ECO220Y1Y, Test #1, Prof. Murdock

October 29, 2021, 9:10 – 11:00 am

U of T E-MAIL: _____@MAIL.UTORONTO.CA

SURNAME (LAST 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 8 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, 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 (which shows your work and reasoning), a fully labelled graph, and/or sentences.
    - 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. Questions give more blank space than is needed for an answer (with typical handwriting) worth full marks. ***Follow the answer guides and avoid excessively long answers.***

(1) See ***Supplement for Question (1)***: *Competitiveness Survey in ECO220Y, September 2021.*

(a) [5 pts] For the competitiveness variable in these data, fill in the selected summary statistics below. Fill in the blanks. When relevant, round to the nearest second decimal place.

The mean is _____, the median is _____, the standard deviation is _____, the range is _____,

the interquartile range is _____, the coefficient of variation is _____, and the 5^{th} percentile is _____.

(b) [4 pts] What percent of respondents self-report being an eight or less on the competitiveness scale? Give the best approximation with the given information. Answer with a quantitative analysis that makes your reasoning clear.

(c) [6 pts] All 630 people in our course had a link to the survey. What is the likely impact of sampling error on the two summaries in the *Supplement*? What about non-sampling errors? Explain. Answer with 3 – 4 sentences.

(2) See ***Supplement for Question (2)***: *Jobs for Sale: Corruption and Misallocation in Hiring.*

(a) [4 pts] For a *relative frequency* histogram, what would be the height of the tallest bar? Answer with a quantitative analysis that makes your reasoning clear.

(b) [8 pts] *Interpret* the histogram in Figure 1. (Include shape, units, range, and *context*.) Answer with 2 – 3 sentences.

(3) [8 pts] See ***Supplement for Question (3)***: *Racial Disparities in the Health Effects from Air Pollution: Evidence from Ports.* Approximate and *interpret* the *difference* between the last two bars. Answer with 2 – 3 sentences.

(4) A project studies the relationship among variables, including **air pollution**, **economic output**, and **population**.

Recall the three data structures:

- cross-sectional,
- time series, and
- panel (longitudinal).

To create the three datasets below, fill in the cells, making up things as necessary.

Write the **variable names** in the **first row**.

To avoid wild guesses for the other cells – for example, if you have no idea about the population of Toronto in 2010 and that is part of your made-up data – use the placeholder **#** to indicate that some number would be in that spot.

The "…" means that there are more rows not shown for conciseness.

(a) [6 pts] Give a concrete example of cross-sectional data for this project. Answer by filling in the cells & the blanks.

…	…	…	…	…

In my made-up data for the described project, there are five variables, and the number of observations is _____.

The unit of observation is _____.

(b) [6 pts] Give a concrete example of times series data for this project. Answer by filling in the cells & the blanks.

…	…	…	…	…

In my made-up data for the described project, there are five variables, and the number of observations is _____.

The unit of observation is _____.

(c) [6 pts] Give a concrete example of panel data for this project. Answer by filling in the cells & the blanks.

…	…	…	…	…

In my made-up data for the described project, there are five variables, and the number of observations is _____.

The unit of observation is _____.

(5) Below is a cross tabulation of two dummy variables. Four numbers are replaced with #A, #B, #C, and #D.

x	y = 0	y = 1	Total
0	#A	#B	429
1	#C	#D	414
Total	417	426	843

(a) [2 pts] If #A were 218, then what would be the values of #B, #C, and #D? Answer with a quantitative analysis.

(b) [4 pts] For each separate scenario below, assess the impact on the correlation between x and y. Fill in the blanks with either "weaker negative," "weaker positive," "stronger negative," or "stronger positive."

- Suppose #A were 100. Changing #A to 125 would mean a _____ correlation.
- Suppose #C were 75. Changing #C to 50 would mean a _____ correlation.

(6) See ***Supplement for Question (6)***: *Can Competitiveness Predict Education and Labor Market Outcomes?*

(a) [2 pts] Use the relevant result(s) in Table 1 to complete the following interpretation. Answer by filling in the blank.

> Of the 1,424 people invited to participate in the matrix-solving experiment in this study using Dutch data in 2018, _____ percent chose to compete (the tournament) in the third round.

