ECO220Y1Y, Test #2, Prof. Murdock: SOLUTIONS

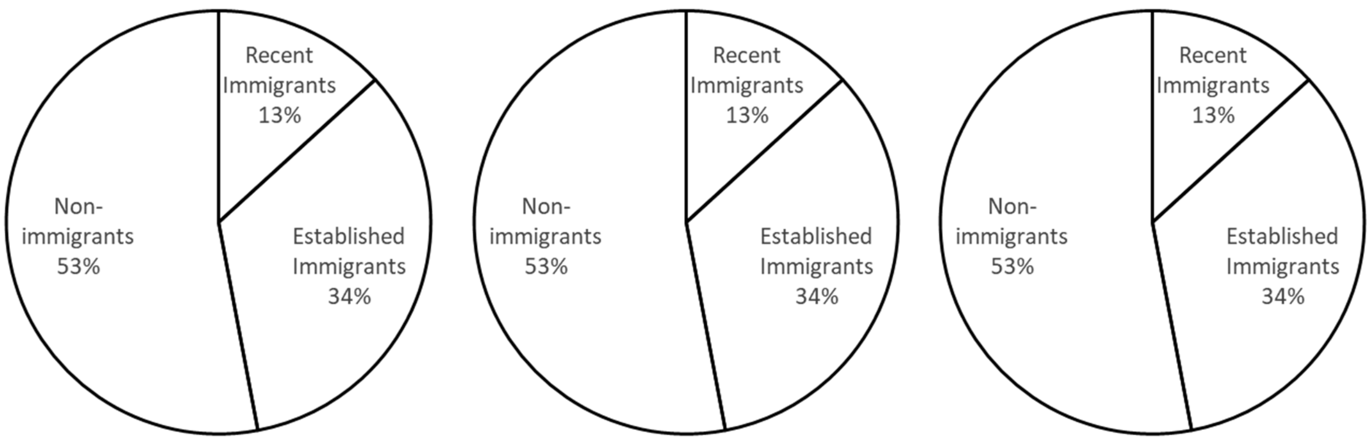
November 16, 2018, 9:10 – 11:00 am

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

(1) (a) $P(L) = 0.51 = \frac{1{,}368{,}000}{1{,}368{,}000+757{,}000+568{,}000}$ [Note: The probability is *not* 0.48. That is the probability that a randomly selected *census tract* is a low income neighbourhood. The question asked about a randomly selected person. Not all census tracts have the same number of people so these two probabilities do not need to be equal. The difference between 0.51 and 0.48 is a *real difference* and does *not* simply arise because of rounding.]

(b) $P(R \,\&\, H) = P(R \mid H) * P(H) = 0.08 * \frac{568{,}000}{1{,}368{,}000+757{,}000+568{,}000} = 0.08 * 0.21 = 0.017$

(c) **Low Income Neighbourhoods** **Middle Income Neighbourhoods** **High Income Neighbourhoods**



[From the article titled "Toronto is segregated by race and income," the provided figure is titled "Toronto's Segregated Immigrant Population." The figure *shows* the issue of segregation: immigrants are segregated into lower income neighborhoods. In other words, immigration status and neighbourhood income are *not independent*. If there were no segregation then these would be independent. In formal notation, this means that: $P(N \mid L) = P(N)$, $P(E \mid L) = P(E)$, $P(R \mid L) = P(R)$. Similarly, for middle income neighbourhoods: $P(N \mid M) = P(N)$, $P(E \mid M) = P(E)$, $P(R \mid M) = P(R)$. Finally, for high income neighbourhoods: $P(N \mid H) = P(N)$, $P(E \mid H) = P(E)$, $P(R \mid H) = P(R)$. In other words, the conditional probabilities would be the same as the marginal probabilities. From the *Supplement*, right below the pie charts, you can simply copy the marginal probabilities. You could compute more precise numbers: $P(N) = \frac{1{,}425{,}700}{1{,}425{,}700+910{,}300+355{,}700} = 0.530$; $P(E) = \frac{910{,}300}{1{,}425{,}700+910{,}300+355{,}700} = 0.338$; $P(R) = \frac{355{,}700}{1{,}425{,}700+910{,}300+355{,}700} = 0.132$. Showing what the pie charts would look like, as done above, expresses the idea of independence most clearly.]

