MATH 338

FINAL EXAM – LECTURE PORTION

THURSDAY, DECEMBER 20, 2018

Your name: ________________________

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

Problem 1:	____/6
Problem 2:	____/6
Problem 3:	____/5
Problem 4:	____/3.5
Problem 5:	____/6.5
Total:	____/26.5

You have 75 minutes to complete this exam.

You may refer to your (prepared in advance) formula sheets. You may ask Dr. Wynne to clarify what a question is asking for. You may not ask other people for help or use any other resources.

For full credit, show all work except for final numerical calculations (which can be done using a scientific or graphing calculator).

1. Suppose Dr. Wynne brings in a guessing game for “extra credit” on the final exam. A large prize is worth 4 extra credit points and Dr. Wynne claims you have a 5% chance of winning it. A small prize is worth 2 extra credit points and Dr. Wynne claims you have a 20% chance of winning it. However, if you don’t win a prize, you lose 1 point on the final exam.

a) [2 pts] Assuming Dr. Wynne is telling the truth about the game, let the random variable X be the number of points your exam grade is adjusted after playing the game. Write the distribution of X (using a table) and compute its (theoretical) mean.

x	P(X = x)
-1	0.75
2	0.2
4	0.05

$\mu_X = (0.75)(-1) + 2(0.2) + 4(0.05) = -0.15$ points

1 pt for the table, 1 pt for the mean

b) [1 pt] Based on your answers to part (a), would you play this game? Why or why not?

No, because I would expect to lose points.

Alternatively: Yes, because I need the points and am willing to risk the 75% chance of losing more points.

1 pt for any reasonable explanation

c) [1.5 pts] Suppose you play the game. Let Event A be “you win a small prize” and Event B be “you win a prize of any size.” Are Events A and B independent, disjoint, both independent and disjoint, or neither independent nor disjoint? Explain your answer.

Events A and B are neither independent nor disjoint. Clearly, if you win a small prize, you win a prize of any size. Therefore $P(B|A) = 1 > P(B) = 0.25$ and they are not independent. Since $P(A \text{ and } B) = 0.2$, they are not disjoint.

0.5 pts for “neither” and 0.5 pts each for proving “not independent” and “not disjoint”

d) [0.5 pts] What would be an appropriate test for determining whether Dr. Wynne is telling the truth about the probabilities of winning a large prize/small prize/no prize? (Just give the name the test)

(Chi-square) goodness of fit test

e) [0.5 pts] If you observed 100 students play the game and then did the test in part (d), how many degrees of freedom would the distribution of the test statistic have?

groups – 1 = 3 – 1 = 2 degrees of freedom

2. In the lab exam, you investigated a random sample of 200 adult males who were randomly assigned to one of two groups. One group received the PMC complex tablet (treatment) and the other received a sugar tablet (control). The systolic blood pressure was measured before and after the study and the change in blood pressure (to the nearest mmHg) for each subject was recorded.

a) [2 pts] Specifically for this study, why would a control group be necessary? What was the point in giving the sugar tablet to the men in that group?

0.5 pts We need to show that the PMC complex is the important factor in lowering the blood pressure

1.5 pts The tablet is administered in order to account for any placebo effect that might exist; for example, if subjects feel less anxious because they are receiving treatment, then their blood pressure might go down no matter whether the treatment actually works

b) [1.5 pts] Describe in context what the Type I and Type II errors are for this study. Don't perform any computations to compute their probabilities, just explain what they are in the context of the study.

0.5 pts Type I Error would be to conclude that the PMC tablet is effective when it is not

0.5 pts Type II Error would be to conclude that the PMC tablet is not effective when it actually is

0.5 pts for keeping in context (1 pt max for just stating the definitions)

c) [1 pt] In the context of this study and your answer to part (b), which error – Type I or Type II – would be the more detrimental (i.e., worse) error to commit? Defend your answer.

Any reasonable answer is fine. For example, if we consider that subjects might try literally anything to reduce their blood pressure, a Type I Error would convince them to pay lots of money for an ineffective treatment, while a Type II Error would be effectively telling them not to try an effective treatment.

d) [1.5 pts] On the lab exam, you should have found that the treatment group experienced a significant decrease in blood pressure of, on average, about 6.5 mmHg (7.4 mmHg better than control). However, medical doctors claim that the blood pressure needs to decrease by 20 mmHg for a blood pressure lowering treatment to provide a significant clinical advantage. Explain this discrepancy. Do you think the PMC complex is a practically significant treatment for lowering blood pressure?

1 pt The difference is statistically significant, but statistical significance does not imply practical significance

OR

1 pt With a large enough sample size, any difference can be shown to be statistically significant

0.5 pts No, the PMC complex is not a practically (clinically) significant treatment for lowering blood pressure

3. You are performing a genetic analysis to see if any of 18 different genes is associated with increased risk of Alzheimer's disease. For each gene, you perform a hypothesis test with significance level 0.05 and power = 0.90.

a) [2.5 pts] If <u>none</u> of the genes is associated with Alzheimer's disease and the genes are all independent, what is the probability of obtaining at least one significant result in your 18 tests?

