ECON 3040 - Midterm Answer Key (Fall 2015)

Part A - Multiple Choice

Version 1 (END)	Version 2 (end)
1. C	1. D
2. B	2. C
3. C	3. D
4. A	4. B
5. B	5. C

6. No correct answer. (The s.e.$(\hat{\beta}_1) = -0.52$)

Part B - Short Answer

7. $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$$\text{var}(\bar{Y}) = \text{var}\left(\frac{1}{n}\sum y_i\right) = \frac{1}{n^2}\,\text{Var}\left(\sum y_i\right)$$

By the independence of the y_is:

$$\frac{1}{n^2}\text{var}\left(\sum y_i\right) = \frac{1}{n^2}\sum \text{var}(y_i) = \frac{1}{n^2}\sum \sigma_Y^2 = \frac{1}{n^2}\, n\sigma_Y^2 = \frac{\sigma_Y^2}{n}$$

Two reasons why we want to know this variance:

(i) To compare this variance to the variance of other estimators of μ_Y (efficiency), and to show $\bar{y}$ is consistent.

(ii) To construct an estimator for the variance of $\bar{Y}$ (e.g. S_Y^2/n).

8. L.S.A. #1 : $E(u \mid X = x) = 0$ (on formula sheet)

The expected value of the random error term (u), conditional on observing a value of X, is zero.

This assumption is required in order for OLS to be <u>unbiased</u>.

9. From the discussion in class, there are two ways to go about this. Either:

i) define a dummy variable D, where $D = 1$ if the individual is male, and $D = 0$ if female; or

ii) $D = 1$ if female and $D = 0$ if male.

The choice of how the dummy variable is defined is not substantive. ~~Following (i),~~ An appropriate population model is:

$$y_i = \beta_0 + \beta_1 D_i + u_i \, , \quad \text{where } y \text{ is hourly wages.}$$

Given the sample averages for men and women:

$$\hat{\beta}_0 = 14.5 \, , \quad \hat{\beta}_1 = 1.7 \qquad \text{for (i)}$$

$$\hat{\beta}_0 = 16.2 \, , \quad \hat{\beta}_1 = -1.7 \qquad \text{for (ii)}$$

The estimated wage gender gap is $\hat{\beta}_1$.

10. The basic idea behind the CLT is that ~~when~~ the sum of random variables, regardless of how they are distributed, tend to be Normal. This implies that estimators whose formulas entail summing the random sample data, tend to have a Normal sampling distribution.

If we look at the formula for $\hat{\beta}_1$, we see that there is a summation term involving the random variable, y_i:

$$\hat{\beta}_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

So, the CLT determines that the sampling distribution of $\hat{\beta}_1$ is Normal (provided the sample size is large enough).

Part C - Long Answer

11. a) $R^2 = 0.029$ means that the variable "str" only explains 2.9% of the variation in "score". It does not mean that $\hat{\beta}_1$ is statistically insignificant, nor that "str" is unimportant in a policy sense.

b) The null and alternative hypotheses are:

$$H_0: \beta_1 = 0 \quad ; \quad H_A: \beta_1 \neq 0 .$$

The t-statistic for this test is:

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\widehat{s.e.}(\hat{\beta}_1)} = \frac{-1.06 - 0}{0.42} = -2.52 .$$

Since the sample size is relatively large ($n = 200$), we can assume $t \sim N(0,1)$.

The p-value associated with this null hypothesis is

$$.00587 \times 2 = 0.012 .$$

We reject the null hypothesis at the 5% significance level, but not at the 1% level.

c) $s_Y^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$

$TSS = \sum (y_i - \bar{y})^2$

$TSS = (n-1) s_Y^2 = 199 \times 128.6 = 25591.4$

$R^2 = \frac{ESS}{TSS}$, $ESS = R^2 \times TSS = 0.029 \times 25591.4 = 742.15$

d) $\widehat{score} = 658.65 - 1.06(20) = 637.45$

e) β_0 is the expected value of "score" for a class size of zero. In this model the intercept does not have any economic meaning.

f) $C.I. = -1.06 \pm 1.96 \times 0.42 = (-1.88, -0.24)$