MATH 338

MIDTERM 1 - LECTURE PORTION

THURSDAY, OCTOBER 4, 2018

Your name: _____KEY_____

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

Problem 1: ____/3.5

Problem 2: ____/7

Problem 3: ____/5

Problem 4: ____/5

Problem 5: ____/6.5

Total: ____/27

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 graph types would not be appropriate for displaying the distribution of a categorical variable?

a) Pie chart

b) Bar chart

c) Stem plot

d) All of these would be appropriate

B. The variable *y* has a mean of 50 and a standard deviation of 10. The variable $z = 100 + 3y$ has…

a) Mean 150 and standard deviation 30

b) Mean 250 and standard deviation 30

c) Mean 150 and standard deviation 130

d) Mean 250 and standard deviation 130

C. Which of the following numbers is not a valid probability?

a) 0

b) 20%

c) 3/8

d) -0.6

D. If the random variable *X* takes the value 3 with probability 0.4 and the value 6 with probability 0.6, then the expected value of *X* is:

a) 2.16

b) 4.5

c) 4.8

d) 6.0

E. Which of the following statements is not true about the normal density curve?

a) The total area under the curve is 1

b) The mean and median of the curve are equal

c) The curve never goes below the x-axis

d) All of these statements are true

F. If $W \sim N(200, 50)$, then the standard deviation of the sample mean $\overline{W}$ of 25 iid draws from *W* is:

a) 2

b) 10

c) 50

d) we don't have enough information to find it

G. Which of the following statements is true about the Central Limit Theorem?

a) It applies to both continuous and discrete random variables

b) It applies to populations with infinite standard deviations

c) It guarantees that the sampling distribution of the sample mean is approximately normal, as long as the sample size is at least 30

d) It guarantees that the sample mean is a biased estimator of the population mean

2. In a certain distribution, a value of 80 corresponds to a z-score of -2.

A) [1.5 pts] You do not have enough information to find the population mean exactly. Is it possible that the population mean is 70? Explain your answer.

0.5 pts: No

1 pt: Since the sign of the z-score is negative, the value of 80 must be below the population mean.

Therefore, the mean must be above 80 and so the population mean cannot be 70.

B) [2.5 pts] If you knew that the population distribution was normal, could you estimate the probability of randomly selecting an individual with a value less than 80? If so, estimate it or (if you need software to do it) explain how you would estimate it. If not, explain why not.

0.5 pts: Yes

Remaining 2 pts for either of the following:

- By the 68-95-99.7 rule, 95% of values are within 2 standard deviations of the population mean, or equivalently, they have z-scores between -2 and 2. Therefore, about 2.5% of values have a z-score less than -2, or equivalently, about 2.5% of values are less than 80
- Drawing the standard normal curve, marking a z-score of -2, and shading the area to the left

2 pts total (out of 2.5) for drawing the original normal curve and indicating that we don't have enough information. We don't have enough information to solve the problem this way, but we do have enough information to solve the problem another way.

C) [3 pts] If you did not know the shape of the distribution, but did know the value and corresponding z-score for a second case, could you find the population mean? If so, explain how. If not, explain why not.

0.5 pts: Yes

1 pts: Set up the equation $z = \frac{x-\mu}{\sigma}$ or $x = \sigma z + \mu$

1.5 pts: When we plug in our known x-values and corresponding z-scores, we have two linear equations in two unknowns (σ and μ). Thus, we can use any method for solving a system of linear equations to find the population mean μ.

3. Let Event A and Event B be arbitrary independent events. The probability of Event A is 0.3 and the probability of Event B is 0.8.

A) [1 pt] What is $P(A^C)$?

0.5 pts for setup and 0.5 pts for computation: $P(A^C) = 1 - P(A) = 1 - 0.3 = 0.7$

B) [1.5 pts] What is $P(A \cup B)$?

0.5 pts for setup: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

1 pt because A and B are independent, we have:

$$P(A \cup B) = 0.3 + 0.8 - (0.3)(0.8) = 0.86$$

C) [1 pt] What is $P(A \mid B)$?

1 pt: Because A and B are independent, $P(A \mid B) = P(A) = 0.3$

D) [1.5 pts] Can A and B be disjoint? Justify your answer.

0.5 pts: No

1 pt for either of the following explanations

- $P(A \cap B) = (0.3)(0.8) = 0.24 > 0$
- A and B are independent, and both are possible, so they cannot be disjoint

4. A game on the HQ Trivia app consists of 12 multiple-choice questions. Each question has 3 possible answers. The game ends as soon as you get a question wrong; for example, if you get question 6 wrong, you don't see questions 7-12. Your young nephew has gotten a hold of your phone and is touching answers completely at random.

Let X be the number of questions your nephew answers correctly.

A) [1 pt] Explain why X cannot be a binomial random variable.

0.5 pts for any attempt to check the BINS assumptions

Remaining 0.5 pts for explaining why the I and/or N assumptions are violated

B) [1 pt] Explain how to fix the HQ Trivia rules so that X can be a binomial random variable.

1 pt for any reasonable explanation. The simplest is to have everyone view all 12 questions.

C) [1.5 pts] Using your solution to part (B), identify the parameters n and p of the distribution of X.

0.5 pts: n = number of questions answered = 12 (or a different number depending on part B)

1 pt: p = probability of a correct answer = 1/3 (or a different number depending on part B)

D) [1.5 pts] Using your solution to part (C), how many questions would you expect your nephew to answer correctly?

0.5 pts: E(X) = np

1 pt for correctly plugging in your answers to part (C)

The most likely answer is np = (12)(1/3) = 4

5. Read the following excerpt from an Ig Nobel Prize-winning paper:

"The analysis included 64 participants who had travelled over speed bumps on their journey to hospital. Of these, 34 had a confirmed histological diagnosis of appendicitis, 33 of whom reported increased pain over speed bumps. The sensitivity was 97%...the specificity was 30%...the positive predictive value was 61%..."

A. [1 pt] What was a case in this study?

1 pt: A case is a patient (with pain, who traveled over speed bumps on the way to the hospital)

Partial credit (0.5 pts) for anything too specific, for example, patients with appendicitis

B. [0.5 pts] The proportion of the 64 patients who were diagnosed with appendicitis is a… (circle one)

a) sample proportion

b) population proportion

c) conditional probability

d) population proportion *and* conditional probability

C) [1 pt] What proportion of patients who reported increased pain over speed bumps were diagnosed with appendicitis?

0.5 pts: We want an estimate of P(appendicitis | increased pain over speed bumps)

0.5 pts: In this situation, that is an estimate of the positive predictive value, or 61%

D) [4 pts] From the data, estimate the negative predictive value of the speed bump test for appendicitis. Give your answer rounded to the nearest percent.

0.5 pts for using the information in the paragraph to set up a two-way table, tree diagram, or Bayes's Rule

2.5 pts for filling in the counts/probabilities correctly, e.g.,

	Increased Pain	No Increased Pain
Appendicitis	33	1
No Appendicitis	21	9

1 pt: negative predictive value = TN/(TN + FN) = 9/(9 + 1) = 90%