# ECO220Y1Y, Test #1, Prof. Murdock

### October 18, 2024, 9:10 am – 11:00 am

- Keep ALL pages closed and face up on your desk until we announce the start, and only then you may detach the *Supplement*, which has the aid sheets and readings, figures, tables, and other materials for some test questions.
- There are 8 test pages with 8 questions, with varying numbers of parts, worth a total of 95 points.
- You have 110 minutes. You must stay for a minimum of 60 minutes.

### Instructions:

- For each question referencing the *Supplement*, carefully review *all* materials. The *Supplement is* <u>NOT</u> *collected*: write your answers on the test papers. At the end, hand in your test papers (you keep the *Supplement*).
- Write your answers clearly, completely, and concisely in the designated space provided immediately after each question. An <u>answer guide</u> ends each question to let you know what is expected. For example, a <u>quantitative</u> <u>analysis</u>, a <u>fully labelled graph</u>, and/or <u>sentences</u>. Any answer guide asking for a <u>quantitative analysis</u> *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. Focus on those expectations.
  - 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 <u>PENCIL</u> and use an <u>ERASER</u> 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 <u>answer guides</u> and avoid excessively long answers.

(1) [7 pts] See Supplement for Question (1): Credit Card Choice. Consider investigating:

- A. the relationship between whether a participant picked the best card and which video they were shown
- B. the relationship between whether a participant picked the best card and how much time they spent choosing

Can you infer causality when interpreting the results for each of the above? Explain. Answer with 3 sentences.

## (2) See Supplement for Question (2): 2023 Salaries at the LCBO.

(a) [2 pts] Describe the shape of the distribution. Circle all that apply.

Normal (Bell Shaped)	Bimodal	Multimodal
Uniform	Somewhat positively skewed	Extremely positively skewed
Symmetric	Somewhat negatively skewed	Extremely negatively skewed

(b) [5 pts] Approximately what percent of these employees have a salary that lies within one standard deviation of the mean? <u>Answer with a quantitative analysis & make your reasoning clear.</u>

(3) See Supplement for Question (3): Happiness and Competitiveness in ECO220Y, 2024/25.

(a) [3 pts] Among those who rate themselves a 10 in competitiveness% [#] rate themselves a 10 in	
happiness. Among those who rate themselves a 10 in happiness% [#] rate themselves a 10 in	
competitiveness. In the class,% [#] rate themselves a 10 in both competitiveness and happiness. If am	iong
those who rate themselves a 10 in competitiveness, 100% rate themselves a 10 in happiness, then the current pos	sitive
correlation would become [stronger / weaker]. <u>Fill in the blanks.</u>	

(b) [4 pts] Compute the 25<sup>th</sup> percentile of competitiveness. <u>Answer with a quantitative analysis & make your reasoning clear.</u>

(c) [5 pts] What is the average level of happiness among those responding? <u>Answer with a quantitative analysis.</u>

# (4) See Supplement for Question (4): Happiness and Competitiveness Across Groups.

(a) [8 pts] Complete the table below with a title, the dependent variable, the explanatory variable, and the six missing numeric values. The title should communicate both the content and main message of the table. <u>Fill in the 9 blanks.</u>

Title:			
Data Source:	2017, Netherlands, Buser et al. 2021 (1)	2023/24, ECO220Y, Prof. Murdock (2)	2024/25, ECO220Y, Prof. Murdock (3)
Dependent variable:			
	0.080		
Constant	not reported		
Observations	3030	410	291
R-squared	0.016		

Notes: Each column reports a separate regression.

(b) [6 pts] Interpret the difference in the constant in Columns (2) and (3) in the table. Answer with 2 – 3 sentences.

### (5) See Supplement for Question (5): The Evolution of Work from Home.

(a) [4 pts] All numbers in Table 1 are conditional on \_\_\_\_\_

\_\_\_\_\_. In Column (2) of Table 1, the number 22.7 is conditional on \_\_\_\_\_\_. In Column (4) of Table 1, the number 19.7 \_\_\_\_\_ [is / is not]

conditional on being an employee (as opposed to another kind of worker). Fill in the blanks.

