ECO220Y, Term Test #2: SOLUTIONS

November 17, 2017, 9:10 – 11:00 am

(1) (a) Two points determine a line. When y = 10, x ≈ 26 and when y = 15, x ≈ 54. Using the point slope formula, $(y - y_1) = m(x - x_1)$, where $m = \frac{15-10}{54-26} \approx 0.18$, gives $(y - 10) = 0.18(x - 26)$. Rewrite as: $y = 5.4 + 0.18x$. (Alternatively, you can read the intercept from the provided figure and just do the m calculation to obtain the slope.)

Equation for the OLS line: pct_police_fatal_shoot-hat = 5.4 + 0.18*int_hom_per_100000.

The slope means that in countries with an extra 10 intentional homicides per 100,000 people, on average we observe an extra 1.8 percentage points of the overall firearm homicides being fatal shootings committed by police. [Note that we *cannot* give a causal interpretation of this OLS line obtained from observational data.]

(b) rate_police_fatal_shoot-hat = 0.054 + 0.0018*int_hom_per_100000

(c) pct_police_fatal_shoot-hat = 5.4 + 1.8*int_hom_per_10000

(2) This is a grouped data question. We must use approximation techniques by selecting a representative value for each possible iClicker response. A sensible choice is the midpoint.

iClicker Response	Total	Fraction	Midpoint	Fraction*Midpoint	Fraction*(Midpoint - Mean)^2
(A) 0, 1 or 2	73	0.1730	1	0.1730	3.3768
(B) 3 or 4	54	0.1280	3.5	0.4479	0.4709
(C) 5 or 6	137	0.3246	5.5	1.7855	0.0022
(D) 7 or 8	115	0.2725	7.5	2.0438	1.1810
(E) 9 or 10	43	0.1019	9.5	0.9680	1.6977
Total	422		**SUM:**	**5.4182**	**6.7285**

Hence, the mean is approximately 5.42 (which is not particularly happy and much lower than overall average reply for Canada). The variance is approximately 6.73. However, the question asked for the s.d., which is approximately 2.6. There is quite a bit of variation in happiness across people in our class.

(3) $V[X_1 + X_2 + \cdots + X_{10}] = V[X_1] + V[X_2] + \cdots + V[X_{10}] = 10 * 28^2 = 7840$ (Note: We can add variances in this case because it is reasonable to assume that the random variables recording the donation for each conversation are independent.)

$SD[X_1 + X_2 + \cdots + X_{10}] = \sqrt{7840} = \$88.54$

(4) Starting with the Correlation: $r = \frac{s_{xy}}{s_x s_y} = \frac{cov(x,y)}{sd(x)*sd(y)} = \frac{0.00006553}{\sqrt{0.00070829}*\sqrt{0.00028613}} = \frac{0.00006553}{0.026614*0.016915} = 0.146$

The Rank Correlation: It cannot be computed with the given information because we are given a variance-covariance matrix for the growth rate variables, not the variables *ranking* the growth rates, which is what is needed for the rank correlation.

The Regression Coefficient: $b_1 = r\frac{s_y}{s_x} = \frac{s_{xy}}{s_x^2} = \frac{cov(x,y)}{var(x)} = \frac{0.00006553}{0.00070829} = 0.093$ (Note: It is very important that you correctly identified the x variable in this calculation.)

The R-squared: $R^2 = (r)^2 = (0.146)^2 = 0.021$

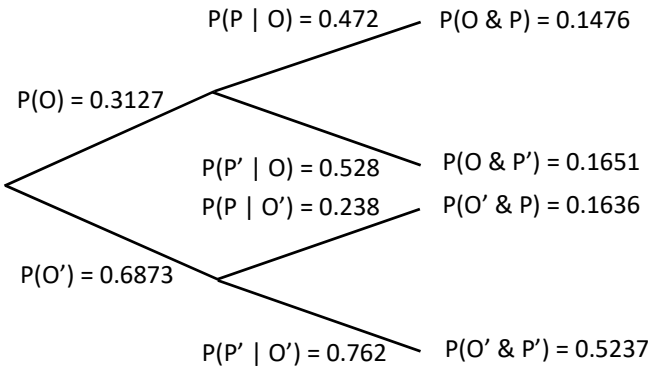
(5) (a) $P(V \text{ and } P') = P(P' \mid V) * P(V) = (1 - 0.386) * \left(\frac{529{,}845}{6{,}858{,}075}\right) = 0.614 * 0.077259 = 0.0474$

(Note: You *absolutely cannot* use the multiplication rule for independent events because the event of living in Vancouver is *not* independent of the event of living with parents. Clearly there is a relationship between where young adults live and whether or not they live with their parents: that is the whole point of Table 1.)

