Resampling: Bootstrapping and Randomization

Presented by:

Jenn Fortune, Alayna Gillespie, Ivana Pejakovic, and Anne Bergen

Outline

1. The logic of resampling and bootstrapping

2. Bootstrapping: confidence intervals

3. Bootstrapping and t-tests

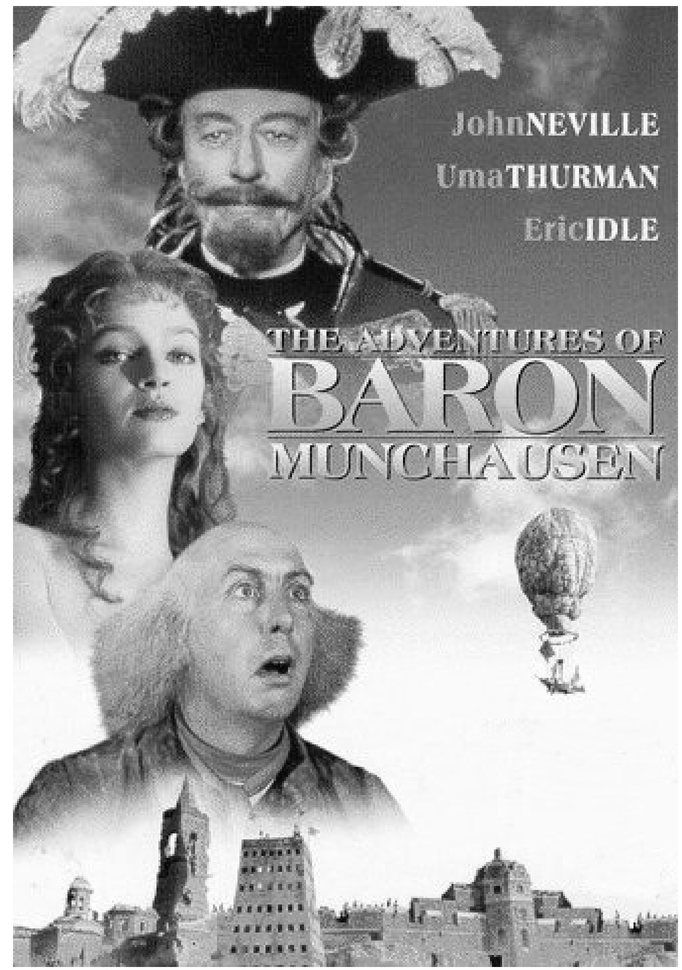
4. Bootstrapping, ANOVA,
   and multiple regression

What is Resampling?

- Draw one sample of some population
- Repeatedly sample from the sample
- E.g., Bootstrapping, Randomization, Jackknifing

Bootstrapping: Getting Something from Nothing

- Baron Munchausen (1720-1797)
- "Pulling oneself up by one's bootstraps"
- *The Surprising Adventures of Baron Munchausen* *(Rudolf Erich Raspe, 1785)*

