ECO220Y1Y, Test #1, Prof. Murdock

May 29, 2023, 11:10 am – 1:00 pm

U of T E-MAIL:																			

@MAIL.UTORONTO.CA

SURNAME (FAMILY NAME):																			
GIVEN NAME (FIRST NAME):																			

UTORID: (e.g. LIHAO118)								

Instructions:

- You have 110 minutes. Keep these test papers and the *Supplement* closed and face up on your desk until the start of the test is announced. You must stay for a minimum of 60 minutes.
- You may use a **non-programmable calculator**.
- There are 6 questions (most with multiple parts) with varying point values worth a total of 95 points.
- This test includes these 8 pages plus the *Supplement*. The *Supplement* contains the aid sheets and readings, figures, tables, and other materials for some test questions. For each question referencing the *Supplement*, carefully review *all* materials. ***The Supplement will NOT be collected:*** write your answers on these test papers. When we announce the end of the test, hand these test papers to us (you keep the *Supplement*).
- Write your answers clearly, completely, and concisely in the designated space provided immediately after each question. An answer guide ends each question to let you know what is expected. For example, a quantitative analysis, a fully labelled graph, and/or sentences. Any answer guide asking for a quantitative analysis *always* automatically means that you must show your work and make your reasoning clear.
    - Anything requested by the question and/or the answer guide is required. Similarly, limit yourself to the answer guide. For example, if the answer guide does not request sentences, provide only what is requested (e.g. quantitative analysis).
    - Marking TAs are instructed to accept all reasonable rounding.
- ***Your entire answer must fit in the designated space provided immediately after each question.*** No extra space/pages are possible. You *cannot* use blank space for other questions, nor can you write answers on the *Supplement*. ***Write in PENCIL and use an ERASER as needed*** so that you can fit your final answer (including work and reasoning) in the appropriate space. We give more blank space than is needed for each answer (with typical handwriting) worth full marks. ***Follow the answer guides and avoid excessively long answers.***

(1) [10 pts] Consider some financials, in US dollars, for the quarters ending June 30, 2022, September 30, 2022, December 31, 2022, and March 31, 2023 (most recently). For these four quarters respectively, Alphabet's revenue is 69.69B and its net income is 16.00B, 69.09B and 13.91B, 76.05B and 13.62B, and 69.79B and 15.05B (most recently). Apple's revenue is 82.96B and its net income is 19.44B, 90.15B and 20.72B, 117.15B and 30.00B, and 94.84B and 24.16B (most recently). Microsoft's revenue is 51.87B and its net income is 16.74B, 50.12B and 17.56B, 52.75B and 16.42B, and 52.86B and 18.30B (most recently). Organize this as **panel data in long form**. Remember that data are comprised of variables and observations. Answer by showing the exact and complete data set & by filling in the blanks.

The data above have _____ [numeric value] variables, and of these, _____ [numeric value] are identifier variables.

The data above have _____ [numeric value] observations. The unit of observation is _____.

A subset of these data with only Apple would be _____ [cross sectional, time series, panel].

(2) See ***Supplement for Question (2)***: *OLS Regressions with the 2023 World Happiness Report Data.*

(a) [5 pts] In **Regression #1**, *interpret* 0.1677988. Answer with **ONE** precise sentence.

(b) [3 pts] *Continuing,* in **Regression #1**, *interpret* the square root of the R^2. Answer with **ONE** precise sentence.

(c) [5 pts] In **Regression #2**, *interpret* 5.043696 and 1.714530. Answer with **ONE** precise sentence.

(d) [5 pts] In **Regression #3**, *interpret* 0.1561908. Answer with **ONE** precise sentence.

(e) [6 pts] In **Regression #4**, *interpret* 0.0174692 and 0.7576. Answer with 2 sentences.

(3) See ***Supplement for Question (3)***: *The rise of age-friendly jobs.*

(a) [14 pts] In the last column of **Table 5**, *interpret* 26.5 and *interpret* 5.6. Include an explanation of how these two numbers help illustrate the *overall* message of Table 5. Answer with 4 – 6 sentences.

