MATH 338

MIDTERM 2
WEDNESDAY, APRIL 11, 2018

Your name: _____

Your scores (to be filled in by Dr. Wynne):

Problem 1: ____/10

Problem 2: ____/6

Problem 3: ____/6

Problem 4: ____/6

Problem 5: ____/6

Total: ____/22

You have 60 minutes to complete and upload this exam.

You may refer to any notes/code you wrote, any files on Titanium, and software help/documentation. You may ask for clarification on a question, or for help troubleshooting Rguroo errors and/or debugging R code. You may not use online resources other than those local to Titanium or the software itself.

For full credit, include all R code (if using R/RStudio), graphs, and output. Save your answers as a .docx or .pdf file and upload the file to Titanium.

1. We are planning to obtain a simple random sample of size 50 from a population with known standard deviation $\sigma = 1$. We wish to perform a test of H_0: $\mu = 7$ against H_a: $\mu > 7$ at the 8% significance level.

A) [1.5 pts] Find the sampling distribution of (all possible) sample means when the null hypothesis is true. You may assume the Central Limit Theorem holds.

B) [1.5 pts] Suppose that we consider increasing our significance level from 8% to 10%. If this is the only change in the study methods, which of the following would also be guaranteed to increase? Indicate one or more correct answers below by using **BOLD AND UNDERLINE**, highlighting, and/or red text.

a. The probability of committing a Type I Error

b. The probability of committing a Type II Error

c. The power of the test

d. The p-value

e. The critical value

C) [1.25 pts] Suppose that, instead, we want to compute a 95% (two-sided) confidence interval for the population mean. Can we find the margin of error for this confidence interval prior to obtaining sample data? If so, compute it. If not, explain why not.

D) [1 pt] Which of the following changes to our study design would result in a smaller margin of error for our confidence interval, assuming other factors remain the same? Indicate one or more correct answers below by using **BOLD AND UNDERLINE**, highlighting, and/or red text.

a. Increasing the sample size to 100

b. Raising the confidence level to 99%

c. Increasing our measurement precision so that the population standard deviation is lower than expected

E) [0.5 pts] Select the best definition of 95% confidence. Indicate the single correct answer below by using **BOLD AND UNDERLINE**, highlighting, and/or red text.

a. We expect that 95% of all possible sample means will be contained in our confidence interval

b. We expect that 95% of all possible population means will be contained in our confidence interval

c. We expect that 95% of all possible confidence intervals will contain our sample mean

d. We expect that 95% of all possible confidence intervals will contain our population mean

For parts F-K, assume that we have now actually collected our data (under the original study design) and performed our hypothesis test.

F) [1 pt] We obtain a z test statistic of 1.55 and a p-value of 0.06. Are these results statistically significant (at our 8% significance level)? Why or why not?

G) [0.5 pts] Based on your answer to part (F), in which of the intervals below does the z critical value lie? Indicate the single correct answer below by using **BOLD AND UNDERLINE**, highlighting, and/or red text.

a. $(-\infty, -1.55)$

b. $(-1.55, 0)$

c. $(0, 1.55)$

d. $(1.55, \infty)$

e. The z critical value is exactly 1.55

H) [0.5 pts] Based on your answer to part (F), what should we conclude about the null hypothesis? Indicate the single correct answer below by using **BOLD AND UNDERLINE**, highlighting, and/or red text.

a. Reject the null hypothesis

b. Accept the null hypothesis

c. Fail to reject the null hypothesis

d. We do not have enough information to make a conclusion about the null hypothesis

I) [1 pt] Oops! We accidentally compute a t test statistic instead of a z test statistic! Which of the following will change due to this mistake? Indicate one or more correct answers below by using **BOLD AND UNDERLINE**, highlighting, and/or red text.

a. The significance level

b. The p-value

c. The critical value

J) [0.75 pts] Explain why using a z test statistic would give us more accurate results for this hypothesis test than using a t test statistic.

K) [0.5 pts] If we continue to use a t-test, how many degrees of freedom should we use in our calculations? Indicate the single correct answer below by using **BOLD AND UNDERLINE**, highlighting, and/or red text.

a. 49

b. 50

c. 98

d. None of the above

e. We do not have enough information to compute the degrees of freedom

CHOOSE TWO OF PROBLEMS 2-5 TO COMPLETE. YOU DO NOT NEED TO ANSWER ALL FOUR PROBLEMS. HIGHLIGHT ON THE COVER PAGE THE PROBLEMS YOU WISH ME TO GRADE.

For **two** of the problems on the next page:

A) [1 pt] Write the name of the statistical procedure you will use.

B) [1.5 pts] Justify why you are using this procedure (including exploratory data analysis and assumption checking where applicable).

C) [1 pt] Set your own confidence level/significance level, perform the procedure in the statistical software, and paste any appropriate code and output. If you are using a confidence interval, I should be able to read or infer your confidence level from your code/output. If you are using a hypothesis test, I should be able to read or infer your null/alternative hypothesis from your code/output.

Even if you are doing something wrong, paste the code and output you get (including error messages).

D) [1 pt] Restate the relevant part of the output (report the confidence interval as an interval, report the test statistic and p-value, report the power, etc.). Make sure to note your confidence level/significance level.

E) [1.5 pts] Interpret your results and/or make a conclusion in the context of the question asked. If you are skeptical about your results, explain why. Your answer to part (B) may help with this.

2. A sample of 541 Australian students (12-15 years old) took the Short Mood and Feelings Questionnaire (SMFQ), a test commonly used to measure depressive symptoms in young adolescents. Higher scores on the questionnaire indicate greater severity of depressive symptoms.

The file SMFQ.csv contains the gender and SMFQ scores of the 541 students. Assuming this is a representative sample of all young adolescent students, is there evidence that one gender (either boys or girls) experiences greater depression symptoms than the other?

3. A sample of 421 medical school students were asked if they would consider cheating on an exam or helping another student to cheat on an exam.

The file cheaters.csv contains the breakdown of students by gender (Male/Female) and willingness to consider cheating on the exam (Yes/No). The variable Num_Students represents the number of students at each combination of gender and willingness to cheat. Estimate how much more likely male medical students are to consider cheating than female medical students.

4. The file milk.csv contains the concentration of trace elements in human milk fed to a sample of infants in three different countries (Argentina, Poland, USA). All concentrations are in μg/L. All data from the USA were collected from a random sample of Boston-area mothers.

Suppose we believe that the mean arsenic (As) concentration in breast milk from Boston mothers is 3 μg/L, but we are concerned that it is increasing. Is a sample of 20 mothers (the sample size in the file) sufficient to detect an alternative mean concentration of 3.5 μg/L?

5. Gastroenteritis is a viral or bacterial infection that spreads through contaminated food and water. Suppose that inspectors wish to determine if the proportion of public swimming pools nationwide that fail to meet disinfectant standards is different from 10.7%, which was the proportion of pools that failed the last time a comprehensive study was done, 2008.

A simple random sample of 30 public swimming pools was obtained nationwide. Tests conducted on these pools revealed that 26 of the 30 pools had the required pool disinfectant levels. Perform an appropriate statistical procedure to help the inspectors.