ECO220Y1Y, April 2015 Exam, SOLUTIONS

Part 1:

(1) (a) The differences in the mean and standard deviations between the two groups are entirely sampling error: the 85 students were randomly divided into the two groups and any differences between the central tendency and variability are just chance.

(b)

$H_0: \mu_{Goal} - \mu_{Control} = 0$

$H_1: \mu_{Goal} - \mu_{Control} > 0$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(2.91 - 2.46) - 0}{\sqrt{\frac{0.65^2}{45} + \frac{1.06^2}{40}}} = 2.32$$

This meets a 2.5% significance level (and hence also a 5% significance level) but not quite a 1% level (critical value is 2.4 for a 1% level).

(c)

$H_0: \mu_d = 0$

$H_1: \mu_d \neq 0$ or $H_1: \mu_d > 0$ (The excerpt implies a one-tailed test but we cannot check for sure because the P-value will be less than 0.01 whether a one or two-tailed test is conducted.)

These are paired data.

$$t = \frac{\bar{d} - \Delta_0}{s_d/\sqrt{n}}$$

$$4.17 = \frac{0.66 - 0}{s_d/\sqrt{45}}$$

$s_d = 1.0617$

$$s_d = \sqrt{s_1^2 + s_2^2 - 2s_1 s_2 r}$$

$$1.0617 = \sqrt{0.93^2 + 0.65^2 - 2 * 0.93 * 0.65 * r}$$

$r = 0.13$

(2) (a) $b_1 \pm t_{\alpha/2} s.e.(b_1)$ where the degrees of freedom is $\nu = n - 2$

$0.236 \pm 2.576 * 0.010$

$LCL = 0.210$ and $UCL = 0.262$

We are 99% confident that workers with a license (or certification) have wages that are 21% to 26% higher than those without a licence (or certification) on average.

(b)

$H_0: p_M - p_F = 0$

$H_1: p_M - p_F \neq 0$

$$\bar{P} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{3402 + 3813}{13035 + 12668} = 0.2807$$

$$z = \frac{\hat{P}_2 - \hat{P}_1}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1} + \frac{\bar{P}(1-\bar{P})}{n_2}}} = \frac{0.261 - 0.301}{\sqrt{\frac{0.2807(1-0.2807)}{13035} + \frac{0.2807(1-0.2807)}{12668}}} = \frac{-0.04}{0.0056} = -7.14$$

This difference is highly statistically significant at better than a 1% significance level (P-value is virtually zero).

(c) $\ln(wage)_{it} = \beta_0 + \beta_1 License_i + \beta_2 Female * License + \varepsilon_{it}$

$H_0: \beta_2 = 0$

$H_1: \beta_2 > 0$

Note: Some students may write a slightly more complex model (e.g. $\ln(wage)_{it} = \beta_0 + \beta_1 License_i + \beta_2 Female_i + \beta_3 Female * License + \varepsilon_{it}$), which is fine so long as the test matches (in example I gave, test beta3)).

(3) (a) Homes in the sample on average are 1,890 square feet with a standard deviation of 827 square feet, which is likely positively skewed. 27.4% of the homes in the sample have an electric stove. On average a household has 0.226 children aged 0 – 5 with a standard deviation of 0.655 children, which is clearly a highly positively skewed distribution.

(b)

$H_0: \beta_{1940} = 0$

$H_0: \beta_{1940} \neq 0$

$$t = \frac{b_1 - \beta_{10}}{s.e.(b_1)} = \frac{1.060 - 0}{0.519} = 2.04$$

Using the t with the limit of degrees of freedom (given that $\nu = n - 2 = 16{,}301 - 2 = 16{,}299$) the P-value will be between 0.02 and 0.05. (Alternatively, because the Student t converges to the Standard Normal we can use the Normal table to obtain the P-value = 0.0414.) Hence, we get the same answer as the simple rule of thumb: it is statistically significant at the 5% level (but not a 1% level).

(c) The constant term (19.080) is the current annual mean MBTU of homes built before 1940, which is the omitted category. This is perfectly consistent with the figure, which shows that homes built before 1940 use just short of 20 MBTU on average annually.

(d) Current mean electricity consumption is 3.084 MBTU's higher for homes built in the 1950s compared to homes built before 1940. It is 10.250 MBTU's higher for homes built between 2005 – 2008 compared to homes built before 1940. This relates directly to the figure which shows more recently built homes currently use more electricity. These differences are both highly statistically significant ($t = 6.81$ and $t = 10.07$) and large (economically significant), which means that they are significant.

(e) After controlling for home vintage, occupant characteristics, appliances, survey year, climate and some other building characteristics, homes that are 1% larger on average use 0.09626 MBTU more (Specification 2) of electricity per year currently and according to Specification 3, the prediction is 0.306% more.

(f) After controlling for home vintage, building characteristics, appliances, survey year, climate and some other occupant characteristics, homes that are owned by the occupants on average use 3.481 MBTU less (Specification 2) of electricity per year currently compared to homes that are not owned by the occupants and according to Specification 3, the prediction is 11.4% less.

(g) The constant term has no interpretation because all of the x's taking a value of zero is clearly beyond the range of the data (how can a house have 0 square feet?).

