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 σ = 1. We wish to perform a test of H_0: μ = 7 against H_a: μ > 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.

$\bar{x} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right) = N\left(7, \frac{1}{\sqrt{50}}\right)$ or $\bar{x} \sim N(7, 0.14)$

0.5 pts each for normal distribution, mean = 7, sd = 0.14

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 would be guaranteed to increase

b. The probability of committing a Type II Error would be guaranteed to decrease

c. The power of the test would be guaranteed to increase

d. The p-value would be unaffected by choice of significance level

e. The critical value would be guaranteed to decrease

0.75 pts each for highlighting a and c, -0.5 pts for each wrong answer highlighted

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.

0.25 yes we can find the margin of error

1 pt Margin of error $= z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (1.96)\left(\frac{1}{\sqrt{50}}\right) = 0.277$

Partial credit for using a wrong z* value or explaining that z* is unknown

Partial credit for finding the margin of error but then identifying the confidence interval incorrectly as $\mu_0 \pm E = (6.723, 7.277)$.

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 would result in a smaller margin of error

b. Raising the confidence level to 99% would result in a larger margin of error

c. Increasing our measurement precision so that the population standard deviation is lower than expected would result in a smaller margin of error

0.5 points each for highlighting a and c, -0.5 points for highlighting b

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?

1 pt yes, they are significant, because the p-value is less than our significance level.

Full credit if anyone actually computes the z critical value of 1.405 and compares it to $z_{obs} = 1.55$

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. (-∞, -1.55)

b. (-1.55, 0)

c. (0, 1.55) again, full credit if anyone actually computes the z critical value to be 1.405

d. (1.55, ∞)

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 is whatever we want it to be

b. The p-value will change

c. The critical value will change

0.5 points each for highlighting b and c, -0.5 points for highlighting a

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.

Knowing the population standard deviation means that the test statistics will follow a normal distribution under H_0. Thus, the appropriate test statistic is the z-score corresponding to the observed sample mean under H_0.

Full credit as long as something is mentioned about the population standard deviation being known

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.

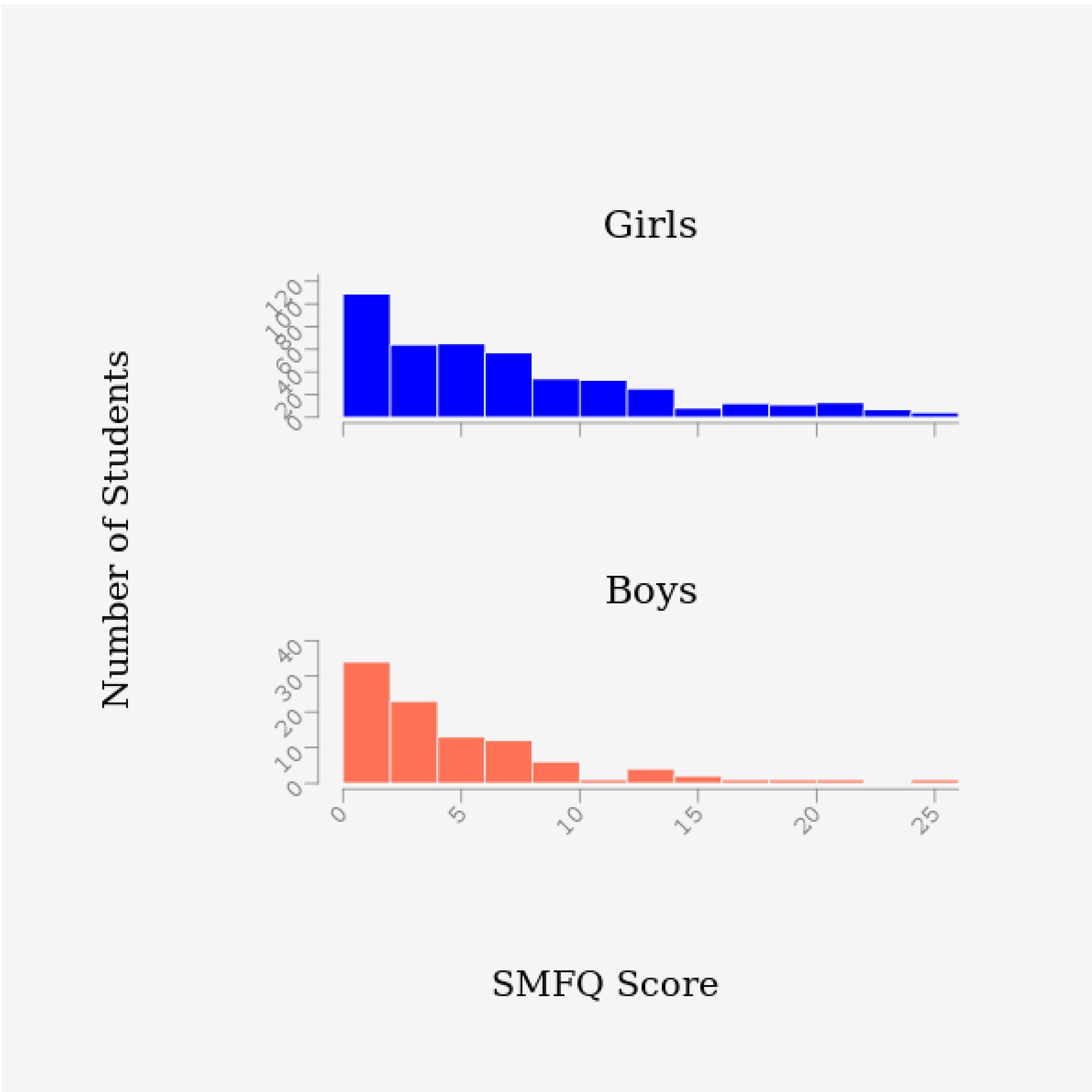
E) [1.5 pts] Interpret your results and/or make a conclusion in the context of the question asked. If you are skeptical about your results, explain why. Your answer to part (B) may help with this.

