Fall 2023 Midterm Solutions

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.

(a) What does the variance of the sample mean signify?

☐ Dispersion in sample data

☑ Sample-to-sample variability

☐ Spread of the population data

☐ Bias in sample mean

(b) We are choosing between two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ for some population parameter θ, both of which are unbiased i.e. $E(\hat{\theta}_1) = \theta$ and $E(\hat{\theta}_2) = \theta$. But the variance of $\hat{\theta}_1$ is lower than that of $\hat{\theta}_2$ i.e. $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$. Which of the following is true?

☐ We are indifferent between the two estimators.

☑ We prefer $\hat{\theta}_1$ over $\hat{\theta}_2$.

☐ We prefer $\hat{\theta}_2$ over $\hat{\theta}_1$.

☐ We need more information to reach any conclusion.

(c) If the average score on a math test is 75 with a standard deviation of 10, and your score is 70, then you are:

☐ 1/2 standard deviation over the mean

☐ 1 standard deviation over the mean

☑ 1/2 standard deviation below the mean

☐ 1 standard deviation below the mean

(d) In a scientific study, researchers are analyzing the correlation between the variable X, which represents the age of trees, and Y, which measures the height of those trees in centimeters. The researchers also create a new variable Z that measures the height of the trees in inches. Which of the following is true?

☐ $\text{cor}(X, Y) > \text{cor}(X, Z)$

☐ $\text{cor}(X, Y) < \text{cor}(X, Z)$

☒ $\text{cor}(X, Y) = \text{cor}(X, Z)$

☐ None of the above

(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) If a study discovers a strong correlation between education level and wealth, can we infer that higher education leads to increased wealth?

No, a strong correlation between education level and wealth does not imply that higher education results in increased wealth. One possible explanation for this correlation could be reverse causality; that is, individuals from wealthy backgrounds may have the resources to pursue higher education.

Alternative answer:

No, a strong correlation between education level and wealth does not imply that higher education results in increased wealth. Other factors, such as parental involvement, can influence both education and wealth, thereby leading to the observed correlation. Parents who are actively involved in their children's lives may provide the necessary resources and support to attain higher education. Additionally, these involved parents may also impart financial literacy skills or even offer initial financial support, setting the stage for future wealth accumulation.

(b) (2.5 pts) Imagine a university reports that the average GPA of its student body is 3.4, based on university-wide data. A student organization surveys a random sample of 100 students and finds that the average GPA is 3.2. Should the student organization be concerned that they found a different average GPA than the university? Explain your reasoning, considering factors such as sampling variation and sample size.

The student organization should not be overly concerned about the observed discrepancy between their sample average GPA and the university's reported average. When taking a sample from a larger population, some degree of variation between sample statistics and the true population parameters is to be expected. In this case, the sample size of 100 is substantial enough for the Central Limit Theorem to be applicable, suggesting that the sample mean is drawn from a Normal Distribution centered around the true mean. Given this, one could use the sample standard deviation to calculate the variance of the sample mean and conduct a formal hypothesis test to see if the true mean is indeed 3.4. Alternatively, confidence intervals could also be constructed to provide a range within which the true population mean is likely to lie, with a specified level of confidence.

Question 3: Investor Behavior (6 pts)

Imagine you're a financial advisor studying two types of investors: *risk-averse* (denoted by $X = 1$) and *risk-taking* (denoted by $X = 2$). You're interested in the amount of money Y they are willing to invest in a new startup: either $10,000, $20,000, or $30,000.

The joint probability distribution table below shows the likelihood of an investor from each risk category investing a particular amount in a new startup:

	Investment Amount (Y)			Total
	$10,000	$20,000	$30,000	
Risk-averse ($X = 1$)	0.35	0.15	0.10	0.60
Risk-taking ($X = 2$)	0.05	0.15	0.20	0.40
Total	0.40	0.30	0.30	1

Answer the following questions. *For each question write down the formula you are using if any and show your work clearly.*

(a) (1 pt) What proportion of all investors invest $10,000?

$$Pr(Y = 10000) = 0.4$$

(b) (1 pt) What is the expected amount of investment across all investors?

$$E(Y) = \sum_y y Pr(Y = y) = 0.4(10000) + 0.3(20000) + 0.3(30000)$$

$$= 4000 + 6000 + 9000 = \$19,000$$

(c) (1 pt) What proportion of investors are risk-averse?

$$Pr(X = 1) = 0.6$$

(d) (1 pt) What *proportion* of risk-averse investors invest $10,000?

$$Pr(Y = 10000|X = 1) = \frac{Pr(Y = 10000, X = 1)}{Pr(X = 1)} = \frac{0.35}{0.60} = 0.5833$$

(e) (1 pt) What is the expected amount of investment from a risk-averse investor?

$$\begin{aligned} E(Y|X = 1) &= \sum_y yPr(Y = y|X = 1) \\ &= \frac{0.35}{0.60}(10000) + \frac{0.15}{0.60}(20000) + \frac{0.10}{0.60}(30000) \\ &= 5833.33 + 5000 + 5000 = \$15,833.33 \end{aligned}$$

(f) (1 pt) Given your answers above, do you think the investment amount is independent of an individual's risk appetite?

No individual's risk appetite is not independent of the investment amount as $E(Y|X) \neq E(Y)$.

Question 4: Estimating Average Income (4 pts)

In an economics study, researchers are interested in estimating the average monthly income of households in a particular city. They take a random sample of 150 households and find the average monthly income to be $4,500 with a standard deviation of $950.

(a) (1.5 pts) Construct a 95% confidence interval for the average monthly income. (Note that: $Pr(Z > 1.96) = 0.025$.)

$$\bar{x} \pm \underbrace{1.96 \frac{S}{\sqrt{n}}}_{\text{Margin of Error}} = 4500 + 1.96 \frac{950}{\sqrt{150}}$$

So the 95% confidence intervals is given by: $[4348, 4652]$.

(b) (1 pt) Interpret the 95% confidence interval you've constructed in part (a). What does it mean in the context of the study?

We are 95% confident that the average monthly income for households in this city lies between $4,348 and $4,652.

(c) (1.5 pts) If the researchers wanted to decrease the width of the confidence interval while keeping the confidence level constant, what could they do?

In the confidence interval, the margin of error is influenced by the variance of the sample mean estimator, which is inversely related to the sample size. If researchers wanted to decrease the width of the confidence interval while keeping the confidence level constant, they could increase the sample size. A larger sample size would reduce the variance of the sample mean estimator, thus narrowing the confidence interval.