(h) After controlling for home vintage, building characteristics, occupant characteristics, appliances, and climate, homes in 2009 on average used 1.980 MBTU more electricity than homes in 2003, which is highly statistically significant. However, this does not necessary mean that average electricity use went up because we are controlling for lots of other factors that changed over that period and are associated with electricity use. (To check if it went up would require a simple regression where the only x variable is RASS 2009.)

(i) 5.2 percent of the variation in current HH electricity usage can be explained by variation in when the home was built, which means the vast majority of variation is explained by other factors. Even in Specification (2), which controls for lots of other factors about the homes and HHs, still only 36% of the variation in HH electricity use is explained. (It is very hard to accurately predict a HH's annual electricity usage with these kind of data. Notably, previous years' use is not an explanatory variable: that would definitely lead to a much bigger R^2.)

(j) While Specification (1) shows, as did the figure, that current annual electricity consumption is higher among more recently constructed homes, that fails to control for other factors such as the location of newer home (hotter areas that require more AC), the size of homes, how many people live in the home, and other factors. This means these other variables are lurking/confounding/unobserved as they are correlated with both home vintage and with electricity use and hence cause endogeneity bias. However, even once we control for those other variables, Specifications (2) and (3) show *very* disappointing efficiency gains. In Specification (2), the most recently constructed homes (2005 – 2008) are somewhat of an exception in that they do use somewhat less (3.9 MBTU less electricity annually on average) than homes constructed prior to 1940 after controlling for building characteristics, occupant characteristics, appliances, survey year and climate and this difference is statistically significant. However, in Specification (2), none of the other post-1980 vintages are statistically different from pre-1940. Further, in Specification (3), we cannot conclude that newer homes of any vintage use less electricity, even though we have controlled for building characteristics, occupant characteristics, appliances, and climate. Note that in Specification (3) the only negative vintage coefficient is for the most recently constructed homes (build in 2005 – 2008), but that is not even close to being statistically different from zero. However, even ignoring the fact it's not statistically significant, the point estimate is only 3.7% less electricity usage (for the most recently constructed homes compared to the oldest homes): nowhere near 80% lower, which was the original goal of the building codes. Further, in Specification (3), the other post-1980 vintage coefficients are *positive*, suggesting that even once we control for the larger modern homes in hotter climates with more occupants (etc.), the modern homes still appear *worse* in terms of energy efficiency compared to the pre-1940 homes. (Remember that a positive coefficient means *more* average annual MBTU usage for that vintage compared to pre-1940 homes after controlling for building characteristics, occupant characteristics, etc.)

ECO220Y1Y, April 2015 Exam, SOLUTIONS

Part 2:

(1) [2pts] Which sample size matches the graph? **(B)**

(2) [2pts] Compared to the situation in the graph, if the sample size were 5,000, which statement is TRUE? **(B)**

(3) [2pts] For one randomly selected ON public sector employee making at least $100,000, what is the probability that that person's salary is over $102,000? **(E)**

(4) [3pts] For a random sample of 180 ON public sector employees making at least $100,000, what is the probability that the sample mean is less than $134,000? **(D)**

(5) [2pts] For Specification 1, if a country's RQ index is 1, what is the 99% prediction interval for its GDP per capita? **(A)**

(6) [2pts] For Specification 1, if 2012 GDP per capita were in dollars (instead of $1,000s), what would change? **(D)**

(7) [2pts] For Specification 1, consider the diagnostic plot shown to the right. Canada has a 2012 GDP per capita of $52,400 (current US$) and an RQ index of 1.7. Where is Canada on the graph? **(E)**

(8) [2pts] For Specification 2, approximately what is the critical value for the test of overall statistical significance at a 5% significance level? **(C)**

(9) [2pts] For Specification 2, the STATA output says "`Prob > F = 0.0000`." What does this imply? **(A)**

(10) [2pts] For Specification 2, if <u>2002 GDP</u> were in dollars (instead of $1,000s), but everything else was the same as the original specification, which statement is TRUE? **(E)**

(11) [2pts] In 1960, for households with a husband and wife, which of these scenarios is the *most likely*? **(A)**

(12) [2pts] In 1960, if a husband has a college degree (C), what is the chance his wife has a college degree (C)? **(C)**

(13) [3pts] In 2005, if husbands and wives were randomly assigned to each other – in other words, if there is no assortative mating – what number would replace 0.104 (i.e. the cell C, C)? **(B)**

(14) [2pts] For 3 randomly selected wives in 2005, what is the chance exactly 1 has a post-college degree? **(D)**

(15) [2pts] If 3M were dropped, the coefficient on 1M Treasury would ___. **(D)**

(16) [1pts] If 3M were dropped, the value 477.328512 (Total, SS) in the *ANOVA* table would ___. **(C)**

(17) [1pts] If 3M were dropped, the value 0.078 (Standard Error) in the *Regression Statistics* table would ___. **(A)**

(18) [3pts] Is there a statistically significant difference between grocery stores that eventually face Walmart entry versus those that do not in terms of the mean grocery selling space in 1990? **(B)**

(19) [3pts] Is there a statistically significant difference between grocery stores that eventually face Walmart entry versus those that do not in terms of the percent of grocery stores affiliated with a chain in 1990? **(B)**

(20) [2pts] In the excerpt above, what does the statement "Stores that would and would not eventually face competition from Walmart were mostly similar across these dimensions" mean? **(C)**