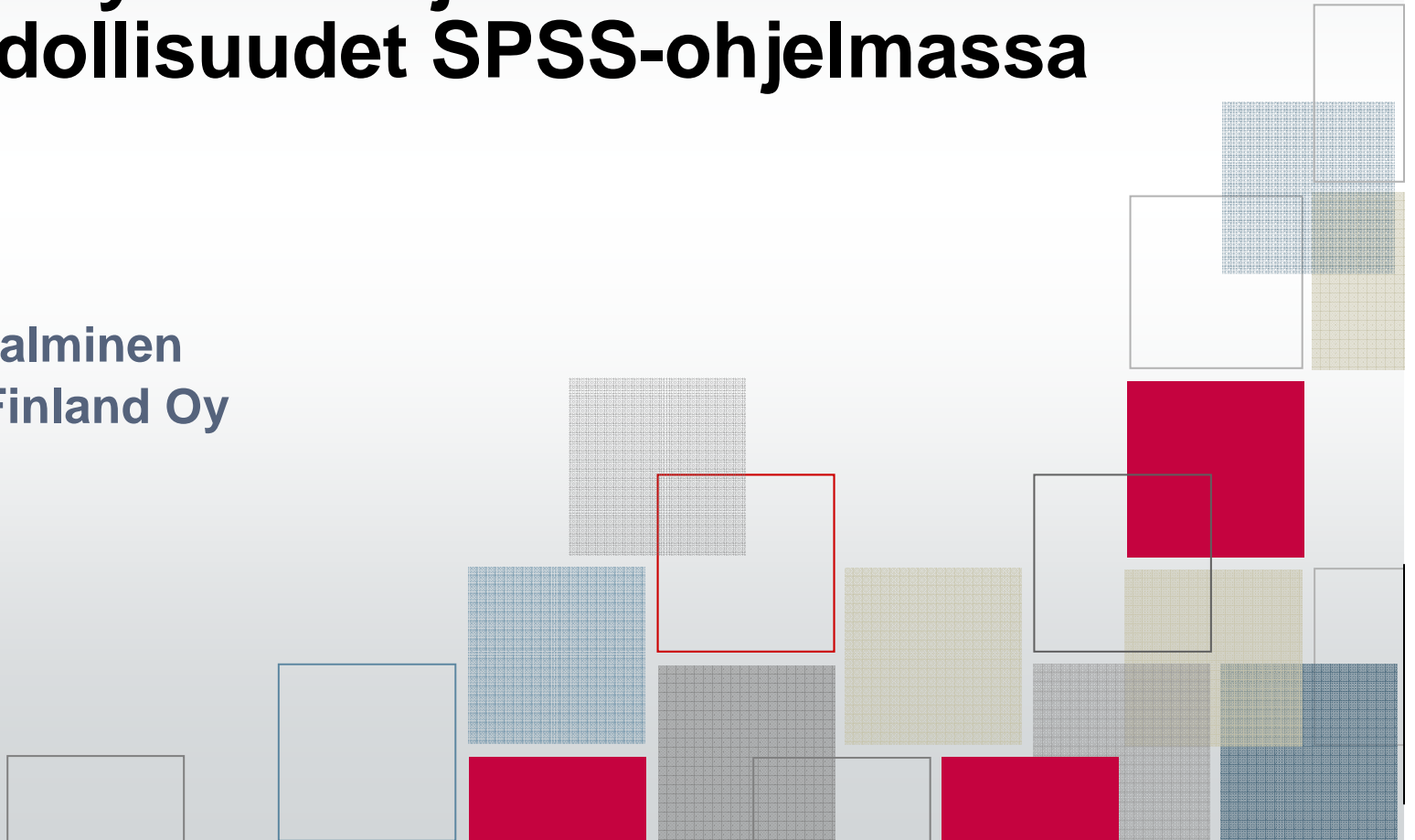


# > R- ja Python-ohjelmoinnin mahdollisuudet SPSS-ohjelmassa

Lauri Salminen  
SPSS Finland Oy



## > Esityksen rakenne

- Esittäytyminen ja taustaa
- Mitä on Python? Mitä on R?
- Integraation syyt ja tavoitteet
- Esimerkkejä
- Miten päästä alkuun?
- Kysymyksiä & keskustelua

## > Mitä on Python?

- Selkeä, helppolukuinen ja nopeasti opittava ohjelmointikieli
- Python ja sen kirjastot perustuvat avoimeen lähdekoodiin; myös kaupallinen käyttö ja uudelleenlevittäminen sallittu
- Valmiit toiminnot ja kirjastot tekevät kielestä tehokkaan ja nopean käyttää

