Midterm Exam

EC031: Introduction to Econometrics — Spring 2026

Name: ______________________________

Instructions:

- **Part 1 (20 points):** Answer all 10 True/False questions.
- **Part 2 (80-85 points):** There are 5 questions. **Choose and answer only 4 out of 5.** Clearly cross out the question you are skipping. You will not receive extra credit for answering all five.
- Show all work for math questions. Writing only the final answer will not receive full credit.
- For conceptual questions, 3–5 clear sentences is sufficient.
- The amount of space provided is not indicative of how much is needed to answer a question.

> **i Tip**
>
> Before starting, read through all five questions in Part 2.

Part 1: True or False [20 points]

Circle **T** *if the statement is true or* **F** *if the statement is false. Each question is worth 2 points. No justification required.*

1. **T / F** The interquartile range (IQR) is resistant to outliers.

Answer: True. The IQR is based on the 25th and 75th percentiles, which are not affected by extreme values.

2. **T / F** If events A and B are mutually exclusive and both have positive probability, then they are also independent.

Answer: False. Mutually exclusive means $P(A \cap B) = 0$, but independence requires $P(A \cap B) = P(A)P(B)$. If both have positive probability, $P(A)P(B) > 0 \neq 0$, so they cannot be independent.

3. **T / F** The expected value of the sample mean $\bar{x}$ equals the population mean μ.

Answer: True. The sample mean is an unbiased estimator of the population mean: $E(\bar{x}) = \mu$.

4. **T / F** The variance of a Binomial(n, p) random variable is np.

Answer: False. The variance of a Binomial(n, p) random variable is $np(1-p)$, not np. The expression np is the expected value.

5. **T / F** The Central Limit Theorem states that the sampling distribution of $\bar{x}$ approaches a normal distribution as sample size increases, regardless of the shape of the population distribution.

Answer: True. This is the statement of the Central Limit Theorem (assuming finite variance).

6. **T / F** The standard error of $\bar{x}$ is $\sigma/\sqrt{n}$, which decreases as the sample size n increases.

Answer: True. As n increases, $\sqrt{n}$ increases, so $\sigma/\sqrt{n}$ decreases.

7. **T / F** For a normal distribution, approximately 68% of observations fall within one standard deviation of the mean.

Answer: True. This is the 68-95-99.7 rule (empirical rule) for normal distributions.

8. **T / F** If $P(A \mid B) = P(A)$, then events A and B are independent.

Answer: True. This is the definition of independence: knowing B occurred does not change the probability of A.

9. **T / F** Increasing the sample size increases the standard error of the sample mean.

Answer: False. Increasing the sample size *decreases* the standard error, since $SE = \sigma/\sqrt{n}$.

10. **T / F** If the mean of a dataset is greater than the median, the distribution is likely skewed to the right.

Answer: True. Right skew pulls the mean above the median due to extreme values in the right tail.

Part 2: Choose 4 of 5 Questions [80-85 points]

*Answer **4 out of the following 5 questions**. Each question is worth 20 points.*

*Clearly cross out the question you are **not** answering.*

Question 1: Descriptive Statistics [20 points]

A researcher collects annual salary data (in thousands of dollars) for 6 recently hired economics graduates:

$$40, \quad 42, \quad 44, \quad 46, \quad 48, \quad 80$$

a. [5 points] Calculate the sample mean.

> ! Solution
>
> $$\bar{x} = \frac{40 + 42 + 44 + 46 + 48 + 80}{6} = \frac{300}{6} = 50$$

b. [5 points] Calculate the sample median.

> ! Solution
>
> With $n = 6$, the median is the average of the 3rd and 4th values (already sorted):
>
> $$\text{Median} = \frac{44 + 46}{2} = 45$$

c. [5 points] Calculate the sample variance. Show your work.

! Solution

Deviations from $\bar{x} = 50$: -10,-8,-6,-4,-2,$+30$

Squared deviations: 100,64,36,16,4,900

$$s^2 = \frac{100 + 64 + 36 + 16 + 4 + 900}{6 - 1} = \frac{1120}{5} = 224$$

d. [5 points] Based on (a) and (b), what does the shape of this distribution appear to be? Is the mean or the median a more appropriate measure of central tendency here? Explain.

! Solution

Since mean (50) $>$ median (45), the distribution appears **right-skewed**. The salary of 80 is an outlier pulling the mean upward. The **median** is the more appropriate measure of central tendency because it is resistant to outliers, whereas the mean is distorted by the extreme value.