(d) Absolutely not. $P(M \mid H) = 0$, which is NOT equal to $P(M) = \frac{757{,}000}{1{,}368{,}000+757{,}000+568{,}000} = 0.28$, whereas the definition of independence requires equality. [Note: Alternatively, and equally correct, you can say $P(H \mid M) = 0$, which is NOT equal to $P(H) = \frac{568{,}000}{1{,}368{,}000+757{,}000+568{,}000} = 0.21$.] The Events **M** and **H** are *mutually exclusive (disjoint)* – living in a middle-income neighborhood means you definitely do not live in a high-income neighborhood – and hence they cannot be independent.

(e) Absolutely not. $P(N \mid L) = 0.42$, which is not equal to $P(N) = 0.53$: if they were independent, these two probabilities would be equal. The fact that 0.42 is much lower than 0.53 means that low income neighborhoods have less than their share of non-immigrants because the non-immigrants disproportionally tend to live in more wealthy neighborhoods. This is describing segregation (the opposite of random allocation across neighborhoods: i.e. independence) by immigration status.

(2) (a) The intercept for the ***Line for 2013-2015*** is <u>well below $20,000</u>. [To get an estimate (which is not requested by the question), use the point-slope formula: $y - y_1 = m(x - x_1)$. Slope is $m = \frac{\Delta y}{\Delta x} \approx \frac{80{,}000-60{,}000}{205-145} = 330$. Find any point (approximately) on the line via visual inspection using the *Supplement*. For example, (102, 40,000), obtain $y - 40{,}000 = 330(x - 102)$ to yield an intercept of around $a = 6{,}340$ $(y = 6{,}340 + 330x)$.]

(b) From 2011/12 through 2017/18, for the relationship between base MSRP and battery range of electric vehicles, the intercept is <u>dramatically declining</u> and the slope is <u>staying about constant</u>.

(c) $\hat{y} = -34{,}580 + 205x$

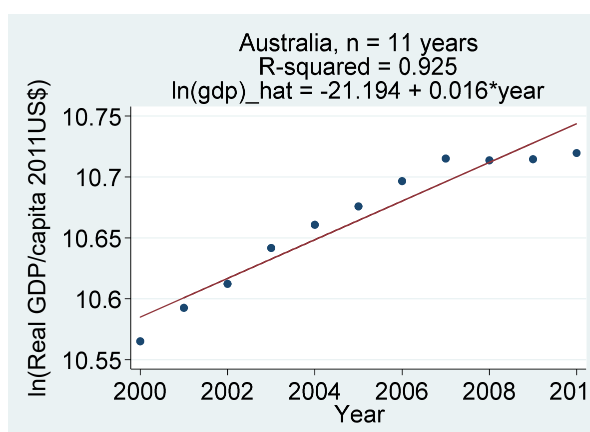
(d) $\hat{y} = -44{,}906 + 429x$

(3) (a) In the United States in 2010, county groups with a one unit higher poverty ranking – for example, going from rank 65 to 66, which means higher poverty – on average have income per capita levels that are $116 lower. The negative sign is not surprising: we would expect counties with higher rates of poverty to tend to have lower per capita income levels.

(b) Dropping the outlier would lead to a substantial drop in the value of the s_e, which measures the amount of scatter about the OLS line, because the outlier has a very large residual (is far from the OLS line) and single-handedly increases the amount of scatter. In contrast, the value of the R^2 would increase because it measures the strength of the relationship between per capita income and the poverty ranking and the outlier is bucking the clear negative relationship: it is a high poverty rank county group with a very *high* per capita income level (highly irregular and pointing to a positive relationship).

(4) (a) In Australia from 1960-1970, real GDP per capita on average increased annually by $644 (2011 US$).