***Lue lisää: <http://www.python.org>***

## > Mitä on R?

- Avoimeen lähdekoodiin perustuva ohjelmointiympäristö tilastoanalyysiin ja grafiikkaan
- Lukemattomia valmiita menetelmiä ilmaisissa kirjastoissa
- Mahdollisuus kehittää ja julkaista omia menetelmiä
- Laajasti käytetty tilastollisten menetelmien opetus- ja tutkimuskielenä

***Lue lisää: <http://www.r-project.org/>***

# > ”SPSS Programmability Extension” edut ja merkitys käyttäjille

## Python

- Ohjauslogiikan liittäminen syntaksiin
- Integraatio muihin ohjelmiin ja järjestelmiin
- Rutiiniajajojen automatisointi skripteinä

## R

- Uusien ja/tai eksoottisten menetelmien käyttö tutussa ympäristössä
- Omien menetelmien tai toimintojen kehittäminen, julkaiseminen ja jakelu muille käyttäjille

## > Esimerkki 1A: R-kieltä syntaksissa

\* Tämä on SPSS-syntaksia.

```
GET FILE='C:\Aineistot\Cars.sav'.
```

\* Siirrytään käyttämään R-kieltä.

```
BEGIN PROGRAM R.
```

```
casedata <- spssdata.GetDataFromSPSS()
```

```
model <-glm(mpg~engine+horse+weight,data=casedata)
```

```
print(model)
```

```
END PROGRAM.
```

## > Esimerkki 1A: Tuloste

```
Call: glm(formula = mpg ~ engine + horse +
weight, data = casedata)
```

Coefficients:

(Intercept)	engine	horse	weight
44.015491	-0.005527	-0.055633	-
0.004617			

Degrees of Freedom: 391 Total (i.e. Null);  
388 Residual

(14 observations deleted due to missingness)

Null Deviance:

23820

Residual Deviance: 7721 AIC: 2291

## > Esimerkki 1B: R-kieltä syntaksissa

\* Tämä on SPSS-syntaksia.

```
GET FILE='C:\Aineistot\Cars.sav'.
```

\* Siirrytään käyttämään R-kieltä.

```
BEGIN PROGRAM R.
```

```
casedata <- spssdata.GetDataFromSPSS()
```

```
model <- glm(mpg~engine+horse+weight,data=casedata)
```

```
res <- summary(model)
```

```
spsspivottable.Display(res$coefficients,
```

```
title="Model Coefficients",
```

```
format=formatSpec.GeneralStat)
```

```
print(plot(model))
```

```
END PROGRAM.
```

# > Esimerkki 1B: Tuloste

The screenshot displays the SPSS Statistics Viewer interface. The left pane shows a tree view of the output, including R, RGraph, and RGraphic objects. The main window shows the following content:

**R**

[DataSet1] C:\Program Files\SPSSInc\Statistics17\Samples\English\Cars.sav

**Model Coefficients**

	Estimate	Std. Error	tvalue	Pr(> t )
(Intercept)	44,015	1,272	34,597	,000
engine	-,006	,007	-,786	,432
horse	-,056	,013	-4,153	,000
weight	-,005	,001	-6,186	,000

