STA 35B: Homework 6

Instructor: Akira Horiguchi

Student name: ABCDE FGHIJ; Student ID: 123456789

Due date: May 12, 2025 at 9 PM (PT)

Instructions

Upload a PDF file, named with your UC Davis email ID and homework number (e.g., ahoriguchi_hw6.pdf), to Gradescope (accessible through Canvas). You will give the commands to answer each question in its own code block, which will also produce output that will be automatically embedded in the output file. All code used to answer the question must be supplied, as well as written statements where appropriate.

All code used to produce your results must be shown in your PDF file (e.g., do not use `echo = FALSE` or `include = FALSE` as options anywhere). Rmd files do not need to be submitted, but may be requested by the TA and must be available when the assignment is submitted.

Students may choose to collaborate with each other on the homework, but must clearly indicate with whom they collaborated.

Problem 1: [IMS] 17.2 Malaria vaccine effectiveness (variant)

With no currently licensed vaccines to inhibit malaria, good news was welcomed with a recent study reporting long-awaited vaccine success for children in Burkina Faso. With 450 children randomized to either one of two different doses of the malaria vaccine or a control vaccine, 89 of 292 malaria vaccine and 106 out of 147 control vaccine children contracted malaria within 12 months after the treatment.

1,000 bootstrapped differences
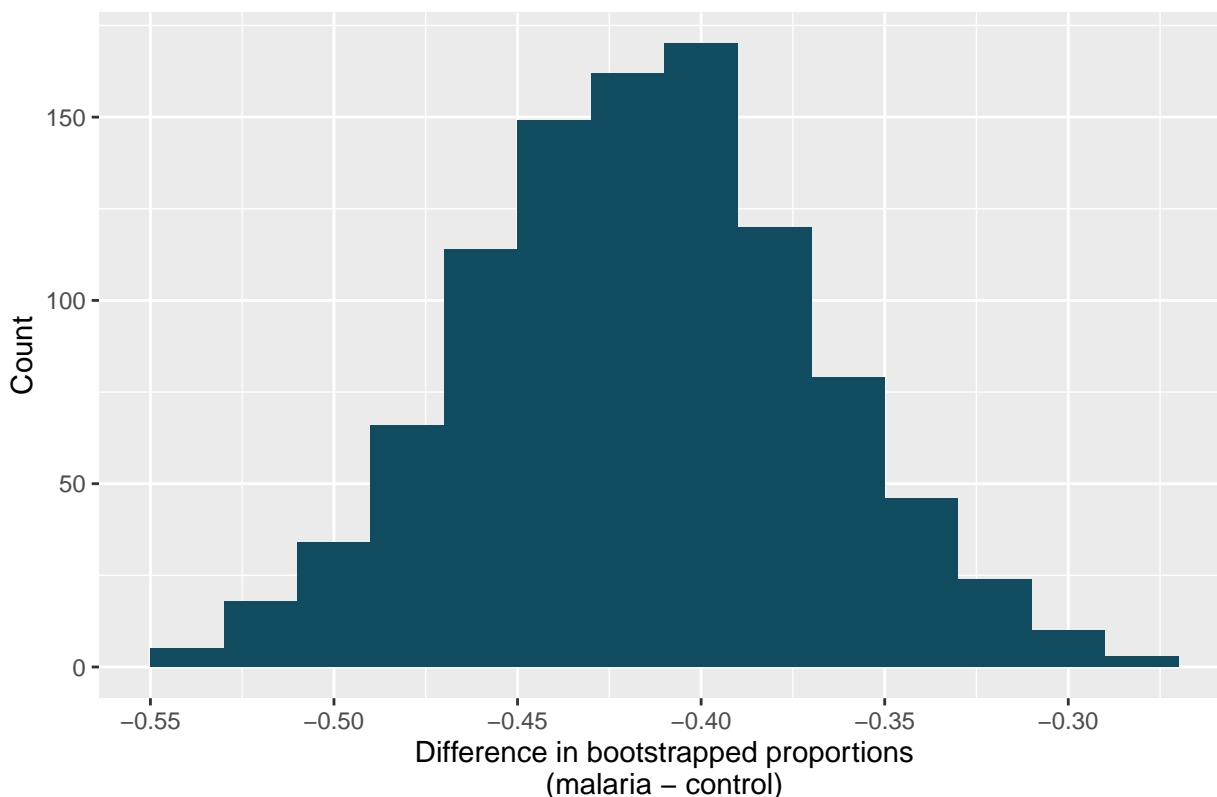
Using the entire bootstrap distribution, find a 95% bootstrap percentile confidence interval for the true difference in proportion of children who contract malaria (malaria vaccine minus control vaccine) in the population. Interpret the interval in the context of the problem.

