ECO220Y, Term Test #4 SOLUTIONS

March 11, 2016, 9:10 – 11:00 am

(1) (a) For each of the eight budgets:

$$H_0: \mu_{male} - \mu_{female} = 0$$

$$H_1: \mu_{male} - \mu_{female} \neq 0$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(\bar{X}_{male} - \bar{X}_{female}) - 0}{\sqrt{\frac{s_{male}^2}{n_{male}} + \frac{s_{female}^2}{n_{female}}}}$$

For each budget we have the sample means and sample sizes, so the only thing missing is the sample standard deviation of money passed for females and the sample standard deviation of money passed for males (for each budget).

(b) No, they do not. Two things can lead to bigger t-statistics: (1) larger differences between males and females and (2) smaller standard errors. Because the sample sizes in ECO220Y for both males and females are much bigger than the original study we would expect smaller standard errors and hence bigger t-statistics (other things equal). There are a couple of budgets where the difference between males and females is bigger in ECO220Y than the original study (Budgets 2 and 5) so this and the larger sample sizes both work towards bigger t-statistics (in absolute value). However, for most budgets there is actually a bigger difference between males and females in the original study so the key factor is the sample sizes.

(c) For Budget 2:

$$H_0: \left(p_{male,A\&V} - p_{male,220}\right) = 0$$

$$H_1: \left(p_{male,A\&V} - p_{male,220}\right) \neq 0$$

A P-value of 0.1938 means that we do not have sufficient evidence to conclude that there is a difference in how males behaved – in terms of the fraction choosing to pass any money to the other player when confronted with Budget 2 – between the original A&V study and the ECO220Y replication study.

(d) For Budget 3:

$$H_0: \left(\mu_{female,A\&V} - \mu_{female,220}\right) = 0$$

$$H_1: \left(\mu_{female,A\&V} - \mu_{female,220}\right) \neq 0$$

A P-value of 0.0023 means that we have very strong evidence (even at a 1% significance level) to conclude that there is a difference in how females behaved – in terms of the mean money passed among those choosing to pass any money to the other player when confronted with Budget 3 – between the original A&V study and the ECO220Y replication study.

(e) MARKING RUBRIC:

+2 Understands that Tables 1 and 2 compare sexes within each study while Tables 3 and 4 compare each sex across the studies

> (Tables 1 and 2 compare the average generosity between sexes *within* each study. Tables 3 and 4 look for differences in male behaviour *across* the studies and differences in female behavior *across* the studies: in other words, did the males in A&V (2001) behave differently that the males in ECO220Y (2014, 2015, and 2016).)

+2 Understands that Tables 1 and 2 look at average money passed whereas Tables 3 and 4 decompose this into fraction passing any money and average money passed conditional on passing some.

(Further, rather than just look at average generosity by sex, by budget, Tables 3 and 4 break generosity down into two components: giving some money (yes/no) and how much money you pass (conditional on yes). Table 3 compares the fraction of males passing any money across the two studies and the fraction of females passing any money across the two studies. Table 4 compares the average generosity among those choosing to pass at least some money across the two studies and by sex.)

+3 Recognizes that in ECO220Y there a higher fraction passing money than the original study for both males and females (i.e. by this measure both sexes are more generous in ECO220Y than the original study). But, some of these differences are not statistically significant. Some differences are fairly large: more than 10 percentage points.

+3 Recognizes that in ECO220Y both sexes are generally also more generous if you look at the mean amount of money passed conditional on choosing to pass some (except for females for Budgets 6 – 8). But, again, some of these differences are not statistically significant. Some differences are fairly large: more than $1.00 (which is non-trivial in this particular game where not that much money was at stake).

NOTE: Total points cannot exceed 8 (the marking rubric gives students a chance to earn "bonus" points for a good part of an answer to make up for something else they missed.)

(2) (a) $R^2 = \frac{SSR}{SST} = \frac{30.8083359}{49.4794791} = 0.6226$

About 62 percent of the variation in the price of a Big Mac across countries is explained by variation in the GDP per capita.

