ECO220Y1Y, Test #2, Prof. Murdock

November 15, 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 8 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) [10 pts] See ***Supplement for Question (1)***: *Asiaphoria and PWT 9.1.* Following Pritchett and Summers (2014), what are the steps to get from the PWT data given in the *Supplement* to the average annual growth rate of real GDP per capita (in percentage terms) for Taiwan for the half decade period from ***2012 to 2017***? Answer with a list of steps that match the methods in Pritchett and Summers (2014) and that someone could follow with any software package.

(2) See ***Supplement for Question (2)***: *Variance-Covariance Matrix and OLS Regressions.*

(a) [3 pts] In the regression $\widehat{var3} = a + b * var1$, what is the value of b? Answer with a quantitative analysis.

(b) [3 pts] In the regression $\widehat{var2} = a + b * var3$, what is the value of the SST? Answer with a quantitative analysis.

(3) [12 pts] See ***Supplement for Question (3)***: *Segregation.* Of the two cities, which has substantial racial segregation across neighborhoods: **(A)** only City A, **(B)** only City B, **(C)** both City A and City B, **(D)** neither City A nor City B? *Explain.* Answer with 4 – 6 sentences with the relevant quantitative analysis woven in.

(4) See ***Supplement for Question (4)***: *Decile Income Transition Matrix in Canada.*

(a) [3 pts] Correctly fill in the blanks with "is" or "is not".

Event **A** _____ mutually exclusive (i.e. disjoint) of Event **B**.

Event **A** _____ independent of Event **B**.

Event **A** _____ mutually exclusive (i.e. disjoint) of Event **C**.

Event **B** _____ mutually exclusive (i.e. disjoint) of Event **D**.

Event **B** _____ independent of Event **D**.

(b) [8 pts] Table 1 reports which probability: $P(B \mid C)$, $P(C \mid B)$, or $P(B \,\&\, C)$? What is the numeric value of that probability? What is its interpretation? Is this a large or small probability in this context? Explain. Answer with the probability in formal notation & 2 – 3 sentences.

(c) [8 pts] Consider a random sample of 20 taxfilers in the 10th income decile in 2007. What is the chance that more than two of them are below the 8th income decile in 2012? Answer with a quantitative analysis.

(5) [7 pts] See ***Supplement for Question (5)***: *Change in GDP/capita from 2007 to 2017.* What is the coefficient of correlation between 2007 real GDP/capita and 2017 real GDP/capita? Answer with a quantitative analysis.

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

(a) [10 pts] In **Fig. 1**, what do 28.277 and -1.814 mean? Relatedly, in **Fig. 2**, what does -0.056 mean? Answer with 2 – 3 sentences that offer an interpretation of these three values.

(b) [6 pts] For **Fig. 3** and **Fig. 4**, *why* is the R^2 bigger for Fig. 4? Would the value of the s_e be bigger for Fig. 3 or Fig. 4? Explain. Answer with 2 – 3 sentences.

(c) [5 pts] In **Fig. 4**, what does 0.347 mean? Answer with 1 precise sentence that offers an interpretation.

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

(a) [2 pts] In Figure A1 for **85 year olds**, when $PM_{2.5}$ exposure is 1 ***standard deviation*** higher, how much higher is the dementia rate on average? Answer by correctly filling in the blank.

When $PM_{2.5}$ exposure is 1 standard deviation higher, the dementia rate is _____ higher on average.

(b) [8 pts] In Figure A1 for **80 year olds**, what do the values 1.09 and 5.82 mean? Answer with 2 sentences interpreting those values.

(8) [10 pts] See ***Supplement for Question (8)***: *Educational Attainment of Toronto's Population, 2016.* Consider a Torontonian with a university degree. What is the chance that that person lives in a high income neighborhood? Answer with a quantitative analysis using formal notation and the definitions of events given in the *Supplement*.

