



2008 PREDICTIVE MODELING SEMINAR



Software and Utilities: Free or Inexpensive – R utilities/add-ons

Presentation by Louise Francis,
Francis Analytics and Actuarial
Data Mining, Inc.

Many slides contributed by Matt
Flynn, ISO

A graphic featuring three interlocking gears of different sizes and colors (orange, green, red) set against a background of binary code (0s and 1s). The text 'DATA • ANALYTICS' is written in a circular path around the top gear, and 'DECISION SUPPORT' is written in a circular path around the bottom gear.

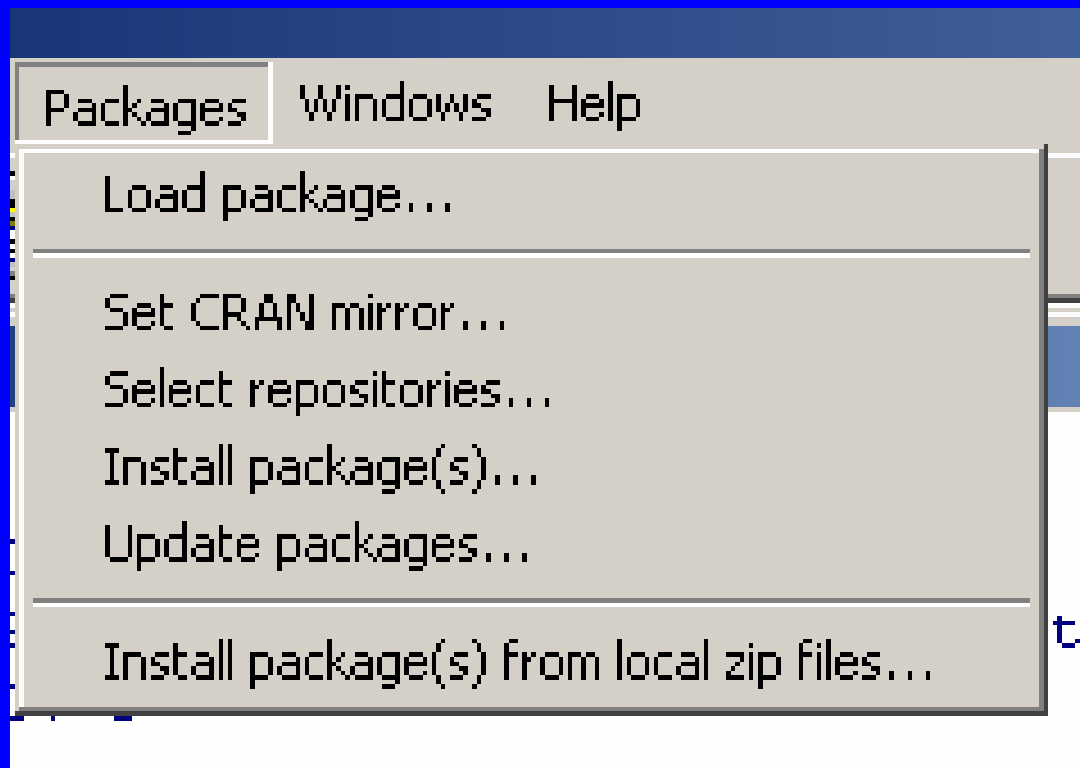
DATA • ANALYTICS
DECISION SUPPORT

ISO Confidential Material

Some Useful R utilities/add-ons

- **Calling R code from SAS**
- **R Commander – Rcmdr**
- **actuar**
- **rpanel**
- **iplots**
- **RExcel**
- **JGR – “Jaguar”**

Install Package from CRAN Mirror



Packages

rateratio.test
rattle
RBGL
RBloomberg
rbugs
Rcapture
rcdd
rcdk
Rcmdr
RcmdrPlugin.epack
RcmdrPlugin.Export
RcmdrPlugin.FactoMineR
RcmdrPlugin.HH
RcmdrPlugin.IPSUR
RcmdrPlugin.TeachingDemos
RColorBrewer
rcom
rcompgen
rda
RDieHarder
realized
ref
regress
regsubseq
regtest
relaimpo
relations
relax
relaxo
reldist
Reliability
relimp
relsurv
repolr
reshape
ResistorArray
resper
reweight

OK

Cancel

Load Package



Packages Windows Help

Load package...

Set CRAN mirror...

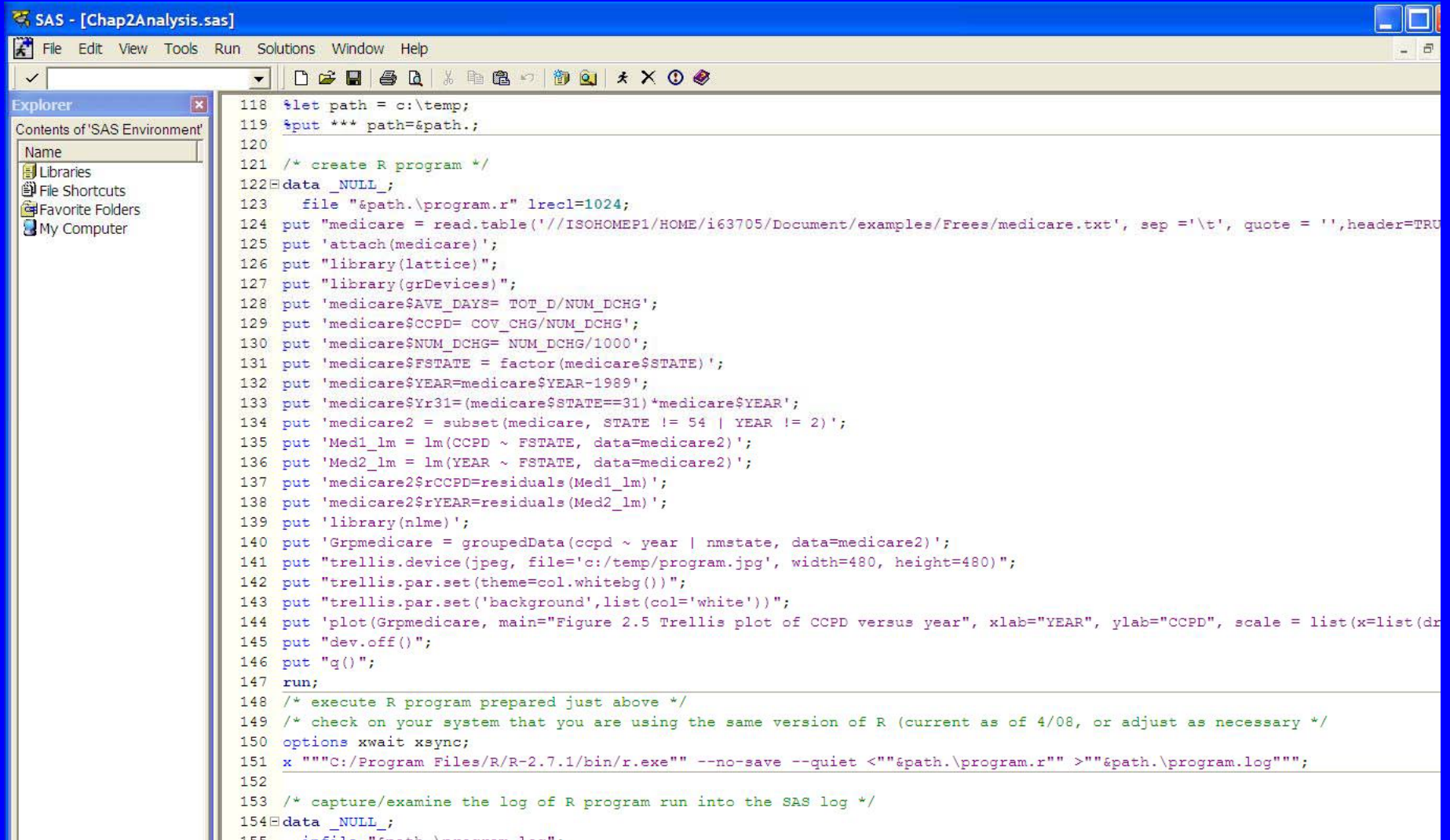
Select repositories...

