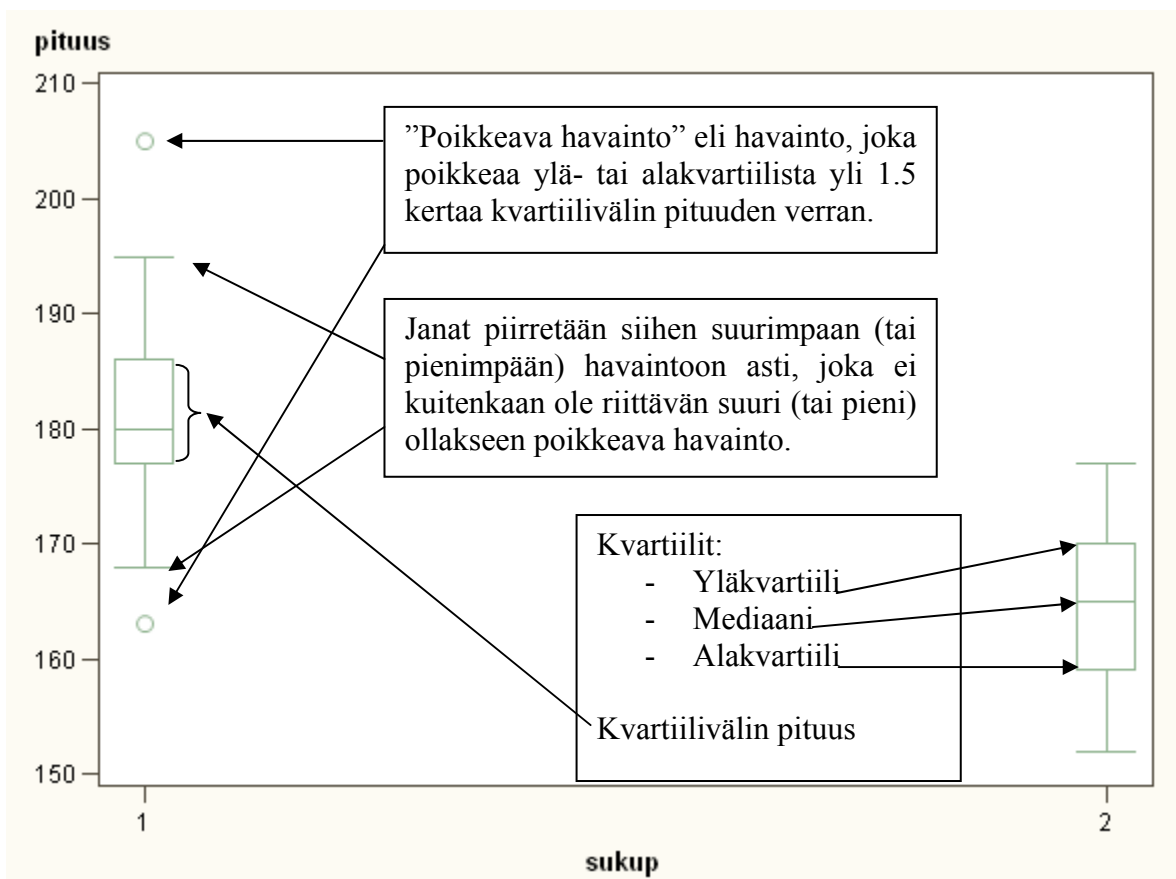
Enterprise Guidessa boxplot piirretään klikkaamalla käytettävään aineistoon liittyvästä asiayhteysvalikosta ”Graph” -> ”Boxplot” ja valitsemalla ”Box Plot” -välilehdeltä ”Box Plot” ja sitten ”Data”-välilehdellä:



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



Muilla välilehdillä tehtävät valinnat koskevat muun muassa kuvan akseleiden, referenssiiviivojen, otsikoiden ja koon asetuksia. Tuloste näyttää valituilla asetuksilla seuraavalta:



Yllä olevassa boxplot-kuvassa on erikoista se, että ohjelma piirtää eri sukupuolten kuviot kuvan ääreläidoille. Tämä selittyy sillä, että Boxplot-työkalu on pääasiassa tarkoitettu sellaisiin tilanteisiin, joissa ryhmiä on selvästi enemmän kuin kaksi, jolloin yksittäiset kuvat jakautuvat tasaisemmin kokonaiskuvaan. Yksi keino kiertää tämä ongelma on määrittää x-akseli esimerkiksi välille 0–3, jolloin arvot 1 ja 2 tulevat keskeemmälle kuvaa. Tämän voi tehdä Boxplot-työkalun ”Horizontal Axis” kohdan alla olevan ”Major Ticks” -välilehden ”Specify”-valinnalla.

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 tulkitsemasi boxplot-kuva on piirretty.

Vaihtoehtoinen ratkaisu SAS-koodilla:

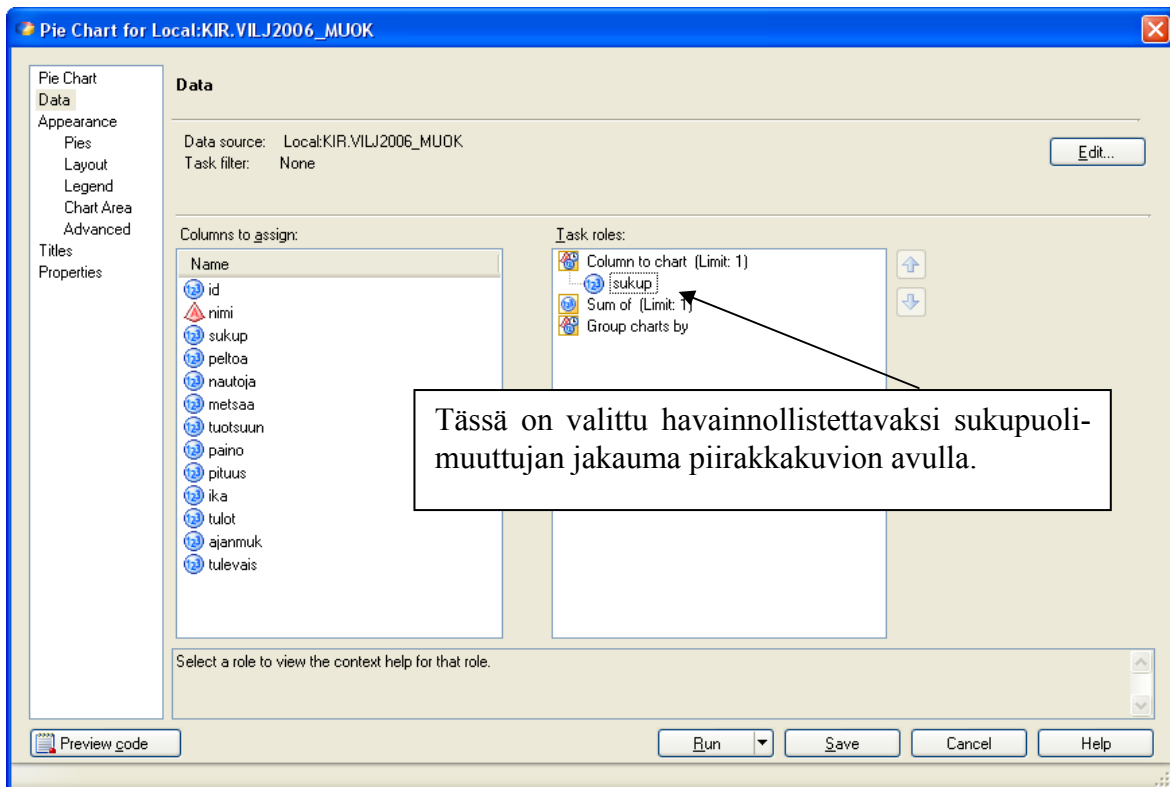
Laatikko-janakuvio voidaan piirtää käyttäen BOXPLOT-proseduuria:

```
GOPTIONS RESET=all;
PROC BOXPLOT DATA=kir.vilj2006_muok;
  PLOT pituus*sukupuollet / BOXSTYLE=schematic IDSYMBOL=circle;
RUN;
```

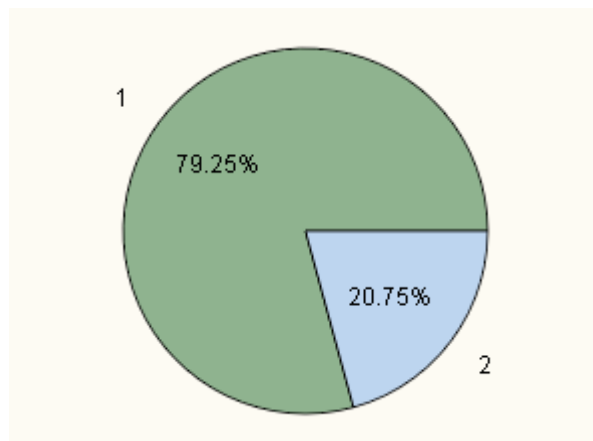
Tehtävä: Piirrä päätuotantosunnittain tulojen jakaumia havainnollistavat boxplot-kuvat Viljelijä 2006 -aineistosta (**vilj2006_muok**). Voit myös tarkastella muiden numeeristen muuttujien jakaumia.

7.5 Piirakkakuviio (Pie Chart)

Samoin kuin pylväskuvion avulla, myös piirakkakuviio avulla voidaan havainnollistaa kategoriisen muuttujan jakaamaa. Niinpä piirakkakuviioo laatiminen toimii hyvin pitkälle samoilla periaatteilla kuin pylväskuvion tekeminen. Ensin valitaan tarkasteltavaan aineistoon liittyvästä asiayhteysvalikosta ”Graph” -> ”Pie Chart”, valitaan ”Pie Chart”-välilehdellä kuvan tarkempi tyyppi ja sitten valitaan ”Data” -välilehdellä muuttujien roolit:



”Advanced”-välilehdellä voidaan valita esitetäänkö kuvassa frekvenssejä vai prosentiosuuksia. Valmis kuva näyttää seuraavalta:



Vaihtoehtoinen ratkaisu SAS-koodilla:

Piirakkakuviio voidaan piirtää esimerkiksi käyttäen GCHART-proseduuria:

```
GOPTIONS RESET=all;  
PROC GCHART DATA=kir.vilj2006_muok;  
  PIE sukupuoli / TYPE=pct;  
RUN;
```

7.6 Tile Chart -kuva

Tile Chart -työkalu on Enterprise Guide 4.2:n mukana tullut uusi työkalu, jolla voidaan piirtää niin sanottuja Tile Chart -kuvia. Tile Chart -kuvassa voidaan esittää samanaikaisesti kahden numeerisen muuttujan arvoja osaryhmittäin tai havaintoyksiköittäin siten, että kullekin ryhmälle tai havaintoyksikölle muodostuu kuvaan oma laatikko eli tiili. Jos kuva piirretään havaintoyksiköittäin, tiilien koolla kuvataan yhden muuttujan arvoa ja värillä toisen muuttujan arvoa. Jos kuva piirretään osaryhmittäin, tiilien koolla kuvataan osaryhmään kuuluvien havaintojen yhden muuttujan arvojen summaa ja vastaavasti värillä osaryhmään kuuluvien havaintojen toisen muuttujan arvojen summaa.

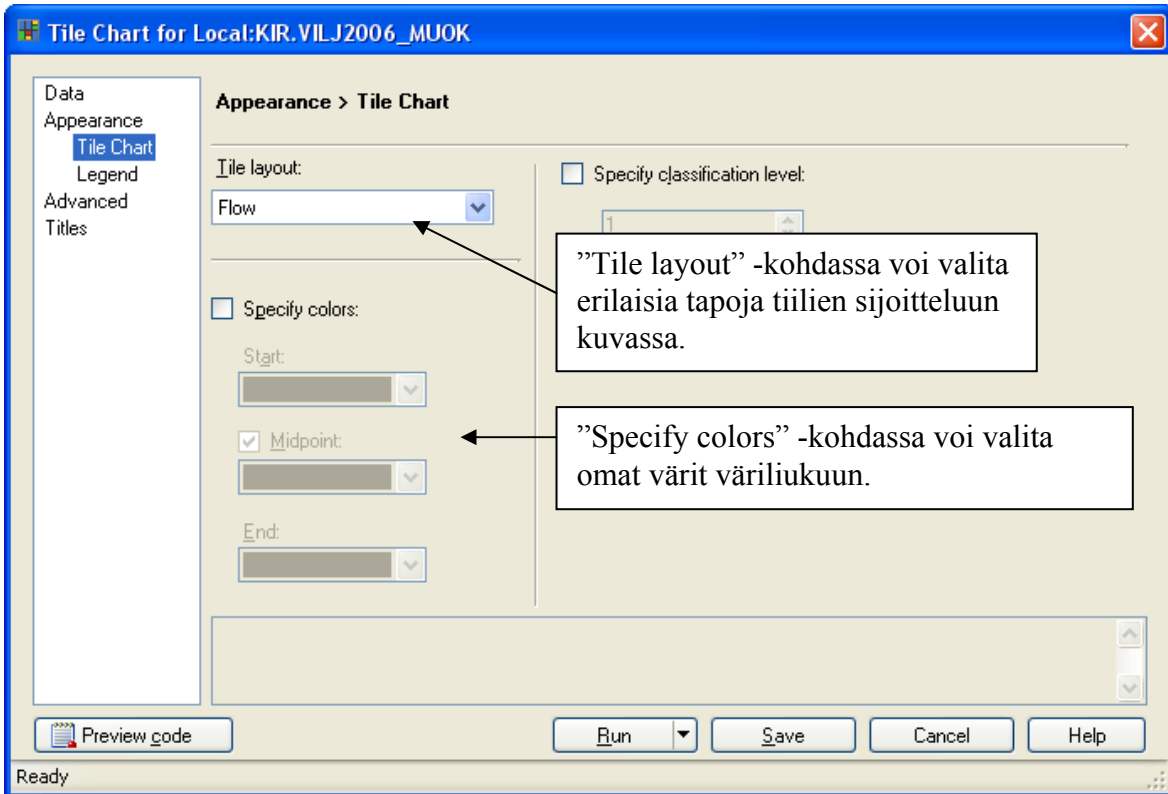
Kuvaa päästään laatimaan siten, että valitaan käytettävään aineistoon liittyvästä asiayhteysvalikosta ”Graph” -> ”Tile Chart”, jolloin päädytään ”Data”-välilehdelle:

”Classification variables” -kohtaan sijoitetaan muuttuja tai muuttujat, jotka määrittelevät osaryhmät tai yksilöt, joille kullekin muodostetaan oma tiili tuloksena syntyvään kuvaan. Jos tähän sijoitetaan useampi ryhmittelevä muuttuja, tiilet jaotellaan tuloksena syntyvään kuvaan hierarkkisesti ja hierarkiajärjestyksen määrää muuttujien järjestys tässä kohdassa.

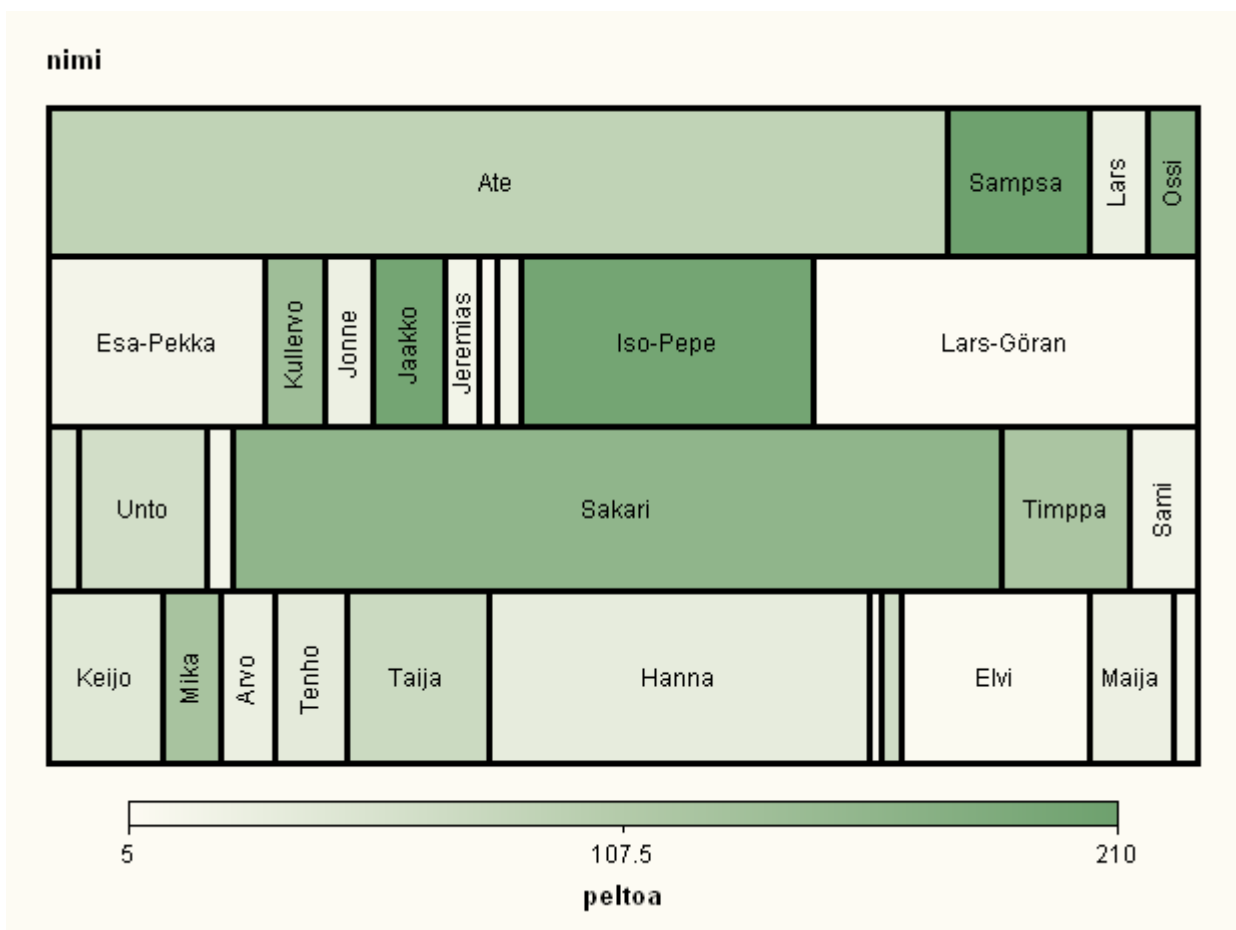
”Size analysis variable” on pakollinen ja siihen sijoitettavan muuttujan arvot määrittävät piirrettävien laatikoiden koot. Kannattaa huomioida, että jos muuttujan arvoihin sisältyy nollia, puuttuvia tietoja tai negatiivisia arvoja, kyseisiä tietoja ei käytetä kuvan piirtoon.

”Color analysis variable” kohtaan sijoitettavan muuttujan arvot määrittävät piirrettävien laatikoiden värit. Värien selitykseksi kuvan alle tulostuu väriliuku. Jos tähän kohtaan ei sijoiteta muuttujaa, myös värit kuvaavat ”Size analysis variable” arvoja.

”Tile Chart” -välilehdellä voi vaikuttaa kuvan ulkoasuun:



Lisäksi muun muassa ”Legend”-välilehdellä voidaan valita väriliu’un keskikohta ja ”Titles”-välilehdellä otsikot ja alaviitteet kuvalle. Lopuksi painetaan ”Run” ja seuraava kuva tulostuu:



Kuvassa tiilen koko kuvaa viljelijän omistamien nautojen lukumäärää ja tiilen väri viljelijän hallussa olevien peltohehtaarien määrää. Kaikilla aineiston viljelijöillä ei ollut nautoja, joten siksi kaikki viljelijät eivät ole mukana tässä kuvassa. Enterprise Guidessa kuvasta saadaan selville myös tarkempi tieto kuhunkin tiileen liittyvistä muuttujien arvoista viemällä hiiri kiinnostuksen kohteena olevan tiilen päälle.

Vaihtoehtoinen ratkaisu SAS-koodilla:

Tile chart -kuvia voidaan piirtää käyttäen GTILE-proseduuria:

```
PROC GTILE DATA=kir.vilj2006_muok;  
  FLOW nautoja TILEBY=(nimi) / COLORVAR=peltoa;  
RUN;
```

Tile chart täytyy piirtää ActiveX tai Java-tyyppisenä kuvana ja ohjata soveltuvaan tulostuskohteeseen. Enterprise Guidessa tämä tehdään automaattisesti oletusasetuksilla, mutta perus-SASissa voidaan käyttää seuraavia DEVICE=activex -optiota ja ODS HTML -lausetta kuvan piirtämisen yhteydessä:

```
ODS HTML;  
GOPTIONS RESET=all DEVICE=activex;  
PROC GTILE DATA=kir.vilj2006_muok;  
  FLOW nautoja TILEBY=(nimi) / COLORVAR=peltoa;  
RUN;  
ODS HTML CLOSE;
```

Tehtävä: Piirrä Viljelijä 2006 -aineistosta (**vilj2006_muok**) tile chart -kuva, jossa kullekin viljelijälle muodostetaan oma tiili. Aseta käytettävät muuttujat siten, että tiilen koko kuvaa viljelijän pituutta ja tiilen väri kuvaa viljelijän painoa.

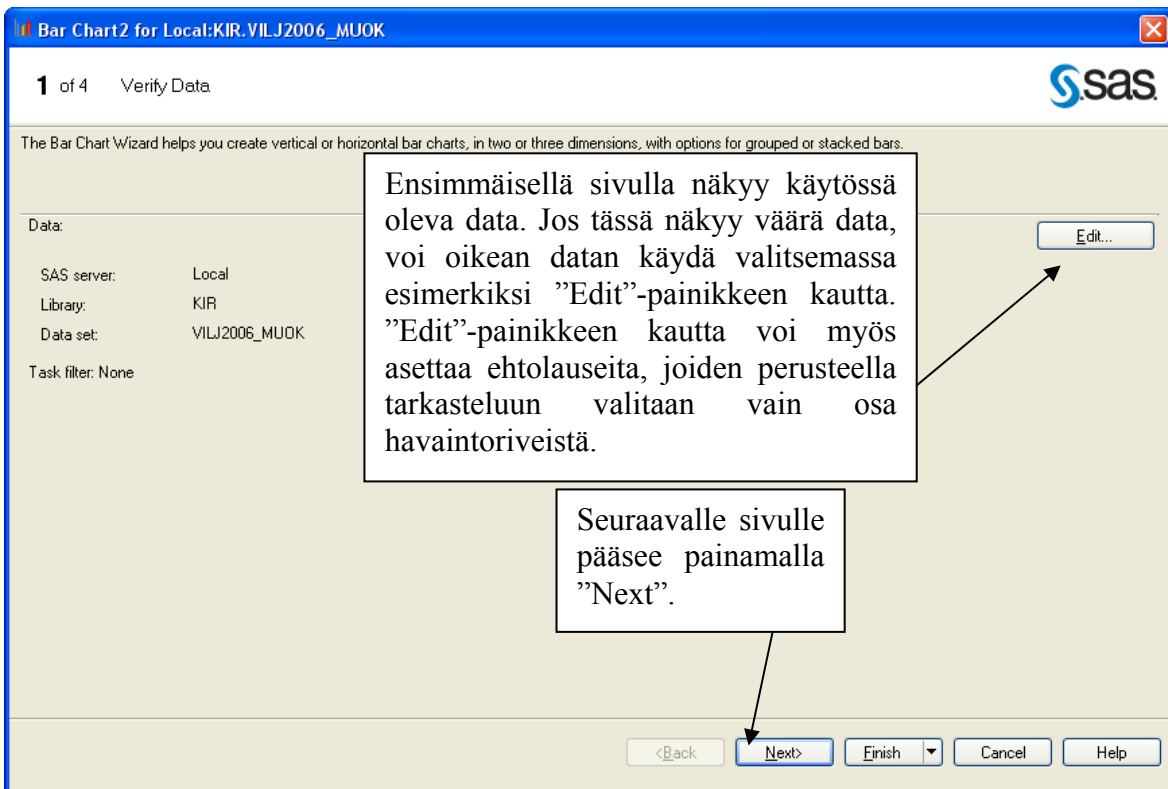
7.7 Wizard-kuvatyoökalut

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 kuvatyoökalut ovat usein käyttökelpoisempia, kun halutaan käyttää kuvatyoökalun ominaisuuksia mahdollisimman monipuolisesti. Wizard-kuvatyoökalut voivat puolestaan olla käyttökelpoisia esimerkiksi sellaisissa tilanteissa, joissa kuvia käytetään aineiston alustavaan tarkasteluun ja käyttäjä kokee wizard-kuvatyoökalujen logiikan tavallisten kuvatyoökalujen logiikkaa selkeämpänä. Lisäksi wizardista on mahdollista siirtyä tavalliseen vastaavaan työkaluun. Tämä onnistuu klikkaamalla määriteltyä wizard-tehtäväsolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Open” -> ”Open in Advanced View”. Tällöin wizardissa tehdyt määrittelyt näytetään tavallisessa työkalussa ja käyttäjä voi jatkaa määrittelyä siellä.

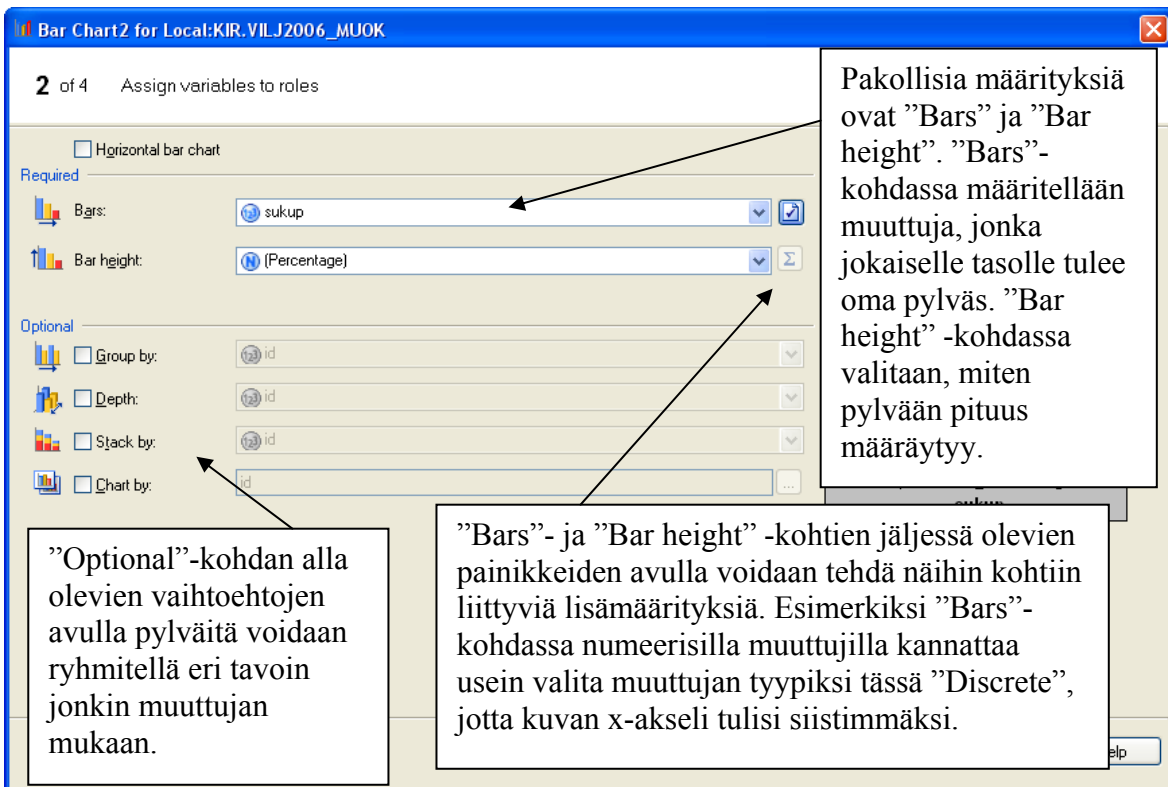
Toimintatapa wizardeissa on sellainen, että ensin avataan prosessidiagrammista tuplaklikkaamalla käytettävä data työtilaan ja sitten käynnistetään wizard-työkalu asiayhteysvalikon ”Graph”-valikosta. Wizardissa määrittelyt on jaettu useammalle sivulle, joiden läpi edetään järjestyksessä, ja lopuksi valitaan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

Wizard-työkalu pylväskuville

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



Toisella sivulla määritellään muuttujien roolit:



Kolmannella sivulla määritellään ulkoasuun liittyviä asioita:

The screenshot shows the 'Specify appearance' step of the SAS Bar Chart2 dialog. The window title is 'Bar Chart2 for Local:KIR.VILJ2006_MUOK'. The SAS logo is in the top right. The dialog is divided into several sections:

- 3D chart:** A checkbox labeled '3D chart' is checked. An arrow points to it from a box labeled 'Kolmiulotteisuus'.
- Colors:** A dropdown menu for 'Color bars by:' is set to '[All bars the same]'. A 'Bar color:' selector is set to a light blue color. An arrow points to this section from a box labeled 'Pylväiden värit'.
- Labels:** A 'Legend:' dropdown is set to 'South'. The 'Data labels:' checkbox is checked, and the dropdown is set to 'Percentage'. An arrow points to this section from a box labeled 'Selitykset pylväiden päihin (esimerkiksi prosentit)'. There is also an 'Axis Labels' button.
- Lines and ticks:** The 'Use reference lines' checkbox is unchecked. The 'Tick marks' checkbox is checked.

On the right, a 'Sample chart' is displayed. It is a bar chart titled '(Percentage)' with the x-axis labeled 'sukupu' and three bars representing categories 1, 2, and 3. The values are 33.33%, 16.67%, and 50.00% respectively.

At the bottom, there are navigation buttons: '<Back', 'Next>', 'Finish' (with a dropdown arrow), 'Cancel', and 'Help'.

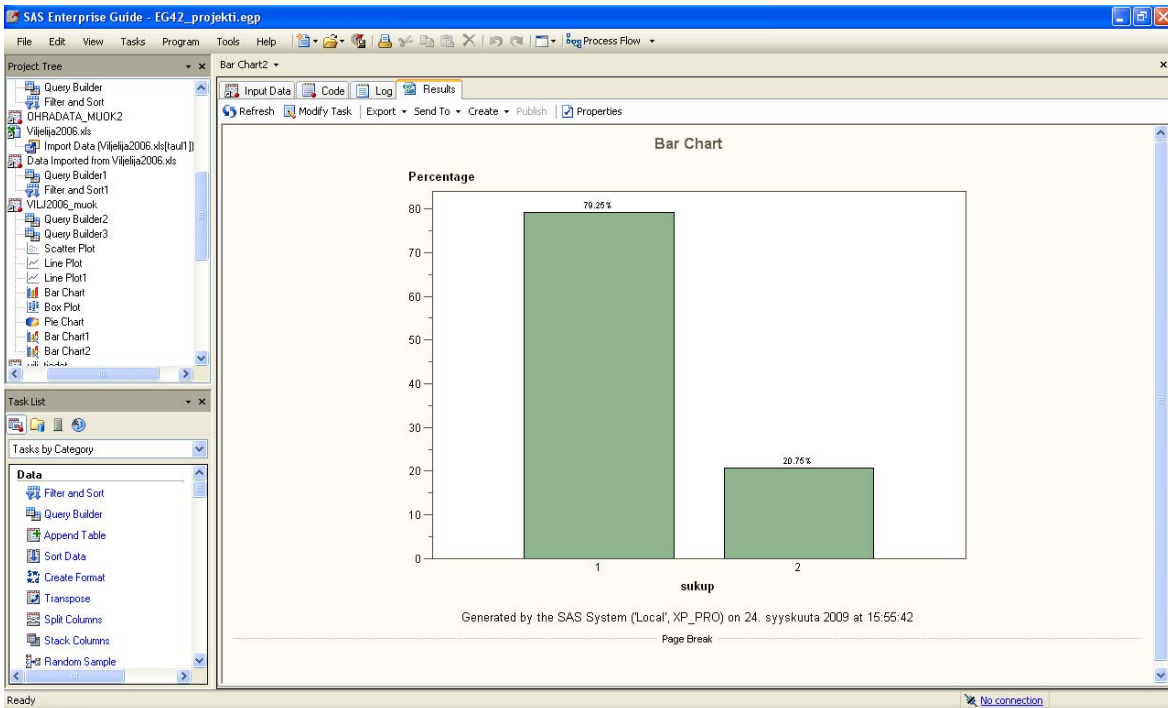
Viimeisellä sivulla voi vielä määritellä kuvien otsikoita ja alaviitteitä:

The screenshot shows the 'Provide a title and footnote' step of the SAS Bar Chart2 dialog. The window title is 'Bar Chart2 for Local:KIR.VILJ2006_MUOK'. The SAS logo is in the top right. The dialog has two main sections:

- Graph:** A text field contains 'Bar Chart'. A 'Reset' button is to the right.
- Footnote:** A text field contains the default SAS footer: 'Generated by the SAS System (&_SASSERVERNAME, &SYSSCPL) on %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) at %TRIM(%SYSFUNC(TIME(), NLTIMAP20.))'. A 'Reset' button is to the right.

At the bottom, there are navigation buttons: '<Back', 'Next>', 'Finish' (with a dropdown arrow), 'Cancel', and 'Help'.

Lopuksi painetaan ”Finish”, jolloin valmis kuva tulee samalla tavalla osaksi projektia kuin tavallisilla kuvatyökaluilla:



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 avaamalle ensin datamatriisi työtilaan ja klikkaamalla asiayhteysvalikosta ”Graph” -> ”Line Plot Wizard”. Tällöin aukeaa ensimmäinen sivu:

1 of 5 Verify Data

The Line Plot Wizard helps you create two-dimensional line plots, with options for markers and line interpolations.

Data:

SAS server:	Local
Library:	KIR
Data set:	KETUT_KAAN_MUOK
Task filter:	None

Samoin kuin pylväskuvalla, tästä voi tarkistaa onko oikea data käytössä. Jos data on oikea, painetaan ”Next”.

Jos näkyvässä on väärä data, voi oikean datan käydä valitsemassa esimerkiksi ”Edit”-painikkeen kautta. ”Edit”-painikkeen kautta voi myös asettaa ehtolauseita, joiden perusteella tarkasteluun valitaan vain osa havaintoriveistä.

<Back Next> Finish Cancel Help

Seuraavalla sivulla määritellään muuttujien roolit:

The screenshot shows the 'Line Plot4 for Local:KIR.KETUT_KAAN_MUOK' 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. The 'Function' is set to 'Sum'. A text box on the right explains that 'Horizontal' refers to the x-axis and 'Vertical' to the y-axis, and that these are mandatory. It also notes that in the 'Chart by' section, one can specify that separate plots are drawn for each level of a certain variable, and in the 'Summarize for each distinct horizontal value' section, one can specify that the values of the y-variable are listed from a list of possible values, with the calculated values for each x-variable plotted separately. Navigation buttons at the bottom include '<Back', 'Next>', 'Finish', 'Cancel', and 'Help'.

Kolmannella sivulla määritetään viivakuvan asetuksia:

The screenshot shows the 'Line Plot4 for Local:KIR.KETUT_KAAN_MUOK' 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 'None', 'Height (points)' set to 10.00000, and 'Color' set to 'Default color'. A text box on the right explains that in the 'Interpolation' section, the plot type is specified. In addition to the 'Line' plot, other options include 'Scatter' (scatter plot), 'Regression' (regression model), and 'STD' (mean deviation plot). Additional settings for each plot type can be accessed via the 'Detail...' button. A text box at the bottom left notes that the appearance of lines and symbols can also be affected. Navigation buttons at the bottom include '<Back', 'Next>', 'Finish', 'Cancel', and 'Help'.

