Fall 2023 Makeup Midterm

ECON 340: Economic Research Methods   Instructor: Div Bhagia

Print Name: _____

This is a closed-book test. You may not use a phone or a computer.

Time allotted: 70 minutes

Total points: 20

Please show sufficient work so that the instructor can follow your work.

I understand and will uphold the ideals of academic honesty as stated in the honor code.

Signature: _____

Question 1: Multiple Choice Questions (1 pt each, total 5 pts)

Choose a single correct response for all questions except for (a), which allows for the selection of multiple answers.

(a). (Select all that apply) Suppose a survey of 500 people found that 300 of them prefer coffee over tea. Suppose we define a variable X that takes a value of 1 if a person said they prefer coffee and 0 otherwise. The sample mean of X is given as:

  □ $1/500(1 \cdot 300 + 0 \cdot 200)$

  □ $(1/500) \sum_{i=1}^{500} X_i$

  □ $1 \cdot (3/5) + 0 \cdot (2/5)$

  □ None of the above

(b). Which of the following statements explains why a random sample is preferred when calculating the sample mean?

  □ A random sample ensures that the sample mean equals the population mean.

  □ A random sample reduces the variance of the sample mean.

  □ A random sample avoids bias in the selection of individuals for the sample.

  □ A random sample guarantees that the sample mean is normally distributed.

(c). Which of the following statements correctly explains the concept of independence between two random variables?

  □ Two random variables are independent if they have no outcomes in common.

  □ Two random variables are independent if they have exactly the same outcomes.

  □ Two random variables are independent if the value of one random variable does not affect the probability distribution of the other random variable.

  □ Two random variables are independent if they have the same probability distribution.

(d). If the average score on a math test is 75 with a standard deviation of 5, and your score is 85, how many standard deviations above the mean are you?

□ 1

□ 2

□ 3

□ 4

(e). Which of the following statements correctly explains the concept of p-values in hypothesis testing?

□ The p-value is the probability that the null hypothesis is true.

□ The p-value is the probability of observing the null hypothesis given the sample data.

□ The p-value is the probability of observing the sample data or more extreme values, assuming the null hypothesis is true.

□ The p-value is the probability of observing a statistically significant result in a hypothesis test.

Question 2: Things you can explain. (5 pts)

(a). (2.5 pts) Suppose that in a certain population, 50% of likely voters are women. If a survey is conducted using a random sample of 1000 landline telephone numbers and the results show that 57% of respondents are women, is there any indication that the survey may be biased? Explain.

(b). (2.5 pts) If a study finds a strong positive correlation between the number of houses and house prices across US cities, can we conclude that more housing supply leads to higher house prices?

Question 3: Exercise and Sleep. (10 pts)

Let X be the average hours of sleep per day you got last week, and let Y be the average hours you exercised per day last week. You want to look at the relationship between these two variables over the last three weeks.

Note: *Use the population formulas. You can use the table and fill in values or directly apply the formulas. In both cases, write down the formula you are using and show your work.*

Week	X_i	Y_i	$(X_i - \mu_X)$	$(Y_i - \mu_Y)$	$(X_i - \mu_X)^2$	$(Y_i - \mu_Y)^2$	$(X_i - \mu_X)(Y_i - \mu_Y)$
1	6	0.5					
2	9	0.3					
3	9	1					

(a). (2 pts) Calculate the variance of X and Y.

(b). (1 pt) In the formula for variance, why do we use squared deviations from the mean instead of just using deviations from the mean?

(c). (2 pts) Calculate the covariance between X and Y, denoted by σ_{XY}.

(d). (2 pts) In class, we learned that covariance is positive when two variables move together, meaning that they increase or decrease together. Can you explain how the formula you used in (c) ensures that this is the case?

Now say instead of recording the exercise in hours, you had recorded it in minutes. Then your data would look as below, where Z is the average minutes of exercise per day.

Week	X_i	Z_i	$(X_i - \mu_X)$	$(Z_i - \mu_Z)$	$(X_i - \mu_X)^2$	$(Z_i - \mu_Z)^2$	$(X_i - \mu_X)(Z_i - \mu_Z)$
1	6	30					
2	9	18					
3	9	60					

(e). (1 pt) Calculate the covariance between X and Z, denoted by σ_{XZ}.

(f). (1 pt) Why do you think it is the case that $\sigma_{XZ} > \sigma_{XY}$?

(g). (1 pt) There is an alternative statistic that you could calculate, which would ensure that you accurately capture that the relationship between X and Y is as strong as the relationship between X and Z. What is this statistic?

Extra credit (1 pt): Calculate this statistic for X and Y and for X and Z.