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?

A) The appropriate inference framework is two independent samples t for a difference of population means. 1 pt for either a two-sample t CI or a two-sample t test.

B) 0.5 pts each for 2 of the following 3 things: making the histograms (e.g., below), commenting on the extremely skewed distributions, commenting on the uneven sample size



0.5 pts for indicating that since sample size is large (combined sample size is 541), t procedures are still okay to use

C) 1 pt for, at minimum, one of the following tables:

Confidence Interval - t Distribution

95% Confidence interval

Variable	DF	Lower CL	Upper CL	Mean	Margin of Error
SMFQ (Female) - SMFQ (Male)	172.629	1.15270	3.42502	2.28886	1.13616

Test of Hypothesis (t-Test): 'SMFQ (Female) - SMFQ (Male)', Unequal Population Variances

Research Hypothesis Ha: Mean of 'SMFQ (Female) - SMFQ (Male)' is not equal to 0

Diff Means	Standardized Obs Stat	DF	P-value	97.5% Lower CL	97.5% Upper CL
2.28886	3.97633	172.629	0.000102771	1.15270	3.42502

Test is significant at 5% level.

D) 1 pt for one of the following:

95% Confidence interval: (1.153, 3.425) or equivalent

$t = 3.976$, $p = 0.0001 < 0.05$ so reject H0

E) 1.5 pts for one of the following, or similar interpretation:

We are 95% confident that the mean SMFQ score for girls is between 1.153 and 3.425 points higher than the mean SMFQ score for boys. Since this interval does not include 0, we suspect that there is a gender difference; in particular, that girls experience greater depression symptoms than boys.

We reject our null hypothesis of no gender difference. Therefore, we did find evidence of a gender difference in depression symptoms.

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.

A) The appropriate inference framework is for comparing a difference of population proportions. The question asks us to estimate the difference in proportions. 1 pt for a two-sample z CI.

B) 0.5 pts for checking the BINS assumptions in at least 1 population.

