

Advanced Use Cases of the Bootstrap Feature in JMP[®] Pro

Michael Crotty
Research Statistician
JMP Division, SAS Institute



THE
POWER
TO KNOW[®]

Objectives

- Introduce the bootstrap feature in JMP Pro
- Discuss examples of some non-standard use cases
- Examples are meant to be illustrative, but not necessarily realistic
- Start conversations with users on ways that they are or hope to be using the bootstrap feature for their problems
- Hope to inspire customers to share their uses of the bootstrap feature at Discovery 2013

Outline

- Introduce the bootstrap feature in JMP
- Bootstrap as an intro stat teaching tool
- Bootstrapping Examples
 - Non-standard quantities
 - Functions of the output
 - Multiple tables in one bootstrap run
 - Model from the Fit Curve platform
- Conclusions

Introduction to the Bootstrap

- Introduced by Brad Efron in 1979; grown in popularity as computing power increases
- Resampling technique that allows you to estimate the variance of statistics, even when analytical expressions for the variance are difficult to obtain
- You want to know about the population, but all you have is one sample
- Treat the sample as a population and sample from it with replacement
 - This is called a *bootstrap sample*
 - Repeating this sampling scheme produces *bootstrap replication*
 - For each bootstrap sample, you can calculate the statistic(s) of interest

The Bootstrap in JMP

- Possible to do a bootstrap analysis prior to JMP 10 using a script
- “One-click bootstrap” added to JMP Pro in Version 10
 - Available in most Analysis platforms
 - Takes advantage of the Automatic Recalc feature
- Results can be analyzed in Distribution platform, which will know to provide Bootstrap Confidence Limits, based on percentile interval method (Efron & Tibshirani 1993)

Bootstrap in an Intro Stat Course

- Allows students to see the effects of sampling variation without being bogged down with distribution theory
- Bootstrap method does not require distributional assumptions on your data or sample
- Avoids some of the feeling of introductory statistics courses being “just a bunch of formulas”
- (As far as I know,) bootstrap is not widely used in introductory statistics courses

Non-standard quantities

- By non-standard, I mean statistics for which we don't readily have standard errors
 - Could be unavailable in JMP
 - Could be difficult to obtain analytically
- Example: Adjusted R^2 value in linear regression

Summary of Fit

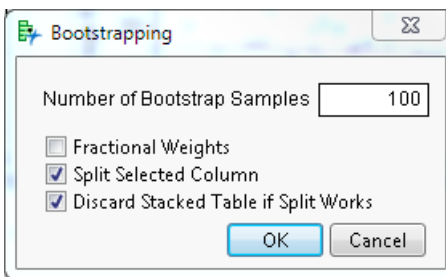
RSquare	0.816083
RSquare Adj	0.779299
Root Mean Square Error	2.50
Mean of Response	47.3
Observations (or Sum Wgts)	

Analysis of Variance

Source	DF	Sum of Squares	Me
Model	5	694.79784	
Error	25	156.58371	
C. Total	30	851.38154	

Parameter Estimates

Term	Estimate	Std Error
Intercept	82.393605	9.178563
Runtime	-2.951817	0.385969



Bootstrap Confidence Limits

Coverage	Pct Lower	Pct Upper
0.95	0.61401	0.91339
0.90	0.67733	0.90573
0.80	0.71797	0.894
0.50	0.76577	0.86202

Original Estimate 0.779299

Functions of output

- Useful for situations where the statistic of interest does not appear in a JMP report
 - Could be a business-defined function of JMP-reported statistics
- Example: defining a function using a bootstrap sample

		BootID*	Stock A	Stock B	Stock C	Portfolio
x	1	0	0.4572	0.3122	2.7372	624.372255
	2	1	0.1922	-2.4137	6.2182	1021.47875
	3	2	0.0656	-1.5596	2.4377	338.157156