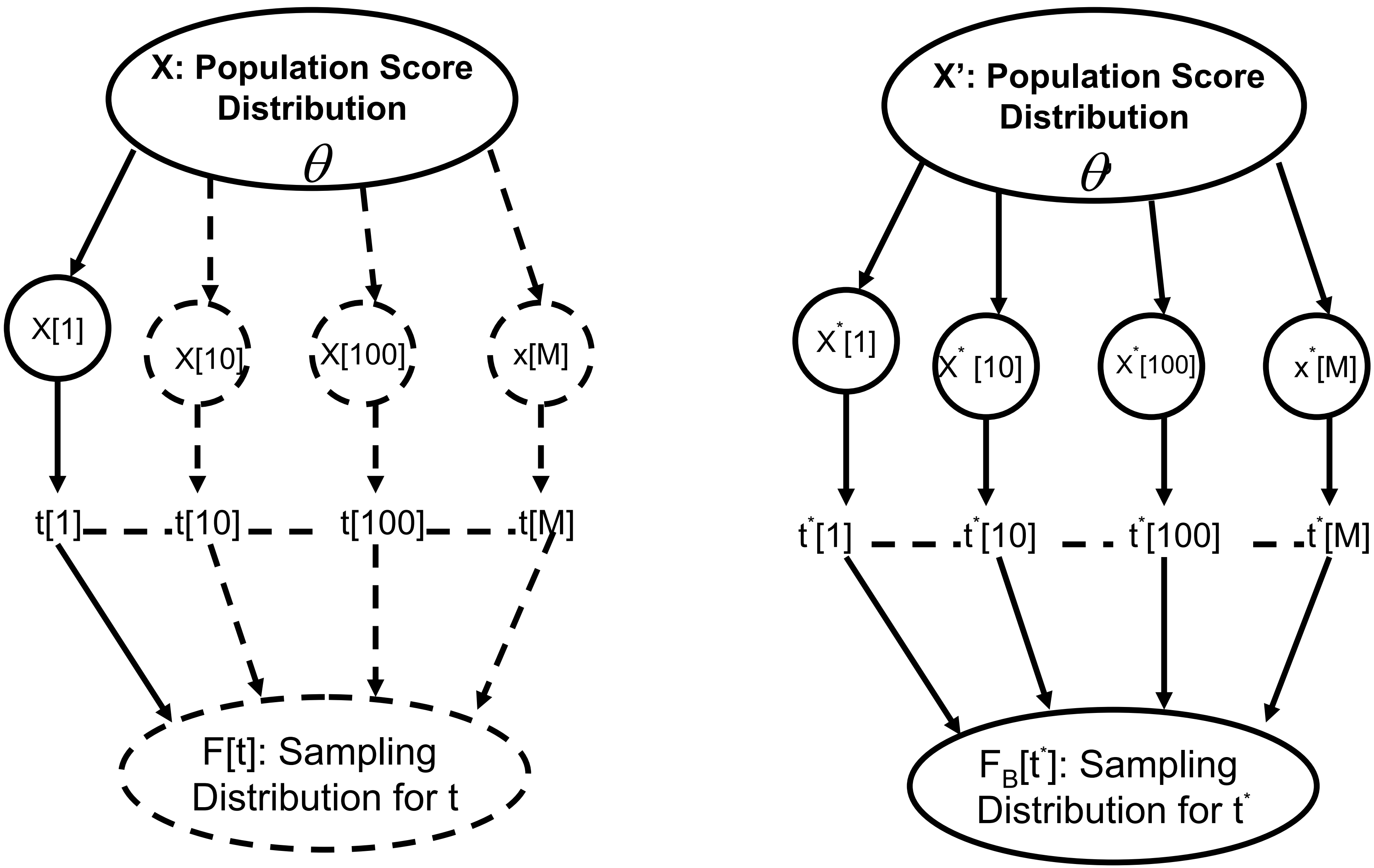


Bootstrapping

- "To bootstrap a statistic such as the sample mean, draw hundreds of resamples with replacement from a single original sample, calculate the statistics for each resample, and inspect the bootstrap distribution of the resampled statistic" (Hesterberg, et al., 2003)

- A sample is to a population as a bootstrap resample is to a sample

Real vs. Bootstrap World



Logic of Bootstrapping

Population: $X_1, X_{2,} X_{3,} X_{4,} X_{5,\ldots} X_N$

- Parameter: θ
- Draw n observations
- Estimate θ : θ_1', θ_2'…
- Sampling distribution of θ

Sample: $x_1, x_{2,} x_{3,} x_{4,} x_{5,\ldots} x_N$

- Parameter: θ'
- Draw n observations from sample with replacement
- Estimate θ'*
- Sampling distribution of θ'*

Resampling without Replacement

If $N/n < 20$:

- Drawing without replacement
  - E.g., $N = 100$ and $n = 20$

- Use the sample to create a pseudo-population with the same proportions as the sample
- Resample from pseudo-population

Did the Stats Grinch Steal Christmas??

- Resampling with Replacement
  - If $N/n \geq 20$

Example: did the Stats Grinch tamper with our Christmas Hugs?

- Null hypothesis: There are 50% red and 50% green Hugs
- Alternative hypothesis: The Grinch tampered with the candy so that the proportion of red and green are no longer equal

- Original sample taken from the population
- *n* = 20
- Assume the sample reflects our population
- Resample with replacement 1000 times
- Obtain the confidence intervals

	A	B	C	D	E	F	G	H	I
1	**Sample (Bucket 1)**								
2	*If your data is in another spreadsheet file, open this, and then copy and paste the data into the green cells.*								
3	Sample size:	20							
4									
5	(The sample reference numbers and random numbers will appear automatically. They are for the resampling process.)								
6	**Sample Reference no**	**Random number**	**Variable 1**	**Variable 2**					
7	1	0.017762947	1						
8	2	0.1214134	1						
9	3	0.354359014	1						
10	4	0.698838821	1						
11	5	0.327121819	1						
12	6	0.939663546	0						
13	7	0.694717656	0						
14	8	0.775573247	0						
15	9	0.841009286	0						
16	10	0.081520258	0						
17	11	0.257003334	0						
18	12	0.584652433	0						
19	13	0.945285804	0						
20	14	0.352788927	0						
21	15	0.006575976	0						
22	16	0.920750274	0						
23	17	0.846666258	0						
24	18	0.244639072	0						
25	19	0.8895796	0						
26	20	0.847755988	0						
27									
28									
29									

Excel spreadsheet by Michael Wood,
http://userweb.port.ac.uk/~woodm/programs.htm

A	B	C	D	E	F	G	H
Single Resample							
Resample size	20						
Resample statistic	2						
	RESAMPLING WITHOUT REPLACEMENT			RESAMPLING WITH REPLACEMENT			
Resample reference no	Sample ref no	Variable 1	Variable 2	Sample ref no	Variable 1	Variable 2	
1	19	0	0	13	0	0	
2	17	0	0	14	0	0	
3	12	0	0	17	0	0	
4	9	0	0	19	0	0	
5	14	0	0	11	0	0	
6	2	1	0	6	0	0	
7	10	0	0	18	0	0	
8	8	0	0	13	0	0	
9	7	0	0	2	1	0	
10	18	0	0	15	0	0	
11	15	0	0	17	0	0	
12	11	0	0	16	0	0	
13	1	1	0	6	0	0	
14	13	0	0	9	0	0	
15	20	0	0	19	0	0	
16	3	1	0	8	0	0	
17	6	0	0	16	0	0	
18	16	0	0	7	0	0	
19	4	1	0	2	1	0	
20	5	1	0	7	0	0	

Lots of resamples (Bucket 2) - Press F9 to calculate

Number of resamples: 996

Resample number	Resample statistic	
	2	
1	7	-
2	6	-
3	3	-
4	5	-
5	5	-
6	5	-
7	2	-
8	5	-
9	5	-
10	7	-
11	3	-
12	4	-
13	5	-
14	3	-
15	6	-
16	6	-
17	5	-
18	3	-
19	2	-
20	5	-
21	4	-
22	7	-
23	4	-
24	4	-
25	4	-
26	6	-
27	8	-
28	6	-
29	4	-
30	7	-
31	11	+
32	2	-
33	8	-
34	6	-
35	8	-
36	5	-
37	9	-
38	3	-
39	2	-
40	5	-
41	4	-
42	8	-
43	5	-
44	5	-
45	5	-

