University of Oregon
Department of Economics
EC320 – Midterm

Instructor: Emmett Saulnier

April 28th, 2022

Name: ____
UO ID: ____

This exam contains 8 pages (including this cover page) and 15 questions. Total of points is 100.

Question	Points	Score
1	5	
2	5	
3	5	
4	5	
5	5	
6	5	
7	5	
8	5	
9	5	
10	5	
11	10	
12	10	
13	10	
14	10	
15	10	
Total:	100	

Multiple Choice

Each question has a single answer and is worth 5 points. Clearly mark your response. No explanation required.

1. (5 points) Suppose we have a sample mean $\bar{x} = 1.75$ with standard error $SE(\bar{x}) = 1$. We are conducting a one-sided hypothesis test at the 5% significance level with the null hypothesis $H_0 : \mu = 0$ and alternative $H_a : \mu > 0$. What is the conclusion of the test? *Hint: Some z-statistics are $z_{0.025} = 1.96$, $z_{0.05} = 1.64$, and $z_{0.10} = 1.28$, where z_α gives the value from the standard normal distribution where α proportion of the probability density function is above z_α.*
   (a) Reject the alternative that $\mu > 0$.
   (b) Reject the null hypothesis that $\mu = 0$ in favor of the alternative that $\mu > 0$
   (c) We cannot come to a conclusion with the given information.
   (d) Fail to reject the null hypothesis that $\mu = 0$.
   (e) None of the above

2. (5 points) Suppose we have data from a survey collected by the University of Oregon, where the same $1,000$ students respond to the questionnaire about their living situation each year for the four years they attend the university. What type of data is this?
   (a) Pooled cross sectional data.
   (b) Panel data
   (c) Experimental data
   (d) Cross sectional data
   (e) Time series data

3. (5 points) Which of the following is **not** a reason for the existence of an error term in a regression model?
   (a) Model misspecification
   (b) Aggregation of variables
   (c) Measurement error
   (d) Sampling variation
   (e) Omitted explanatory variables

4. (5 points) Suppose you have OLS estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ from the regression model $y_i = \beta_0 + \beta_1 x_i + u_i$. The OLS estimates are obtained by
   (a) minimizing the sum of the absolute value of the residuals.
   (b) maximizing the sum of the residuals.
   (c) maximizing the sum of the squared residuals.
   (d) minimizing the sum of the residuals.
   (e) minimizing the sum of the squared residuals.

5. (5 points) Which of the following describes the best estimator for a population parameter
   (a) The estimator is biased and has the lowest variance.
   (b) The estimator is unbiased and has the lowest variance.
   (c) The estimator is biased and has the highest variance.
   (d) The estimator is unbiased and has the highest variance.
   (e) None of the above.

True or False

Circle either True or False, no explanation required. Each question is worth 5 points.

6. (5 points) **True or False?** The population mean does not change based on how we choose a sample.

7. (5 points) **True or False?** We can observe both the treated and untreated potential outcomes for a given individual.

8. (5 points) **True or False?** We can solve the Omitted Variable Bias problem by adding all of the variables we have in our data as controls in the regression.

9. (5 points) **True or False?** We don't need any variation in X to calculate OLS estimates.

10. (5 points) **True or False?** If we run a regression where $R^2 = 1$, then the regression model has explained all of the variation in the outcome variable.

Short Answer

Provide a written response to each question, where it is clearly marked which question you are responding to. Use the back of the page if you need additional space. Show all of your work.

11. Suppose we are interested in measuring the treatment effect τ. We have that $y_{0,i}$ is the potential outcome for individual i if they did not receive treatment, $y_{1,i}$ is the potential outcome for individual i if they did receive treatment, and D_i is an indicator for whether individual i was actually treated or not. Additionally, suppose the treatment effect is constant for all individuals, that is $y_{1,i} = \tau + y_{0,i}$.

    (a) (4 points) Derive an expression that shows difference in means for the treated and untreated groups equals the treatment effect plus selection bias.

    (b) (2 points) Suppose we have the following *ideal* data.

i	D_i	$y_{0,i}$	$y_{1,i}$
1	0	95	109
2	0	96	110
3	0	101	115
4	1	81	95
5	1	86	100
6	1	75	89

    Calculate the average treatment effect (ATE).

    (c) (2 points) Calculate the difference-in-means between the treated and untreated groups for the actual outcome of each group.

    (d) (2 points) Use the equation you derived in (a) to explain why your answers from (b) and (c) either differ or are the same.

12. Suppose we have three random variables y_i, x_i, and u_i for which we know the data generating process. u_i is a normal random variable with a mean of 0 and standard deviation of 1. x_i is a normal random variable with a mean of 5 and standard deviation of 2. Finally, $y_i = 6 + 5x_i + u_i$.
    (a) (2 points) Calculate the expected value of y_i.
    (b) (2 points) Calculate the variance of y_i.
    (c) (3 points) Suppose we collected a $1,000$ observation sample of these variables, then regressed y_i on x_i, and got OLS estimates $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Will $\hat{\beta}_1 = 5$? Briefly explain.
    (d) (3 points) Will $E[\hat{\beta}_1] = 5$? Briefly explain.

13. Suppose we estimate the following model $wage_i = \beta_0 + \beta_1 educ_i + u_i$, where $wage_i$ is annual income in thousands of dollars and $educ_i$ is years of education.

	Dependent variable:
	wage
(Intercept)	14.695*
	(7.772)
educ	6.021***
	(0.5695)
Observations	935
R^2	0.107
Note:	*p<0.1; **p<0.05; ***p<0.01

(a) (2 points) What is the estimate for β_0 and what is its interpretation?

(b) (4 points) What is the estimate for β_1 and what is its interpretation?

(c) (4 points) What are two reasons why the estimate for β_1 may not measure the causal effect of education on earnings?

14. Recall that $TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$, $ESS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$, and $RSS = \sum_{i=1}^{n} \hat{u}_i^2$.

    (a) (3 points) Describe intuitively what TSS, ESS, and RSS measure.

    (b) (5 points) Show that $TSS = ESS + RSS$. Be sure to show your work. *Hint: Remember that you showed* $\sum_{i=1}^{n} \hat{u}_i = 0$ *and* $\sum_{i=1}^{n} x_i \hat{u}_i = 0$ *as a part of your homework.*

    (c) (2 points) Describe how TSS, ESS, and RSS relate to R^2 by writing out the formula we use to calculate R^2.

15. Suppose we have the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + u_i$.
    (a) (1 point) What does the error term represent?
    (b) (3 points) Which of the six assumptions are required for the OLS estimate $\hat{\beta}_1$ to be an unbiased estimate for β_1?
    (c) (3 points) Give the definitions of each of those assumptions.
    (d) (3 points) Give an example of a violation of each of those assumptions.