Skip Navigation Links

Page One

Campus Computing <u>News</u>

Academic Mainframe Services to be Terminated

This Just In . . .

Conference Proceedings Available Online

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks <u>Online</u>

Research and Statistical Support University of North Texas

RSS Matters

This article is a reprint from the March 2000 <u>issue</u> of Benchmarks Online. The title has been changed slightly. Link to the last RSS article here: <u>A New Face in RSS</u>. Link to the last RSS article in this series here: <u>Interactive Graphics in R (Part II - cont.)</u>: <u>Kernel Density Estimation in One and Two Dimensions</u>. - Ed.

Resampling Based Statistics in S-Plus for Windows: An Example Using the MANOVA Procedure

By Dr. Rich Herrington, Research and Statistical Support Services Manager

This month we take a look at the bootstrap resampling capabilities of S-Plus. S-Plus has general bootstrapping functionality available so that nearly all statistical functions and expressions can be bootstrapped. S-Plus provides both parametric and nonparametric bootstrap confidence intervals.

From the main menu bar, we access the resampling menu from: Statistics - Resample - Bootstrap.

Statistics	<u>G</u> raph Options	Mindow Help	
<u>⊅</u> ata Sumn <u>C</u> ompare S Power and		Linear	
_ ⊅esi <u>a</u> n) Commands	
<u>R</u> egressio	n	king data will be	ir
ANOVA		<pre> err_Data </pre>	
Mixed <u>E</u> fi	fects	•	
Generaliz	ed <u>L</u> east Squares	5 🕨	
S <u>u</u> rvival		•	
Iree		1980. 1980.	
C <u>o</u> mpare M	lodels	2	
<u>C</u> luster /	nalysis	- > • •	
<u>M</u> ultivari	late	- 1	
<u>@</u> uality (Control Charts	<u>+ .</u>	13
		<u>B</u> ootstrap	
<u>S</u> moothing	1	▶ <u>J</u> ackknife	
T <u>i</u> me Seri	les	•	1.1

The menu for the Bootstrap facilities has five entry areas for initializing the Bootstrap analysis: **Model, Options, Results, Plot,** and **Jack After Boot**. Each of these option tabs are initialized with default values. However, the critical entry field which does not have a default entry is the **Expression** entry field. Entering an expression to bootstrap can be tricky as this assumes that the user has some knowledge of the syntax of the S-Plus language.

Model	Options	Results	Plot	Jack After Boo
Data Data Set:	randsamp	<u>Save As</u>		Indiana
Statistic to Est	imate	I Sav	e Resampling	Indices
Expression:				
Cublic assicit:				
Zubi Cassion 1.				
-opression.				
-obression:				
-obression:				

One way of avoiding having detailed knowledge of the syntax used to generate a particular analysis, is to generate the analysis before hand from the drop down menu system. Once this analysis has been run, the syntax used to generate the analysis is displayed. Essentially, the drop down menu system generates the syntax as entry fields are filled in. After an analysis is run from the menu system, this syntax can be saved, cut and pasted back into the **Expression** entry field. In the following example we will perform a four-group MANOVA with four dependent measures.

Example

The data set we will use for our analysis will have four groups: a control group and three experimental groups (c1, e1, e2, e3). We see a screen capture of the object browser and the data worksheet:

皇皇皇 ‰ ⅔ 『 ¥0bject Explorer1	₩•₩₩ <u>₩</u> ΞΞ		ੈ Z → Å					
Match (All)	Contents of: Data	S]	anova		é	-	a 71	
🕀 🕞 Data	Object	Ē		1	2	3	4	5
Graphs	begin.alge		15 - 18 E	q1	q2	q3	q4	group
Scripts	🔄 cars.data	1	1	29.00	40.00	41.00	40.00 <mark>c</mark>	
	dropout.surv	2	2	31.00	38.00	45.00	35.00 c	
	fit fit	3	3	31.00	41.00	42.00	36.00 c	
	🕞 manova	4	4	21,00	53.00	45.00	31,00 c	
	andsamp	5	5	45.00	40.00	35.00	30.00 c	
	wafer.manova	6	6	34.00	39.00	35.00	42.00 c	
		7	7	38.00	53.00	36.00	43.00 c	
		8	8	30.00	45.00	35.00	40.00 c	
		9	9	42.00	42.00	39.00	27.00 c	
		10	10	39.00	41.00	37.00	33.00 c	
		11	11	45.00	34.00	36.00	35.00 c	
	1 H H							
	1							

