

STAT 113: Lab 7 (Counts for SLOs H2, H3, I2-I5)

Tests and Intervals With a Normal Distribution

Due Date TBD (Will decide at the end of class today)

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 Normal distribution approximations.

Soon, we will dispense with simulation entirely and find standard errors using theoretical approximations too. But not just yet.

What to Turn In

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

- (a) **Plot the data** or provide a summary table (whichever you think is more informative given the type of data) and compute relevant descriptive statistics
- (b) Construct a **bootstrap distribution** and use it to estimate the **standard error** of the statistic in question
- (c) Construct 95% **confidence intervals** for the parameter of interest in two ways:
 - (i) Directly **using the bootstrap distribution**
 - (ii) **Using a Normal approximation** to the bootstrap distribution

- (d) **Interpret** the confidence intervals **in context**
- (e) State **null and alternative hypotheses**, being explicit as to what population parameter they are statements about.
- (f) Construct a **randomization distribution** based on the null hypothesis in question, and use it to estimate the standard error of the statistic in question.
- (g) **Compute a P -value** for the data and the test in two ways
 - (i) Directly **using the randomization distribution**
 - (ii) **Using a Normal approximation** to the randomization distribution
- (h) **State your conclusion** and interpret it **in context**

Scenarios

1. (SLO H2) 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. (SLO H3) In the dataset `ICUAdmissions` (also in `Lock5Data`, 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 survival rates (proportion who survived) between infected and non-infected patients, and test whether this difference is non-zero.