Math 159 – Spring 2010
Jo Hardin
Bootstrap example

Example 0.1. Consider a population of 82 law schools. Two measurements were made on the entering class of each school (in 1973!). LSAT, the average score for the class on a national law test, and GPA, the average undergraduate grade-point average for the class. A random sample of 15 schools is selected from the population, and the correlation between GPA and LSAT score was found to be 0.776.

To investigate the coverage of various bootstrap confidence intervals, we calculate 6 different 95% confidence intervals for $\rho = 0.776$. That is, we simulated data from a bivariate normal population with true correlation of 0.776.

Fish Using Fisher's transformation, we use the standard normal theory confidence interval and back transform for ρ.

t Using a t distribution, we find the SE associated with bivariate normal correlation $((1 - r^2)/\sqrt{n})$.

t BS SE Using a t distribution, we find the SE from the bootstrap distribution of r^*.

BS-t We use the bootstrap to estimate the "T^*" quantiles (instead of using the t distribution) as well as the SE.

perc Using the bootstrap distribution of r^*, we pull off the appropriate quantiles.

BCa As with the percentile interval, we use the bootstrap distribution of r^*. However, we use the bias correction and acceleration factor to adjust the percentage for the appropriate quantiles.

CI	lower	upper	width	asym
Fish	0.452 (0.035)	0.919 (0.010)	0.467	0.435
t	0.558 (0.095)	0.989 (0.000)	0.4318	1
t BS SE	0.531 (0.075)	1.016 (0.000)	0.485	1
BS-t	0.249 (0.025)	0.972 (0.005)	0.723	0.415
perc	0.493 (0.050)	0.923 (0.005)	0.430	0.555
BCa	0.441 (0.040)	0.912 (0.020)	0.471	0.418

lower represents the average of the lower bounds in 200 BS samples. The number in parentheses is the percent of times the true parameter was lower than the lower bound (should be 0.025).

upper represents the average of the upper bounds in 200 BS samples. The number in parentheses is the percent of times the true parameter was above the upper bound (should be 0.025).

width is the average width of the estimated confidence interval (across 200 bootstrap samples).

asym represents the asymmetry of the interval. It is the average value of the following statistic: (upper - estimate) / (estimate - lower). A shape > 1 indicates greater distance from the estimate to the upper bound. A shape < 1 indicates greater distance from the estimate to the lower bound.