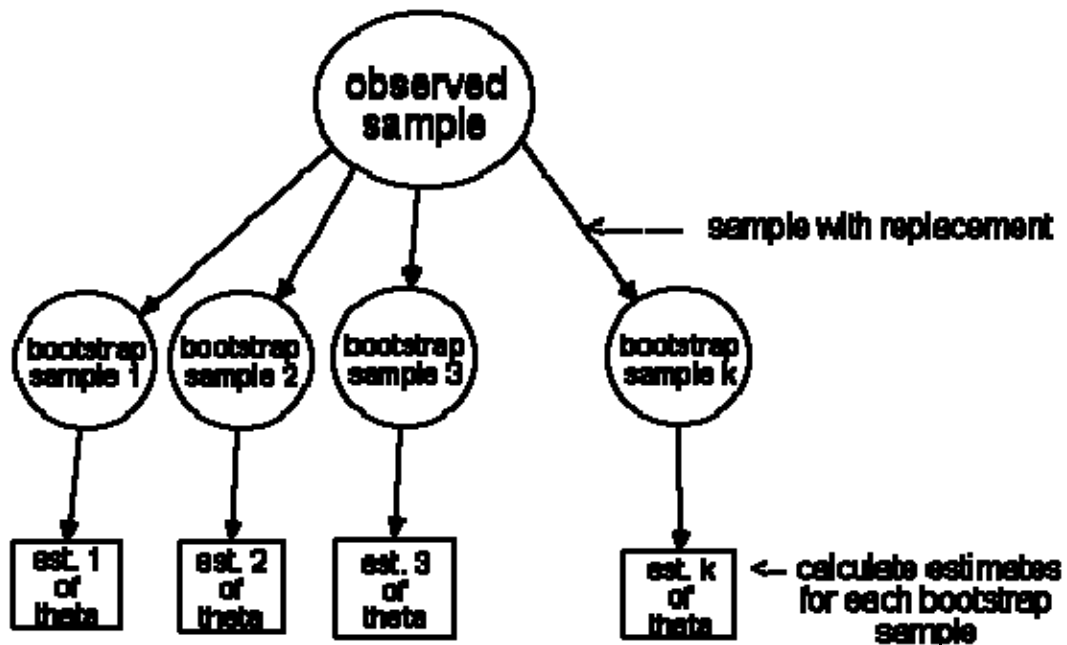


# BOOTSTRAPPING

- In traditional inference, standard errors and confidence intervals are based on the theoretical sampling distribution of parameter estimates (rely on distributional and asymptotic theory)
- One way to deal with small samples and non-normality (violations of theoretical sampling distribution) is the bootstrap.
- Bootstrap was developed by Efron (1982)
  - Efron (1982) "The Jackknife, Bootstrap, and Other Resampling Plans". *Siam monograph* No. 38, CBMS-NSF. Philadelphia
  - Another reference: Efron, B. and Tibshirani, R. (1993) *An introduction to the bootstrap*, New York: Chapman and Hall
- Provides a way to evaluate the empirical sampling distribution of parameter estimates. This empirical sampling distribution can be used in similar manner to the theoretical sampling distribution.
- How does the **bootstrap** work?
  1. Given an observed data set of size  $n$ , fit the model and obtain parameter estimates  $\theta$ .
  2. Take a sample of size  $n$  from the observed data (with replacement) - this is a bootstrap sample
  3. Using the bootstrap sample, fit the model and obtain parameter estimates  $\theta$ .
  4. Go to step 2 and repeat  $k$  times
  5. The distribution of the  $k$  estimates of  $\theta$  represent the empirical sampling distribution.
    - Take the standard deviation of the  $k$  estimates of  $\theta$ . This is an estimate of the standard error of  $\theta$ .
    - OR take 2.5 and 97.5 percentiles of the empirical sampling distribution to form a 95% confidence interval (Efron suggests a bias correction).



- When resampling of the observed data is done, this is called bootstrapping or "non-parametric bootstrapping".

- Another method can be used when the observed data is unavailable, i.e. only the observed covariance matrix is available. The method is called **Monte Carlo Simulation**, or the **Parametric bootstrap**.
  - Steps are same as above EXCEPT STEP 2. Replace STEP 2 with: Generate a sample of size  $n$  from the multivariate Normal distribution with covariance matrix equal to the sample covariance matrix of the original data.
  - So instead of resampling the data, we generate new data with the same covariance structure as the original data.
  - Because this method of bootstrapping relies heavily on the Normal distribution, it is not useful for cases where the normal distribution is violated. Having said that, this method is usually used when the observed data is unavailable so actually there would not be any way to check the normality assumption.
- Bootstrapping is very useful for obtaining standard errors when we do not have a theoretical formula, e.g. standardized loadings, indirect effects.
- The more bootstrap samples taken (i.e. the larger  $k$  is) the smoother the empirical sampling distribution will be. Usually take  $k$  to be larger than 500.

AMOS can perform the bootstrap procedure.

- Check the "Perform Bootstrap" option in the Analysis Properties window.
- Check the Bias-corrected confidence intervals to get Efron's bias corrected percentile confidence intervals
- Check the Bollen-Stine bootstrap to get a bootstrap corrected Chi-squared value
- The columns of the tables for the bootstrap estimates are:

**SE** Bootstrap estimates of standard error.

**SE-SE** An approximate standard error for the standard error in the preceding column, given by  $s/\sqrt{2B}$  where  $s$  is the standard error from the preceding column and  $B$  is the number of bootstrap samples.

**Mean** The mean across bootstrap samples of the quantity being estimated.

**Bias** The difference between the average of  $B$  estimates obtained from  $B$  bootstrap samples, and the single estimate obtained from the original sample.

**SE-Bias** An approximate standard error for the bias estimate in the preceding column. The formula used is  $s/\sqrt{B}$ , where  $s$  is the approximate standard error in the S.E. column and  $B$  is the number of bootstrap samples.

When you request bias-corrected confidence intervals and significance tests, the following columns appear under the subheading BC Confidence.

Lower Lower bound on the bias-corrected confidence interval.

Upper Upper bound on the bias-corrected confidence interval.

P Estimated probability of getting a sample value this far from zero if the population value is zero (Efron and Tibshirani, 1993).

When you request percentile-based confidence intervals and significance tests, the same columns appear under the subheading PC Confidence.