

# Econometrics Midterm Exam

ECON 480 – Safner  
Fall 2018

Name: \_\_\_\_\_

I pledge that I have neither given nor received unauthorized aid on this exam

Please read all of the following information and wait until authorized to start the exam:

- You have **75 minutes** to complete this exam.
- This test consists of **10** required questions worth **100** points (20% of final course grade), and up to 2 bonus questions, worth up to 10 additional points.
- Please read all instructions and question prompts carefully.
- **Show all work.** You will *not* earn full credit if you only write the answer, even if correct.
- Describing your thought process will give you a better chance to earn partial credit for wrong answers.
- You can ace this test without use of a calculator, but you can use one. **Do not use your phone.**
- Good luck!

1. [5 points] Why is OLS called Ordinary Least Squares? No need to calculate anything, but how are the estimators  $\hat{\beta}_0$  and  $\hat{\beta}_1$  chosen?
2. [5 points] There are two sources of randomness in OLS estimates. What is the difference between sampling randomness and modeled randomness?
3. [5 points] In your own words, explain what the regression  $R^2$  means. Explain verbally two different ways that it can be calculated. What does it mean to have a low  $R^2$ ?

4. [5 points] Explain, in your own words, what a  $p$ -value is, and how it is used to establish statistical significance.

5. [10 points] Both using formulas and your own words, describe what exogeneity and endogeneity mean, and how they are related to bias. What can we learn about the bias?

6. [10 points] Suppose a professor wants to estimate the effect of class attendance on grades. The professor randomly selects students and collects data on the number of classes that student attends (*Attendance*) and that student's final grade (*Grade*). The professor then estimates an OLS model of the form

$$Grade_i = \beta_0 + \beta_1 Attendance_i + \epsilon_i$$

and finds a large positive value for  $\hat{\beta}_1$ . Is this an unbiased estimate of the impact of attendance on grades? Why or why not? Do you expect the estimate to overstate or understate the true relationship between  $Grade_i$  and  $Attendance_i$ ?

7. [10 points] A discrete random variable  $X$  has the following pdf:

$x_i$	$P(x_i)$
10	0.1
20	0.2
30	0.3
40	0.4

Calculate the standard deviation of  $X$ .

8. [10 points] Suppose exam grades are normally distributed with a mean of 75 and standard deviation of 10. Convert to the standard normal distribution and estimate the following probabilities:
- (a) [3 points] What is the probability of a student's exam grade being between a 65 and an 85?
  - (b) [3 points] What is the probability of a student's exam grade being above an 85?
  - (c) [3 points] What is the probability of a student's exam grade being lower than a 55?
  - (d) [1 point] What is the probability of a student's exam grade being exactly a 75?

9. [20 points] The results of a sample regression of the number of violent crimes (per 1,000 people) on the number of police officers in a city are reported as follows:

(1) Crimes	
Police	-2.00 (4.00)
Constant	45*** (10)
<div style="display: flex; justify-content: space-between;"> <span><math>n</math></span> <span>100</span> </div> <div style="display: flex; justify-content: space-between;"> <span><math>R^2</math></span> <span>0.150</span> </div> <div style="display: flex; justify-content: space-between;"> <span><math>SER</math></span> <span>7</span> </div>	
Standard errors in parentheses *** $p < 0.01$ , ** $p < 0.05$ , * $p < 0.1$	

- (a) [1 point] Write out the regression equation (i.e. using the estimates of the coefficients, and standard errors below them in parentheses).

- (b) [2 points] What does the estimate for  $\hat{\beta}_0$  mean in terms of the regression line? Interpret what the estimate means in context.

- (c) [2 points] What does the estimate for  $\hat{\beta}_1$  mean in terms of the regression line? Interpret what the estimate means in context.

(d) [2 points] What does the SER mean in terms of the regression line? What does it mean in context?

(e) [2 points] Suppose a small town has 10 police officers. What is the predicted number of violent crimes?

