Stat 6336 Exam Fall 2022 (ChatGPT: 48 out of 100)

Answer all questions to the best of your ability. Point values are beside each question. You may use a calculator and your notes. Good Luck!

1. (10 points) Is exposure to a high cholesterol diet associated with heart disease? Examine the data from the table below and perform an appropriate analysis.

Table

Description automatically generated

Chat GPT – a general answer based on literature. It can’t read the table.

-10

1. (4 points) In a geology course, students were learning to use a balance scale to make accurate weightings of rock samples. Cameron plans to weigh a rock 20 times and then calculate the average of the 20 measurements to estimate her rock's true weight. Jordan plans to weigh a rock 5 times and calculate the average of the 5 measurements to estimate his rock's true weight. Will one student’s weight be closer to the rock’s true weight than the other student’s? If so, which student’s rock and why? If not, explain why not.

ChatGPT gives a rather long-winded, redundant, but correct answer. – 0 points.

1. (25 points) NAEP The table below shows a summary page from the National Assessment of Educational Progress (NAEP) website. It is a custom page that can be called up to perform a comparison of the user’s choice. I asked for information about the performance of Texas 8th graders for reading assessments in 2013 and 2015. The data given there allows the user to compare the performance of male and female students on these assessments. NAEP results consist of test scores from a probability sample of students from each state. Each year of testing, schools are selected at random from the state and students are selected at random from within each school. Note that estimates of means are given in the table, as well as their standard errors, which are computed by NAEP analysts, taking the sample design into account. Of great concern to policy makers is the “gap,” meaning the difference in mean achievement, between important reporting groups. Is there evidence that the gap between males and females in reading grew from 2013 to 2015? Support your argument with a data analysis. By the way, the sample size is greater than 1000 for each year.

Table

Description automatically generated

ChatGPT very apologetically says it can’t do this problem without specific numbers. It gives a method for calculating an answer to the question. (-20 points)

1. (10 points) In their article “Moving to a World Beyond p < 0.05”, Wasserstein, Schirm, and Lazar, state that researchers and practitioners should stop using the phrase “statistically significant” and its variants (p < 0.05, significantly different, nonsignificant, etc). Why do they say this and what would be better? You don’t have to list ALL their suggestions for improvement. Two or three suggestions will suffice.

ChatGPT gives a very nice answer. Full marks. 4% TurnitIn score.

1. (10 points) Rishi's project for her introductory statistics course was to compare the selling prices of textbooks at two Internet bookstores. He first took a random sample of ten textbooks used that term in courses at her college, based on the list of texts compiled by the college bookstore. Then, he obtained prices of the 10 randomly selected textbooks at the two Internet sites to test if the average prices of textbooks sold at the two internet bookstores are different. Advise Rishi on the type of test he should use to compare the selling prices from the two bookstores. Note: this question is more about your decision-making process for determining an appropriate test than it is about finding THE correct test.

Very nice and thorough answer, mentioning assumptions and caveats for various statistical tests.

Does not mention that a paired test, where the same 10 textbooks are taken from the two different websites, would be a better design.

Mentions that a larger sample size would be better.

1. (10 points) Table 1 of Rochon, et. al (2018) is reproduced as an appendix to this exam. For ONE distribution, summarize the results for that distribution. To give a valid summary, you will need to give some context to the table and explain the terms on the second column (sample size is in the first column).

Very politely explains that it can’t analyse data from tables. It needs the data input into its editor. -10 points.

1. (3 points) Explain the difference between robustness and resistance. Give an example of each property from one of the methods that we have studied.

Gives a good definition of each. Examples are wrong. Gives the median as a “robust” measure of central tendency, when it is really resistant. Robustness is not so much about outliers. I would give 1 point here.

1. (5 points) In class, we said that the null hypothesis for the runs test is where F and G are CDFs. Why is this equivalent to testing whether the data are randomly sampled?