Install package(s)...

Update packages...

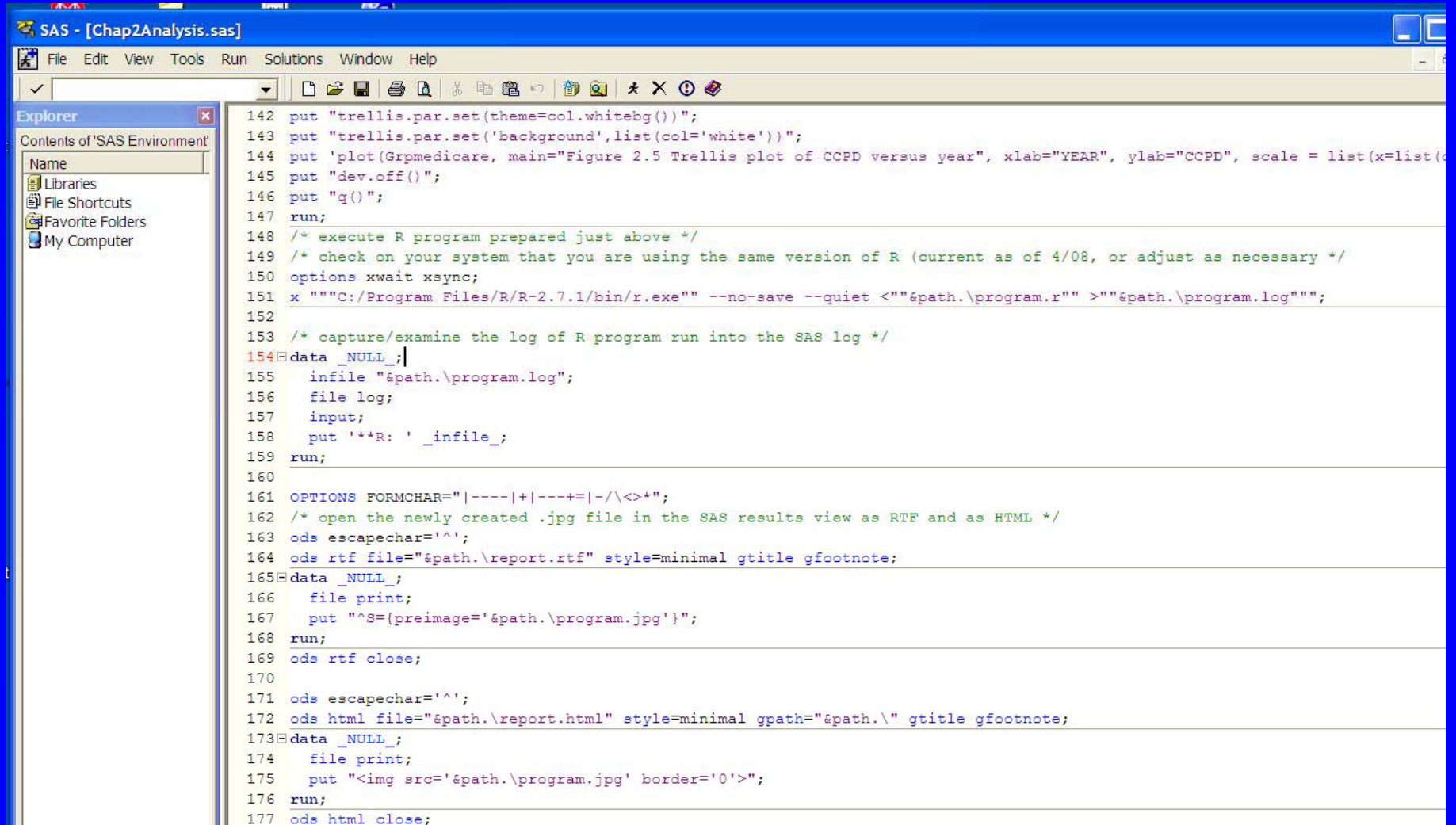
Install package(s) from local zip files...

Calling R from SAS



```
SAS - [Chap2Analysis.sas]
File Edit View Tools Run Solutions Window Help
Explorer
Contents of 'SAS Environment'
Name
Libraries
File Shortcuts
Favorite Folders
My Computer
118 %let path = c:\temp;
119 %put *** path=%path.;
120
121 /* create R program */
122 data _NULL_;
123   file "%path.\program.r" lrecl=1024;
124   put "medicare = read.table('///ISOHOMEPI/HOME/i63705/Document/examples/Frees/medicare.txt', sep = '\t', quote = '', header=TRUE);";
125   put 'attach(medicare)';
126   put "library(lattice)";
127   put "library(grDevices)";
128   put 'medicare$SAVE_DAYS= TOT_D/NUM_DCHG';
129   put 'medicare$CCPD= COV_CHG/NUM_DCHG';
130   put 'medicare$NUM_DCHG= NUM_DCHG/1000';
131   put 'medicare$FSTATE = factor(medicare$STATE)';
132   put 'medicare$YEAR=medicare$YEAR-1989';
133   put 'medicare$Yr31=(medicare$STATE==31)*medicare$YEAR';
134   put 'medicare2 = subset(medicare, STATE != 54 | YEAR != 2)';
135   put 'Med1_lm = lm(CCPD ~ FSTATE, data=medicare2)';
136   put 'Med2_lm = lm(YEAR ~ FSTATE, data=medicare2)';
137   put 'medicare2$rCCPD=residuals(Med1_lm)';
138   put 'medicare2$rYEAR=residuals(Med2_lm)';
139   put 'library(nlme)';
140   put 'Grpmedicare = groupedData(ccpd ~ year | nmstate, data=medicare2)';
141   put "trellis.device(jpeg, file='c:/temp/program.jpg', width=480, height=480)";
142   put "trellis.par.set(theme=col.whitebg())";
143   put "trellis.par.set('background',list(col='white'))";
144   put 'plot(Grpmedicare, main="Figure 2.5 Trellis plot of CCPD versus year", xlab="YEAR", ylab="CCPD", scale = list(x=list(dr
145   put "dev.off()";
146   put "q()";
147   run;
148   /* execute R program prepared just above */
149   /* check on your system that you are using the same version of R (current as of 4/08, or adjust as necessary */
150   options xwait xsync;
151   x ""C:/Program Files/R/R-2.7.1/bin/r.exe" --no-save --quiet <"%path.\program.r" >"%path.\program.log"";
152
153   /* capture/examine the log of R program run into the SAS log */
154 data _NULL_;
155   file "%path.\program.log";
```