(b) [6 pts] In **Column (2)** locate the number **(1766)** near the bottom. What is the *interpretation* of that number? Answer with 2 – 3 sentences.

(c) [6 pts] In the row **Competitiveness (questionnaire)**, compare the descriptive statistics reported in **Column (1)** with those in **Column (2)**. What should we conclude? Answer with 2 – 3 sentences.

(7) [8 pts] See ***Supplement for Question (7)***: *Asiaphoria Meets Regression to the Mean.* A researcher seeks to recompute the result of **0.143** in Table 1 using the revised and updated Penn World Table 10.0 data (PWT 10.0). PWT 10.0 has 12,810 observations (rows) and many variables including those recording the country, year, GDP, population, and so on. Using the PWT 10.0 data, list the analysis steps to obtain the new result (to replace 0.143 in Table 1). The steps should not be specific to a software package (e.g. Excel), but instead should state each step, in order, that the researcher must take with any software. Answer with an enumerated list.

(8) See ***Supplement for Question (8)***: *Gender Differences in Job Search and the Earnings Gap.*

(a) [7 pts] In **Table 5**, what is the *interpretation* of -6719? Answer with 1 – 2 sentences.

(b) [4 pts] In **Table 5**, what is the *interpretation* of the R-squared? Answer with 1 – 2 sentences.

(c) [3 pts] Continuing, in **Table A.6**, what is the *interpretation* of -0.097? Presume your reader has just read your previous answers and this is a continuation: you do not need to repeat yourself. Answer with 1 short sentence.

The pages of this *Supplement* will *NOT* be graded: write your answers on the test papers.

This *Supplement* contains the aid sheets, readings, figures, tables, and other materials for some test questions. For each question referencing this *Supplement*, carefully review *all* materials.

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:** $s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{n-1} = \frac{\sum_{i=1}^{n} x_i y_i}{n-1} - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(n-1)}$

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

<u>SIMPLE REGRESSION</u>:

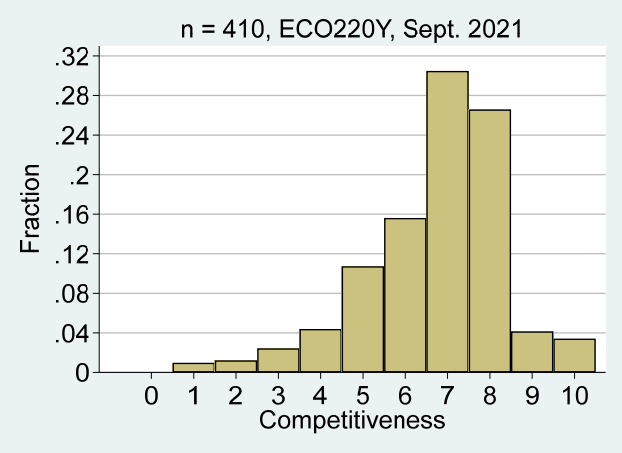
OLS line: $\hat{y}_i = b_0 + b_1 x_i \quad b_1 = \frac{s_{xy}}{s_x^2} = r\frac{s_y}{s_x} \quad b_0 = \bar{Y} - b_1 \bar{X}$

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

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

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

Supplement for Question (1): In September 2021, 410 people in our course answered this 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'." Here are two summaries: a discrete histogram and a STATA summary.



competitiveness

	Percentiles	Smallest		
1%	2	1		
5%	4	1		
10%	5	1	Obs	410
25%	6	1	Sum of Wgt.	410
50%	7		Mean	6.731707
		Largest	Std. Dev.	1.643307
75%	8	10		
90%	8	10	Variance	2.700459
95%	9	10	Skewness	-.8335755
99%	10	10	Kurtosis	4.204276

Supplement for Question (2): Consider a 2021 journal article "Jobs for Sale: Corruption and Misallocation in Hiring" (https://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.20201062). Below are two excerpts and Figure 1, Panel A.

Excerpt, Abstract: Corrupt government hiring is common in developing countries. This paper uses original data to document the operation and consequences of corrupt hiring in a health bureaucracy. Hires pay bribes averaging 17 months of salary.