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

Addition rule: $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$ **Conditional probability:** $P(A|B) = \frac{P(A\ and\ B)}{P(B)}$

Complement rules: $P(A^C) = P(A') = 1 - P(A)$ $P(A^C|B) = P(A'|B) = 1 - P(A|B)$

Multiplication rule: $P(A\ and\ B) = P(A|B)P(B) = P(B|A)P(A)$

Expected value: $E[X] = \mu = \sum_{all\ x} xp(x)$ **Variance:** $V[X] = E[(X-\mu)^2] = \sigma^2 = \sum_{all\ x}(x-\mu)^2 p(x)$

Covariance: $COV[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] = \sigma_{XY} = \sum_{all\ x}\sum_{all\ y}(x-\mu_X)(y-\mu_Y)p(x,y)$

Laws of expected value:

$E[c] = c$

$E[X + c] = E[X] + c$

$E[cX] = cE[X]$

$E[a + bX + cY] = a + bE[X] + cE[Y]$

Laws of variance:

$V[c] = 0$

$V[X + c] = V[X]$

$V[cX] = c^2 V[X]$

$V[a + bX + cY] = b^2 V[X] + c^2 V[Y] + 2bc * COV[X,Y]$

$V[a + bX + cY] = b^2 V[X] + c^2 V[Y] + 2bc * SD(X) * SD(Y) * \rho$

where $\rho = CORRELATION[X,Y]$

Laws of covariance:

$COV[X,c] = 0$

$COV[a + bX, c + dY] = bd * COV[X,Y]$

Combinatorial formula: $C_x^n = \frac{n!}{x!(n-x)!}$ **Binomial probability:** $p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$ for $x = 0,1,2,\ldots,n$

If *X* is Binomial $(X \sim B(n,p))$ **then** $E[X] = np$ **and** $V[X] = np(1-p)$

SIMPLE REGRESSION:

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

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

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

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

Supplement for Question (1): Recall the Penn World Tables (PWT) and PWT 9.1 (released April 30, 2019, DOI: 10.15141/S50T0R). Below are the data for Taiwan from 2000 through 2017, the most recent year available. Recall that the variable rgdpna is "Real GDP at constant 2011 national prices (in mil. 2011US$)" and the variable pop is "Population (in millions)." Also, recall the methods in Pritchett and Summers (2014): "Asiaphoria Meets Regression to the Mean."

country	countrycode	oecd	continent	year	rgdpna	pop
Taiwan	TWN	0	Asia	2000	573629.5	22.18453026
Taiwan	TWN	0	Asia	2001	566402.25	22.34111977
Taiwan	TWN	0	Asia	2002	597976.75	22.46317291
Taiwan	TWN	0	Asia	2003	622616.9375	22.56266212
Taiwan	TWN	0	Asia	2004	663124.4375	22.64683533
Taiwan	TWN	0	Asia	2005	699038.375	22.72975349
Taiwan	TWN	0	Asia	2006	738348.5	22.82345581
Taiwan	TWN	0	Asia	2007	786466.125	22.91744423
Taiwan	TWN	0	Asia	2008	792004	22.99769592
Taiwan	TWN	0	Asia	2009	779596.375	23.07840157
Taiwan	TWN	0	Asia	2010	862475.3125	23.14094734
Taiwan	TWN	0	Asia	2011	895288.625	23.19351768
Taiwan	TWN	0	Asia	2012	913765.1875	23.27036667
Taiwan	TWN	0	Asia	2013	933890.3125	23.34466934
Taiwan	TWN	0	Asia	2014	971442.5	23.40363503
Taiwan	TWN	0	Asia	2015	979276.125	23.46291351
Taiwan	TWN	0	Asia	2016	994081.375	23.51594543
Taiwan	TWN	0	Asia	2017	1024689.125	23.55552292

Supplement for Question (2): Consider data with 50 observations and three interval variables named var1, var2, and var3. See the variance-covariance matrix below.

```
. correlate var1 var2 var3, covariance;
(obs=50)
             |     var1     var2     var3
-------------+---------------------------
        var1 |  76.9928
        var2 |  140.151  977.224
        var3 |  61.6322  183.694  92.3089
```

Supplement for Question (3): Consider the joint probability tables below for two hypothetical cities. For each city, for a randomly selected resident, the definitions of events are below. Note that W' (aka W^C) is the complement of event W.