(b) [6 pts] In Column (3) of Table 1, *interpret* the difference between 48.3 and 14.9, which is 33.4. <u>Answer with 2 – 3</u> <u>sentences.</u>

(c) [6 pts] Is Figure 4 a histogram? Explain. Further, interpret the shortest bar. Answer with 2 – 3 sentences.

(d) [6 pts] In Column (1) of Table 2, *interpret* 0.9. (The first sentence is started for you.) <u>Answer with 2 sentences.</u> In the United States in 2021 to 2022 and among those who were working the week the survey asked about,

(e) [4 pts] Continuing, in Column (1) of Table 2, interpret 0.01. Answer with 1 sentence.

(6) See Supplement for Question (6): Analyzing Nigeria using the PWT 10.0 data.

(a) [4 pts] Using the appropriate regression output, what is the approximate level of real GDP per capita in Nigeria in 2008? Include the units of measurement. Answer with a quantitative analysis & an approximation with units.

(b) [6 pts] Focusing *only* on the parts of the regression output that we have studied in ECO220Y thus far, *interpret* the *key differences* in the two sets of regression results. <u>Answer with 2 - 3 sentences</u>.

(7) [4 pts] To measure the strength of a relationship between two variables, what **advantage(s)** does the rank correlation have over the coefficient of correlation (r)? In other words, good feature(s) that r lacks. <u>Circle all that apply.</u>

It is unit-free	It is robust to endogeneity biases
It measures elasticity	It is robust to nonlinear relationships
It is robust to outliers	It does not go up or down systematically with sample size
It is robust to lurking variables	It has a useful interpretation in terms of standard deviations

(8) See Supplement for Question (8): Happiness and Life Expectancy.

(a) [6 pts] Interpret the slope estimate for Regression #1. Answer with 1 sentence.

(b) [3 pts] Continuing, *interpret* the slope for Regression #2. <u>Answer with 1 sentence</u>.

(c) [3 pts] Continuing, *interpret* the slope for Regression #3. <u>Answer with 1 sentence</u>.

(d) [3 pts] Continuing, interpret the slope for Regression #4. Answer with 1 sentence.

Sample mean:  $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$  Sample variance:  $s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{(\sum_{i=1}^{n} x_i)^2}{n(n-1)}$  Sample s.d.:  $s = \sqrt{s^2}$ Sample coefficient of variation:  $CV = \frac{s}{\overline{x}}$  Sample covariance:  $s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})}{n-1} = \frac{\sum_{i=1}^{n} x_i y_i}{n-1} - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(n-1)}$ Sample interquartile range: IQR = Q3 - Q1 Sample coefficient of correlation:  $r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1}$ SIMPLE REGRESSION: OLS line:  $\hat{y}_i = b_0 + b_1 x_i$   $b_1 = \frac{s_{xy}}{s_x^2} = r \frac{s_y}{s_x}$   $b_0 = \overline{Y} - b_1 \overline{X}$ Residuals:  $e_i = y_i - \hat{y}_i$  Standard deviation of residuals:  $s_e = Root MSE = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n} (e_i - 0)^2}{n-2}}$   $SST = \sum_{i=1}^{n} (y_i - \overline{Y})^2 = SSR + SSE$   $SSR = \sum_{i=1}^{n} (\hat{y}_i - \overline{Y})^2$   $SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$  $s_y^2 = \frac{SST}{n-1}$  Coefficient of determination:  $R^2 = (r)^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$ 

**Supplement for Question (1):** Recall Carlin et al. (2017) "Millennial-Style Learning: Search Intensity, Decision Making, and Information Sharing" and cred\_card.xlsx. The researchers study peoples' ability to choose the best credit card (dominant card) from among four offers. Recall that the researchers randomly divided the participants into groups: they show a short video (baseline) to one group and a longer video (implemental) to the other group.

*Supplement for Question (2):* Consider the Stata summary below of the 2023 salaries of employees of the Liquor Control Board of Ontario (LCBO) earning at least \$100,000.