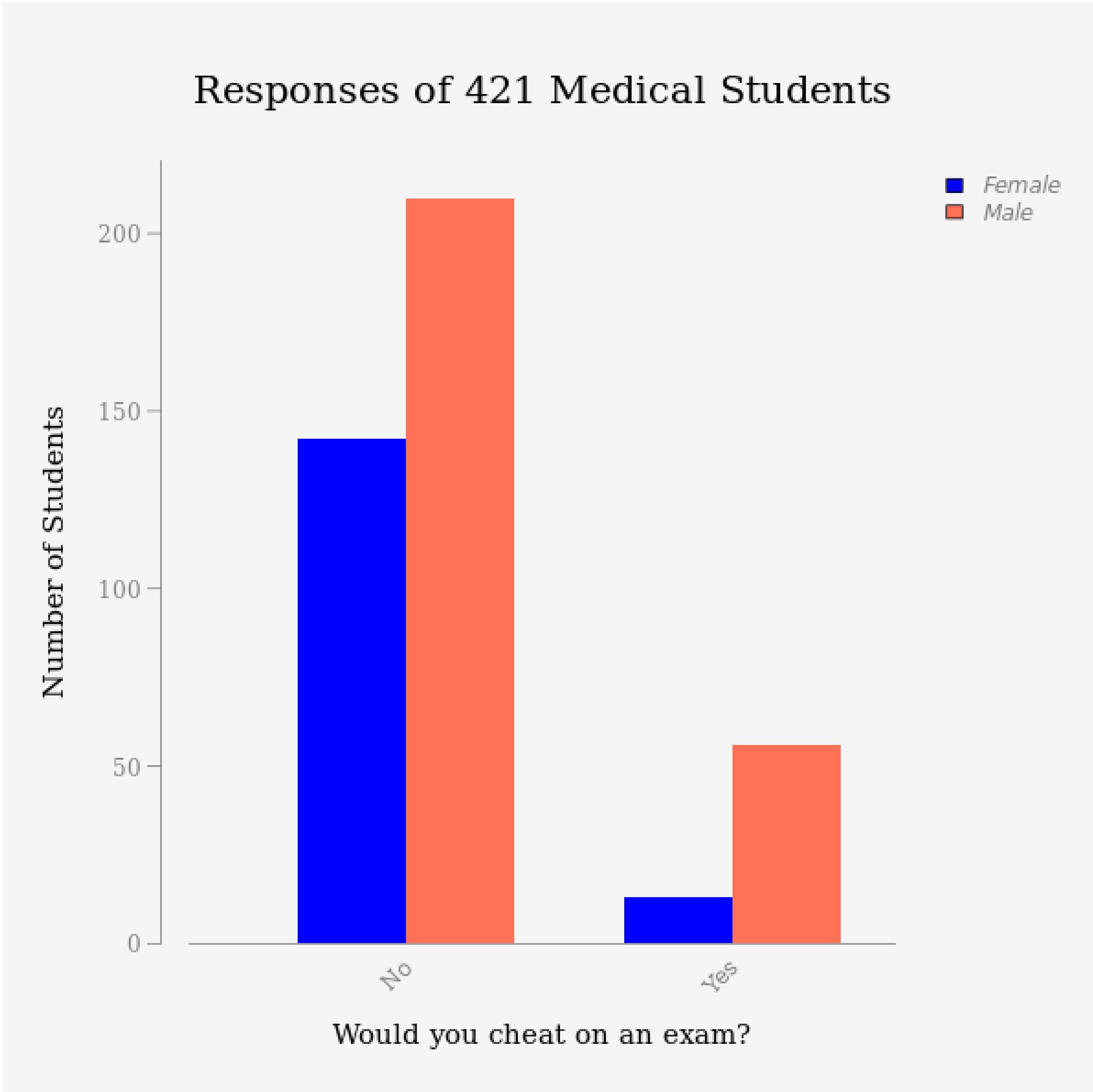
1 pt for checking the sample size assumptions: our smallest count is 13, so we have at least 10 successes and 10 failures in each sample, and so a z procedure is appropriate.

Output, such as the following table or bar graph, is nice but not strictly necessary.

Joint Totals of Would_Cheat and Gender

Row Variable is Gender
Column Variable is Would_Cheat

	Yes	No	Total
Male	56	210	266
Female	13	142	155
Total	69	352	421



C) 1 pt for, at minimum, the following table with one of the four rows listed below:

Confidence Interval for Difference of Two Population Proportions

Success = Yes
Population 1 = Male, Population 2 = Female
Sample Size: Male = 266, Female = 155
Number of Successes: Male = 56, Female = 13
Proportion of Success: Male = 0.2105, Female = 0.08387
Confidence level = 95%

Method	Lower CL	Upper CL	Midpoint	Width
Large Sample z	0.0610463	0.192264	0.126655	0.131218
Large Sample z with cc	0.0559408	0.197370	0.126655	0.141429
Wilson-Score	0.0636298	0.197008	0.130319	0.133378
Wilson-Score with cc	0.0625630	0.197256	0.129909	0.134693

cc: Continuity correction is used in computing the interval.