(b) $b_1 \pm t_{\alpha/2} s.e.(b_1)$ with $\nu = n - 2$

$b_1 \pm t_{\alpha/2} s.e.(b_1) = 0.0348673 \pm 2.012 * 0.0039593 = \ 0.035 \pm 0.008$; $LCL = 0.027$ and $UCL = 0.043$

We are 95% confident that on average countries with GDP per capita that is $1,000 USD higher have Big Mac prices that are between 2.7 and 4.3 cents higher (USD).

(c) $\hat{y}_{x_g} \pm t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{X})^2}{(n-1)s_x^2}}$ or $\hat{y}_{x_g} \pm t_{\alpha/2} \sqrt{(s.e.(b_1))^2 (x_g - \bar{X})^2 + \frac{s_e^2}{n} + s_e^2}$ with $\nu = n - 2$

$$= (2.522506 + 0.0348673 * 15) \pm 2.012 * 0.63028 \sqrt{1 + \frac{1}{49} + \frac{(15 - 30.55949)^2}{(49-1)(22.97713)^2}}$$

$$= 3.046 \pm 2.012 * 0.63028 \sqrt{1 + \frac{1}{49} + \frac{(15 - 30.55949)^2}{(49-1)(22.97713)^2}} = 3.046 \pm 2.012 * 0.64 = 3.05 \pm 1.29$$

The point prediction is that the price of a Big Mac will be $3.05 USD for this country and the margin of error (for a 95% confidence level) is $1.29 USD.

(d) Draw a histogram of the regression residuals $e_i = y_i - \hat{y}_i$ and verify that they are approximately Normal (Bell shaped). (Note: The textbook also talks about a more advanced method – the Normal probability plot (see p. 652). While this topic is not a required part of the course, it is also an acceptable answer so long as it is described (a plot of the regression residuals versus the expected Normal scores for a sample from N(0,1) containing the same number of observations).)

(3) (a)

country	growth_1950_60	growth_1960_70	growth_1970_80	…	growth_2000_10
China	#	#	#	…	#
…	…	…	…	…	…
Thailand	#	#	#	…	**0.036**

MARKING RUBRIC:

+2 Understands that the estimated average annual growth rate in Thailand from 2000 through 2010 is 3.6%.

+2 Understands that 0.036 is one piece of *datum* (observation for a particular variable) in the data used to estimate the regressions in Table 1

+2 Understands that there were also regressions for Thailand (and each other country) for each other decade (i.e. many, many regressions like that show in the figure such as Thailand from 1990 – 2000, Japan from 1970 – 1980, etc.) and these form the rest of the observations in the data used to estimate the regressions in Table 1

+2 Understands structure of the data: variables growth_1950_60, …, growth_2000_2010 and observations are a country (see page 3 in "Reading Guide for "Asiaphoria Meets Regression to the Mean" for Test #1 (and beyond)")

(b) MARKING RUBRIC:

+2 Understands there are 8 regression reported in Table 1 and what the x and y variables are in each regression

+2 Understands that PANEL A uses each country's growth in the immediately preceding decade to predict a decade's growth whereas PANEL B uses each country's growth two decades earlier to predict a decade's growth

+2 Correctly notes that the R^2/slope are basically zero in PANEL B

+2 Explains these results further support the caution against the idea that current performance is good predictor of future performance

(4)

$$\bar{X}_d \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

$$s_d = SD[X_{local} - X_{online}] = \sqrt{V[X_{local}] + V[X_{online}] - 2 * SD[X_{local}] * SD[X_{online}] * CORR[X_{local}, X_{online}]}$$

$$s_d = \sqrt{230.80^2 + 251.99^2 - 2 * 230.80 * 251.99 * 0.88} = 120.03$$

$$\nu = n - 1 = 12 - 1 = 11$$

$$\bar{X}_d \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}} = (842.87 - 729.54) \pm 2.201\frac{120.03}{\sqrt{12}} = 113.33 \pm 2.201 * 34.65 = 113.33 \pm 76.26$$

$LCL = 37.07$ and $UCL = 189.59$

Our estimate of the size of the difference between online and local quotes is that the local quotes are on average $113 more expensive and the margin of error on that estimate is $76: we are 95% confident that the local quotes are on average $37 to $190 more expensive than online quotes for all customers. The modest sample size of 12 explains why the interval is quite wide (the ME quite big), which leaves us unable to make a very precise inference about the real difference between local and online prices.