STA 35B: Homework 7

Instructor: Akira Horiguchi

Student name: ABCDE FGHIJ; Student ID: 123456789

Due date: June 4, 2025 at 9 PM (PT)

Instructions

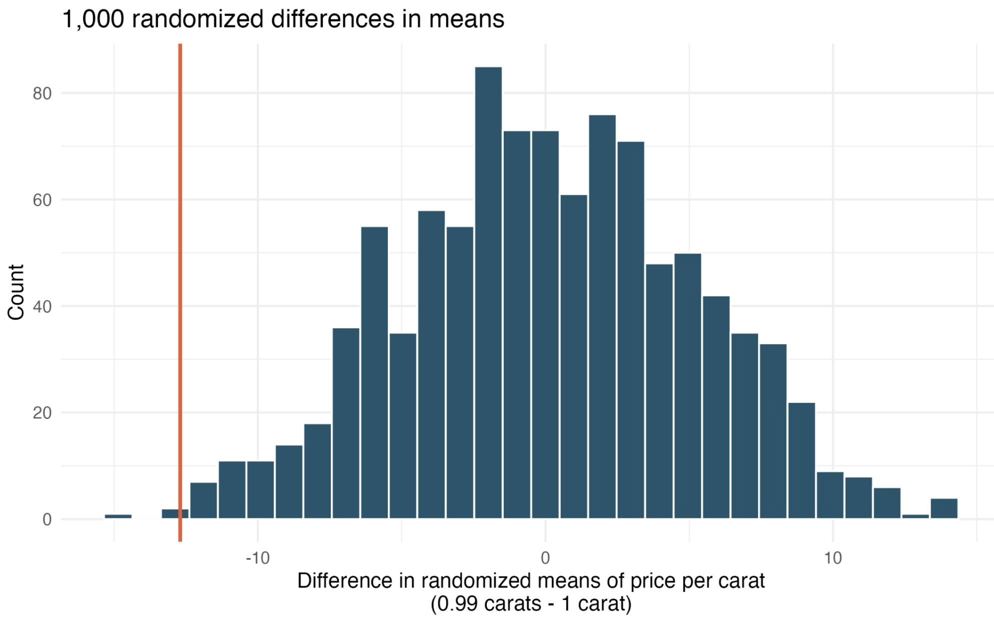
Upload a PDF file, named with your UC Davis email ID and homework number (e.g., ahoriguchi_hw6.pdf), to Gradescope (accessible through Canvas). You will give the commands to answer each question in its own code block, which will also produce output that will be automatically embedded in the output file. 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] 20.6.3 Diamonds, randomization test

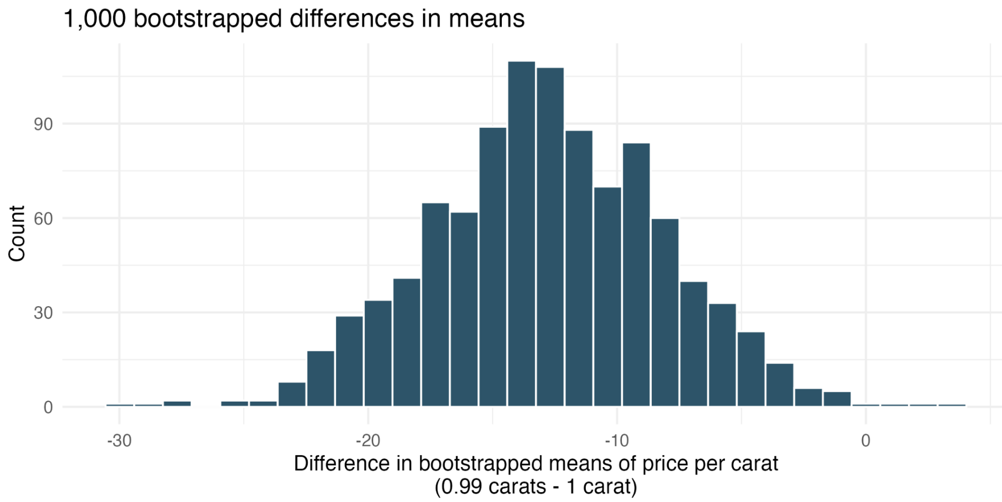
The prices of diamonds go up as the carat weight increases, but the increase is not smooth. For example, the difference between the size of a 0.99 carat diamond and a 1 carat diamond is undetectable to the naked human eye, but the price of a 1 carat diamond tends to be much higher than the price of a 0.99 carat diamond. We have two random samples of diamonds: 23 0.99 carat diamonds and 23 1 carat diamonds. In order to be able to compare equivalent units, we first divide the price for each diamond by 100 times its weight in carats. That is, for a 0.99 carat diamond, we divide the price by 99 and for a 1 carat diamond, we divide it by 100. Then, we randomize the carat weight to the price values in order simulate the null distribution of differences in average prices of 0.99 carat and 1 carat diamonds. The null distribution (with 1,000 randomized differences) is shown below and depicts the distribution of differences in sample means (of price per carat) if there really was no difference in the population from which these diamonds came.7 (Wickham 2016)



Using the randomization distribution, conduct a hypothesis test to evaluate if there is a difference between the prices per carat of diamonds that weigh 0.99 carats and diamonds that weigh 1 carat. Make sure to state your hypotheses clearly and interpret your results in context of the data. (Wickham 2016)

