Stats 506: PS3 Resampling Methods

Diana Liang

11/8/2019

Motivation

Finding the confidence interval is often difficult in real life analyses of complex problems when data is sparse or unavailable. In model simulation, finding a confidence interval can be just as difficult. Different resampling methods have been developed to solve this problem, and four of them will be compared below: *Jackknife, percentile method, basic bootstrap, and normal approximation*.

The simulations in this study will be based on finding ratio of means, as in the mean(x)/mean(y). The first part will not include Monte Carlo replicates while the second part will.

Part 1: Confidence Intervals for No Monte Carlo Replicates

The below table uses the ToothGrowth data to find the ratios of mean odontoblast length between the "OJ" supplement and the "VC" supplement.

Mean Ratios Based on Dosage and Method (95% Confidence Interval)

Dose	Basic Bootstrap	Jackknife	Normal Approx	Percentile
0.5	1.66 (1.10, 2.06)	1.66 (1.15, 2.17)	1.66 (1.18, 2.14)	1.66 (1.25, 2.21)
1.0	1.35 (1.17, 1.53)	1.35 (1.16, 1.55)	1.35 (1.17, 1.54)	1.35 (1.17, 1.54)
2.0	1.00 (0.86, 1.11)	1.00 (0.86, 1.13)	1.00 (0.87, 1.12)	1.00 (0.88, 1.13)

The confidence intervals seem to be similar between the different methods, either meaning that the methods verify each other's accuracy or meaning that this analysis was not enough to demonstrate the differences.

Part 2: Confidence Intervals with Monte Carlo Replicates

Monte Carlo replicates allow for more variability in the data without changing the underlying information.

Set Up:

Two distributions were chosen with known means. The first distribution is normal with a *mean of 10* and a variance of 5 while the second distribtuion is normal with a *mean of 20* and a variance of 15. By this definition the *actual mean ratio should be 0.5*.

1000 Monte Carlo replicates (and 10,000 bootstraps) were created for n_x, n_y = 25. Then the same analyses were done for n_x, n_y = 10.

Tables:

Mean Ratio Analyses for Different Resampling Methods (n = 25)

Method	Avg Length	Avg Shape	Coverage Prob
ci_jk	0.3788	1.00	0.945
ci_pct	0.4032	1.69	0.933
ci_bbs	0.4032	0.61	0.910
ci_norm	0.4938	1.00	0.944

Mean Ratio Analyses for Different Resampling Methods (n = 10)

Method	Avg Length	Avg Shape	Coverage Prob
ci_jk	0.7253	1.00	0.942
ci_pct	1.0635	2.48	0.937
ci_bbs	1.0635	0.48	0.894
ci_norm	22.4312	1.00	0.944

Interpretation:

All the methods provided a coverage probability of slightly below 95%, as expected for using a significance level of 0.05. Jackknife and normal approximation had a greater coverage probability than the percentile method which had a greater coverage probability than the basic bootstrap. Overall, all the methods grew closer to the expected 95% with greater n.

The percentile method tends to underestimate the mean ratio while basic bootstrap overestimates the mean ratio. Jackknife and normal approximation have very symmetrical

confidence intervals around the point estimate, possibly due to the way these confidence intervals were calculated. In the analyses for n = 10, normal approximation has a greater coverage probability but also has an intensely larger average length compared to the other methods, suggesting that it is more influenced by extreme values. With greater n, all the average shapes became more symmetrical and the average lengths became shorter.