EC 282 — Midterm 1 Mock Exam

Introduction to Econometrics

Spring 2026

Professor Altindag

Instructions:

- You have 80 minutes to complete this exam.
- The exam consists of 15 multiple choice questions and 3 short answer problems.
- You may round all calculations to 2 decimal places.
- Show your work on short answer problems for partial credit.
- You may use a calculator. A formula sheet is provided at the end.

NAME: ____

Part I: Multiple Choice (3 points each, 45 points total)

Circle the best answer for each question.

1. A **random variable** is best described as:
   (a) A variable whose value is always unknown
   (b) A numerical summary of a random outcome
   (c) A variable that follows a normal distribution
   (d) A sample statistic computed from data

2. Let Y be a Bernoulli random variable with $\Pr(Y = 1) = 0.3$. What is $\text{Var}(Y)$?
   (a) 0.30
   (b) 0.09
   (c) 0.21
   (d) 0.49

3. Two random variables X and Y are **independent** if and only if:
   (a) $\text{Cov}(X, Y) = 0$
   (b) $\Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$ for all x, y
   (c) $E[X] = E[Y]$
   (d) $\text{Corr}(X, Y) = 1$

4. If $\text{Cov}(X, Y) = 0$, which of the following is true?

(a) X and Y are independent
(b) X and Y have no linear relationship, but may have a nonlinear one
(c) X and Y have no relationship of any kind
(d) $\text{Corr}(X, Y) = 1$

5. The **correlation** between X and Y is defined as $\rho_{XY} = \sigma_{XY}/(\sigma_X \sigma_Y)$. Which of the following is **NOT** a property of correlation?

   (a) $-1 \leq \rho_{XY} \leq 1$
   (b) Correlation is unitless
   (c) $\rho_{XY} = 0$ implies X and Y are independent
   (d) $\rho_{XY} = 1$ implies a perfect positive linear relationship

6. If $Y_1, Y_2, \ldots, Y_n$ are i.i.d. with $E[Y_i] = \mu_Y$ and $\text{Var}(Y_i) = \sigma_Y^2$, then $\text{Var}(\bar{Y})$ equals:

   (a) σ_Y^2
   (b) σ_Y^2/n^2
   (c) σ_Y^2/n
   (d) $\sigma_Y/\sqrt{n}$

7. The **Central Limit Theorem** states that:

   (a) The population distribution is approximately normal for large n
   (b) The sample mean $\bar{Y}$ is always exactly normally distributed
   (c) For large n, $\bar{Y}$ is approximately normally distributed regardless of the population distribution
   (d) The sample variance converges to σ_Y^2 as $n \to \infty$

8. The **Law of Large Numbers** tells us that:

   (a) Large samples always have larger variance
   (b) $\bar{Y} \xrightarrow{p} \mu_Y$ as $n \to \infty$
   (c) The sample mean equals the population mean when $n > 30$
   (d) The sampling distribution of $\bar{Y}$ is normal for large n

9. An estimator $\hat{\mu}_Y$ is **unbiased** if:

   (a) $\hat{\mu}_Y = \mu_Y$ for every sample
   (b) $E[\hat{\mu}_Y] = \mu_Y$
   (c) $\text{Var}(\hat{\mu}_Y) = 0$
   (d) $\hat{\mu}_Y \xrightarrow{p} \mu_Y$

10. The sample variance formula uses $n-1$ in the denominator (instead of n) because:
    (a) It makes the variance smaller
    (b) It produces an unbiased estimator of the population variance (Bessel's correction)
    (c) It is required by the Central Limit Theorem
    (d) It ensures the sample variance is always positive

11. The **conditional expected value** $E[Y \mid X = x]$ is:
    (a) The expected value of Y ignoring X
    (b) The expected value of Y calculated using the conditional distribution of Y given $X = x$
    (c) Always equal to $E[Y]$
    (d) Only defined when X and Y are independent

12. Suppose X and Y are independent. Which of the following must be true?
    (a) $E[X] = E[Y]$
    (b) $\text{Cov}(X, Y) = 0$
    (c) $\text{Var}(X) = \text{Var}(Y)$
    (d) $\Pr(X = x) = \Pr(Y = y)$ for all x, y

13. A population has mean $\mu_Y = 200$ and variance $\sigma_Y^2 = 400$. If you draw a random sample of $n = 100$, then by the CLT the sampling distribution of $\bar{Y}$ is approximately:
    (a) $N(200, 400)$
    (b) $N(200, 4)$
    (c) $N(200, 2)$
    (d) $N(200, 20)$

14. Among two unbiased estimators of μ_Y, the one with **smaller variance** is said to be:
    (a) More consistent
    (b) More efficient
    (c) Less biased
    (d) More robust

15. A researcher surveys students only from their own lecture section to estimate the average GPA of all students at the university. This is an example of:
    (a) Random sampling
    (b) The Central Limit Theorem in action
    (c) Selection bias (non-random sampling)
    (d) An efficient estimator

Part II: Short Answer Problems

Problem 1. Joint Distribution, Conditional Probability, and Covariance (25 points)

A university surveys students about their **housing type** (X) and **GPA category** (Y). Define:

- $X \in \{0, 1\}$ where $X = 0$ is off-campus and $X = 1$ is on-campus
- $Y \in \{0, 1\}$ where $Y = 0$ is GPA below 3.0 and $Y = 1$ is GPA 3.0 or above

The **joint distribution** is given below:

	$Y = 0$ (GPA < 3.0)	$Y = 1$ (GPA ≥ 3.0)	Marginal of X
$X = 0$ (Off-campus)	0.20	0.25	
$X = 1$ (On-campus)	0.15	0.40	
Marginal of Y			1.00

(a) (5 points) Compute the marginal distributions of X and Y. Fill in the missing values in the table.

