Shrout, PE & Bolger, N (2002)
Mediation Analysis and Bootstrap Methods
Appendix A

Notes on the Bootstrap Method
Corrected March, 2003

Efron and Tibshirani (1993) describe the bootstrap as "a computer-based method for assigning measures of accuracy to statistical estimates". The usual measure of accuracy for statistical quantities such as means, regression coefficients and correlations is the standard error. A standard error is the expected standard deviation of the estimate if the estimation study had been repeated a large number of times. Usually a standard error is estimated from equations that are derived from a series of assumptions and mathematical operations. The bootstrap can be used to estimate the standard error using an empirical approach, rather than an explicit application of a formula.

The bootstrap method involves having a computer program generate a series of data sets that are designed to resemble the ones that would be observed if the estimation study were repeated many times. Each bootstrap data set is obtained by sampling (with replacement) from the original data. Call these data sets bootstrap samples.

For example, suppose N=8 and the following heights in inches are observed: Original sample: (68, 69, 69, 70, 71, 72, 72, 74). We use this example to make it very clear the steps involved in using bootstrap methods. The statistical validity of the bootstrap method is best when the sample size is 20 or more (Polansky, 1999).

Here are five bootstrap samples that were created by randomly sampling with replacement from the original sample (the values are ordered in magnitude to make the relative frequencies easier to see):

Bootstrap sample 1: (68, 68, 68, 69, 69, 71, 72, 72)
Bootstrap sample 2: (69, 69, 72, 72, 72, 72, 72, 74)
Bootstrap sample 3: (69, 69, 69, 70, 71, 72, 72, 74)
Bootstrap sample 4: (69, 70, 70, 72, 72, 72, 74, 74)
Bootstrap sample 5: (69, 69, 69, 70, 71, 72, 72, 72)

Note that the original sample has two instances each of 69 and 72 inch subjects, and the values are more likely than any other to occur more than once in the bootstrap samples. However, this tendency is not a certainty. In sample 1, by chance the value 68 was chosen three times and the value 69 was chosen only once. Note also that the original sample does not contain someone 73 inches tall, and thus none of the bootstrap samples have a person of that height.

The mean of the original sample is 70.625 and the estimated standard error of the mean is .706. The means of the bootstrap samples are 69.63, 71.50, 70.75, 71.63, and 70.50. The standard deviation of these five values is .813, and this is the bootstrap estimated standard error, although we would never use a bootstrap estimate based on only five bootstrap samples.

The usual estimate of the confidence interval for the mean is 70.625±(.706)*(2.365) leading to the bounds (68.96, 72.29), where 2.365 is the *t* value corresponding to a tail area .975 and 7 degrees of freedom. As we expect, all five bootstrap sample means fall within this range. If we had constructed 100 bootstrap samples, however, we would have expected about five sample means to be either smaller than 68.98 or greater than 72.29. The confidence bound is expected to be accurate since it is well known that the sample mean is approximately normally distributed.

If we had created 1000 bootstrap samples, then we could have inferred the confidence region on the mean without calculating the standard error and assuming a normal distribution for the estimate. A percentile estimate of the 95% confidence interval is computed by ordering the 1000 bootstrap sample means from lowest to highest, and marking the 25^{th} out of 1000 as the lower bound, and the 975^{th} out of 1000 as the upper bound. Formally we call the i^{th} ordered estimate $\theta^{(i)}$, and we choose the confidence bounds to be $\theta^{(\alpha)}$ and $\theta^{(1-\alpha)}$, where α is chosen so that total coverage is $1-2\alpha$. For the 95% CI, α is .025.

Although the percentile estimates are easy to calculate, it is known that they tend to be too narrow. Efron and Tibshirani (1993, pp 184-186) describe bias-corrected intervals that use bounds $\{\theta^{(\alpha 1)}, \theta^{(\alpha 2)}\}$ instead of the percentile bounds $\{\theta^{(\alpha)}, \theta^{(1-\alpha)}\}$. The values $\alpha 1$ and $\alpha 2$ are defined in a way that takes into account asymmetry in the distribution of bootstrap estimates. When the distribution of the bootstrap estimates is normal, the bias corrected interval and the percentile intervals are virtually the same.