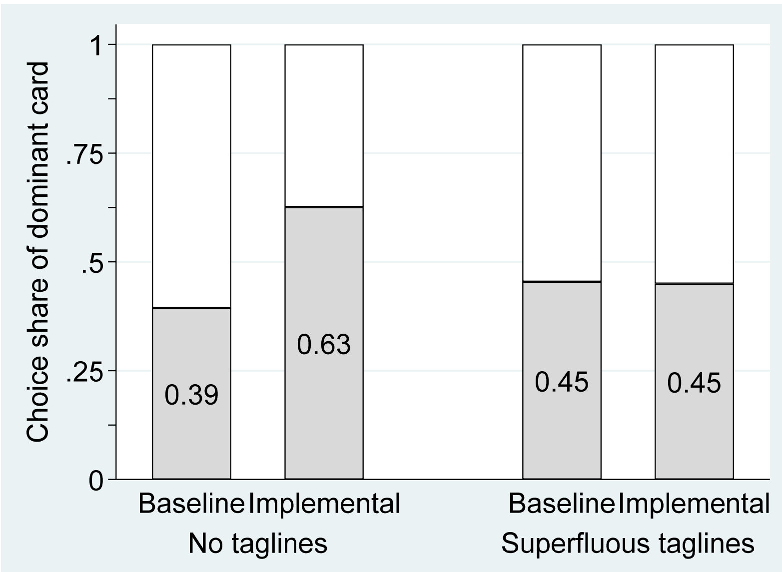
ECO220Y1Y, Test #1, Prof. Murdock: SOLUTIONS

October 5, 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)



13/33=0.39;

77/123=0.63;

15/33=0.45;

63/140=0.45

(b) Those two values – 0.45 and 0.45, which are sample proportions – are NOT equally affected by sampling error. Unlike the original study where participants were evenly divided into the four groups, the new study had far fewer participants watch the baseline video (66) versus the implemental video (263). When we look at the subset that saw the superfluous taglines, the sample size for the third bar is only n = 33 versus n = 140 for the fourth bar. Hence, the first 0.45 is more subject to sampling error because small sample sizes mean we can have a large discrepancy between the sample statistic and the population parameter. [Note: It is NOT valid to support your argument by comparing these results to those in Figure 6 because those are *not* parameters, but rather another set of sample statistics subject to sampling error. Also, the new study samples from a different population.]

(2) $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}} =$

$\sqrt{0.3352(0-.8454)^2 + 0.5034(1-.8454)^2 + 0.1428(2-.8454)^2 + 0.0180(3-.8454)^2 + 0.0006(4-.8454)^2}$
=0.729

[Given that the question specifies a very large sample, we can safely ignore the degrees of freedom correction. Anyway, without knowing the sample size we could not do it.]

(3) (a) Cross-sectional data would be the subset of observations for the month of October 2008 [or any other *one specific month* from October 2008 through December 2011]: this subset would have 36 observations, where the unit of observation is a state. Time series data would be the subset of observations for the state of Florida [or any other *one specific state* listed in Footnote 1 of the *Supplement*]: this subset would have 39 observations, where the unit of observation is month [month-year]. [Note: Discussing a subset of variables in answering is not only irrelevant – the type of data does not depend on the number of variables – but also confusing. There is no reason to say that you would include only the flu index or only the sunlight variable when answering.]

(b) *The histogram that is consistent with the boxplot is* Histogram #1. *We can rule out the other two because:*

- Histogram #2 can be ruled out because it says the 25th percentile is definitely less than 23,000 and the box plot shows that the 25th percentile is about 23,125.

  [Alternatively, this histogram says the 75th percentile is definitely greater than 26,000 but the box plot shows that the 75th percentile is about 25,300.] [We *cannot* rule out Histogram #2 based on the median, the outside values, or the whiskers. The only definitive reasons are the 25th and 75th percentiles.]

- Histogram #3 can be ruled out because it says that there are no observations between the lower outside values and 22,000: if that were true then the box plot's lower whisker should not extend to about 21,000.

  [Alternatively, this histogram says that there no observations between 27,000 and the upper outside values, which contradicts the box plot's upper whisker that extends to about 27,600.] [We *cannot* rule out Histogram #3 based on the median, the outside values, or the 25th and 75th percentiles. The only definitive reasons are the two whiskers.]

(c) Height of first bar would be about 102/108 = 0.944. We know that at least 75% of the observations are exactly equal to 1 and it looks like there are only about 6 observations above 1.9 (108 – 6 = 102). [Note: Of course, some observations could be directly on top of each other for the outlier dots. However, we can be 100% sure that the height of the first bar is at least 84/108 = 0.78 (which is an unlikely scenario where there are many dots on top of each other for the six dots shown for flu indices from 2 to 10) and less than or equal to 102/108 = 0.944.]