(b) $P(M \text{ or } P) = P(M) + P(P) - P(M \text{ and } P) = \frac{821{,}555}{6{,}858{,}075} + 0.347 - P(P \mid M) * P(M) = 0.119794 + 0.347 - 0.331 * 0.119794 = 0.119794 + 0.347 - 0.039652 = 0.427$

(Note: You *absolutely cannot* use the addition rule for mutually exclusive events because the event of living in Montréal is *not* mutually exclusive (i.e. disjoint) from the event of living with parents.)

(c) This is a reversing the conditioning problem. Define O as the event of living in Oshawa and O' as the event of living in Québec (not Oshawa). Define P as the event of living with your parents. We must find P(O | P) = ?.



P(O) = (70,945)/(70,945+155,900) = 0.3127

P(O | P) = P(O & P)/P(P)

P(P) = P(O & P) + P(O' & P) = 0.1476 + 0.1636 = 0.3112

P(O | P) = P(O & P)/P(P) = 0.1476/0.3112 = 0.474

(d) $n = 8$; $p = 0.285$; Binomial probability: $P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$ for $x = 0,1,2, \dots, n$

$P(0) = \frac{8!}{0!(8-0)!} 0.285^0 (1 - 0.285)^{8-0} = 0.0683;$

$P(1) = \frac{8!}{1!(8-1)!} 0.285^1 (1 - 0.285)^{8-1} = 0.2178;$

$P(2) = \frac{8!}{2!(8-2)!} 0.285^2 (1 - 0.285)^{8-2} = 0.3039;$

$P(X > 2) = 1 - P(X \leq 2) = 1 - 0.0683 - 0.2178 - 0.3039 = 0.41$

(6) (a) The SST, which is $SST = \sum_{i=1}^{n} (y_i - \bar{Y})^2$ and measures the amount of variation in the number of mass shooters between 1966 and 2012 across the included countries, would be MUCH smaller if the U.S. were excluded because the U.S. is an extreme outliers with far more mass shooters (90) than any other country (the next highest is 18).

(b) The R^2, which measures how much of the variation across countries in the number of mass shooters between 1966 and 2012 can be explained by variation in the number of guns (i.e. how well an OLS line fits the scatter diagram), would be MUCH smaller (close to zero) if the U.S. were excluded. Without the U.S. there is no visible relationship at all between these variables. The U.S. is single-handedly causing an apparent positive relationship between these variables because it has an extraordinarily high number of mass shootings and an extraordinarily high number of guns.

(c) In countries where the total number of guns is 10 percent higher, on average we observe the total number of mass shooters between 1966 and 2012 is (approximately) 3.9 percent higher. [Note: While the log-log functional form often brings outliers into fold with the other observations, in this case, the U.S. is such an extreme outlier (both in x and in y and in the pair) that it is still an outlier even in the log-log form (although not as extreme).]

(d) The point of Figure 2 is that it controls for population size differences across countries to allow direct comparisons of smaller countries like Canada with bigger countries like the U.S. In Canada there are about 30 guns for every 100 hundred people and about 9 mass shooters for every 100 million people (i.e. 0.9 per every 10 million people) whereas in the U.S. there are about 88 guns for every 100 people – nearly a gun for every man, woman, and child in the country! – and about 28 mass shooters per 100 million people. Hence, the U.S. has a higher concentration of guns and a higher rate of people becoming mass shooters.

(e) Without Yemen and the U.S. there does not appear to be any relationship between the concentration of guns and the rate at which people become mass shooters.