ECO220Y, Term Test #2: SOLUTIONS

December 4, 2015, 9:10 – 11:00 am

(1) $(\hat{P}_2 - \hat{P}_1) \pm z_{\alpha/2}\sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{n_2} + \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}}$

$\hat{P}_2 = \frac{89}{177} = 0.503$ and $\hat{P}_1 = \frac{42}{180} = 0.233$

$$(0.503 - 0.233) \pm 1.96\sqrt{\frac{0.503(1-0.503)}{177} + \frac{0.233(1-0.233)}{180}}$$

$(0.269) \pm 1.96 * 0.049$

0.269 ± 0.096

$LCL = 0.173$ and $UCL = 0.366$. We are 95% confident that among *all* potential non-mobility impaired users of the elevator posting the "electricity" message will cause the percent using the stairs to be between 17.3 and 36.6 percentage points higher compared to the "exercise" message. The point estimate of the difference is 26.9%, which is a big difference in the percent using the stairs depending on the message shown, with a margin of error of 9.6%. The electricity message is much more effective than the exercise message in prompting people to use the stairs. (The reason the causal interpretation is warranted is because these are experimental data: the x-variable in this case is the message shown, which has been *randomly* set.)

(2) (a) MARKING RUBRIC:

+1 Recognizes that the natural log straightens the scatter plot

+3 Correctly interprets the two OLS slopes and is context-specific (e.g. speaking about China during the relevant time period)

+2 Correctly interprets the R^2 (noting how high it is)

+1 Correctly notes that the OLS intercepts have no interpretation in this case (far outside range of the data)

+3 Correctly points out that despite huge R^2 values, which would seem to imply the obvious but wrong conclusion that the line can continue to make good predictions into the future, extrapolating beyond the range of the data is highly inadvisable (cannot conclude that China and India will continue to experience annual growth rates of 8.5% and 4.1% respectively)

(b) MARKING RUBRIC:

+3 Correctly points out that Japan's growth rate is very different during these three consecutive time periods: correctly interprets the slope coefficients (fast, medium, and, most recently, non-existent)

+3 Correctly notes that, just like the earlier graphs, Japan's fast growth was very steady: i.e. notes the huge R^2 in the two earlier time periods just like we see today with India and China

+3 Pulls this together to correctly point out that despite huge R^2 values extrapolating beyond the range of the data is highly inadvisable: i.e. understands why Japan is a perfect example to support the author's main warning about forecasting future performance based on past performance

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).

(c) MARKING RUBRIC:

+3 Demonstrates understanding of what the x and y variables are in the reported regressions

+3 Correctly interprets the OLS slope

+3 Correctly interprets the R^2 (correctly noting that it very small)

+2 Draws a valid conclusion overall

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).

(3) From the Normal table obtain two things:

$P(Z < -1.89) = 0.029$

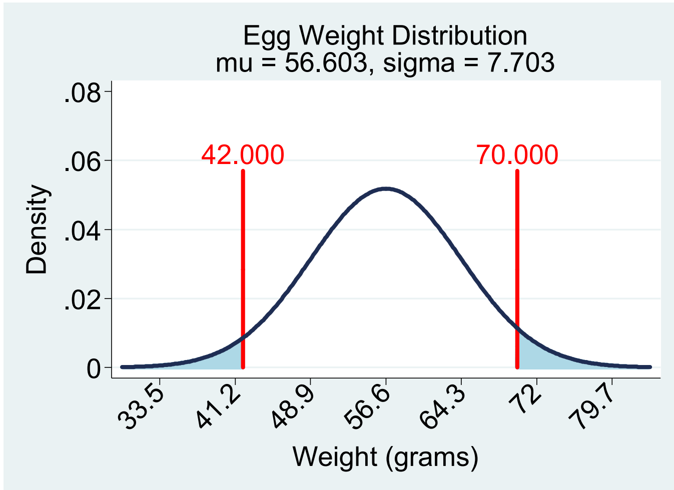
$P(Z > 1.74) = 0.041$

Use this to get two equations with two unknowns:

$$-1.89 = \frac{(42 - \mu)}{\sigma}$$

$$1.74 = \frac{(70 - \mu)}{\sigma}$$

Solve for μ and σ to obtain $\mu = 56.6$ grams and $\sigma = 7.7$ grams.



