ECO220Y1Y, Test #4, Prof. Murdock SOLUTIONS

(1) (a)

$H_0\colon \beta_{50\_59} = 0$

$H_1\colon \beta_{50\_59} \neq 0$

$t = 1.79$ and given the very large sample size of over 6,000 homes, treat ν as infinity and use the Normal table:

$P(t < -1.79) + P(t > 1.79) = 2 * (0.5 - 0.4633) = 0.073$

Hence, after controlling for house size, number of residents, presence of central air conditioning, and climate zone, we can conclude that 2009 annual electricity use in California homes built from 1950 to 1959 on average differs from homes built before 1940 at a 10% significance level (but not a 5% level).

(b) That is the LCL and UCL of the 95% CI estimate of the slope: $b_j \pm t_{\alpha/2} s_{b_j}$ (where, again, the degrees of freedom can be approximated as infinity given the very large sample size of over 6,000 homes)

$-.1454542 \pm 1.960 * 0.0330391$

$-.1454542 \pm 0.064757$

$LCL = -0.210$ and $UCL = -0.081$

(c) ln_elec_mmbtu-hat = -0.3212621 + 0.2460722*0 + 0.4137596*ln(1750) + 0.2681087*ln(3) + 0.054859*1 = 3.117841

exp(3.117841) = 22.6 MMBTUs

(d) <u>14</u>; <u>23</u>; <u>smaller</u>; <u>the same</u>; <u>larger</u>

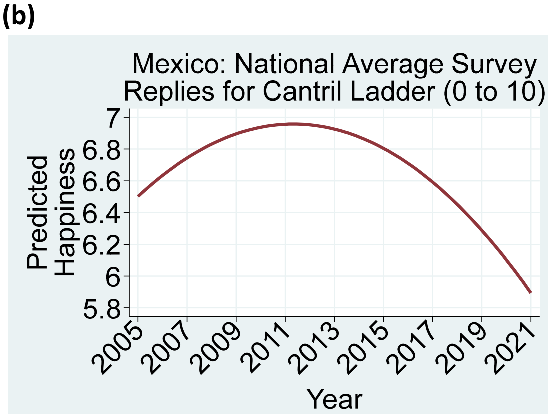
(2) (a)

$H_0\colon \beta_1 = \beta_2 = 0$

H_1: not all slopes are zero

Excel reports the P-value as 0.001644677 which means that this multiple regression is statistically significant overall at a 1% level, but not a 0.1% level.

(b)



(c) $s_e = Root\ MSE = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{0.947504494}{16-2-1}} = 0.27$ This measures the amount of scatter for the figure in Part (a) imagining adding the dots for each observation. Given that average happiness in Mexico reaches a maximum of less than 7 and a minimum above 5.8, having a standard deviation of the residuals of about 0.3 is large: there is fair bit of scatter and the predicted happiness in each year is only roughly matching the actual data point.

(3) (a) b_0: <u>positive</u>; <u>intercept for 2018</u>

b_1: <u>negative</u>; <u>slope for 2018</u>

b_2: <u>negative</u>; <u>difference in intercept for 2006 versus 2018</u>

b_3: <u>positive</u>; <u>difference in slope for 2006 versus 2018</u>

(b) $\hat{y}_{x_g}$ is the predicted share for 2018 when real provincial income per capita in 2006 is 20 million VND: Figure 3 shows it is roughly 0.61. *Interpretation:* For a province in Vietnam with real provincial income per capita of 20 million Vietnamese dollars in 2006, we predict that 61 percent of households operate a farm or non-farm business in 2018.

(c) As the real provincial income per capita in 2006 rises by 5 million Vietnamese dollars we observe that the annual rate of entry of informal businesses from 2006 to 2008 declines by 2.8 percentage points on average.

[In other words, more economically developed provinces have a substantially lower entry rate than less developed provinces where the least developed have predicted entry rates around 24% versus only 11% for the most developed.] [Alternatively: As the real provincial income per capita in 2006 rises by 1 million Vietnamese dollars we observe that the annual rate of entry of informal businesses from 2006 to 2008 declines by 0.6 percentage points on average.]

(4) (a) Using the slope estimate from Column (1), which estimates the *annual* change in first year GPA, yields 11*0.019 = 0.209. Hence, from 1990 to 2000 across these nine large public universities in the United States, first year GPAs, which are measured on a four-point scale, increased by 0.21 on average over this 11-year period. This is a sizable increase for just over a decade: for example, an average GPA of 3.11 versus 2.90 is a notable increase for students and institutions, which means it is economically significant.

(b) In Column (1) we should still see a positive and statistically and economically significant result: if students are getting stronger, grades in first year university courses would rise over time. However, in Column (7) we should see a slope estimate that is not statistically different from zero: once we control for students' rising ability as measured by standardized test scores, there is no change in first year GPAs.

[Column (7) also controls for any changes in students' choice of majors or courses because they are academically stronger.]

(5) (a) <u>1</u>; <u>1</u>; <u>2</u>; <u>larger</u>

(b) Column (1) finds that red slides on average receive a bid that is about 3.9% higher than black slides, and this is statistically significant at a 1% level. In contrast, Column (3) finds that after we control for the quantity of slides available, red slides get a bid about 1.3% higher, but this difference relative to black slides is not statistically significantly different from zero. The key reason for the dramatically different results is that Column (1) is a simple regression – a simple comparison of red to black slides – whereas Column (3) is a multiple regression that controls for the quantity of slides available. This fits with the authors argument that it is not the color but rather the rarity – there were fewer red than black slides put to the auction and bidders knew this information – that is driving higher bids for the red slides.