- Event **W:** Resident's race is white
- Event **L:** Resident lives in a low income neighborhood
- Event **M:** Resident lives in a middle income neighborhood
- Event **H:** Resident lives in a high income neighborhood

City A			
	L	**M**	**H**
W	0.0659	0.0530	0.0712
W'	0.2810	0.2257	0.3032

City B			
	L	**M**	**H**
W	0.0930	0.0992	0.1465
W'	0.3125	0.1996	0.1492

Supplement for Question (4): Recall Table 1 and the excerpt below from Statistics Canada (2016) "The evolution of income mobility in Canada" https://www150.statcan.gc.ca/n1/pub/75f0002m/75f0002m2016001-eng.htm.

> **Excerpt:** Table 1 shows the estimated transition probabilities for the 2007-to-2012 panel of taxfilers. Consider the third row. The 10 elements in this row show the 5-year transition probabilities for taxfilers who were in the third decile in 2007. The [value] 36.3%, shows the proportion of taxfilers from the third decile in the 2007 distribution who stayed in the same decile in 2012. [This measures immobility.] … Upward mobility indicates how many of them moved to the fourth or higher deciles in 2012. Downward mobility indicates how many of them moved to the second or the first deciles in 2012.

For a randomly selected taxfiler from the 2007-to-2012 panel of taxfilers, define events as:

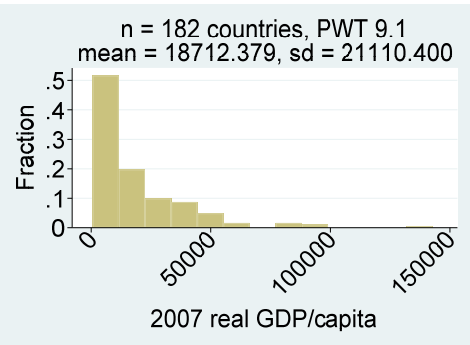
- Event **A:** Being in the **1st income decile in 2007**
- Event **B:** Being in the **9th income decile in 2007**
- Event **C:** Being in the **10th income decile in 2012**
- Event **D:** Being **downwardly mobile** from 2007 to 2012

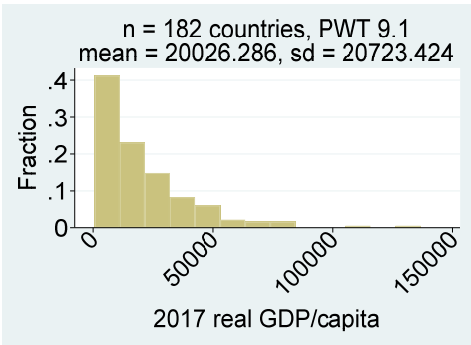
Table 1. Decile Income Transition Matrix for the 2007-to-2012 panel of Taxfilers

