Bootstrapping for Correlation and Simple Linear Regression

As we encountered in the chapter on assocation, there are two ways to obtain a sample of paired data (X_i, Y_i), $i = 1, 2, ..., n$, known as bivariate sampling and fixed$-X$ sampling.

Bootstrap Confidence Interval for ρ

As outlined on page 267, obtain many bivariate bootstrap samples (sampling the pairs (X_i, Y_i) with replacement) of size n, calculate a bootstrap correlation coefficient r_b for each sample, then obtain the confidence interval using a method such as the BCA or percentile method.

Fixed$-X$ Bootstrap sampling

Suppose that we have a model

$$Y_i = h(X_i) + \varepsilon_i,$$

where $h(X_i)$ is some function of X_i, and the errors ε_i are i.i.d. random variables with mean 0 and standard deviation σ. As outlined on page 268, the key idea in fixed$-X$ bootstrap sampling is that the X_i values remain the same (are fixed), but we obtain new Y_i values in each bootstrap iteration by sampling the observed errors e_i and adding them to the predicted values $\widehat{h}(X_i)$ to obtain new Y_i values.

Bootstrap Inference for the Slope in Simple Linear Regression

In linear regression models we have a $t-$pivot for the slope:

$$t = \frac{\widehat{\beta}_1 - \beta_1}{SE(\widehat{\beta}_1)},$$

whose distribution is independent of β_0, β_1 and σ. As in earlier problems, we use the bootstrap distribution of the $t-$pivot to conduct tests and construct confidence intervals for β_1. Specifically we

1) Calculate the least squares estimates of β_0 and β_1 and calculate the residuals $e_i = Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i$ for each observation.

2) We take a bootstrap sample of the e_i's and use the sample of $(X_i, e_{i,b})$ values to calculate the least squares estimate $\widehat{\beta}_{1e}$, and form the bootstrap $t-$pivot:

$$t_e = \frac{\widehat{\beta}_{1e}}{SE(\widehat{\beta}_{1e})}.$$

3) We repeat the bootstrap sampling process many times and form a bootstrap distribution of t_e values which is used to create confidence intervals, such as the 95% bootstrap interval:

$$\widehat{\beta}_1 - t_{e,.975} SE(\widehat{\beta}_1) < \beta_1 < \widehat{\beta}_1 - t_{e,.025} SE(\widehat{\beta}_1).$$

We can perform bootstrap tests for $H_0 : \beta_1 = 0$ by seeing if 0 is in the associated bootstrap confidence interval.