Standard statistics		
mean		4.987951807
median		5
sd		1.985337059
ave dev from mean		1.565950549
lower quartile		4
upper quartile		6
interquartile range		2
percentiles:		
	2.5	1
	97.5	9

Cut value:	
10	
Probability below cut	98.09%
probability equal to cut	1.20%
probability above cut	0.70%
Total	100.00%

Values of interest	Probability
7	10.74%
8	6.22%
9	2.31%
10	1.20%
11	0.40%
12	0.20%
13	0.10%
14	0.00%
15	0.00%
Total	21.18%

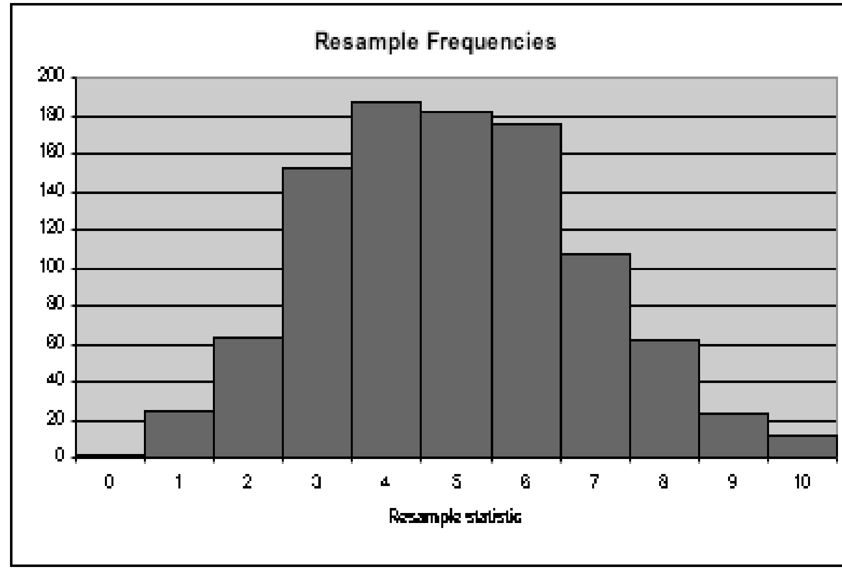


Table for graph			
Middle of bottom bar	0		
Middle of next bar	1		

Bottom	Top	Middle	Frequency
below	-0.5		0
-0.5	0.5	0	1
0.5	1.5	1	25
1.5	2.5	2	63
2.5	3.5	3	152
3.5	4.5	4	187
4.5	5.5	5	182
5.5	6.5	6	175
6.5	7.5	7	107
7.5	8.5	8	62
8.5	9.5	9	23
9.5	10.5	10	12
10.5	upwards		7
Total			989

- Did the Stats Grinch steal Christmas?

- YES: the 95% confidence interval for the sum of red candies is 2 to 9, which means that we reject the null hypothesis that our sample comes from a population with 50% red and 50% green candies

Why do Bootstrapping?

- If we are concerned that assumptions are not met
  - Non-normal populations
    - e.g., Memory of older adults
- Small sample size
- Wish to estimate parameters without an analytical solution

Assumption free testing???

Critical Assumption of Bootstrapping

- The sample data reasonably represents the population

Advantages

1. Method is more transparent, simpler, more general, requires little knowledge of mathematics or probability theory
2. Assumptions are less restrictive
3. Can be applied in situations where conventional methods cannot be used

(Wood, 2005)

Resample Size

- Experiment with different numbers of resamples
- At least 1000 is recommended, but often 10,000 or more are used

A Statistics Revolution?

- Popularity of resampling methods and software increasing
- Computer software programs:
  - AMOS
  - Excel
  - SPSS
  - Resampling Stats
  - S-Plus
  - Howell
  - Etc….

Confidence Intervals

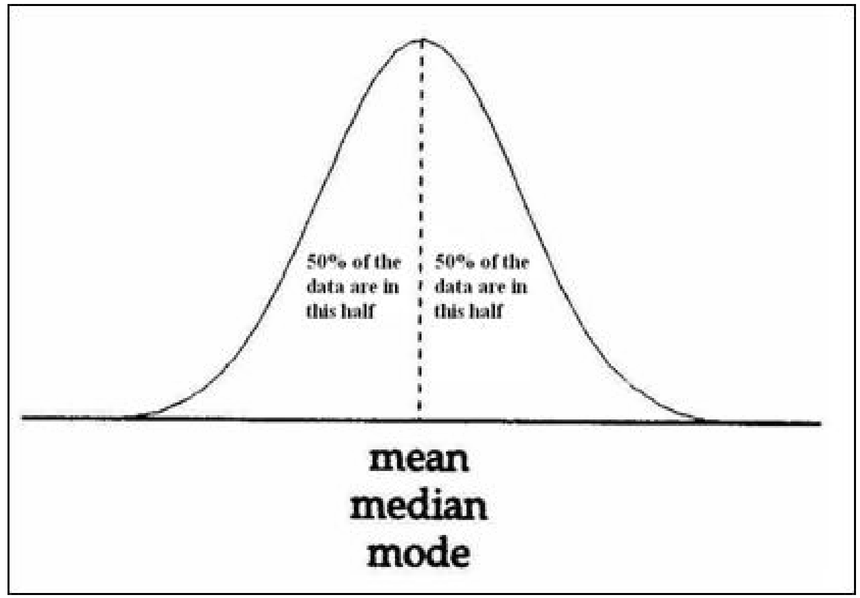
- Confidence Intervals:
  - “How big is the effect?”
    - Significance level
    - Shows the effect size

Parametric Confidence Intervals

- $CI_{.95} = M \pm t_{.025}(df) * s_x$
- What if sample does not meet assumptions?
- What do we do?
- Bootstrapping!