Problem 2: [IMS] 17.10

According to a report on sleep deprivation by the Centers for Disease Control and Prevention, the proportion of California residents who reported insufficient rest or sleep during each of the preceding 30 days is 8.0%, while this proportion is 8.8% for Oregon residents. These data are based on simple random samples of 11,545 California and 4,691 Oregon residents. Calculate a 95% confidence interval for the difference between the proportions of Californians and Oregonians who are sleep deprived and interpret it in context of the data. (CDC, 2008)

Problem 3: [IMS] 19.4

Researchers studying anthropometry collected body measurements, as well as age, weight, height and gender, for 507 physically active individuals. The histogram below shows the sample distribution of bootstrapped means from 1,000 different bootstrap samples.

1,000 bootstrap means
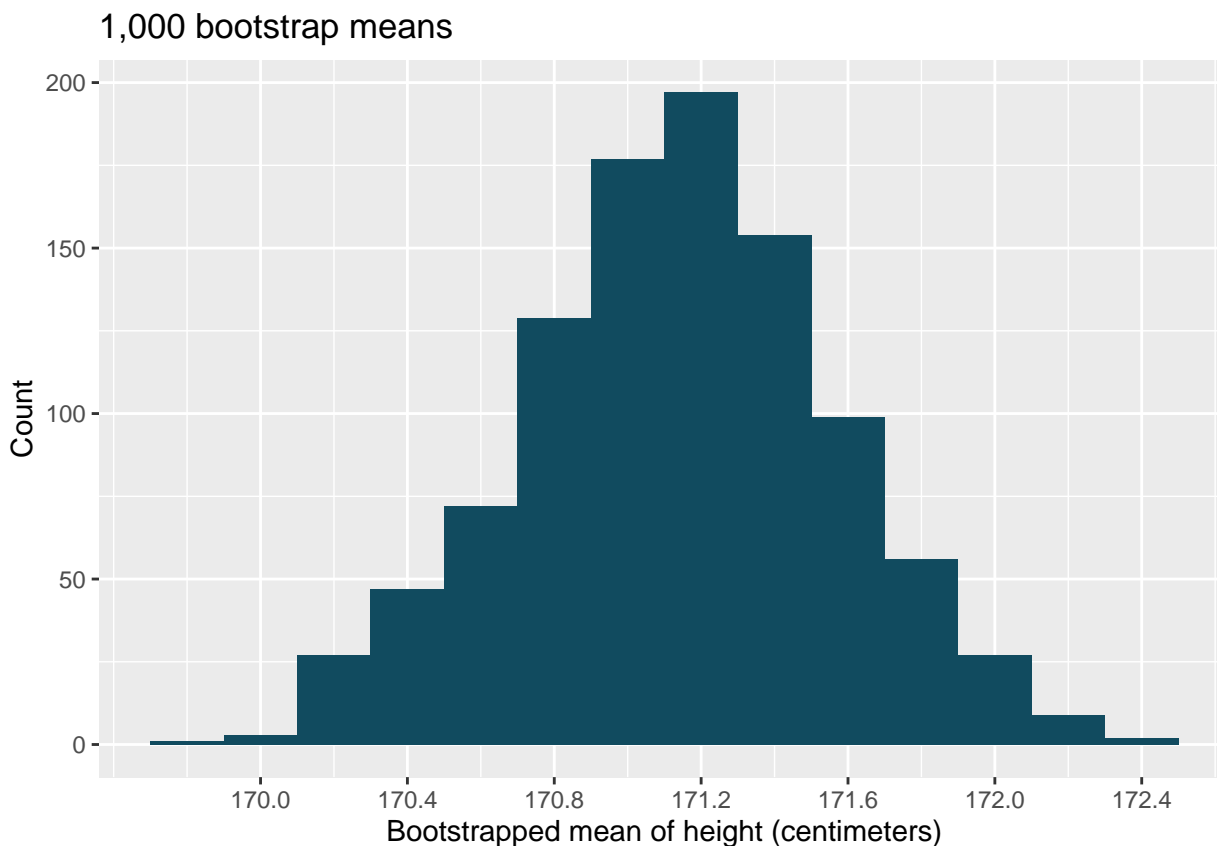
By looking at the bootstrap sampling distribution (1,000 bootstrap samples were taken), find an approximate 90% bootstrap percentile confidence interval for the true average adult height in the population from which the data were randomly sampled. Provide the interval as well as a one-sentence interpretation of the interval.

