Econ 3180 – Midterm, March 1st 2013

You may use a calculator. Answer all questions in the answer book provided. The exam is 50 minutes long and consists of 100 marks.

A formula sheet, and a table of probabilities from the standard Normal distribution, are provided at the back of the exam booklet.

Exam version 1.

You must write your exam version number in your answer booklet.

Easy Question

[12 marks]

1.)

Consider the data:

$$Y = \{1,2,3,4\}$$

Calculate the sample average, $\bar{Y}$

Part A – Multiple Choice

[28 marks – 4 marks each]

2.) The sample average is a random variable and

a. is a single number and as a result cannot have a distribution.
b. has a probability distribution called its sampling distribution.
c. has a probability distribution called the standard normal distribution.
d. has a probability distribution that is the same as for the $Y_1,\ldots,Y_n$ i.i.d. variables.

3.) If Y is distributed $N(3,9)$, what is $\Pr(Y > 0)$?

a. 0.1587
b. 0.3707
c. 0.6293
d. 0.8413

4.) The Central Limit Theorem implies that, with a large enough sample size

a. $\bar{Y}$ cannot be Normally distributed
b. $\hat{\beta}_1$ is Normally distributed, if L.S.A. #2 holds
c. the outcome of a single die roll will follow a Normal distribution
d. all of the above

5.) In the simple linear regression model, the regression slope

a. indicates by how many percent Y increases, given a one percent increase in X.
b. when multiplied with the explanatory variable will give you the predicted Y.
c. indicates by how many units Y increases, given a one unit increase in X.
d. represents the elasticity of Y on X.

6.) When the OLS predicted values, $\hat{Y}_i$, exactly match the Y_i data,

a. $R^2 = \bar{Y}$.
b. $0 < R^2 < 1$.
c. $R^2 = 0$.
d. $R^2 = 1$.

7.) The OLS estimator is derived by

a. minimizing the distance between the fitted line, and the actual data points.
b. minimizing the explained sum of squares.
c. minimizing the total sum of squares.
d. minimizing the sum of squared residuals.
e. magic.

8.) Finding a small value of the *p*-value (e.g. less than 5%)

a. indicates evidence in favor of the null hypothesis.
b. implies that the *t*-statistic is less than 1.96.
c. indicates evidence against the null hypothesis.
d. will only happen roughly one in twenty samples.

Part B – Short Answer

[24 marks – 6 marks each]

9.) Consider a sample of three observations collected from the random variable, *Y*:

$$Y = \{1,4,7\}$$

Estimate the variance of *Y*. Why did you divide by "2" instead of "3"?

10.) The formula for the OLS estimator $\hat{\beta}_1$ may be rewritten as:

$$\hat{\beta}_1 = \beta + \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Briefly explain why Least Squares assumption #1 is required for $\hat{\beta}_1$ to be unbiased.

11.) Consider the estimated regression line:

$$\bar{Y} = 41.2 + 5.2X$$

What is the predicted value of *Y*, and OLS residual, when $X = 4$?

12.) In large samples, the sampling distribution for $\bar{Y}$ may be written as:

$$\bar{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

What happens to the variance of $\bar{Y}$ when the sample size, *n*, becomes very large? What desirable property of $\bar{Y}$ does this lead to?

Part C – Long Answer

[36 marks total]

13.)

Suppose that a researcher, using wage data on 180 randomly selected workers *with* a university education and 200 workers *without* a university, estimates the OLS regression,

$$\widehat{Wage} = \underset{(1.45)}{10.24} + \underset{(4.20)}{8.52} \times UNI, \qquad R^2 = 0.18$$

Where *Wage* is measured in $/hour and *UNI* is a "dummy" variable that is equal to 1 if the person has a university education and 0 if the person does not have a university education.

a. In the sample, what is the mean wage for workers who have obtained a university education?

[10 marks]

b. Conduct a formal hypothesis test to determine whether or not obtaining a university education will affect hourly wage.

[10 marks]

c. Suppose that the researcher forgot to use heteroskedastic-robust standard errors. Why might heteroskedasticity arise? Explain the problem with heteroskedasticity. How might it affect your answers to part (a) and part (b)?

[10 marks]

d. A different researcher uses the same data, but instead defines *UNI* = 1 for a person who *does not* have a university education, and *UNI* = 0 for a person who does have a university education. What will this researcher estimate for the intercept?

[6 marks]

END.