**RGraphic**

[DataSet1] C:\Program Files\SPSSInc\Statistics17\Samples\English\Cars.sav

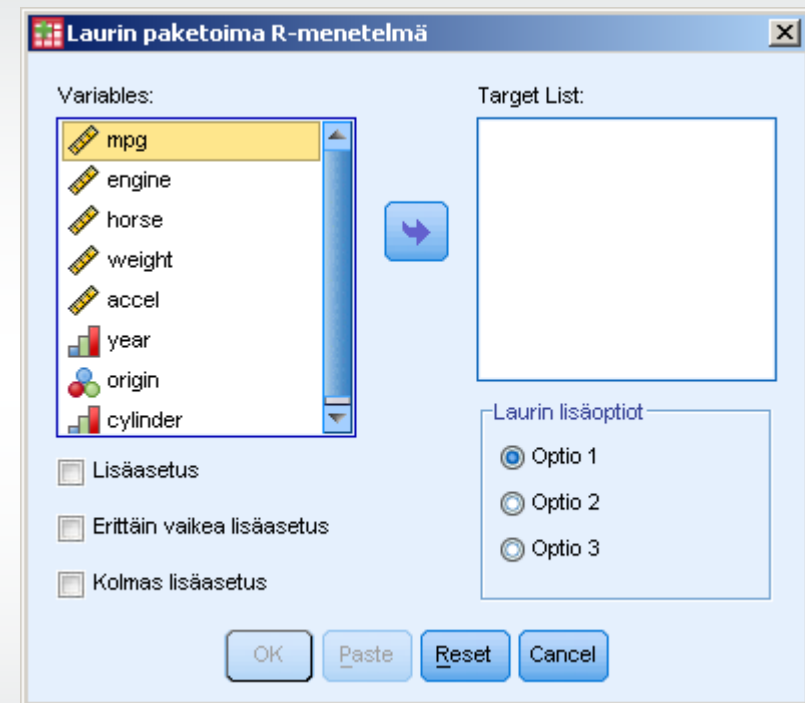
**Residuals vs Fitted**

The residual plot shows a scatter of points around a horizontal line at zero. The x-axis is labeled "Predicted values glm(mpg ~ engine + horse + weight)" and ranges from 5 to 35. The y-axis is labeled "Residuals" and ranges from -30 to 20. A red trend line is visible, showing a slight upward slope. Two points are labeled: 330 and 35.

PASW Statistics Processor is ready

## > SPSS Custom Dialog

- Sisältää seuraavat tiedot:
  - Menetelmän nimi
  - Sijainti valikoissa
  - Muuttujien tyyppi ja valinta
  - Tarvittavat parametrit
  - Ajettava ohjelma
  
- Mahdollistaa R/Python menetelmien helpon ”paketoimisen” ja käytön



## > Esimerkki 2: Paketoitu R-menetelmä\*

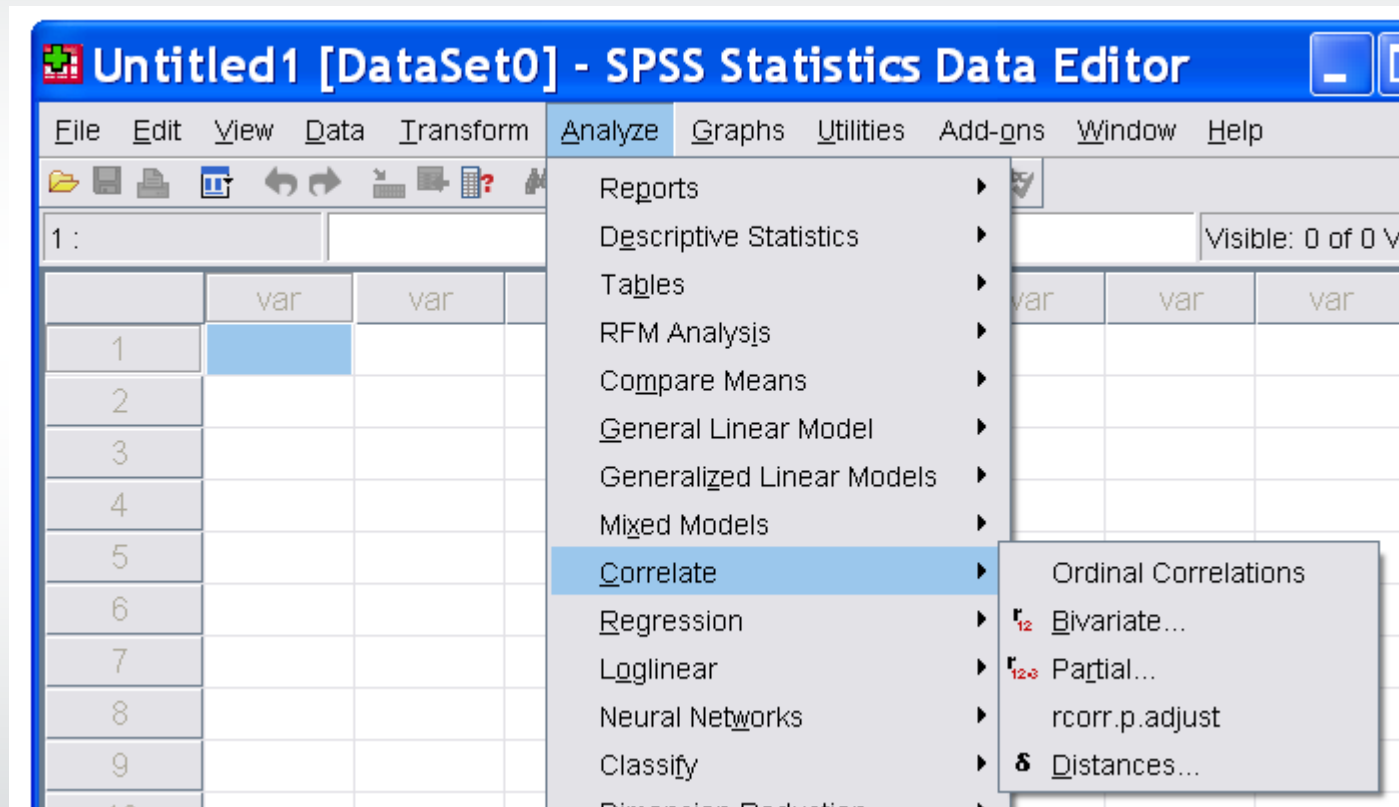
- Paketointi sisältää
  - Yksinkertainen R-ohjelma: "Adjusted P-values via Sequential Bonferroni Method Applied" (.r-tiedosto)
  - SPSS Custom Dialog-tiedosto (.spd-tiedosto)
  