Neljännellä sivulla on vielä muutama kuvan ulkoasua koskeva lisämääritys:

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", jolloin edellä tehdyillä valinnoilla työtilaan tulostuu kullekin ketulle oma painokäyrä:

SAS Enterprise Guide - EG42_projekti.egg

Line Plot ID=1

paino

kuuk_num	paino
8	2.5
9	5
10	7.5
11	10
12	12.5

kuuk_num

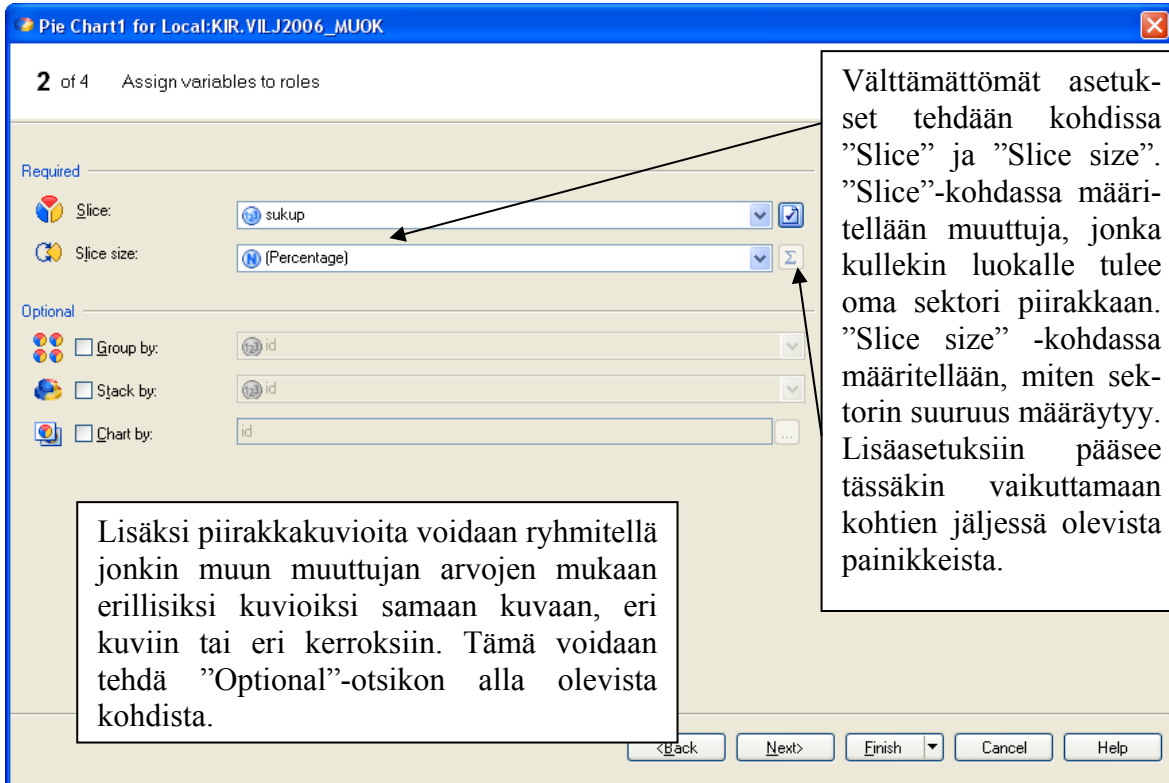
Generated by the SAS System (Local, XP_PRO) on 09. lokakuuta 2009 at 09:23:15

Page Break

Line Plot

Wizard-työkalu piirakkakuviolle

Piirakkakuvioiden wizard-työkalussa lähdetään liikkeelle samalla tavalla kuin muissa wizard-työkaluissa eli avataan data työtilaan ja klikataan asiayhteysvalikosta ”Graph” -> ”Pie Chart Wizard”. Ensimmäinen ikkuna on sama kuin muissa wizard-työkaluissa, eli siitä voi tarkistaa, onko käytössä oikea data ja tarvittaessa vaihtaa tai karsia dataa ”Edit”-painikkeen kautta. Toisella sivulla määritetään muuttujien roolit:



2 of 4 Assign variables to roles

Required

Slice: sukup

Slice size: (Percentage)

Optional

Group by: id

Stack by: id

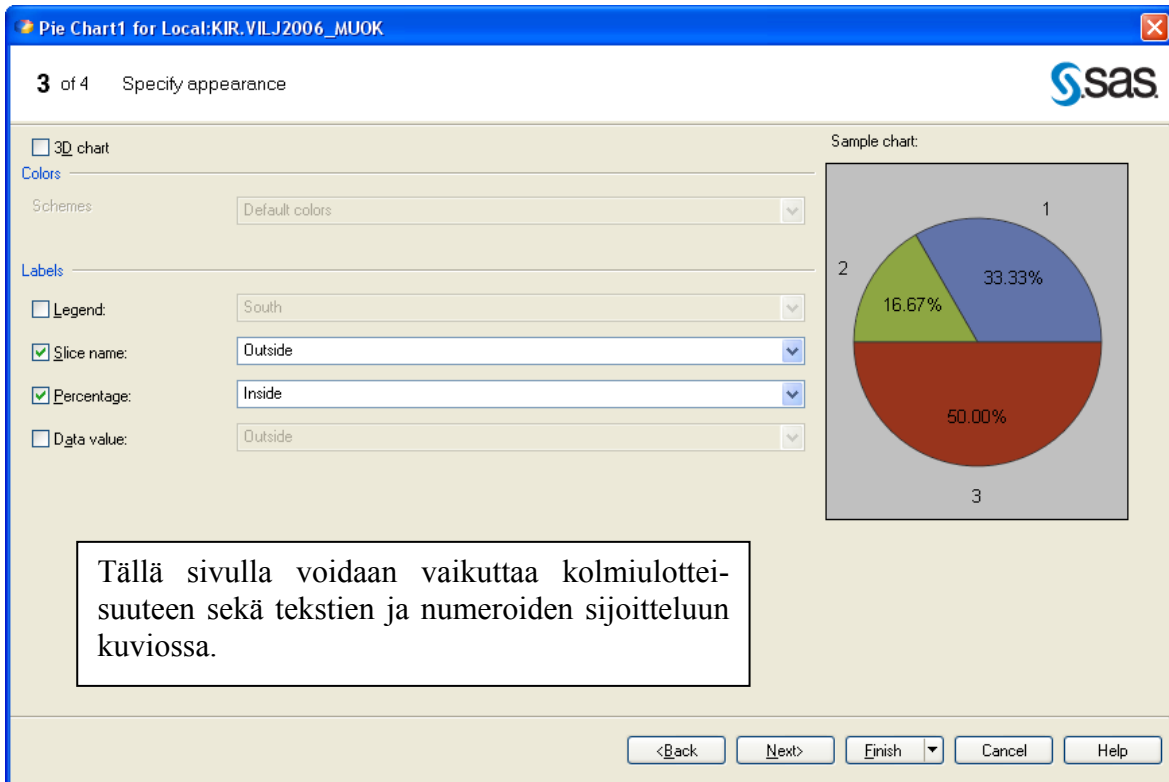
Chart by: id

Lisäksi piirakkakuviota 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:



3 of 4 Specify appearance

3D chart

Colors

Schemes: Default colors

Labels

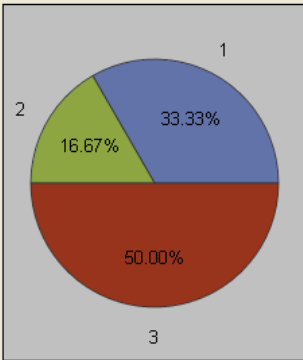
Legend: South

Slice name: Outside

Percentage: Inside

Data value: Outside

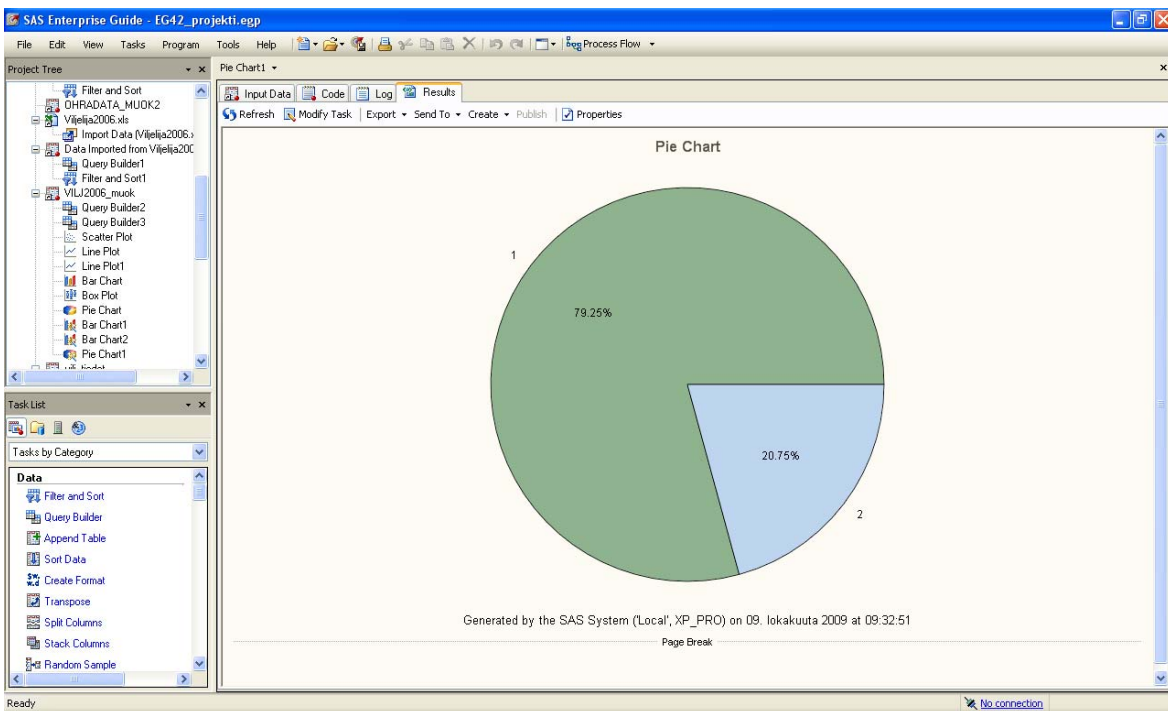
Sample chart:



Tällä sivulla voidaan vaikuttaa kolmiulotteisuuden sekä tekstien ja numeroiden sijoitteluun kuviossa.

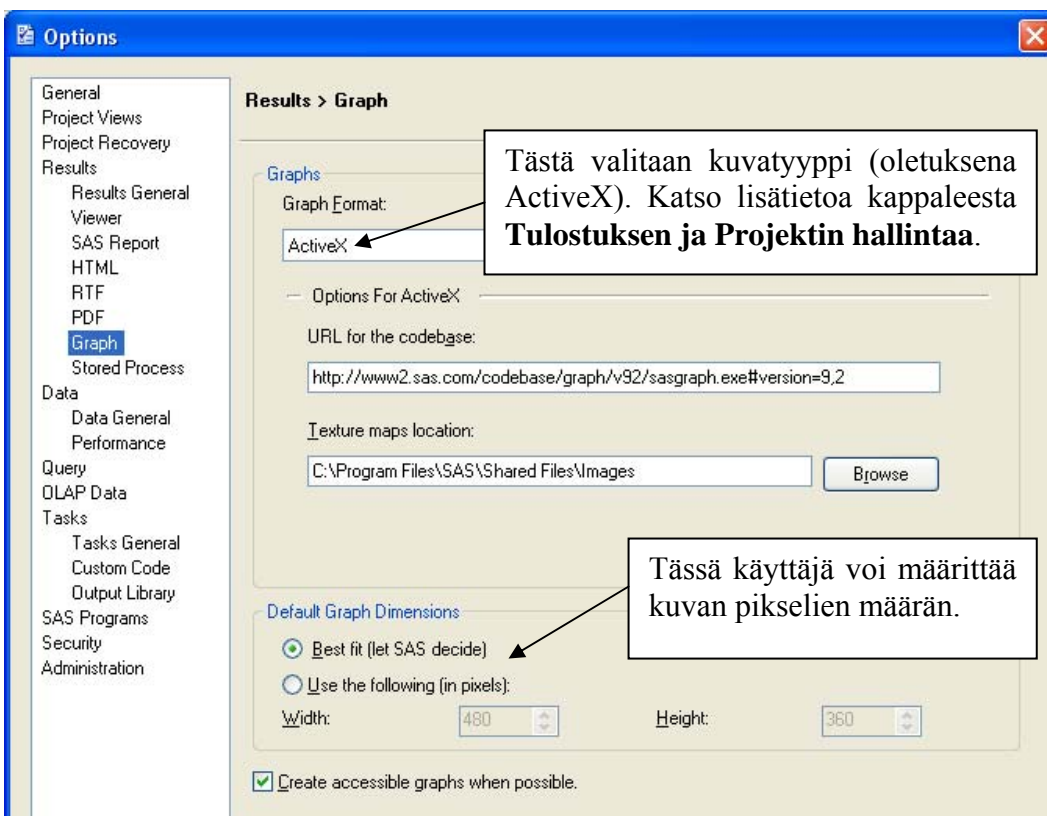
<Back Next> Finish Cancel Help

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



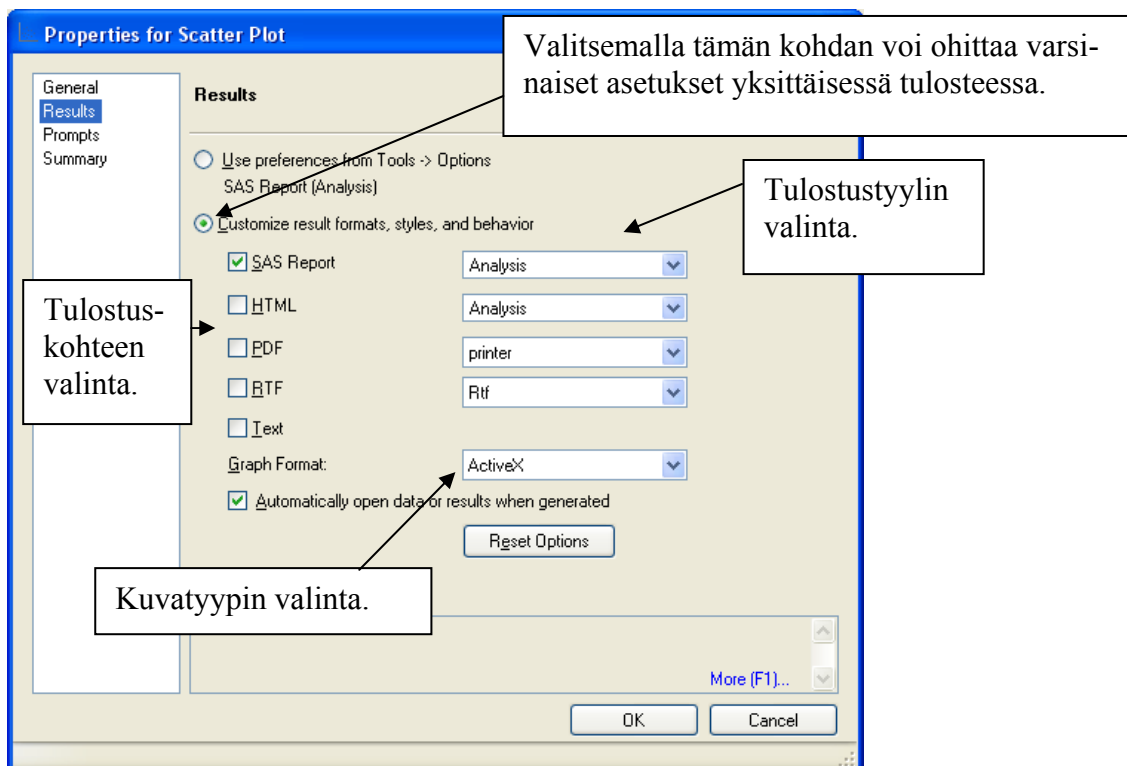
7.8 Kuvien tulostuskohte ja yleisasetukset

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



Yksittäiseen kuvaan liittyvät asetukset muokataan ensisijaisesti kuvatyökalussa kuvan määrittelyn yhteydessä. Tiettyjä tulostuskohteita käytettäessä kuvia voidaan kuitenkin muokata myös jälkikäteen, jolloin päästään osittain vaikuttamaan sellaisiin asioihin, joita ei kuvaa määrittäessä saa muokattua. Tällainen muokkaus onnistuu suoraan tulosteesta käytettäessä HTML-tulostuskohdetta tai SAS Report -oletustulostuskohteeseen tehdyillä kuvilla, jos tulosteesta on ensin tehty niin sanottu SAS-raportti (esimerkiksi valitsemalla tulosteeseen liittyvästä asiayhteysvalikosta ”Create” -> ”Create Report”). Jos kuva on tulostettu HTML-kohteeseen, sitä voidaan muokata klikkaamalla kuvan päällä hiiren oikealla painikkeella ja valitsemalla ”Graph Properties”. SAS-raporttiin vietyä kuvaa pääsee muokkaamaan klikkaamalla raportissa kuvan päällä hiiren oikealla painikkeella ja valitsemalla ”Properties”. ”Graph Properties”/”Properties”-valinnalla aukeaa valikko, josta päästään vaihtamaan kuvan asetuksia. HTML-tulostuskohteeseen liittyen 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 on mahdollista tavallisilla ActiveX-kuvilla EG:ssä ja Internet Explorerissa sekä Java-kuvilla Javaa tukevilla selaimissa. Kuvien tulostuslaatua voi parantaa muun muassa suurentamalla kuvien pikselimäärää (”Tools” -> ”Options” -> ”Graph”) tai kuvatyökalun sisällä ”Chart Area”-välilehdellä.

Kuvan yleisasetusten määrittäminen ennen kuvan tuottamista ei ole ainoa tapa määrittää kuville asetuksia. Enterprise Guide 4.2:ssa yksittäisen tulosteen kohde, tyyli ja/tai kuvatyypin voidaan määrittellä yleisasetuksista poikkeavaksi tulostuksen määrittävän työkalun sisältä. Useista työkaluista löytyy vasemmassa reunassa olevasta listasta ”Properties”-välilehti, josta voidaan valita ”Edit” -> ”Results”, jolloin päästään seuraavalle yksittäisen tulosteen asetusten määrittelysivulle:



Yleisasetuksista poikkeavat määrittelyt kohdistuvat kyseisen yksittäisen tehtäväsolmun tulostukseen, kun tehtävä suoritetaan normaalisti (valitaan ”Run”). Samaisia asetuksia voidaan myös muokata aiemmin määrittelylle yksittäiselle tehtäväsolmulle seuraavasti: klikataan tehtäväsolmua prosessidiagrammissa hiiren oikealla painikkeella, valitaan ”Properties” ja välilehti ”Results”. Tällöin aukeaa vastaava ikkuna kuin edellä. Tehtyjen valintojen jälkeen kyseinen tehtä-

väsolmu pitää vielä suorittaa uudelleen. Se onnistuu esimerkiksi klikkaamalla tehtäväsolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla aukeavasta valikosta ”Run”.

Yksittäisen tulostusikkunaan tulostetun kuvan voi myös tallentaa kuvatiedostoon (ainoastaan HTML-tulostuskohteesta) tai siirtää leikepöydän kautta (”copy-paste”) esimerkiksi Wordiin. Kuvan tallentaminen *.jpeg-tiedostoksi onnistuu klikkaamalla valmista HTML-tulostuskohteeseen tehtyä 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”. Leikepöydän kautta kopioiminen onnistuu SAS Report -, HTML- ja RTF-tulostuskohteisiin tehdyillä kuvilla.

Kuvien tallentaminen kuvatiedostoon SAS-koodilla:

Tallennettaessa kuvia tiedostoksi SAS-koodilla Enterprise Guidessa, kannattaa ensin tehdä valinta ”Tools” -> ”Options” -> ”Results General” ja valita siellä ainoaksi tulostuskohteeksi ”Text output”. Tällöin tulostus syntyy perinteisessä SAS-tulostusmuodossa ja edistyneemmät tulostusmuodot eivät pääse vaikuttamaan koodissa määriteltyihin kuva-asetuksiin.

Kuvien ajaminen kuvatiedostoksi onnistuu kätevästi esimerkiksi seuraavalla koodilla, joka tuottaa tuoksena eps-muotoisen kuvatiedoston:

```
FILENAME GFILE 'C:\kansio\polku\kuvatiedosto.eps';
GOPTIONS RESET=all KEYMAP=winansi FTEXT=duplex GSFNAME=gfile
GSFMODE=replace DEVICE=psepsf DISPLAY HTEXT=2;
```

* Tähän väliin kuvan piirtävä SAS-koodi *;

```
GOPTIONS RESET=all;
```

Option ”psepsf” tilalle voi laittaa myös muita ”deviceja”, jolloin kuvan saa tallennettua kuvatiedostoksi muussakin kuin eps-muodossa. Tyypillisiä vaihtoehtoja ovat muun muassa erilaiset pdf-vaihtoehdot. Vaihtoehtoja voi tarkastella kätevästi perus-SASin Explorer-ikkunasta (vasemman reunan ikkuna) menemällä kirjastoviitteen ”Sashelp” alle (tuplaklikkaa ensin ”Libraries” ja sitten ”Sashelp”). Siellä klikataan ”Devices”-kuvaketta hiiren oikealla painikkeella ja valitaan ”Explore from here”. Tällöin aukeaa ikkuna, jossa on listattuna kaikki käytettävissä olevat ”Devicet”. Huomaa, että vaihdattaessa koodissa DEVICE=-optiota, tulee vaihtaa myös FILENAME-rivillä tiedoston päätteeksi vastaavaksi (esim. PDF-deviceillä pdf-pääte).

Ohessa vielä esimerkki kuvan piirtämisestä tiedostoksi (sirontakuvi):

```
FILENAME GFILE 'C:\temp\pylvaskuva_painopituus.eps';
GOPTIONS RESET=all KEYMAP=winansi FTEXT=duplex GSFNAME=gfile
GSFMODE=replace DEVICE=psepsf DISPLAY HTEXT=2;
```

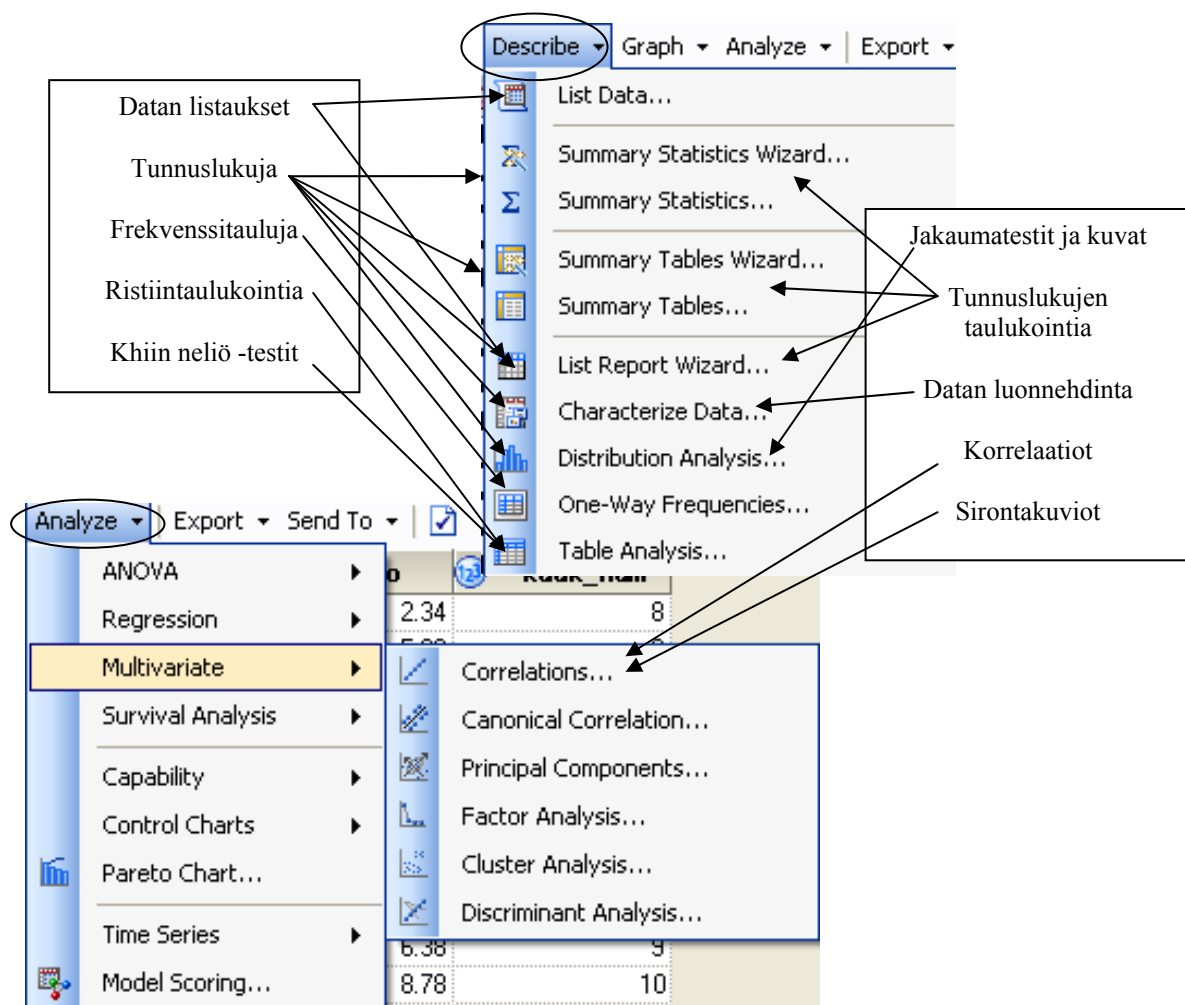
```
TITLE 'Ääkkösetkin tulostuvat oikein';
PROC Gplot DATA=kir.vilj2006_muok;
  PLOT paino*pituus=sukup;
  SYMBOL1 COLOR=blue VALUE=dot INTERPOL=rl;
  SYMBOL2 COLOR=red VALUE=circle INTERPOL=rl;
RUN;
```

```
GOPTIONS RESET=all;
```

Tarkempaa tietoa tässä esitetyistä koodin yksityiskohdista voi hakea esimerkiksi valitsemalla EG:ssä ”Help” -> ”SAS Syntax Help” ja kirjoittamalla mielenkiinnon kohteena olevan koodin yksityiskohdan ”Etsi”-välilehden hakukenttään.

8 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 kuvaillevien 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:



8.1 Datan listaukset ja raportit

Yksinkertaisin tapa tarkastella dataa on tulostaa koko data tai osia siitä. Tällaista listausta voidaan myös muokata lukijaystävällisemmäksi. Listaukseen voidaan valita vain osa muuttujista ja suorittaa erilaista datan järjestämistä, ryhmittelyä ja summausta. Tähän tarkoitukseen soveltuvia työkaluja löytyy Enterprise Guidesta kaksi: List Report Wizard ja List Data. List Report Wizard on näistä kahdesta monipuolisempia ja sitä voidaan käyttää myös muun muassa tunnuslukujen taulukointiin.

List Data -työkalu

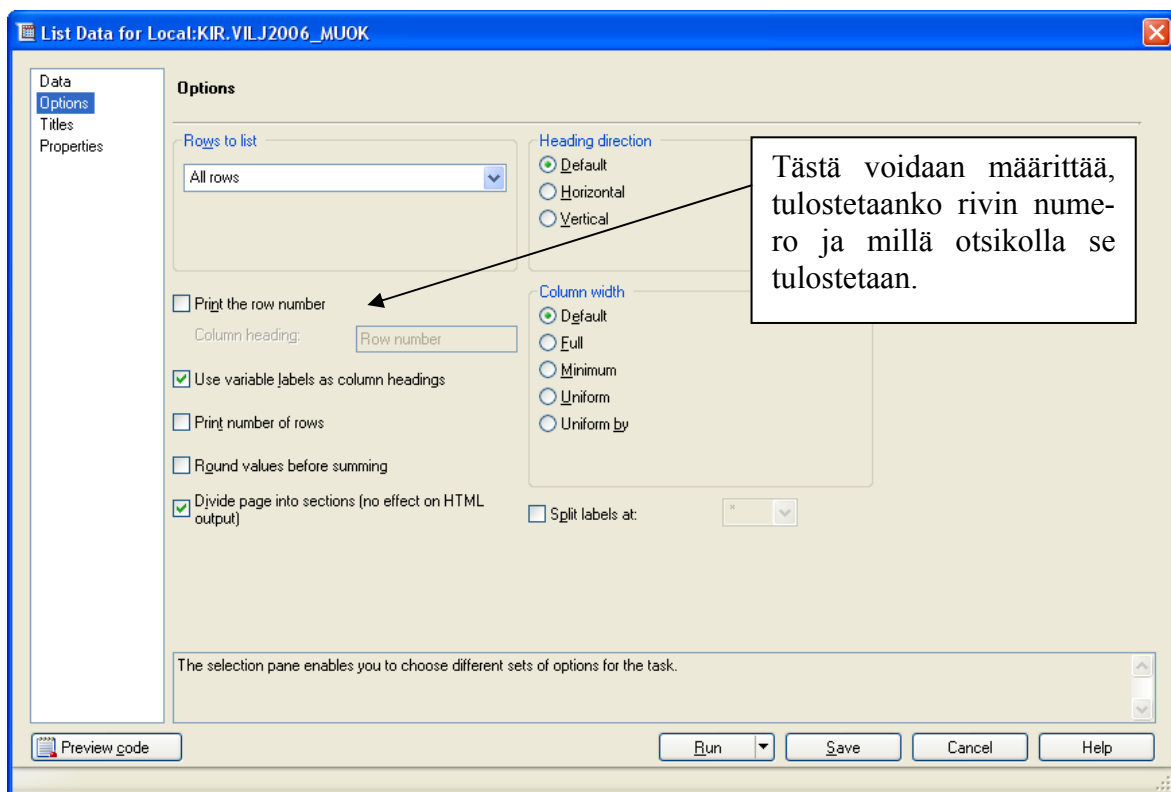
Listaukset voidaan tuottaa List Data -työkalulla, valitsemalla työtilassa tarkasteltavana olevaan aineistoon liittyvästä asiasyhteysvalikosta **”Describe”** -> **”List Data”**, jolloin aukeaa näkyviin **”Data”**-välilehti:

”List variables” -kohdassa luetellaan muuttujat, joiden arvot halutaan tulostaa.

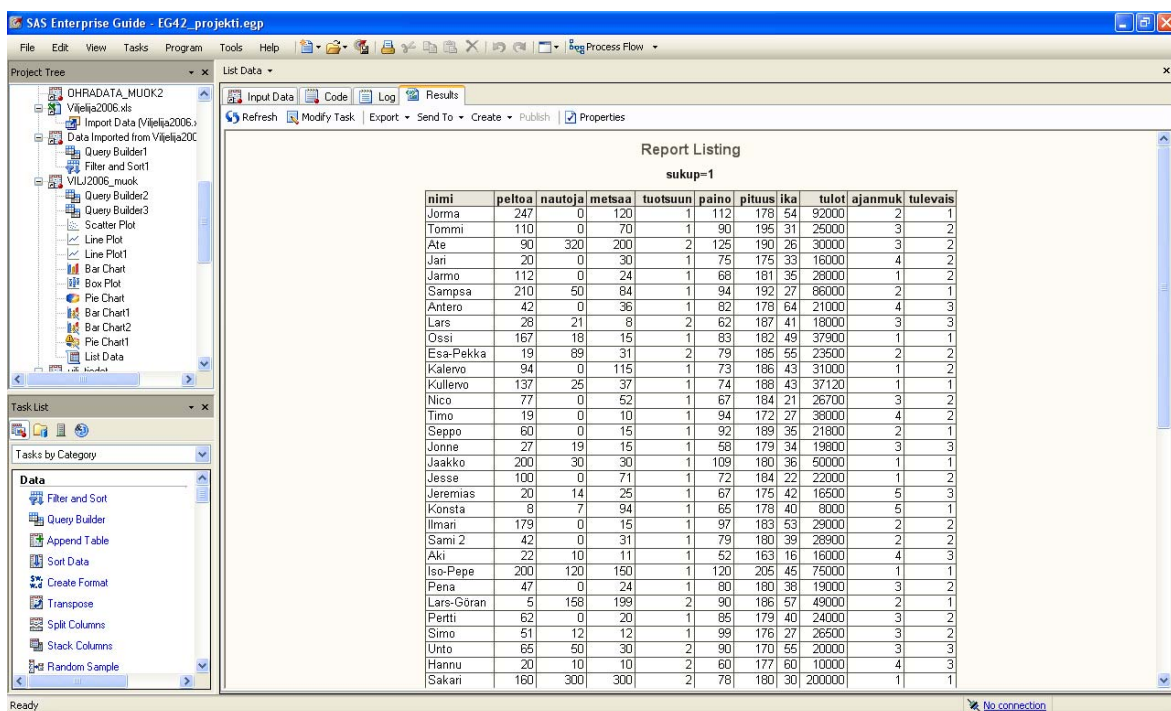
”Group analysis by” -kohtaan sijoitetaan sellainen muuttuja tai muuttujat, joiden mukaan tulostus halutaan järjestää osaryhmiin. Tällöin kukin osaryhmä tulostuu omaksi taulukkokonaksi.

Jos halutaan käyttää osaryhmiin jakoa, mutta tulostaa siitä huolimatta kaikki rivit yhteen taulukkokon, ryhmittelevä muuttuja tai muuttujat laitetaan ”Group analysis by” -kohdan lisäksi ”Identifying label” -kohtaan.

Klikkaamalla vasemmalla olevasta ikkunasta ”Options”, päästään määrittelemään tulostusta tarkemmin:



Painamalla ”Run” saadaan seuraava tuloste:



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

Vaihtoehtoinen ratkaisu SAS-koodilla:

Aineisto pitää ensin järjestää BY-muuttujan arvojen mukaan järjestykseen:

```
PROC SORT DATA=kir.vilj2006_muok;  
  BY sukupuoli;  
RUN;
```

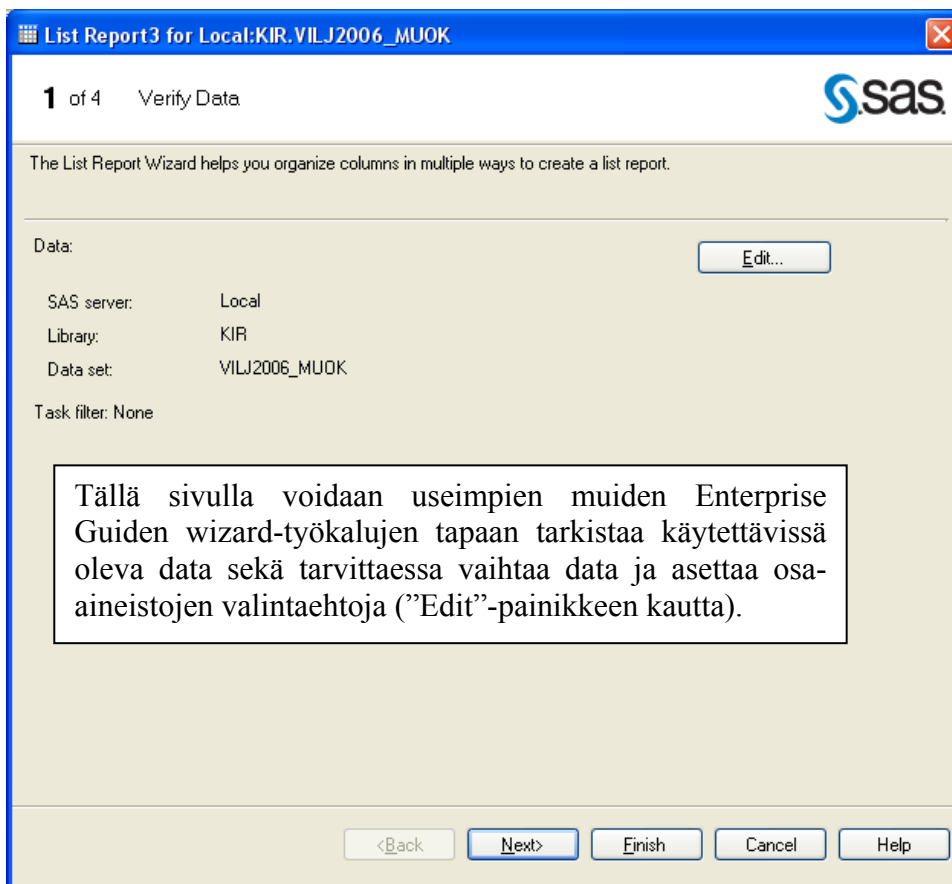
Listaus voidaan tuottaa käyttäen PRINT-proseduuria:

```
PROC PRINT DATA=kir.vilj2006_muok NOOBS;  
  VAR nimi peltoa nautoja metsaa tuotsuun paino pituus ika tulot  
      ajanmuk tulevais;  
  BY sukupuoli;  
RUN;
```

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

List Report Wizard -työkalu

Monipuolisemmat listaukset ja raportit voidaan tuottaa List Report Wizard -työkalulla. Tässä esimerkissä tuotetaan listaustyyppinen tarkastelu valitsemalla työtilassa tarkastelussa olevaan aineistoon liittyvästä asiayhteysvalikosta ”Describe” -> ”List Report Wizard”, jolloin aukeaa wizardin ensimmäinen sivu:



Toiselle sivulle päästään painamalla ”Next”:

List Report Wizard ottaa kaikki aineiston muuttujat oletuksena tarkasteltaviksi, jos muuttujia on alle 20. Jos muuttujia on yli 20, muuttujat pitää valita tarkasteluun.