(b) (5 points) Calculate $E[X]$, $E[Y]$, $\text{Var}(X)$, and $\text{Var}(Y)$.

(c) (5 points) Calculate the following conditional probabilities:

- $\Pr(Y = 1 \mid X = 1)$ — the probability of a high GPA given on-campus housing.
- $\Pr(Y = 1 \mid X = 0)$ — the probability of a high GPA given off-campus housing.

(d) (5 points) Based on your answers, are X and Y independent? Justify your answer using the definition of independence.

(e) (5 points) Calculate $\mathrm{Cov}(X, Y)$ and $\mathrm{Corr}(X, Y)$. Interpret the sign of the correlation.

Problem 2. Sampling Distribution and Estimation (15 points)

A large population of monthly apartment rents has mean $\mu_Y = 950$ and standard deviation $\sigma_Y = 300$. A researcher draws a random sample of $n = 225$ apartments.

(a) (3 points) What is the expected value of the sample mean $\bar{Y}$? Is the sample mean an unbiased estimator of μ_Y? Explain.

(b) (3 points) Calculate the standard deviation of $\bar{Y}$ (i.e., the standard error $\sigma_{\bar{Y}}$).

(c) (3 points) Using the Central Limit Theorem, write down the approximate sampling distribution of $\bar{Y}$.

(d) (3 points) Calculate $\Pr(\bar{Y} > 960)$. Show your work by standardizing.

Hint: $\Phi(0.50) = 0.691$.

(e) (3 points) If the researcher instead drew a sample of $n = 900$, would $\Pr(\bar{Y} > 960)$ be larger or smaller? Explain without doing the full calculation.

Problem 3. Conditional Expectation and the Law of Iterated Expectations (15 points)

An insurance company classifies drivers into two risk groups based on age:

- $X = 0$: drivers aged 25 and older (70% of all drivers, i.e., $\Pr(X = 0) = 0.70$)
- $X = 1$: drivers under 25 (30% of all drivers, i.e., $\Pr(X = 1) = 0.30$)

Let Y denote the number of accidents per year. The conditional distributions are:

	$Y = 0$	$Y = 1$	$Y = 2$
$\Pr(Y = y \mid X = 0)$ (age ≥ 25)	0.80	0.15	0.05
$\Pr(Y = y \mid X = 1)$ (age < 25)	0.50	0.35	0.15

(a) (4 points) Calculate $E[Y \mid X = 0]$ and $E[Y \mid X = 1]$. Interpret these values in plain language.

(b) (4 points) Use the **Law of Iterated Expectations** to compute $E[Y]$, the overall expected number of accidents per year:

$$E[Y] = E[Y \mid X = 0] \cdot \Pr(X = 0) + E[Y \mid X = 1] \cdot \Pr(X = 1)$$

(c) (3 points) Calculate $\mathrm{Var}(Y \mid X = 1)$.

(d) (4 points) Can we conclude from this data that being under 25 *causes* more accidents? What other factors might explain the difference in $E[Y \mid X = 0]$ and $E[Y \mid X = 1]$?

Formula Sheet

Probability & Random Variables

Expected Value:

$$E[Y] = \mu_Y = \sum_i y_i \cdot \Pr(Y = y_i)$$

Variance and Standard Deviation:

$$\mathrm{Var}(Y) = \sigma_Y^2 = E[(Y - \mu_Y)^2] = \sum_i (y_i - \mu_Y)^2 \cdot \Pr(Y = y_i) \qquad \sigma_Y = \sqrt{\sigma_Y^2}$$

Bernoulli Random Variable ($Y \in \{0, 1\}$, $\Pr(Y = 1) = p$):

$$E[Y] = p \qquad \mathrm{Var}(Y) = p(1 - p)$$

Conditional Probability:

$$\Pr(Y = y \mid X = x) = \frac{\Pr(X = x, Y = y)}{\Pr(X = x)}$$

Conditional Expected Value:

$$E[Y \mid X = x] = \sum_i y_i \cdot \Pr(Y = y_i \mid X = x)$$

Law of Iterated Expectations:

$$E[Y] = E\big[E[Y \mid X]\big] = \sum_i E[Y \mid X = x_i] \cdot \Pr(X = x_i)$$

Covariance and Correlation:

$$\mathrm{Cov}(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] \qquad \mathrm{Corr}(X, Y) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Independence: X and Y are independent if $\Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$ for all x, y.

If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$. The converse is **not** true.

Sampling & Estimation

Sample Mean:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \qquad E[\bar{Y}] = \mu_Y \qquad \mathrm{Var}(\bar{Y}) = \frac{\sigma_Y^2}{n} \qquad \sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$$

Sample Variance and Standard Error:

$$S_Y^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \qquad SE[\bar{Y}] = \frac{S_Y}{\sqrt{n}}$$

Central Limit Theorem: For large n, $\bar{Y} \overset{a}{\sim} N\left(\mu_Y, \, \frac{\sigma_Y^2}{n}\right)$

Law of Large Numbers: $\bar{Y} \xrightarrow{p} \mu_Y$ as $n \to \infty$

Standardization: If $Y \sim N(\mu_Y, \sigma_Y^2)$, then $Z = \dfrac{Y - \mu_Y}{\sigma_Y} \sim N(0, 1)$

Estimator Properties

- **Unbiased:** $E[\hat{\mu}_Y] = \mu_Y$
- **Consistent:** $\hat{\mu}_Y \xrightarrow{p} \mu_Y$ as $n \to \infty$
- **Efficient:** Smallest variance among all unbiased estimators
- **BLUE:** The sample mean is the Best Linear Unbiased Estimator of μ_Y