ECO220Y1Y, Test #1, Prof. Murdock

October 11, 2019, 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 7 questions (many 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 required for test questions. For each question referencing the *Supplement*, carefully review *all* materials. ***The Supplement will NOT be graded:*** 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), 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)***: *Changes in German Local Business Tax Rates.*

(a) [6 pts] Which of the following is a reasonable estimate of the number of tax rate changes that fall in the range from a 1 percentage point increase to a 2 percentage point increase? **(A)** 1,035 **(B)** 2,070 **(C)** 4,140 **(D)** 8,280 Support your choice with a quantitative analysis and rationale. Answer with a clear choice & supporting work.

(b) [6 pts] Describe the shape of this distribution. Explain what it means in this context. Answer with 2 – 3 sentences.

(2) [8 pts] See ***Supplement for Question (2)***: *Fare Evasion on Public Transit.* A comparison of overall fare evasion rates between Cities A and B suffers from Simpson's Paradox (composition effects). Create a table to best illustrate the paradox. Next, state the paradox in this specific case. Answer with a complete table & 1 sentence.

(3) [6 pts] See ***Supplement for Question (3)***: *Beliefs about the safety of vaccines.* What is the difference in the percent strongly agreeing that vaccines are safe in South Asia versus Western Europe? What is the *interpretation* of that *difference*? Answer with 1 precise sentence that states & *interprets* the difference.

(4) See ***Supplement for Question (4)***: *Corporate Culture.*

(a) [4 pts] For **Q2** in **Table 3**, what is the value that belongs in the shaded grey box in Column (2)? Answer with a quantitative analysis.

(b) [3 pts] For **Q2** in **Table 3**, aside from the median, percentiles are not reported. What is the value of the 10^{th} percentile? Answer with a value & your reasoning.

(c) [5 pts] For **Q4c** in **Table 3**, what are the values that belong in the two shaded grey boxes in Columns (2) and (3)? Answer with a quantitative analysis.

(d) [9 pts] **Table D.1** mentions "non-response bias." What is non-response bias? Also, is it a type of sampling error or non-sampling error? Further, how does Table D.1 help check for substantial issues related to non-response bias? Overall, what should we conclude from Table D.1 about this potential issue? Answer with 3 – 4 sentences.

(e) [6 pts] **Panels D and E** of **Table D.1** summarize two variables: profitability and industry. You wonder if these two variables are related with each other. Which methods, tools, or formulas should you use to best describe that relationship? Why? Which *visual* summary of the relationship should you also include? Answer with 1 – 2 sentences.

(5) [8pts] See ***Supplement for Question (5)***: *Variance-Covariance Matrix.* Which pair of variables has the *strongest positive correlation*? Answer with a quantitative analysis.

(6) See ***Supplement for Question (6)***: *Household Electricity and Natural Gas Use.*

(a) [9 pts] Draw an exact box plot of **electricity** usage for the sample of 68 homes in **San Francisco County**. Label all important numeric values. Answer with a fully-labelled box plot.

(b) [4 pts] For the the **natural gas** usage distribution for **California**, both Histogram #1 and Histogram #2 have 38 bins. Why should we expect them to have the same (or a very similar) number of bins? Answer with 1 sentence.

(c) [6 pts] Which important feature about the shape of the **natural gas** usage distribution for **California** is clear in Histogram #2 but is not clear in Histogram #1? Why? What is it about these data and how histograms are constructed that hides that important information in Histogram #1? Answer with 2 – 3 sentences.

(7) See ***Supplement for Question (7)***: *Hazed and Confused.*

(a) [3 pts] What is the authors' *research question*? Answer with 1 question.

(b) [4 pts] For answering the authors' research question, how would we imagine the gathering of *experimental data*, even if it is not realistic in this case? Answer with 1 – 2 sentences.

(c) [8 pts] What does it mean to say that pollution exposure is *endogenous*? *Explain.* How does this affect the assessment of the reported coefficient of correlation (0.66)? Answer with 3 – 4 sentences.