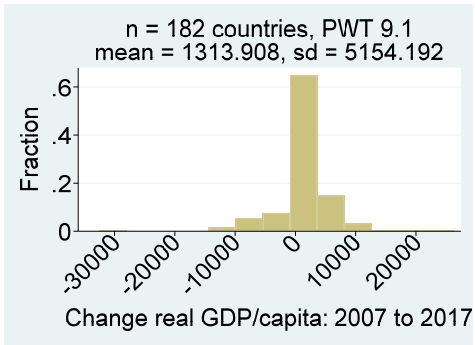
2007 decile	2012 decile										Total mobility statistics		
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Immobility	Upward	Downward
	Percentage										Percentage		
1st	39.7	22.9	11.2	7.6	5.4	4.0	3.1	2.5	2.0	1.5	39.7	60.3	0.0
2nd	13.5	39.4	18.5	10.0	6.2	4.4	3.1	2.2	1.6	1.0	39.4	47.0	13.5
3rd	6.4	14.9	36.3	16.9	9.7	6.1	4.1	2.8	1.8	1.0	36.3	42.4	21.4
4th	4.5	7.2	17.5	27.6	17.5	10.7	6.8	4.3	2.6	1.3	27.6	43.2	29.2
5th	3.1	4.4	8.2	17.0	25.6	17.6	11.3	6.9	4.0	1.8	25.6	41.6	32.7
6th	2.3	3.0	5.1	9.0	16.9	24.3	18.3	11.7	6.5	2.7	24.3	39.3	36.4
7th	1.8	2.1	3.4	5.9	9.5	16.9	24.3	19.6	11.8	4.6	24.3	36.0	39.7
8th	1.4	1.6	2.3	4.0	6.4	9.9	17.2	26.3	22.0	8.8	26.3	30.9	42.8
9th	1.2	1.2	1.6	2.7	4.1	6.4	10.1	18.1	32.5	22.1	32.5	22.1	45.4
10th	1.2	0.9	1.1	1.6	2.4	3.4	5.3	8.6	18.2	57.4	57.4	0.0	42.6

Source: Statistics Canada, Longitudinal Administrative Databank 2007 and 2012, authors' calculations.

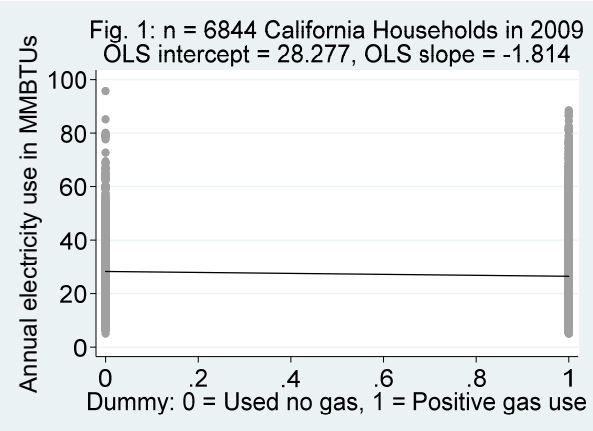
Supplement for Question (5): See the three histograms below. Note the summary statistics in the titles of each.

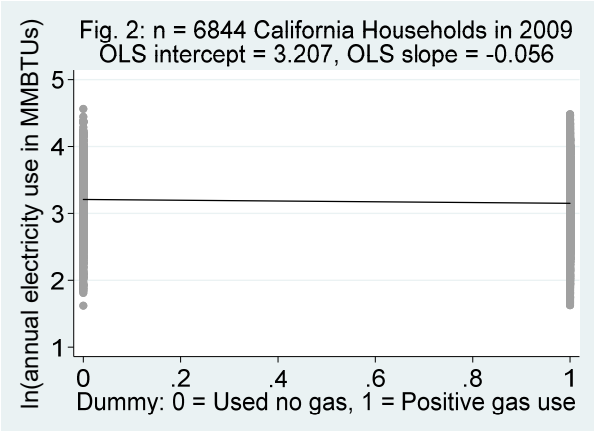






Supplement for Question (6): Consider data from the 2016 paper "How Much Energy Do Building Energy Codes Save? Evidence from California Houses" (https://www.aeaweb.org/articles?id=10.1257/aer.20150102). A survey of households provides many variables describing each house, its owners, location, annual household electricity use in MMBTUs, and annual household natural gas use in MMBTUs. Figures 1 and 2 (below) use these data to address a simple descriptive question: how does electricity usage compare between households with no natural gas use versus those that use natural gas? See the OLS results in the titles. Notice that in Figure 2 the y variable has had a natural logarithm applied.