	Salary	paid in \$1,000s	CAN dollars	
	Percentiles	Smallest		
18	100.837	100.025		
5%	102.262	100.053		
10%	103.146	100.162	Obs	766
25%	105.021	100.325	Sum of Wgt.	766
50%	117.8125		Mean	126.6229
		Largest	Std. Dev.	34.79833
75%	134.933	315.324		
90%	160.791	319.356	Variance	1210.924
95%	194.44	327.261	Skewness	4.468313
99%	273.2	563.291	Kurtosis	40.36134

*Supplement for Question (3):* The class answered some survey questions in Workshop 6 via an anonymous MS Form. Consider the responses to the Cantril ladder (happiness) question and the competitiveness question.

					comp	petitive	eness					
cantril	0	1	2	3	4	5	6	7	8	9	10	Total
0	1	1	0	0	0	0	0	0	1	0	2	5
2	0	0	0	1	0	1	0	0	0	0	2	4
3	0	0	0	0	0	1	1	2	1	1	1	7
4	2	1	0	2	1	1	3	6	5	0	0	21
5	1	0	3	0	2	11	5	11	5	2	1	41
6	0	1	1	1	2	7	17	14	5	4	6	68
7	0	0	1	2	1	7	14	10	1	8	6	60
8	1	0	1	0	4	2	8	18	9	8	7	58
9	1	0	0	1	0	2	1	2	7	3	2	19
10	0	0	0	0	0	0	0	2	2	1	3	8
Total	6	3	6	7	10	32	49	65	56	27	30	291

This Supplement will NOT be collected or graded: write your answers on the test papers. Supplement: Page 2 of 4

**Supplement for Question (4):** Each year in ECO220Y the class answers some survey questions in Workshop 6 via an anonymous MS Form. The variable named *cantril* is for the question: "Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?". The variable named *comp* is for the question: "How competitive do you consider yourself to be? Please choose a value on the scale below, where the value 0 means 'not competitive at all' and the value 10 means 'very competitive'." From Excel, for the academic years of 2023/24 and 2024/25, the coefficient of correlation between *cantril* and *comp* is 0.31199298 and 0.20709925, respectively. Next are the OLS results from Excel.

**ECO220Y, 2023/24**: cantrul = 4.119839316 + 0.337523939 \* comp

**ECO220Y, 2024/25:** cantrul = 5.199494783 + 0.176086548 \* comp

**Supplement for Question (5):** Consider a 2023 paper "The Evolution of Work from Home" published in the *Journal of Economic Perspectives*. It analyzes survey data from the United States. Consider two tables and a figure from it.

Tuble 1. Full time Working Arrangements in the Onited s		020, Tereentage Dis	stributions	
	Fully	Hybrid	Fully	Percent of
	onsite	arrangement	remote	all workers
	(1)	(2)	(3)	(4)
All workers	55.9	28.6	15.5	100.0
Self-employed, excluding contractors and gig workers	24.9	26.8	48.3	7.4
Contractors and gig workers	32.9	22.7	44.4	3.3
All employees	59.3	29.0	11.8	89.3
In firms with 1 to 9 employees	67.5	17.6	14.9	7.1
In firms with 10 to 49 employees	68.3	24.1	7.6	14.2
In firms with 50 to 99 employees	57.2	34.1	8.7	13.3
In firms with 100 to 499 employees	56.5	32.4	11.2	19.7
In firms with 500 to 4,999 employees	50.7	37.7	11.6	19.6
In firms with 5,000 or more employees	63.5	18.8	17.8	15.3

**Table 1.** Full-Time Working Arrangements in the United States as of 2023, Percentage Distributions

Source: Survey of Working Arrangements and Attitudes (Barrero et al. 2020–2023), January to June 2023. N = 25,014.

*Note:* This table considers full-time American workers who are 20–64 years old as of the survey, where "full-time" means working for pay five or more days in the survey reference week. "Fully Onsite" refers to those who worked at their employer's worksite (or a client's location) each workday in the reference week. "Fully Remote" refers to those who worked from home on all workdays in the reference week. "Hybrid Arrangement" refers to those who split the workweek between home and their employer's worksite (or client locations). Column (4) reports the sample percentage of persons in the indicated row.

# Supplement for Question (5), continued:



Figure 4. Work-from-Home Intensity Peaks among Persons in Their Thirties

group in the Survey of Working Arrangements and Attitudes (Barrero et al. 2020–2023). *Note:* The sample runs from January 2022 through June 2023. N = 71,000.