This *Supplement* contains the aid sheets and readings, figures, tables, and other materials for test questions. For each question referencing this *Supplement*, carefully review *all* materials. You keep this Supplement: please do *not* hand it in.

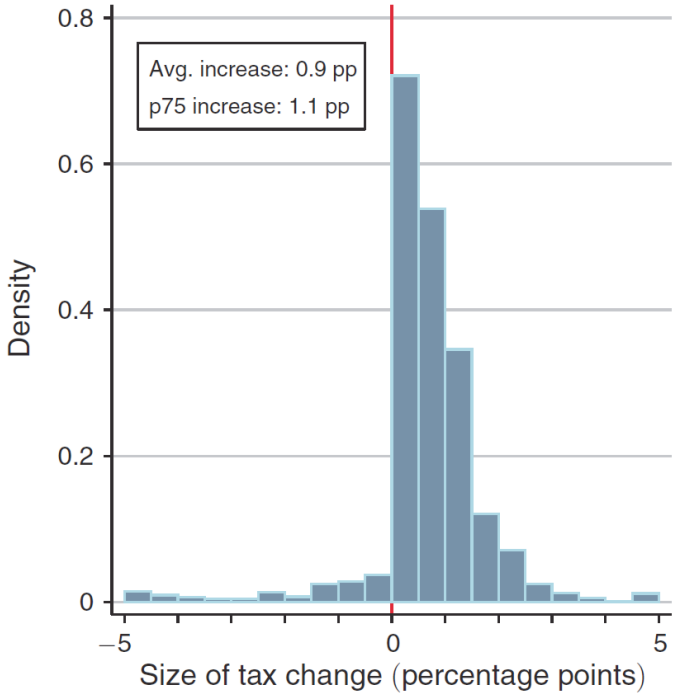
Sample mean: $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ **Sample variance:** $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$ **Sample s.d.:** $s = \sqrt{s^2}$

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:** $r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1}$

Supplement for Question (1): Recall the 2018 journal article "Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany" (https://doi.org/10.1257/aer.20130570). An excerpt from Figure 2 "Distribution of Local Business Tax Changes" is to the right. This histogram shows how the German local business tax rates changed between 1993 and 2012 for a sample of municipalities.

> **Excerpt discussing Figure 2 (histogram), pp. 399-400:** [In this sample,] 93 percent of the tax changes are increases. The mean increase is 0.9 percentage points (or 5 percent) and the seventy-fifth percentile of the tax increase distribution is equal to 1.1 percentage points (6 percent).



Notes: The histogram shows the distribution of changes in the local business tax rate. The sample consists of 17,999 tax rate changes in 10,001 municipalities. We omit 0.1 percent of the observations with absolute changes larger than 5 percentage points for illustrative purposes.

Supplement for Question (2): Suppose two cities are comparing their public transit fare evasion problems. Each city has two modes of public transit: streetcars and buses. The table below shows the results of their fare evasion audits where each city observed a random sample of rides (observations) and recorded how many of those rides were instances of fare evasion (invalid payments). For example, City A audited 3,000 streetcar rides and found 750 instances of evasion.

	City A		**City B**	
	Invalid payments	Observations	Invalid payments	Observations
Streetcars	750	3,000	2,400	12,000
Buses	600	5,000	400	4,000

Supplement for Question (3): Consider a June 2019 post in *The Economist*, which includes the following excerpt and figure.

> **Excerpt:** A report from the Wellcome Trust, a charity, covering 140 countries discovered that only 80% of people trust vaccines to some degree.

For those wishing to agree with a statement that vaccines are safe, the survey gave two options: "strongly agree" and "somewhat agree." Here are two of the exact values for the figure: 36% of people in Western Europe strongly agreed and 85% in South Asia strongly agreed.