From the main menu bar select: **Statistics - Multivariate - MANOVA**. Select the **Create Formula** tab. Fill out the create formula tab with the following specifics. First select q1 through q4 and click **Add Response**. Then select group and click **Add Main Effect**:

ormula			
Variable		Special Term	
	q1 q2 q3 q4 group	Term Category: <u>Option:</u> Format:	*
Iransformation		Add	
Add		Remove	
<u>R</u> esponse	Main Effect: (•	F Remove Intercept	
Interaction: (;)	Mai <u>n</u> +Interact.:		oup 💌
с	Alighted sectors		
Eormula:	cbind(q1,q2,q3,	q4) gioup	
OK Cancel	Apply K	current	Help

Select **OK** to return to the previous menu. Select **OK** once more to actually run the analysis. In the report window we see the following:

```
*** Multivariate Analysis of Variance Model ***
Short Output:
Call:
  manova(formula = cbind(ql, q2, q3, q4) ~
      group, data = manova, na.action =
      na.exclude)
Terms:
               group Residuals
Deg. of Freedom
                          123
                  3
Estimated effects may be unbalanced
Analysis of Variance Table:
          Df Pillai Trace approx. F
    group 3 0.0775 0.8089
Residuals 123
        num df
                 den df
                            P-value
   group
            12
                     366
                            0.6413
Residuals
```

The calling function is listed under **Call**. Copy the **manova(formula....)** and paste this into your **Commands window**. Use the **summary** function to summarize the call to the manova function. Assign this summary to an object, **man.out**, for example:

```
> man.out<-summary(manova(formula = cbind(ql, q2, q3, q4) ~
       group, data = manova, na.action =
       na.exclude))
> man.out
         Df Pillai Trace approx. F num df
                                             den df
                                                      P-value
   group 3 0.0775
                         0.8089
                                       12
                                                      0.6413
                                              366
Residuals 123
names(man.out)
                 "$$"
                                "Df"
                                               "Eigen.values" "Stats"
                                                                            "test"
[1] "row.names"
```

Typing **man.out** by itself displays the contents of this object.**names** displays the components of this list. We have six components to this list. To extract the fifth element "**Stats**". We have to index the list in the following fashion:

```
> man.out[[5]]
 , Statistic
         pillai wilks lambda hotelling-lawley roy largest
group 0.07751211 0.9240134 0.08059128 0.05130897
, , approx. F
        pillai wilks lambda hotelling-lawley roy largest
                0.8029447
                              0.7969582 1.564923
group 0.8089407
, , num df
     pillai wilks lambda hotelling-lawley roy largest
group
        12
                   12
                                     12
                                                 4
, , den df
    pillai wilks lambda hotelling-lawley roy largest
       366
             317.7817
                                   356
group
                                              122
, , P-value
        pillai wilks lambda hotelling-lawley roy largest
group 0.6412999 0.6474587
                            0.6537033 0.1879375
> man.out[[5]][2]
[1] 0.9240134
> man.out[[5]][3]
[1] 0.08059128
> man.out[[5]][4]
[1] 0.05130897
> man.out[[5]][5]
[1] 0.8089407
> man.out[[5]][12]
[1] 4
ÞI
```

We see that wilks lambda (.9240) is the second index for the fifth element of the list, **man.out**. So the complete calling function to the bootstrap function will be:

```
> summary(manova(formula = cbind(q1, q2, q3, q4) ~
+ group, data = manova, na.action =
+ na.exclude))[[5]][2]
[1] 0.9240134
> |
```

This calling function returns a value of .9240 for wilks lambda for this particular data set. We need to copy this function call: **summary(manova.....))[[5]][2]**, into the **Expression** window on the bootstrap menu.

