MATH 338

MIDTERM 2 - LECTURE PORTION

TUESDAY, NOVEMBER 6, 2018

Your name: ___________KEY_______________

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

Problem 1:	____/3
Problem 2:	____/6.5
Problem 3:	____/5.5
Problem 4:	____/5
Total:	____/20

You have 75 minutes to complete this exam.

You may refer to your (single-sided, prepared in advance) formula sheet. 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. Circle the most correct answer to the following multiple choice questions [0.5 pts each].

A) Which of the following statements is true of a single, properly done observational study?

a) It can prove that the changes in the explanatory variable cause changes in the response

b) It is not affected by lurking variables

c) It randomly assigns cases to control and treatment groups

d) None of these statements is true of a single observational study

B) Let z* represent the upper (positive) critical value for a two-sided hypothesis test. Let z** represent the critical value for a one-sided (right-sided) hypothesis test with the same significance level. Which of the following comparisons is true?

a) z* < z** b) z* > z** c) z* = z** d) we don't have enough information to know

For parts C-D, assume we perform a hypothesis test with significance level 0.01 and power 0.90.

C) What is the probability of committing a Type I Error on this hypothesis test?

a) 0.01 b) 0.10 c) 0.90 d) 0.99 e) we don't have enough information to find it

D) What is the probability of committing a Type II Error on this hypothesis test?

a) 0.01 b) 0.10 c) 0.90 d) 0.99 e) we don't have enough information to find it

E) Let t* be the critical value for a 90% confidence interval coming from a t-distribution with 12 degrees of freedom. Let t** be the critical value for a 90% confidence interval coming from a t-distribution with 20 degrees of freedom. Which of the following comparisons is true?

a) t* < t** b) t* > t** c) t* = t** d) we don't have enough information to know

F) Bootstrap t procedures are typically used when we check our assumptions and find that:

a) We do not have a simple random sample from our population

b) We are comparing more than two groups (more than two values of the explanatory variable)

c) We do, in fact, know the population standard deviation

d) The sampling distribution of the sample mean may not look anywhere close to normally distributed

2. In Bayesian statistics, the equivalent of a confidence interval is a credible interval. If the 95% credible interval for a parameter μ is (2, 8), Bayesian statisticians say that there is a 95% chance that μ is in the interval (2, 8).

a) [2 pts] Explain why this is an invalid (wrong) interpretation for a 95% (frequentist) confidence interval.

Various answers are acceptable including: μ is a fixed value and so the probability makes no sense (or is 100% or 0%); the language of confidence does not make probability statements about parameters; the random quantity in a confidence interval is not the parameter but the lower and upper bounds of the interval, etc.

b) [1 pt] If a 95% z confidence interval for the population mean μ is (2, 8), what is the margin of error for this confidence interval?

The margin of error is half the width, or margin of error = 3

c) [2 pts] Suppose that the population standard deviation is known to be $\sigma = 10$. How many people would you need to sample to obtain a margin of error of at most the value you calculated in part (b)? [If you did not answer part (b), make up a maximum margin of error and solve below]

1 pt for using the correct formula: $n \geq (z^*\sigma/E)^2$

1 pt for correctly plugging in: $n \geq ((1.96)(10)/3)^2$ or $n \geq 42.68 \rightarrow$ 43 subjects.

d) [1.5 pts] If you obtain a single sample and, based on that sample, compute a 95% z confidence interval for μ to be (2, 8), what is the probability that your observed sample mean $\bar{x}$ is in the interval? Explain your answer.

0.5 pts: The probability that the observed sample mean is in the interval is 1, or 100%

1 pt: The observed sample mean is exactly at the center of the interval, so it is guaranteed to be in the interval

OR

1 pt: The observed sample mean is 5

0.5 pts: Therefore, it must be in the interval with probability 100%

3. Govers and colleagues (2018) investigated the heredity and effect of inheriting protein aggregates (PA) in *E. coli*. They subjected *E. coli* cells to a 15-minute heat shock, causing PA to form in the parent cells. However, not all children cells received the PA.

The authors performed a t-test to show a significant difference in growth rate ($p = 0.00167$) between cells that contained PA and cells that were PA-free. However, they found this difference was not statistically significant once the age of the cells was accounted for ($p = 0.56$). Cells were assumed to be independent (i.e., this is not a paired comparison).

a) [1 pt] Draw lines connecting each variable on the left with the term on the right that best describes it.

Presence/absence of PA — Explanatory Variable

Growth Rate — Response Variable

Age — Covariate

b) [2 pts] Write the null and alternative hypothesis for the t-test the researchers performed as an equation/inequality. If you use numeric subscripts (e.g., μ_1), explain what the means represent.

Null Hypothesis (H_0):

1 pt: $\mu_{PA} = \mu_{PA\ free}$, or $\mu_{PA} - \mu_{PA\ free} = 0$, or equivalent with labels

Alternative Hypothesis (H_a):

1 pt: $\mu_{PA} \neq \mu_{PA\ free}$, or $\mu_{PA} - \mu_{PA\ free} \neq 0$, or equivalent with labels

c) [1.5 pts] Explain in your own words what "p = 0.00167" means in the context of this particular experiment. Note that we have not specified a significance level!

If there is no difference in the population mean growth rate, then there is a 0.167% chance that we would observe a t-statistic (or a difference in sample mean growth rates) greater in magnitude (absolute value) than the actual t-statistic we observed.

0.5 pts for each component of the statement: correct conditioning (on H_0 being true), correct probability statement, correct context

0.5 pts total for comparing to an unknown significance level

d) [1 pt] What is the smallest significance level at which the difference would still be significant?

1 pt 0.00167 (if $p \leq \alpha$ then we have a significant difference, so for any $\alpha \geq 0.00167$ the observed p-value of 0.00167 will give us a significant result)

4. The chickwts dataset we used in the “Practice Lab Exam” for Midterm 1 recorded the weights of 71 chicks (in g) after six weeks. The chicks’ diet included one of six different feed supplements (casein, horsebean, linseed, meatmeal, soybean, or sunflower).

a) [2 pts] Write the null and alternative hypothesis for an appropriate one-way ANOVA F test for determining whether there is no effect of feed supplement on the population mean chick weight. If you use numeric subscripts (e.g., μ_1), explain what the means represent.

Null Hypothesis (H_0):

1 pt: $\mu_{casein} = \mu_{horsebean} = \mu_{linseed} = \mu_{meatmeal} = \mu_{soybean} = \mu_{sunflower}$, or equivalent labels

Alternative Hypothesis (H_a):

1 pt: At least one of the feeds produces a different population mean weight (or equivalent statement)

b) [2 pts] I performed the appropriate one-way ANOVA F test and obtained the ANOVA table below. Fill in the missing values in the table.

Source	df	Sum of Sq.	Mean Sq.	F	Pr (>F)
feed (Group)	5	231129	46226	15.37*	5.94e-10
Residuals (Error)	65	195556**	3009		
Total	70	426685			

0.5 pts per number

*Due to typo on original exam, 15.49 is also acceptable

**Obvious typos with the right conceptual approach were given credit – I can’t judge for this problem

c) [1 pt] Assume my significance level for the one-way ANOVA test is 0.01. Should I perform post hoc tests to investigate the pairwise differences in population mean chick weights? Why or why not?

1 pt: It is appropriate to perform post hoc tests here because we have rejected our null hypothesis that the population means are all equal; therefore, we want to test which population means are different.