The Economist

Supplement for Question (4): Consider the 2019 working paper "Corporate Culture: Evidence from the Field" (https://dx.doi.org/10.2139/ssrn.2805602). It seeks to answer questions that include "How do we measure corporate culture?" and "Is it possible to assign a dollar value to culture?" using a survey of 1,348 corporate executives in North America. An excerpt of **Table 3** from the paper is below. Table 3 it summarizes the replies to two survey questions: Q2 and Q4c. It includes column numbers in parentheses for easy reference. The values from three cells, shaded grey, have been intentionally erased in Table 3. An excerpt of **Table D.1** from the paper is on the next page.

Table 3: The value of corporate culture

This table provides descriptive statistics on the value placed on corporate culture by surveyed executives at public and private North American firms. The question is listed along with the percentage of responses in each category.

Q2, "How important do you believe corporate culture is at your firm?"

Obs.	Mean	Std. dev.	Median	1 = Not important	2 = Somewhat Important	3 = Important	4 = Very important
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1335		0.77	4	4.2%	4.9%	25.4%	65.5%

Q4c, "Do you believe that improving your corporate culture would increase your firm's value?"

Obs.	Mean	Std. dev.	Median	0= No	1= Yes
(1)	(2)	(3)	(4)	(5)	(6)
1104			1	8.1%	91.9%

Supplement for Question (4), continues on the next page >>>>>

Supplement for Question (4), cont'd: In the excerpt of **Table D.1** below, Panels B and C are omitted to save space. They are consistent with the rest of the table.

Table D.1: Test of non-response bias: Respondents versus universe of invited firms

This table compares the demographic information for people who respond to the culture survey and the universe of firms invited to participate for which we knew demographic information (those that respond to the Duke Quarterly CFO survey). Column 1 summarizes responses from those that took the culture survey. Column 2 summarizes responses from Duke Quarterly CFO survey respondents since 2011 who we asked to take the culture survey.

Panel A. Revenue	**Culture Survey Respondents**	**CFO Survey Respondents**
1 = Less than $25 million	33%	27%
2 = $25-$99 million	24%	25%
3 = $100-$499 million	19%	24%
4 = $500-$999 million	7%	7%
5 = $1-$4.9 billion	8%	8%
6 = $5-$9.9 billion	3%	3%
7 = More than $10 billion	6%	5%
Mean	2.67	2.74
Panel D. Profitability		
0 = No after-tax profit	15%	12%
1 = After-tax profit	85%	88%
Mean	0.85	0.88
Panel E. Industry		
Communication	2%	3%
Energy	2%	6%
Finance	14%	12%
Healthcare	5%	5%
Manufacturing	23%	26%
Mining	3%	5%
Retail	12%	15%
Services	15%	14%
Technology	8%	5%
Other	16%	10%

Supplement for Question (5): Consider data with 1,000 observations and four interval variables named w, x, y, and z. See the variance-covariance matrix below.

```
. correlate w x y z, covariance;
(obs=1,000)

             |        w        x        y        z
-------------+------------------------------------
           w |  70.8024
           x |  88.1797  1038.88
           y |  21.1834  59.7227  216.618
           z |  52.6201  129.192  25.5046  85.6173
```

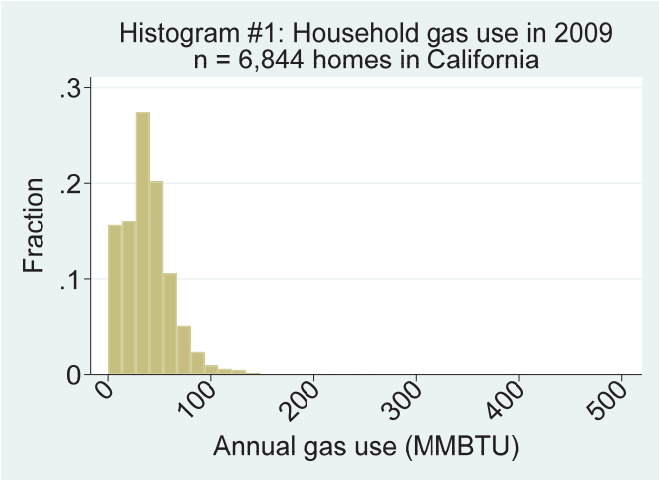
Supplement for Question (6): Consider data from the 2016 paper "How Much Energy Do Building Energy Codes Save? Evidence from California Houses" (https://www.aeaweb.org/articles?id=10.1257/aer.20150102). A survey of households provides many variables describing each house, its owners, and location. The key dependent variables are annual household electricity use in MMBTUs and annual household natural gas use in MMBTUs.