Problem 4

The figure below shows three unimodal and symmetric curves: the standard normal (z) distribution, the t-distribution with 5 degrees of freedom, and the t-distribution with 1 degree of freedom. Determine which is which, and explain your reasoning.
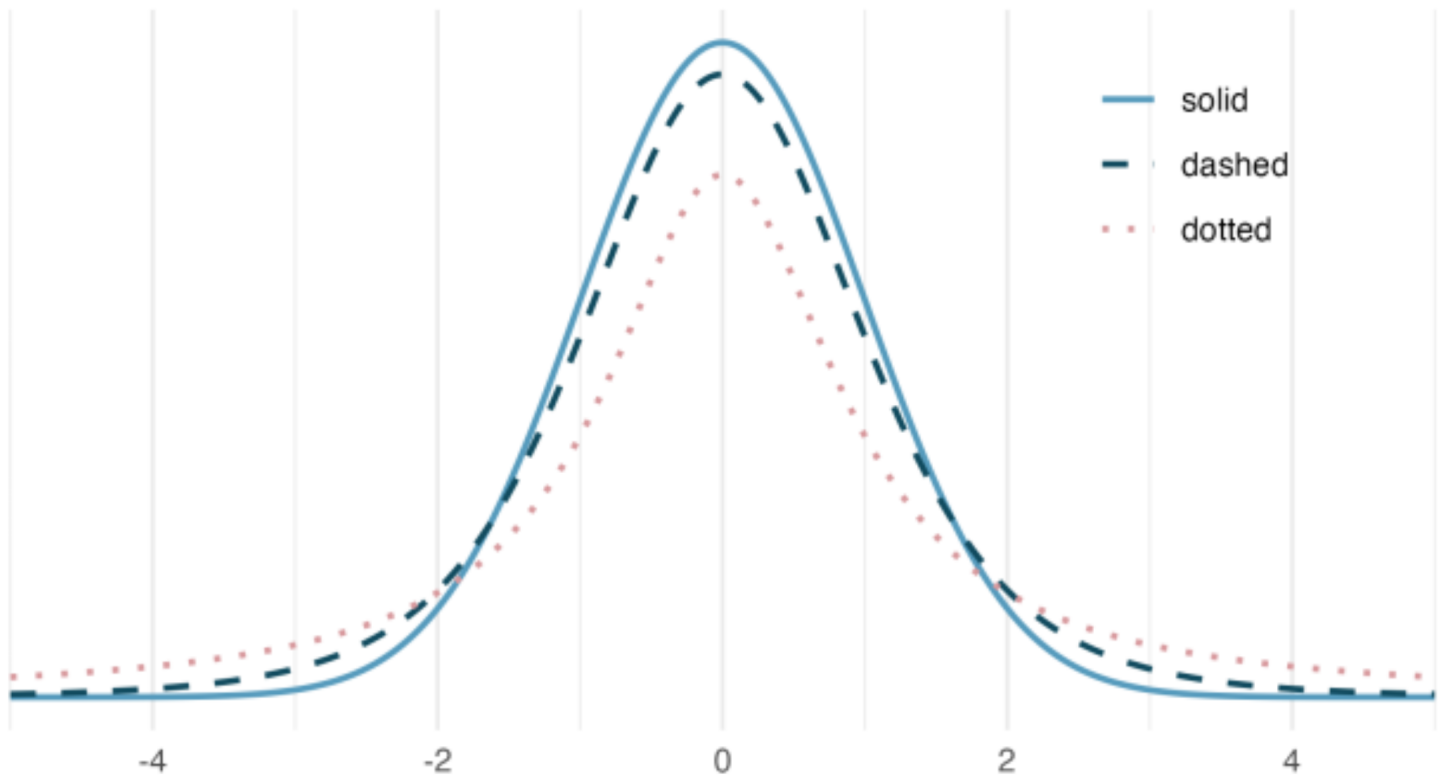
Problem 5

In this exercise we work with a random sample of 1,000 cases from the dataset released by the United States Department of Health and Human Services in 2014. Provided below are sample statistics for gestation (length of pregnancy, measured in weeks) of births in this sample.

Min	Q1	Median	Mean	Q3	Max	SD	IQR
21	38	39	38.7	40	46	2.6	2

a. What is the point estimate for the average length of pregnancy for all women?

b. You might have heard that human gestation is typically 40 weeks. Using the data, describe why the assumptions required to perform a hypothesis test are satisfied, perform a complete hypothesis test, using mathematical models, to assess the 40 week claim. State the null and alternative hypotheses, find the T score, find the p-value, and provide a conclusion in context of the data.

Problem 6

A 90% confidence interval for a population mean is (65, 77). The population distribution is approximately normal and the population standard deviation is unknown. This confidence interval is based on a simple random sample of 25 observations. Calculate the sample mean, the margin of error, and the sample standard deviation. Assume that all conditions necessary for inference are satisfied. Use the t-distribution in any calculations.