ISSUE 35 | APRIL 2010

SOCIETY OF ACTUARIES Technology Section

CompAct

ELECTRONIC NEWSLETTER

CONTENTS

QUICK LINKS

R CORNER–GRAPHICS

by Steve Craighead

Editor's note: R Corner[1] is a series by Steve Craighead introducing readers to the "R" language used for statistics and modeling of data. The first column was published in the October 2008 issue, and explains how to download and install the package, as well as providing a basic introduction to the language. Refer to each CompAct issue since then for additional articles in the series. The introductory article can be found on p. 24 of the October 2008 issue on the SOA Web site.

The graphical capabilities of R are extensive. R is truly an object-oriented language with classes built upon classes. Most classes that are available in R and its associated packages already have graphical methods. The majority of these are simply the standard method "plot." This method may be all that you need. For example Markus Gesmann has designed the ChainLadder package to determine insurance claims reserves. You can examine how R can be used graphically by installing the ChainLadder Package using the various Package options from R.

After loading the package, use the following commands (borrowed and modified from the Bootstrap-Chain-Ladder Model help screen):

```
B <- BootChainLadder(RAA, R=999, process.distr="gamma")
B
#Here is the ubiquitous plot() method
plot(B)
```

Section Specialist

Sam Phillips, Staff Editor

```
# Compare to MackChainLadder
MackChainLadder(RAA)
quantile(B, c(0.75,0.95,0.99, 0.995))
```

See Figure 1 for the result of these commands.

To demonstrate a few of the other capabilities of graphics, we will plot the empirical density of the total IBNR and the density of the data fit to a LogNormal distribution use these commands:

```
#Notice how the title is added by the "main" variable in the plot
command
plot(density(B$IBNR.Totals),main="Empirical IBNR Density with
LogNormal")

# fit a distribution to the IBNR library(MASS)

fit <- fitdistr(B$IBNR.Totals[B$IBNR.Totals>0], "lognormal")
#display the fit results to the R window
fit

#now let's layer-on the fitted lognormal curve onto the graph.
Observe the use
#of "col=red" to plot the lognormal distribution and the "add=TRUE"
to overlay
#the curve on the empirical density
curve(dlnorm(x,fit$estimate["meanlog"], fit$estimate["sdlog"]),
col="red", add=TRUE)
```

R displays the fit statistics:

meanlog	sdlog
10.821176062	0.374878187
(0.011860621)	(0.008386725)

Figure 2 displays the empirical distribution with the LogNormal distribution fitted to the data.

To change the axis labels, you use the "xlab" and "ylab" commands within the plot command as shown here:

```
plot(density(B$IBNR.Totals),main="Empirical IBNR
Density",xlab="Sorted IBNR Data",ylab="Probability")
```

See Figure 3, for the results

If you want to add a legend to the graphic you can use these commands:

```
#replot the original plot(density(B$IBNR.Totals),main="Empirical IBNR Density",xlab="Sorted IBNR Data",ylab="Probability")
#add the overlay
curve(dlnorm(x,fit$estimate["meanlog"], fit$estimate["sdlog"]), col="red", add=TRUE)
#add the legend
#the first two numbers give the location of the legend within the plot box, since the x values range from 0 to over 150K #and the y values range from 0 to just over 2.0e-5, the legend will fit on an empty area of the plot
#the following c("Empirical","LogNormal") list gives the elements in the legend
#the two setting for col=c(1,2) will make sure that the line style will have the colors corresponding to the plot.
#the lty=c(1,1) sets the line type of the graph to be a solid line. Other line types can be used for dashed lines, etc.
legend(100000,0.000015, c("Empirical", "LogNormal"), col = c(1,2),lty=c(1,1))
```

Now observe the legend in Figure 4.

For further examples and features that can be set with the legend command type:

```
library(help="legend")
```

For finer control you can control any component on the graphic, by changing other features documented for the "plot" command or by using the "par" command. Especially pay attention to the "mfrow" setting in the par command. For instance if you execute "par(mfrow=c(2,2))" you can place 4 graphs on one graphics device like in figure 1 above. "par(mfrow=c(1,1)" restores the graphic device to the original setting.

If you want to layer two separate "plot" commands on top of each other, use the "par(new=TRUE)" command between the two plots (this forces the second plot not to clear the graphic. To make both plots have the same xy ranges, you would use the "xlim=c(,)" and "ylim=c(,)" construct in both plots.

For my last example, I will limit the display of the data on the x-axis by using "xlim" command:

```
plot(density(B$IBNR.Totals),main="Empirical IBNR
Density",xlim=c(50000,150000))
curve(dlnorm(x,fit$estimate["meanlog"], fit$estimate["sdlog"]),
col="red", add=TRUE,xlim=c(50000,150000))
```

Figure 5 Restricted X Axis

Graphic Resources

As I said before, the graphical capabilities of R are extensive and I have barely scratched the surface. In this last section, I'll leave you with several excellent additional resources that lead you to discover more of the capabilities of R.

J. H. Maindonald has written an excellent article titled, "Using R for Data Analysis and Graphics Introduction, Code and Commentary." It is available here. In it he instructs the reader on the fundamentals of R as well as demonstrates the graphics capability of R.

Nicholas Lewin-Koh maintains the Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization CRAN Task View. On this web site, he outlines the major topics of how R handles graphics, for both the user and the developer.

The home page of the R Graph Gallery is here. You can enter the gallery by choosing one of four different collections on the right-hand side of the Web page. Once you choose a gallery and then choose an example graphic, the Web page will display the graphic and under Requirements on the left-hand side of the screen, you can choose to either download the associated R code or examine the code in your Internet browser.

The last resource that I recommend is the R Graphics Manual, which demonstrates the collection of R graphics from the extensive list (now over 2000) of all R packages. The link is here.

Conclusions

This article is the last basic article on R. In future articles I will model various actuarial models or discuss the use of especially promising R packages. My next article will be a toy ERM model demonstrating the use of R in modeling copulas for the aggregation of stand-alone capital distributions to determine the diversified (economic) capital distribution at the enterprise level.

Steve Craighead, ASA, CERA, MAAA is a consultant at TowersWatson. He can be contacted at steve.craighead@towerswatson.com.

1Craighead, S. (2000), "Insolvency Testing: An Empirical Analysis of the Generalized Beta Type 2 Distribution, Quantile Regression, and a Resampled Extreme Value Technique," ARCH, pp. 13–149.

Actuaries
Risk is Opportunity.®

475 North Martingale Road, Suite 600 Schaumburg, Illinois 60173
Phone: 847.706.3500 Fax: 847.706.3599 www.soa.org