

ECO220Y1Y, Test #1, Prof. Murdock SOLUTIONS

(1)

company	quarter	year	revenue_billions	net_income_billions
Alphabet	2	2022	69.69	16.00
Alphabet	3	2022	69.09	13.91
Alphabet	4	2022	76.05	13.62
Alphabet	1	2023	69.79	15.05
Apple	2	2022	82.96	19.44
Apple	3	2022	90.15	20.72
Apple	4	2022	117.15	30.00
Apple	1	2023	94.84	24.16
Microsoft	2	2022	51.87	16.74
Microsoft	3	2022	50.12	17.56
Microsoft	4	2022	52.75	16.42
Microsoft	1	2023	52.86	18.30

The data above have 5 variables, and of these, 3 are identifier variables. The data above have 12 observations. The unit of observation is a company in a quarter. A subset of these data with only Apple would be time series. [Note: It can be argued that quarter and year are not identifier variables as they measure time, so 1 is also an acceptable answer. Also, some may combine year and quarter into one time variable, which affects the total number of variables. The blanks filled in must match with correctly constructed data above it.]

(2) (a) In 2022 for 111 countries worldwide, those with a healthy life expectancy one year higher (e.g. 72 years vs. 71), on average have a mean Cantril ladder score, which measures happiness on a scale from 0 to 10, that is 0.17 higher, which is a notable amount.

(b) In countries where the healthy life expectancy is one standard deviation higher, we observe happiness that is 0.75 standard deviations higher on average. [Note: The order of variables can be switched.]

(c) Across 114 countries in 2022, members of the OECD have mean national happiness, as measured by the Cantril ladder score on a scale from 0 to 10, that is on average 1.71 higher than non-OECD countries where mean happiness is 5.04, which is a very large difference.

(d) In 2022 for 108 countries worldwide, those with GDP per capita that is 10% higher, on average have a mean Cantril ladder score, which measures happiness, that is 1.6% higher.

(e) From 2006 through 2022 in the country of Nicaragua, happiness as measured by the Cantril ladder on average grew at an impressive rate of 1.75% annually. The R-squared of 0.76 is quite high which means that this growth is quite steady during this period: not too much scatter over those years.

(3) (a) In the United States, compared to 1990, among college-educated females aged 15 to 49 years, the percent employed in an occupation that is among the top 25% of all occupations for being “age-friendly,” is a whopping 26.5 percentage points higher: 25.7% in 1990 versus 52.2% in 2020. These relatively *young* and well-educated females are *much* more likely to be in an age-friendly occupation at the end of this 30-year period. In contrast, among relatively *old* males without a college education, the increase is only 5.6 percentage points, which is the smallest increase across education levels, sex, and age. Overall, Table 5 is about the increasing fraction of people employed in age-friendly jobs from 1990 to 2020 and interesting differences about that increase across demographic groups: (1) females are benefiting more than males, (2) more educated people are benefitting more than less educated people, and (3) relatively young people are generally benefiting somewhat less than older people, *but not much less*. There is an exception for the least educated males where the young benefit *more* than the old.

(b) In the United States from 1990 to 2020, the share of people employed in occupations that are among the worst 10% of all occupations for being age-friendly, dropped by about 6 percentage points from about 9% in 1990 to about 3% in 2020. This is a very large drop of 66% as workers move away from occupations that are extremely age unfriendly.

(4) If #A is 120, then the value of #D is 20.

If #A is 200, then x and y have a negative correlation.

If #A is 280, then x and y have zero correlation.

(5)(a) When AI is turned on for this large sample of 3,127 cruises by Japanese taxi drivers, the distribution of minutes between jobs is highly positively skewed with most cruises being less than 20 minutes but some being nearly one hour.

(b) The bin width is about 1 minute and adding the height of the first three bars we get about $4.2 + 6.8 + 7 = 18$. Hence, the 15th percentile is about 3 minutes.

(c) These data are observational, and the excerpts explain that the explanatory variable – using AI – is endogenous. They discuss two important lurking/unobserved/confounding/omitted variables: the skill of the driver and the difficulty of finding customers (i.e. the demand for taxis). These lurking variables are correlated with both the *choice* to use AI and the dependent variable – how long it takes to find a customer. Further, the mean and median show that AI use is associated with *longer* times between jobs – less efficiency – which is the opposite of what we would expect and this is due to severe endogeneity bias. The less skilled drivers and all drivers in difficult situations use AI: it is *not* that AI is causing the longer time between jobs.

(6) (a)

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{253*1 + 321*2 + 799*3 + 1,191*4 + 929*5}{3,493} = \frac{12,701}{3,493} = 3.6361$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1} = \frac{253(1-3.6361)^2 + 321(2-3.6361)^2 + 799(3-3.6361)^2 + 1,191(4-3.6361)^2 + 929(5-3.6361)^2}{3,493-1} = 1.3822$$

$$s = \sqrt{1.382} = 1.18$$

(b) stay the same

(c) [3 pts] Table 3 reports **six** separate regressions, and of these, **five** are multiple regressions. These data are **observational**. Of the explanatory variables listed **four** are dummy variables.

(d) Among Swedish households in 2019, those with a household income that is 10% higher on average gave an answer that is 0.01 lower to the survey question asking how important the subsidy is for the decision to buy an E-bike on a scale from 1 to 5 where five is critical important. This is a tiny estimate. Further, the R-squared is only 0.003: only about 0.3% of variation in the answer to this survey question is explained by variation in income. Income is a terrible predictor of replies, contrary to what we may have expected.

(e) stay the same