

ST 551 Midterm Exam

Due on Wednesday, November 2nd by 11:59PM

Student's Name Goes Here

Fall 2022

Instructions

Please complete this exam on your own and without assistance from anyone else except for the course instructor or GTA. Exams which are deemed to appear “too similar” will be investigated and a score of zero given for exams where academic dishonesty is suspected. **Exams should be typed using the provided template, not handwritten**, and you may use any materials that are posted on the course Canvas page. If you have any questions regarding the wording of the exams, please post those questions on the Midterm Exam Clarification Questions discussion board (But please do not post solutions or partial solutions in your posts).

Write all conclusions/interpretations of hypothesis tests using complete sentences and in the context of the question. That is, explain what the conclusion means for someone not well-versed in statistics.

You are encouraged to “show your work” so that even if your answer is wrong you may still receive partial credit.

Please do not make major modifications to the exam template (Except adding your solutions).

Submit your exam to Gradescope as a PDF before the deadline. Be sure to indicate which pages contain solutions to each question or group of sub-questions in Gradescope. Students that fail to correctly indicate the pages will be penalized as will those whose exams are “messy”, “disorganized”, or otherwise hard to read.

Question 1 (10 points)

For each question below, indicate whether the statement is TRUE or FALSE and provide a 1-2 sentence explanation as to why.

a. (2 points) In order for any hypothesis test to be considered *valid*, we must be able to show that the reference distribution of the test statistic is asymptotically normal as the sample size goes to infinity.

- Answer:
- Reason:

b. (2 points) The *power* of a statistical test describes the probability in which we reject a *true* null hypothesis.

- Answer:
- Reason:

c. (2 points) Researchers conduct a randomized experiment and find that tomatoes which grown using Fertilizer A were statistically larger (i.e. had a larger mean weight) than tomatoes grown using Fertilizer B. However, because the plants themselves were sampled in a non-random way, we can not make *causal* claims/inferences.

- Answer:
- Reason:

d. (2 points) In order to verify that two random variables X_1 and X_2 are independent we need to compute the covariance to check that it's equal to zero.

- Answer:
- Reason:

e. (2 points) Any test statistic which rejects the null hypothesis in favor of a one-sided alternative will also reject the null in favor of a two-sided alternative so long as the significance levels (α) are the same.

- Answer:
- Reason:

Question 2 (17 points)

Energy researchers are interested in measuring the typical residential power usage in a city. They conduct a small pilot study in which they randomly sample 6 newly built residences of similar size and measure the power usage of each residence for one month. The power used at each residence, in kilowatt hours (kWh), are shown below.

458	473	504	515	566	689
-----	-----	-----	-----	-----	-----

a. (2 points) The researchers would like to assess whether the *mean* power usage of residences in the city exceeds 470 kWh/month. Based on the researcher's question of interest, write out the appropriate null and alternative hypotheses for a statistical hypothesis test that addresses their question using correct statistical notation (No sentences please).

b. (4 points) Conduct a hypothesis test, by hand (i.e. you can use R code or mathematical notation, but do not use any built-in test functions like `t.test()`, `wilcox.test()`, `binom.test()`, etc.), to answer the researcher's question of interest from Part a. at the 5% significance level. Specifically, write out the value of the test statistic, the p-value, and a conclusion in the spaces provided below. For the conclusion, be sure to state the statistical outcome of the test, why you arrived at that conclusion, and what this conclusion means for the researchers using the context of the question. Please show all relevant work in the space indicated below.

- Test statistic:
- p-value:
- Conclusion:

Show your work for this question below. If you write R code, please make sure the computed test statistic and p-value are shown as outputs (And please don't leave a bunch of extraneous code/outputs in your solutions).

c. (3 points) Compute, by hand, and interpret a two-sided 95% confidence interval for the population mean power usage per residence (in kWh/month). Write your confidence interval in "interval form", i.e. (*lower bound*, *upperbound*), and interpret the interval in the context of this particular question. Show your work in the space indicated below.

- 95% confidence interval:
- Interpretation:

Show your work for this question below. If you write R code, please make sure the computed test statistic and p-value are shown as outputs (And please don't leave a bunch of extraneous code/outputs in your solutions).

d. (2 points) In discussing the outcome of your hypothesis test in part b. with the researchers, they inform you that they believe that the population distribution is likely to be heavily (and I mean *heavily*) right-skewed. Why does this new knowledge about the population distribution make you doubt the validity of the test you conducted in part b.? Be specific.

e. (4 points) Conduct a signed-rank test, by hand (i.e. you can use R code or mathematical notation, but do not use any built-in test functions like `t.test()`, `wilcox.test()`, `binom.test()`, etc.), to answer the researchers question of interest (About typical power usage exceeding 470 kWh/month) at the 5% significance level in light of the new information from Part d. Specifically, write out the value of the test statistic, the p-value, and a conclusion in the spaces provided below. For the conclusion, be sure to state the statistical outcome of the test, why you arrived at that conclusion, and what this conclusion means for the researchers using the context of the question. Please show all relevant work in the space indicated below and be specific about the parameter the test is addressing.

- Test statistic:
- p-value:
- Conclusion:

Show your work for this question below. If you write R code, please make sure the computed test statistic and p-value are shown as outputs (And please don't leave a bunch of extraneous code/outputs in your solutions).

f. (2 points) Are the results of the hypothesis test you conducted in Part e. valid inferences for the population of interest? Justify your answer.

Question 3 (13 points)

Every year, the U.S. Bureau of Labor Statistics (BLS) conducts the *American Time Use Survey* (ATUS) which asks respondents to monitor how much time they spend participating in various activities throughout a single 24-hour period. Inspired by ATUS, a high school math teacher asks their class of 20 students to monitor their time-usage for one day. After the students collect their data, the math teacher would like to determine if the *median* amount of time, in minutes, student spend on hygiene is different than 25 minutes. The time, in minutes, students reported spending on hygiene related activities is shown below.

8	12	12	12	13	13	14	16	17	17
17	18	21	22	24	26	28	29	31	48

a. (3 points) Based on the scenario above, what is the population, variable, and parameter of interest for this study?

- Population of interest:
- Variable of interest:
- Parameter of interest:

b. (2 points) Based on the teacher's question of interest, write out the appropriate null and alternative hypotheses for a statistical hypothesis test using correct statistical notation (No sentences please).

c. (4 points) Conduct an appropriate statistical hypothesis test, by hand (i.e. you can use R code or mathematical notation, but do not use any built-in test functions like `t.test()`, `wilcox.test()`, `binom.test()`, etc.), to answer the teacher's question of interest at the 5% significance level. Specifically, write out the value of the test statistic, the p-value, and a conclusion in the spaces provided below. For the conclusion, be sure to state the statistical outcome of the test, why you arrived at that conclusion, and what this conclusion means for the teacher using the context of the question. Please show all relevant work in the space indicated below.

- Test statistic:
- p-value:
- Conclusion:

Show your work for this question below. If you write R code, please make sure the computed test statistic and p-value are shown as outputs (And please don't leave a bunch of extraneous code/outputs in your solutions).

d. (2 points) What would making a Type I error mean in the context of this particular problem. Write your answer as if you were explaining what a Type I error would mean for this particular problem and not just *in general*.

e. (2 points) When asked about their interest in the using the median, the teacher explains that they did not believe that the population distribution was *exactly* normally distributed and that median was “more robust than the mean to outliers”. Is this *necessarily* a reason to avoid conducting a hypothesis test for the population mean? Explain your reasoning.