10.4.6 Estimating Variablility of Location Estimates by the Bootstrap

Let $x_1, x_2 \ldots, x_n$ be realizations of independent random variables with common distribution function F. Suppose we are interested in the estimate of $\hat{\theta}$ (mean, median, etc.).

The boostrap solution is to view the empirical cdf F_n as an approximation to F. Recall, F_n is a discrete probability distribution that gives probability $1/n$ to each observed value $x_1, x_2, \ldots, x_n$.

A sample of size n from F_n is a sample of size n drawn with replacement from the collection $x_1, x_2, \ldots, x_n$.

Draw B samples of size n from the observed data producing $\theta_1^*, \theta_2^*, \ldots, \theta_B^*$. These are your boostrap estimates.

Question: What can you do with the boostrap estimates?

The standard deviation of $\hat{\theta}$ is estimated by

$$s_{\hat{\theta}} = \sqrt{\frac{1}{B}\sum_{i=1}^{B}(\theta_i^* - \bar{\theta}^*)^2}$$

where $\bar{\theta}^*$ is the mean of $\theta_1^*, \theta_2^*, \ldots, \theta_B^*$.

Example (p. 361) Heat of Sublimation of Platinum (kcal/mol)
26 measurements from an experiment by Hampson and Walker (1961).

An approximate $100(1-\alpha)\%$ CI for population mean is: $\bar{x} \pm z(\alpha/2)s_{\bar{x}}$
.
Let's find an approximate 95% CI.

What do you notice about this interval?

This example illustrates the sentivity of the sample mean to outlying observations.
What about other measurees of location that are robust or insenstive to outliers?