D) 1 pt for any of the following 95% CIs, or equivalent with a different confidence level:

(0.061, 0.192)

(0.056, 0.197)

(0.064, 0.197)

(0.063, 0.197)

E) 1.5 pts for the following, or similar interpretation:

We are 95% confident that male medical students are between 6.1 and 19.2 percentage points more likely than female medical students to consider cheating.

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?

A) 1 pt for a power analysis in the one mean t framework.

B) 1.5 pts for any reasonable explanation for why t power analysis is appropriate. It should include, at minimum, a description that we are in the process of study design and want to check whether our sample size is large enough (has enough power to detect our desired difference).

C) 1 pt for, at minimum, the following table. Note that this problem does not necessarily require you to subset the dataset, since you can choose Variable/By Factor and Population 1 Inference.

Power: t-Test for Mean; As (USA)

Research Hypothesis Ha: Mean of 'As (USA)' is greater than 3
Sample Size = 20
Standard Deviation = 0.842590813737963
Significance Level = 5%

Null	Alternative	Effect Size	Approx. Power	Exact Power
3	3.50000	0.593408	0.822431	0.819273

Approximate Power is computed via normal approximation.

D) 1 pt for reporting either the approximate power of 0.822 or the exact power of 0.819.

E) 1.5 pts for comparing the power to the standard of 0.8 and finding that, indeed, a sample of 20 subjects is sufficiently large to detect the alternative mean 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.

A) 1 pt for a one-proportion confidence interval or hypothesis test.

B) 0.5 pts for checking BINS assumptions.

1 pt for checking sample size assumptions. Define one of the following:

Success = having required levels, then we have 26 successes and 4 failures

Success = not having required levels, then we have 4 successes and 26 failures

Or, under the null hypothesis, we expect (0.107)(30) = 3.21 pools to fail to meet disinfectant standards and (0.893)(30) = 26.79 pools to meet disinfectant standards

We do not meet the sample size assumption for either a CI or HT and must use binomial exact procedures.

C) 1 pt for, at minimum, one of the following tables:

Confidence Interval for One Population Proportion

Success = Failed
Sample Size = 30
Number of Successes = 4
Proportion of Success = 0.1333
Confidence level = 95%

Method	Lower CL	Upper CL	Midpoint	Width
Binomial (Exact)	0.0375535	0.307218	0.172386	0.269665

Test of Hypothesis: Failed
Method: Binomial Exact Test

Research Hypothesis Ha: Proportion of 'Failed' is not equal to 0.107

Sample Size	No. of Successes	Sample Proportion	P-value	95% Lower CL	95% Upper CL
30	4	0.133333	0.556299	0.0375535	0.307218

Test is not significant at 5% level.

Confidence Interval for One Population Proportion

Success = Met
Sample Size = 30
Number of Successes = 26
Proportion of Success = 0.8667
Confidence level = 95%

Method	Lower CL	Upper CL	Midpoint	Width

Binomial (Exact)	0.692782	0.962447	0.827614	0.269665

Test of Hypothesis: Met
Method: Binomial Exact Test

Research Hypothesis Ha: Proportion of 'Met' is not equal to 0.893

Sample Size	No. of Successes	Sample Proportion	P-value	95% Lower CL	95% Upper CL
30	26	0.866667	0.556299	0.692782	0.962447

Test is not significant at 5% level.

D) 1 pt for any of the following, or equivalent:

Our 95% confidence interval is (0.693, 0.962) if success = met disinfectant levels, and (0.038, 0.307) is success = failed to meet disinfectant levels

Our hypothesis test gives a p-value of 0.556 as long as the correct alternative hypothesis is given (p != 0.893 if success = met disinfectant levels, and p != 0.107 if success = failed to meet disinfectant levels)

E) 1.5 pts for either of the following interpretations, or similar:

We are 95% confident that between 3.8% and 30.7% of pools fail to meet required disinfectant levels (or between 69.3% and 96.2% of pools meet disinfectant levels). Since this interval includes 10.7% (or 89.3%), we do not believe there is a significant difference from the last time the study was done.

Since our p-value is greater than our significance level (0.05 or any other reasonable number), we fail to reject our null hypothesis. The proportion of pools that fail to meet required disinfectant levels is not significantly different from 10.7%.