Problem 2: [IMS] 20.6.5 Diamonds, bootstrap interval

We have data on two random samples of diamonds: 23 0.99 carat diamonds and 23 1 carat diamonds. Provided below is a histogram of bootstrap differences in means of price per carat of diamonds that weigh 0.99 carats and diamonds that weigh 1 carat. (Wickham 2016)

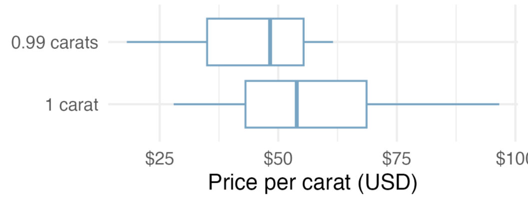


a. Using the bootstrap distribution, create a (rough) 95% bootstrap percentile confidence interval for the true population difference in prices per carat of diamonds that weigh 0.99 carats and 1 carat.
b. Using the bootstrap distribution, create a (rough) 95% bootstrap SE confidence interval for the true population difference in prices per carat of diamonds that weigh 0.99 carats and 1 carat. (The standard error of the bootstrap distribution is 4.64.)

Problem 3: [IMS] 20.6.9 and 20.6.11 Diamonds, mathematical test and mathematical interval

We have data on two random samples of diamonds: one with diamonds that weigh 0.99 carats and one with diamonds that weigh 1 carat. Each sample has 23 diamonds. Sample statistics for the price per carat of diamonds in each sample are provided below.

	Mean	SD	n
0.99 carats	\$44.51	\$13.32	23
1 carat	\$57.20	\$18.19	23



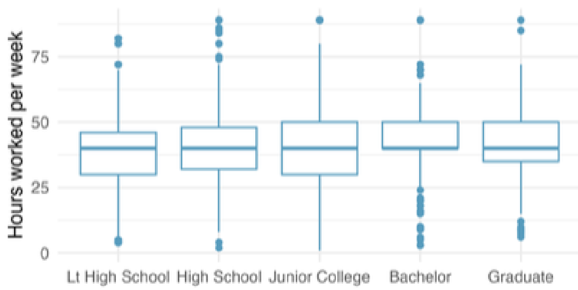
a. Conduct a hypothesis test using a mathematical model to evaluate if there is a difference between the prices per carat of diamonds that weigh 0.99 carats and diamonds that weigh 1 carat. Make sure to state your hypotheses clearly, check relevant conditions, and interpret your results in context of the data. (Wickham 2016)
b. Assuming that the conditions for conducting inference using a mathematical model are satisfied, construct a 95% confidence interval for the true population difference in prices per carat of diamonds that weigh 0.99 carats and 1 carat. (Wickham 2016)

Problem 4: [IMS] 22.5.10 Work hours and education

The General Social Survey collects data on demographics, education, and work, among many other characteristics of US residents. (NORC 2010) Using ANOVA, we can consider educational attainment levels for all 1,172 respondents at once. Below are the distributions of hours worked by educational attainment and relevant summary statistics that will be helpful in carrying out this analysis.

term	df	sumsq	meansq	statistic	p.value
`degree`	4	2,006	502	2.19	0.07
`Residuals`	1,167	267,382	229		

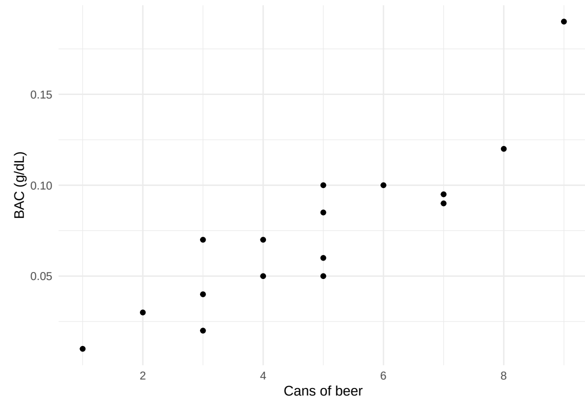
a. Write hypotheses for evaluating whether the average number of hours worked varies across the five groups.
b. Check conditions and describe any assumptions you must make to proceed with the test.
c. Below is the output associated with this test. What is the conclusion of the test?



Educational attainment	Mean	SD	n
Lt High School	38.7	15.8	121
High School	39.6	15.0	546
Junior College	41.4	18.1	97
Bachelor	42.5	13.6	253
Graduate	40.8	15.5	155

Problem 5: [IMS] 24.8.17 Beer and blood alcohol content

Many people believe that weight, drinking habits, and many other factors are much more important in predicting blood alcohol content (BAC) than simply considering the number of drinks a person consumed. Here we examine data from sixteen student volunteers at [The] Ohio State University who each drank a randomly assigned number of cans of beer. These students were evenly divided between men and women, and they differed in weight and drinking habits. Thirty minutes later, a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood. The scatterplot and regression table summarize the findings. (Malkevitch and Lesser 2008)



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.013	0.013	-1.005	0.332
beers	0.018	0.002	7.480	0.000

a. Describe the relationship between the number of cans of beer and BAC.
b. Write the equation of the regression line. Interpret the slope and intercept in context.
c. Do the data provide convincing evidence that drinking more cans of beer is associated with an increase in blood alcohol? State the null and alternative hypotheses, report the p-value, and state your conclusion.
d. The correlation coefficient for number of cans of beer and BAC is 0.89. Calculate R^2 and interpret it in context.
e. Suppose we visit a bar in our own town, ask people how many drinks they have had, and also measure their BAC. Would the relationship between number of drinks and BAC would be as strong as the relationship found in the Ohio State study? Why?