2 of 4 Define List

Preview: <Select a column to move or edit>

id (SUM)	nimi	sukup (SUM)	peltoa (SUM)	nautoja (SUM)	metsaa (SUM)	tuotsuun (SUM)	paino (SUM)	pit (S

Muuttujien valintaa ja niiden roolien määrittelyä voidaan tehdä valitsemalla "Edit" -> "Assign Columns"

"Edit"-valikon kautta tehdään tässä yhteydessä myös valinta "Table Properties" -> "Show every row", koska tällä kertaa on tavoitteena tuottaa listaus-tyyppinen aineiston tarkastelu.

Tältä sivulta voidaan tehdä tulostaulukon asetusten määrittelyä monella eri tavalla. "Add" ja "Edit"-painikkeiden lisäksi määrittely onnistuu muun muassa klikkaamalla muuttujan nimeä hiiren oikealla painikkeella "Preview"-ikkunassa.

<Back Next> Finish Cancel Help

Edellisessä ikkunassa tehtävällä "Edit" -> "Assign Columns" -valinnalla päästään seuraavaan ikkunaan:

Assign Columns

Available columns:

- id
- nimi

Selected columns:

- id
- nimi
- peltoa
- nautoja
- metsaa
- paino
- pituus
- ika
- tulot
- ajanmuk
- tulevais

Display all values

Group rows by e

- tuotsuun

Create a separa

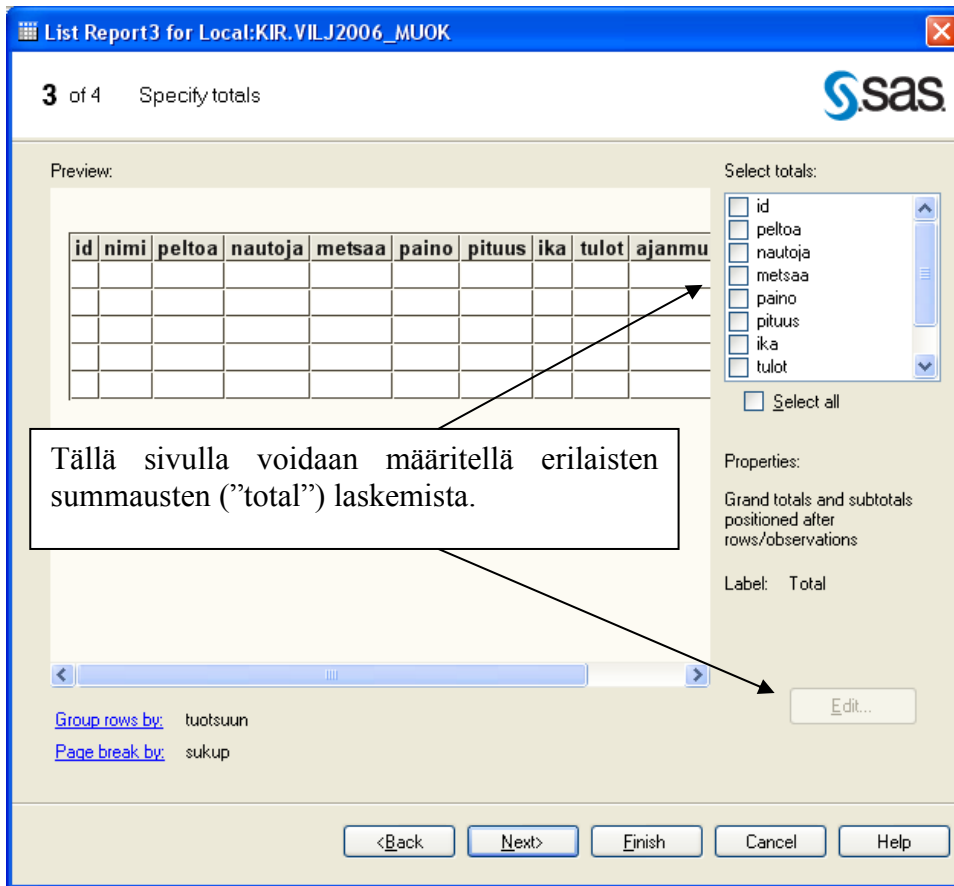
- sukup

"Selected columns" -ikkunassa on lueteltu tuloksena tuotettavassa taulukossa tarkasteltavat muuttujat. Kaikille muuttujille on valittu alasetus "Display all values", koska halutaan tehdä listaus-tyyppinen tarkastelu. Valittavana olisi myös erilaisia kuvailevia tunnuslukuja (kuten keskiarvo).

tuotsuun-muuttuja on sijoitettu ryhmittelevään rooliin siten, että eri tuotantosuunnat on ryhmitelty saman taulukon sisällä.

sukup-muuttuja on sijoitettu ryhmittelevään rooliin siten, että sukupuolittain tuotetaan erilliset taulukot

Painamalla "OK" ja "Next" päästään työkalun kolmannelle sivulle:



Painamalla "Next" päästään työkalun viimeiselle sivulle, jossa voidaan määrittää otsikko ja alaviite ja painamalla "Finish" saadaan aikaan seuraava tuloste:

SAS Enterprise Guide - FG42_projekti.egg

List Report3

List Report

sukup = 1

id	nimi	peltoa	nautoja	metsaa	paino	pituus	ika	tulot	ajanmuk	tulevais
tuotsuun = 1										
1	Jorma	247	0	120	112	178	54	92000	2	1
2	Tommi	110	0	70	90	195	31	25000	3	2
4	Jani	20	0	30	75	175	33	16000	4	2
6	Jarmo	112	0	24	68	181	35	28000	1	2
7	Sampsa	210	50	84	94	192	27	68000	2	1
8	Antero	42	0	36	82	178	64	21000	4	3
11	Ossi	167	18	15	88	182	49	37900	1	1
14	Kalervo	94	0	115	73	186	43	31000	1	2
15	Kullervo	137	25	37	74	188	43	37120	1	1
17	Nico	77	0	52	67	184	21	26700	3	2
18	Timo	19	0	10	94	172	27	38000	4	2
19	Seppo	60	0	15	92	189	35	21800	2	1
22	Jonne	27	19	15	58	179	34	19800	3	3
24	Jaakko	200	30	30	109	190	36	50000	1	1
26	Jesse	100	0	71	72	184	22	22000	1	2
27	Jeremias	20	14	25	67	175	42	16500	5	3
28	Konsta	8	7	94	65	178	40	8000	5	1
29	Ilmari	179	0	15	97	183	53	29000	2	2
30	Sami 2	42	0	31	79	180	39	28900	2	2
31	Aki	22	10	11	52	163	16	16000	4	3
32	Iso-Pepe	200	120	150	120	205	45	75000	1	1
33	Pena	47	0	24	80	180	38	19000	3	2
36	Pertti	62	0	20	85	179	40	24000	3	2
37	Simo	51	12	12	99	176	27	26500	3	2
42	Hassan	50	0	10	90	180	53	40000	3	2
43	Sampo	120	0	150	90	190	45	200000	2	1
48	Mika	125	24	11	77	181	47	22600	3	1
49	Petteri	26	0	78	85	179	58	16300	4	2
51	Kalle	37	0	48	73	168	42	17450	4	2
53	John	100	0	500	90	192	50	30000	2	1

Vaihtoehtoinen ratkaisu SAS-koodilla:

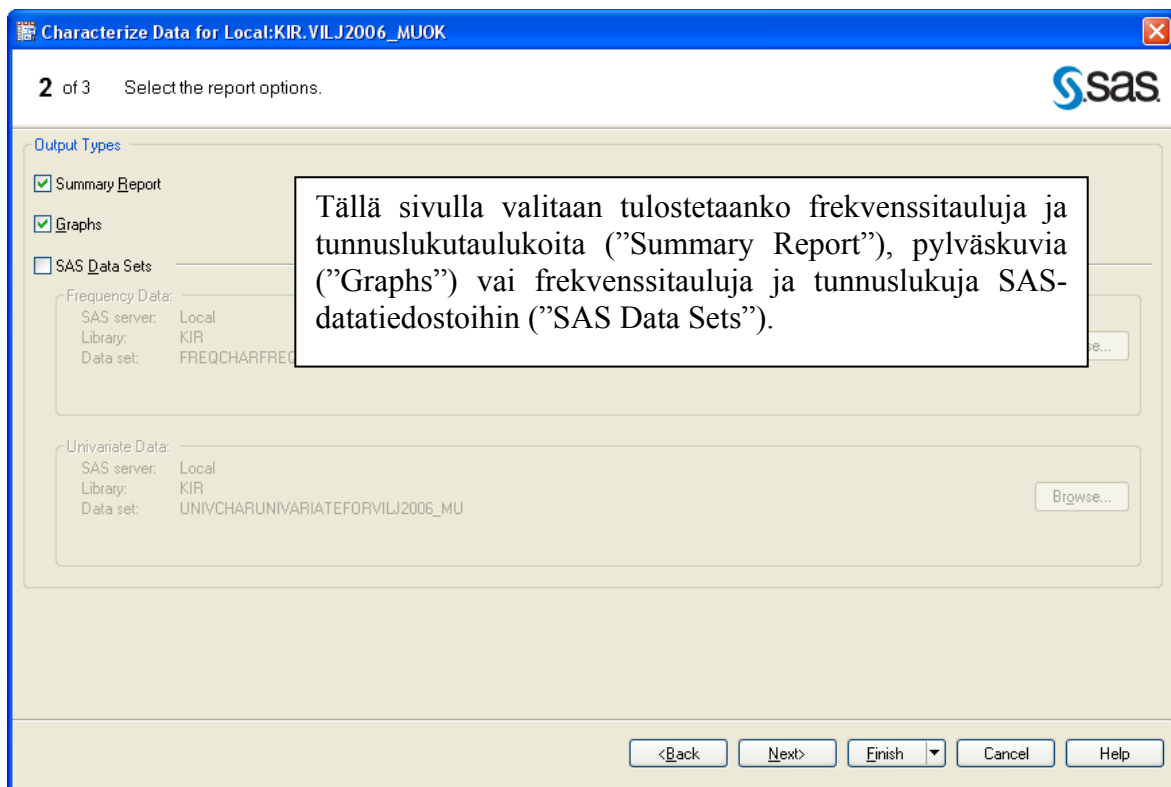
REPORT-proseduurilla voidaan tuottaa monipuolisesti listauksia ja taulukointeja. Tässä yksinkertainen esimerkki:

```
PROC REPORT DATA=kir.vilj2006_muok;  
  COLUMN sukup tuotsuun nimi peltoa nautoja metsaa;  
  DEFINE sukup / GROUP;  
  DEFINE tuotsuun / GROUP;  
RUN;
```

Lisätietoa REPORT-proseduurista löytyy muun muassa hakemalla laajemmasta SASin ohjeesta: ”Help” -> ”SAS Syntax Help”.

8.2 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 avaamalla data työtilaan ja valitsemalla asiayhteysvalikosta ”Describe” -> ”Characterize Data”. Ensimmäisellä sivulla voidaan vaikuttaa käytettävään dataan. Toisella sivulla tehdään tärkeimmät valinnat:



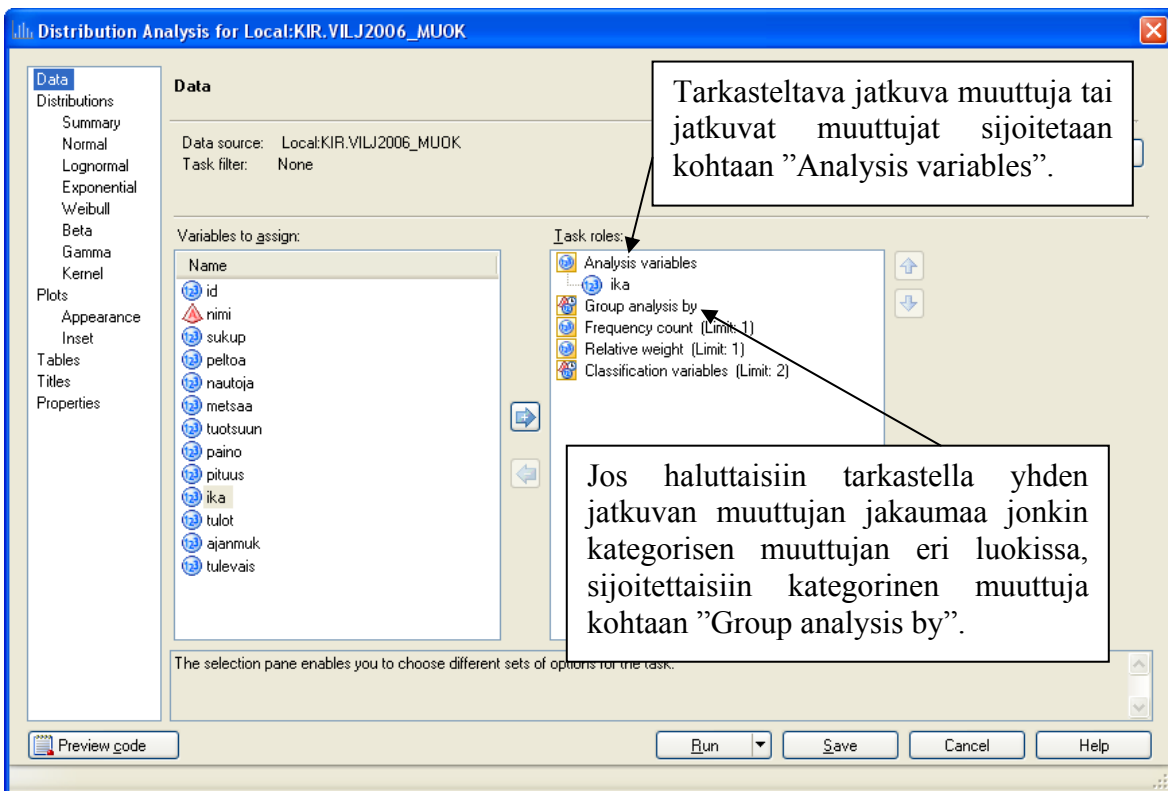
Viimeisellä sivulla voi vielä valita, montako kategorisen muuttujan luokkaa maksimissaan huomioidaan erillisinä luokkina. Lopuksi painetaan "Finish".

8.3 Jatkuvan muuttujan jakauman tarkastelu

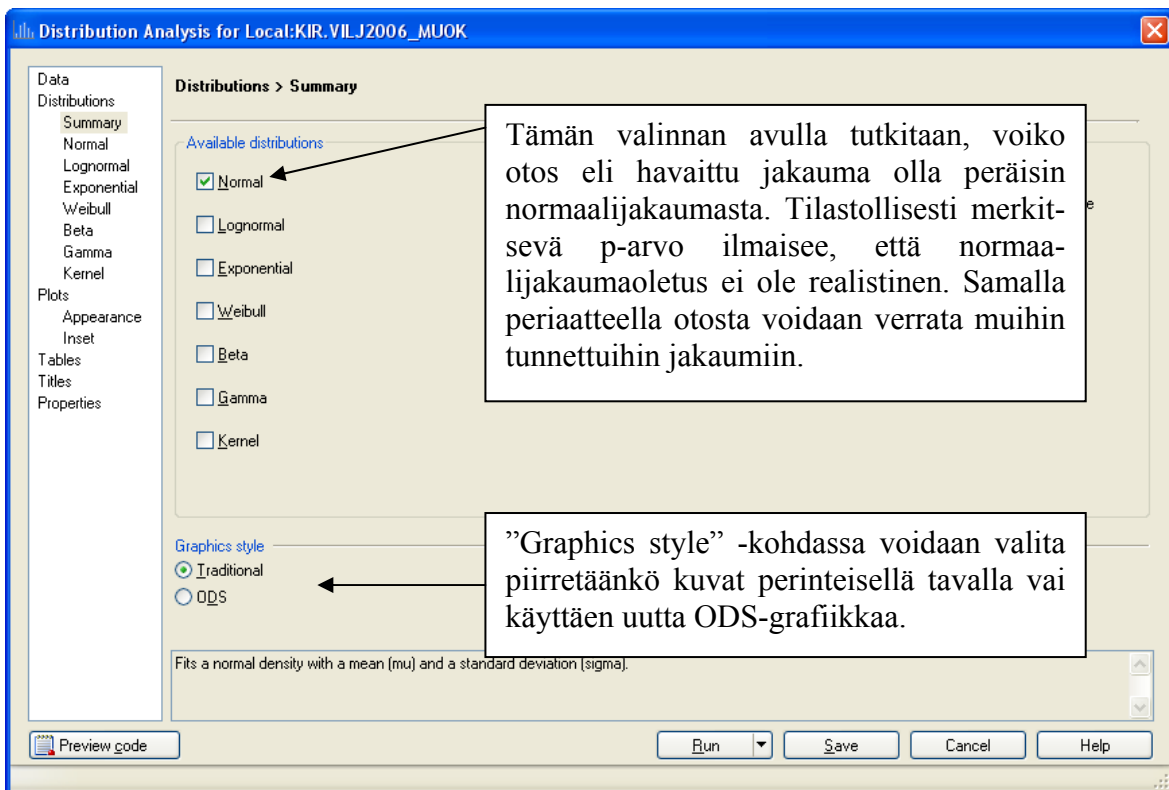
Datassa olevan jatkuvan muuttujan jakauman tarkastelua varten esitellään kolme erillistä SAS Enterprise Guiden työkalua sekä kahden työkalun wizard-vastineet.

8.3.1 Distribution Analysis -työkalu

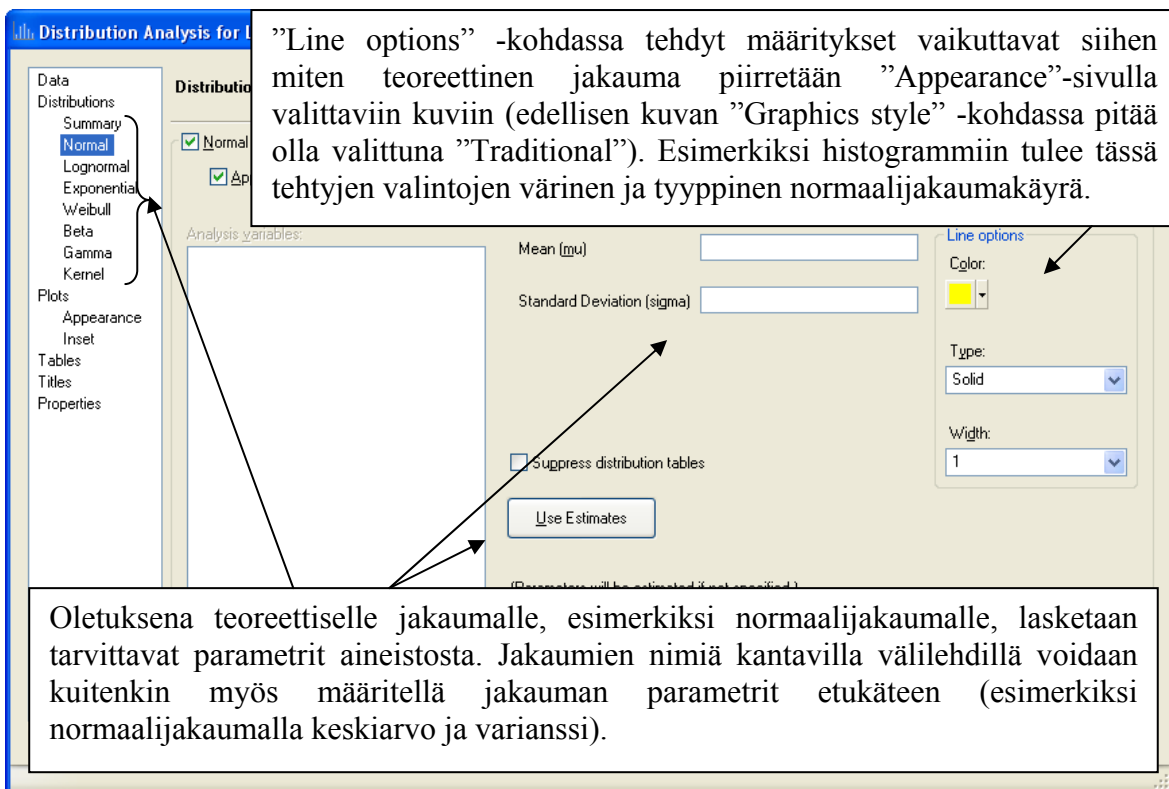
Yhden jatkuvan muuttujan jakauman tarkastelua on kätevää tehdä ”Distribution Analysis” työkalulla, jolla saadaan tulostetuksi taulukoita jakaumaa kuvaavista tunnusluvuista. Lisäksi voidaan testata, onko havaittu otos peräisin jostakin teoreettisesta jakaumasta, esimerkiksi normaalijakaumasta. Myös erilaisia jakaumia havainnollistavia kuvia saa tulostettua kätevästi. Alkuun päästään avaamalla datamatriisi työtilaan ja klikkaamalla asiayhteysvalikosta ”Describe” -> ”Distribution Analysis”. Tällöin aukeaa ”Data” -välilehti:



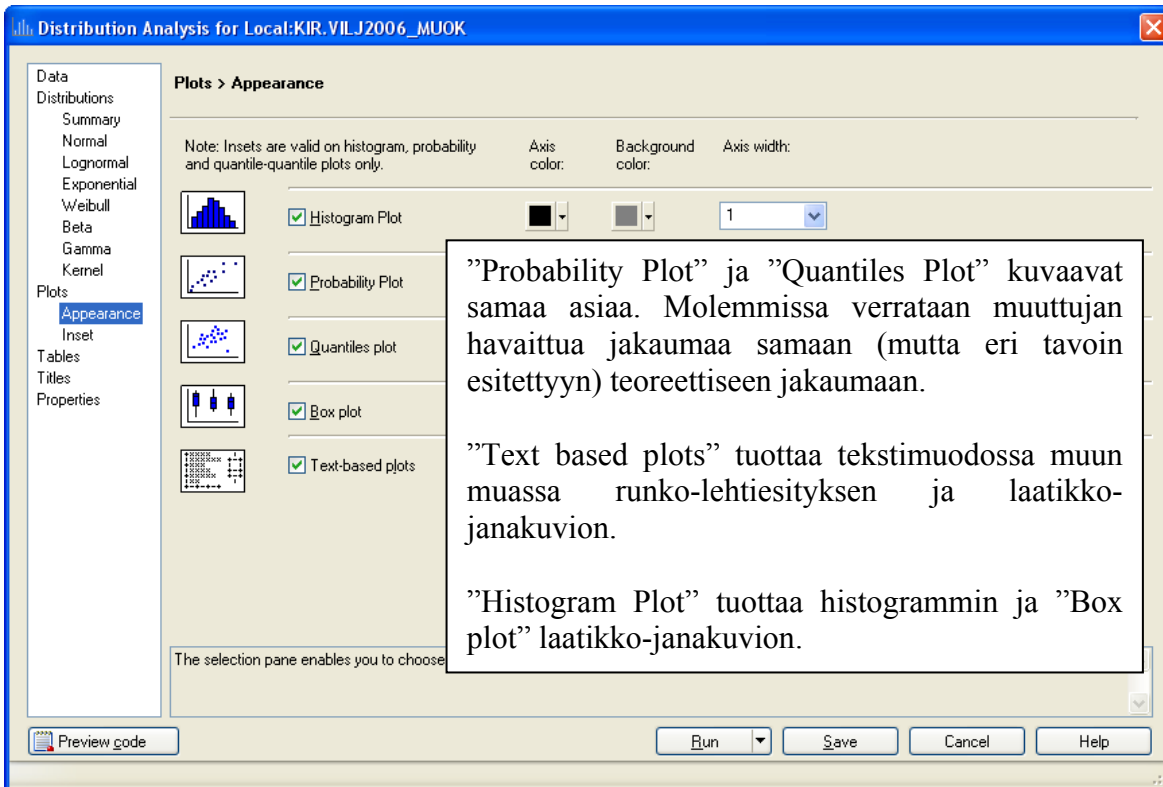
Välilehdellä ”Summary” voidaan määrittellä testit, joilla tutkitaan ovatko havaitut muuttujan arvot otos jostakin tunnetusta jakaumasta:



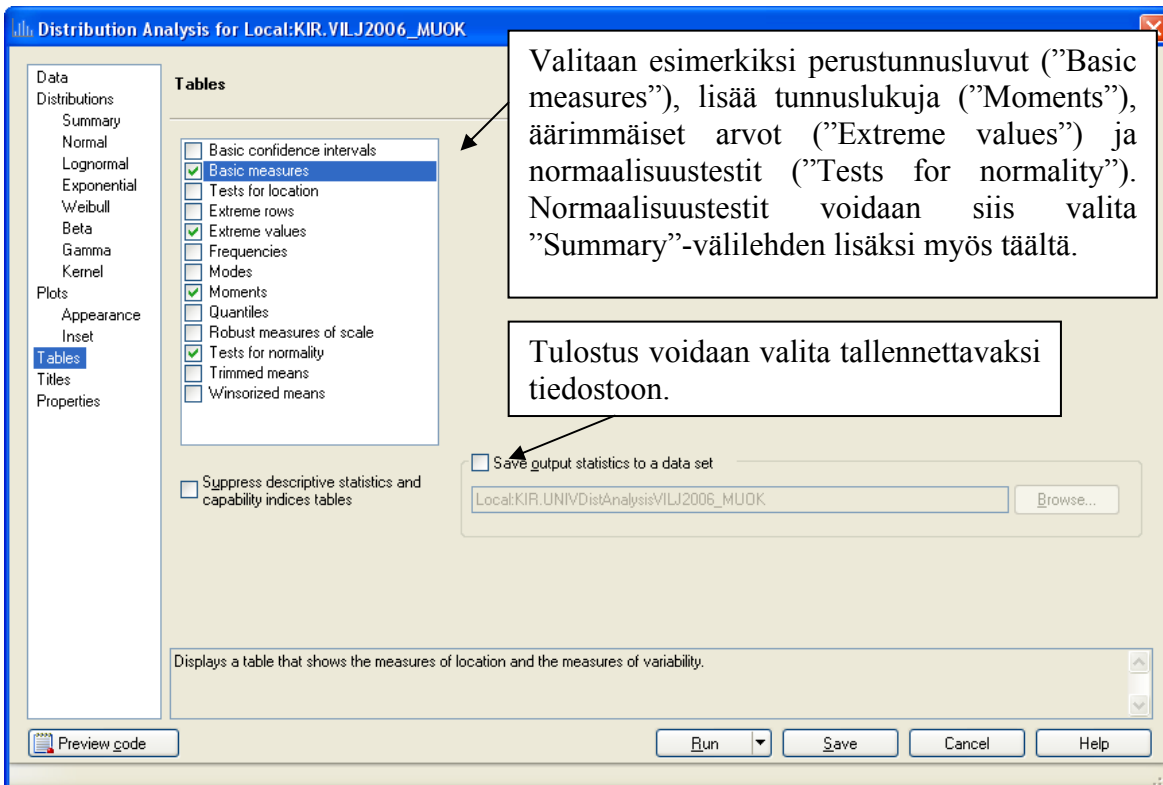
Jakaumatarkasteluihin liittyviä yksityiskohtia voi määrittää kyseisen tunnetun jakauman nimeä kantavalla välilehdellä:



”Plots” -> ”Appearance”-välilehdellä määritellään muuttujan jakaamaa havainnollistavat kuvat:



”Tables”-välilehdellä puolestaan määritellään tulostettavat tunnuslukutaulut:



Lopuksi valitaan ”Run”.

Vaihtoehtoinen ratkaisu SAS-koodilla:

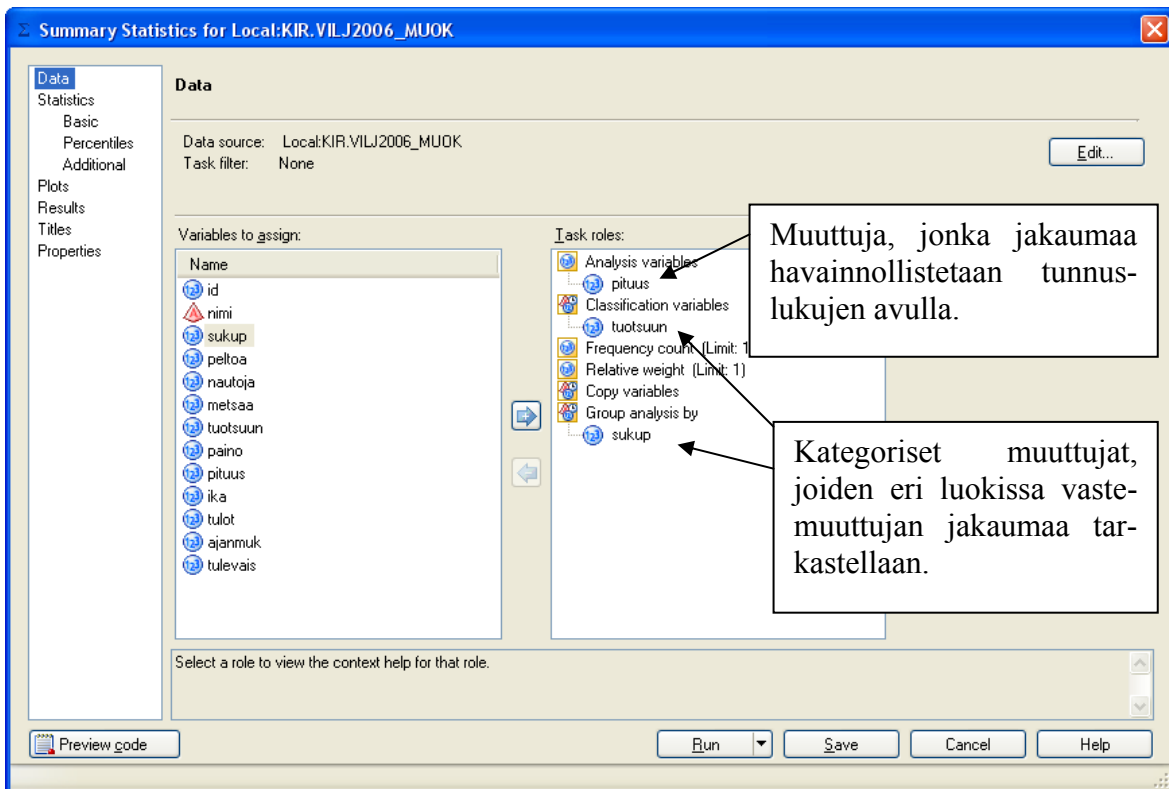
Distribution Analysis -työkalun vastine koodikäytössä on UNIVARIATE-proseduuri:

```
PROC UNIVARIATE DATA=kir.vilj2006_muok PLOTS NORMAL;  
VAR ika;  
HISTOGRAM ika / NORMAL;  
PROBPLOT ika / NORMAL (MU=EST SIGMA=EST);  
QQPLOT ika / NORMAL (MU=EST SIGMA=EST);  
RUN;
```

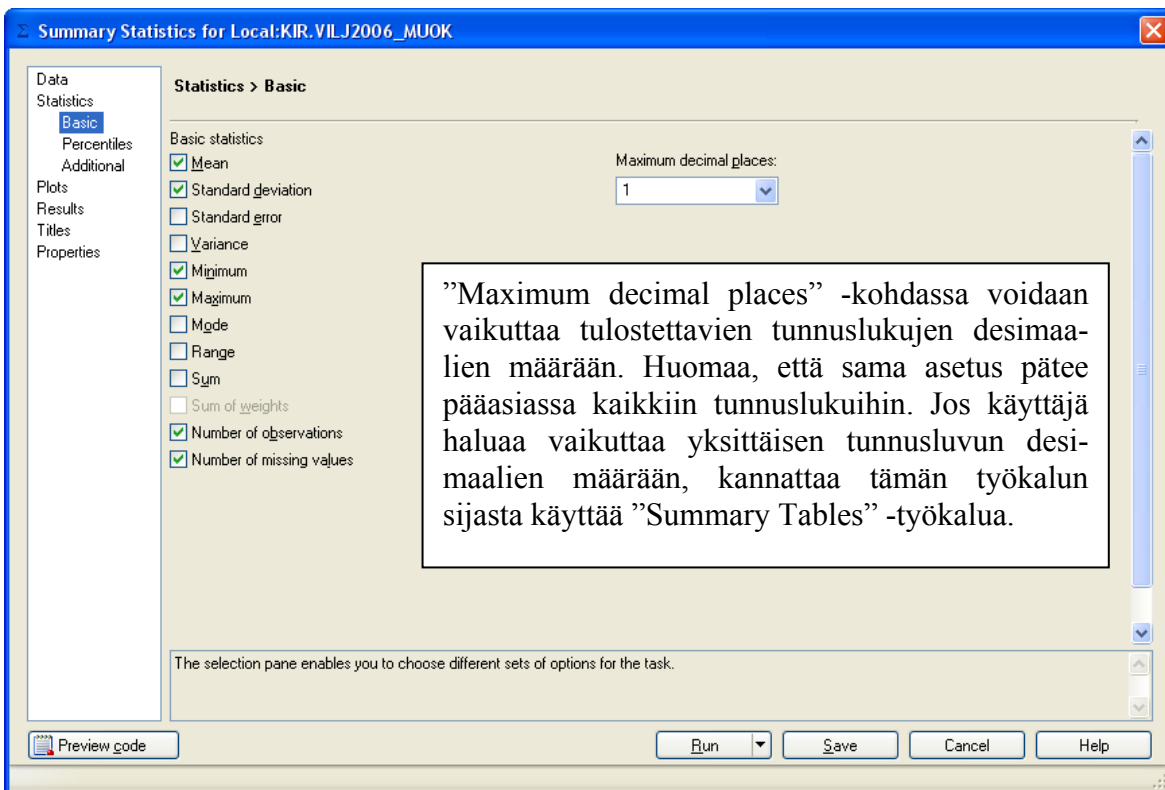
Tehtävä: Tutki Viljelijä 2006 -aineistosta (**vilj2006_muok**) ikä-muuttujan (ika) jakaumaa. Tarkastele sijaintia, hajontaa, muotoa ja poikkeavia havaintoja.

8.3.2 Summary Statistics -työkalu

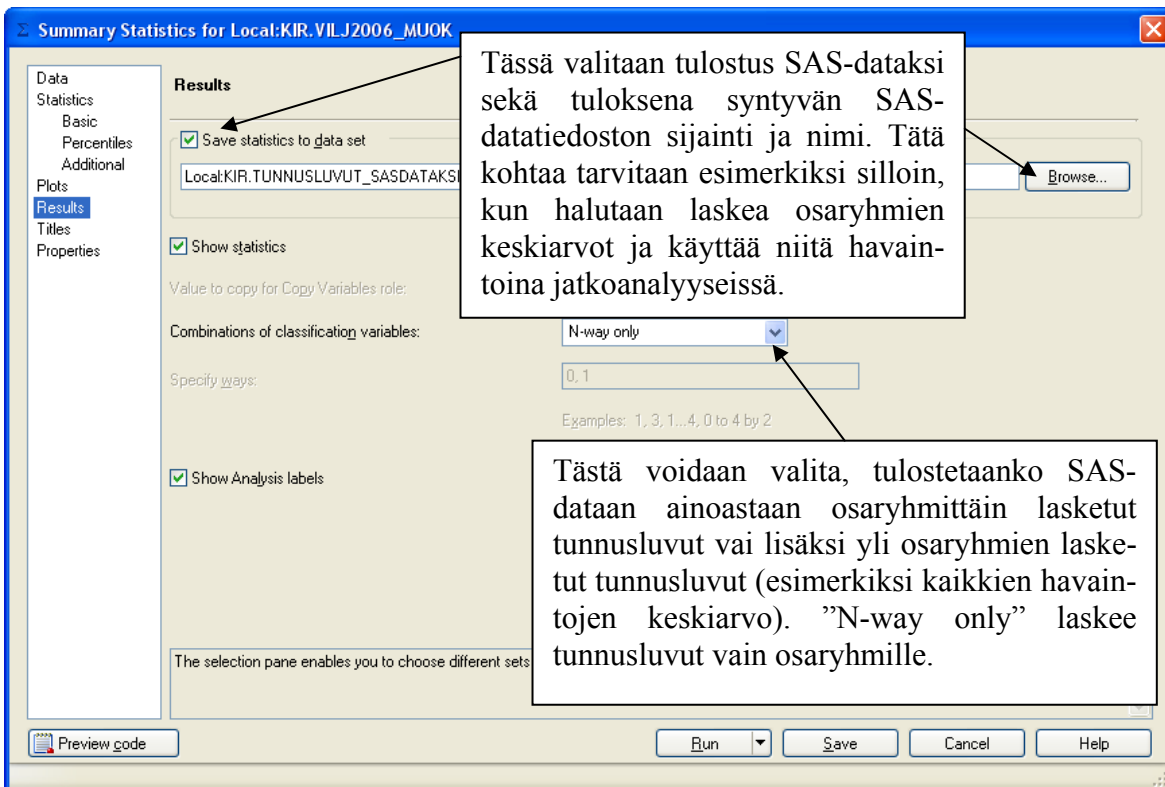
Enterprise Guiden Summary Statistics -työkalulla on mahdollista tulostaa jatkuvan muuttujan jakaumaa kuvaavia tunnuslukuja taulukkomuodossa. Tämä on erityisen kätevä työkalu, jos halutaan laskea tunnuslukuja erikseen yhden tai useamman kategorisen muuttujan luokissa. Työkalu on erityisen tarpeellinen myös, jos halutaan tallentaa tunnuslukuja SAS-dataksi jatkokäyttöä varten. Työkalu saadaan käyttöön avaamalla käytettävä datamatriisi työtilaan ja valitsemalla asiayhteysvalikosta: **"Describe" -> "Summary statistics"**, jolloin aukeaa "Data" -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 (valitsemalla ”Run”) saadaan seuraava taulukko:

”Classification variables” tuottaa erilliset luokat samaan taulukkoon.

