ECO220Y1Y, Test #2, Prof. Murdock: SOLUTIONS

November 15, 2019, 9:10 – 11:00 am

NOTE: The parts of the solutions [in brackets] are extra explanations and are not required parts of your answer.

(1) These are the steps (note that you may combine steps 1) and 2)):

1) Construct a new variable (column) for GDP per capita = rgdpna/pop
2) Construct a new variable (column) for the natural log of GDP per capita = ln(rgdpna/pop)
3) Run a simple regression where the y-variable is ln(rgdpna/pop) and the x-variable is year using only the six observations from 2012 through 2017: the slope coefficient for the year variable estimates the annual growth rate of real GDP per capita during that period: multiply by 100 to get the percentage growth rate

(2) (a) $b_1 = \frac{s_{xy}}{s_x^2} = \frac{61.6322}{76.9928} = 0.8005$

(b) Recall $s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n-1} = \frac{SST}{n-1}$. From var-cov matrix, $s_y^2 = 977.224$, hence, $SST = s_y^2 * (n-1) = 977.224 * 49 = 47884$

(3) Only City B is experiencing racial segregation across neighborhoods. In low income neighborhoods in City B, about 23% are white, whereas in high income neighborhoods nearly 50% are white, which are computed as $P(W \mid L) = 0.0930/0.4055 = 0.2293$ and $P(W \mid H) = 0.1465/0.2957 = 0.4954$, respectively. Hence, richer neighborhoods are disproportionately white whereas lower income neighborhoods are disproportionately non-white: race and neighborhood are *not* independent because neither of the conditional probabilities above is even close to the overall proportion white, which is about 34%. In contrast, in City A, race and neighborhood are independent, which means no segregation. In City A, 19% of the city is white, which is computed as $P(W) = 0.0659 + 0.0530 + 0.0712 = 0.19$, and 19% of each neighborhood is white, regardless of whether it is low or high income: $P(W \mid L) = 0.0659/0.3469 = 0.19$ and $P(W \mid H) = 0.0712/0.3744 = 0.19$.

(4) (a) Event **A** ____IS______ mutually exclusive (i.e. disjoint) of Event **B**.

Event **A** ___IS NOT___ independent of Event **B**.

Event **A** ___IS NOT___ mutually exclusive (i.e. disjoint) of Event **C**.

Event **B** ___IS NOT___ mutually exclusive (i.e. disjoint) of Event **D**.

Event **B** ___IS NOT___ independent of Event **D**.

(b) $P(C \mid B) = 0.221$. Among Canadian taxfilers who were in the 9th income decile in 2007, 22.1 percent of them moved up to the 10th (richest) income decile five years later, which is a large probability. Taxfilers in the 9th decile had by far the biggest chance of moving into the top income decile at 22%. Those below the 6th decile had less than a 2% chance of moving to the top decile and even those in the 8th decile in 2007 had only a 8.8% chance. Alternatively, you could compare the likely destination in 2012 of someone in the 9th decile in 2007. While the most likely outcome is that they would remain in the 9th decile (a 32.5% chance), the second most likely outcome is that they would move up to the 10th decile (a 22.1% chance).

(c) Define the Binomial random variable X as a count of those below the 8th income decile in 2012.

$$P(X > 2 \mid n = 20, p = 0.159) = 1 - P(X = 0) - P(X = 1)$$

$$P(X = 0) = \frac{20!}{0!\,(20-0)!} 0.159^0 (1 - 0.159)^{20-0} = 0.0313$$

$$P(X = 1) = \frac{20!}{1!\,(20-1)!} 0.159^1 (1 - 0.159)^{20-1} = 0.1185$$

$$P(X = 2) = \frac{20!}{2!\,(20-2)!} 0.159^2 (1 - 0.159)^{20-2} = 0.2128$$

$$P(X > 2) = 1 - 0.0313 - 0.1185 - 0.2128 = 0.6374$$

(5) $V[a + bX_1 + cX_2] = b^2 V[X_1] + c^2 V[X_2] + 2bc * SD[X_1] * SD[X_2] * CORR[X_1, X_2]$. We know that $V[X_{2017} - X_{2007}] = 5154.192^2$ from the title of the third histogram. Further, we know that $SD[X_{2007}] = 21110.400$ and $SD[X_{2017}] = 20723.424$ from the titles of the first and second histograms. Hence,

$$5154.192^2 = (1)^2 20723.424^2 + (-1)^2 21110.400^2 + 2(1)(-1) * 20723.424 * 21110.400 * CORR[X_{2007}, X_{2017}]$$

Hence, $CORR[X_{2007}, X_{2017}] = \frac{20723.424^2 + 21110.400^2 - 5154.192^2}{2 * 20723.424 * 21110.400} = 0.9698 \approx 0.97$.