Gives a very round-about answer, some of which is correct. It’s talking like a student who is just spitting out everything it knows about runs tests and random samples without answering the question. (-4)

1. (3 points) In a study of consumer choice conducted by the advertising department on the SMU campus, students were randomly assigned to one of two treatments: (1) small assortment and (2) large assortment. In each treatment, the participants were allowed to select a candy from an assortment of candy types once each day for 10 days. In (1), the assortment contained 6 types; in (2), the assortment contained 120 types. The number of different types (which we will call variety) of candy selected by each participant over the 10 days was the measured response. A test of the null hypothesis that the mean of the responses is the same in the two groups against a two-sided alternative was conducted. Suppose the p-value for the test was 0.21. Which of the following is the most accurate description of a valid conclusion?
2. That the size of the assortment offered consumer participants does not cause them to change the variety of their choices.
3. That the average variety of choices that would be selected by SMU students under the two conditions is the same.
4. That there is insufficient evidence that the size of the assortment caused consumer participants to change their variety seeking behavior.
5. That there is evidence that the size of the assortment caused consumer participants to alter their variety-seeking behavior.
6. (3 points) A Type I error occurs when:
7. We conclude that there is not an effect in the population when in fact there is.
8. We conclude that the test statistic is significant when in fact it is not.
9. The data we have typed into SAS is different from the data collected.
10. We conclude that there is an effect in the population when in fact there is not.

ChatGPT gets this correct with a nice (long winded) explanation.

1. (3 points) A 95% confidence interval for the difference in the means of two populations was calculated as (0.25, 2.40) using t-tools. Suppose that a t-test of the hypothesis vs. was carried out for the same data. The p-value must be which of the following?

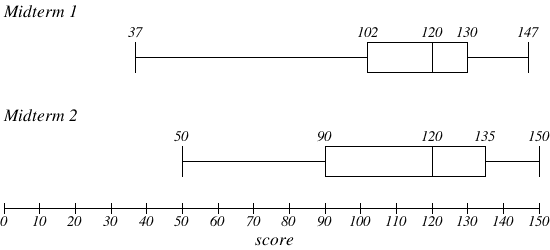
(i) > 0.05 (ii) < 0.05 (iii) < 0.025 (iv) > 0.025

States the correct answer, but does not explain why (iii) is not the correct answer.

1. (3 points) The t-tools applied to log-transformed data provide inferences about the difference in the means of logged measurements. Why is this problematic for interpreting the results on the original scale?
2. The median of the logged values does not equal the log of the median.
3. The log-transformed data are not symmetric.
4. The original data is not symmetric; therefore, the median does not equal the mean.
5. The mean of the logged values does not equal the log of the mean.

ChatGPT is correct here, although it is incorrect that option (a) is irrelevant because the t-test is a test for the mean rather than the median.

Use the boxplots below to answer questions 12 and 13.



1. (3 points) The boxplot above shows the grades of students in a statistics class on two midterms. Which midterm has a greater percentage of students with scores at or above 120?
2. Section A
3. Section B
4. Both sections are about equal.
5. It is impossible to tell this level of detail from a boxplot

ChatGPT is incorrect. It says you can’t tell, but 120 is. The median in both data sets. Of course, Chat GPT can’t see the plot, either. -3

1. (3 points) Refer again to the boxplot above. Which of the following is correct?
2. The means of both midterms are larger than their medians
3. The means of both midterms are smaller than their medians
4. The means of both midterms are about the same as their medians
5. There is no way to tell the relationship between mean and median from a boxplot

ChatGPT does not consider the skewness of the plot in answering this question -3

1. (3 points) What is one reason to perform an observational study instead of an experiment?
2. Observational studies tend to have better external validity than experiments.
3. Observational studies tend to have better internal validity than experiments.
4. Fewer subjects are required for observational studies.
5. Researchers do not need to worry about representative samples for observational studies.

ChatGPT is correct with good, concise explanations for each answer.

1. (3 points) How should confidence intervals be interpreted in reporting statistical results?
2. As another measurement of uncertainty like p-values
3. To show the effect sizes that are most compatible with the data under the given model.
4. Instead of p-values by determining whether the null value is in the interval.
5. They should not be used because they lead to the same bad interpretations as p-values.

ChatGPT is correct and explains each option correcty.