Question 2: Probability and Bayes' Theorem [20 points]

A medical test detects a rare disease. The disease affects **1%** of the population. The test has:

- **Sensitivity**: 95% — if a person has the disease, the test returns positive 95% of the time.
- **False positive rate**: 10% — if a person does *not* have the disease, the test incorrectly returns positive 10% of the time.

For this question, D = 1 is the event that a person has the disease, D=0 is the event that a person does not have the disease, $+$ = the event that the test is positive and $-$ = the event that the test is negative.

a. [2 points] What is the prior probability that a randomly selected person has the disease?

! Solution

$P(D = 1) = 0.01$

b. [2 points] What is the probability of a positive test result given the person *has* the disease? Write your answer as a conditional probability.

! Solution

$P(+ \mid D = 1) = 0.95$

c. [2 points] What is the probability of a positive test result given the person does *not* have the disease? Write your answer as a conditional probability.

! Solution

$P(+ \mid D = 0) = 0.10$

d. [10 points] A person receives a **positive** test result. Using Bayes' Theorem, calculate the probability that this person actually has the disease.

! Solution

Total probability of a positive test:

$$P(+) = P(+ \mid D = 1)\, P(D = 1) + P(+ \mid D = 0)\, P(D = 0)$$

$$= (0.95)(0.01) + (0.10)(0.99) = 0.0095 + 0.099 = 0.1085$$

By Bayes' Theorem:

$$P(D = 1 \mid +) = \frac{P(+ \mid D = 1)\, P(D = 1)}{P(+)} = \frac{0.0095}{0.1085} \approx 0.088$$

e. [4 points] Many people assume a positive result from a 95%-sensitive test means they very likely have the disease. Does your answer to (d) support this? Briefly explain.

! Solution

No. Even after a positive result, the probability of actually having the disease is only about **8.8%**. Because the disease is very rare (1% base rate), most of the positive tests in the population come from the large pool of healthy people who test falsely positive. This is the **base rate fallacy** — people overlook the rarity of the condition when interpreting test results.

Question 3: Discrete Random Variables [20 points]

Part A. A local coffee shop tracks the number of customer complaints per hour. Based on past data:

x	0	1	2	3
$P(x)$	0.50	0.25	0.15	0.10

a. [2 points] Verify that this is a valid probability distribution.

! Solution

All probabilities are between 0 and 1, and $0.50 + 0.25 + 0.15 + 0.10 = 1.00$. ✓

b. [5 points] Calculate the expected number of complaints per hour, $E(X)$.

! Solution

$$E(X) = 0(0.50) + 1(0.25) + 2(0.15) + 3(0.10) = 0 + 0.25 + 0.30 + 0.30 = 0.85$$

c. [5 points] Calculate Var (X).

! Solution

$$E(X^2) = 0(0.50) + 1(0.25) + 4(0.15) + 9(0.10) = 0 + 0.25 + 0.60 + 0.90 = 1.75$$

$$\text{Var } (X) = E(X^2) - [E(X)]^2 = 1.75 - (0.85)^2 = 1.75 - 0.7225 = 1.0275$$

Part B. A small business estimates it has a **50% chance** of being awarded each grant it applies for, independently across applications. It submits **3 applications**.

d. [4 points] Let Y be the number of grants awarded. What distribution does Y follow? Write the probability mass function.

! Solution

$Y \sim \text{Binomial } (n = 3, p = 0.5)$

$$P(Y = y) = \binom{3}{y}(0.5)^3, \quad y = 0, 1, 2, 3$$

e. [4 points] What is the probability the business receives **at least 2 grants**?

! Solution

$$P(Y \geq 2) = P(Y = 2) + P(Y = 3) = \binom{3}{2}(0.5)^3 + \binom{3}{3}(0.5)^3 = 3(0.125) + 1(0.125) = 0.5$$

Question 4: Normal Distribution [20 points]

SAT scores for college applicants are approximately normally distributed with a **mean of 1050** and a **standard deviation of 200**.

(Use the normal distribution values on the equation sheet.)

a. [5 points] What is the probability that a randomly selected applicant scores **less than 850**?

! Solution

$$z = \frac{850 - 1050}{200} = -1 \qquad P(X < 850) = P(Z < -1) = 0.1587$$

b. [5 points] What is the probability that a randomly selected applicant scores **more than 1250**?

! Solution

$$z = \frac{1250 - 1050}{200} = 1 \qquad P(X > 1250) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587$$

c. [5 points] What is the probability that a randomly selected applicant scores **between 850 and 1250**?