Calling R from SAS

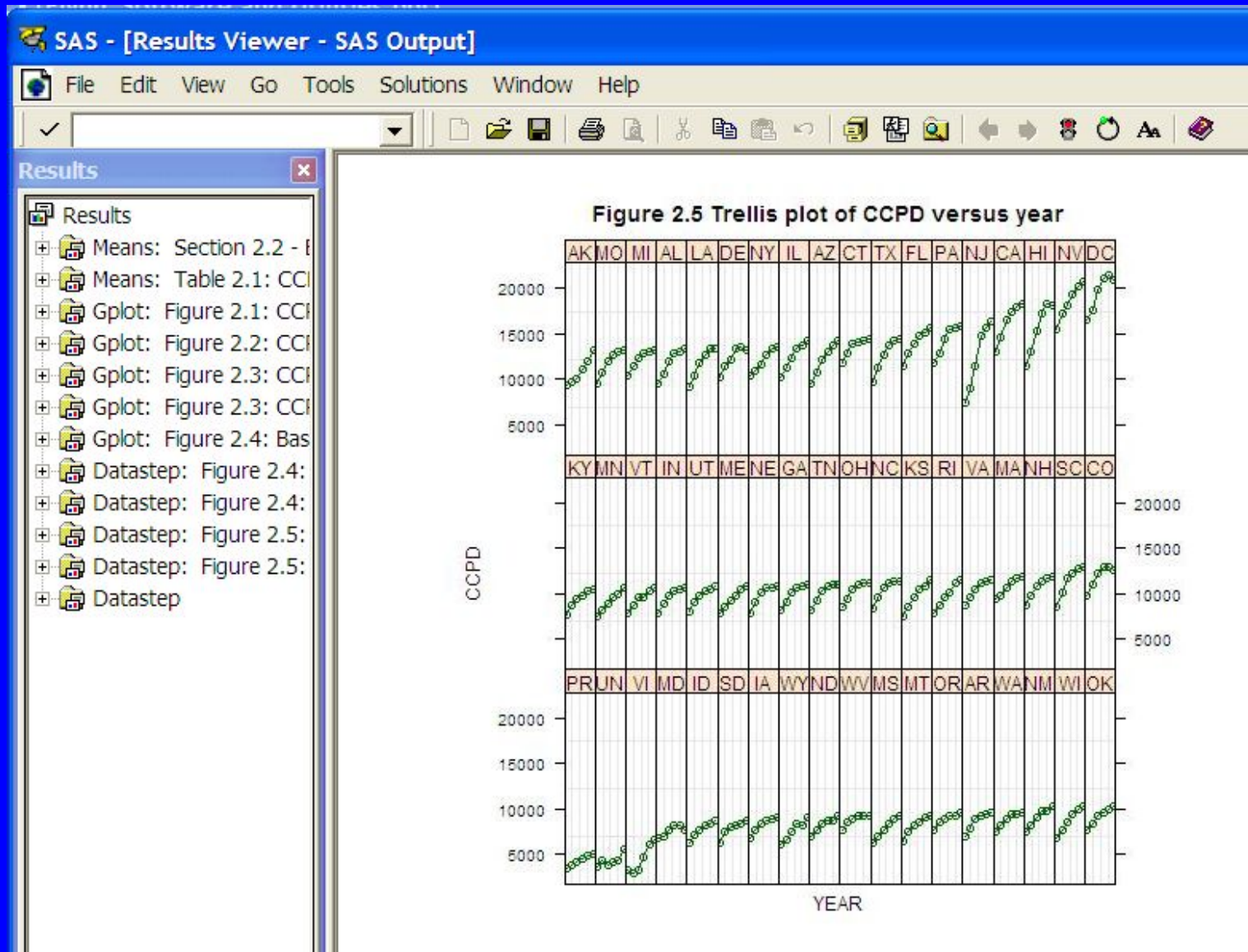


```
SAS - [Chap2Analysis.sas]
File Edit View Tools Run Solutions Window Help

Explorer
Contents of 'SAS Environment'
Name
Libraries
File Shortcuts
Favorite Folders
My Computer

142 put "trellis.par.set(theme=col.whitebg())";
143 put "trellis.par.set('background',list(col='white'))";
144 put 'plot(Grpmedicare, main="Figure 2.5 Trellis plot of CCPD versus year", xlab="YEAR", ylab="CCPD", scale = list(x=list(
145 put "dev.off()";
146 put "q()";
147 run;
148 /* execute R program prepared just above */
149 /* check on your system that you are using the same version of R (current as of 4/08, or adjust as necessary */
150 options xwait xsync;
151 x ""C:/Program Files/R/R-2.7.1/bin/r.exe" --no-save --quiet <"&path.\program.r" >"&path.\program.log""";
152
153 /* capture/examine the log of R program run into the SAS log */
154 data _NULL_;
155   infile "&path.\program.log";
156   file log;
157   input;
158   put '**R: ' _infile_;
159 run;
160
161 OPTIONS FORMCHAR="|----|+|----+|=|<>*";
162 /* open the newly created .jpg file in the SAS results view as RTF and as HTML */
163 ods escapechar='^';
164 ods rtf file="&path.\report.rtf" style=minimal gtitle gfootnote;
165 data _NULL_;
166   file print;
167   put "^S={preimage='&path.\program.jpg'}";
168 run;
169 ods rtf close;
170
171 ods escapechar='^';
172 ods html file="&path.\report.html" style=minimal gpath="&path.\" gtitle gfootnote;
173 data _NULL_;
174   file print;
175   put "<img src='&path.\program.jpg' border='0'>";
176 run;
177 ods html close;
```