- Asentaminen
  - Utilities -> Install Custom Dialog
  - Browse -> rcorr.p.adjust.spd -> OK
  - Käynnistä SPSS uudelleen

\*) Ohjelmoinut ja julkaissut:

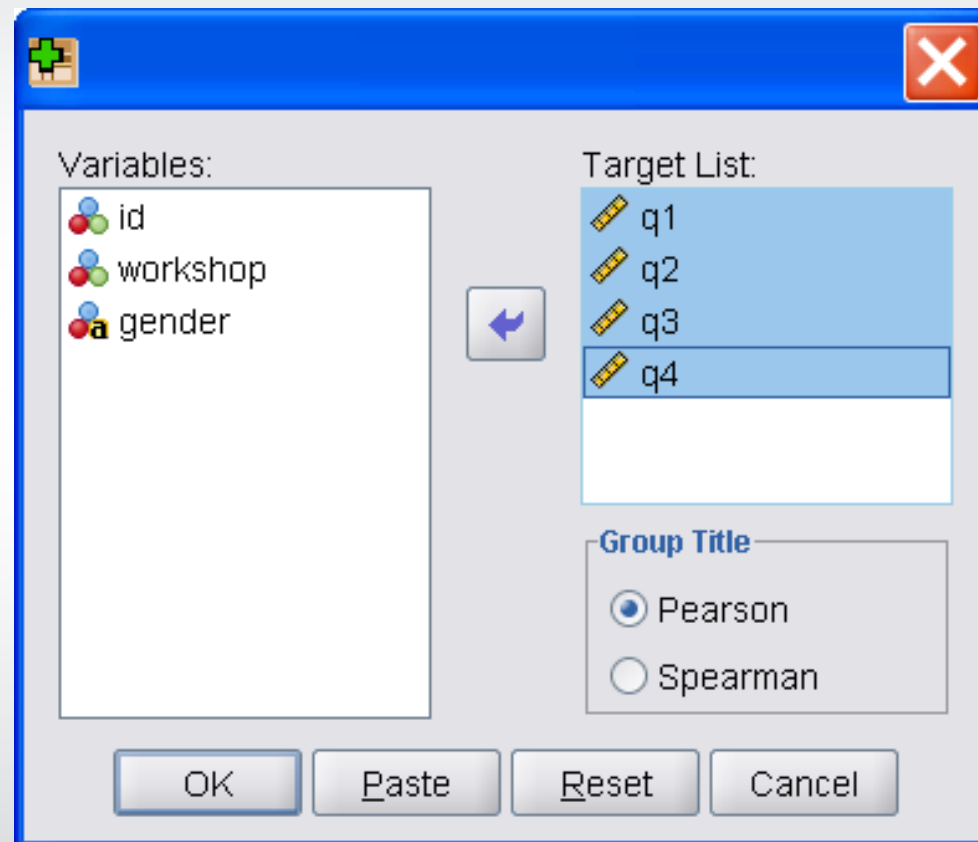
Robert A. Muenchen, University of Tennessee

Saatavilla: <http://RforSASandSPSSusers.com>

# > Paketoitu R-menetelmä valikossa



# > Paketoidun R-menetelmän käyttö



## > Paketoidun R-menetelmän tuloste

Adjusted via Sequential Bonferroni Method Applied

P	q1	q2	q3	q4
q1		0.0947	0.8564	0.0072
q2	0.0947		0.5581	0.0074
q3	0.8564	0.5581		0.9556
q4	0.0072	0.0074	0.9556	