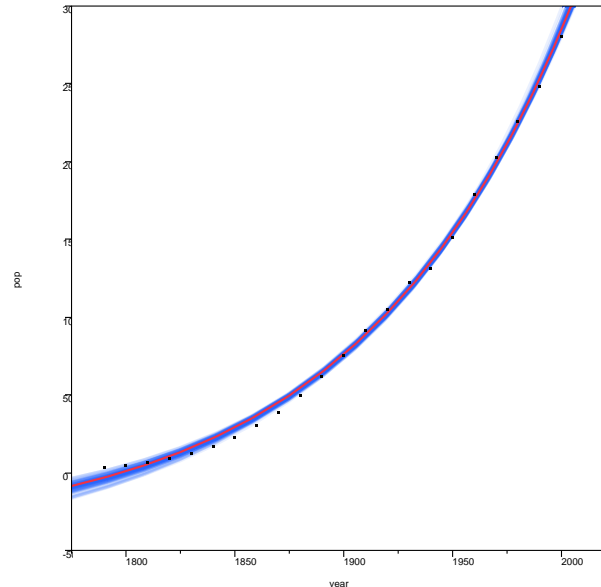
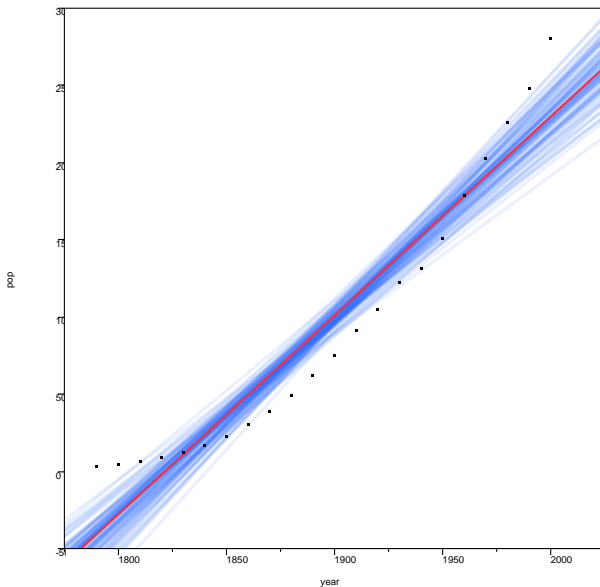
Multiple tables at one time

- “Trick” to use JSL to select multiple tables in a report and bootstrap them
 - Can also be accomplished by setting a random seed prior to each separate bootstrap analysis for multiple tables in a report
- Example: Fixed scale parameter in lifetime data

```
rpt["Parametric Estimate - Normal"][NumberColBox("Estimate")] << select;  
rpt["Fix Parameter"]["Parametric Estimate - Normal"][NumberColBox("Estimate")] << select;  
dtlst = rpt["Parametric Estimate - Normal"][NumberColBox("Estimate")] << Bootstrap(100);
```

Bootstrapping curves

- Similar to the example for functions of output
- We want to use the bootstrap feature to get a sense of the variability of a curve fit in the new Nonlinear platform
- Example: Fit Curve



Conclusions

- Bootstrap is a powerful new feature with many uses
- Primarily a UI feature, but capability is enhanced when scripted in JSL
- Allows us to get confidence intervals for statistics, functions of statistics and curves
- Review objectives:
 - Examples are meant to be illustrative, but not necessarily realistic
 - Start conversations with users on ways that they are or hope to be using the bootstrap feature for their problems
 - Hope to inspire customers to share their uses of the bootstrap feature at Discovery 2013

Acknowledgements, References

- JMP Statistical R&D
- Sam Gardner, Clay Barker, Bill Meeker
- Efron, B. “Bootstrap methods: Another look at the jackknife.” *Annals of Statistics*, 7(1):1-26, 1979.
- Efron, B. and Tibshirani, R. *An Introduction to the Bootstrap*. Chapman and Hall, 1993.
- Stine, R. “Bootstrap Resampling.” Course notes from ICPSR Blalock Lectures 2003.
<http://www-stat.wharton.upenn.edu/~stine/mich/index.html#bootstrap>

Questions?



THE
POWER
TO KNOW.

Advanced Use Cases of the Bootstrap Feature in JMP[®] Pro

Michael Crotty

michael.crotty@sas.com

Research Statistician

JMP Division, SAS Institute

Thank You!