The essence of bootstrapping:

Assume a quantitative variable, that we might *characterize* or describe using numerous different statistics (means, medians, trimmed means, variances, sds, skewness, etc.). ***Our goal is to make inferences about 'parent population' parameters using confidence intervals that have in turn been constructed using the information from a reasonably large number of computer generated bootstrap samples.*** (Take note: we will NOT introduce any mathematical theory here; all that follows involves computer intensive computation, but no 'theory' as such.)

1. Begin from an initial sample, not too small (say, 40 or 50 cases at a minimum); this should be a random sample, or a sample for which we can reasonably argue that it reasonably 're-presents' some 'larger universe' of scores that we shall think of as our *parent population.*

2. Decide what feature(s) of the parent population we would like to make inferences about ('center', 'spread', 'skewness', etc.); then, given one or two choices, say center and spread, *decide on what statistics we want to use for inferences.* We might have two, three or more alternative measures of each feature (*e.g.,* four 'means' for center; s.d.s and IQRs to assess spread, etc), a total of S statistics, say. One goal here is to compare various estimators with one another with respect to their purposes in helping to make inferences.

3. *Compute and save each of these 'statistics' for our initial sample*; we shall call them by a special name: bootstrap parameters (which are also statistics, see below). Reflect on this point since it is easy for anyone to be confused here.

3. Choose or write functions that we will be able to 'apply' to each bootstrap sample, where *each bootstrap sample is simply a sample drawn w/out replacement from the initial sample*. The initial sample will now be regarded – for the purposes of bootstrapping – as our 'bootstrap population'. (*Note carefully that we must take care in what follows to distinguish the parent population from the bootstrap population. The latter population can be said to have bootstrap parameters that are also properly labeled as 'conventional sample statistics'*.)

4. *Generate bootstrap samples a substantial number of times* (say B = 500 to 1000 of these), where we save these bootstrap replicates (those that 'measure' center, spread, skewness) for each of the bootstrap samples. Best to generate an array (matrix, of order B x S) that contains all of these; they shall be called 'replicate values' for the respective statistics, and they will be the basis for the ultimate inferences.

5. Summarize the preceding table *by columns,* one for each statistic that relates to a particular feature of the initial bootstrap population (recalling that our bootstrap population began as our initial sample). Both numeric and graphical methods should usually be employed.

6. *Compute and compare the (conventional)* ***means*** *of the replicate statistics (columns) with the bootstrap population parameters*; the differences may positive or negative, and *these differences measure 'bias'*. Ideally, we might seek zero bias,

but small amounts of bias are usually tolerated, particularly if the biased statistics have compensating virtues, especially relatively small variation across the set of bootstrap samples.

7. *Then compute and compare the s.d.s of the respective statistics*; often the main goal of the entire bootstrapping study is to find which statistics have the smallest s.d.s (which is to say bootstrap standard errors) since these are the statistics that will have the narrowest confidence intervals. If a statistic is found to be notably biased, we may want to 'adjust' the statistics (nominally used as 'centers' of our ultimate confidence intervals).

8. *Generate the density distributions (histograms ok) and, more importantly, **selected quantiles**, of any or all bootstrap statistics* we choose. For, example if we aim to generate a 95% interval for a trimmed mean, we find the 2.5% and the 97.5% quantile points of the distribution of that trimmed mean, and (supposing it has minimal bias) *these become our confidence limits for a 95% interval.* We will surely want to compare these limits with those for the conventional mean. *Statistics with the narrowest CIs can be said to be 'best', particularly if they were found to be 'minimally biased.'* Similar methods are used for 99% CIs, etc. Graphics can be useful in this context, but be sure to note that all the information is based on a (rather arbitrary) initial sample, so care has to be taken not to misinterpret, or over-interpret results.

9. Summarize by comprehensively describing the main results, also noting that this methodology has bypassed normal theory methods – that strictly speaking, apply only when normality assumptions can be invoked; moreover, we have made no assumptions about shapes or other features of the (putative [look it up!]) parent population. In particular, we have not assumed normality at any point.

10. Finally, recognize that the interpretation of any such bootstrap CI is essentially the same as that for a conventional CI gotten by normal theory methods. Be sure you review (see Cumming and Finch) how properly to interpret either kind of interval, bootstrap or conventional.

These ideas readily generalize to statistics that are vectors, such as vectors of regression coefficients. This means we are all free to invent and use bootstrap methods to study the comparative merits and demerits of a wide variety of statistics, without regard for whether they are supported by 'normal theory.' We need not invoke normality assumptions, nor make any other so-called parametric assumptions in the process. *The main thing to note is that any bootstrap sample drawn from a matrix, is just a sample drawn with replacement from the ROWS of an initial sample data matrix*; and (vector) bootstrap statistics are computed for each bootstrap matrix, analogous to what has been described above. The computation may be 'intense' but with modern computers such operations are readily carried out for rather large matrices (thousands of rows, hundreds of columns) if efficient computational routines are used. Conventional statistics can be notably inferior to certain new counterparts, a point that needs often to be considered seriously.