”Group charts by” tuottaa muuttujan luokille erilliset taulukot.

Summary Statistics
Results
The MEANS Procedure

sukupuoli=1

Analysis Variable : pituus							
tuotsuun	N Obs	Mean	Std Dev	Minimum	Maximum	N	N Miss
1	30	181.7	8.2	163.0	205.0	30	0
2	12	180.2	7.2	169.0	190.0	12	0

sukupuoli=2

Analysis Variable : pituus							
tuotsuun	N Obs	Mean	Std Dev	Minimum	Maximum	N	N Miss
1	6	161.2	6.0	152.0	168.0	6	0
2	5	168.8	6.1	161.0	177.0	5	0

Generated by the SAS System (Local, XP_PRO) on 09. lokakuuta 2009 at 15:45:37

Page Break

Vaihtoehtoinen ratkaisu SAS-koodilla:

Summary Statistics -työkalun vastine koodikäytössä on MEANS-proseduuri:

```
PROC MEANS DATA=kir.vilj2006_muok N MEAN STD MIN MAX;
  CLASS tuotsuun sukupuoli;
  VAR pituus;
RUN;
```

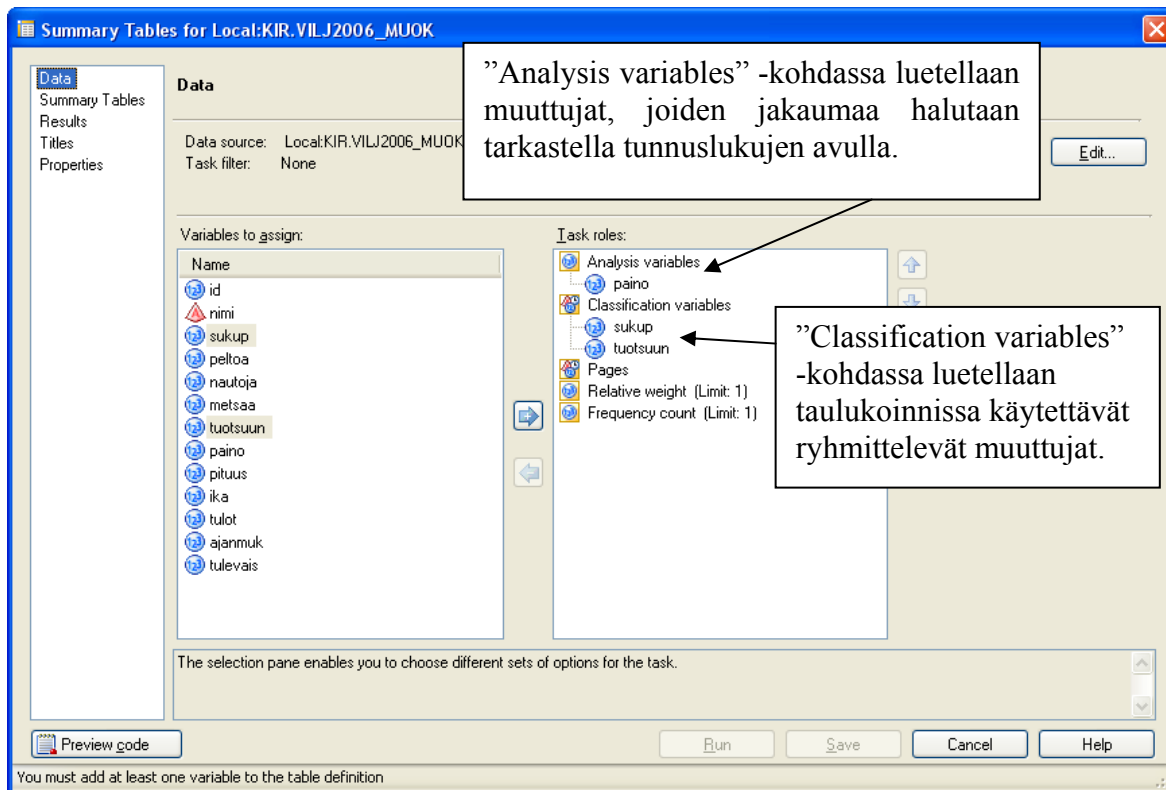
Alla SAS-koodi tilanteelle, jossa tulokset halutaan ruudulle tulostuksen sijaan tallentaa SAS-dataksi:

```
PROC MEANS DATA=kir.vilj2006_muok NWAY NOPRINT;
  CLASS tuotsuun sukupuoli;
  VAR pituus;
  OUTPUT OUT=ka_data N= MEAN= STD= MIN= MAX= / AUTONAME AUTOLABEL;
RUN;
```

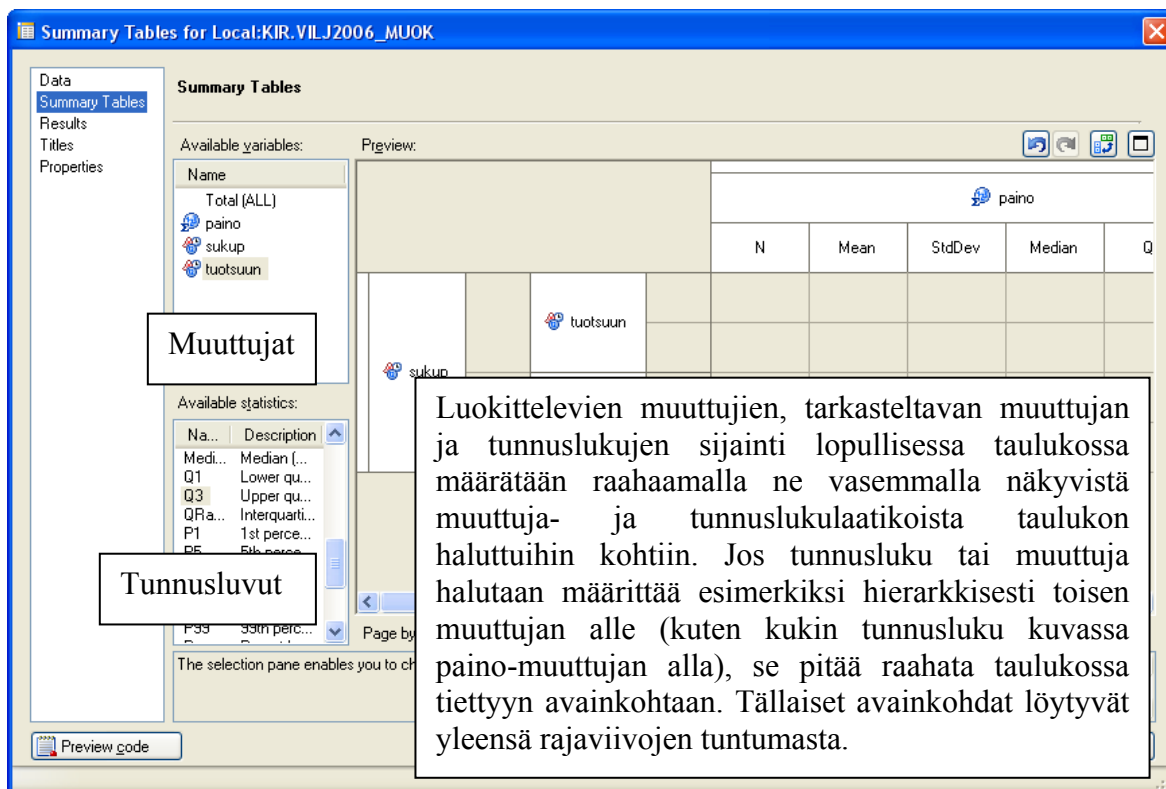
Tehtävä: Tutki Viljelijä 2006 -aineistosta (**vilj2006_muok**) pituus-muuttujan jakaamaa sukupuolen ja tuotantosuunnan eri luokkayhdistelmissä. Käytä tähän tarkasteluun tilastollisia tunnuslukuja ja mahdollisesti myös kuvia.

8.3.3 Summary Tables -työkalu

Kolmas tapa laskea tunnuslukuja on käyttää taulukointityökalua ”Summary Tables”. Tämä työkalu saadaan käyttöön avaamalla käytettävä data työtilaan ja valitsemalla asiayhteysvalikosta ”Describe” -> ”Summary Tables”. Muuttujat esitellään ”Data”-välilehdellä:



Varsinaisen taulukon määrittelyyn päästään ”Summary Tables” -välilehdellä:



Tunnuslukuihin voi tehdä lisämääryksiä valikosta, joka aukeaa klikkaamalla hiiren oikealla painikkeella 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
sukup	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-tilukko on siirretty EG:n tulostusikkunasta Wordiin kopioimalla leikepöydän kautta. Samalla tekniikalla voidaan siirtää muitakin EG:n RTF-tulostusobjekteja Wordiin. Tällöin taulukkoa päästään jatkamuokkaamaan Wordissa, muun muassa siirtämällä sarakkeiden rajoja ja muuttamalla fonttia. Taulukosta saadaan Wordissa muokattua esimerkiksi seuraavanlainen:

		Paino (kg)					
		N	Keski-arvo	Keskihajonta	Mediaani	Alakvartiili	Yläkvartiili
Sukupuoli	Tuotantosuunta						
Mies	Peltoviljely	30	83.1	15.58	82.5	73.0	92.0
	Kotieläimet	12	84.1	17.15	80.5	76.5	92.5
Nainen	Peltoviljely	6	54.3	9.95	57.0	43.0	63.0
	Kotieläimet	5	73.8	15.75	80.0	64.0	81.0

Taustavärit saa poistettua Wordissa (Office 2003) esimerkiksi valitsemalla koko taulukko aktiiviseksi (klikkaamalla vasemmassa yläkulmassa toisinaan näkyvää laatikkoa), klikkaamalla hiiren oikeaa painiketta ja valitsemalla ”Reunat ja sävytys”. Aukeavassa ikkunassa mennään ”Sävytys”-välilehdelle ja valitaan ”Täyttökuvio”-kohdassa ”Ei täyttöä”.

Vaihtoehtoinen ratkaisu SAS-koodilla:

Summary Tables -työkalun vastine koodikäytössä on TABULATE-proseduuri:

```
PROC TABULATE DATA=kir.vilj2006_muok;
  VAR paino;
  CLASS sukup tuotsuun;
  TABLE sukup*tuotsuun,paino*(N Mean StdDev Median Q1 Q3);
RUN;
```

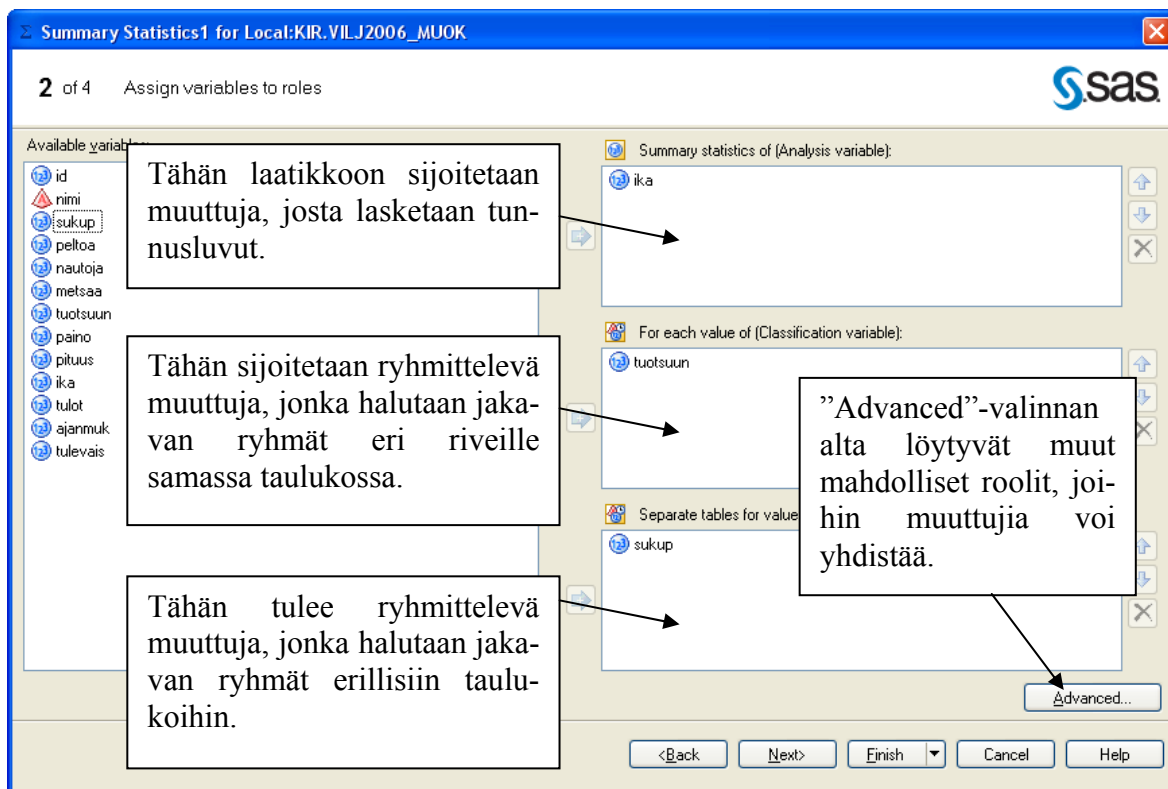
Tehtävä: Tee Viljelijä 2006 -aineistosta (vilj2006_muok) taulukko, jossa ika-muuttujalle on laskettu soveltuvia tunnuslukuja sukupuolen ja tuotantosuunnan eri luokkakombinaatioissa. Vie taulukko myös Wordiin tai muuhun tekstinkäsittelyohjelmaan ja harjoittele taulukon muokkaamista siellä.

8.3.4 Tunnuslukujen taulukointi wizard-työkaluilla

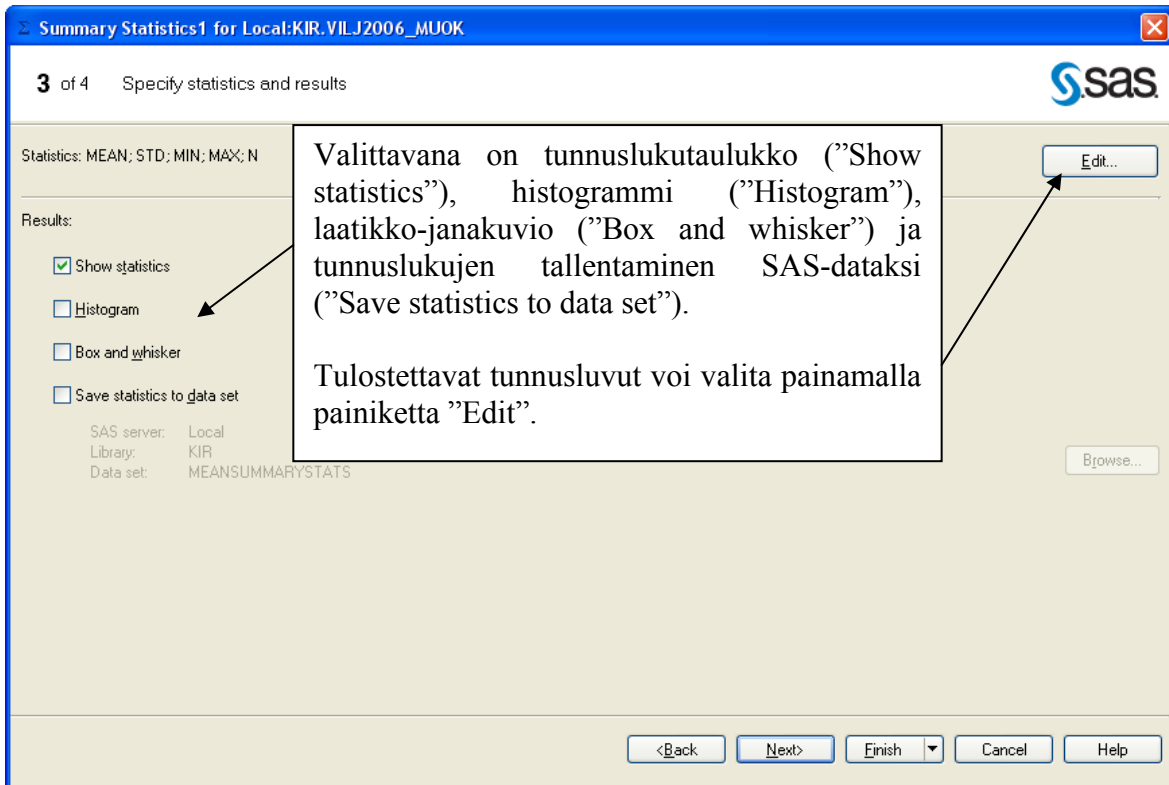
Kuvien tapaan myös jatkuvan muuttujan jakauman tarkasteluun löytyy Enterprise Guidesta muutama wizard-työkalu. Näiden työkalujen tarkoituksena on tarjota entistä helppokäyttöisempi vaihtoehtoinen tapa tunnuslukujen taulukointiin. Wizardissa määrittelyt on jaettu useammalle sivulle, joiden läpi edetään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

Summary Statistics Wizard

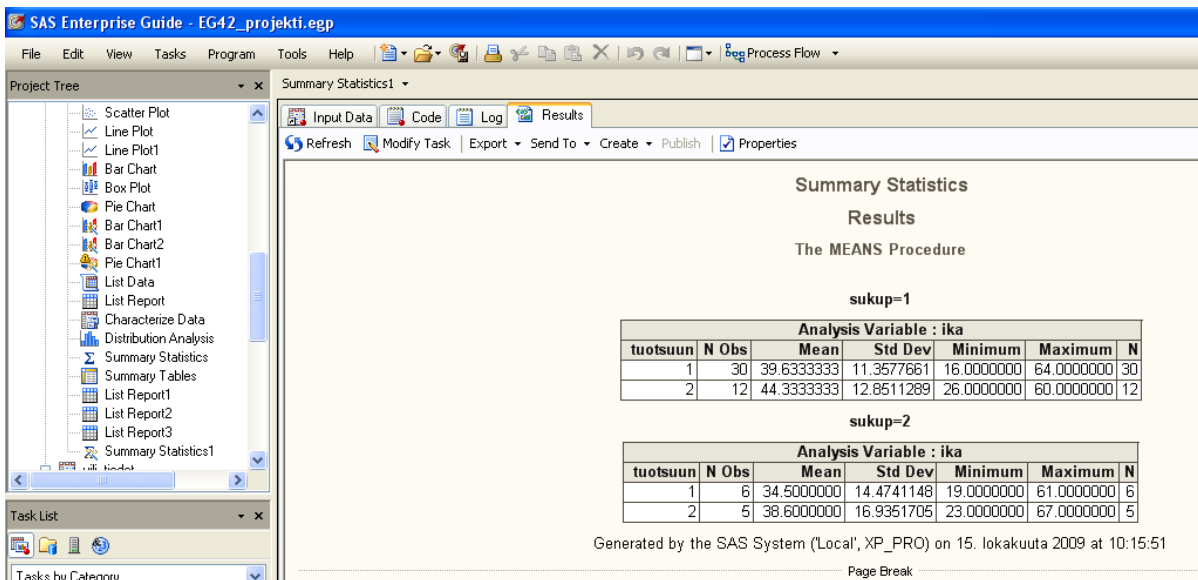
Tarkastellaan ensin ”Summary Statistics Wizard” -työkalua, joka saadaan käyttöön, kun avataan käytettävä data työtilaan ja valitaan asiayhteysvalikosta ”Describe” -> ” Summary Statistics Wizard”. Ensimmäisellä sivulla voidaan tarkistaa, että käytössä on haluttu data sekä tarvittaessa vaihtaa dataa ja valita tarkasteluun osa-aineisto valintaehtoien avulla. Toiselle välilehdelle päästään klikkaamalla ”Next”. Toisella välilehdellä määritellään muuttujien roolit:



Kolmannella sivulla valitaan mitä tulostusta halutaan:



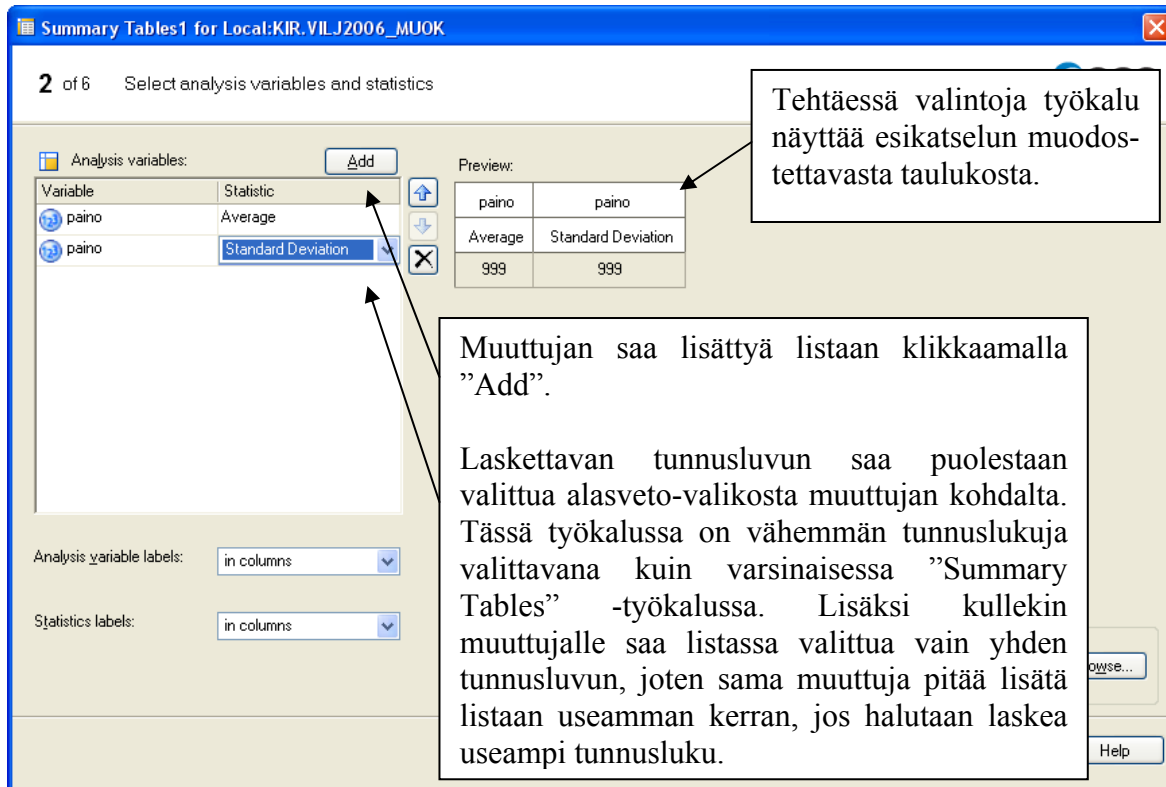
Viimeisellä sivulla voidaan vaihtaa kuvaan liittyvät otsikot ja alaviitteet. Lopuksi painetaan ”Finish”, jolloin saadaan samanlainen tuloste kuin vastaavilla valinnoilla tavallista ”Summary Statistics” -työkalua käyttäen:



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 avaamalla tarkasteltava data työtilaan ja valitsemalla asiayhteysvalikosta ”Describe” -> ” Summary Tables Wizard”. Ensimmäisellä sivulla voidaan tässäkin työkalussa varmistaa, että käytössä on haluttu data ja tarvittaessa vaihtaa dataa tai valita tarkasteluun vain osa riveistä. Toisella sivulla puolestaan määritellään muuttuja tai muuttujat, joiden arvoista lasketaan tunnuslukuja, sekä laskettavat tunnusluvut:



Summary Tables1 for Local:KIR.VILJ2006_MUOK

2 of 6 Select analysis variables and statistics

Analysis variables:

Variable	Statistic
paino	Average
paino	Standard Deviation

Preview:

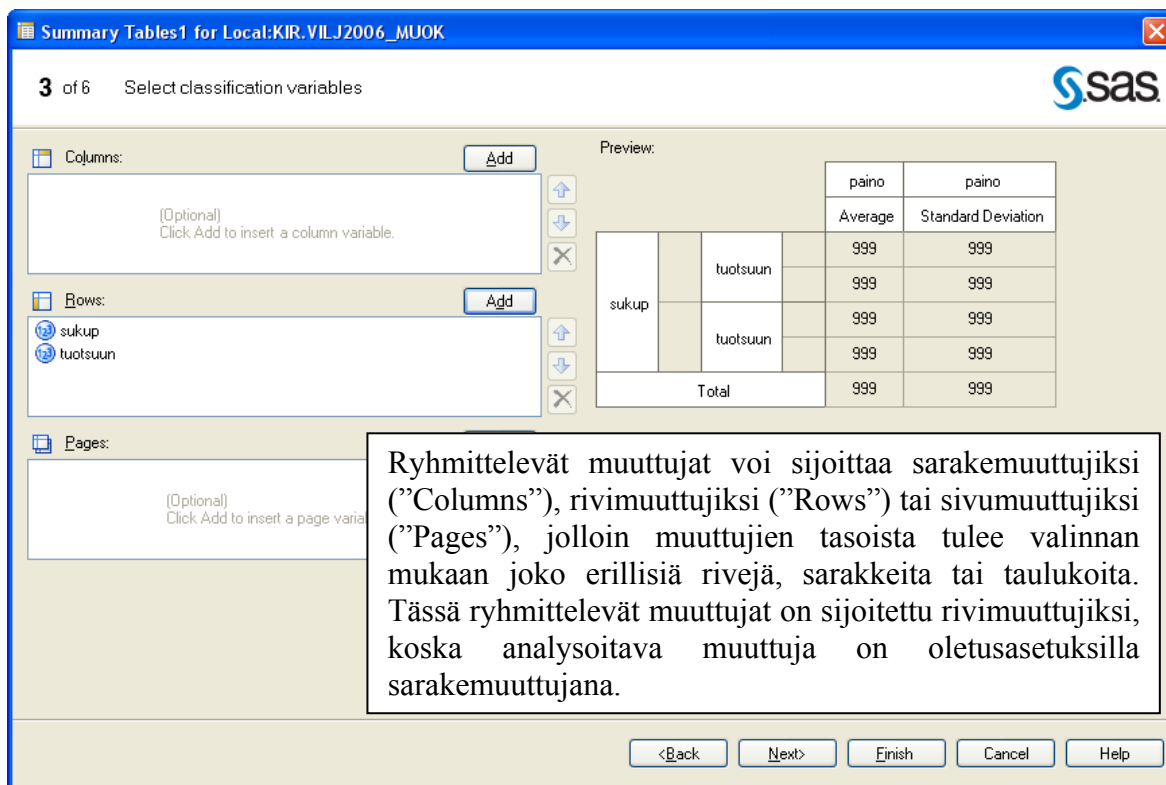
	paino	paino
Average	999	999
Standard Deviation	999	999

Tehtäessä valintoja työkalu näyttää esikatselun muodostettavasta taulukosta.

Muuttujan saa lisättyä listaan klikkaamalla ”Add”.

Laskettavan tunnusluvun saa puolestaan valittua alasveto-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:



Summary Tables1 for Local:KIR.VILJ2006_MUOK

3 of 6 Select classification variables

Columns:

(Optional)
Click Add to insert a column variable.

Rows:

sukup	tuotsuun	999	999
tuotsuun	tuotsuun	999	999
Total		999	999

Pages:

(Optional)
Click Add to insert a page variable.

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.

<Back Next> Finish Cancel Help

Neljännellä sivulla voi määrittellä, lasketaanko tunnuslukuja taulukkoon yli osaryhmien:

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
		Average	Standard Deviation
sukup	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 tuloste, joka vastaa tavallisen "Summary Tables" -työkalun tulostetta samoilla asetuksilla:

SAS Enterprise Guide - EG42_projekti.egp

File Edit View Tasks Program Tools Help

Project Tree

Summary Tables1

Summary Tables

		paino	paino
		Average	Standard Deviation
1	1	83.07	15.58
	2	84.08	17.15
2	1	54.33	9.95
	2	73.80	15.75

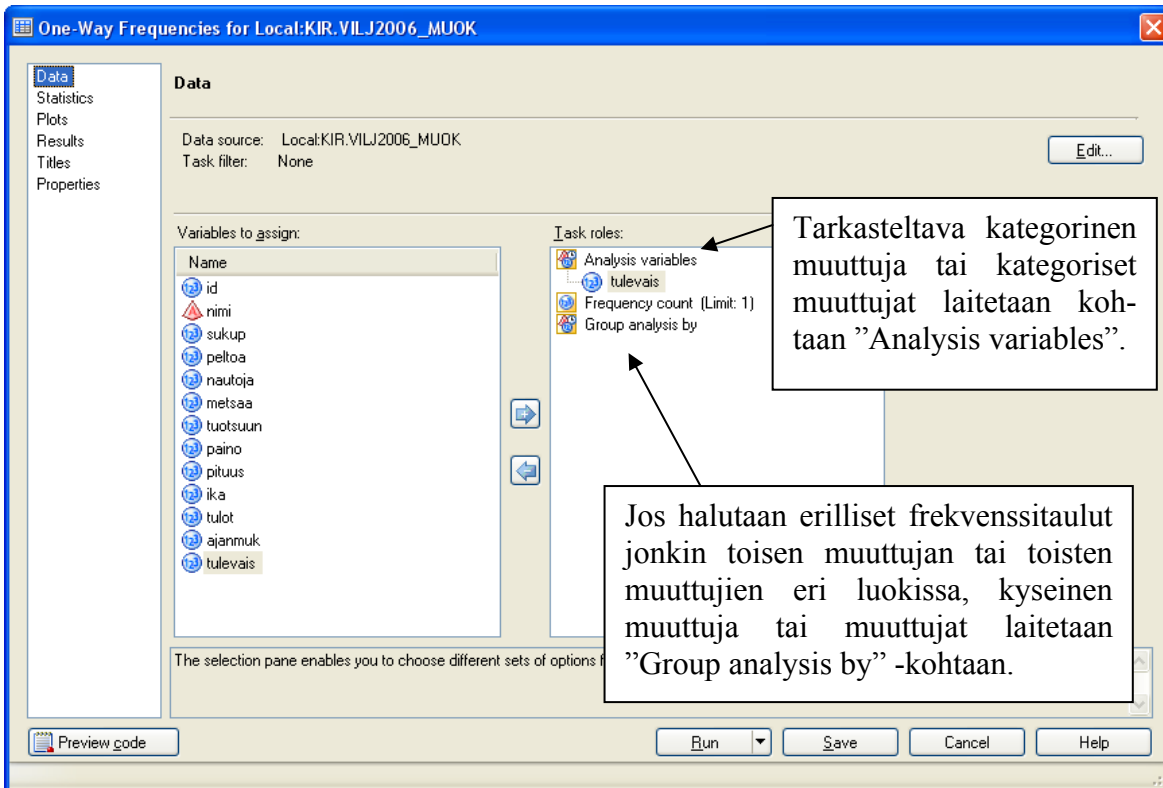
Generated by the SAS System (Local, XP_PRO) on 15. lokakuuta 2009 at 10:38:59

Page Break

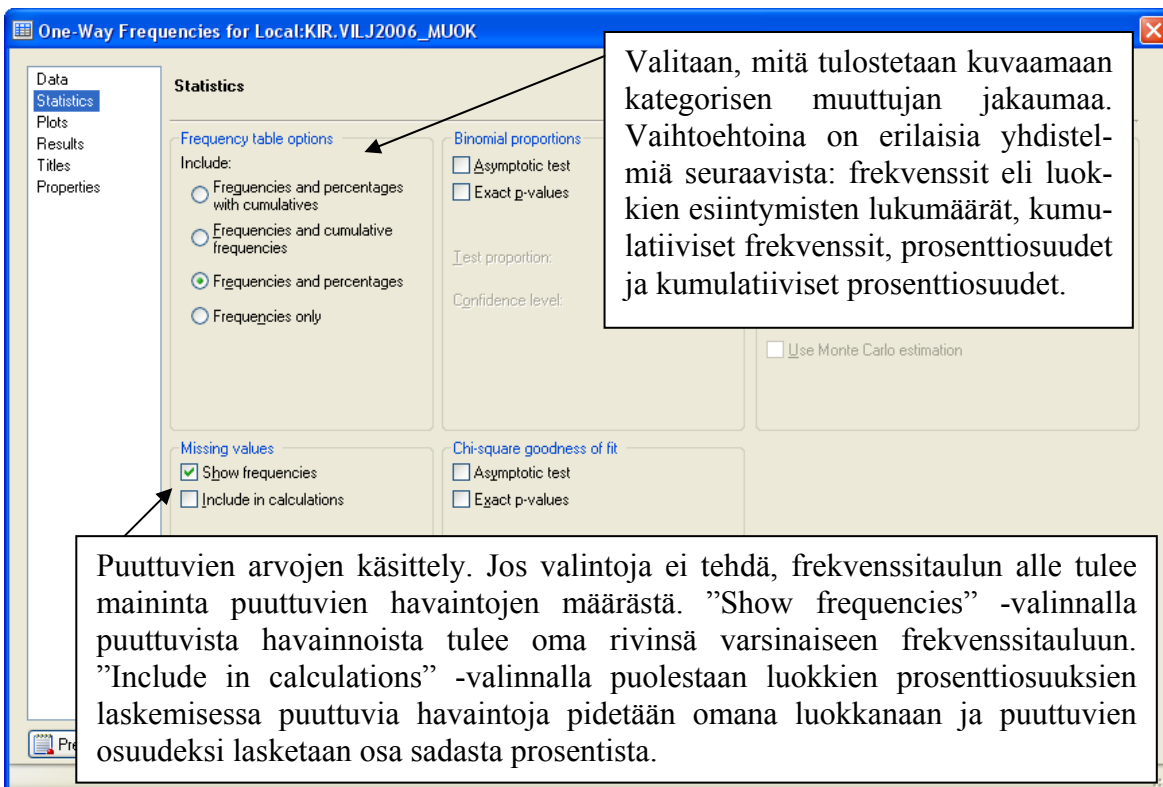
Tehtävä: Tutki Viljelijä 2006 -aineistosta (**vilj2006_muok**) paino-muuttujan jakaumaa sukupuolen ja tuotantosuunnan eri luokkakombinaatioissa. Laske tarvittavat tilastolliset tunnusluvut Summary Tables Wizard -työkalulla.