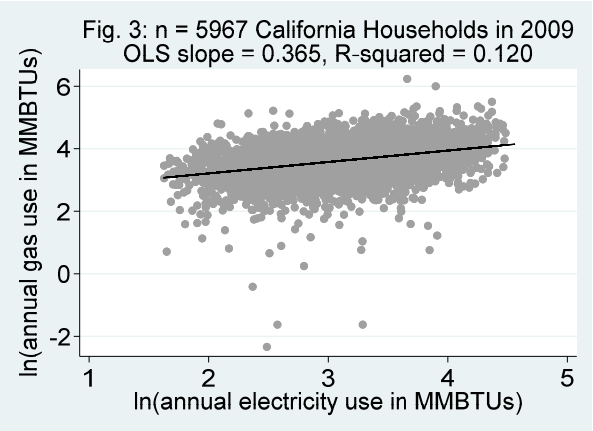


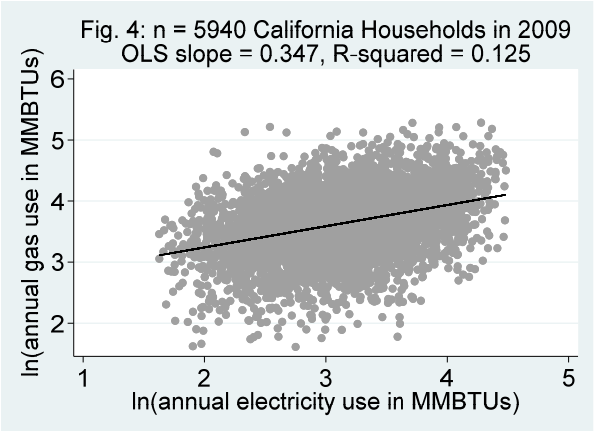
Next, see the STATA summary of annual household natural gas usage for a sample of 5,967 homes in California in 2009 for homes reporting positive (i.e. non-zero) natural gas usage. It includes some outliers in both the left and right tail.

```
          Annual household natural gas use in MMBTUs
              (millions of British thermal units)
-------------------------------------------------------------
      Percentiles      Smallest
 1%     8.487006       .0967511
 5%     15.69719        .196828
10%     20.30206       .1972256       Obs                5,967
25%     28.33918       .6623706       Sum of Wgt.        5,967

50%     38.74644                      Mean            43.08604
                        Largest       Std. Dev.       23.74111
75%     52.58042       211.5951
90%     69.85884       244.8814       Variance        563.6405
95%     82.91546       402.0947       Skewness        3.407511
99%     125.7113       509.4124       Kurtosis        41.75801
```

Figures 3 and 4 (below) use these data to address another simple descriptive question: how well does electricity use predict natural gas use among households with non-zero natural gas use? The analysis in Figure 4 excludes 27 outliers. Notice that in Figures 3 and 4 both the x variable and the y variable have had a natural logarithm applied.