P Adjusted (pearson)	q1	q2	q3	q4
q1	NA	0.3789	1	0.0432
q2	0.3789	NA	1	0.0432
q3	1.0000	1.0000	NA	1.0000
q4	0.0432	0.0432	1	NA

## > Extension Bundles

- R- ja Python-menetelmiä, joista on tehty vakio muotoinen jakelupaketti
- SPSS -> Utilities -> Extension Bundles
- Hieman työläämpiä rakentaa, mutta helposti käytettäviä/jaettavia

## > Esimerkki: Robust Regression

- This procedure estimates a linear regression model, robustly, using an M estimator. It uses the rlm function from the R MASS package. The procedure allows you to save both the residuals from the regression and the estimated coefficients.
- <http://www.stats.ox.ac.uk/pub/MASS4/>
- Analyze>Regression>Robust Regression
- R-packages used: MASS

## > Esimerkki: Robust Regression

**Robust Regression**

**Variables:**

- year
- origin
- cylinder

**Dependent:** mpg

**Independent Variables:**

- engine
- horse
- weight
- accel

Categorical variables will be converted automatically to factors.

**Missing Values**

- Listwise deletion
- Fail

This dialog requires the R Integration Plug-in and the R MASS package.

OK Paste Reset Cancel Help Save...

# > Robust Regression: Tuloste

SPSSINC ROBUST REGR DEPENDENT = mpg ENTER = engine horse weight  
/OPTIONS MISSING=LISTWISE  
/SAVE.

→ **Robust Regression**

[DataSet1] C:\Data\Automatisointi\Cars.sav

	Value	Std. Error	tvalue
(Intercept)	42,516	1,136	37,414
engine	-,012	,006	-1,904
horse	-,021	,012	-1,778
weight	-,005	,001	-7,531

rim(formula = mpg ~ engine+horse+weight, data = dta, na.action = na.exclude, method = "MM", model = FALSE)  
Residual standard error: 3.53851  
Degrees of freedom: 388

PASW Statistics Processor is ready

## > Esimerkki 4: Python toimintojen automatisoinnissa ja hallinnassa

- Tavoitteena mahdollistaa käyttö osana automaattista prosessia
- Tyypillisiä automatisoitavia toimintoja
  - Ajantasaisen aineiston lataaminen ja muokkaaminen
  - Vakioraporttien ja taulukointien tuottaminen
  - Tulosteiden ulkoasun muokkaaminen
  - Ennusteiden päivittäminen ajastetusti

## > Esimerkki 4: Python toimintojen automatisoinnissa ja hallinnassa

```
import SpssClient

# avataan ohjelma, aineistot ja talletetut analyysit
SpssClient.StartClient()
aineisto=SpssClient.OpenDataDoc('c:/Data/data.sav')
vakioajo=SpssClient.OpenSyntaxDoc('c:/Data/Automatisointi/vakioajo.sps')
vakioajo.RunSyntax()

# talletetaan tuotetut taulukot PDF dokumenttiin lukemista varten
OutputDoc = SpssClient.GetDesignatedOutputDoc()
OutputDoc.SelectAllTables()
OutputDoc.ExportDocument(SpssClient.SpssExportSubset.SpssSelected
                        "c:/Data/Automatisointi/output.pdf",
                        SpssClient.DocExportFormat.SpssFormatPdf)

aineisto.CloseDocument()
vakioajo.CloseDocument()
SpssClient.StopClient()
```

## > Miten päästä alkuun?

- SPSS Developer Central ([www.spss.com/devcentral](http://www.spss.com/devcentral) )
  - "Python Essentials"-paketti
  - "R-Essentials"-paketti
  - Keskustelufoorumi ja esimerkkejä
  
- Python ( [www.python.org](http://www.python.org) )
  - Dokumentaatio
  - Ohjeet & oppaat
  
- R-kieli ([www.r-project.org](http://www.r-project.org) )
  - Dokumentaatio
  - Ohjeet & oppaat
  
- Robert A. Muenchen (<http://RforSASandSPSSusers.com>)
  - Esimerkkiohjelmia
  - Hyvä R-opas SPSS ja SAS-käyttäjille

## > **Kommentteja ja kysymyksiä?**

**[lauri.salminen@spss.fi](mailto:lauri.salminen@spss.fi)**

**P. 09 4355 9222**