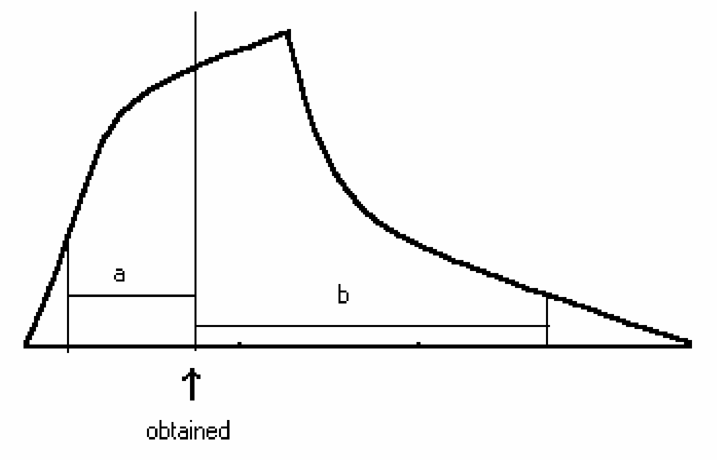
Bootstrapped Percentile CI for the Mean: Normal Distribution

- Find mean
- Resample *with replacement* 1000 times or more
- Calculate bootstrapped sampling distribution of the mean
- 2.5th and 97.5th percentile cut-off



Bootstrapped Bias--Corrected CI for the Mean: Non-Normal Distribution

- What to do here?
- a $\neq$ b
- Normal distribution confidence interval method:
  - (M – a) and (M + b)
- Biased correction confidence interval method:
  - (M + a) and (M - b)

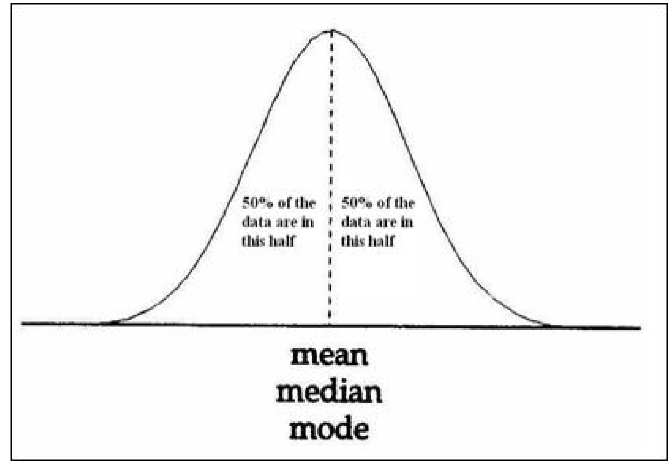


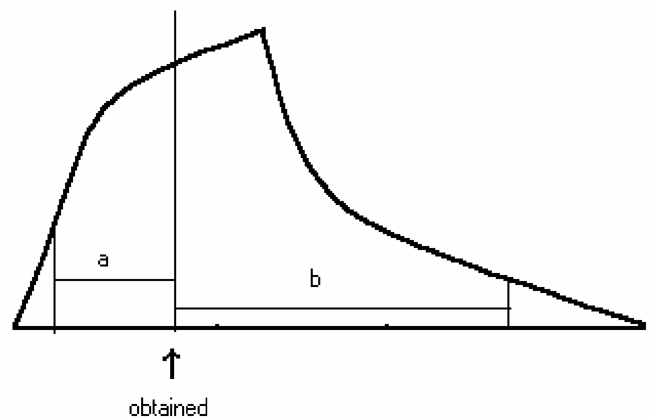
Bootstrapping CIs for Medians

- Median for normal distribution
  - Analytic solution for CI
- Median for non-normal distribution
  - NO analytic solution for CI
- Confidence interval on the difference between 2 medians
  - NO formula to use to get CI regardless of the shape of the distribution

Bootstrapping CI for 1 Median

- **Percentile method:**
  - We take our original sample of *n* observations, and sample from it, *with replacement* to create new samples
  - Calculate median
  - Obtain sampling distribution of median
  - 2.5th and 97.5th percentile cut-offs are (M-a) and (M+b)
- **Bias Correction method:**
  - We take our original sample of *n* observations, and sample from it, *with replacement* to create new samples
  - Calculate median
  - Obtain sampling distribution of median
  - 2.5th and 97.5th percentile cut-offs are (M-b) and (M+a)