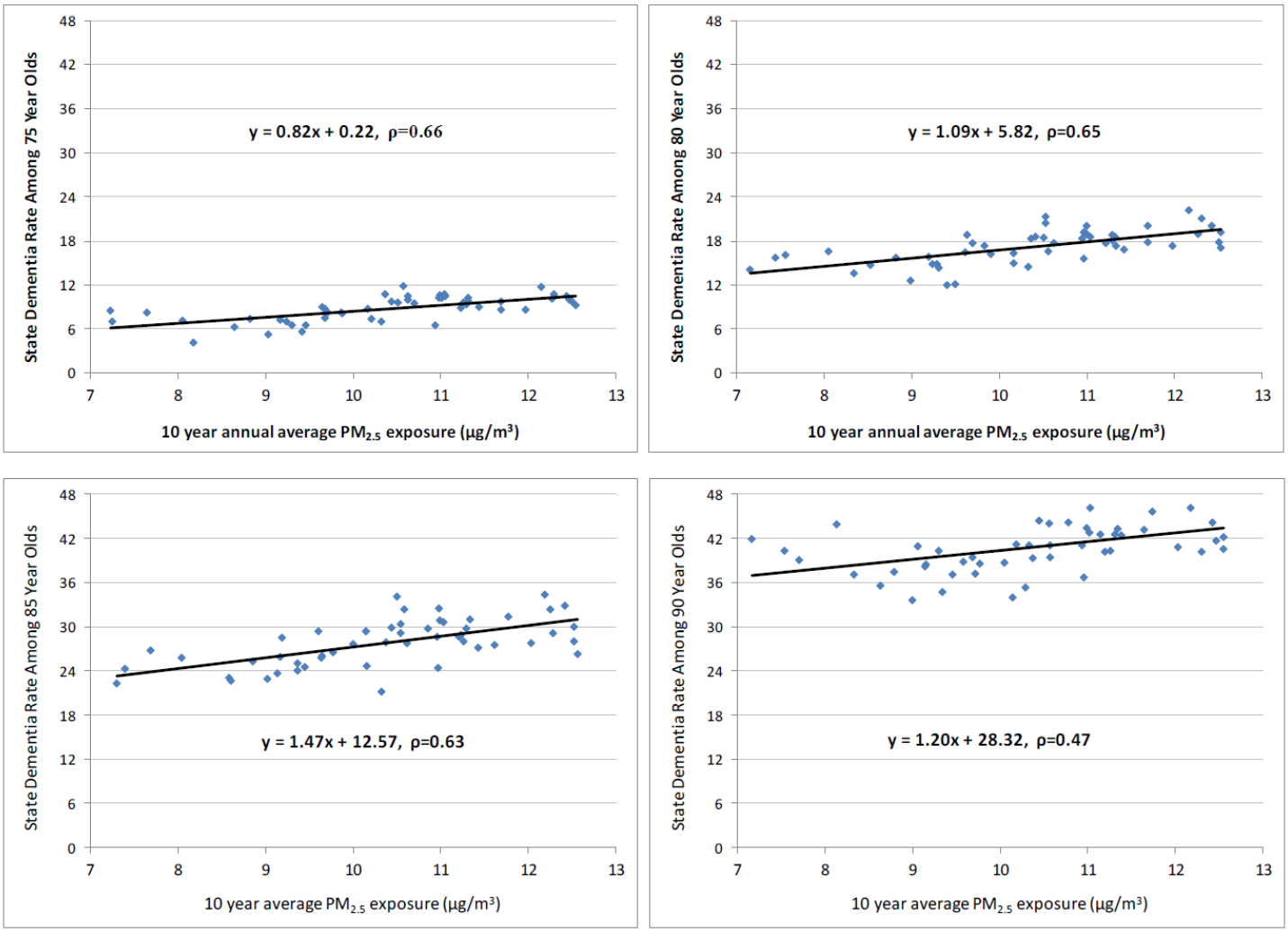




Supplement for Question (7): An August 2018 *NBER Working Paper* "Hazed and Confused: The Effect of Air Pollution on Dementia" (https://www.nber.org/papers/w24970.pdf) examines how cumulative long-term exposure to air pollution – concentration of particulate matter (PM2.5) – affects the chance of being diagnosed with Alzheimer's disease and related dementias later in life.

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 [measured in micrograms per cubic meter, $\mu g/m^3$] over a decade. Figure A1 [shown below] illustrates this association by plotting state-level dementia rates among 75, 80, 85 and 90-year-old individuals in 2013 against their average residential PM2.5 exposures from 2004 through 2013. Correlation coefficients range from 0.47 to 0.66."

FIGURE A1: ASSOCIATION BETWEEN $PM_{2.5}$ AND DEMENTIA AMONG MEDICARE ENROLLEES, 2013



Note: Each data point represents the fraction of individuals living in a state who had been diagnosed with dementia prior to the end of 2013 plotted against their average decadal exposure to PM2.5 based on place of residence. The figures are conditional on integer age: 75 (upper left), 80 (upper right), 85 (lower left) and 90 (lower right). Each figure also shows linear regression equations and correlation coefficients. The figures are based on dementia diagnoses observed for all enrollees in traditional Medicare in 2013.