(4) $\mu \approx 0.142 * 24.5 + 0.203 * 34.5 + 0.177 * 44.5 + 0.167 * 54.5 + 0.147 * 64.5 + 0.161 * 80 = 49.8$

Hence, the mean age of all U of T alumni is about 50 years old: this is μ, which is a population parameter.

[These are grouped data. We must use the last column because the question asks about *all* U of T alumni (i.e. it asks about the population). Picking the mid-point age for each group (a reasonable approximation) and 80 for the oldest group (this is the hardest to pick: any reasonable estimate of a typical age for the 70+ group is acceptable).]

(5) (a) The height of the bar for the 0 to 0.5 bin is about 0.72, which means that about 36 percent of the sample had a tax rate change in that range (36 = 100*0.72*0.5).

(b) Two ways to solve: (1) Using "the mean increase is 0.9 percentage points (or 5 percent)," we solve for x in $\frac{(x+0.9)}{x} = 1.05$ to obtain an initial local business tax rate of about 18% OR (2) Using "the seventy-fifth percentile of the tax increase distribution is equal to 1.1 percentage points (6 percent)," we solve for x in $\frac{(x+1.1)}{x} = 1.06$ to obtain an initial local business tax rate of about 18%. [Note that these are approximate because the excerpt rounds the numbers off.]

(6) (a) Two points determine a line: (2010, 14,850) and (2012, 15,600). (The y-values are approximations.)

$Slope \approx \frac{(15{,}600-14{,}850)}{(2012-2010)} = 375$

Point-slope formula: $y - 14{,}850 = 375(x - 2010)$, giving the approximate equation for the line:

GDP_per_capita = -738,900 + 375*year

(b) $Slope\ for\ 1960\ to\ 1970 \approx \frac{(8{,}800-6{,}325)}{(1970-1960)} = 247.5$

This means that from 1960 to 1970 in Mexico, real GDP per capita measured in US dollars was increasing by about $250 each year. During the two-year period from 2010 to 2012, absolute growth levels were faster: $375 per year. [Of course, Mexico was significantly poorer in the 1960s – real GDP per capita levels were about half of what they were in the early 2010s – so comparing $250 directly with $375 is somewhat misleading. Instead, economists typically measure GDP growth in *percent* terms.]

(7) (a)

Range: 115 $\mu g/m^3$ = (155 – 40)

Interquartile Range: 27 $\mu g/m^3$ = (99 – 72)

Coefficient of Variation: 0.24 (unit-free) = 20.86533/87.31765

(b)

- About 68.3% of the cities should have PM10 concentrations between 66.45 $\mu g/m^3$ and 108.18 $\mu g/m^3$.
- About 95.4% of the cities should have PM10 concentrations between 45.59 $\mu g/m^3$ and 129.05 $\mu g/m^3$.
- About 99.7% of the cities should have PM10 concentrations between 24.72 $\mu g/m^3$ and 149.91 $\mu g/m^3$.

(c) The STATA output is entirely consistent with all three parts of the Empirical Rule. Somewhere between 50% and 80% fall between 66.45 μ/m^3 and 108.18 μ/m^3 (which includes 68.3%), exactly 95.3% (=81/85) fall between 45.59 μ/m^3 and 129.05 μ/m^3 (which is about 95.4%), and exactly 98.8% (=84/85) fall between 24.72 μ/m^3 and 149.91 μ/m^3 (which is about 99.7%).

(8) (a) In the 2017/18 recruiting season, the share of hiring employers on the American Economic Association's Job Openings for Economists platform that were economics departments [among employers that are economics departments, tech companies, business schools, and policy schools] dropped by 10 percentage points compared to the 2014/15 recruiting season [i.e. dropped from 57.1% down to 47.4%]. [Note: There are other non-trivial types of employers of PhD economists, including banks, consulting firms, government agencies, and international organizations. However, the table did not include a note and the excerpt did not explain this so we accept answers that imply that the four employer types in the table constitute all of the major types of hiring employers of PhD economists on that platform.]

(b) In the 2017/18 recruiting season for hiring PhD economics on the American Economic Association's Job Openings for Economists platform, the share of hiring employers [among employers that are economics departments, tech companies, business schools, and policy schools] that are tech companies is about 5%. That number is misleading because it underestimates the importance of tech companies in this job market because each tech company is typically hiring a lot of people each year (more than the other categories of employers) so their effective share (thinking in terms of job openings for PhD economics) is far bigger than 5%.