8.4 Kategorisen muuttujan jakauman tarkastelu

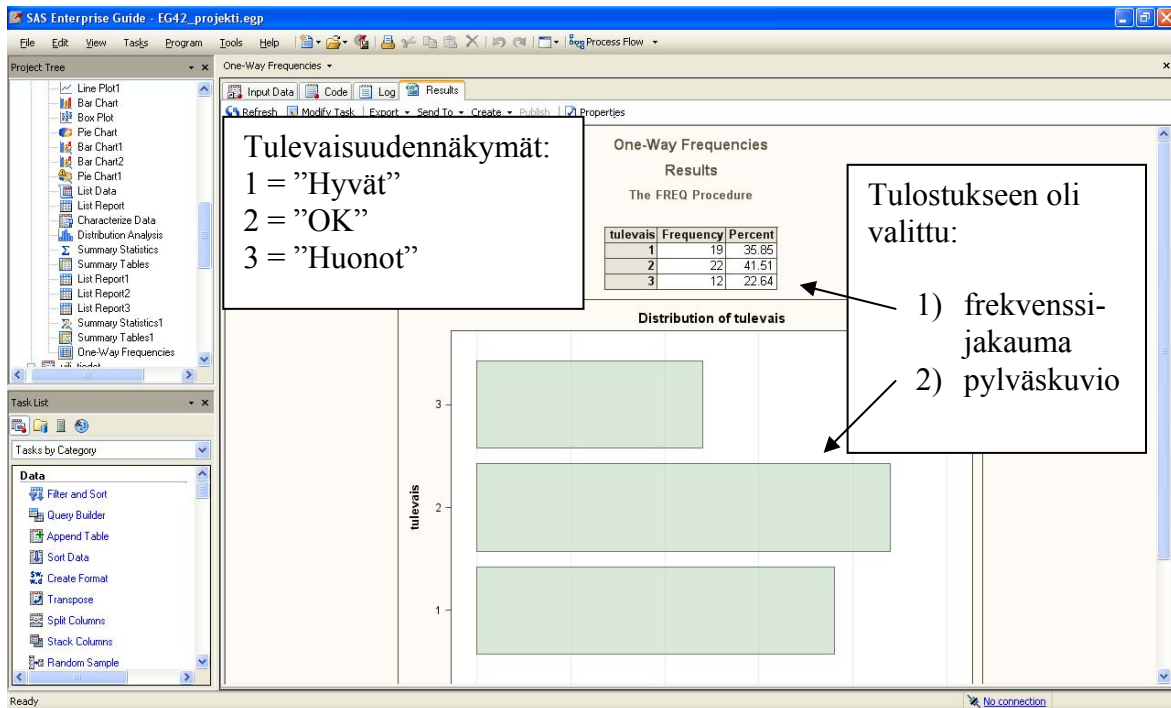
Kategorisen muuttujan jakaamaa voidaan tarkastella Enterprise Guidessa avaamalla käytettävä data työtilaan ja klikkaamalla asiayhteysvalikosta ”Describe” -> ”One-Way Frequencies”, jolloin aukeaa työkalun ”Data”-välilehti:



”Statistics”-välilehdellä määritellään frekvenssijakauman tarkasteluun liittyviä asetuksia:



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



Vaihtoehtoinen ratkaisu SAS-koodilla:

Frekvenssitauluja voi tuottaa FREQ-proseduurilla:

```
ODS GRAPHICS ON;
PROC FREQ DATA=kir.vilj2006_muok;
  TABLES tulevais / plots(only orient=horizontal)=freq;
RUN;
ODS GRAPHICS OFF;
```

ODS GRAPHICS -lauseet samoin kuin TABLES-lauseen optiot liittyvät pylväskuvan tuottamiseen ods-grafiikalla. Jos yllä olevaa koodia käytetään perus-SASissa, tulostus pitää lisäksi ohjata johonkin muuhun tulostuskohteeseen (esimerkiksi ODS RTF -lauseilla) kuin oletuksena olevaan SASin listaustulostuskohteeseen. EG:ssä tulostus tehdään oletuksena SAS Report -tulostusmuodossa, jolloin vastaavaa ongelmaa ei ole. Alla esimerkkinä sama koodi täydennettynä ODS RTF -lauseilla:

```
ODS GRAPHICS ON;
ODS RTF FILE='c:\temp\freq_testi.rtf';
PROC FREQ DATA=kir.vilj2006_muok;
  TABLES tulevais / plots(only orient=horizontal)=freq;
RUN;
ODS RTF CLOSE;
ODS GRAPHICS OFF;
```

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

8.5 Kahden muuttujan yhteyden tarkastelu

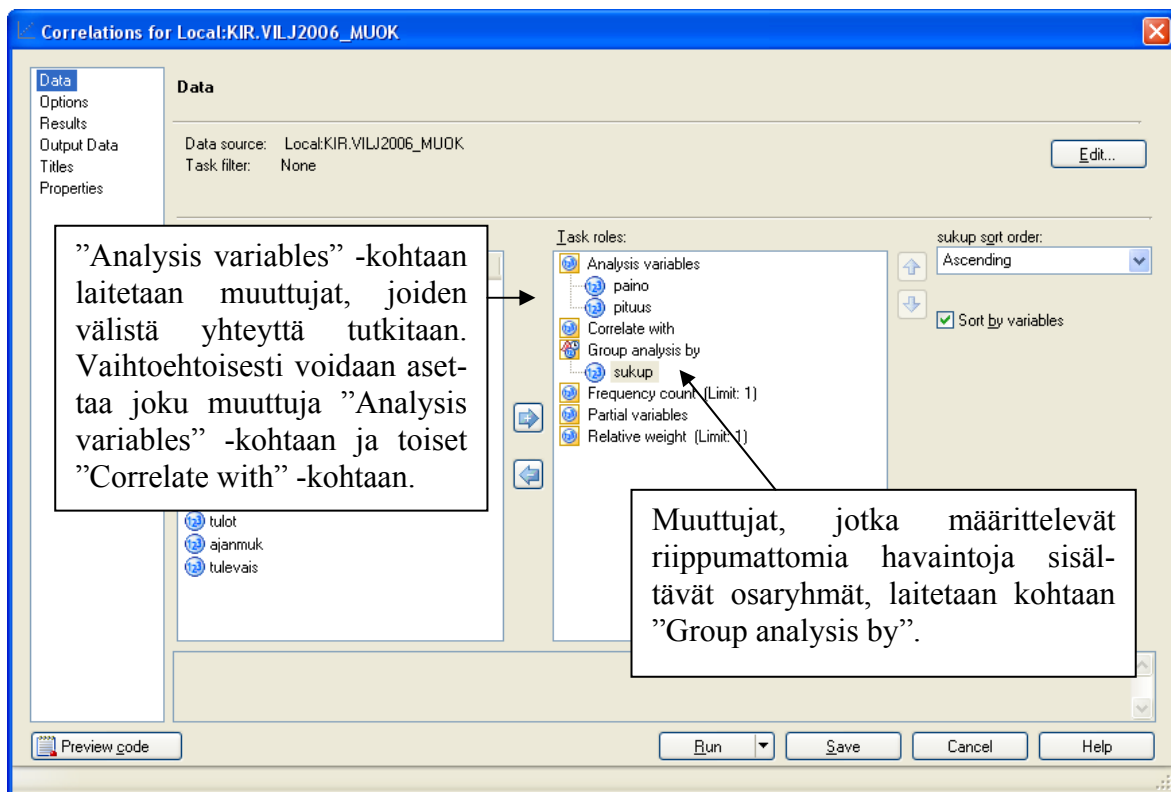
Kahden muuttujan yhteyttä on jo tarkasteltu aiemmin sirontakuvioiden avulla, mutta nyt otetaan käyttöön kuvien lisäksi tilastolliset tunnusluvut.

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. SAS Enterprise Guiden työkaluista löytyy myös tilastollisia menetelmiä, joiden avulla voidaan ottaa huomioon riippuvuudet ja siten analysoida koko aineisto yhdessä. Kyseisiä menetelmiä ei kuitenkaan käsitellä tässä oppaassa. Lisätietoa edistyneempien menetelmien soveltamisesta SAS Enterprise Guidessa löytyy muun muassa kirjoista Helenius & Vahlberg (2008) ja Davis (2007).

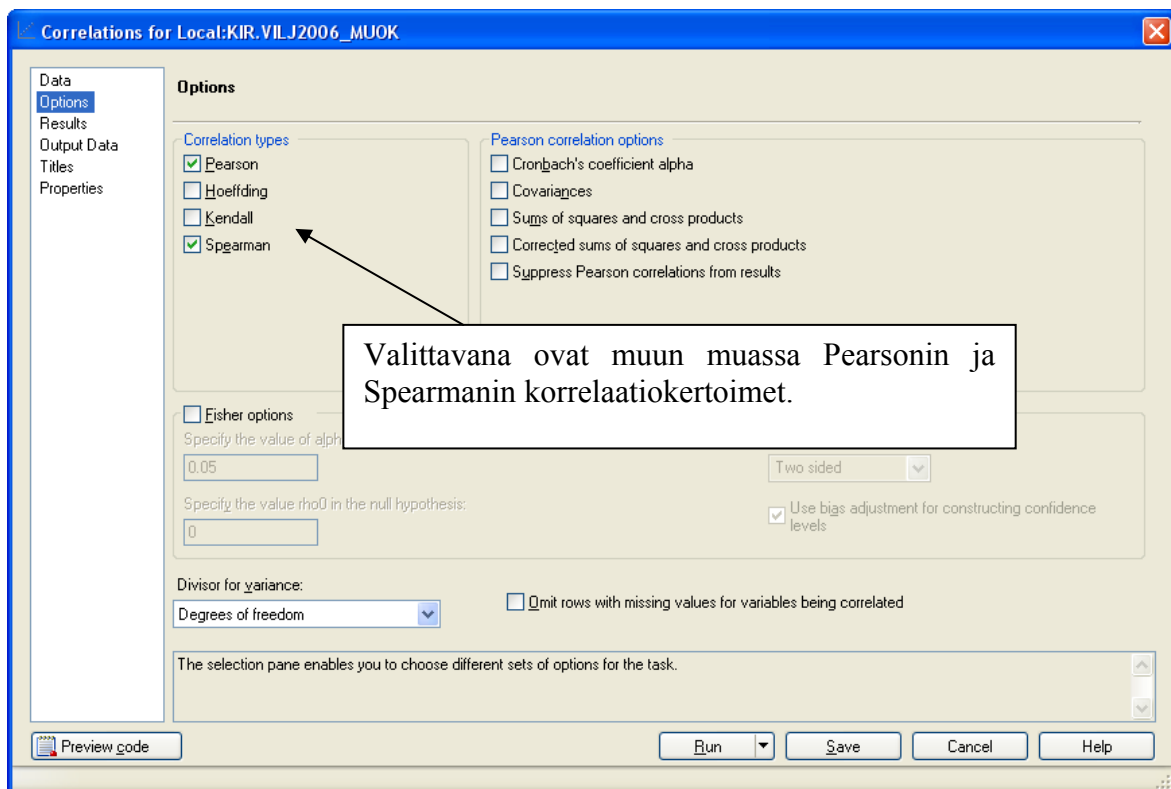
8.5.1 Kahden numeerisen muuttujan välinen yhteys

Kahden numeerisen muuttujan välistä yhteyttä voidaan luonnehtia korrelaatiokerroimen avulla. Tavallinen eli Pearsonin korrelaatiokerroin on kahden muuttujan lineaarisen yhteyden voimakkuuden ja suunnan mitta. Oletuksena Pearsonin korrelaatiokerroimen käytölle on, 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 korrelaatiokerroimet 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 korrelaatiokerroimia käytettäessä on kuvasta syytä tarkastella poikkeavia havaintoja, jotka saattavat vaikuttaa merkittävästi korrelaatiokerroimen suuruuteen.

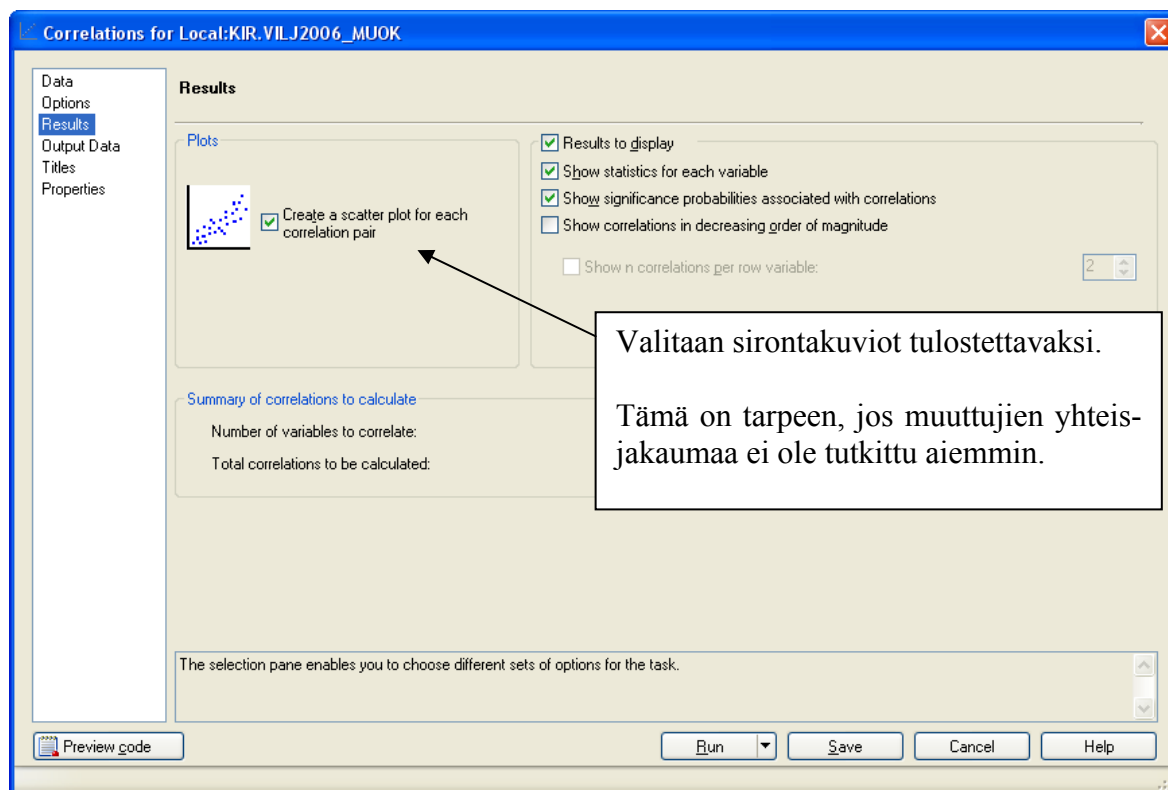
Korrelaatiokertoimet saadaan laskettua Enterprise Guidessa avaamalla käytettävä data työtilaan ja valitsemalla asiayhteysvalikosta ”Analyze” -> ”Multivariate” -> ”Correlations”, jossa ”Data” -välilehdellä määritellään muuttujien tehtävät:



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

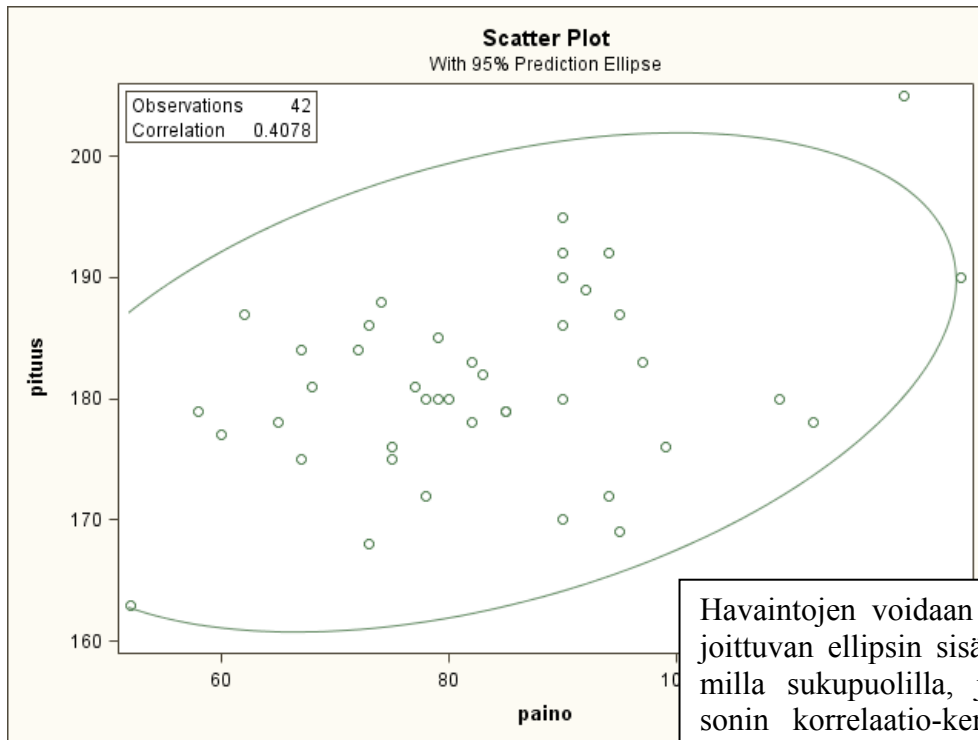


”Results”-välilehdellä voidaan valita muun muassa sirontakuviot mukaan tulostukseen:



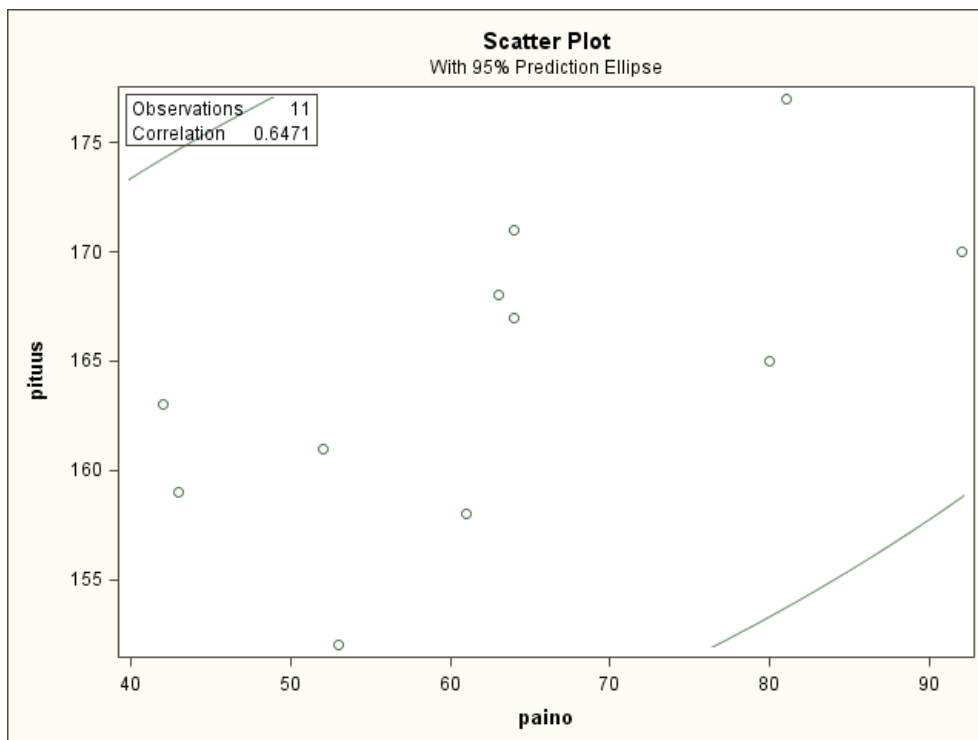
”Output Data” -välilehdellä voidaan valita tulokset tulostettavaksi SAS-dataksi. Kun painetaan lopuksi ”Run”, saadaan tuloksena sukupuolittain seuraavat sirontakuviot (kopioitu leikepöydän kautta Wordiin):

”Miehet”:



Havaintojen voidaan ajatella sijoittuvan ellipsin sisälle molemmilla sukupuolilla, joten Pearsonin korrelaatio-kertoimen aineistoa koskevat oletukset täyttyvät.

”Naiset”:



Enterprise Guide tuottaa myös sirontakuviomatriisin, jota ei tässä esitetä, mutta joka on hyvin käyttökelpoinen, jos tutkitaan samanaikaisesti useamman kuin kahden muuttujan parittaisia yhteyksiä.

Korrelaatiokerroin miehille:

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

Korrelaatiokerroin naisille:

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

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 nollassa oletuksena on, että korrelaatiokerroin on nolla. Pieni p-arvo tarkoittaa, että korrelaatiokerroin poikkeaa tilastollisesti merkittävästi nollassa. 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 selkeää sääntöä. Tulokset vaihtelevat tutkittavan ilmiön mukaan, ja eri asioissa käytetäänkin erisuuruisia 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 & Vahlberg (2008).

Vaihtoehtoinen ratkaisu SAS-koodilla:

Aineisto pitää ensin järjestää BY-muuttujan mukaan:

```
PROC SORT DATA=kir.vilj2006_muok;  
  BY sukupuoli;  
RUN;
```

Korrelaatiokertoimet ja sirontakuviot voi tuottaa CORR-proseduurilla:

```
ODS GRAPHICS ON;  
PROC CORR DATA=kir.vilj2006_muok PLOTS=(SCATTER MATRIX)  
  PEARSON SPEARMAN;  
  BY sukupuoli;  
  VAR paino pituus;  
RUN;  
ODS GRAPHICS OFF;
```

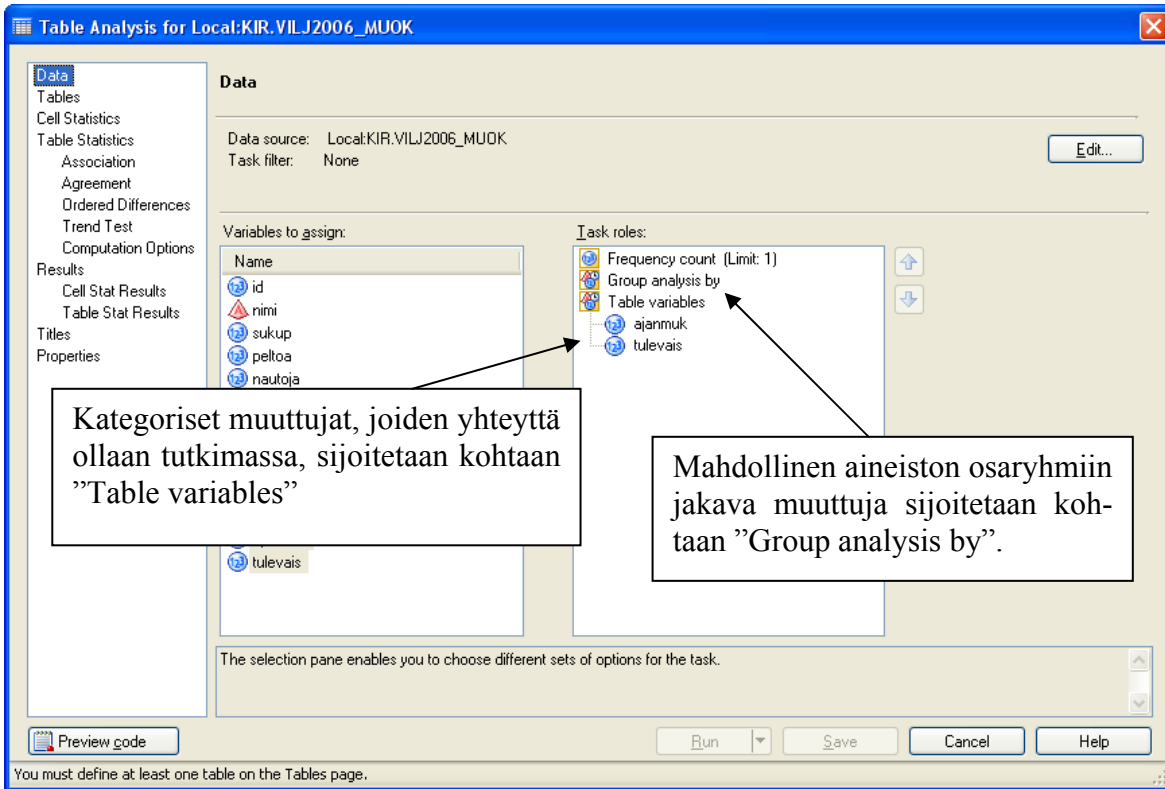
ODS GRAPHICS -lauseet samoin kuin PLOTS-optio liittyvät kuvien tuottamiseen ods-grafiikalla. Jos yllä olevaa koodia käytetään perus-SASissa, tulostus pitää lisäksi ohjata johonkin muuhun tulostuskohteeseen (esimerkiksi ODS RTF -lauseilla) kuin oletuksena olevaan SASin listaustulostuskohteeseen. EG:ssä tulostus tehdään oletuksena SAS Report -tulostusmuodossa, jolloin vastaavaa ongelmaa ei ole. Esimerkki ODS RTF -lauseiden käytöstä löytyy SAS-koodiesimerkistä luvussa **Kategorisen muuttujan jakauman tarkastelu**.

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

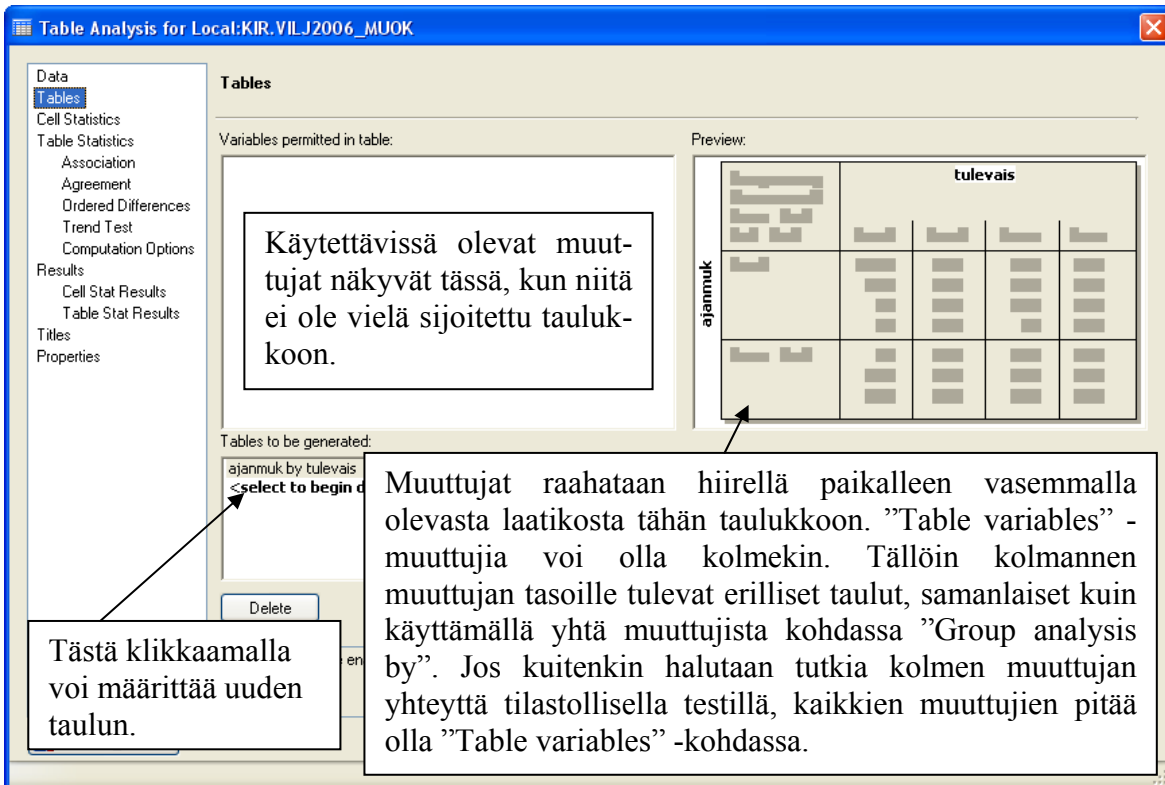
8.5.2 Kahden kategorisen muuttujan välinen yhteys

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.

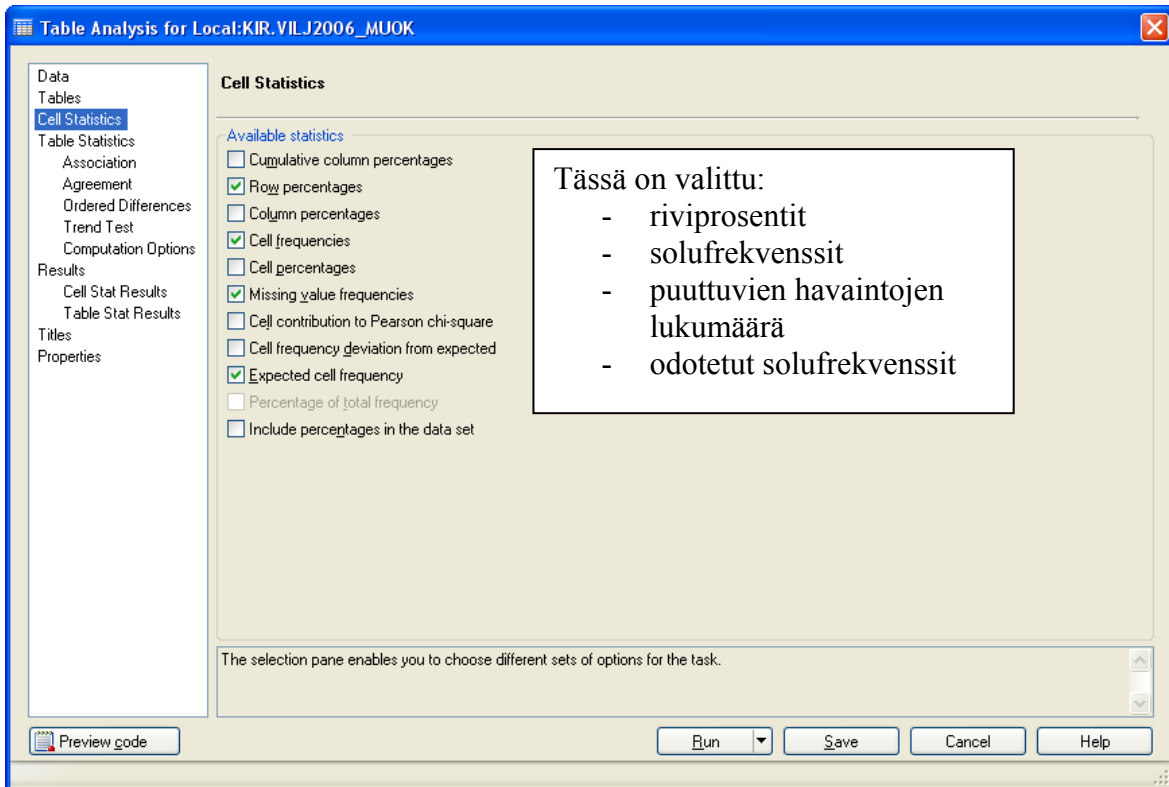
SAS Enterprise Guiden työkalu ristiintaulukointiin ja Khiin neliö -testin tekemiseen löytyy klikkaamalla tarkasteltavaan aineistoon liittyvästä asiayhteysvalikosta ”Describe” -> ”Table Analysis”, jolloin ensin aukeaa ”Data” -välilehti:



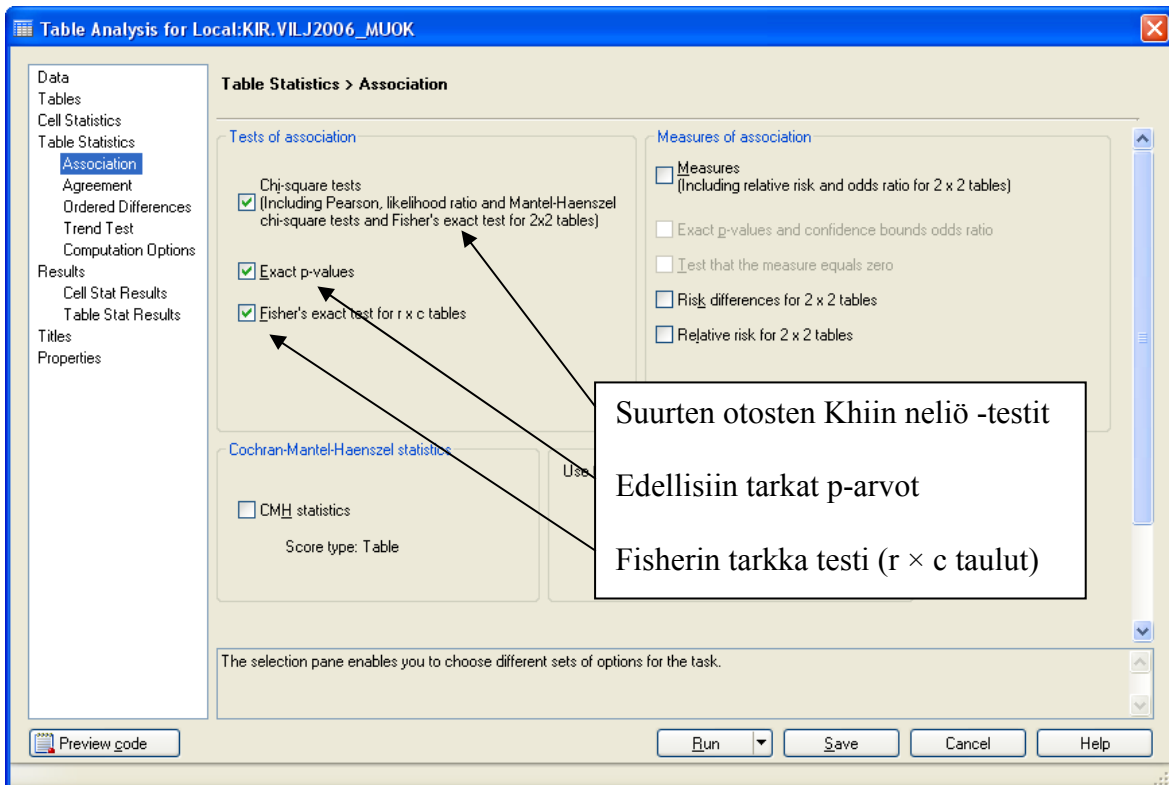
”Tables”-välilehdellä tehdään tarkempi taulukon määrittely:



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



Erilaisia tilastollisia testejä voidaan valita ”Association”-välilehdellä:



”Resultsin” alla olevilla välilehdillä puolestaan voidaan valita, että ristiintaulukko ja tilastollisten testien tulokset tuotetaan SAS-datana. Painamalla ”Run” saadaan tuloste:

Ristiintaulukko

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

Suurten otosten Khiin neliö -testit

	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

Tarkat testit

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

Vaihtoehtoinen ratkaisu SAS-koodilla:

Ristiintaulukko ja siihen liittyvät testit voidaan tuottaa **FREQ**-proseduurilla:

```
PROC FREQ DATA=kir.vilj2006_muok ORDER=INTERNAL;  
  TABLES ajanmuk*tulevais / MISSPRINT EXPECTED CHISQ FISHER;  
  EXACT CHISQ;  
RUN;
```

Optioiden tarkkoihin selityksiin ja muihin käytettävissä oleviin optioihin kannattaa tutustua avaamalla EG:ssä ”Help” -> ”SAS Syntax Help” ja laittamalla Etsi-sivulle hakusanoiksi mielenkiinnon kohteena olevat optiot.

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

9 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 omat rajoituksensa, joten joidenkin tavallisen SASin monipuolisempien ominaisuuksien hyödyntäminen Enterprise Guidessa voi edellyttää koodin käyttöä.

Tässä oppaan Enterprise Guiden versioon 4.2 liittyvässä painoksessa on pyritty lisäämään asiaa liittyen SAS-koodin hyödyntämiseen. Kunkin työkalun yhteydessä on nyt lyhyt kuvaus vaihtoehtoisesta ratkaisusta SAS-koodia käyttäen. Kaikki koodiratkaisut löytyvät myös yhdestä SAS-kooditiedostosta (EG42opas_kaikkikoodit.sas) oppaaseen liittyvästä tiedostokansiosta (<http://www.mtt.fi/mttraportti/pdf/egdata/EG42datat.zip>).

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

9.1 Koodin kirjoittaminen Enterprise Guidessa tai tuominen muualta

Koodin kirjoittaminen onnistuu EG:ssä koodi-ikkunassa. Uuden koodi-ikkunan saa avattua valitsemalla ”File” -> ”New” -> ”Program”. Koodi-ikkunat tulevat projektiin erillisiksi 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”.

Esimerkiksi perus-SASissa tehty valmis kooditiedosto (*.sas) voidaan myös avata Enterprise Guideen. Tämä onnistuu valitsemalla ”File” -> ”Open” -> ”Program” ja etsimällä SAS-kooditiedosto tietokoneella olevasta 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”.

SAS-koodisolmuja käytettäessä kannattaa muistaa myös mahdollisuus lisätä omia linkkejä solmujen väleille prosessidiagrammissa. SAS-koodisolmuja ei nimittäin automaattisesti yhdistetä prosessidiagrammissa niiden käyttämien datojen kuvakkeisiin. Käyttäjän lisäämien linkkien avulla koodisolmut saadaan laajemmissa ajoissa suoritettua oikeassa kohtaa kokonaisuuden suoritusta ja samalla projektikokonaisuuden hahmottaminen prosessidiagrammia tarkastellen helpottuu. Linkittämisestä löytyy lisätietoa luvusta **Projektin pitäminen järjestyksessä**.

9.2 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 prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Open” -> ”Open Last Submitted Code”. Tällöin ruudulle aukeaa koodieditori-ikkuna, 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” -> ”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.

