STAT 113: Lab 6

Tests and Intervals With a Normal Distribution

Last revised November 3, 2017

Overview

The goal of this lab is to build a bridge between purely simulation-based inference and purely analytic inference using a "hybrid" approach. Namely, we will do the simulations in order to get standard errors, but then the rest of the process of finding confidence intervals and computing P-values will rely on analytic approximations.

What to Turn In

Turn in Markdown documents with analyses of two scenarios, each consisting of the following steps.

1. Plot the data or provide a summary table (whichever you think is more informative given the type of data) and compute relevant descriptive statistics
2. Construct a bootstrap distribution and use it to estimate the standard error of the statistic in question
3. Construct 95% confidence intervals for the parameter of interest in three ways
   (a) Directly from the bootstrap distribution
   (b) Using a Normal approximation to the bootstrap distribution
   (c) First finding standardized endpoints (z^* values), and converting these to values on the original scale, based on the mean and standard deviation of the Normal from (b).

4. Construct a randomization distribution based on the null hypothesis in question, and use it to estimate the standard error of the statistic in question.
5. Compute a P-value for the data and the test in three ways
   (a) Directly using the randomization distribution
   (b) Using a Normal approximation to the randomization distribution
   (c) Converting the observed statistic to a z score using the mean and standard deviation of the Normal from (b), and finding the P-value from the z score using a standard Normal

Scenarios

1. Using the `FloridaLakes` dataset (in `Lock5Data`), estimate the mean pH of all lakes, and test whether the mean pH is different from 7 (that of pure water).
2. In the dataset `ICUAdmissions`, from a sample of 200 patients visiting a particular ICU. The variable `Infection` indicates whether patients in an ICU (Intensive Care Unit) had an infection (1) or not (0). The variable `Status` indicates whether the patient died (1 = died, 0 = survived) in the ICU or not. Estimate the difference in death rates between infected and non-infected patients, and test whether this difference is non-zero.