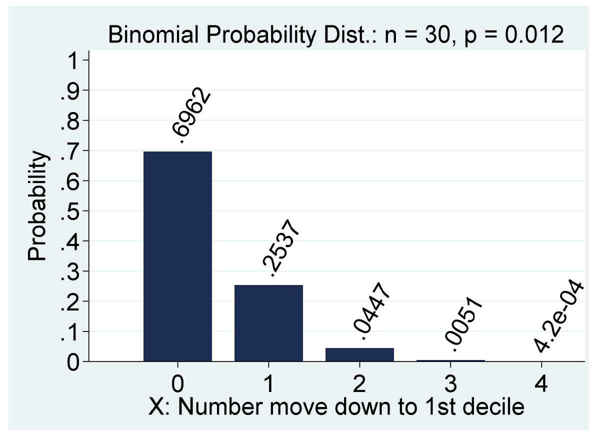
(b) Recalling that b measures the annual growth *rate*, it will be approximately $0.016 = \frac{673}{43000}$.
[For your reference, the graph to the right illustrates the answer.]



(c) In terms of growth *levels*, Australia's real GDP per capita (in 2011 USD) grew more quickly in the 2000s than the 1960s: \$673 annually in the 2000s versus \$644 annually in the 1960s. However, it would be much better to compare growth *rates*. Australia had much lower levels of GDP per capita (around \$19,000) in the 1960s so that \$644 annual growth was quite fast: a 3.4% annual growth rate $\left(0.034 = \frac{644}{19000}\right)$. In the 2000s, Australia had much higher levels of GDP per capita (around \$43,000) so that \$673 is far less impressive: only a 1.6% annual growth rate $\left(0.016 = \frac{673}{43000}\right)$. Hence, Australia's GDP per capita actually grew more slowly in the 2000s compared to the 1960s, which is common for developed countries.

(5) (a) Answering requires finding the Binomial distribution for $n = 30$ and $p = 0.012$. Use the Binomial probability formula:

$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$



$$P(X = 0) = \frac{30!}{0!\,(30-0)!} 0.012^0 (0.988)^{30-0} = 0.012^0 (0.988)^{30} = 0.6962$$

$$P(X = 1) = \frac{30!}{1!\,(30-1)!} 0.012^1 (0.988)^{30-1} = 30 * 0.012^1 (0.988)^{29} = 0.2537$$

$$P(X = 2) = \frac{30!}{2!\,(30-2)!} 0.012^2 (0.988)^{30-2} = 15 * 29 * 0.012^2 (0.988)^{28} = 0.0447$$

$$P(X = 3) = \frac{30!}{3!\,(30-3)!} 0.012^3 (0.988)^{30-3} = 10 * 29 * 14 * 0.012^3 (0.988)^{27} = 0.0051$$

$$P(X = 4) = \frac{30!}{4!\,(30-4)!} 0.012^4 (0.988)^{30-4} = 5 * 29 * 7 * 27 * 0.012^4 (0.988)^{26} = 0.00042$$

(b) A Binomial random variable X with $n = 100$ and $p = 0.988$ has a mean of 98.8 $(E[X] = np = 100 * 0.988)$ and a s.d. of 1.09 $\left(SD[X] = \sqrt{np(1-p)} = \sqrt{100 * 0.988(0.012)}\right)$. The shape will be (extremely) negatively skewed: there is a very high mean and we cannot have values of X above 100 so there can only be a left tail.

(6) Law of variance: $V[a + bX_1 + cX_2] = b^2 V[X_1] + c^2 V[X_2] + 2bc * SD[X_1] * SD[X_2] * CORR[X_1, X_2]$.

We're given: $V[X_{1990-2010} - X_{1970-1990}] = 0.0268^2$, $SD[X_{1990-2010}] = 0.0167$ and $SD[X_{1970-1990}] = 0.0260$.

Hence, $0.0268^2 = (1)^2 0.0167^2 + (-1)^2 0.0260^2 + 2(1)(-1) * 0.0167 * 0.0260 * CORR[X_{1990-2010}, X_{1970-1990}]$

Solving for the correlation, we obtain a positive correlation of 0.27 $\left(= \frac{(0.0268^2 - 0.0167^2 - 0.0260^2)}{-2*0.0167*0.0260}\right)$.