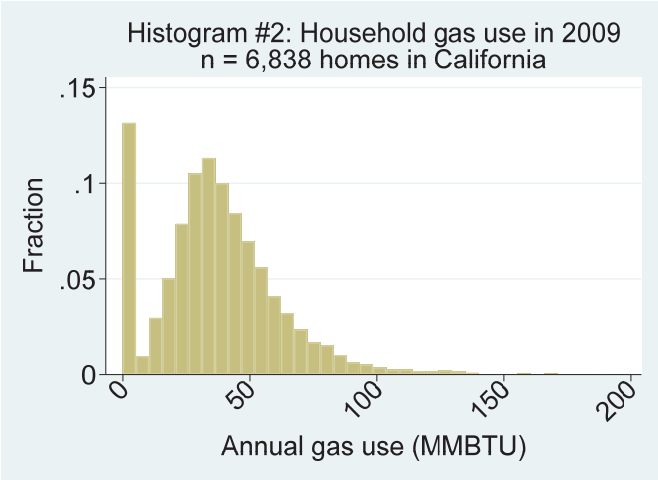
See the STATA summary of annual household electricity usage for a sample of 68 homes in San Francisco County in 2009.

Annual household electricity use in MMBTUs
(millions of British thermal units)

Percentiles		Smallest		
1%	5.703708	5.703708		
5%	6.999839	6.441251		
10%	7.960055	6.835079	Obs	68
25%	10.70018	6.999839	Sum of Wgt.	68
50%	16.09537		Mean	18.10568
		Largest	Std. Dev.	9.32831
75%	24.0922	37.04711		
90%	29.39358	39.34836	Variance	87.01737
95%	37.04711	43.58534	Skewness	1.072756
99%	49.35115	49.35115	Kurtosis	4.049082

The two histograms below summarize the 2009 annual household natural gas usage for a sample of homes in California. Histogram #1 shows all homes. Histogram #2 excludes the six homes with gas usage between 200 and 510 MMBTUs.





Supplement for Question (7): An August 2018 *NBER Working Paper* "Hazed and Confused: The Effect of Air Pollution on Dementia" (https://www.nber.org/papers/w24970.pdf) examines how cumulative long-term exposure to air pollution – concentration of particulate matter (PM2.5) – affects the chance of being diagnosed with Alzheimer's disease and related dementias later in life. Alzheimer's leads to continued deterioration of memory and thinking skills. While the biggest risk factor is age, other risk factors include things people can work to change or manage such as "diabetes, high blood pressure, obesity, smoking, depression, cognitive inactivity or low education, and physical inactivity."[1] On page 2 of the *NBER Working Paper*, the authors' state: "Like the prior observational studies, we observe strong, positive relationships between the prevalence of dementia and the average concentration of PM2.5 over a decade." Among 75 year olds, the coefficient of correlation between dementia rates and PM2.5 exposure is 0.66. In other words, people who have spent time living in places with high levels of air pollution are more likely to suffer from dementia later in life.

[1] Retrieved from the risk factors page on the Alzheimer Society Canada website (https://alzheimer.ca/en/Home/About-dementia/Alzheimer-s-disease/Risk-factors) on September 26, 2019.