

Economics 345
Mid-Term Exam
October 8, 2003

Name: _____

Directions: You have the full period (7:20-10:00) to do this exam, though I suspect it won't take that long for most students. You may consult any materials, including textbooks and class notes. Please attempt to write clearly, and show your work for mathematical calculations (use back of page if necessary). Write your answers directly on the test. The point value for each question is indicated.

1. Use the data from the following table to answer the questions that follow: (10 points)

OBSERVATION	Sex	Weight (in pounds)
1	M	150
2	F	120
3	F	130
4	F	130
5	M	180
6	M	200
7	F	150
8	M	170
9	M	190
10	F	160
11	F	90
12	M	210
13	F	120
14	F	130
15	F	130
16	M	140
17	M	250

- 1.a What is the mean of the weight data?
- 1.b What is the median of the weight data?
- 1.c What is the standard deviation of the weight data?
- 1.d Is the variance of the male weights larger than the variance of the female weights (Show calculation for each)?
- 1.e If you perform an OLS regression of $Weight = \hat{a}$ (i.e., just estimate a constant) what will your estimate of the constant be?
- 1.f If you perform an OLS regression of $Weight_i = \hat{b} \cdot Male_i + \hat{c} \cdot Female_i$ where male is a dummy variable taking the value of 1 if the individual is male (0 otherwise) and Female is an indicator taking the value of 1 if the individual is female (0 otherwise), what will your estimates of \hat{a} & \hat{b} be (derive them)?

2. When we call an estimator unbiased, what does that mean (the definition in the book's glossary is not sufficient; I want some intuition as well)? (2 points)

3. If an estimator is biased, can we solve the problem by collecting more data? Why or why not? (2 points)

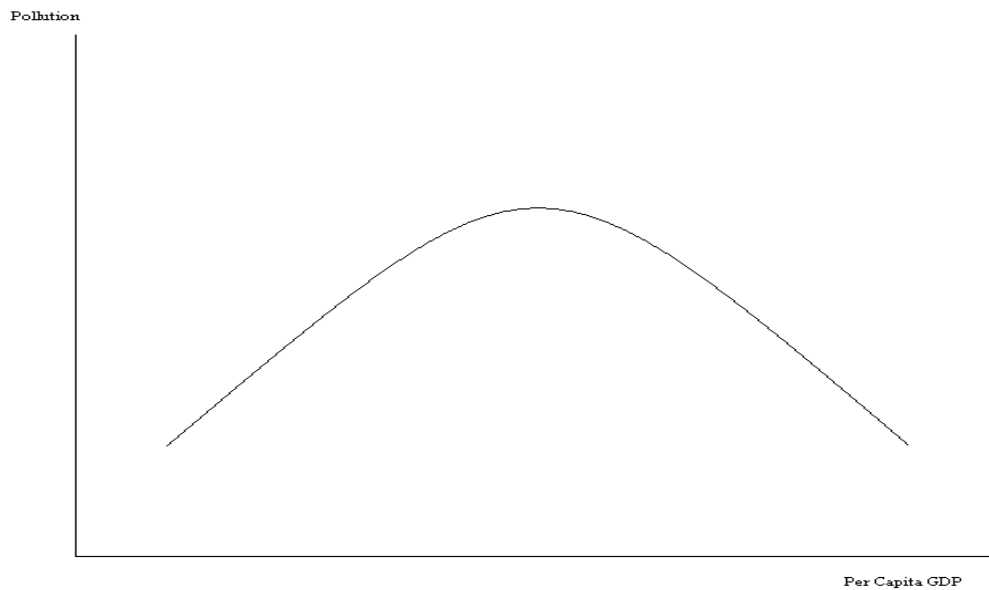
4. What two conditions are required for a bias to arise due to omitted variables? (3 points)

5. Does heteroskedasticity bias our estimates? What problems does it create for our estimates? (3 points)

6. What is the relationship between the mean and the median in a normal distribution? Intuitively, why is the distribution of income in a population unlikely to be normal? What is the likely relationship between mean and median of the distribution of income in virtually any country (even relatively egalitarian ones)? Why? (5 pts)

7. Assume you wish to estimate the following relationship empirically $y = e^{\alpha t}$ where y is a country's income, t is a time variable indicating years since democracy was adopted, e is a constant, and α is the parameter you wish to estimate. If you could only use OLS regression techniques and you had data on y and t , how would you estimate α ? Provide an explanation of how you would go about doing the regression and then, explain what α represents with respect to income. (5 points)

8. Suppose theory predicts that the relationship between pollution and per capita output (GDP) is represented by the following graph:



If you have data on every country's pollution level (e.g., carbon dioxide emissions) and per capita GDP how would you perform your regressions? What signs would you expect on your x variables if the graph above represents reality? Intuitively, what is the relationship between output and pollution? How would one determine the level at which pollution is maximized? (5 points)

9. Assume you have data on the lifespan (i.e., how many years each individual lives) of 1,000 individuals as well as data on the income levels of those individuals' parents while the individuals were growing up (i.e., average household income during the ages 0-18 for each individual). You also have the individual's race (Hispanic, black, or Asian to simplify things), and all the individuals are female (again to simplify things). What regression would you run if you think that the effect of parental income on lifespan is constant across races but there is a constant gap in lifespan among the various races (write out the regression with annotations explaining what you're doing, and graph the assumed relationship between parental income and lifespan)? What regression would you run if there is no constant gap, but the effect of parental income on lifespan varies by race (again, write it out, annotate, and graph)? (10 points)

10. Assume that the study above (in question #9) also includes information on the individuals' (not their parents) own education levels. Further assume that all the individuals studied lived all their lives in Utah (i.e., were born, educated, and died in Utah). What threats are there to the validity (internal and external) of the results you obtain in question 9's analysis (don't just name the kinds of threats described in the text; explain specifically how they might manifest themselves in this context)? Also, talk about how you could improve your confidence in the internal and external validity of the analysis. (10 points)

11. Explain how sample selection bias and simultaneity (endogeneity) bias can both be understood as omitted variables biases. (5 points)

12. Does measurement error in one of your independent variables bias your coefficient estimates? For measurement error in your dependent variable, when will your results be biased? When will they not be biased? (5 points)

13. You run the following regression: $crime_i = \hat{\alpha} + \hat{\beta} \cdot income_i + \hat{\delta} \cdot police_i + \hat{\phi} \cdot youngmen_i + \varepsilon_i$ where crime is the number of violent crimes committed in city i per 100,000 city residents; income is the per capita income in the city; police is a measure of how many police officers the city employs; and youngmen measures the fraction of the city's population that is male and between the ages of 15 and 30. You get the following Stata output: (20 points)

Variable	Coefficient	t-statistic	Standard Error	P value
Constant	100.00	5.00	20.00	0.00
Income	-0.50		0.24	
Police	2.00	4.00		0.00
Youngmen	10.00			0.10
R Squared	0.750			
Adj R Squared	0.680			
F Statistic	10.00			0.00

13.1 Construct a 95% confidence interval for your estimate of the effect of police on crime.

13.2 Is income a statistically significant determinant of crime at the 5% Type I error level (show calculations) using a two-tailed test?

13.3 Argue that you should use a one tailed test for 13.2 (explain argument). Is income a statistically significant determinant of crime at the 5% Type I error level (one-tailed test)? How about at the 1% level?

13.4 Why might we be estimating a positive effect of police on crime? What are the threats to the validity of this result?

13.5 What is the p value for our estimate of the effect of income on crime (it's missing in the table above, but you can figure it out)? What does this number represent?

13.6 What is the difference between the adjusted R squared and the R squared (explain; don't just do the arithmetic)?

13.7 How much of the variation in crime can we explain with this regression? Does the regression as a whole explain a statistically significant portion of the variation in crime? How do you know?

14. You have data on the alcohol consumption (gallons of ethanol consumed in a one year period, regardless of the form of alcohol) habits of 10,000 individuals and run a regression generating the following results: (15 points)

Variable	Coefficient	Standard Error	t statistic	P value
Constant	1.00	0.25	4.00	0.00
Smoker	1.50	0.50		
Income	0.25	0.10	2.50	
Education	-0.80	1.25		0.74
Adj R Squared	0.53			
F Statistic	85.00			0.00

14.1 Can we tell from this regression whether smoking (note: the smoker variable is an indicator = 1 if the individual smokes on a daily basis) causes drinking (i.e., smoking creates a desire to drink)? Why or why not?

14.2 Why might we have a multicollinearity problem with regard to income and education? Intuitively, what does multicollinearity do to our regression results?

14.3 Suppose you believe men and women have different drinking patterns, so you re-estimate the regression that was run above, adding a male dummy variable (= 1 if individual is male; 0 otherwise) **and** a female dummy variable (= 1 if individual is female; 0 otherwise). Do you need to make any other changes in the regression in order to estimate a male and a female effect? Why or why not?

14.4 Is income a statistically significant determinant of drinking at the 5% type I error level (two tailed test; show calculation)? What does type I error mean in the context of this problem? What is type II error? Do we control for type II error explicitly in this regression (explain)?