Supplement for Question (8): Recall the September 30th, 2018 article in *The Toronto Star* "Toronto is segregated by race and income. And the numbers are ugly."

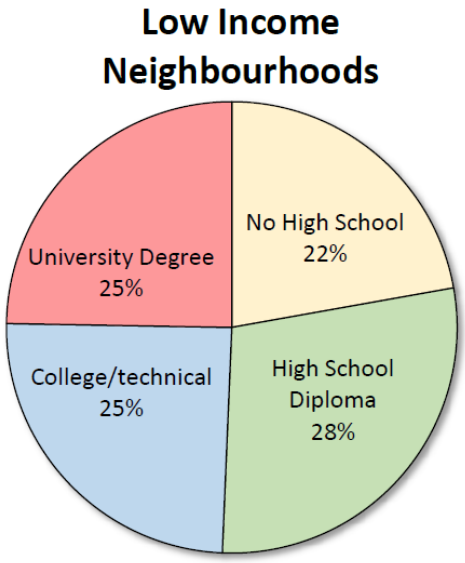
An excerpt of the figure "Educational Attainment of Toronto's Population, 2016" is below.

The figure focuses on Torontonians that are over 15 years old. For that group, we can find the percent that live in each type of neighborhood:

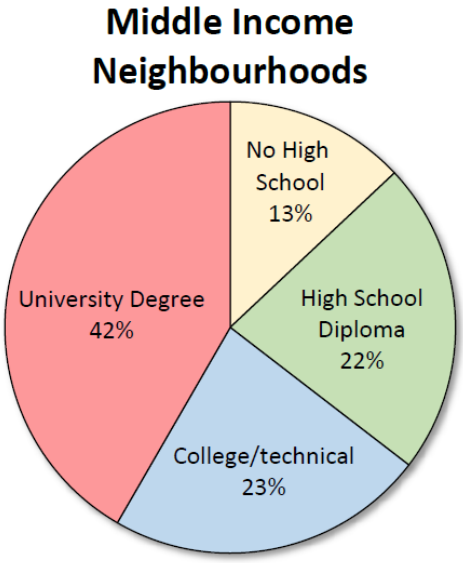
- 50% live in low income neighborhoods
- 29% live in middle income neighborhoods
- 21% live in high income neighborhoods

Define events as follows for a randomly selected Torontonian who is over 15 years old. **When organizing your quantitative analysis with formal notation, your answer must use the definitions below.**

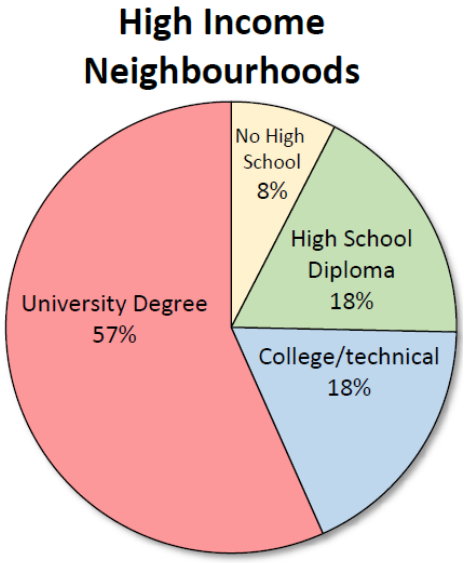
- Event **U** : Has a **university degree**
- Event **U'**: Complement of Event U
- Event **L** : Lives in a **low income** neighborhood
- Event **M** : Lives in a **middle income** neighborhood
- Event **H** : Lives in a **high income** neighborhood



$32,000 average income



$49,000 average income



$102,000 average income