(6) (a) In a sample of 6,844 California households, the average 2009 electricity usage for those households using no natural gas is 28.3 MMBTUs. Households that did use natural gas on average used about 1.8 MMBTUs less electricity, hence using an average of 26.5 MMBTUs. In other words, households using natural gas on average used approximately 5.6% less electricity, which makes sense as they may substitute gas for electricity for items like stoves, ovens, water heaters, and home heating systems.

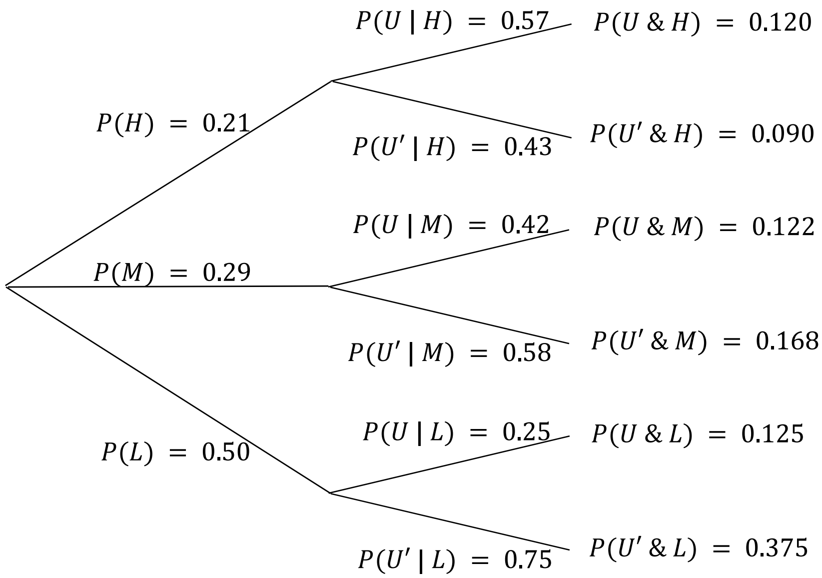
(b) The R-squared is slightly bigger for Figure 4 because it excludes the 27 outliers that lie very far below the middle of the OLS line. [It only makes a slight difference because the sample size is so huge that even 27 outliers do not have very much influence.] The R-squared measures how well the OLS line fits the data and it fits a bit worse when there are outliers like these. The s_e measures scatter (the opposite of fit) and it will be bigger for Figure 3 because of the extra scatter induced by the 27 outliers.

(c) In a sample of 5,940 California households in 2009, which excludes outliers that use very little natural gas, on average households that use 10 percent more electricity use approximately 3.5 percent more natural gas.

(7) (a) When $PM_{2.5}$ exposure is 1 standard deviation higher, the dementia rate is <u>0.63 standard deviations</u> higher on average.

(b) For 80 year olds, for states with a 10 year annual average $PM_{2.5}$ exposure that is 1 microgram per cubic meter ($\mu g/m^3$) higher, the observed state dementia rate is on average 1.1 percentage points higher. The intercept (i.e. constant term) of 5.82 has no interpretation because zero pollution exposure is far outside the range of the data: no U.S. states have perfectly pure air.

(8)



$$P(H \mid U) = ?$$

$$P(H \mid U) = \frac{P(U \,\&\, H)}{P(U)}$$

$$P(U) = P(U \,\&\, H) + P(U \,\&\, M) + P(U \,\&\, L)$$

$$P(U) = 0.120 + 0.122 + 0.125 = 0.367$$

$$P(H \mid U) = \frac{P(U \,\&\, H)}{P(U)} = \frac{0.120}{0.367} = 0.327$$