Table 2: How	Work-from-Home	Rates Covary	v with Individu	ual Characteristics
		nates covar		

	Percentage of paid full days worked from home						
Explanatory variables:	(1)	(2)	(3)	(4)	(5)		
1/Eamala)	0.9*	1.0**	1.9***	-0.2	0.6		
I(Feiliale)	(0.5)	(0.5)	(0.6)	(0.6)	(0.6)		
1(Lives with shild under 14)		4.5	5.5***	2.6***	1.6**		
I(LIVES WITH CHING UNDER 14)		(0.5)	(0.7)	(0.7)	(0.7)		
$1(\text{Formale}) \times 1(\text{Fixes with shild under } 14)$			-2.3**	-0.0	0.5		
1(Female) × 1(Lives with child under 14)			(1.0)	(0.9)	(0.9)		
1(one to three years of college)				7.0***	5.1***		
I (one to three years of conege)				(0.7)	(0.6)		
1/four year college degree)				16.5***	11.4***		
I (Tour-year college degree)				(0.6)	(0.7)		
1/graduate degree)				19.1***	13.4***		
I (graduate degree)				(0.7)	(0.8)		
Industry and occupation fixed effects					$\checkmark$		
N	48,244	48,244	48,244	48,244	48,244		
$R^2$	0.01	0.01	0.01	0.04	0.11		

*Source:* Survey of Working Arrangements and Attitudes (Barrero et al. 2020–2023), October 2021 to October 2022 (inclusive). *Notes:* We regress the percentage of paid full days worked from home on indicators for sex, for whether the respondent lives with a child under 14, and education categories. 1(. . .) denotes the indicator function. The sample includes respondents who worked during the reference week and have non-missing data on occupation and industry of the current or most recent job. We report standard errors in parentheses. \*p < 0.1, \*\*p < 0.05, \*\*\*p < 0.01.

*Note from Prof. Murdock:* You have not yet studied standard errors, statistical tests, and OLS estimates in Columns (2) through (5). Those are in grey font, and you should *not* discuss those results in your answers.

**Supplement for Question (6):** With a current population exceeding 230 million people, Nigeria is the largest country in Africa. Using the 10.0 Penn World Tables data – asiap\_pwt\_100\_all.xlsx in DACM – we create a real GDP per capita variable by dividing the variable rgdpna (real GDP at constant 2017 national prices, in mil. 2017US\$) by the variable pop (population, in millions). The Excel regression output below regresses real GDP per capita on year for Nigeria for two different time periods: 1985 through 2000 and 2000 through 2015.

### Regression of real GDP per capita on year for Nigeria from 1985-2000:

Regressio	n Statistics					
R Squared	0.006243453					
Observations	16					
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	6993.834005	13689.7141	0.510882401	0.617394373	-22367.68255	36355.35056
year	-2.037658198	6.870603494	-0.296576305	0.771144268	-16.77363711	12.69832071

Regression of real GDP per capita on year for Nigeria from 2000-2015:

Regressio	n Statistics					
R Squared	0.990299461					
Observations	16					
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-338087.4303	9055.828675	-37.33368225	2.02361E-15	-357510.251	-318664.6095
year	170.5378794	4.510986202	37.8050102	1.70039E-15	160.8627762	180.2129825

Supplement for Question (7): N/A (all information given with the question)

Supplement for Question (8): Recall the data in the 2024 World Happiness Report. Consider 134 countries, excluding Afghanistan, with non-missing data for 2023. The variable named cantril is the national average Cantril life ladder (0 worst to 10 best) survey answer. The variable named  $healthy_le$  is healthy life expectancy (in years) at birth. The variables with a  $ln_p$  prefix apply a natural log transformation to the original variable. Consider the four sets of OLS results from Excel below.

**Regression #1:**  $\widehat{cantrul} = -4.049812145 + 0.14848334 * healthy_le$ 

**Regression #2:**  $ln_{cantril} = -0.106040709 + 0.027822082 * healthy_le$ 

**Regression #3:**  $\widehat{cantrul} = -33.73867247 + 9.432628835 * ln_healthy_le$ 

**Regression #4:**  $ln_{cantrul} = -5.681250942 + 1.77037576 * ln_healthy_le$