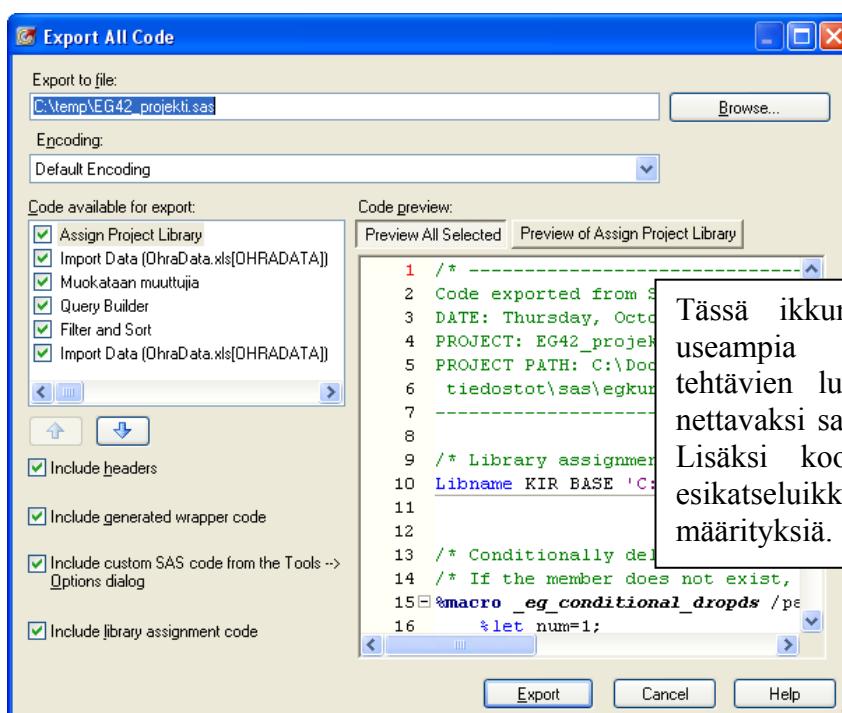
Koodia voidaan myös tarkastella tehtävän määrittelyn ollessa vielä kesken. Useimmissa tehtävämäärittelyikkunoissa löytyy vasemmasta alakulmasta ruutu ”Preview code”. Sitä klikkaamalla saadaan siihen asti tehdyillä valinnoilla generoitu koodi esikatseluikkunaan. Klikkaamalla ”Insert code” pääsee koodin 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, kun halutaan varmistaa, millä menetelmällä EG toteuttaa jonkin toimenpiteen. Koodista voi tarkistaa EG:n käyttämän menetelmän tai valinnan (eli avainsanan koodissa) ja hakea sen tarkemman merkityksen ”Help” -> ”SAS Syntax Help” ohjeen hakutoiminnoilla.

9.3 Koodin tallentaminen SAS-kooditiedostoon

SAS-kooditiedosto voidaan tallentaa tietokoneelle omana tiedostonaan klikkaamalla koodisolmua hiiren oikealla painikkeella ja valitsemalla ”Export” -> ”Export koodisolmun_nimi” tai ”Save koodisolmun_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 projektitiedoston 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 In Process Flow” (yhden Process Flow -sivun koodit) tai ”Export All Code In Project” (koko EG-projektin koodit). Molemmilla valinnoilla aukeaa seuraava ikkuna:

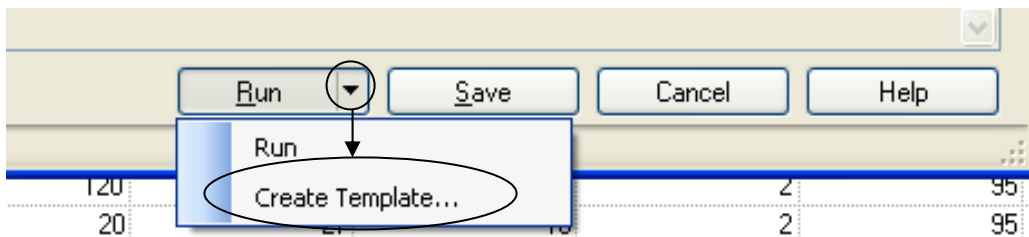


10 Käyttöä helpottavia toimintoja

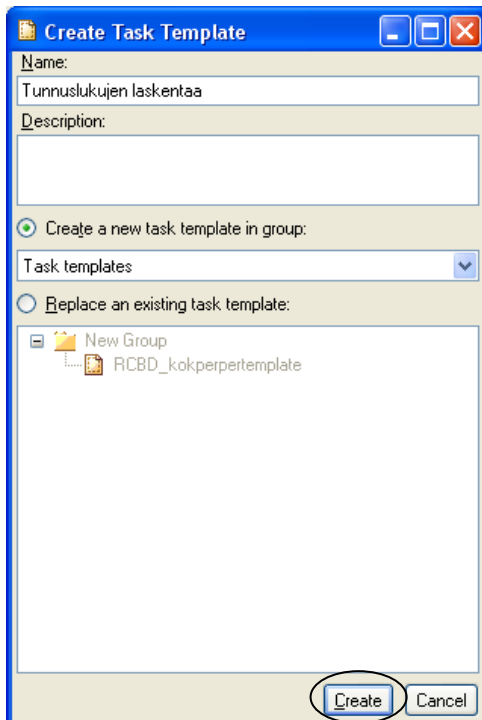
Enterprise Guide 4.2:een on tullut parannuksia ja uusia toimintoja liittyen työn helpottamiseen suoraviivaistamisen ja automatisoinnin kautta. EG 4.2:ssa on muun muassa mahdollista tallentaa valmiiksi määritetty tehtäväsolmu ajopohjaksi myöhempää käyttöä varten. Lisäksi käyttäjäkehoitteita eli promptteja (aikaisemmin parametrit tai makromuuttujat) on monipuolistettu ja nyt myös prosessidiagrammin haaroihin voidaan määrittää ehtolauseita. Näistä toiminnoista on lisätietoa tässä luvussa.

10.1 Työkalussa tehtyjen määritysten tallentaminen ajopohjaksi Create Template -toiminnolla

Enterprise Guide 4.2:ssa on mahdollista tallentaa työkalun sisällä tehdyt määritykset ajopohjaksi ("Template"), jota voidaan jatkossa käyttää vastaavan tehtävän määrittelyn nopeuttamiseksi. Kaikki tehdyt määritykset eivät kuitenkaan tallennu ajopohjaan, sillä esimerkiksi käytettävät muuttujat joudutaan valitsemaan joka kerta uudelleen. Määritykset voidaan tallentaa työkalun "Run"-painikkeen oikeassa reunassa olevan nuolenpäälävinnan kautta. Aukeavasta valikosta valitaan "Create Template":



Aukeavassa ikkunassa määritetään muun muassa ajopohjan nimi:



Kohdassa "Name" annetaan ajopohjan nimi.

Lisäksi ikkunan alaosan valinnoilla voidaan valita mihin ryhmään ajopohja tallennetaan ja tehdäänkö uusi ajopohja vai korvataanko vanha.

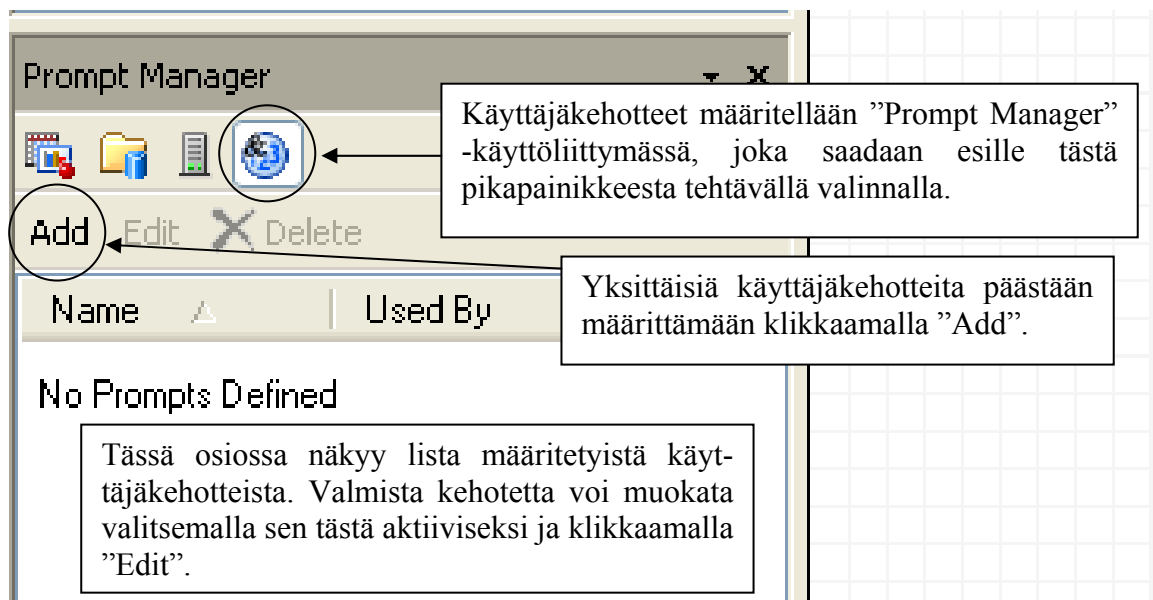
Valitsemalla "Create" luodaan ajopohja. Valmiin ajopohjan voi ottaa käyttöön avaamalla ensin käytettävän aineiston työtilaan ja sitten valitsemalla ylävalikosta "Tasks" -> "Task Templates" ja sieltä haluttu ajopohja.

10.2 Käyttäjäkehotteiden (Prompts) käyttäminen

Käyttäjäkehotteiden eli promptien ja niiden käyttämien makromuuttujien avulla pystytään lisäämään uusia ulottuvuuksia Enterprise Guiden tehtäväsolmujen käyttämiseen. Esimerkiksi osaryhmittäin keskiarvoja laskevassa tehtäväsolmussa voitaisiin asettaa valittaviksi sekä analysoitava muuttuja että ryhmittelevä muuttuja käyttäjäkehotteen kautta, jolloin käyttäjä saisi valita todelliset käytettävät muuttujat vasta tehtäväsolmun suorituksen yhteydessä kehotteeseen liittyvän kysymysikkunan kautta. Tällöin kerran määritellyllä tehtäväsolmulla saataisiin laskettua keskiarvoja käyttäen monia eri muuttujia. Käyttäjän suorittaessa käyttäjäkehotetta hyödyntävän tehtäväsolmun, Enterprise Guideen aukeaa ikkuna, jossa kysytään kehotteen määrittelyn yhteydessä ennakkoon muodostettu kysymys ja annetaan tietyt vastausvaihtoehdot. Valittu arvo sijoitetaan taustalla makromuuttujan tai makromuuttujien arvoiksi ja tehtävä suoritetaan EG:ssä näillä parametreilla. Arvot voivat olla esimerkiksi muuttujien nimiä, lukuarvoja tai merkkijonoja. Käyttäjäkehotteiden avulla voidaan siis rakentaa eräänlaisia aliohjelmia, jolloin yhdellä tehtäväsolmulla saadaan suoritetuksi useita tehtäviä. Niitä voidaan käyttää muun muassa jonkin tehtäväsolmun käyttämien muuttujien määrittelyyn tai Query Builderissä aineiston osa-aineistoiksi jakavan ehdon määrittelyyn. Käyttäjäkehotteisiin viittaava nimi ”Prompt” on uutta Enterprise Guiden versiossa 4.2. Aiemmissa versioissa vastaavaan aihealueeseen viitattiin nimellä ”Parameter” tai ”Macro Variable”. Uusimman version myötä EG:n käyttäjäkehotteissa voidaan käyttää myös lukuisia uusia makromuuttujatyyppejä. Niissä voidaan käsitellä nyt entistä monipuolisemmin muun muassa aikaan ja päivämääriin liittyvää tietoa sekä esimerkiksi väri voidaan antaa arvoksi. Uutta on myös käyttäjäkehotteiden käyttäminen prosessidiagrammin ehdollistamisen yhteydessä (katso lisätietoa luvusta **Prosessidiagrammin haarojen ehdollistaminen (Condition)**). Tarkempaa tietoa käyttäjäkehotteista löytyy klikkaamalla ”Help” -> ”SAS Enterprise Guide Help” -> ”Working with Prompts”.

Käyttäjäkehotteiden määrittely

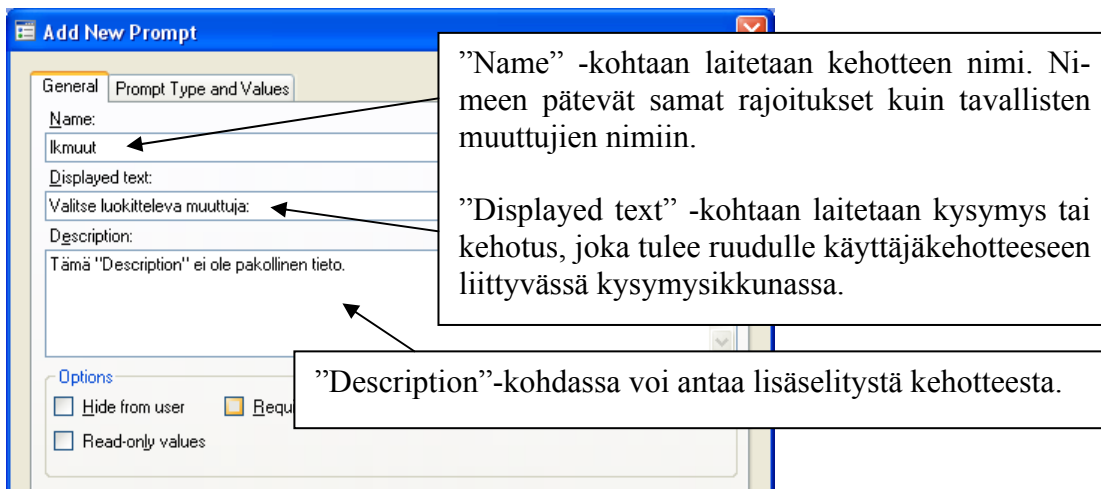
Käyttäjäkehotteita päästään määrittelemään valitsemalla EG:n käyttöliittymän vasemman alakulman ikkunakokonaisuuden neljäs vaihtoehto:



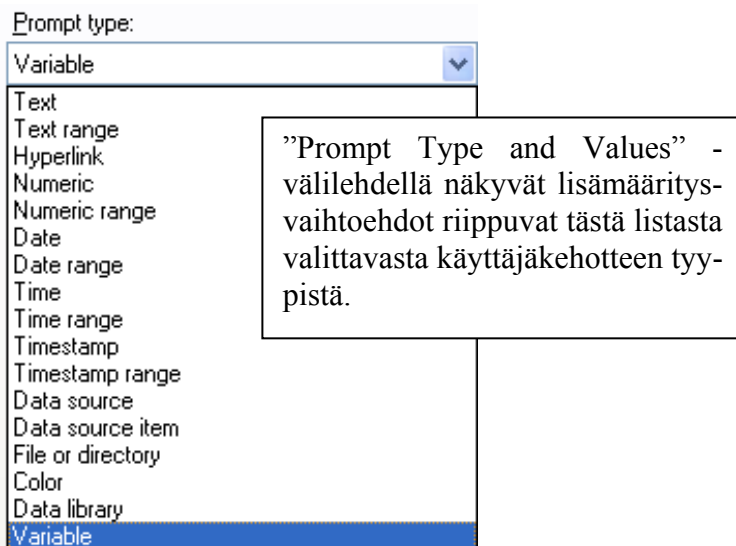
Vaihtoehtoisesti ”Prompt Manager” -käyttöliittymä voidaan tarvittaessa ottaa käyttöön Query Builderin käyttöliittymän ”Prompt Manager” -valinnalla tai SAS-koodisolmuun liittyen klikkaamalla solmua hiiren oikealla painikkeella ja valitsemalla ”Properties” -> ”Prompts” -> ”Prompt Manager”.

Edellisen ikkunan ”Add”-valinnan jälkeen päästään määrittelemään käyttäjäkehotteen tyyppi ja siihen liittyvä kysymys. Seuraavassa esimerkissä määritetään kehoite, joka saa arvoksi muuttujan

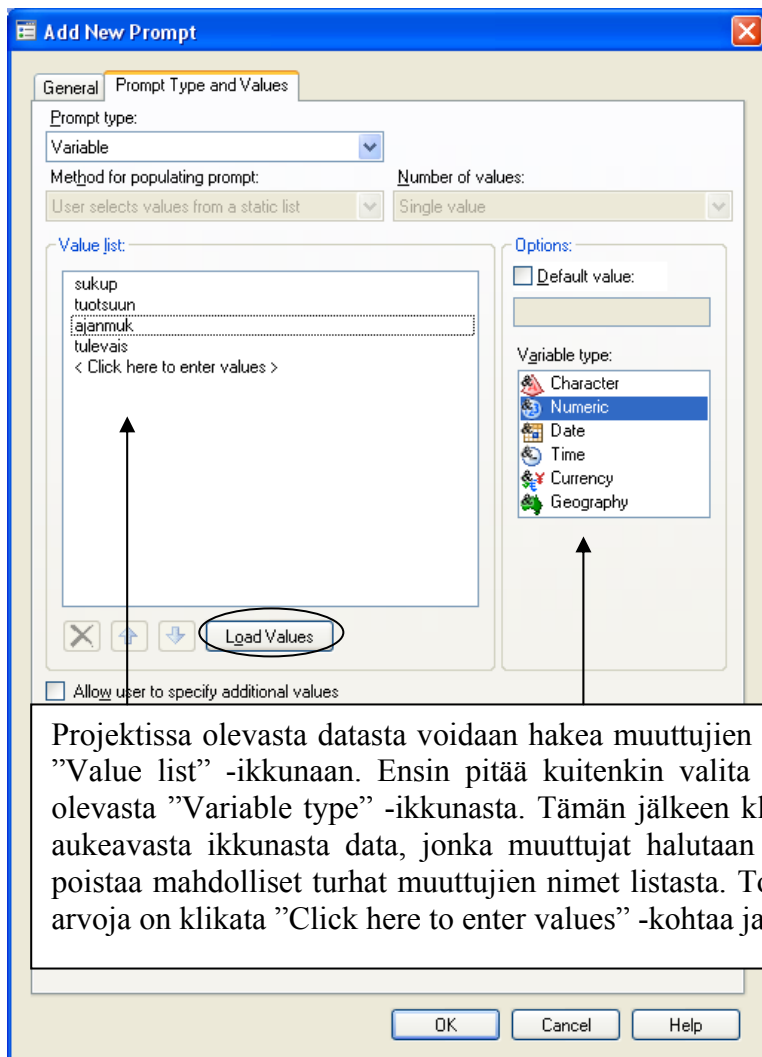
nimen ja sitä tullaan käyttämään jatkossa luokittelevan muuttujan valintaan aineiston tarkastelussa:



"Prompt Type and Values" -välilehdellä tärkein määrittäminen on käyttäjäkehoitteen tyyppi eli käyttäjän antamien parametrien tyyppiä valinta. Valittavana ovat seuraavat vaihtoehdot:



Tässä esimerkissä valitaan "Variable", joka tarkoittaa, että käyttäjä antaa käyttäjäkehoitteen kysymykseen vastaukseksi muuttujan nimen. Tällä valinnalla aukeaa "Prompt Type and Values" -välilehdelle seuraava näkymä:



Projektissa olevasta datasta voidaan hakea muuttujien nimiä mahdollisten arvojen listaan ”Value list” -ikkunaan. Ensin pitää kuitenkin valita haettavien muuttujien tyyppi yllä olevasta ”Variable type” -ikkunasta. Tämän jälkeen klikataan ”Load Values” ja valitaan aukeavasta ikkunasta data, jonka muuttujat halutaan listaan. Tämän jälkeen kannattaa poistaa mahdolliset turhat muuttujien nimet listasta. Toinen vaihtoehto antaa mahdollisia arvoja on klikata ”Click here to enter values” -kohtaa ja syöttää arvot käsin.

Lopuksi painetaan ”OK”, jolloin uuden käyttäjäkehoteen nimi tulee näkyviin ”Prompt Manager” -ikkunaan.

Kun käyttäjäkehote on määriteltä, se on valmis käytettäväksi tehtäväsolmuissa tai SAS-koodissa. Seuraavassa esitellään kuitenkin ensin, miten numeroarvojen määrittely käyttäjäkehoteelle eroaa edellä läpikäydystä muuttujan nimi -arvojen määrittelystä. Liikkeelle lähdetään ihan samalla tavalla eli avataan ”Prompt Managerista” määrittelytyökalu ”Add”-valinnalla ja ryhdytään määrittelemään uutta kehotetta. Tässä esimerkissä luodaan kehoite, jonka avulla käyttäjää voidaan pyytää antamaan vuosituloraja, jonka perusteella valitaan osa-aineisto jatkotarkasteluihin.

Uuden kehotteen määrittelyssä ”General”-välilehdellä toimitaan samoin kuin edellä eli annetaan kehoitteelle nimi ja määritetään kysymysikkunassa esitettävä kysymys:

”Prompt Type and Values” -välilehti näyttää seuraavalta, kun käyttäjäkehotteen tyyppiä valitaan ”Numeric”:

Lopuksi painetaan ”OK” ja käyttäjäkehote on valmis jatkokäyttöön.

Käyttäjakehotteiden käyttäminen

Edellä on käyty läpi muutaman erilaisen käyttäjakehotteen eli promptin muodostaminen. Valmiita kehotteita voidaan käyttää esimerkiksi erilaisissa tulostusta tuottavissa tehtäväsolmuissa muuttujien valinnassa sekä Query Builderin ”Filter Data” -välilehdellä osa-aineiston valintaan liittyvissä ehtolauseissa. Myös SAS-koodissa voidaan käyttää käyttäjakehotteita.

Tässä esitellään ensin esimerkkinä tilanne, jossa valitaan ryhmittelevä muuttuja käyttäjakehotteen avulla. Tätä varten avataan ensin ”Summary Statistics” -työkalu siten, että Viljelijä 2006 -aineisto on käytössä:

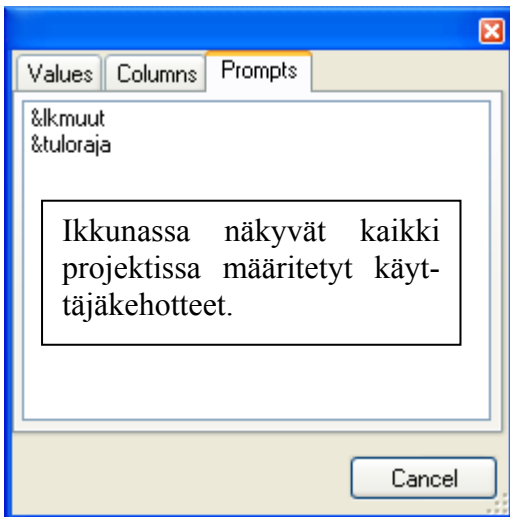
Muuttujalistassa näkyvät tavallisten aineiston muuttujien lisäksi myös käytettävissä olevat käyttäjakehotteet.

Asetetaan analysoitavaksi muuttujaksi tavallinen ika-muuttuja. Sen sijaan ryhmitteleväksi muuttujaksi laitetaan nyt käyttäjakehote ikmuut, jonka mahdolliseksi arvoiksi on edellä määriteltä muutamat ryhmitteleviksi muuttujiksi soveltuvat aineiston muuttujat.

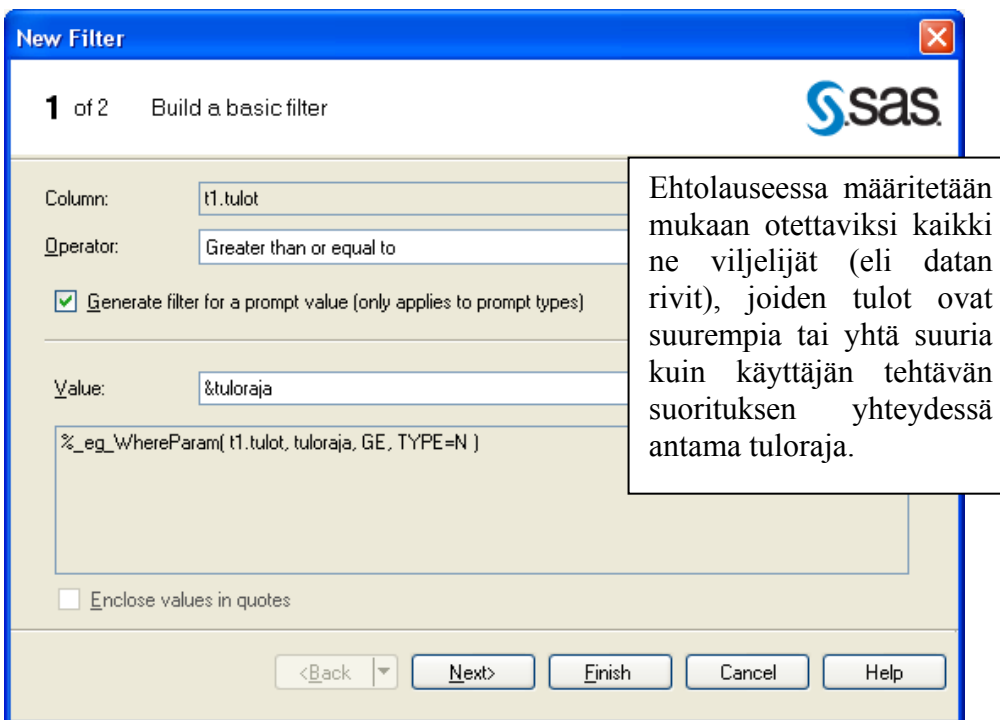
Kun painetaan ”Run”, aukeaa valintaikkuna, jossa käyttäjän pitää valita millä kehotteen arvolla ajo suoritetaan:

Valitaan esimerkiksi sukupuoli ja painetaan ”Run”, jolloin EG laskee ika-muuttujan tunnusluvut sukupuolittain. Käyttäjakehote hyödyntävä tehtäväsolmu voidaan suorittaa myöhemmin uudelleen, ja kehotteelle voidaan valita jokaisella kerralla eri arvo.

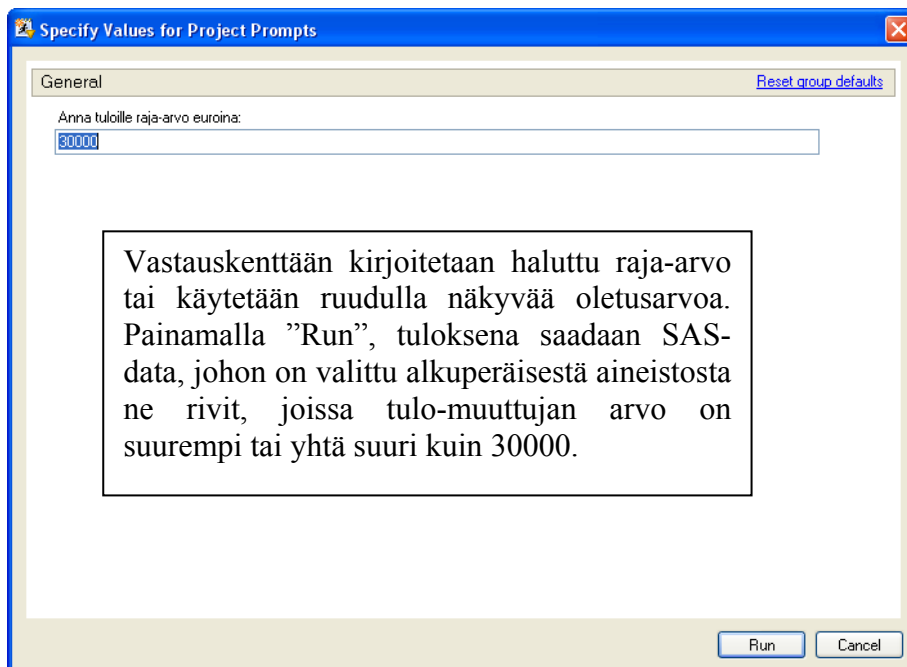
Toinen esimerkkutilanne, jossa käyttäjäkehoteet ovat hyödynnettävissä, liittyy osa-aineiston valintaan. Avataan ensin Query Builder siten, että käytössä on aiemmassakin esimerkissä käytetty Viljelijä 2006 -aineisto ja mennään "Filter Data" -välilehdelle. Raahataan "Filter Data" -välilehdelle muuttuja, jonka arvoihin liittyvä käyttäjäkehote on aiemmin määritelty (tässä esimerkissä "tulot"). 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 "Prompts"-välilehdellä voidaan määrittää käyttäjäkehote käytettäväksi ehtolauseessa:



Kun listasta valitaan käyttäjäkehote, EG palaa edelliseen ikkunaan, jossa näkyy käyttäjäkehoteen avulla määritetty ehtolause:



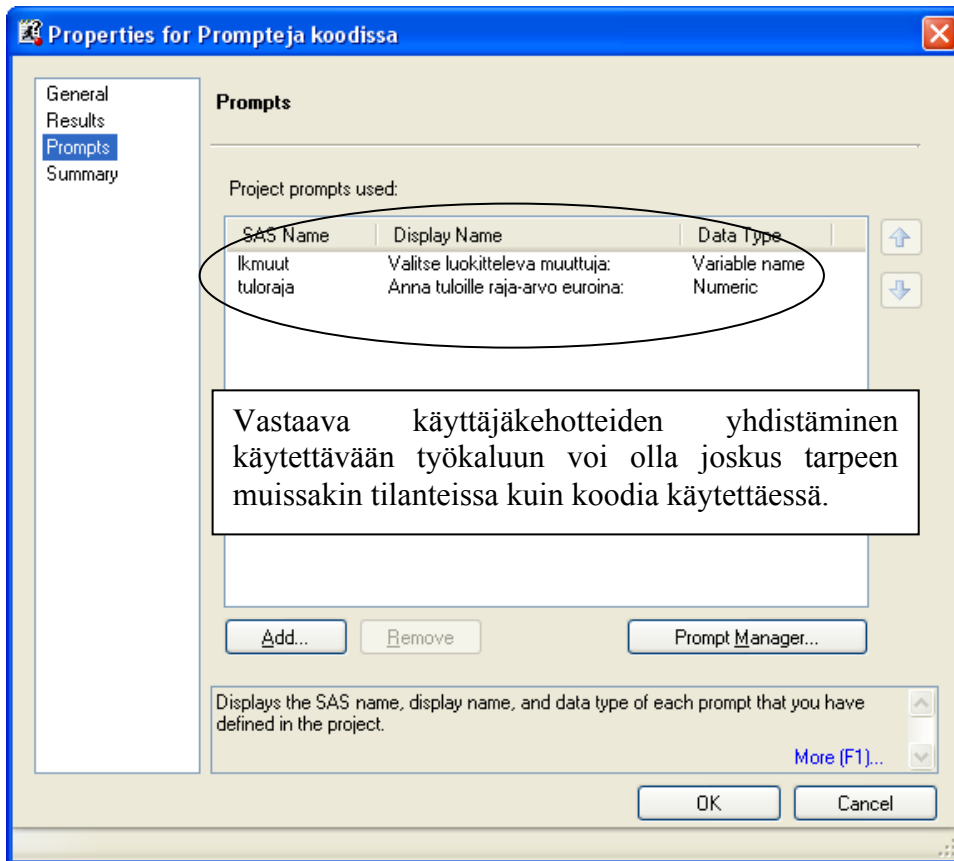
Lopuksi painetaan ”Finish” ja suoritetaan Query Builder normaalisti. Suorituksen yhteydessä aukeaa seuraava ikkuna:



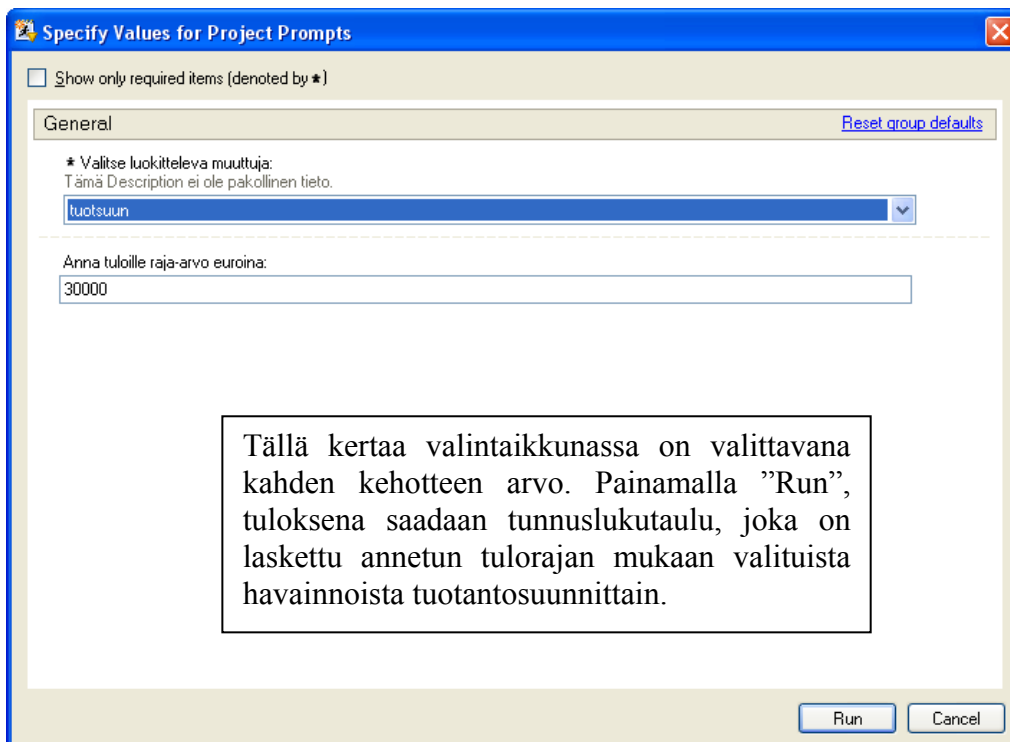
SAS-koodia käytettäessä käyttäjäkehoitteisiin liittyviin makromuuttujiin voidaan viitata niiden nimellä sillä lisäyksellä, että nimen eteen tulee makromuuttujiin viittaava &-merkki. Samassa tehtävässä tai SAS-koodissa voi myös käyttää useammasta kehoitteesta peräisin olevia makromuuttujia. Alla on esimerkki SAS-koodista, jossa on käytetty molempia edellä määriteltyjä kehoitteita:

```
Program Log Results
Save Run Stop Select Server Export Send To
1 PROC MEANS DATA=kir.vilj2006_muok;
2   WHERE tulot>=&tuloraja;
3   CLASS &lkmuut;
4   VAR ika;
5 RUN;
```

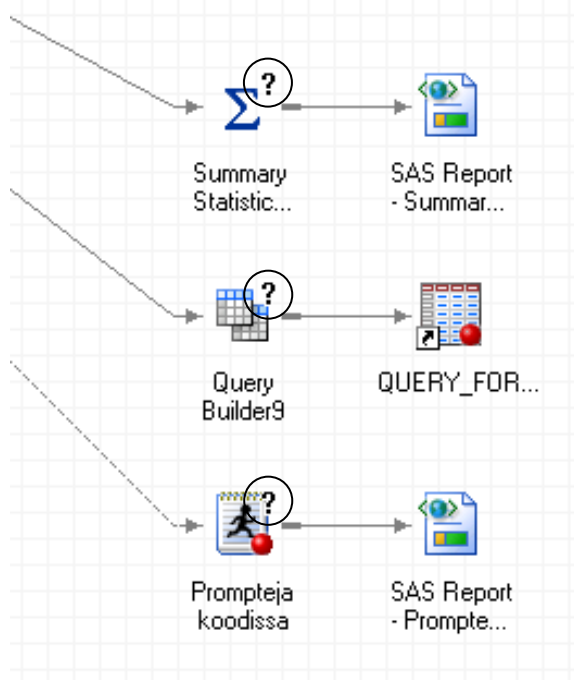
Ennen koodin suorittamista käyttäjäkehoitteet pitää vielä erikseen yhdistää kyseiseen koodiin, koska ne on määritelty ”Prompt Managerissa” eikä tässä koodissa. Se onnistuu klikkaamalla koodisolmuun liittyvästä asiayhteysvalikosta ”Properties” tai klikkaamalla koodisolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Properties”. Tämän jälkeen mennään ”Prompts”-välilehdelle, klikataan ”Add” ja valitaan koodissa käytettävät kehoitteet. Kun muuttujat on valittu, ne näkyvät seuraavassa ikkunassa:



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



Prosessidiagrammissa käyttäjেকেhotteita hyödyntävät tehtäväsolmut erottuvat pienestä kysymysmerkkisymbolista kuvakkeen oikeassa yläkulmassa:



10.3 Prosessidiagrammin haarojen ehdollistaminen (Condition)

Enterprise Guide 4.2:ssa projektiin saadaan joustavuutta myös käyttämällä prosessidiagrammin haarojen ehdollistamista (condition). Tämän toiminnon avulla voidaan esimerkiksi määrittää prosessidiagrammin suoritus siten, että se ohjautuu tietyn ehdon mukaan kahdesta haarasta vain jompaankumpaan. Ehto voi ratketa esimerkiksi aineiston, ajan/päivämäärän tai käyttäjältä kysyttävän kysymyksen perusteella. Seuraavassa esimerkissä prosessidiagrammin suoritus ohjautuu jompaankumpaan haaraan käyttäjältä saadun vastauksen perusteella.