Excerpt, p. 3097: The data in this study are from a rural area of a large developing country. Due to concerns about respondents and individuals who assisted in the data collection being identified and retaliated against, the identity of the country was removed from the paper.

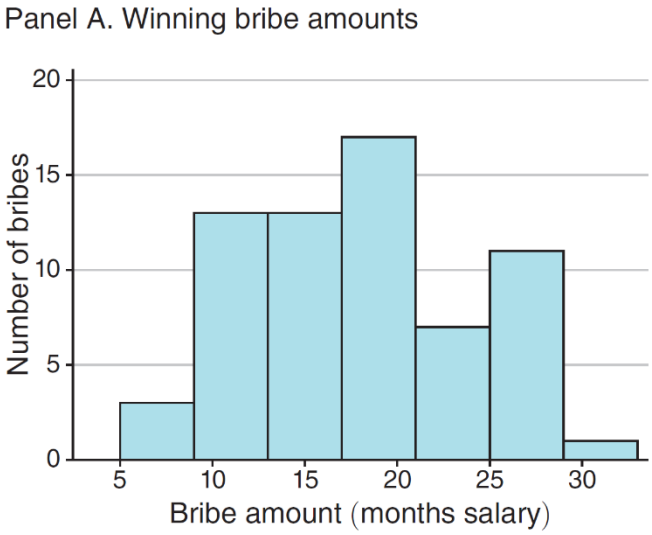


Figure 1: Bribe Payments

Notes: Panel A is the distribution of bribe payments that were made by those hired, where the payment unit is months of salary in the job that they are hired for.

Supplement for Question (3): Recall the 2021 working paper "Racial Disparities in the Health Effects from Air Pollution: Evidence from Ports" (https://www.nber.org/papers/w29108). In Figure 2(b), the last two bars are noted with large arrows.

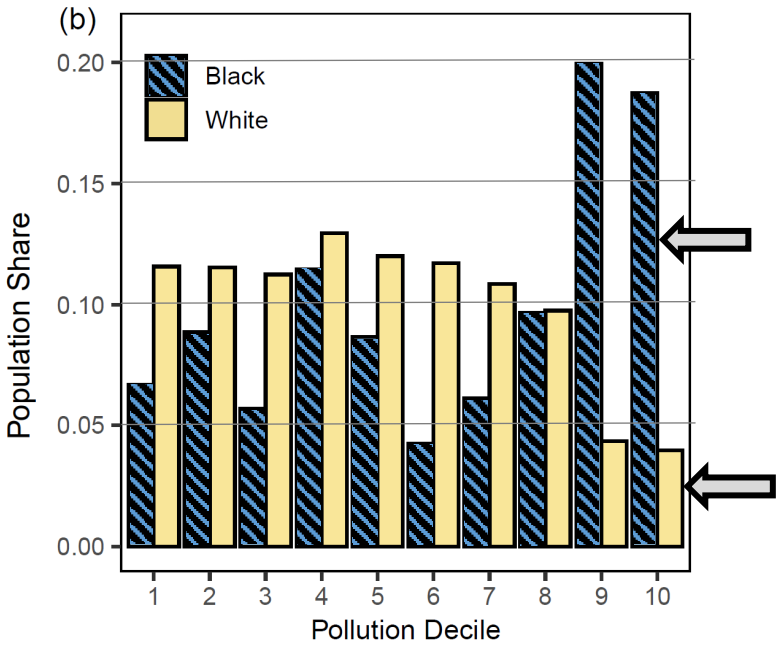


Figure 2(b): Distribution of the population in California port areas by decile of $PM_{2.5}$ concentration. Plots population distribution in the California port areas by decile of PM2.5 concentration, separately for non-Hispanic Black and white population. Larger pollution deciles represent higher pollution exposures. The data are acquired from the U.S. 2010 Decennial Census and U.S. EPA Air Quality System.

Supplement for Question (6): Recall the 2021 working paper "Can Competitiveness Predict Education and Labor Market Outcomes? Evidence from Incentivized Choice and Survey Measures" (https://www.nber.org/papers/w28916). Below are excerpts from Table 1: select columns and select rows. The authors invited a random subset of the "Full sample" to do the incentivized experiment (solve matrices for payment) and hence join the "Experimental sample."