Model	Options	Results	Plot	Jack After Bo
Data <u>D</u> ata Set:	manova	Save N	lodel Object- s:	
Statistic to Est	imate	3		
Statistic to Est	10.0000	novalformula – chir	odial a2 a3	all ~ aroun dat
	10.0000	nova(formula = cbir	nd(q1, q2, q3,	q4) ~ group, dat
	10.0000	nova(formula = cbir	nd(q1, q2, q3,	q4) ~ group, dat
Statistic to Esl Expression:	10.0000	nova(formula = cbir	nd(q1, q2, q3,	q4) ~ group, dat
	10.0000	nova(formula = cbir	nd(q1, q2, q3,	q4) ~ group, dat
	10.0000	nova(formula = cbir	ıd(q1, q2, q3,	q4) ~ group, dat

Data Data Set: manova Save Model Object Save As: Save Resampling India Statistic to Estimate	ndices	2000-2000-000-00 <u>7</u>			Data
	ndices			manova	
Statistic to Estimate		Resampling Inc	☐ Sav		
			1977	mate	Statistic to Est
Expression: $[, q4] \sim$ group, data = manova, na.action = na.excl	xclude))[[5]][2]	a.action = na.ex	, data = manova, r	, q4) ~ group	Expression;

For the **Options** tab we need to select the grouping variable and how many bootstrap iterations we need:

Options			>
opuons	Results	Plot	Jack After Boo
is			
iles:			
100			
group	<u>×</u>		
eed:			
, 			
umbers			
led Data to Fr	ame 1		
	les: 100 group eed: umbers	les: 100 group 💌 eed:	les: 100 group 💌 eed: umbers:

For the **Results** tab we select **empirical percentiles**:

Model	1 October	Results	Plot	Jack After Boo
Moder	Options	nesuits	EI01/	Jack Aner Dou
Printed Result	ts	1,000,000	ntile Options	
✓ Summary S	Statistics	Percer	ntile Levels: 🛛 🖸	(0.025,0.05,0.95,
Empirical F	Second and		8 3	
	rercentiles			
✓ <u>B</u> Ca Perce	entiles			
- Corrolation	Matrix of Estimate			
	Matrix of Estimate	S:		
			-	

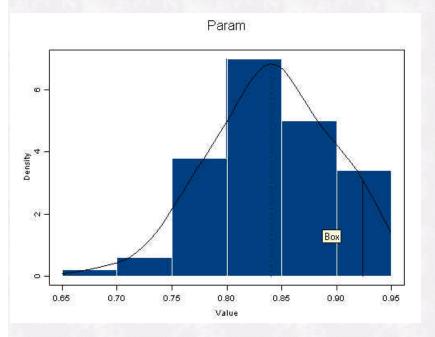
For the **Plot** tab we select **Normal Quantile-Quantile** to see how well the sampling distribution matches with "normal distribution" theory.

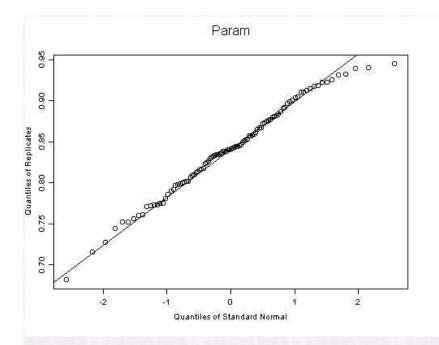
Model	Options	Results	Plot	Jack After Bool
Plots		· · · · · · · · · · · · · · · · · · ·		1
Distribution	of Replicates			
Normal Qui	antile-Quantile			
Samana Silan	namen and a second s			

Selecting **OK** generates the following report window:

```
*** Bootstrap Results ***
Call:
bootstrap(data = manova, statistic = summary(manova(formula = cbind(q1, q2, q3, q4) ~ group, data
       = manova, na.action = na.exclude))[[5]][2], B = 100, group = group, trace = F,
       assign.framel = F, save.indices = F)
Number of Replications: 100
Summary Statistics:
     Observed Bias Mean
                                   SE
Param 0.924 -0.08357 0.8404 0.05491
Empirical Percentiles:
          2.5%
                     5%
                               95%
                                       97.5%
Param 0.7348601 0.7518553 0.9250432 0.9352212
BCa Percentiles:
                      5$
                               95%
                                       97.5%
          2.5%
Param 0.9100873 0.9213088 0.9448852 0.9448857
```

And the following plots:





We see that the empirically resampled sampling distribution for wilks lambda follows normal theory fairly closely except for the right tail region. We see that the upper and lower cut-offs for the 2.5/97.5th and 5/95th percentiles both contain the observed value of wilks lambda. We take this as a failure to reject the null hypothesis for wilks lambda. In general the BCa percentiles will be more accurate than the empirical percentiles.

Further Reading

Davison, A.C. and Hinkley, D.V. (1997). Bootstrap Methods and Their Application. Cambridge University Press.

Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap. San Francisco: Chapman & Hall.