(b) [7 pts] In **Figure 4**, *interpret* the *difference* in the first two bars, called out with arrows. Answer with 2 sentences.

(4) [4 pts] In the cross tabulation below, four numbers are replaced with #A, #B, #C, #D. Answer by filling in the blanks.

x	y = 0	y = 1	Total
0	#A	#B	700
1	#C	#D	300
Total	400	600	1000

If #A is 120, then the value of #D is _____ [numeric value].

If #A is 200, then x and y have _____ [zero / a positive / a negative] correlation.

If #A is 280, then x and y have _____ [zero / a positive / a negative] correlation.

(5) See ***Supplement for Question (5)***: *AI, Skill, and Productivity: The Case of Taxi Drivers.*

(a) [3 pts] In **Panel (a)** of **Figure A2**, *describe* – in this context – the shape of the histogram. Answer with 1 sentence.

(b) [3 pts] In **Panel (b)** of **Figure A2**, compute the 15th percentile, including its units. Answer with a quantitative analysis.

(c) [9 pts] Applying relevant course concepts, explain why it will be challenging to answer the research question: "How much does using AI improve efficiency by reducing cruising time for taxi drivers?" Answer with 3 – 4 sentences.

(6) See ***Supplement for Question (6)***: *Welfare Implications of Electric-Bike Subsidies: Evidence from Sweden.*

(a) [8 pts] Compute the mean and s.d. of replies to the E-bike subsidy question. Answer with a quantitative analysis.

(b) [1 pt] If this study had used a larger sample size of 30,000, we should expect the s.d. computed in Part (a) to _____ [increase / decrease / stay the same]. Answer by filling in the blank.

(c) [3 pts] Table 3 reports _____ [numeric value] separate regressions, and of these, _____ [numeric value] are multiple regressions. These data are _____ [experimental / observational]. Of the explanatory variables listed _____ [numeric value] are dummy variables. Answer by filling in the blanks.

(d) [8 pts] In **Column (1)** of **Table 3**, *interpret* -0.104. Next, *interpret* 0.003. Answer with 2 – 3 sentences.

(e) [1 pt] If this study had used a smaller sample size of 300, we should expect the R^2 in Column (1) of Table 3 to _____ [increase / decrease / stay the same]. Answer by filling in the blank.

This *Supplement* will *NOT* be collected or graded: write your answers on the test papers. ***Supplement*: *Page 1 of 4***

Sample mean: $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ **Sample variance:** $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$ **Sample s.d.:** $s = \sqrt{s^2}$

Sample coefficient of variation: $CV = \frac{s}{\bar{X}}$ **Sample covariance:** $s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{n-1} = \frac{\sum_{i=1}^{n} x_i y_i}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n(n-1)}$

Sample interquartile range: $IQR = Q3 - Q1$ **Sample coefficient of correlation:** $r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1}$

<u>SIMPLE REGRESSION</u>: **OLS line:** $\hat{y}_i = b_0 + b_1 x_i$ $b_1 = \frac{s_{xy}}{s_x^2} = r\frac{s_y}{s_x}$ $b_0 = \bar{Y} - b_1\bar{X}$

Residuals: $e_i = y_i - \hat{y}_i$ **Standard deviation of residuals:** $s_e = Root\ MSE = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(e_i - 0)^2}{n-2}}$

$SST = \sum_{i=1}^{n}(y_i - \bar{Y})^2 = SSR + SSE$ $SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{Y})^2$ $SSE = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

$s_y^2 = \frac{SST}{n-1}$ **Coefficient of determination:** $R^2 = (r)^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

Supplement for Question (2): Recall the *2023 World Happiness Report*. The variable cantril is the mean national Cantril ladder reply and ln_cantril is the natural log of it. The variable healthy_le is national healthy life expectancy in years. The variable oecd is 1 if a country is a member of the OECD and 0 otherwise. The variable ln_gdp_pc is the natural log of GDP per capita. Regressions #1 to #3 use the 2022 data. Regression #4 uses the data for Nicaragua from 2006 through 2022.