1.5 pts P(no significant result) = $(18!)/(0!\ 18!) * (0.05)^0(0.95)^{18} = 0.397$

1 pt P(at least one significant result) = 1 – P(no significant result) = 0.603

1 pt total for working in the binomial framework even if everything is wrong

b) [1 pt] Based on your answer to part (a), why would running 18 different tests using the same data be a problem?

When we do 18 tests instead of 1, we have increased the probability of making a Type I error from 5% to over 60%!

1 pt for any reasonable description of the multiple comparisons problem

c) [1.5 pts] One method of correcting for running many tests on the same set of data – the Bonferroni correction – sets a new significance level for each test by dividing the original significance level by the number of tests. If you performed the Bonferroni correction, how would the following change for a <u>single</u> test? (Circle one answer per line)

Power of the Test:	INCREASE	DECREASE	NO CHANGE
Magnitude of Critical Value:	INCREASE	DECREASE	NO CHANGE
Observed Test Statistic Value:	INCREASE	DECREASE	NO CHANGE

4. In the straight_jeans datasets from the lab section, we were looking at the difference in price between men's and women's jeans. The people who collected the data did so with the intention of proving – using statistics – what everyone suspects: the "pockets" of women's pants are a sham.

They took a random sample of 20 brands, and for each brand, collected data for one pair of men's jeans and one pair of similar-size and similar-style women's jeans. The table below summarizes their results for the question, "Can an average-sized woman's hand fit in the front pocket of these pants?"

	Yes	No
Women's Jeans	1	19
Men's Jeans	20	0

A) [0.5 pts] In their sample, what proportion of women's jeans do <u>not</u> have front pockets that an average-sized woman can stick her hand in?

19/20 = 0.95 = 95%

Suppose the investigators wish to test the null hypothesis H_0: $p_1 = p_2$ against the alternative hypothesis H_a: $p_1 < p_2$, where p_1 is the proportion of women's jeans with a front pocket that a woman's hand can fit in and p_2 is the proportion of men's jeans with a front pocket that a woman's hand can fit in.

B) [2 pts] They would like to perform a large-sample z hypothesis test to compare the two proportions. However, when doing background and exploratory analysis, they find <u>at least two very serious violations</u> of the assumptions for this hypothesis test. Which assumptions are violated? How they are violated?

1 pt the samples are not independent

1 pt there are not at least 5 successes and 5 failures in each sample

C) [1 pt] Would these assumptions (or the equivalent assumptions) still be violated if they analyzed the data using a chi-square test of independence? Why or why not?

Yes – given these marginal totals we would not expect at least 5 counts in each cell, and we know that the samples are dependent

5. One extremely common use of regression models in science is to prepare a calibration curve for an instrument. Although for most instruments the calibration curve should be sigmoidal (S-shaped), over a narrow range the curve is approximately linear, and thus simple linear regression can be used.

To calibrate a real-time quantitative PCR platform, one introduces a solution with known concentration of DNA (represented by the variable "copy number") into the machine and measures the Ct value, which represents the time until a specific threshold is reached. Suppose that for a RT-QPCR instrument, the relationship between copy number and Ct is given by the equation

$$Ct = -5.576 \times \log_{10}(copy\ number) + 42.843$$

and that the R^2 value associated with this equation was computed to be 0.9863.

A) [1.5 pts] What is the <u>correlation</u> between $\log_{10}$(copy number) and Ct value?

1 pt square root of $R^2 = 0.9931 \rightarrow r = \pm 0.9931$

0.5 pts since the slope is negative, r = -0.9931

B) [1 pt] The $\log_{10}$ notation represents the base 10 logarithm; that is, for a copy number of 10,000, $\log_{10}$(copy number) = 4. For this instrument, predict the Ct value for a copy number of 10,000.

Ct = -5.576(4) + 42.843 = 20.539

A sample of 7 solutions was used to fit the calibration curve. For parts C-D, refer to the ANOVA table associated with this simple linear regression model, partially completed below.

Source	Df	Sum of Sq.	Mean Sq.	F	Pr (>F)
Regression (Model)	1	217.62	217.62	360.5	7.47e-06
Residuals (Error)	5	3.02	0.604		
Total	6	220.64			

C) [2 pts] Fill in the missing values in the table.

D) [2 pts] What are the null and alternative hypothesis associated with a hypothesis test that would use this ANOVA table? If you can write them mathematically, do so.

1 pt H_0: the mean of the Ct value is constant → therefore Ct is not linearly dependent on $\log_{10}$(copy number) → $\beta_1 = 0$

1 pt H_a: the mean of the Ct value is not constant → therefore Ct is linearly dependent on $\log_{10}$(copy number) → $\beta_1 \neq 0$