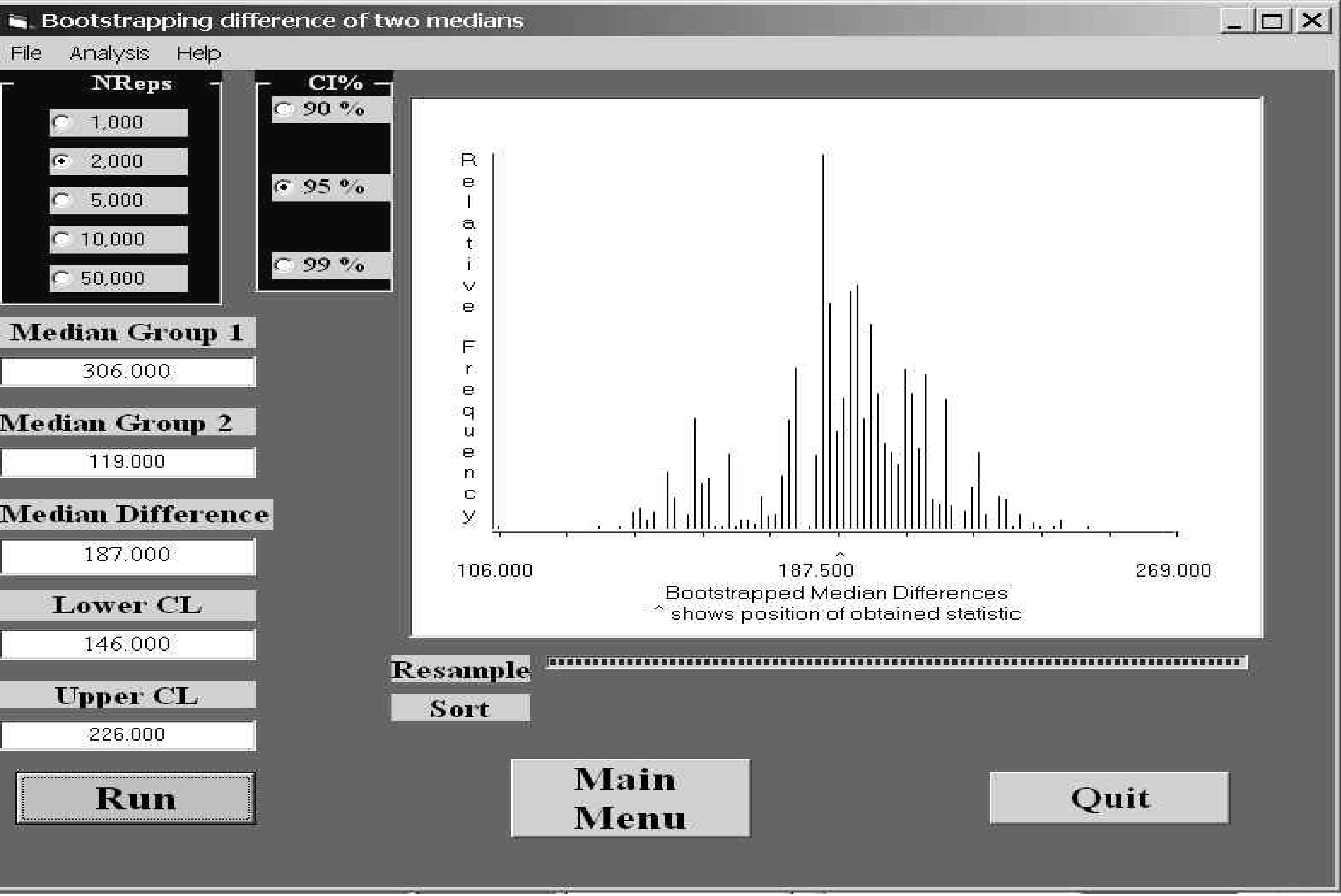




Bootstrapping CI for 2 Medians

Two Independent Samples:

1. Bootstrap each sample to create a bootstrap sampling distribution for each median
2. **Calculate the difference between the medians and create sampling distribution of these differences**
3. Establish confidence interval



Software by D.C. Howell,
http://www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html

The t-test

- The t-test provides a good, simple demonstration of how statistics we're familiar with can be modified for non-parametric techniques.
- There are two approaches: pure randomization and randomization with bootstrapping

Randomization

- Also known as “permutation tests.”
- There are two approaches: the shuffle (resample without replacement) and the bootstrapped shuffle (resample with replacement)

(Hesterberg et al., 2003; Howell, 2002; Wood, 2005)

Randomization

- Will the same pattern of results be observed if the observed data are randomly assigned to experimental groups?
- If H_0: $\mu'_1 = \mu'_2$ is true, then it should not matter which group each data point falls in.

The Shuffle vs. The Candy Dish

- The data points are “shuffled,” or randomly assigned to the experimental groups (without replacement)

- The data points are tossed into one big pseudo-population and samples are then drawn from there with replacement

Either way, the idea is the same! If the null hypothesis is true, it should not matter which experimental group each score came from.

The bootstrapping with randomization approach (“candy dish”) is more popular; this is approach used in the rest of our discussions on randomization techniques.

An example of bootstrapping with randomization

22, 25, 19, 23, 20, 20, 22, 20, 18, 24, 23, 20, 25

19, 17, 15, 15, 17, 16, 18, 14, 15, 16, 18, 13, 14

We calculate a t-statistic for our obtained data

22, ***15***, 25, 19, ***19, 17***, 23, ***15, 16***, 20, ***14***, 20, 22, ***16*** 20, 18, ***18***, 24, ***18, 14***, 23,***13***, 20, 25, ***15, 17***