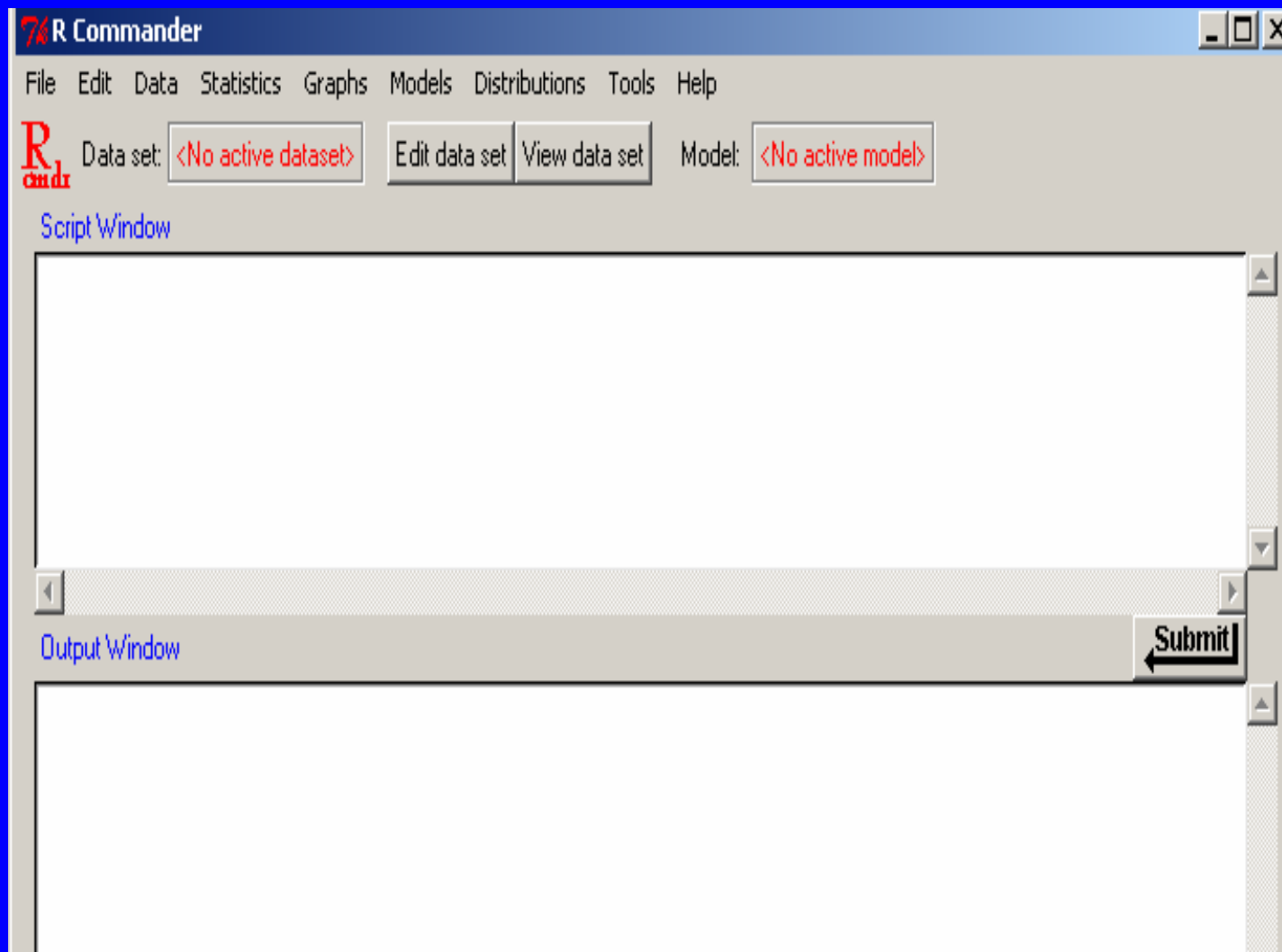
Calling R from SAS



Rcmdr utility

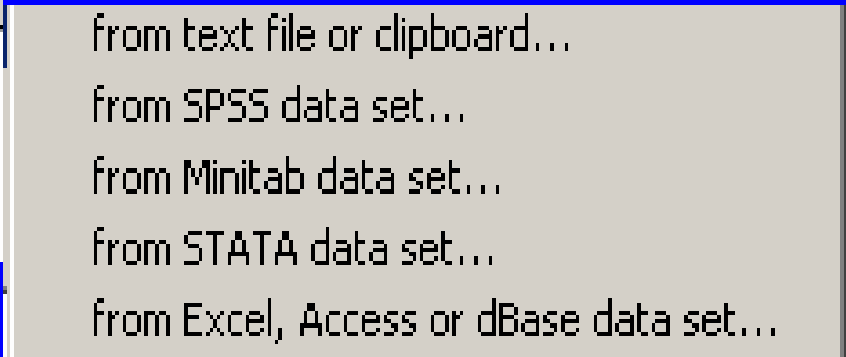
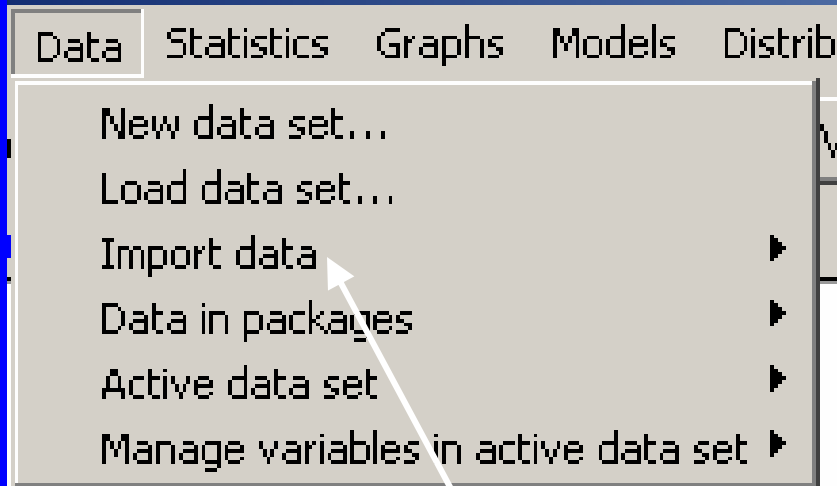
- **A GUI Inter face for R**
- **Developed by John Fox**
- **Makes basic data importing and statistical procedures available by menu**

Rcmdr GUI



Import Data from Other Software (Excel, Access, SPSS) from Rcmdr

mander



Rcmdr

The screenshot displays the Rcmdr R GUI interface. The main window is titled "R Commander" and contains several panes:

- R Console:** Shows the execution of R code. The code includes reading a dataset, fitting a Poisson GLM, and plotting a histogram.
- Script Window:** Contains the R code being executed, including:

```
Dataset <- read.table("//ISOHOMEPI/HOME/I63705/Document/R/bioChemists.csv", head=10, as.is=T)
summary(Dataset)
Dataset <- read.table("//ISOHOMEPI/HOME/I63705/Document/R/bioChemists.csv", head=10, as.is=T)
summary(Dataset)
showData(Dataset, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80, maxheight=100)
GLM.1 <- glm(art ~ fem + kid5 + mar, family=poisson(log), data=Dataset)
summary(GLM.1)
Hist(Dataset$art, scale="frequency", breaks="Sturges", col="darkgray")
```
- Output Window:** Displays the results of the GLM fit:

```
Coefficients:
(Intercept)  0.93045  0.07745 12.013 < 2e-16 ***
fem[T.Women] -0.28549  0.05433 -5.255 1.48e-07 ***
kid5         -0.16118  0.03934 -4.097 4.19e-05 ***
mar[T.Single] -0.13271  0.06092 -2.179  0.0294 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

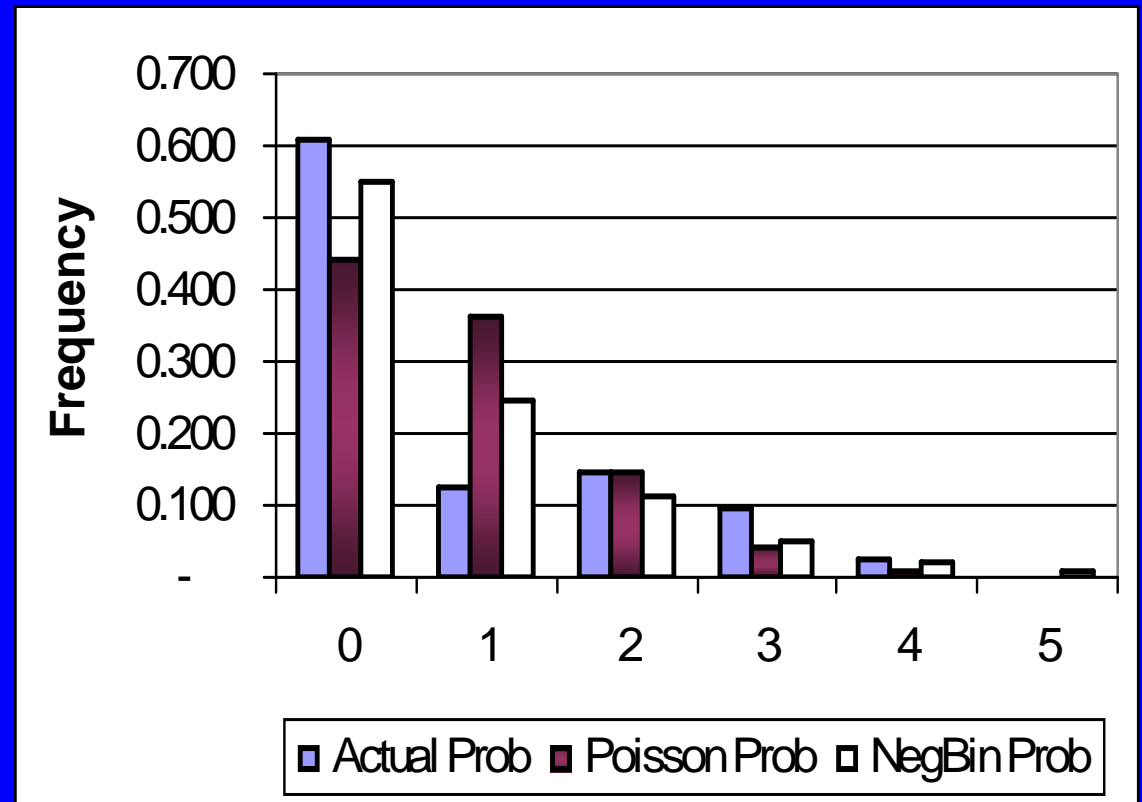
Null deviance: 1817.4  on 914  degrees of freedom
Residual deviance: 1776.7  on 911  degrees of freedom
AIC: 3452.5

Number of Fisher Scoring iterations: 5
```
- Device 2 (ACTIVE):** Shows a histogram of the variable "art" from the dataset. The x-axis is labeled "Dataset\$art" and ranges from 0 to 20. The histogram shows a right-skewed distribution with a peak at 0.

Alternative Distributions: ZIP (zero inflated Poisson)

$$\varphi + (1 - \varphi)e^{-\lambda} \quad x=0$$

$$(1 - \varphi) \frac{\lambda^x}{x!} e^{-\lambda} \quad x > 0$$



Tinn-R: An R Editor

```
Tinn-R - [\\ISOHOMEP1\HOME\i63705\Document\R\rpanel_zaign.r]
File Project Edit Format Marks Insert Search Options Tools R View Window Web Help
R complex
rpanel_zaign.r
# rpanel_zaign.r      # Matt Flynn mflynn@iso.com
#####
# Plots of Zero-Adjusted Inverse Gaussian (ZAIG) distributions
# Heller, Stasinopoulos and Rigby (2006)
# The zero-adjusted Inverse Gaussian distribution as a model for insurance claims, wp
# http://studweb.north.londonmet.ac.uk/~stasinom/papers/ZAIG.pdf
# http://www.gamlss.com/

library(rpanel)      # 1.0-4 Bowman, Bowman and Crawford
library(gamlss)      # 1.8-0 Rigby and Stasinopoulos

plot.zaig <- function(panel) {
  with(panel, {
    x<-seq(0,n,0.01)
    if (dist == "Inverse Gaussian"      ") {
      nu<-0;
    }
    y <- dZAIG(x, mu, sigma, nu)
    e_y <- (1-nu)*mu
    var v <- (1-nu)*mu^2*(nu+mu*sigma^2)
  }
}
```

Lin 9/43: Col 1 | Normal mode smNormal | Size: 1.95 KB | Tinn-R hotkeys inacti

RExcel

- **An Add-in for Excel**
- **Lets you run R from within Excel**
- **You must first run an execute file from R to install it**

RExcel

The screenshot shows the Microsoft Excel interface with the RExcel add-in menu open. The menu options are:

- Close R
- Run Code
- Get R Value
- Put R Var
- Get R output
- Set R working dir
- Load R file
- Copy Code
- Debug R
- Error Log
- Options
- Set R server
- RExcel Help
- R Help
- RCommander
- Demo Worksheets
- About RExcel

The worksheet contains the following R code in cells A13 to A15:

```
A13: za <- 1:12
A14: za <- matrix(za,3,4)
A15: zb <- za%%t(za)
```

The worksheet also contains instructional text in cells A9 to A27, and a data table in cells F13 to H27.

| | F | G | H |
|----|----|----|-----------|
| 13 | | | |
| 14 | | | |
| 15 | | | |
| 16 | | | |
| 17 | | | |
| 18 | | | |
| 19 | | | t(za)%%za |
| 20 | | | |
| 21 | | | |
| 22 | | | |
| 23 | | | |
| 24 | | | |
| 25 | 13 | 17 | 23 |
| 26 | 23 | 11 | 121 |
| 27 | 2 | 3 | 1 |

RPanel & ZIGP

```
# rpanel_zigp.r ····· # Matt Flynn mflynn@iso.com
#####
# ··· Plots of Zero-Inflated Generalized Poisson (ZIGP) distributions
# Czado, Claudia and Aleksey Min
# Zero-inflated generalized Poisson regression models:
# Asymptotic theory and applications
# Munich University of Technology WP 2005
# http://www-m4.ma.tum.de/Papers/Min/Czado-Min.pdf
#
library(rpanel) ····· # 1.0-4 Bowman, Bowman and Crawford
library(ZIGP) ····· # 1.2 Vinzenz Erhardt
#
plot.zigp <- function(panel) {
  with(panel, {
    if (dist == "Poisson") {
      phi <- 1; omega <- 0;
    } else
    if (dist == "Generalized Poisson") {
      omega <- 0;
    } else
    if (dist == "ZIP") {
      phi <- 1;
    }
    probs <- dzigp(0:n, mu, phi, omega)
    plot(0:n, probs, type='n', ylim=c(0,0.5), xlab="x", ylab="Probability")
    segments(0:n, rep(0, n+1), 0:n, probs, col="blue", lwd=3)
    abline(h=0, col="grey")
    mutext <- as.character(round(mu, 3))
    phitext <- as.character(round(phi, 3))
    omegatext <- as.character(round(omega, 3))
    title(paste("Dist:", dist, ", mu=", mutext, ", phi=", phitext, ", omega=", omegatext))
    text(10,0.4, expression(P(Y==y, mu, phi, omega) == (1[ 'x=0' ])*(omega + (1 - omega)*exp(frac(-mu, phi))) + " ····· "))
    text(10,0.35, expression((1[x>0])*(1 - omega)*frac(mu*(mu + (phi - 1)*x)^(x - 1), "x!")*phi^(-x)*))
  })
}
```

actuar

- **Package of actuarial functions**
 - Loss distributions
 - Risk theory
 - Some kinds of simulation
 - Credibility theory