Ensin määritellään tekstiarvoinen käyttäjäkehote, jota hyödynnetään prosessidiagrammin ehdollistamisessa (käyttäjäkehotteet esitellään tarkemmin luvussa **Käyttäjähötteiden (Prompts) käyttäminen**):

"General"-välilehdellä on määritelty käyttäjäkehotteen nimi ja kysymys, johon käyttäjä voi vastata jonkin tässä määritellyssä listassa olevista tekstimuotoisista vastausvaihtoehdoista.

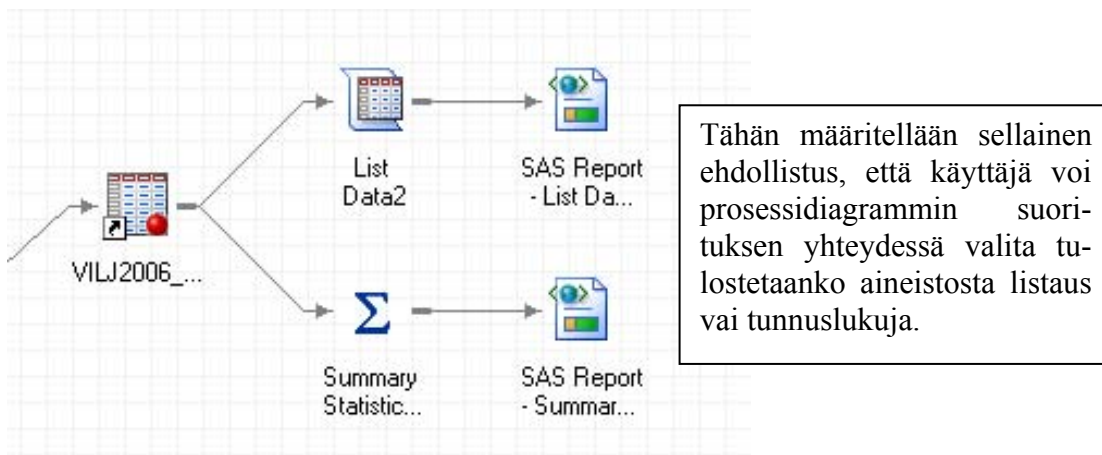
Staattinen lista eli tässä yhteydestä määriteltävä lista on valittuna.

Tässä on määritelty kaksi vastausvaihtoehtoa: "Listaus" tai "Tunnusluvut". Listaus on valittu oletukseksi.

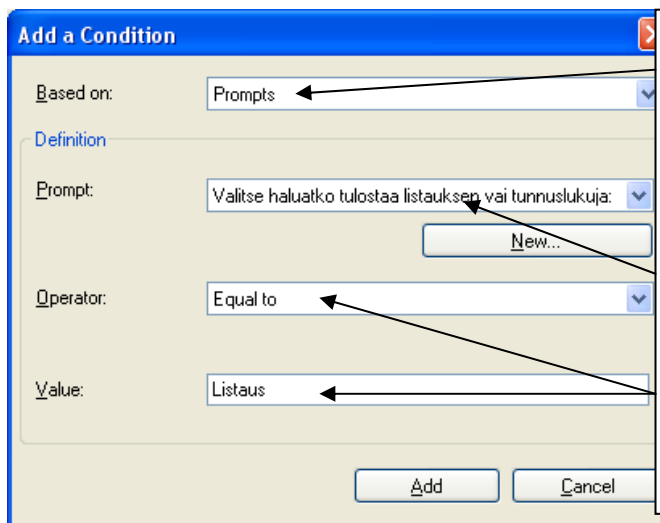
Valitsemalla "OK" saadaan käyttäjäkehote valmiiksi.

Unformatted Value	Formatted [Displayed] Value	Default
Listaus	Listaus	<input checked="" type="radio"/>
Tunnusluvut	Tunnusluvut	<input type="radio"/>

Yksirivisen tekstimuotoisen käyttäjäkehotteen eli promptin lisäksi voidaan prosessidiagrammin ehtolauseiden määrittelyssä hyödyntää joko numeerisia, päivämäärään tai kellonaikaan liittyviä prompteja. Nyt kun käyttäjäkehote on määritelty, voidaan määritellä varsinainen prosessidiagrammin ehdollistus. Alla näkyy osa prosessidiagrammia, johon määritellään edellä muodostetun käyttäjäkehotteen avulla tehtävä ehdollistus:



Ehdollistus päästään määrittämään klikkaamalla ehdollistettavaa tehtäväsolmua (tässä "List Data2") hiiren oikealla painikkeella ja valitsemalla "Condition" -> "Add" ja aukeavasta ikkunasta "Add", jolloin voidaan tehdä seuraavat määriytykset:

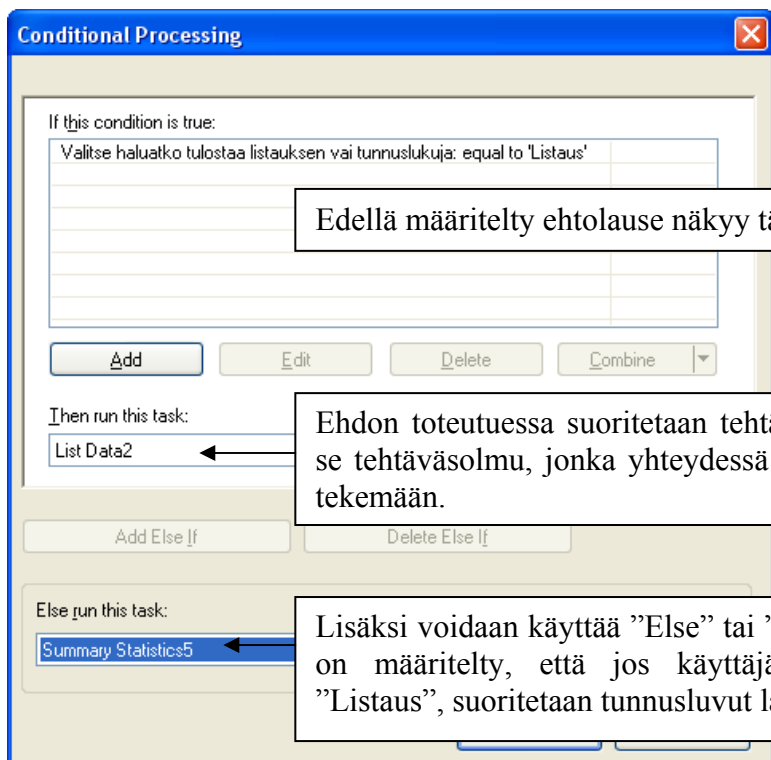


"Based on" -kohdassa ehtolauseiden on valittu perustuvan käyttäjäkehoteisiin ("Prompts"). Muut vaihtoehdot ovat käytettävään aineistoon tai ajankohtaan perustuvat ehdot.

"Prompt" -kohdassa valitaan tässä käytettävä käyttäjäkehote.

Kahdessa viimeisessä kohdassa määritetään, että jos käyttäjän antama arvo on yhtä kuin "Listaus" niin tämä tehtäväsolmu suoritetaan.

Valitsemalla "Add" päästään takaisin edelliseen ikkunaan:



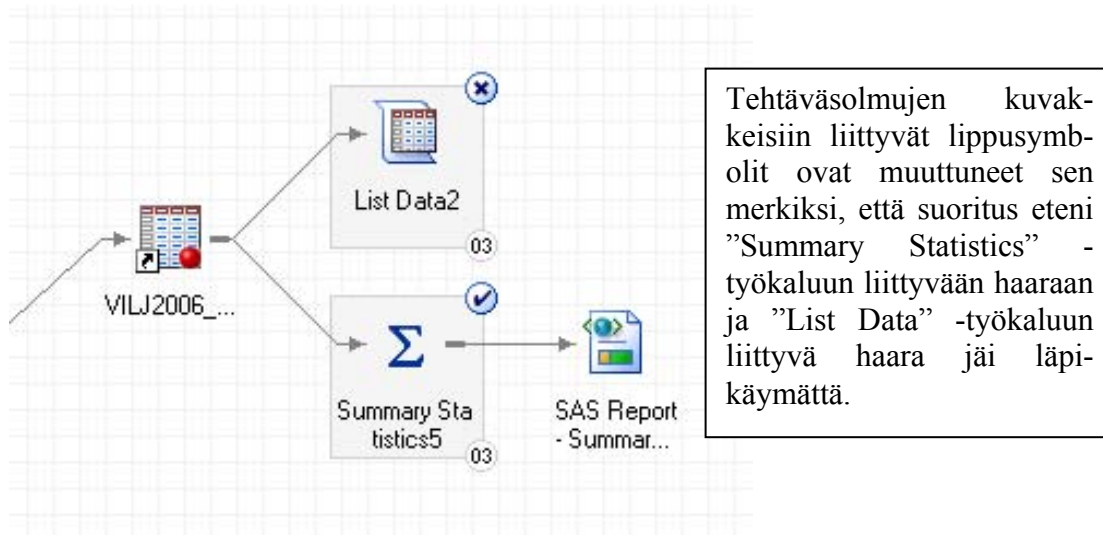
Edellä määritelty ehtolause näkyy tässä ikkunassa.

Ehdon toteutuessa suoritetaan tehtäväsolmu "List Data2" eli se tehtäväsolmu, jonka yhteydessä tätä määrittystä lähdettiin tekemään.

Lisäksi voidaan käyttää "Else" tai "Else If" -rakenteita. Tässä on määritelty, että jos käyttäjä ei valitse vaihtoehtoa "Listaus", suoritetaan tunnusluvut laskeva tehtäväsolmu.

Valitsemalla "OK", saadaan prosessidiagrammin ehdollistus määriteltyä ja sen merkiksi ehdollistettujen tehtäväsolmujen kuvakkeisiin tulee pieni lippusymboli. Prosessidiagrammin haara voidaan nyt suorittaa valitsemalla ehdollistettuja tehtäväsolmuja edeltävä aineisto aktiiviseksi ja valitsemalla asiayhteysvalikosta "Run" -> "Run Branch From...". Tällä valinnalla aukeaa käyttäjäkehoteeseen liittyvä valintaikkuna ja esimerkiksi valinnalla "Tunnusluvut" ajo ohjautuu ainoastaan prosessidiagrammin alempaan tunnuslukujen laskentaan liittyvään haaraan ja tulokseksi saadaan siihen liittyvä tulostus.

Ajon jälkeen prosessidiagrammi näyttää seuraavalta:



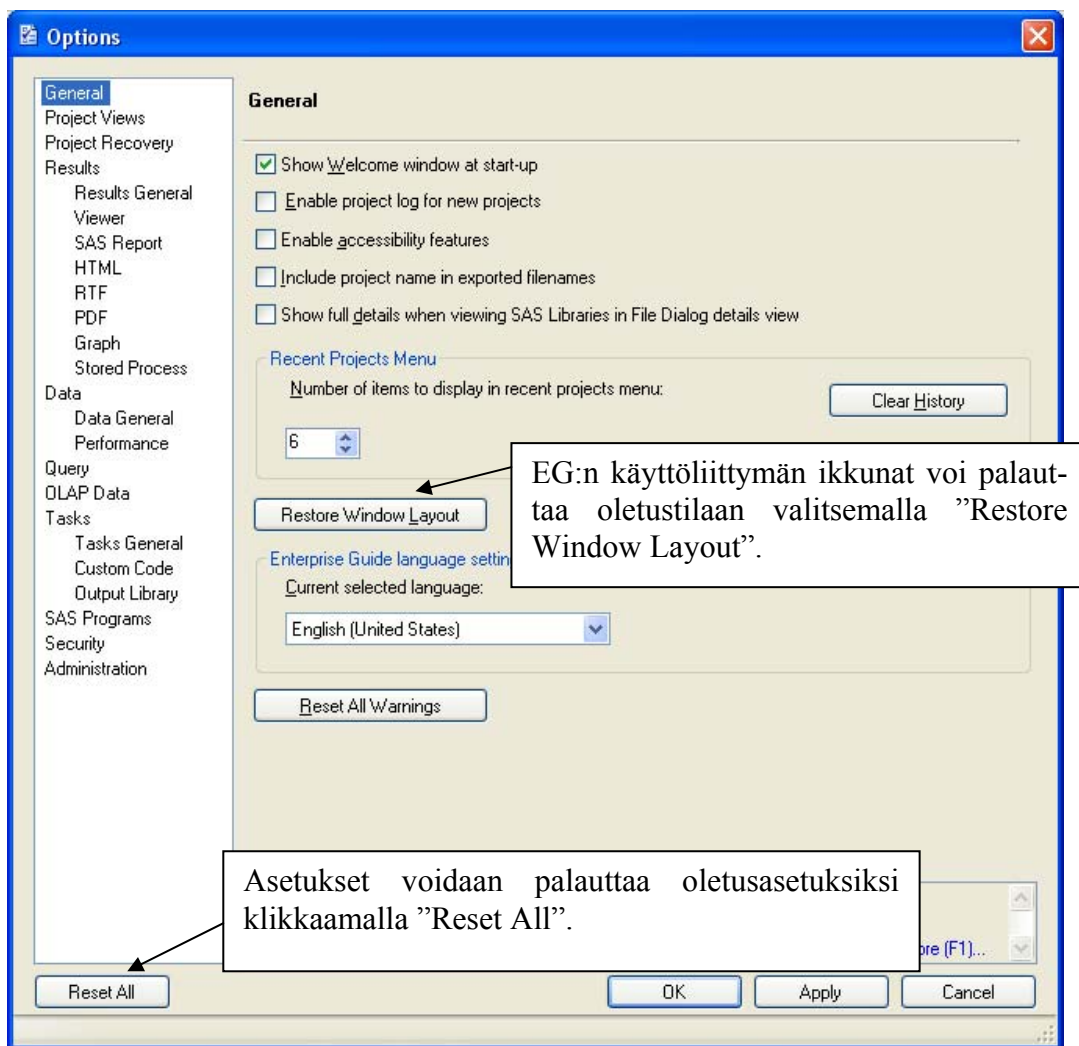
Tehtäväsolmujen kuvakeisiin liittyvät lippusymbolit ovat muuttuneet sen merkiksi, että suoritus eteni "Summary Statistics" -työkaluun liittyvään haaraan ja "List Data" -työkaluun liittyvä haara jäi läpikäymättä.

11 Tulostuksen ja projektin hallintaa

Tässä kappaleessa on esitetty yleisiä asioita, jotka pätevät Enterprise Guideen ja sen moniin eri työkaluihin. Esitystapa on hieman erilainen kuin aiemmissa kappaleissa. Asiat on pyritty ilmaistamaan 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”:



11.1.1 Tulostuskohteen valinta

Tulostuskohteella tarkoitetaan tulostedokumentin tiedostomuotoa. Tuloste voidaan tehdä SAS Report-, HTML-, RTF-, PDF- ja Text-muodossa. SAS Report on Enterprise Guiden oma tiedostomuoto ja se on EG:ssä oletuksena oleva tulostuskohte. Yksittäisistä SAS Report -tulosteista voidaan koota niin sanottu SAS-raportti, johon on koottu useita tulosteita (”File” -> ”New” -> ”Report”). 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). RTF:llä tehtyä tulostetta on melko kätevä muokata Wordissa. PDF-tulostetta sen si-

jaan 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 tuloste 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 tulostuskohteen 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)
- PNG

Yleensä kannattaa käyttää oletuksena olevaa ActiveX-kuvatyyppeä, jolloin käytössä ovat monipuoliset kuvan muokkausominaisuudet. Kaikkien kuvatyyppeiden ominaisuuksista saa kätevästi lisätietoa painamalla "F1"-näppäintä "Tools" -> "Options" -> "Graph" -ikkunan ollessa auki.

11.1.2 Tulostustyylin valinta

Tulostustyyllillä tarkoitetaan tulostetaulukoiden ja taustan graafista ulkoasua. Tulostustyyllillä voidaan vaikuttaa muun muassa 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 voi luoda kuvia tai taulukoita, jotka ovat ulkoasun puolesta sellaisenaan kelvollisia esimerkiksi joihinkin julkaisuihin.

Oletuksena olevan tulostustyylin voi vaihtaa toiseksi valitsemalla **"Tools"** -> **"Options"** ja käytössä olevan tulostuskohteen mukaan **"SAS Report"**, **"HTML"**-, **"RTF"**- tai **"PDF"**-välilehden. "Text"-tulostuskohteen tyyliä sen sijaan ei pysty muokkaamaan. Kaikista näistä välilehdistä löytyy kohta "Style", josta tulostustyylin voi vaihtaa. Hyviä vaihtoehtoisia tulostustyyliä (yksinkertaisia) ovat ainakin "printer", "journal" ja "minimal". Muun muassa tehtäessä tulostusta paperille on hyvä käyttää sellaista tulostustyyliä, joka ei käytä taustavärejä, koska esimerkiksi kuvien yhteydessä haaleat taustavärit saattavat näyttää paperilla epäsiisteiltä.

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

Style Manager

Style List:

Style	Location	URL
42testi	My Style	C:\Docu
AMODefault	Built-in Style	C:\Prog
Analysis	Built-in Style	C:\Pro
Astronomy	Built-in Style	C:\Prog
Banker	Built-in Style	C:\Prog
BarrettsBlue	Built-in Style	C:\Prog
beige	Built-in Style	C:\Prog
blockPrint	Built-in Style	C:\Prog
brick	Built-in Style	C:\Prog
brown	Built-in Style	C:\Prog
Curve	Built-in Style	C:\Prog
d3d	Built-in Style	C:\Prog
default	Built-in Style	C:\Prog
Education	Built-in Style	C:\Prog
EGDefault	Built-in Style	C:\Prog
EGDefault82	Built-in Style	C:\Prog
Electronics	Built-in Style	C:\Prog

Preview of Analysis:

SAS System Title

SAS Procedure Title

Column 1	Column 2	Column 3
Row 1	Data (Num)	Data (Char)
Row 2	Data (Num)	Data (Char)

GRAPH RESULTS

SAS System Footnote

Buttons: Set as Default, Edit..., Add...

Callout 1: Täällä voi myös valita oletustyylin painamalla ”Set as Default”.

Callout 2: Uuden tyylin pääsee luomaan valitsemalla ”Add”, jolloin ohjelma kysyy, luodaanko kokonaan uusi tyyli vai käytetäänkö olemassa olevaa tyyliä pohjana.

Callout 3: Painamalla ”Edit” pääsee muokkaamaan valittua omaa tyyliä (valmiita tyyliä ei pysty muokkaamaan).

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

Style Editor

Click to select an element:

SAS System Title

SAS Procedure Title

Column 1	Column 2	Column 3
Row 1	Data (Num)	Data (Char)
Row 2	Data (Num)	Data (Char)

GRAPH RESULTS

SAS System Footnote

Attributes for Body:

Text Borders Images Custom

Selected fonts: arial, palatino, times roman

Text size: 10pt

Text style: Regular

Text color: [Color selection]

Background color: Color: [Color selection] Transparent Background

Horizontal alignment: Default

Vertical alignment: Default

Buttons: Apply to Other Elements..., Undo Apply

Preview of Body: Body

Selected element: Body

Buttons: Add..., Delete

Callout: Täällä voi muokata tyyli pohjan ominaisuuksia, kuten fonttia, fonttikokoa, tekstityyliä, värejä, yms.

11.1.3 Muut tulostusasetukset

Tulosteeseen 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.

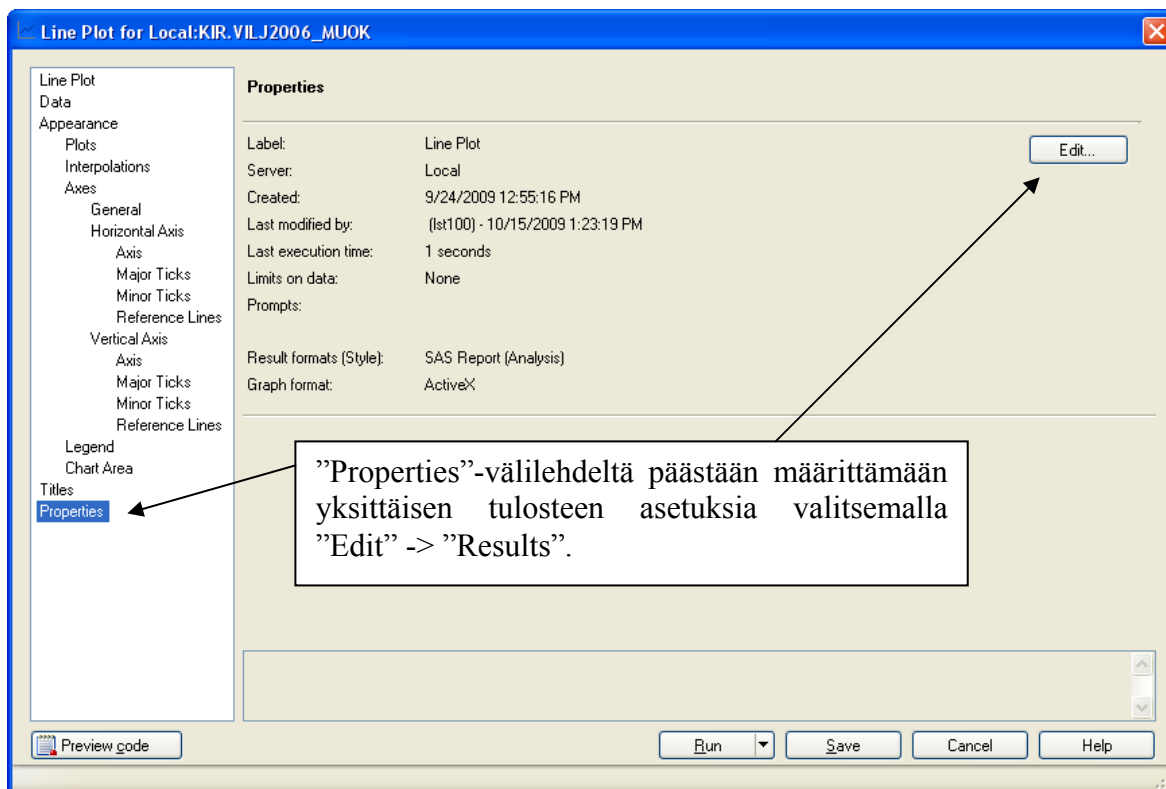
Työkalujen käyttämien SAS-proseduurien nimet saa poistettua tulostuksesta. Tämä onnistuu myös ”Tools” -> ”Options” -> ”Tasks” -> ”Tasks General” -välilehdellä, kun poistetaan merkintä kohdasta ”Include SAS procedure titles in results”.

11.2 Yksittäisen tulosteen hallinta ja jatkokäyttö

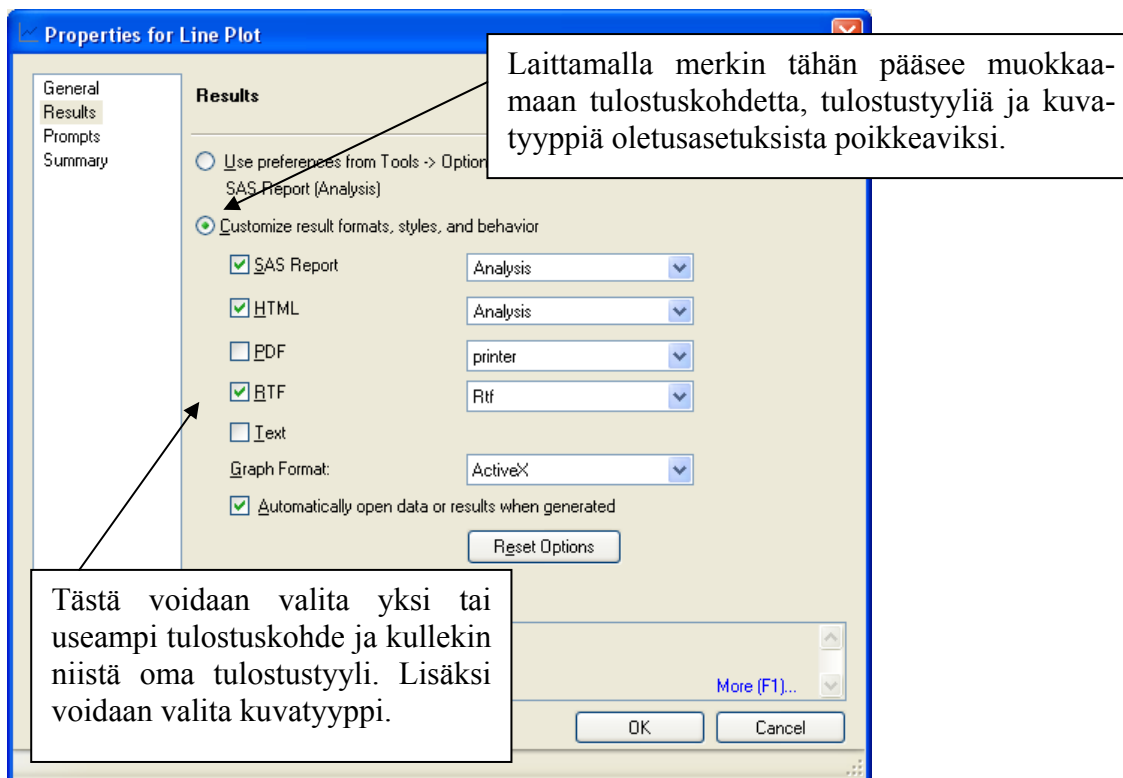
Enterprise Guidessa voidaan muuttaa yleisasetusten lisäksi monia yksittäisen tulosteen asetuksia. Tulosteita voidaan helposti myös muokata ja siirtää toisiin ohjelmiin. Seuraavassa tarkastellaan näitä asioita lähemmin.

11.2.1 Kohteen ja tyylin vaihtaminen yksittäiselle tulosteelle

Useimmissa Enterprise Guiden työkaluissa kyseisen yksittäisen tulostuksen kohde ja tyyli voidaan vaihtaa oletusasetuksista poikkeavaksi ”Properties”-välilehdeltä. Tässä esimerkkinä ”Line Plot” -työkalun ”Properties”-välilehti:



Kun valitaan ”Edit” -> ”Results”, aukeaa seuraava ikkuna:



Tämän jälkeen voidaan palata normaaliin tehtävän määrittelyyn painamalla ”OK” ja tulostus ohjautuu valittuun kohteeseen valitulla tyyllillä, kun tehtävä suoritetaan normaalisti.

Yksittäiselle tulosteelle voidaan vaihtaa tulostuskohdetta ja tulostustyyliä myös kun tulostuksen tekevä tehtäväsolmu on jo olemassa prosessidiagrammissa. Asetuksia päästään muokkaamaan painamalla tulostuksen tuottavaa tehtäväsolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Properties” ja sieltä ”Results”. Tällöin aukeaa samanlainen ikkuna kuin edellä. Tässäkin tapauksessa määrittelyn jälkeen painetaan ”OK”, jolloin palataan prosessidiagramminäkymään. Tällöin tehtäväsolmu pitää vielä suorittaa uudelleen, jotta muutokset menevät tulosteeseen asti. Tämä onnistuu esimerkiksi klikkaamalla tehtäväsolmua prosessidiagrammissa hiiren oikealla painikkeella ja valitsemalla ”Run”.

Yksittäiselle tulosteelle voidaan vaihtaa tyyliä myös asiayhteysvalikon ”Properties”-painikkeen kautta tulosteen ollessa auki työtilassa (SAS Report tai HTML-tulostuskohteita käytettäessä). Tämä ”Properties”-valinta ei johda samaan ikkunaan kuin hiiren oikealla painikkeella prosessidiagrammista tehtävä valinta edellä, mutta sen sijaan se on kätevä ja nopea tapa vaihtaa tulosteen tyyliä ja tutkia tulosteen yleistietoja.

11.2.2 Yksittäisen tulosteen tallentaminen tiedostoon

Yksittäinen työtilassa auki oleva tuloste voidaan tallentaa tiedostoon klikkaamalla asiayhteysvalikosta ”Export” -> ”Export *tulostuskohde – työkalun nimi*”. Tällöin aukeaa ikkuna, josta voidaan valita tallennettavalle tulostetiedostolle tallennuskansio ja nimi. Eri tulostuskohteet tallentuvat pääasiassa seuraavan tyyppisinä tiedostoina:

SAS Report ->	SAS Report -tiedosto (*.srx), HTML-tiedostot (*.htm tai *.html), XML-tiedosto (*.xml) tai Portable Document Format -tiedosto (*.pdf)
HTML ->	HTML-tiedostot (*.htm tai *.html)
RTF ->	Rich Text Format -tiedosto (*.rtf)
PDF ->	Portable Document Format -tiedosto (*.pdf)
Text ->	List-tiedosto (*.lst) tai tekstityiedosto (*.txt)

”Export”-kohdasta löytyy myös vaihtoehto ”Export *tulostuskohde – työkalun nimi* As A Step In Project”. Tämän vaihtoehdon avulla tulosteen tallentaminen tiedostoksi saadaan pysyväksi vaiheeksi prosessidiagrammiin, jolloin se voidaan milloin tahansa tehdä helposti uudelleen suorittamalla kyseinen ”Export File” -tehtäväsolmu. Enterprise Guide 4.2:n uutena ominaisuutena ”Export As A Step In Project” voidaan ”Overwrite existing output” -valinnasta riippuen määrittää korvaamaan aiempi tuloste tai tuottamaan kunkin tulosteen päivämäärä- ja aikaleimalla, jolloin aiempi tuloste ei korvau.

Kuvan voi tallentaa joko edellä mainitulla tavalla koko tulosteen mukana tai HTML-tulostuskohdetta käytettäessä 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” tai ”Tallenna kuva nimellä” ja aukeavasta ikkunasta tallennuskansio ja annetaan kuvatiedostolle nimi. Käytävissä 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 tuloste toiseen ohjelmaan (esim. Wordiin), on avata EG:ssä tallennettu tiedosto (esimerkiksi *.rtf) kyseisessä ohjelmassa. Kuvatiedostoina tallennetut kuvat voidaan puolestaan lisätä Wordiin käyttäen Wordin toimintoa ”Lisää” -> ”Kuva” -> ”Tiedostosta”. Toisaalta kaikki Enterprise Guiden tulostus voidaan myös ohjata kokonaan toiseen ohjelmaan (”Tools” -> ”Options” -> ”Results” -> ”Viewer”).

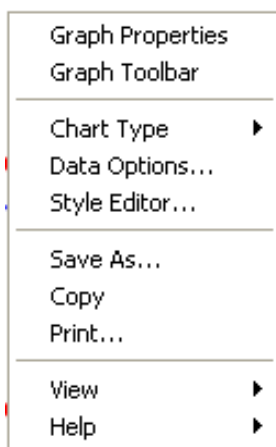
Nopein tapa siirtää dataa tai tulosteita toiseen ohjelmaan on klikata dataan tai tulosteeseen liittyvästä asiayhteysvalikosta ”Send To”, jonka jälkeen valitaan kohde. Yksi vaihtoehto on ”E-mail Recipient”, jolloin tuloste tai data menee suoraan uuden sähköpostiviestin liitetiedostoksi. Tämä toimii ainakin Mozilla Thunderbirdin kanssa. ”Microsoft Word” ja ”Microsoft Excel” -vaihtoehdot puolestaan avaavat kohteen Wordissa tai Excelissa. Tämä ”Send To” -toiminto toimii parhaiten datoina. Tulosteilla kaikki kolme edellä mainittua vaihtoehtoa eivät ole aina valittavissa. Neljäntenä vaihtoehtona on ”E-mail Recipient as a Step in Project”, jonka avulla sähköpostin lähetyksen saadaan pysyväksi osaksi projektikokonaisuutta. Tätä vaihtoehtoa käytettäessä pitää sähköpostiasetusten olla määritettynä Enterprise Guidessa.

Yksittäisiä kuvia tai tulostustaulukoita 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 kopioidava objekti halutaan liittää, ja valitaan ”Muokkaa” -> ”Liitä” (tai ”Edit” -> ”Paste”). Erittäin

käyttökelpoinen tapa siirtää tulostusta Excelliin onkin yksinkertaisesti kopioida RTF-muotoon tulostettu yksittäinen taulukko leikepöydän kautta Excelliin. RTF-tulosteiden käsittelyssä (esimerkiksi tulosteobjektien kopioinnissa) vaikuttaa toisinaan esiintyvän eräs virhetilanne, ainakin käytettäessä EG 4.2:ta yhdessä Office 2003:n kanssa. Nimittäin toisinaan yritettäessä painaa hiiren oikealla painikkeella tulosteobjektin päällä, valikko (josta esimerkiksi kopiointi valittaisiin) ei aukea. Tämä ongelma voidaan ohittaa sulkemalla kyseinen tuloste työtilasta ja avaamalla sen jälkeen uudelleen. Sulkeminen tehdään tulosteen oikean ylänurkan läheisyydessä olevasta mustasta ruksista. Ongelmaa esiintyy ainakin, jos on jo kertaalleen kopioinut tulosteobjektin esimerkiksi Wordiin ja yrittää sen jälkeen kopioida samaista tulosteobjektia uudelleen EG:stä.

11.2.4 Valmiin kuvan muokkaaminen

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

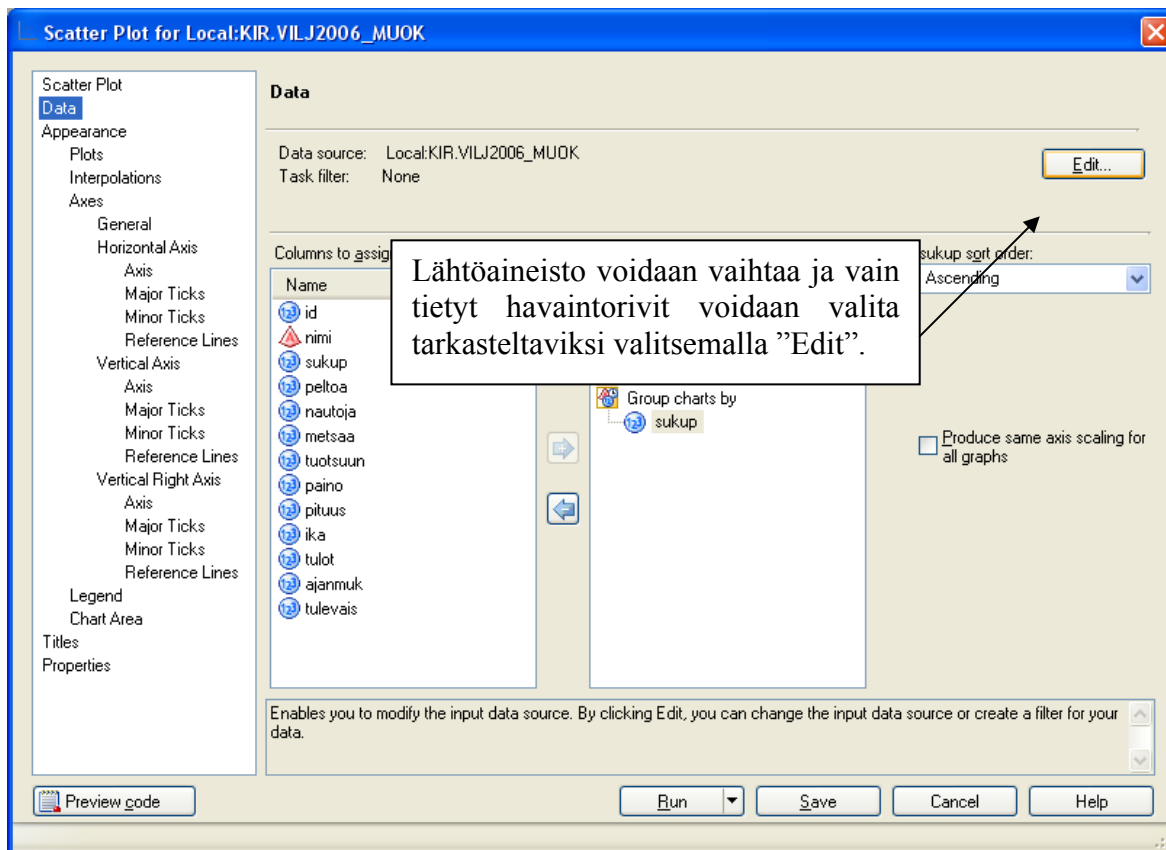


Suurin osa toimenpiteistä löytyy ”Graph Properties” -kohdan alta, mutta muualtakin saattaa löytyä tutustumisen arvoisia toimenpiteitä. ”Graph Properties”-kohdasta voidaan muun muassa vaihtaa kuvan taustan tyyliä kohdassa ”Style”. Tällä voi tehdä esimerkiksi esityksiin näyttäviä kuvia. Oikealla painikkeella 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 tietokoneella, jossa on käyttäjärjestelmänä Windows. Java-kuvaa voidaan puolestaan muokata kaikissa internet-selaimissa, jotka tukevat Javaa.