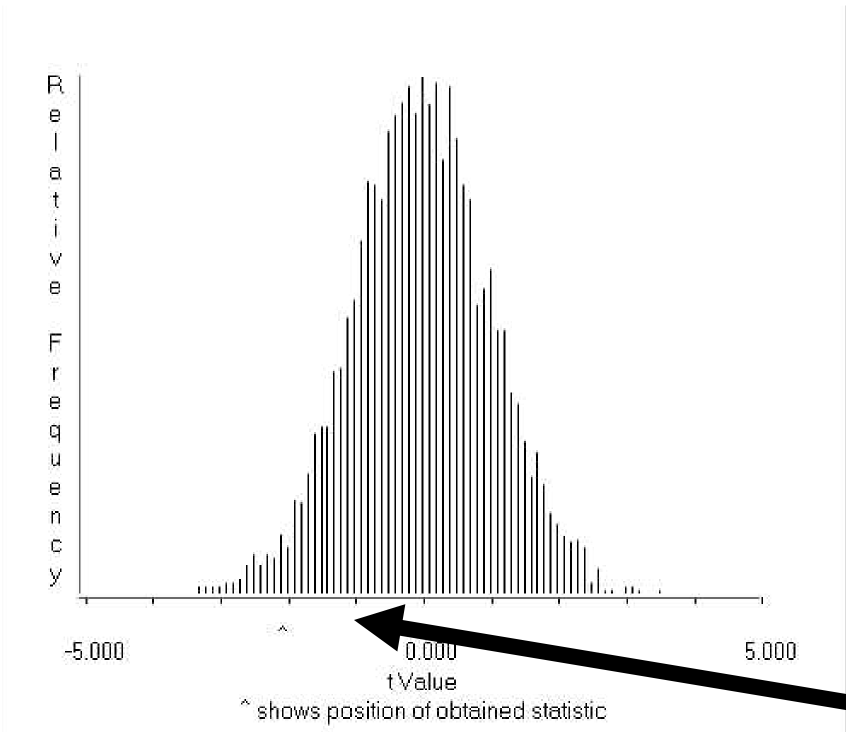
The values get combined into a "pseudo population."

22, ***15***, 25, ***15***, 25, 19, ***18, 14***, 20, 25, ***15, 17***

20, 22, ***16*** 20, 18, ***18***, 24, ***18, 14***, 23,***13, 16,*** 20

We then draw 1000s of pairs of samples from this pseudo-population (with replacement) and calculate a t-value for each.

The resulting distribution of t-values



t_{obt}

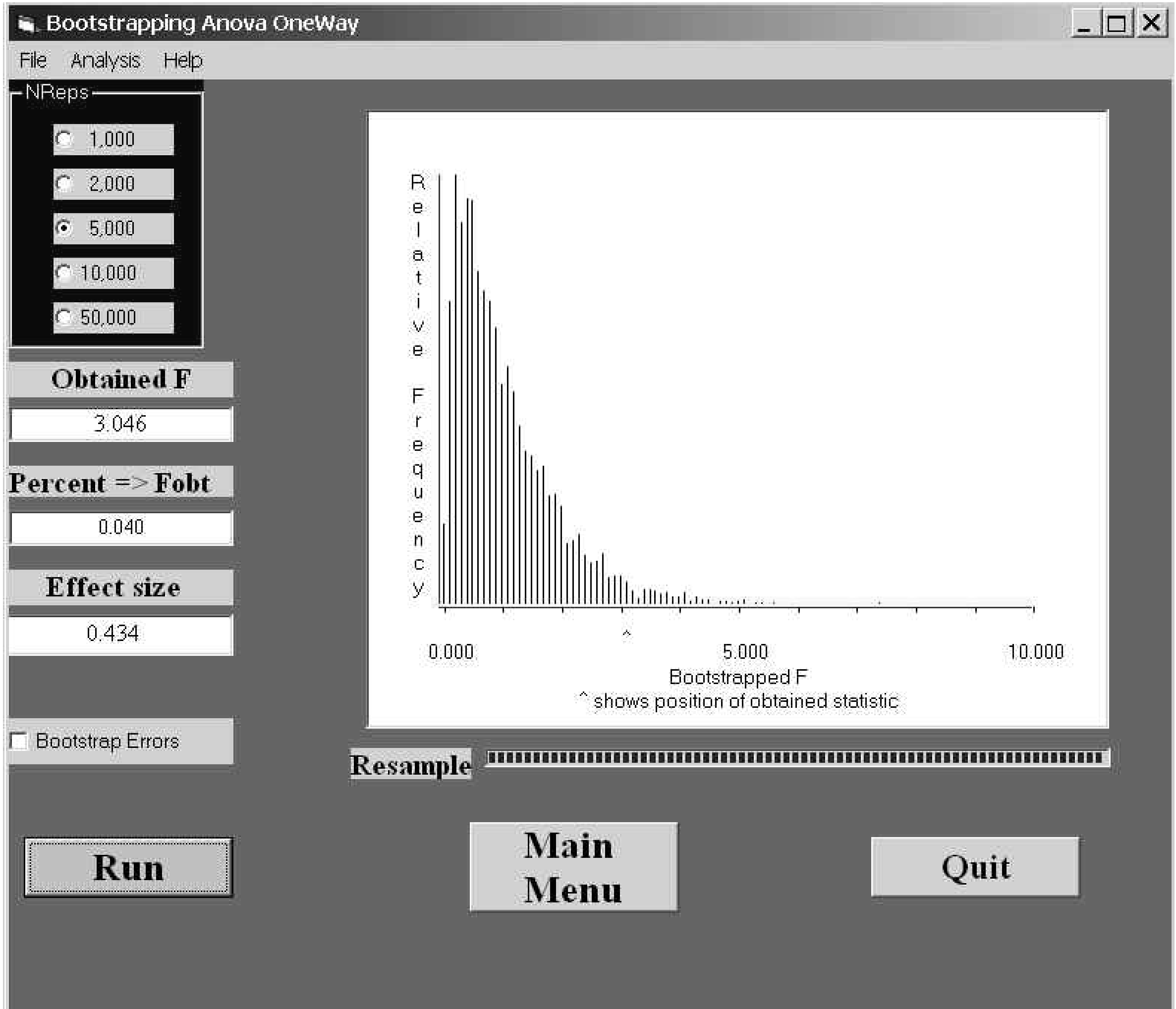
- If the t-value calculated from our obtained data falls outside the 95% confidence limits, then we can conclude that there is a significant difference between the groups.