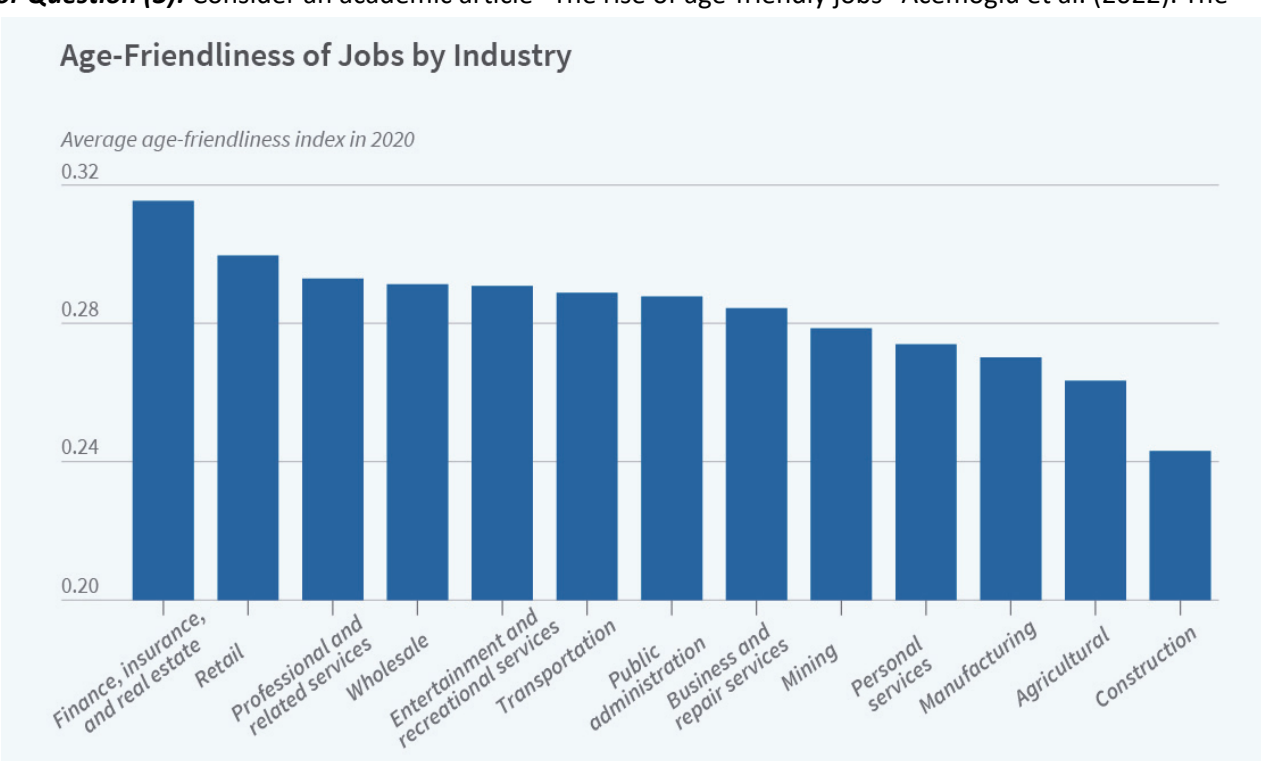
Regression #1: `cantril-hat = -5.378188 + 0.1677988*healthy_le, R-squared = 0.5637, n = 111`

Regression #2: `cantril-hat = 5.043696 + 1.714530*oecd, R-squared = 0.4002, n = 114`

Regression #3: `ln_cantril-hat = 0.2097413 + 0.1561908*ln_gdp_pc, R-squared = 0.5798, n = 108`

Regression #4: `ln_cantril-hat = -33.44057 + 0.0174692*year, R-squared = 0.7576, n = 17`

Supplement for Question (3): Consider an academic article "The rise of age-friendly jobs" Acemoglu et al. (2022). The working paper version is featured in the *NBER Digests*, which includes the figure to the right. The *Supplement* for this question **continues onto the next page.**



Supplement for Question (3) continued:

Excerpt, Abstract: In 1990, one in five U.S. workers were aged over 50 years whereas today it is one in three. One possible explanation for this is that occupations have become more accommodating to the preferences of older workers. We explore this by constructing an ''age-friendliness'' index (AFI) for occupations. We use Natural Language Processing to measure the degree of overlap between textual descriptions of occupations and characteristics which define age-friendliness. We find that between 1990 and 2020 around three quarters of occupations have seen their age-friendliness increase and employment in above-average age-friendly occupations has risen by 49 million. However, older workers have not benefited disproportionately from this rise, with substantial gains going to younger females and college graduates and with male non-college educated workers losing out the most. These findings point to the need to frame the rise of age-friendly jobs in the context of other labour market trends and imperfections. Purely age-based policies are insufficient given both heterogeneity amongst older workers as well as similarities between groups of older and younger workers. The latter is especially apparent in the overlapping appeal of specific occupational characteristics.

Table 5. Percent in a top quartile age-friendly job

Education	Sex	Age	1990	2020	Change
Total	Female	15–49	31.4%	48.2%	16.8
		50–74	31.8%	51.5%	19.7
	Male	15–49	20.6%	31.7%	11.1
		50–74	23.0%	34.6%	11.6
College	Total	15–49	27.5%	48.5%	21.0
		50–74	28.1%	53.2%	25.1
	Female	15–49	25.7%	52.2%	26.5
		50–74	23.0%	56.2%	33.2
	Male	15–49	29.0%	44.3%	15.3
		50–74	30.7%	50.5%	19.8
No college	Total	15–49	24.9%	34.8%	9.9
		50–74	26.5%	36.9%	10.4
	Female	15–49	33.2%	45.7%	12.5
		50–74	33.6%	49.0%	15.4
	Male	15–49	17.8%	26.0%	8.2
		50–74	20.3%	25.9%	5.6

Notes: This table shows proportion of different demographic groups employed in top quartile age-friendly occupations (as per the AFI of each occupation in 2020).

