Midterm Exam

EC 320 - Introduction to Econometrics

Summer 2025

Give every question your best attempt. Best of luck. You got this!

Name: ANSWER KEY 95#: ____________

The maximum amount of points on this exam is 100 points. You have a total of 12h to complete the exam, unless otherwise noted. Once you complete it, upload your answers to Canvas. **Before submitting, be sure that your answers are clear and legible.**
You are allowed full access to your notes, textbook, and class lectures. Any form of cheating will result on a zero on the exam.

There are three sections to be completed:

- **Multiple Choice:** 13 Questions
- **Short Answer Questions:** 3 Questions
- **Multi-Part Analysis Questions:** 1 Question (5 parts)

Point totals and question specific instructions are listed for each section. Please ask for clarification if a question is not clear to you.

The exam is a total of 7 pages. Do not feel limited to the space below each question.

Multiple Choice - 53 Points

Circle or "X" the answer you think most correctly answers the following questions. If you mark a choice and would like to change it, **clearly indicate which one is your correct answer**.

1. [5 points] Consider this model:

$$wage_i = \beta_0 + \beta_1 education_i + \beta_2 sex_i + u_i$$

Where $education_i$ is the number of years of education someone has and sex_i is 0 for female and 1 for male. Then the expected wage for females with 10 years of education is:

A. β_0

B. $\beta_0 + \beta_2 + 10\beta_1$

C. $\beta_0 + \beta_2$

D. $\beta_0 + 10\beta_1$

2. [5 points] Suppose that the true y_i is generated via:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + u_i$$

If z_i is omitted, will $\hat{\beta}_1$ always be biased?

A. No, $\hat{\beta}_1$ will only be biased if z_i also correlates with x_i

B. No, $\hat{\beta}_1$ will only be biased if z_i also correlates with u_i

C. No, $\hat{\beta}_1$ will never be biased if z_i is omitted because β_1 is the coefficients on x_i, not on z_i

D. Yes, $\hat{\beta}_1$ will always be biased because, as long as β_2 is nonzero, z_i has an important effect on y_i that must be accounted for.

3. [5 points] Simplify: $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

A. $n^2\bar{x}\bar{y} - n\bar{x}\bar{y}$

B. $\sum_i x_i y_i$

C. 0

D. $\sum_i x_i y_i - \bar{x}\bar{y}n$

4. [5 points] In a simple regression, the standard error for $\hat{\beta}_1$ depends ____________ on the sample variance of the explanatory variable x.

   A. Positively

   B. Negatively

   C. Not at all

5. [5 points] Under standard OLS assumptions, the key assumption for $\hat{\beta}_1$ to be unbiased is ____________ and the key assumption for $\hat{\beta}_1$ to be consistent is ____________:

   A. $\mathbb{E}[u_i|x_i] = 0$ & $\mathbb{E}[u_i] = 0$

   B. $\mathbb{E}[u_i|x_i] = 0$ & $Cov(x_i, u_i) = 0$

   C. x and u are independent & $Cov(x_i, u_i) = 0$

   D. x and u are independent & $\mathbb{E}[u_i] = 0$

6. [5 points] If our significance level is 0.01 and our p-value is 0.02, then we ____________ the null hypothesis

   A. Reject

   B. Fail to Reject

7. [5 points] $Cov(X, 5X - Y)$ is equal to:

   A. $25Var(X) + Cov(X, Y)$

   B. $25Var(X) + Var(Y)$

   C. $5Var(X) - Cov(X, Y)$

   D. $5Var(X) - Var(Y)$

8. [5 points] Which of the following describes homoskedasticity:

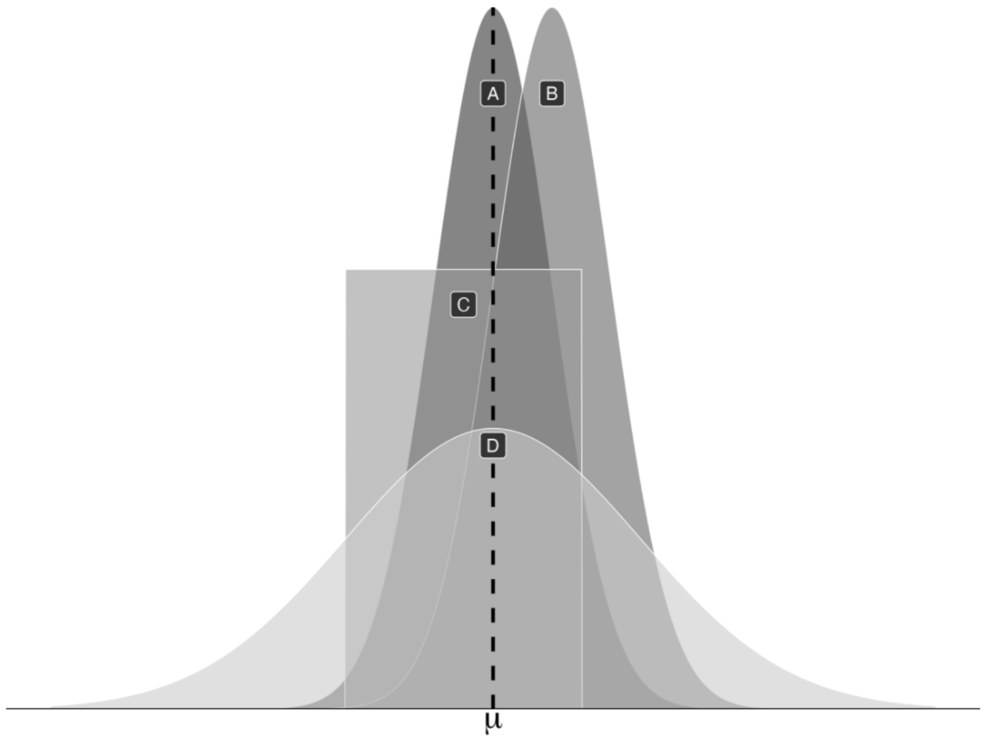
   A. $Var(u_i|x) = \sigma^2$

   B. $Var(u_i|x) = \sigma_i^2$

   C. $Var(u_i|x) = \sigma^2 X$

   D. $Var(u_i|x) = \beta$

9. [5 points] In the figure below, each distribution describes an estimator. Which one is the best unbiased estimator.



A. A

B. B

C. C

D. D

True or False

Select True [T] or False [F] for each statement below:

10. [2 points] ☒ T ☐ F A continuous random random variable is a random variable that takes any real value with *zero probability*

11. [2 points] ☐ T ☒ F $\mathbb{E}[X]^2$ is equivalent to $\mathbb{E}[X^2]$

12. [2 points] ☐ T ☒ F It is typical and important, of regression estimates with a causal interpretation to have a high R^2

13. [2 points] ☐ T ☒ F Greater variation in x_i increases the variance of our OLS slope parameter

Short Answer - 25 Points

Provide a response to each question below. Show all work and clearly mark your answers.

14. [10 points] Consider a linear model without an intercept, that is: $y_i = \beta_1 x_i + u_i$. Recall that OLS **minimizes the sum of squared residuals**. Write down the minimizing problem in this context and take the first order conditions. Show that $\hat{\beta}_1 = \dfrac{\sum_i x_i y_i}{\sum_i x_i^2}$

$$\min_{\hat{\beta}_1} \sum_i u_i^2 \quad \text{where} \quad u_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 x_i$$

$$\min_{\hat{\beta}_1} \sum_i (y_i - \hat{\beta}_1 x_i)(y_i - \hat{\beta}_1 x_i)$$

$$\min_{\hat{\beta}_1} \sum_i (y_i^2 - 2\hat{\beta}_1 y_i x_i + \hat{\beta}_1^2 x_i^2)$$

FOC:

$$\frac{\partial}{\partial \hat{\beta}_1}\left(\sum_i (y_i^2 - 2\hat{\beta}_1 y_i x_i + \hat{\beta}_1^2 x_i^2)\right) = 0$$

$$\Rightarrow \sum_i (-\cancel{2} y_i x_i + \cancel{2}\hat{\beta}_1 x_i^2) = 0$$

$$\Rightarrow \hat{\beta}_1 \sum_i x_i^2 = \sum_i y_i x_i$$

$$\Rightarrow \boxed{\hat{\beta}_1 = \frac{\sum_i y_i x_i}{\sum_i x_i^2}}$$

15. [5 points] Describe in words what the objective of the OLS estimator is and how the first order condition reaches that objective.