! Solution

$$P(850 < X < 1250) = P(-1 < Z < 1) = 0.8413 - 0.1587 = 0.6826$$

d. [5 points] What SAT score corresponds to the **90th percentile**?

! Solution

The 90th percentile has $z = 1.28$.

$$x = \mu + z\sigma = 1050 + 1.28(200) = 1050 + 256 = 1306$$

Question 5: Conditional Probability and Updating Beliefs [25 points] (Not a typo)

An economist models the relationship between **education** (E), **employment** (J), and **high income** (Y) as a chain:

$$E \to J \to Y$$

That is, education affects whether someone is employed, and employment affects whether someone earns a high income. Income does not depend directly on education — only through employment.

You are given the following probabilities:

- $P(E = 1) = 0.4$ (probability of having a college degree)
- $P(J = 1 \mid E = 1) = 0.8$ and $P(J = 1 \mid E = 0) = 0.5$
- $P(Y = 1 \mid J = 1) = 0.7$ and $P(Y = 1 \mid J = 0) = 0.2$

a. [5 points] Using the structure of the network, what can we decompose the joint probability distribution $P(E, J, Y)$ into?

! Solution

The joint distribution can be decomposed as:

$$P(E, J, Y) = P(E) \cdot P(J \mid E) \cdot P(Y \mid J)$$

This follows from the directed acyclic graph structure, where E is a parent of J, and J is a parent of Y.

b. [5 points] Using the structure of the network, intuitively explain why we can write:

> **ℹ Tip**
>
> If you get stuck on this question, skip to the next questions and come back to this one later.

$$P(Y = 1 \mid E = 1) = P(J = 1 \mid E = 1)\, P(Y = 1 \mid J = 1) + P(J = 0 \mid E = 1)\, P(Y = 1 \mid J = 0)$$

> **! Solution**
>
> In the network $E \to J \to Y$, income Y depends only directly on employment J — there is no arrow from E directly to Y. So conditional on J, Y is independent of E. To find $P(Y = 1 \mid E = 1)$ we must average over the two possible values of J, weighted by how likely each is given $E = 1$. This is the law of total probability applied through the intermediate node.

b. [6 points] Compute $P(Y = 1)$, the overall probability that a randomly selected person earns a high income. Show all steps.

! Solution

First compute $P(Y = 1 \mid E = 1)$ and $P(Y = 1 \mid E = 0)$:

$$P(Y = 1 \mid E = 1) = (0.8)(0.7) + (0.2)(0.2) = 0.56 + 0.04 = 0.60$$

$$P(Y = 1 \mid E = 0) = (0.5)(0.7) + (0.5)(0.2) = 0.35 + 0.10 = 0.45$$

Then apply the law of total probability over E:

$$P(Y = 1) = P(E = 1)\,P(Y = 1 \mid E = 1) + P(E = 0)\,P(Y = 1 \mid E = 0)$$

$$= (0.4)(0.60) + (0.6)(0.45) = 0.24 + 0.27 = 0.51$$

c. [5 points] You observe that a randomly selected person earns a **high income**. Using Bayes' Theorem, compute the probability that they have a college degree, $P(E = 1 \mid Y = 1)$.

! Solution

$$P(E = 1 \mid Y = 1) = \frac{P(Y = 1 \mid E = 1)\,P(E = 1)}{P(Y = 1)} = \frac{(0.60)(0.4)}{0.51} = \frac{0.24}{0.51} \approx 0.47$$

d. [4 points] Compared to the prior probability $P(E = 1) = 0.4$, did observing high income increase or decrease your belief that the person has a college degree? Give an intuitive explanation based on the structure of the network.

! Solution

Observing high income **increased** the probability of having a college degree, from 0.40 to 0.47. Intuitively, a college degree raises the chance of being employed, and being employed raises the chance of high income. So learning that someone earns a high income makes it more likely they were employed, which in turn makes it more likely they had a college degree. The signal flows backwards through the chain.

Equation Sheet

Descriptive Statistics

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Expected Value and Variance

$$E(X) = \sum_i x_i\, P(x_i) \qquad \text{Var }(X) = E(X^2) - [E(X)]^2$$

Binomial Distribution

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad E(X) = np \qquad \text{Var }(X) = np(1-p)$$

Combinations

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Standard Normal Distribution Values

$$P(Z \leq -1) = 0.1587, \quad P(Z \leq -2) = 0.0228, \quad P(Z \leq -3) = 0.0013$$

$$P(Z \leq 1.28) = 0.90$$