- "Compete (experiment)" records whether the respondent picked the tournament – a chance to compete to solve more matrices than another random participant – in the third round of the matrix solving task rather than a flat payment per matrix solved. It's the incentivized measure of competitiveness.
- "Competitiveness (questionnaire)" records each respondent's self-assessment of their competitiveness on a scale from 0 to 10, where higher numbers mean more competitive. It's the survey measure of competitiveness.
- "Gross monthly income" is self-reported and is measured in Euros.

Table 1: Descriptive statistics (age between 25 and 65)

	Scale	Experimental sample (1)	Full sample (2)
Compete (experiment)	0 or 1 (Binary)	0.274 (0.446)	
Competitiveness (questionnaire)	0 – 10	6.209 (2.066)	6.211 (2.038)
Female	0 or 1 (Binary)	0.532 (0.499)	0.543 (0.498)
Gross monthly income	0 – 19,960	2456 (1790)	2480 (1766)
N		1424	3082

Note: Columns (1) and (2) report means, with standard deviations in parentheses, for the experimental sample and the full sample, respectively.

Supplement for Question (7): Recall the 2014 working paper "Asiaphoria Meets Regression to the Mean," (http://www.nber.org/papers/w20573). Table 1 is reproduced below with one result – **0.143** – called out in boldface.

Table 1: Little persistence in cross-national growth rates across decades

Period 1	Period 2	Correlation	Rank Correlation	Regression Coefficient	R-squared	N
Adjacent decades						
1950-60	1960-70	0.363	0.381	0.378	0.132	66
1960-70	1970-80	0.339	0.342	0.382	0.115	108
1970-80	1980-90	0.337	0.321	0.323	0.114	142
1980-90	1990-00	0.361	0.413	0.288	0.13	142
1990-00	2000-10	0.237	0.289	0.205	0.056	142
One decade apart						
1950-60	1970-80	0.079	0.192	0.095	0.006	66
1960-70	1980-90	0.279	0.312	0.306	0.078	108
1970-80	1990-00	0.214	0.214	0.163	0.046	142
1980-90	2000-10	0.206	0.137	**0.143**	0.043	142
Two decades apart						
1960-70	1990-00	0.152	0.177	0.152	0.023	108
1970-80	2000-10	-0.022	0.005	-0.015	0.001	142

Supplement for Question (8): Consider a 2021 working paper "Gender Differences in Job Search and the Earnings Gap: Evidence from Business Majors" (https://www.nber.org/papers/w28820). Below are excerpts from the abstract, paper, Table 5, and Table A.6. For Tables 5 and A.6, items you have not yet studied are in grey font and you may ignore them.

> **Excerpt, Abstract:** To understand gender differences in the job search process, we collect rich information on job offers and acceptances from past and current undergraduates of Boston University's Questrom School of Business.

> **Excerpt, p. 7:** We use an online survey administered to the 2013 to 2017 Questrom graduating classes.

Table 5: Gender Gap in Accepted Earnings

	Dependent Variable: Accepted Earnings in the First Job
	(1)
Female	-6719 ***
	(1147)
Mean of dependent variable	61708
R^2	0.026
N	1359
Note: The dependent variable is total accepted earnings in the first year in 2017 dollars. Robust standard errors in parentheses. ***significant at the 1% level, **5% level, *10% level.	
Notes from Prof. Murdock: Column (1) reports the results from a simple regression (OLS). Items you have not yet studied are in grey font and you may ignore them.	

Table A.6: Gender Gap in Log Earnings

	Dependent Variable: Log Accepted Earnings in the First Job
	(1)
Female	-0.097 ***
	(0.018)
Mean of dependent variable	10.98
R^2	0.021
N	1359
Note: The dependent variable is the natural log of total accepted earnings in the first year in 2017 dollars. Robust standard errors in parentheses. ***significant at the 1% level, **5% level, *10% level.	
Notes from Prof. Murdock: Column (1) reports the results from a simple regression (OLS). Items you have not yet studied are in grey font and you may ignore them.	