ECO220Y1Y, Test #1, Prof. Murdock SOLUTIONS

October 29, 2021, 9:10 – 11:00 am

(1) (a) The mean is 6.73, the median is 7, the standard deviation is 1.64, the range is 9, the interquartile range is 2, the coefficient of variation is 0.24, and the 5th percentile is 4.

(b) The fastest way to find the fraction answering eight or less is to compute: (1 – 0.08) = 0.92 because it looks like about eight percent answered either 9 or 10. This is about 92% answering eight or less.

(c) The 410 people choosing to answer the survey are *not* a random sample of the 630 people in our course. Further, 410 is a large sample size. Hence, sampling error will not have any significant impact on either summary. In contrast, non-sampling errors are likely: most notably non-response bias. It may be that those choosing to reply are systematically different: perhaps they are somewhat more competitive as they are trying to keep up and participate to do well in the course.

(2) (a) In this frequency histogram the height of the tallest bar is 17. The total sample size is 65 = (3 + 13 + 13 + 17 + 7 + 11 + 1). Hence, the height would be 0.26 = 17/65 in a relative frequency histogram.

(b) This frequency histogram is roughly Normal (Bell shaped) and shows a stunning amount of bribery to obtain government jobs in this unnamed large developing country. The amount that successful hires pay in bribes ranges from a *minimum* of 5 months salary to a maximum of over 30 months salary. Over a fifth pay more than *two years* salary in a bribe to secure a job.

(3) In these California port areas in 2010, the last two bars focus on the ten percent of neighborhoods where the air quality is worst: the tenth decile of PM2.5 air pollution. Approximately 4 percent of the white population lives in locations with the worst air quality compared to about 18 percent of the black population. The share of the black population living in these locations is more than *four times higher* and 14 percentage points higher than the share of the white population, which indicates a large racial disparity in exposure to poor air quality.

(4) (a)

city_id	year	pm25	gdp_bil	pop_mil
TOR	2010	#	#	#
VAN	2010	#	#	#
…		…	…	…
CAL	2010	#	#	#

In my made-up data for the described project, there are five variables, and the number of observations is 20. The unit of observation is a Canadian city.

(b)

year	city_name	pm25	gdp_bil	pop_mil
2013	Toronto	#	#	#
2014	Toronto	#	#	#
…		…	…	…
2017	Toronto	#	#	#

In my made-up data for the described project, there are five variables, and the number of observations is 5. The unit of observation is a year.

(c)

year	city_name	pm25	gdp_bil	pop_mil
2011	Toronto	#	#	#
2011	Vancouver	#	#	#
…		…	…	…
2016	Calgary	#	#	#

In my made-up data for the described project, there are five variables, and the number of observations is 120. The unit of observation is a city-year (20 cities x 6 years).

(5) (a) If #A = 218, then #B = 211 (=429-218), #C = 199 (=417-218), and #D = 215 (=843-218-211-199).

(b)

- Suppose #A were 100. Changing #A to 125 would mean a weaker negative correlation.
- Suppose #C were 75. Changing #C to 50 would mean a stronger positive correlation.

(6) (a) 27.4

(b) In the full sample of 3,082 Dutch people used in this study, the standard deviation of self-reported monthly income is 1,766 Euros. There is a lot of income variation in this study: relative to the mean of 2,480 Euros, the standard deviation is extremely large.

(c) There is virtually no difference in self-reported competitiveness between the experimental sample and the full sample: both the mean and standard deviation are nearly the same with mean self-reported competitiveness, on a zero-to-ten scale, being about 6.2 and the standard deviation being about 2. Given that the experimental sample is a random subset, this is not surprising and means that it is representative.

(7)

1) Create a variable measuring the GDP per capita by dividing GDP by population
2) Take the natural logarithm of the new GDP per capita variable
3) For each of the 142 countries, run a separate regression using the 11 years of data from 1980 to 1990 where the y-variable is the natural log of GDP per capita and the x-variable is the year (i.e. 142 simple regressions)
    a. Record the slope estimate for each regression: this is each county's estimated growth rate in that decade (1980s)
4) For each of the 142 countries, run a separate regression using the 11 years of data from 2000 to 2010 where the y-variable is the natural log of GDP per capita and the x-variable is the year (i.e. 142 simple regressions)
    a. Record the slope estimate for each regression: this is each county's estimated growth rate in that decade (2000s)
5) Run one regression with 142 observations where the y-variable is the growth rate in the 2000s and the x-variable is the growth rate in the 1980s

(8) (a) Among these graduates from undergraduate programs at Boston University's business school (Questrom) from the years 2013 through 2017, females on average accepted a job that pays $6,719 less in salary compared to males, which is a big difference.

(b) Less than 3 percent of the variation in starting salaries is explained by the person's identified sex, meaning 97 percent is related to other factors. This regression does not control for important factors like the type of job, how well the student did at Questrom, which program they were in, and so on, which is consistent with the pathetic R-squared.

(c) Further, compared to males, females' starting salaries are 9.7 percent lower.