RPanel & ZIGP

```
] ... with(panel, {  
] ..... if (dist == "Poisson") {  
] ..... phi <- 1; omega <- 0;  
] ..... } else  
] ..... if (dist == "Generalized Poisson") {  
] ..... omega <- 0;  
] ..... } else  
] ..... if (dist == "ZIP") {  
] ..... phi <- 1;  
] ..... }  
] ..... probs <- dzigp(0:n, mu, phi, omega)  
] ..... plot(0:n, probs, type='n', ylim=c(0,0.5), xlab="x", ylab="Probability")  
] ..... segments(0:n, rep(0, n+1), 0:n, probs, col="blue", lwd=3)  
] ..... abline(h=0, col="grey")  
] ..... mutext <- as.character(round(mu, 3))  
] ..... phitext <- as.character(round(phi, 3))  
] ..... omegatext <- as.character(round(omega, 3))  
] ..... title(paste("Dist:", dist, ", mu=", mutext, ", phi=", phitext, ", omega=", omegatext))  
] ..... text(10, 0.4, expression(P(Y==y, mu, phi, omega) == (1[x=0]) * (omega + (1 - omega) * exp(frac(-mu, phi))) + "....."))  
] ..... text(10, 0.35, expression((1[x>0]) * (1 - omega) * frac(mu * (mu + (phi - 1) * x)^(x - 1), "x!") * phi^(-x) *  
] ..... exp(frac(-1, phi) * (mu + (phi - 1) * x))))  
] ..... }  
] ..... panel  
] ..... }  
panel <- rp.control("ZIGP Parameters", dist="ZIP", n=20, mu=2, phi=1.5, omega=0.2)  
rp.slider(panel, mu, 0.01, 20,  
] ..... title = "mu:", action = plot.zigp)  
rp.slider(panel, phi, 1, 3,  
] ..... title = "phi:", action = plot.zigp)  
rp.slider(panel, omega, 0, 1,  
] ..... title = "omega:", action = plot.zigp)  
rp.radiogroup(panel, dist, c('ZIP', 'Poisson', 'Generalized Poisson', 'ZIP'), action=plot.zigp, title='Dist')
```

RPanel

ZIG...

mu:

phi:

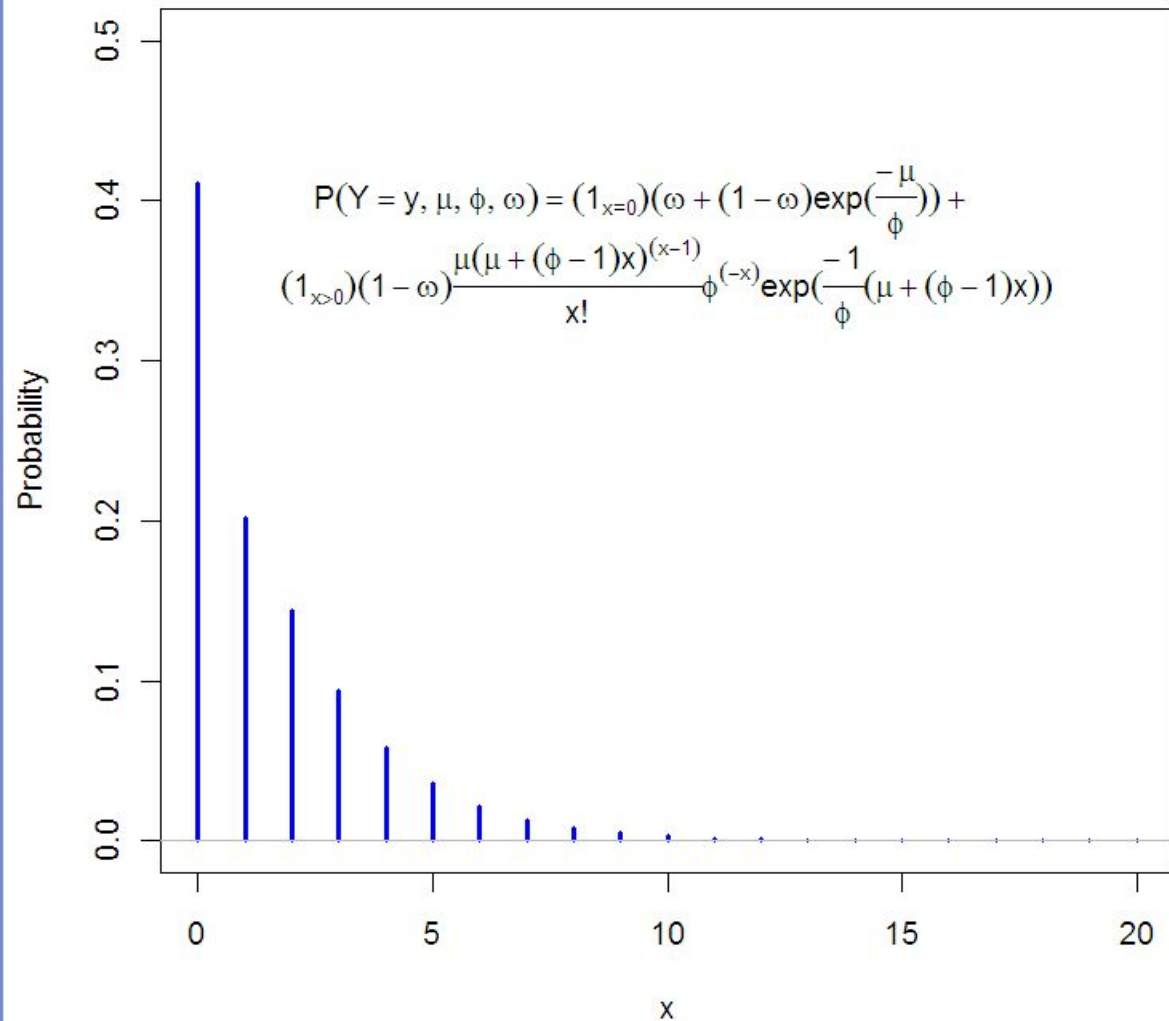
omega:

Dist

- ZIGP
- Poisson
- Generalized Poisson
- ZIP

R Graphics: Device 2 (ACTIVE)

Dist: ZIGP , mu = 2 , phi = 1.5 , omega = 0.2



iPlots: Interactive Plots



JGR – “Jaguar”



The screenshot shows the homepage of the JGR website. At the top, there is a navigation menu with links for "JGR", "Features", "Screenshots", "Download", "JGR Installation", "JGR on Linux", and "FAQ". Below the navigation is a header section with the JGR logo (a jaguar head) and the text "JGR - Java GUI for R". To the right of the header is a small graphic of four colored squares (yellow, red, blue, green). The main content area contains a paragraph describing JGR as a universal and unified GUI for R, mentioning its introduction at the useR! meeting in 2004 and an introductory article in the Statistical Computing and Graphics Newsletter. Below this is a "What's new?" section with a list of updates from 2008, including releases of JGR 1.6-2, 1.6-1, and 1.6, and the launch of the new JGR webpage. At the bottom, there is a link to "JGR's friends at RoSuDa.org" and an Apple logo with the text "Made on a Mac".