Bootstrapping ANOVA

1) Calculate F-obtained for original sample
2) Assume no treatment effect (Ho)
3) Combine data from all conditions into a pseudo-population
4) Sample with replacement from pseudo-population
5) Divide bootstrap samples between conditions
6) Calculate F-statistic for each bootstrapped sample
7) Compare F-obtained to bootstrapped sampling distribution of F's

Bootstrapping ANOVA: Example

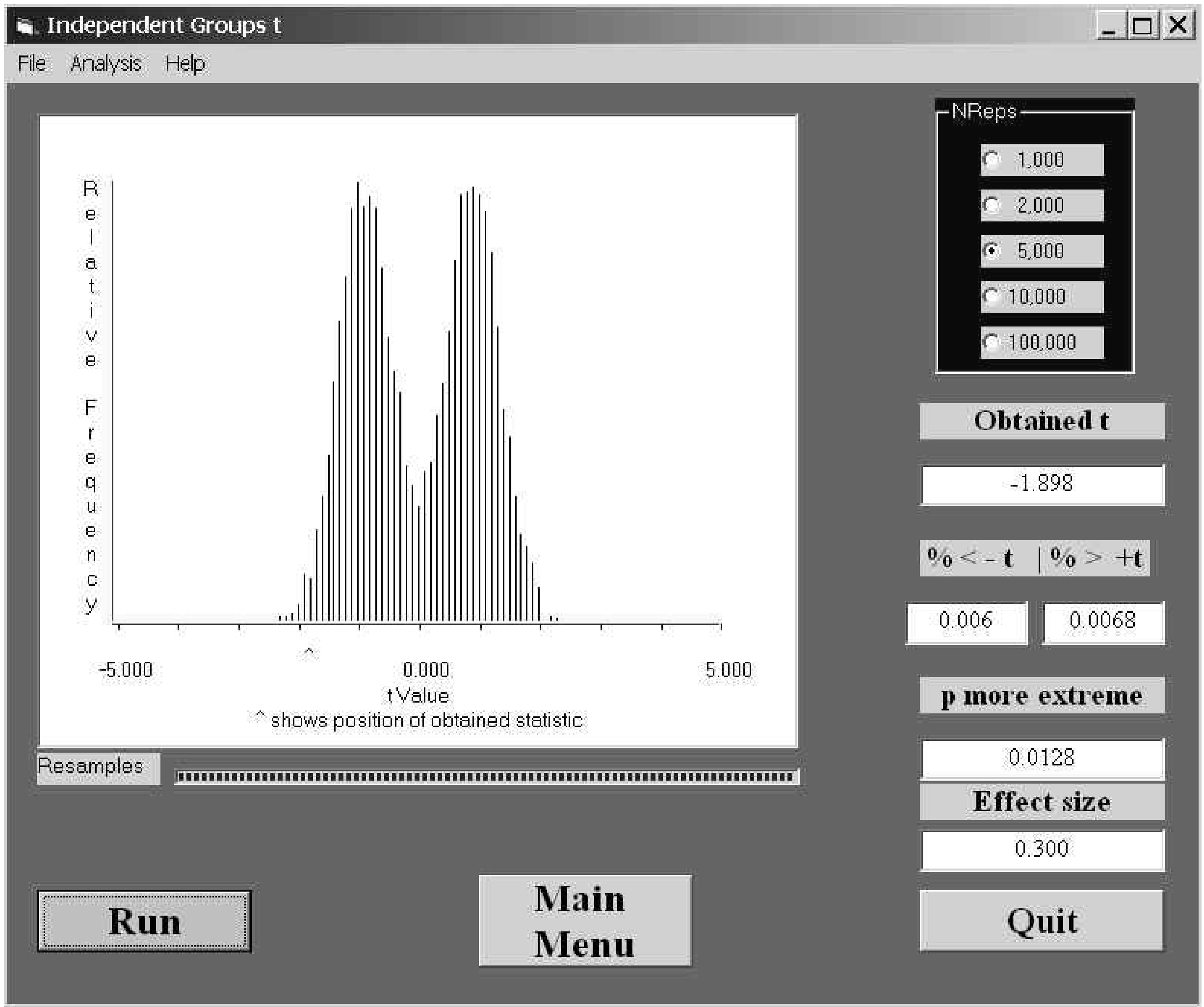
Condition	n	Mean	SD
1	14	11.07	3.95
2	10	15.4	11.12
3	11	18.09	7.13
4	10	19.5	7.11


Bootstrapping Anova OneWay
File
Analysis
Help
NReps
1,000
2,000
5,000
10,000
50,000
Obtained F
3.046
Percent => Fobt
0.040
Effect size
0.434
Bootstrap Errors
Relative Frequency
0.000
5.000
10.000
Bootstrapped F
^ shows position of obtained statistic
Resample
Run
Main Menu
Quit

Problem with Bootstrapping Raw Scores

- If conditions have very different means
  - M1 = 3
  - M2 = 4
  - M3 = 60
  - M4 = 65
- This pseudo-population has a bimodal distribution
- Pseudo-population doesn’t resemble what we think actual population would look like when Ho true (we’re resampling from the wrong pseudo-population)