An OLS estimator seeks to minimize the errors or "misses" from the "guess" of the estimand. The FOC does this because it gives the value of the estimator ($\hat{\beta}_1$) in this case that minimizes the squared errors.

16. Suppose we have three random variables Y_i, X_i, and u_i for which the data generating process is known. u_i is distributed normally with mean 0 and standard deviation of 1. It is also statistically independent. X_i is normally distributed with mean of 5 and standard deviation of 2. And $Y_i = 12 + 5X_i + u_i$.

   (a) [3 points] Find the expected value of Y_i

   Use $E[A+B] = E[A] + E[B]$

   $\hookrightarrow E[Y_i] = E[12] + E[5X_i] + E[u_i]$

   $E[Y_i] = 12 + 5E[X_i] + 0$

   $E[Y_i] = 12 + 5 \cdot 5 = 12 + 25 = 37$

   (b) [3 points] Find the variance of Y_i

   $Var(y_i) = Var(12) + Var(5X_i) + Var(u_i)$

   $= 0 + 5^2 Var(X_i) + Var(u_i)$ $\longrightarrow$ Recall $Var(x) = (\text{Std. Dev})^2$

   $= 25 \cdot (2)^2 + (1)^2$

   $Var(y_i) = 25 \cdot 4 + 1 = 101$

   (c) [4 points] Suppose we collect a 1,000 observation sample of these variables. If we regress Y_i on X_i and find the OLS estimates $\hat{Y}_i = \beta_0 + \beta_1 \hat{X}_i$, if OLS is unbiased, what should $\hat{\beta}_1$ equal to? Explain.

It should equal 5. Because an unbiased OLS estimate gives $E[\hat{\beta}] = \beta_1$ which equals 5.

Multi-Part Question - 22 Points

17. **Intepreting OLS Estimates (5 parts)**

Suppose we want to model **student test scores** based on their **hours studied** and **class attendance rate**. It is specified in the following model:

$$Score_i = \beta_0 + \beta_1 HoursStudied_i + \beta_2 Attendance_i + u_i$$

- $Score_i$: Test score of student i
- $HoursStudied_i$: Hours studied per week
- $Attendance_i$: Percentage of class attended

The results of the regression are reported below:

Dependent Variable:	Test Score
Variables	
Intercept	45.7^{***}
	(4.2)
Hours Studied	1.8^{***}
	(0.45)
Attendance	0.40^{***}
	(0.10)
Fit Statistics	
Observations	150
R^2	0.583

Standard errors in parenthesis
Significance Codes: $^{***}: 0.01$, $^{**}: 0.05$, $^{*}: 0.10$

(a) [6 points] Find the *t-statistics* for each coefficient in the table.

T-Stat can be found by $\frac{\text{coefficient}}{\text{std. error}}$

$\boxed{\text{Intercept : } \frac{45.7}{4.2} = 10.88}$

$\boxed{\text{Hours Studied : } \frac{1.8}{0.45} = 4}$

$\boxed{\text{Attendance : } \frac{0.40}{0.10} = 4}$

(b) [4 points] Interpret the intercept coefficient β_0 in the table above

A student that does not study at all & attends 0% of classes is expected to score 45.7 on a test.

(c) [5 points] Interpret the Hours Studied coefficient. How is Attendance factored in this?

An additional hour of studying per week increases the expected test score by 1.8 points.

Attendance is held constant.

(d) [5 points] Interpret the Attendance coefficient. How are Hours Studied factor in this?

An additional percentage point in attendance increases the expected test score by 0.40 points.

Hours studied is held constant.

(e) [2 points] Intepret the R^2 given in the table above

The model explains 58% of the variation in test scores.