(f) [2 points] Suppose that same small town is actually in our data and has had 20 crimes occur. What is the residual for this town? Is this a reasonably good prediction for the current model (and how do you know)?

(g) [2 points] Write down the null and alternate hypotheses for testing whether Police has *any* effect on Crime.

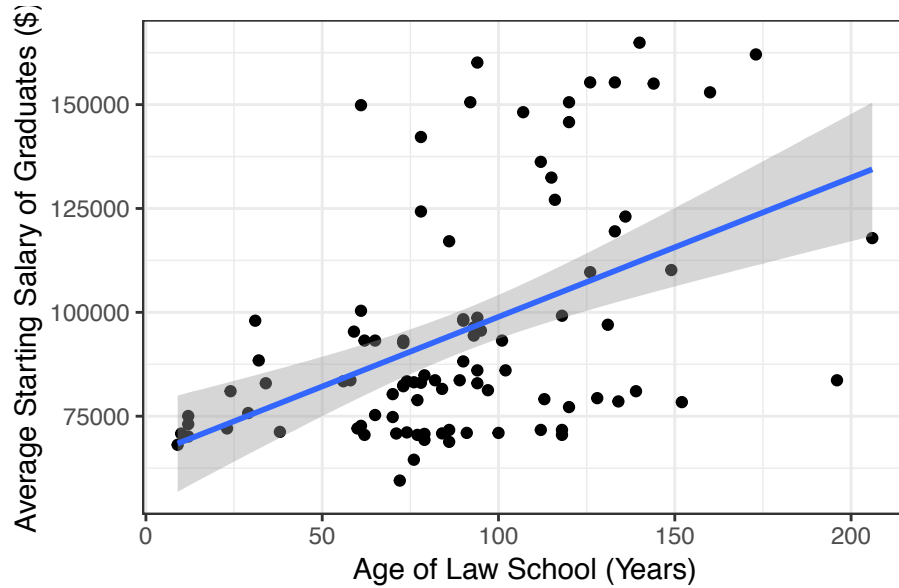
(h) [3 points] Calculate the test-statistic.

(i) [2 points] If the  $p$ -value is 0.618, what can we conclude at the  $\alpha = 0.05$  level?

(j) [2 points] Calculate a 95% confidence interval for  $\beta_1$ . Hint: the critical value is 1.96.



10. [20 points] U.S. News and World Report ranks law schools. Suppose their researchers want to estimate the effect of how old a law school is (in years) on the average starting salary of graduates class in their first legal job (in \$1,000s). They randomly select 95 law schools and create the following scatterplot and graph of their OLS regression line:



- (a) [2 points] Are the errors likely to be homoskedastic or heteroskedastic? Explain your answer both in terms of the graph, and the economic intuition why the errors might be this way.

(b) [4 points] The output for the regression from R is below:

Call:

```
lm(formula = salary ~ age, data = lawschool)
```

Residuals:

Min	1Q	Median	3Q	Max
-47433	-16117	-5761	8550	64019

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	65405.75	6353.98	10.294	< 2e-16 ***
age	335.09	65.51	5.115	1.68e-06 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24890 on 93 degrees of freedom

Multiple R-squared: 0.2196, Adjusted R-squared: 0.2112

F-statistic: 26.17 on 1 and 93 DF, p-value: 1.675e-06

Write an equation for the estimated regression line, and write the standard errors of the estimates beneath them in parentheses.

(c) [2 points] Interpret the  $R^2$  of this regression.

- (d) [2 points] Interpret the standard error of the regression in this context.
- (e) [2 points] For a law school that is relatively new, 10 years old, what is the average graduate's predicted starting salary?
- (f) [3 points] Suppose that this new law school is in our data, and has an *actual* average starting salary of \$90K. Calculate the residual for this law school. Is this a reasonably good prediction for the current model (and how do you know)?
- (g) [5 points] Is Age exogenous or endogenous? Would we have reasons to believe that we have overestimated or underestimated the effect of age on salary?