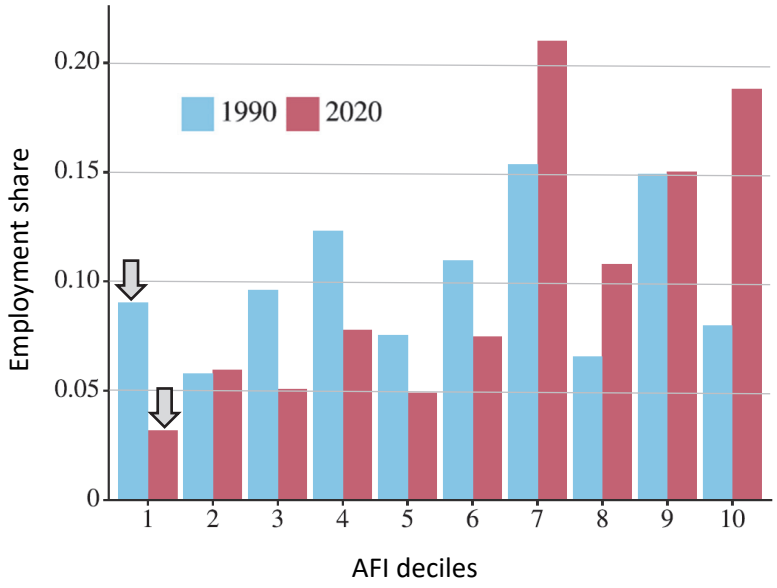


Figure 4. Shifts in employment share by AFI decile

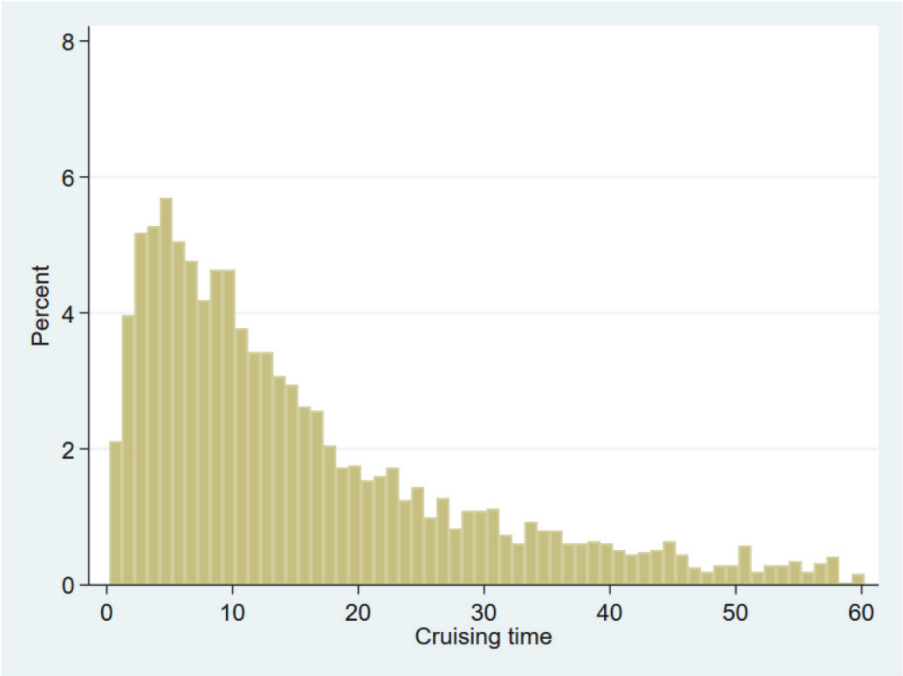
Note: This figure shows employment shares in 1990 and 2020 across occupations by AFI deciles (as per the AFI of each occupation in 2020).

Supplement for Question (5): Consider a 2022 *NBER Working Paper* "AI, Skill, and Productivity: The Case of Taxi Drivers."

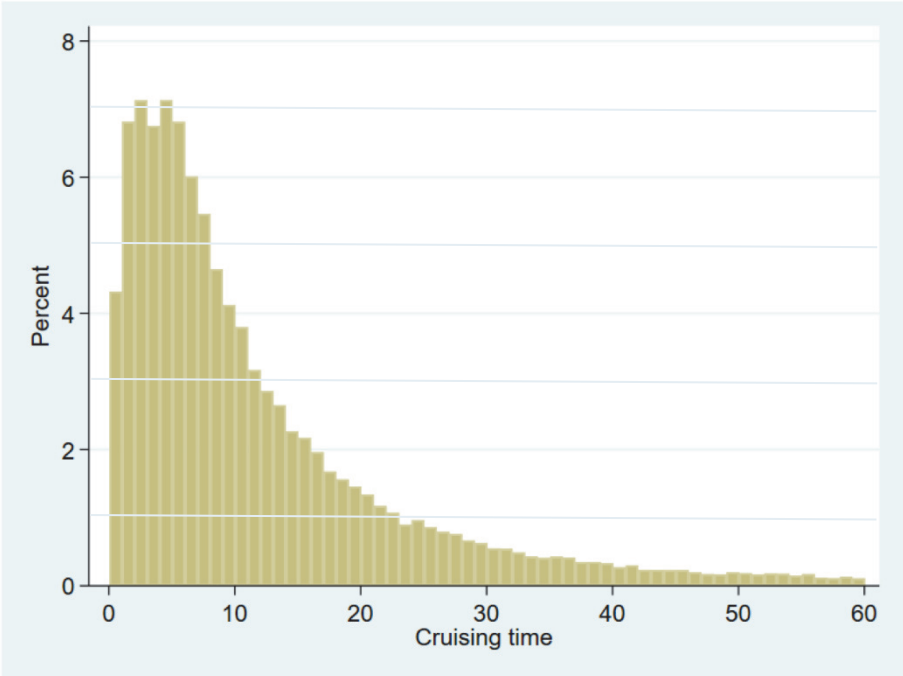
Abstract: We examine the impact of Artificial Intelligence (AI) on productivity in the context of taxi drivers in Japan. The AI we study assists drivers with finding customers by suggesting routes along which the demand is predicted to be high. We find that AI improves drivers' productivity by shortening the cruising time, and such gain is accrued only to low-skilled drivers, narrowing the productivity gap between high-and low-skilled drivers by 14%. The result indicates that AI's impact on human labor is more nuanced and complex than a job displacement story, which was the primary focus of existing studies.