[JGR](#) [Features](#) [Screenshots](#) [Download](#) [JGR Installation](#)
[JGR on Linux](#) [FAQ](#)

JGR - Java GUI for R

JGR (speak 'Jaguar') is a universal and unified Graphical User Interface for R (it actually abbreviates **J**ava **G**ui for **R**). JGR was introduced at the **useR!** meeting in **2004** and there is an introductory article in the **Statistical Computing and Graphics Newsletter Vol 16 nr 2** p9-12

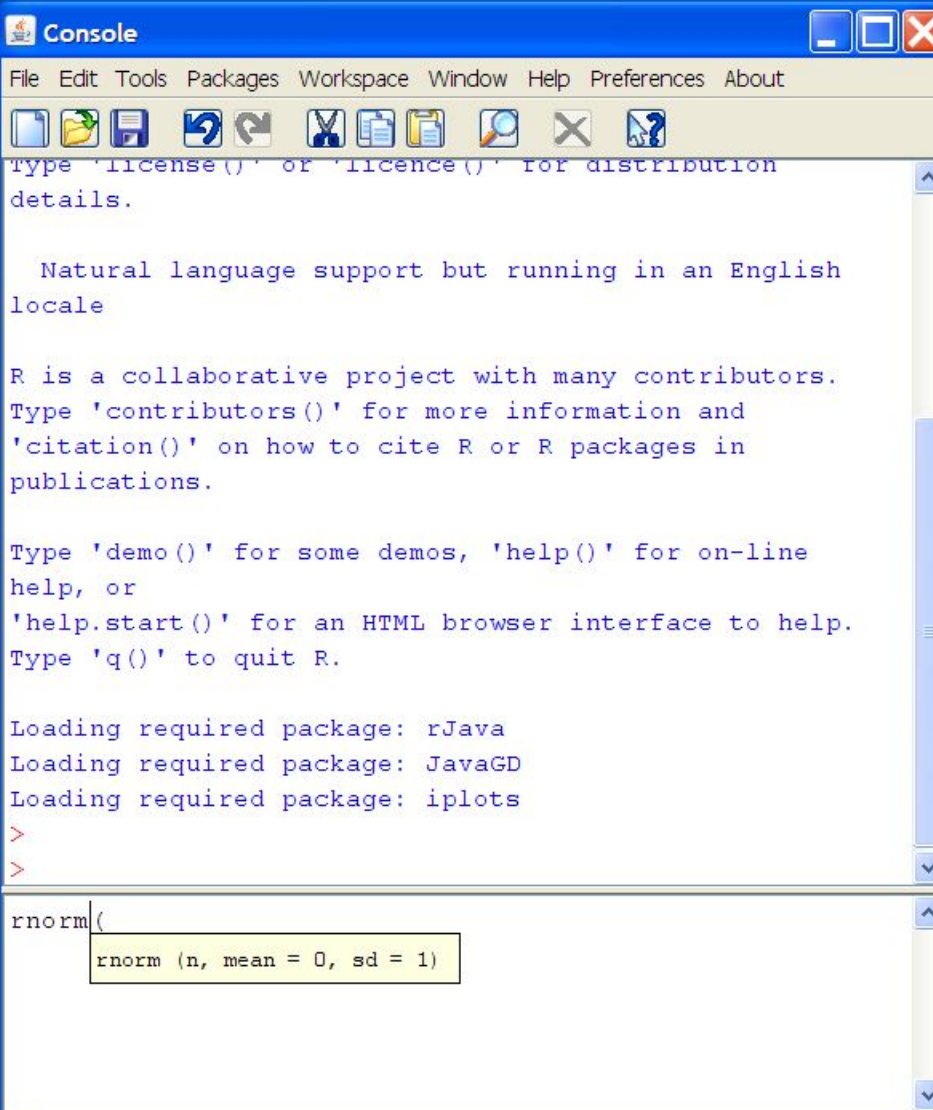
What's new?

- 2008/07/24** JGR 1.6-2 and new launchers 1.61 released
 - please download the new launchers
- 2008/06/18** JGR 1.6-1 released
 - minor fix in editor and object-browser
- 2008/06/15** JGR 1.6 and new launchers released
 - please delete your .JGRprefs file in your home directory before and download the new launchers
- 2008/06/15** Launch of new JGR webpage

JGR's friends at [RoSuDa.org](#)


Made on a Mac

JGR – “Jaguar”



```
Console
File Edit Tools Packages Workspace Window Help Preferences About
Type 'license()' or 'licence()' for distribution
details.

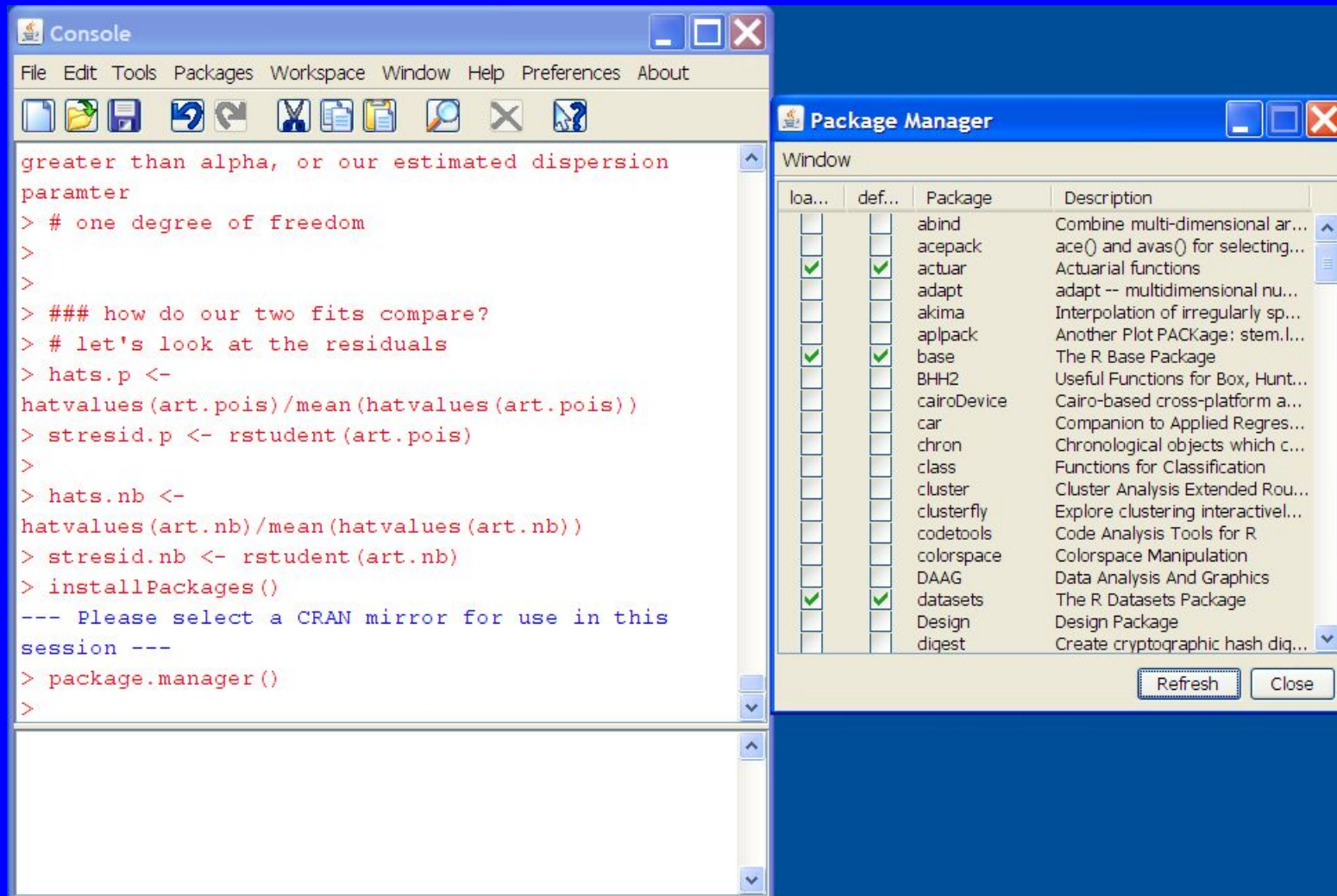
Natural language support but running in an English
locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in
publications.

Type 'demo()' for some demos, 'help()' for on-line
help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Loading required package: rJava
Loading required package: JavaGD
Loading required package: iplots
>
>
rnorm(
  rnorm (n, mean = 0, sd = 1)
```