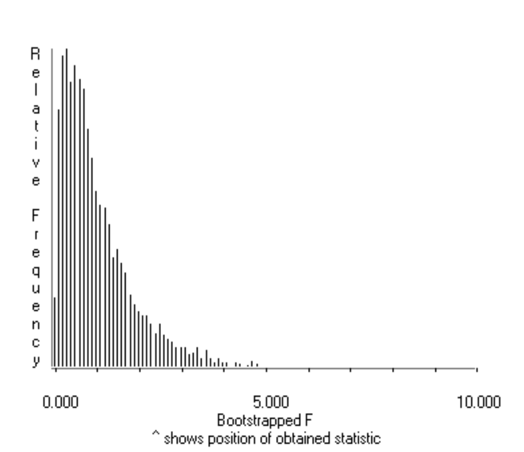
A bimodal resample distribution

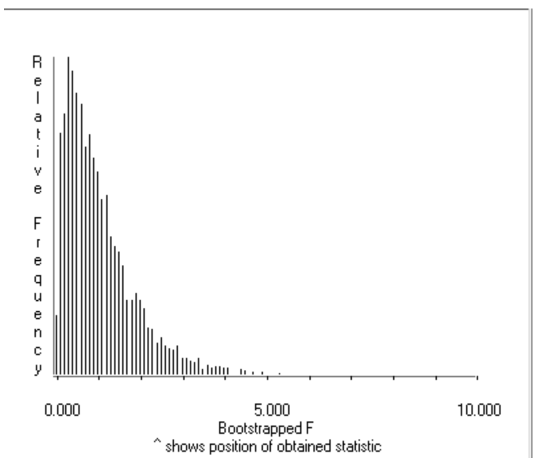


Solution: Bootstrap Errors (Residuals)

1) Calculate error scores (X_{ij} - M_j)
2) Bootstrap sample error scores (X_{ij} - M_j) rather than raw data (X_{ij})
3) Divide error scores between conditions and calculate F-statistics as before to create bootstrapped sampling distribution of F's

Bootstrapping Residuals





- By bootstrapping residuals rather than raw scores, will get F-distribution of bootstrapped samples from a pseudo-population that resembles the actual population (when Ho is true)

Multiple Comparisons and Bootstrapping - one strategy

1) Run a series of randomization tests comparing pairs of independent groups
2) Use Bonferroni adjustment to control for alpha inflation

- Good reference:

  Westfall, R. H. & Young, S. S. (1993) Resampling-based multiple testing. New York: John Wiley & Sons

Randomization and ANOVA

- Same as bootstrapping ANOVA except that you sample *without* replacement from your pseudo-population

Bootstrapping and Linear Regression

When? $\Rightarrow$ Violations of assumptions:

e.g. non-normal residual distribution

Two Options:

1) Model (residual) resampling
   - aka fixed *x* resampling
2) Case resampling
   - aka random *x* resampling

Bootstrapping Regression Coefficients - Residual Resampling

1) Run regression to obtain fit coefficient (β's) and residuals (difference between observed and predicted values)
2) Draw a bootstrap sample of the residuals
3) Add residuals to original regression equation and X (predictor) values to generate new bootstrap values for outcome variable (Y)
4) Run regression analysis to generate new bootstrap regression coefficients for bootstrap sample
5) Repeat process of resampling of the residuals, adding them to the fitted values and estimating the regression coefficients) at least 1000 times

Regression - Residual Resampling

- This process gives an estimate of standard errors and confidence intervals for the regression coefficients from the bootstrap samples

Case Resampling

1) Pair X and Y values at random (bootstrap sample)
2) Get regression coefficients for bootstrap sample
3) Repeat at least 1000 times
4) Calculate confidence intervals for estimated regression coefficients

When to use case resampling (theoretical approach):

1) X,Y pairs are a random sample from population where X and Y are independent
2) X values are randomly assigned to *n* units and experimental response Y is obtained

(Manly, 1997)

When to use case resampling (practical approach):

1) Heteroscedasticity in the residual variance
2) Correlation structure in the residuals
3) Suspect other important parameters are missing from the model

When to use residual resampling:

1) Residuals are independent
2) Residuals are homoscedastic

NB: logic behind bootstrapped residuals is better understood than with case resampling method

Conclusions

References

Chernick, M. R. (1999). *Bootstrap Methods: A Practitioner's Guide.* New York, NY: John Wiley & Sons.

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall, Inc.

Jöckel, K. H., Rothe, G., & Sendler, W. (1992). *Bootstrapping and Related Techniques*. New York, NY: Springer-Verlag.

Lepage, R., & Billard, L. (1992). *Exploring the Limits of Bootstrapping*. Toronto, ON: John Wiley & Sons.

Lunneborg, C. E. (2000). *Data Analysis by Responding: Concepts and Applications*. Pacific Grove, CA: Brooks/Cole.

Manly, B. F. J. (1997). *Randomization, Bootstrap and Monte Carlo in Biology.* New York, NY: Chapman & Hall, Inc.

Hesterberg, T., Monaghan, S., Moore, D. S., Clipson, A., & Epstein, R. (2003). *Bootstrap Methods and Permutation Tests: Companion Chapter 18 to the Practice of Business Statistics.* New York, NY: W.H. Freeman & Company.

Howell, D. C. (2002). Resampling and nonparametric approaches to data. In *Statistical Methods for Psychology* (5th ed., pp. 693-699). Belmont, CA: Thomson Wadsworth.

Howell. D. C. (2002, December 5). *Resampling statistics: Randomization and the Bootstrap*. Retrieved November 1, 2006, from http://www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html

Kelly, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement, 65*, 51-69.

Simon, J. L. (1997). *Resampling: The new statistics.* Retrieved November 11, 2006, from http://www.resample.com/content/text/index.shtml

Wood, M. (2005). Bootstrapped confidence intervals. *Organizational Research Methods, 8,* 454-470.