(4) (a) Define X2 to be annual GDP growth between 2000 and 2010. Define X1 to be annual GDP growth between 1990 and 2000.

$$\mathrm{V[X2 - X1]} = \mathrm{V[X2] + V[X1]} - 2 * \mathrm{CORR[X1, X2]} * \mathrm{SD[X1]} * \mathrm{SD[X2]}$$

$$1.25^2 = 1.00^2 + 1.37^2 - 2 * \mathrm{CORR[X1, X2]} * 1.37 * 1.00$$

$$-1.31 = -2 * \mathrm{CORR[X1, X2]} * 1.37 * 1.00$$

$$0.48 = \mathrm{CORR[X1, X2]}$$

(b) Define X2 to be annual GDP growth between 1990 and 2000. Define X1 to be annual GDP growth between 1980 and 1990.

$$\mathrm{MEAN[X2 - X1]} = \mathrm{MEAN[X2]} - \mathrm{MEAN[X1]} = 1.584 - 0.660 = 0.924 \text{ percentage points}$$

$$\mathrm{V[X2 - X1]} = \mathrm{V[X2] + V[X1]} - 2 * \mathrm{CORR[X1, X2]} * \mathrm{SD[X1]} * \mathrm{SD[X2]}$$

$$\mathrm{V[X2 - X1]} = 2.556^2 + 3.170^2 - 2 * 0.3422 * 3.170 * 2.556 = 11.04$$

$$\mathrm{SD[X2 - X1]} = \sqrt{11.04} = 3.32 \text{ percentage points}$$

(5) (a) [6 pts]

$$P\left(\hat{P} \leq \frac{2}{12} \mid p = 0.2, n = 12\right) = P(X = 0) + P(X = 1) + P(X = 2) =$$

$$= \frac{12!}{0!\,(12-0)!} 0.2^0 (0.8)^{12} + \frac{12!}{1!\,(12-1)!} 0.2^1 (0.8)^{11} + \frac{12!}{2!\,(12-2)!} 0.2^2 (0.8)^{10} = 0.558$$

This is a high probability (55.8% chance) and it is not at all surprising if so few say they plan to take advantage: this could easily just be sampling error.

(b)

$$P\left(\hat{P} \leq \frac{2}{12} \mid p = 0.2, n = 1200\right) = P\left(Z < \frac{\frac{2}{12} - 0.2}{\sqrt{\frac{0.2 * 0.8}{1200}}}\right) = P(Z < -2.89) = 0.0019$$

This is a low probability (0.2% chance) and it is surprising if so few say they plan to take advantage: sampling error is not likely to cause such a big discrepancy between the sample proportion and the (claimed) population proportion.

(Note: The difference between (a) and (b) is that with a large sample size there will be less sampling error in measuring the sample proportion: hence being off by a few percentage points is not surprising in a small sample but would be surprising in a large sample.)

(6) (a) Using the STATA summary of the Monte Carlo simulation (or the graph), we cannot get the exact probability, but we can see that the probability of such a low sample median would definitely be less than 1% because the 1st percentile is $107,395. Hence, obtaining a sample median as low as $105,000 due to sampling error would be very surprising. Sampling error is an implausible explanation.

(b) The answer would not change. Simulation error is already really tiny and hence the picture we had of the sampling distribution was already very clear.

(c) A larger sample size would mean less sampling error for the sample median. Hence, we would expect the sampling distribution to be less spread out (taller and thinner distribution). This would mean that sampling error would be an even more implausible explanation for such a low sample median (i.e. the probability would be even more miniscule).