JGR – “Jaguar”



The screenshot displays two windows from the JGR (Jaguar) software interface. The 'Console' window on the left contains R code for fitting a Poisson regression model and installing packages. The 'Package Manager' window on the right shows a list of available packages with checkboxes for selection.

```
greater than alpha, or our estimated dispersion
paramter
> # one degree of freedom
>
>
> ### how do our two fits compare?
> # let's look at the residuals
> hats.p <-
hatvalues(art.pois)/mean(hatvalues(art.pois))
> stresid.p <- rstudent(art.pois)
>
> hats.nb <-
hatvalues(art.nb)/mean(hatvalues(art.nb))
> stresid.nb <- rstudent(art.nb)
> installPackages()
--- Please select a CRAN mirror for use in this
session ---
> package.manager()
>
```

| loa... | def... | Package | Description |
|-------------------------------------|-------------------------------------|-------------|------------------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | abind | Combine multi-dimensional ar... |
| <input type="checkbox"/> | <input type="checkbox"/> | acepack | ace() and avas() for selecting... |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | actuar | Actuarial functions |
| <input type="checkbox"/> | <input type="checkbox"/> | adapt | adapt -- multidimensional nu... |
| <input type="checkbox"/> | <input type="checkbox"/> | akima | Interpolation of irregularly sp... |
| <input type="checkbox"/> | <input type="checkbox"/> | aplpack | Another Plot PACKage: stem.l... |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | base | The R Base Package |
| <input type="checkbox"/> | <input type="checkbox"/> | BHH2 | Useful Functions for Box, Hunt... |
| <input type="checkbox"/> | <input type="checkbox"/> | cairoDevice | Cairo-based cross-platform a... |
| <input type="checkbox"/> | <input type="checkbox"/> | car | Companion to Applied Regres... |
| <input type="checkbox"/> | <input type="checkbox"/> | chron | Chronological objects which c... |
| <input type="checkbox"/> | <input type="checkbox"/> | class | Functions for Classification |
| <input type="checkbox"/> | <input type="checkbox"/> | cluster | Cluster Analysis Extended Rou... |
| <input type="checkbox"/> | <input type="checkbox"/> | clusterfly | Explore clustering interactivel... |
| <input type="checkbox"/> | <input type="checkbox"/> | codetools | Code Analysis Tools for R |
| <input type="checkbox"/> | <input type="checkbox"/> | colorspace | Colorspace Manipulation |
| <input type="checkbox"/> | <input type="checkbox"/> | DAAG | Data Analysis And Graphics |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | datasets | The R Datasets Package |
| <input type="checkbox"/> | <input type="checkbox"/> | Design | Design Package |
| <input type="checkbox"/> | <input type="checkbox"/> | digest | Create cryptographic hash dig... |

JGR – “Jaguar”

The screenshot displays the JGR (Jaguar) R environment interface. On the left is the Console window, and on the right is the Object Browser window.

Console Window: Shows the following R code and output:

```
paramter
> # one degree of freedom
>
>
> ### how do our two fits compare?
> # let's look at the residuals
> hats.p <-
hatvalues(art.pois)/mean(hatvalues(art.pois))
> stresid.p <- rstudent(art.pois)
>
> hats.nb <-
hatvalues(art.nb)/mean(hatvalues(art.nb))
> stresid.nb <- rstudent(art.nb)
> installPackages()
--- Please select a CRAN mirror for use in this
session ---
> package.manager()
> object.browser()
>
```

Object Browser Window: Shows the structure of the 'data' object. The 'articles' data frame (dim(915:7)) is expanded, showing variables: n (integer), art (integer), fem (factor) levels: 2, mar (fact), kid5 (int), phd (nume), and ment (integer). A tooltip for the 'art' variable displays summary statistics:

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max |
|-----|-------|---------|--------|-------|---------|-------|
| art | 0.000 | 0.000 | 1.000 | 1.693 | 2.000 | 19.00 |

Buttons at the bottom of the Object Browser: Save Data, Refresh, Close.

JGR – “Jaguar”

The screenshot shows the JGR (Jaguar) software interface. The Console window displays the output of a statistical model fit, including deviance, AIC, and parameter estimates. The Object Browser window shows a table of model results for the 'art.pois' model.

```
.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(2.2644) family taken to

Null deviance: 1109.0 on 914
freedom
Residual deviance: 1004.3 on 909
freedom
AIC: 3135.9

Number of Fisher Scoring iterations: 4

          Theta:  2.264
        Std. Err.:  0.271

2 x log-likelihood:  -3121.917
>
```

| Name | Data | Type | family | df | r.squar... | aic | deviance |
|----------|----------|------|---------|----|------------|---------|----------|
| art.pois | articles | glm | poisson | 6 | | 3314.11 | 1634.37 |

Helpful websites

The mothership

<http://cran.r-project.org/>

R FAQ

<http://cran.r-project.org/doc/FAQ/R-FAQ.html>

An introduction to R

<http://cran.r-project.org/doc/manuals/R-intro.pdf>

R News



http://cran.r-project.org/doc/Rnews/Rnews_2008-1.pdf

Kickstarting R

<http://cran.r-project.org/doc/contrib/Lemon-kickstart/index.html>

A Brief Guide to R for Beginners in Econometrics

http://people.su.se/~ma/R_intro/

R Wiki

<http://wiki.r-project.org/rwiki/doku.php>

Contact info:

- **-Matt Flynn**

- Office: (860) 570-0015
- Mobile: (203) 435-7362
- Fax: (201) 748-1900
- Email: mflynn@iso.com

- **Louise Francis**

- 215-923-1567
- Louise_francis@msn.com