Excerpts (pp. 5 – 6): Our unit of observation is each cruise during which drivers are searching for customers. Formally, we define a cruise as the time between when a cruise starts (i.e., dropping off the previous customer) and when it ends (i.e., picking up the next customer) on the street.

Figure A2 shows the distributions of cruising time separately for (a) when AI is turned on, and (b) when AI is turned off. Both the mean and median cruising times are higher when AI is turned on than when it is turned off; the mean (median) time when AI is turned on is 15.6 (11.4) minutes, whereas the time when it is turned off is 11.7 (7.95) minutes, suggesting that drivers are more likely to turn on AI when it is difficult to find customers. This selective usage of AI indicates that a simple comparison of the average cruising time between when AI is turned on and off is problematic, because it reflects the difference in the underlying demand for the taxi rather than the effect of AI.



(a) AI is turned on



(b) AI is turned off

Figure A2. Histograms of cruising time when AI is turned on/off

Notes: These figures show the distributions of cruising time in the sample period (a) when AI is turned on, and (b) when AI is turned off separately. The number of observations for (a) and (b) are 3,127 and 59,182, respectively.

Supplement for Question (6): Consider a 2022 *NBER Working Paper* "Welfare Implications of Electric-Bike Subsidies: Evidence from Sweden." The abstract begins with "We evaluate a large-scale Swedish electric bike (E-bike) subsidy program in 2018, similar to those implemented in many other countries." See the excerpts below and see Table 3, making sure to read the notes below Table 3. Two results are in boldface for easy reference for one part of the question.

> **Excerpt, p. 10:** We invited a representative sample of 10,500 people from the subsidy-takers to take an online survey in March 2019. The survey contains questions about motives to buy the E-bike, commuting distance, and means of transportation before and after the E-bike purchase. Around 3,500 people answered the survey question about how important the subsidy was for their decision to buy an E-bike.

> **Excerpt, p. 13:** In Table 3, we regress the E-bike subsidy survey question responses on the demographics and show that the responses of 1 to 5 line up in intuitive ways with demographics. We have six demographic features including age, gender, having children, whether the respondent has a university degree and whether the household lived in an urban area (defined by Sweden's three main city regions Stockholm, Gothenburg and Malmo), and the Green Party election outcome (percent of votes received) in the municipality in which the respondent lives. In Table 3, we relate the importance of the E-bike subsidy answer (1 through 5) on these demographic attributes using OLS regressions.

Table 3: The Importance of the Subsidy for Purchase

	"How important was the subsidy for your decision to buy an E-bike?"					
Explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
Log family income	**-0.104*****	-0.102***	-0.114***	-0.113***	-0.112***	-0.108***
	(0.033)	(0.033)	(0.035)	(0.035)	(0.035)	(0.035)
Age		-0.010***	-0.010***	-0.010***	-0.010***	-0.010***
		(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Woman			-0.127***	-0.127***	-0.128***	-0.128***
			(0.040)	(0.040)	(0.040)	(0.040)
Have children			0.072	0.072	0.071	0.075
			(0.048)	(0.048)	(0.048)	(0.048)
University				-0.003	0.000	0.013
				(0.043)	(0.043)	(0.043)
Urban					-0.043	0.024
					(0.054)	(0.063)
Green votes						-0.020**
						(0.010)
Constant	4.294***	4.827***	4.889***	4.889***	4.898***	4.973***
	(0.207)	(0.213)	(0.215)	(0.215)	(0.216)	(0.218)
Observations	3,493	3,493	3,493	3,493	3,493	3,493
R-squared	**0.003**	0.018	0.022	0.022	0.022	0.023

Notes: This table reports OLS regressions where the dependent variable is based on the question "How important was the subsidy for your decision to buy an E-bike?" Possible responses range from "Not important at all – 1" to "Crucially important – 5." There are n = 3,493 responses to this survey question: 253 answer 1; 321 answer 2; 799 answer 3; 1,191 answer 4; 929 answer 5. Explanatory variables include various demographics about households. Robust standard errors are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Note from Prof. Murdock: You have not yet studied standard errors, statistical tests, and OLS estimates in Columns (2) through (6). Those are in grey font, and you should *not* discuss items in grey font in your answers.