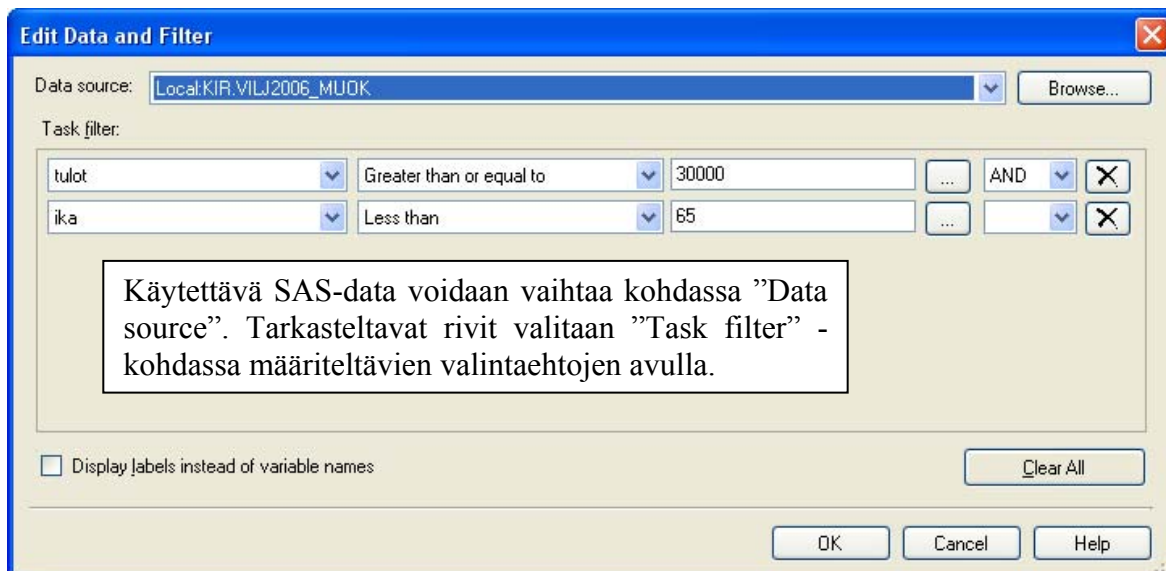
Vastaava kuvien muokkaus jälkikäteen onnistuu myös SAS Report -tulostuskohteeseen tehdyillä kuvilla, mutta ne pitää ensin viedä niin sanottuun SAS-raportti -työkaluun. Tämä onnistuu tulosteen ollessa auki työtilassa valitsemalla asiayhteysvalikosta ”Create” -> ”Create Report” tai ”Create Report Snapshot” (lisätietoa vaihtoehtoista luvussa **SAS-raportti**). Tämän seurauksena kuva aukeaa SAS-raportti -työkaluun. Painettaessa raportissa kuvan päällä hiiren oikealla painikkeella, voidaan tehdä valinta ”Properties”, joka on vastaava kuin edellä HTML-tulostuskohteeseen tehdyillä kuvilla ”Graph Properties” -valinta. Aukeavassa ikkunassa voidaan tehdä monia kuvan muokkaustoimenpiteitä valmiiseen kuvaan. Tämän muokkaustavan etuna on HTML:n ”Graph Propertiesiin” verrattuna se, että tehdyt muutokset tallentuvat automaattisesti projektitiedoston sisälle tehtyyn SAS-raporttiin, jolloin esimerkiksi kuvan alun perin luoneen tehtäväsolumun uudelleensuorittaminen ei johda jälkikäteen muokatun kuvan häviämiseen.

11.2.5 Käytettävän datan vaihtaminen ja rivien valinta tietyissä työkaluissa

SAS Enterprise Guide 4.2:ssa on uutena ominaisuutena useimmissa työkaluissa mahdollisuus vaihtaa ja muokata lähtöaineistona käytettävää SAS dataa työkalun sisältä käsin. Tämä valinta voidaan tehdä ”Data”-välilehden ”Edit”-painikkeesta. Tässä esimerkkinä ”Scatter Plot” -työkalun ”Data”-välilehti:



”Edit”-valinnan jälkeen aukeavassa näkymässä voidaan tehdä seuraavia määryksiä:



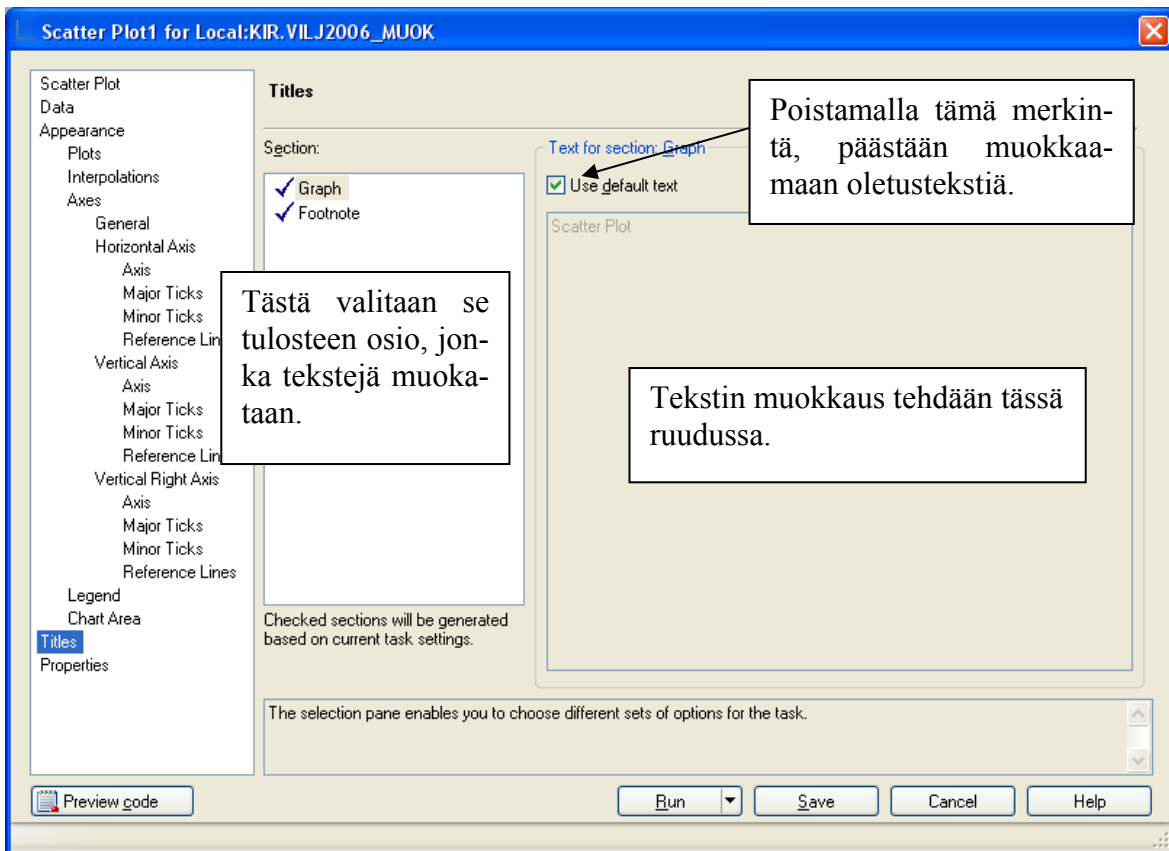
Painettaessa ”OK” tehdyt valinnat tulevat näkyviin ”Data”-välilehdelle ja työkalussa tehtävät tarkastelut koskevat valittua aineistoa tai osa-aineistoa.

11.2.6 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. ”Data” -välilehdellä klikataan käsiteltävän muuttujan nimeä hiiren oikealla painikkeella ja valitaan ”Properties”. Tällöin aukeaa ikkuna, jossa voidaan määrittää kyseiselle muuttujalle seliteteksti (”Label”) tai formaatti (”Format”-kohdassa valintapainike ”Change”). Näin voidaan esimerkiksi formaatin avulla säädellä näytettävien desimaalien määrää tulosteessa.

11.2.7 Otsikoiden ja alaviitteiden muokkaaminen

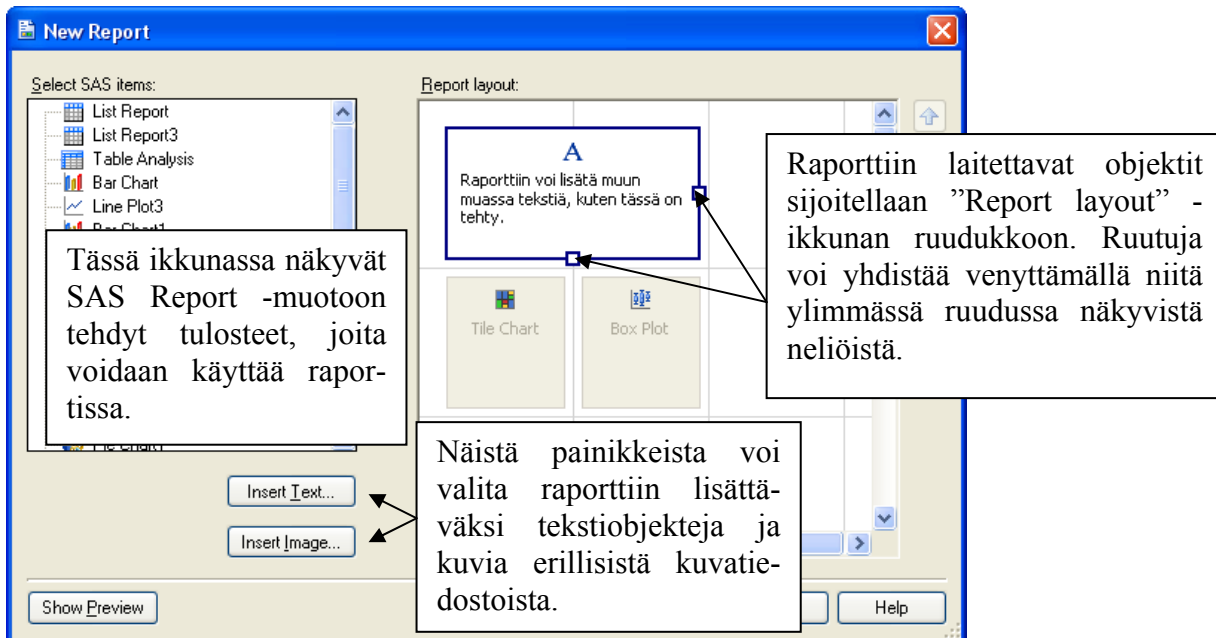
Otsikoiden ja alaviitteiden muokkaaminen toimii samalla tavalla kaikissa EG:n tehtäväso-
luissa, jotka tuottavat tulosteita. Tehtäviä muokatessa on mahdollista valita ”Titles”-välilehti,
joka näyttää seuraavalta:



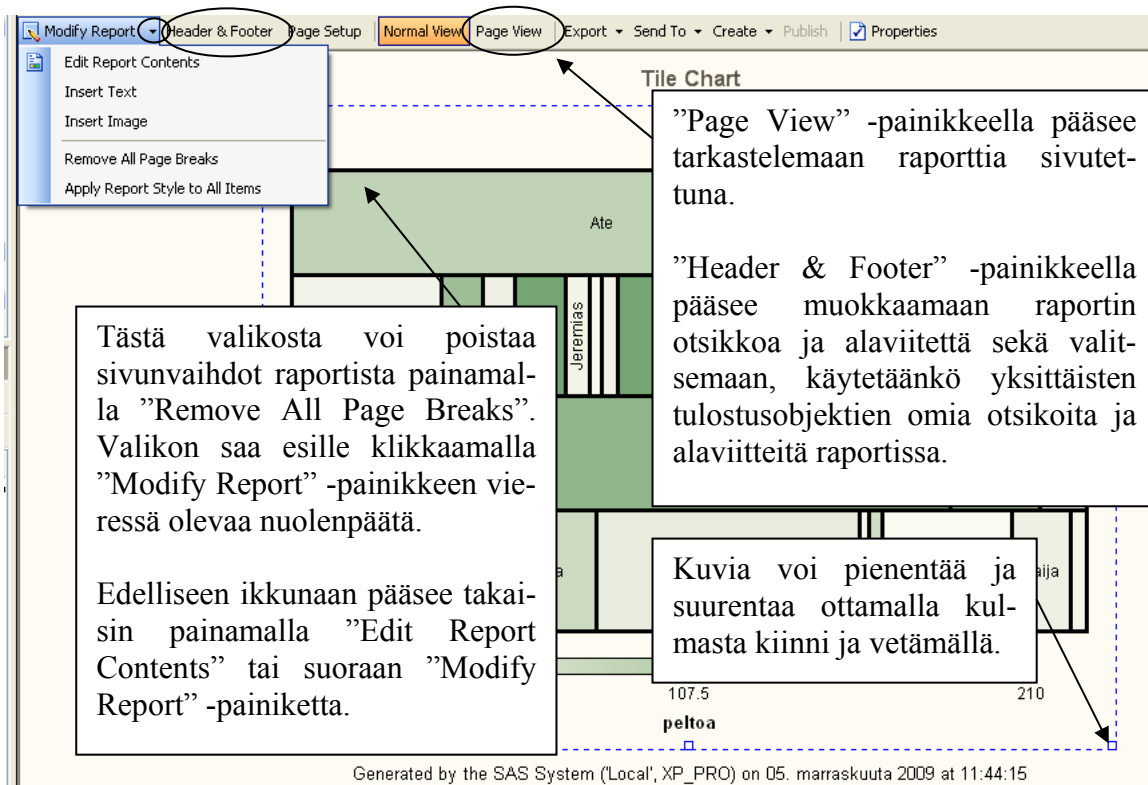
11.3 Yksittäisten tulosteiden kokoaminen yhteen

11.3.1 SAS-raportti

Käyttäjä voi koota SAS Report -tulostuskohteeseen (oletustulostuskohde) 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”:

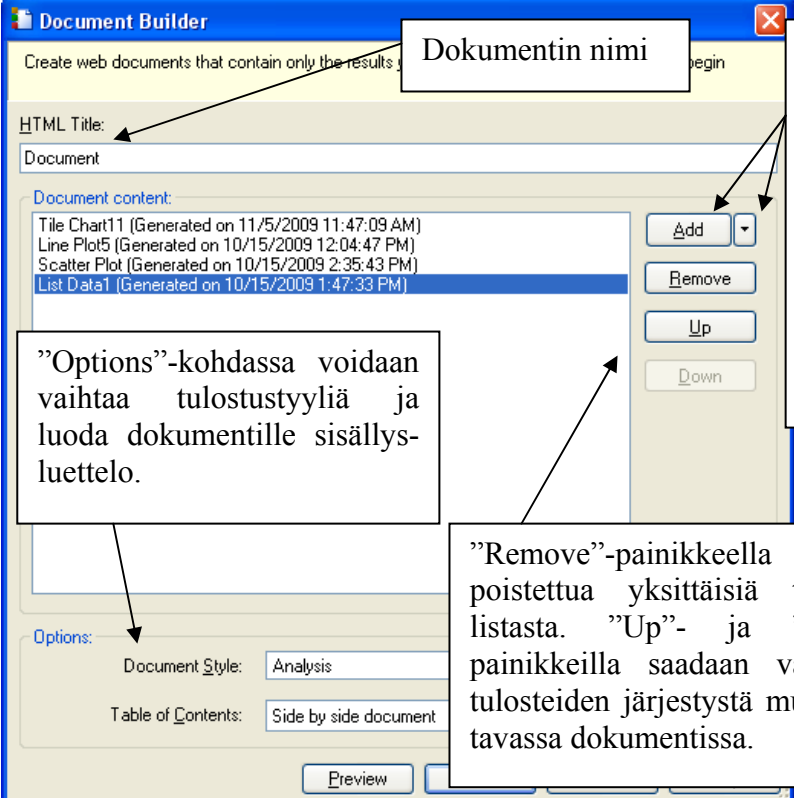


Edellisen kuvan asiayhteysvalikon "Create" -valintapainikkeesta voidaan joko tehdä kopio raportista ("Duplicate Report") tai ottaa tilannekuva raportista ("Create Report Snapshot"). Tavallisen raportin ja tilannekuvan erona on, että tavallinen raportti muuttuu, jos ajo tehdään uudelleen ja sen lähtödata on muuttunut. Vastaavassa tilanteessa raportin tilannekuva sen sijaan ei muutu. Tavallisen raportin luomisen yhteydessä prosessidiagrammiin syntyy myös raporttisolmu. Tilannekuvasta ei sen sijaan synny raporttisolmua prosessidiagrammiin, vaan raportin tilannekuvat näkyvät ainoastaan "Custom Reports" -osiossa. Valinta tavallisen raportin ja tilannekuvavaihtoehdon välillä saattaa tulla EG:ssä vastaan myös yksittäisten SAS Report -muotoon tehtyjen tulosteiden käsittelyn yhteydessä. Tällöin tavalliseen raporttiin liittyvä vaihtoehto voi olla esimerkiksi "Custom Report" ja tilannekuvaan liittyvä vaihtoehto "Custom Report Snapshot". Kaikki SAS-raportit tulevat aina näkyviin myös käyttöliittymän erillisiin "Custom Reports" -osioihin. Prosessidiagrammin ja "Custom Reports" -osioiden välillä voi navigoida kätevästi työkalurivin valintapainikkeen, projektipuun tai asiayhteysvalikon yläpuolella olevan viimeksi käytettyjen osioiden valikon kautta.

Valmiin raportin voi tulostaa paperille tai tallentaa tiedostoksi. Tiedostoksi tallennus onnistuu asiayhteysvalikon "Export"-valinnan kautta. Raportin voi tallentaa SAS Report - (*.srx), HTML- (*.htm tai *.html), XML- (*.xml) tai PDF (*.pdf) -tiedostoksi. Wordiin ja Exceliin raportteja saa siirrettyä "Send To" -toiminnon avulla, jos käytössä on "SAS Add-In for Microsoft Office" -tuote tai muussa tapauksessa kopioimalla leikepöydän kautta. "Send To" -valikosta raportti voidaan lähettää myös sähköpostin liitetiedostona "E-mail Recipient" -valinnalla. Jos raportteja halutaan tarkastella "SAS Web Report Studio"ssa, ne pitää julkaista "Publish"-valinnalla.

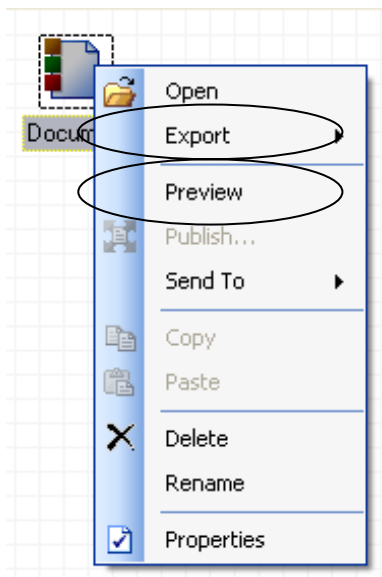
11.3.2 HTML-dokumentti

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



The screenshot shows the Document Builder application window. The title bar reads "Document Builder". Below the title bar, there is a text field for "HTML Title:" with the value "Document". A callout box labeled "Dokumentin nimi" points to this field. The main area is titled "Document content:" and contains a list of items: "Tile Chart11 (Generated on 11/5/2009 11:47:09 AM)", "Line Plot5 (Generated on 10/15/2009 12:04:47 PM)", "Scatter Plot (Generated on 10/15/2009 2:35:43 PM)", and "List Data1 (Generated on 10/15/2009 1:47:33 PM)". To the right of this list are buttons: "Add", "Remove", "Up", and "Down". A callout box on the right explains the "Add" button: "Add"-painikkeella saadaan lisättyä yksittäisiä tulosteita dokumenttiin, jolloin ne näkyvät oheisessa listassa. Nuolipainikkeen kautta dokumenttiin voi lisätä myös muistiinpanosolmuja ("Note"), linkkejä ulkoisiin dokumentteihin ("Other documents") ja erottimia ("Separator"). Below the list, there are "Options:" with "Document Style:" set to "Analysis" and "Table of Contents:" set to "Side by side document". A callout box on the left explains the "Options" section: "Options"-kohdassa voidaan vaihtaa tulostustyyliä ja luoda dokumentille sisällysluettelo. At the bottom, there is a "Preview" button. A callout box at the bottom explains the "Remove" button: "Remove"-painikkeella saadaan poistettua yksittäisiä tulosteita listasta. "Up"- ja "Down"-painikkeilla saadaan vaihdettua tulosteiden järjestystä muodostettavassa dokumentissa.

Painettaessa ”OK” dokumentista tulee projektiin oma kuvake (käyttöliittymän erilliseen ”Custom Reports” -osioon), jota voidaan joko esikatsella (”Preview”) tai tallentaa HTML-tiedostoksi (”Export”):



11.4 Projektin hallinta

Enterprise Guidessa kaikki työ tallennetaan Enterprise Guide projekteina. Tällainen projekti koostuu toisiinsa liittyvistä datoista, tehtäväsolmuista, tulosteista, koodeista, log-tiedostoista ja muistiinpanoista. Projektin osien kertyessä prosessidiagrammiin toisiinsa nuolilla liitettyinä kuvakkeina, kokonaisuus pysyy järjestyksessä, vaikka käytettäisiin useita eri datoja eri tehtävissä. Tämän ansiosta jälkikäteen pystyy kätevästi selvittämään esimerkiksi mitä dataa on käytetty mihinkin analyysiin. Varsinaisia projekteja voi Enterprise Guide -ohjelmassa olla auki kerrallaan vain yksi, mutta samaan projektiin voidaan tehdä useita prosessidiagrammin välilehtiä. Useampaa projektia voi käsitellä samanaikaisesti avaamalla useamman Enterprise Guide -istunnon samalle tietokoneelle.

Enterprise Guide 4.2:ssa on uutena ominaisuutena projektin automaattinen tallennus ja palauttaminen EG:n kaatumisen varalta. Tähän liittyviä asetuksia voi muokata valitsemalla ”Tools” -> ”Options” -> ”Project Recovery”.

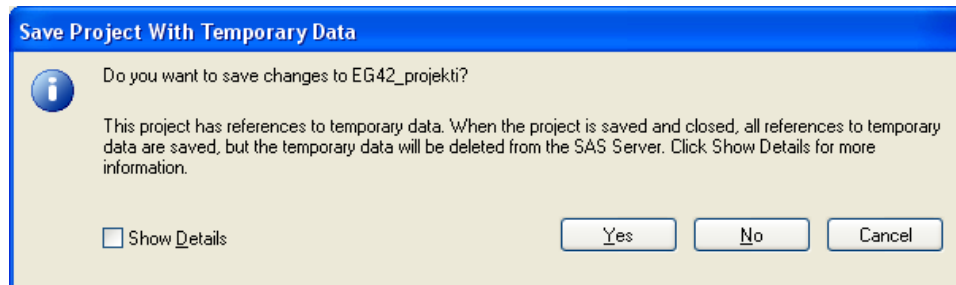
11.4.1 Projektin perustoiminnot

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

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

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

Enterprise Guide -istunnon lopettaminen tehdään tavallisten tietokoneohjelmien tapaan 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 projektikokonaisuuden pitämiseksi järjestyksessä. Ensinnäkin projektikokonaisuutta voi jakaa useammalle prosessidiagrammisivulle. Uuden sivun saa avatuksi valitsemalla ”File” -> ”New” -> ”Process Flow”. Erillisille prosessidiagrammisivuille on kätevä ryhmitellä esimerkiksi samaan projektiin kuuluvien erillisten datojen analyysit. Lisätyn sivun nimeä saa vaihdetuksi klikkaamalla projektipuussa oletusnimen kohdalla hiiren oikealla painikkeella ja valitsemalla ”Rename”. Tehtäväsolmuja voi kopioida toiselle prosessidiagrammisivulle painamalla kuvakkeen päällä hiiren oikealla painikkeella ja valitsemalla ”Copy” ja sitten painamalla kohdesivulla hiiren oikeata painiketta ja valitsemalla ”Paste”. Huomionarvoista on, että kopioitaessa tehtäväsolmu uudelle prosessidiagrammisivulle, sen mukana kopioituu automaattisesti myös viittaus tehtävän käyttämään lähtödataan.

Järjestyksen ylläpitoa prosessidiagrammissa auttaa ennen kaikkea sinne tulostuvien kuvakkeiden eli solmujen organisoiminen loogiseen järjestykseen. Oletuksena Enterprise Guidessa on käytössä toiminto ”Auto Arrange”, joka järjestää solmuja käyttäjän puolesta prosessidiagrammissa. ”Auto Arrange”-toiminnon saa pois käytöstä klikkaamalla prosessidiagrammin tyhjässä kohdassa hiiren oikealla painikkeella ja poistamalla valinta kohdasta ”Auto Arrange”. Tämän jälkeen solmuja saa siirreltyä ottamalla niistä hiirellä kiinni ja raahaamalla ne uuteen paikkaan. Solmujen välillä olevat yhdysviivat venyvät siirrettäessä.

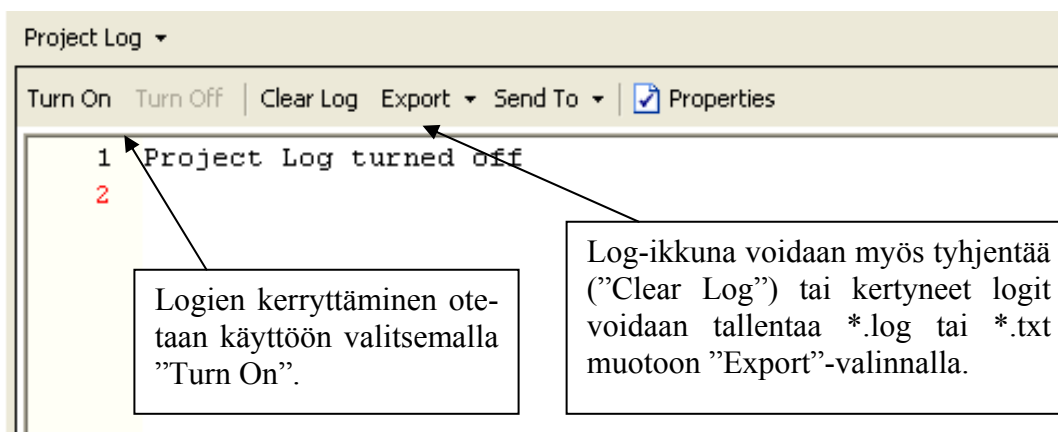
Luotaessa SAS-datoja EG:n työkalujen avulla prosessidiagrammin selkeyden kannalta on myös tärkeää kiinnittää huomiota nimien valintaan ja tallennuskohteeseen. Kannattaa siis tarkistaa onko oletuskirjastoviitteen osoittama kansio hyvä kohde tallennukselle ja tarvittaessa vaihtaa kirjastoviitettä. Kannattaa myös aina valita luotavalle datalle looginen nimi.

Joskus prosessidiagrammin jäsentämistä helpottaa ylimääräisten yhdysviivojen vetäminen sellaisten prosessidiagrammissa olevien solmujen välille, joiden välillä yhdysviivaa ei muuten ole. Tällainen linkitys onnistuu klikkaamalla ensimmäistä linkitettävää solmuja hiiren oikealla painikkeella ja valitsemalla ”Link solmun_nimi to”. Tällöin aukeaa ikkuna, josta voidaan valita solmu, johon yhdysviiva vedetään. Linkittäminen on käyttökelpoista esimerkiksi SAS-koodisolmuja käytettäessä, jolloin voidaan linkittää koodit niiden käyttämiin lähtödatoihin (tällaisia linkkejä ei automaattisesti synny). Selkeyden lisäksi usein tärkeämpi syy yhdysviivojen käyttämiselle on, että niiden avulla solmut voidaan pakottaa suoritettavaksi tietyssä järjestyksessä, kun suoritetaan koko prosessidiagrammi tai jokin sen laajempi haara yhdellä kertaa.

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 tai vastaavasti asiayhteysvalikon ”Zoom”-kohdasta.

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 pidempääkin dokumentaatiota, ja solmut tallentuvat projektin mukana pysyviksi muistiinpanoiksi projektista. Tätä ominaisuutta käyttämällä projektiin palaaminen pitkän ajan kuluttua helpottuu.

Kaikki EG-projektin logit voi myös kerätä yhdeksi kokonaisuudeksi. Tämä ominaisuus vastaa perus-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 tietyistä vastaavista tiedoista tavallisen SAS-koodin suorittamisesta. Enterprise Guidessa koodin kerryttäminen log-ikkunaan pitää kuitenkin ensin ottaa käyttöön. Tämä onnistuu valitsemalla ”View” -> ”Project Log” tai valitsemalla asiayhteysvalikosta ”Project Log”, jolloin työtilaan aukeaa seuraava log-ikkuna, josta kertyneitä logeja voi tarkastella:



11.4.3 Tehtäväsolmujen käsittely

Prosessidiagrammissa olevan valmiin tehtäväsolmun voi suorittaa klikkaamalla sitä hiiren oikealla painikkeella ja valitsemalla ”Run”. Tehtävän suoritus voidaan myös tehdä valitsemalla kyseinen solmu aktiiviseksi prosessidiagrammissa (eli klikkaamalla kuvaketta hiirellä kertaalleen) ja painamalla asiayhteysvalikossa olevaa ”Run”-painiketta. Kolmas vaihtoehto on tuplaklikata tehtäväsolmua ja suorittaa se ”Refresh”-painikkeella. Tehtävän suorituksen pysäyttäminen puolestaan onnistuu klikkaamalla asiayhteysvalikossa olevaa punaista ”Stop-nappulaa”. Edellä mainittujen tapojen lisäksi tehtävien suorittaminen ja suorituksen pysäyttäminen onnistuu myös ”Program”-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 (error), josta yleensä löytyy virheen syy. Pienemmistä rikkeistä tehtäväsolmun kuvakkeeseen tulee keltaisella pohjalla oleva huutomerkki. Nämä huutomerkki eivät yleensä viittaa merkittävään virheeseen, mutta ainakin jos tehtävän tulos ei ole looginen, asia kannattaa tarkistaa logista (warning, login löytää tässä tilanteessa tulosten rinnalta erillisenä välilehtenä).

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*” tai tekemällä saman valinnan asiayhteysvalikon kohdasta ”Run”. Vastaavasti käyttäjä voi suorittaa koko projektin kaikkien prosessidiagrammissivujen tehtäväsolmut valitsemalla asiayhteysvalikosta ”Run” -> ”Run Project”. Toisaalta voi suorittaa tietyn tehtäväsolmun ja kaikki sen perään yhdysviivoilla yhdistetyt tehtäväsolmut klikkaamalla ensimmäistä suoritettavaa solmua hiiren oikealla painikkeella ja valitsemalla ”Run Branch From *tehtäväsolmun nimi*” tai tekemällä vastaavan valinnan asiayhteysvalikon kohdasta ”Run”. Käyttäjä voi myös valita tietyt tehtäväsolmut suoritettaviksi pitämällä näppäimistön Ctrl-painiketta painettuna (poiminnan ajan), poimimalla hiirellä suoritettavat tehtäväsolmut mukaan valintaan ja tämän jälkeen valitsemalla esimerkiksi asiayhteysvalikosta ”Run” -> ”Run Selected Items”.

Jos tehtäväsolmuun liittyviä valintoja puolestaan halutaan muokata ennen suorittamista, tuplaklikataan tehtäväsolmua prosessidiagrammissa ja saatetaan tehtäväsolmu muokkaustilaan valitsemalla asiayhteysvalikosta ”Modify Task”. Muokkausten jälkeen solmun voi suorittaa uudelleen valitsemalla ”Run”. Tällöin EG kysyy, halutaanko edellisen suorituksen tulokset korvata. 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 prosessiketjun alkuun. Datan vaihtaminen onnistuu klikkaamalla prosessidiagrammissa 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 voida 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 valitsemalla ”Delete”. Kun solmu poistetaan, samalla poistuu sen perässä olevia, siihen liittyviä muita solmuja. Kun esimerkiksi poistetaan sirontakuvion tuottava tehtäväsolmu, samalla poistuu siihen liittyvä tulostesolmu, koodisolmu ja log-solmu.

Koko projekti tai sen yksittäinen prosessidiagrammisivu voidaan määrätä suoritettavaksi ajastuksen avulla. Ajastustyökalu löytyy prosessidiagrammiin liittyvästä asiayhteysvalikosta kohdasta "Schedule" -> "Projektin nimi" tai "Prosessidiagrammisivun nimi". Ajastettuja tehtäviä voi tarkastella suomenkielisessä Windows XP:ssä valitsemalla "Käynnistä" -> "Ohjauspaneeli" -> "Ajoitetut tehtävät". "Ordered List" -työkalulla voi puolestaan suorittaa useamman projektissa olevan tehtäväsolmun halutussa järjestyksessä. Työkalussa valitaan projektissa olevia tehtäviä listaan ja tuloksena syntyy "Ordered List" -tehtäväsolmu omaan "Ordered Lists" -osioonsa käyttöliittymässä. Listaan valitut tehtävät voidaan tämän jälkeen suorittaa helposti sarjassa suorittamalla kyseinen "Ordered List" -tehtäväsolmu. Tämän työkalun saa käyttöön klikkaamalla "File" -> "New" -> "Ordered List".

11.5 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 tarvittavat kirjastoviitteet pitää yhdistää 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/HF2/92_all_hosts.html).
- 4) EG 4.2:n myötä SAS-järjestelmänvalvojalla on tietyin ehdoin mahdollisuus muokata käyttöliittymässä saatavilla olevaa työkaluvalikoimaa erilaiseksi eri käyttäjäryhmille. Jos siis jokin työkalu puuttuu työpaikkasi Enterprise Guidesta, ota yhteyttä SAS-järjestelmänvalvojaasi.
- 5) Enterprise Guide 4.2 saattaa toimia hyvin hitaasti, jos tietokoneella on internet-yhteys, mutta kone ei saa yhteyttä verkkolevyihin. Tällainen tilanne saattaa esiintyä esimerkiksi, jos työpaikan kannettava tietokone on kotona internet-yhteydessä. Tällaisessa tilanteessa kannattaa kokeilla nopeuttaisiko internet-yhteyden katkaiseminen koneen toimintaa.
- 6) Tässä oppaassa mahdollisesti havaituista virheistä voi lähettää palautetta kirjoittajalle sähköpostilla osoitteeseen timo.hurme@mtt.fi tai timo.hurme@gmail.com.

- Davis, J. B. 2007. Statistics using SAS Enterprise Guide. Cary, NC: SAS Institute Inc. 772 s.
- Helenius, H. ja Vahlberg, T. 2008. Tilastomenetelmien soveltamisesimerkkejä SAS® Enterprise Guide:lla –versio 4.1. Espoo: Numos Oy. 263 s.
- Hurme, T. 2007. Opas SAS Enterprise Guiden käyttäjälle – EG versio 4.1. MTT:n selvityksiä 140. Jokioinen: MTT. 115 s.
- Laine, M. 2000. SAS peruskurssi -oppaan verkkopainos:
(<http://www.helsinki.fi/~mjlaine/sasopas/sasopas.pdf>). Viitattu 18.2.2010.
- Prairie, K. 2005. The essential PROC SQL handbook for SAS users. Cary, NC: SAS Institute Inc. 572 s.
- SAS 9.2 Help and Documentation. Valittavissa Enterprise Guide 4.2 -ohjelman valikosta “Help” -> “SAS Syntax Help”. Viitattu 18.2.2010.
- SAS Enterprise Guide Help. Valittavissa Enterprise Guide 4.2 -ohjelman valikosta “Help” -> “SAS Enterprise Guide Help”. Viitattu 18.2.2010.
- Slaughter, S. J. ja Delwiche, L. D. 2006. The Little SAS book for Enterprise Guide 4.1. Cary, NC: SAS Institute Inc. 388 s.

MTT TEKEE TIETEESTÄ ELINVOIMAA

MTT RAPORTTI5

www.mtt.fi/julkaisut

MTT Raportti -verkkojulkaisusarjassa julkaistaan maatalous- ja elintarviketutkimusta sekä maatalouden ympäristötutkimusta käsitteleviä tutkimusraportteja. Lukijoille tarjotaan tietoa MTT:n kaikilta tutkimusaloilta eli biologiasta, teknologiasta ja taloudesta.

Tämän julkaisun tilaushinnasto ja -ohje löytyvät internetistä osoitteesta www.mtt.fi/julkaisut

MTT, 31600 Jokioinen.
Puh. (03) 4188 2327, sähköposti julkaisut@mtt.fi

