ECO220Y1Y, Test #3, Prof. Murdock

January 18, 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 (some with multiple parts) with varying point values worth a total of 100 points.
- This test includes these 8 pages plus the *Supplement*. The *Supplement* contains the aid sheets (formulas and Normal table) and readings, figures, tables, and other materials required for some test questions. For each question referencing this *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). Leave yourself time to complete all questions rather than overdoing some questions and running out of time.
    - Marking TAs are instructed to accept all reasonable rounding.
    - For questions with multiple parts (e.g. (a) – (c)), ***attempt each part***.
- ***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) [12 pts] See ***Supplement for Question (1)***: *Stats Canada & Financial Data.* A researcher wishes to prove that *less than* 75% of all adult Canadians either oppose or somewhat oppose the access to personal financial data. What can we conclude? Answer with hypotheses in formal notation, the P-value & 1 – 2 sentences *interpreting* the results.

(2) [9 pts] See ***Supplement for Question (2)***: *U.S. Income Distribution.* See T/F/*Explain*. Answer with 2 – 3 sentences.

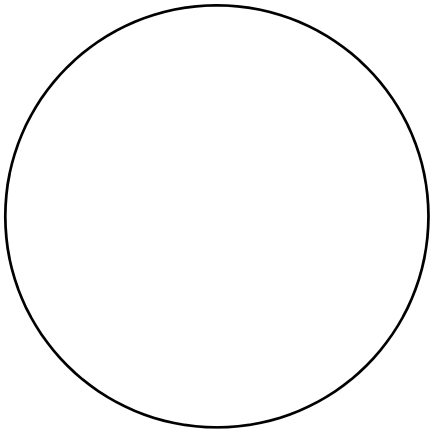
(3) See ***Supplement for Question (3)***: *Toronto's Segregated Immigrant Population.*

(a) [12 pts] For **<u>1981</u>**, complete these pie charts and the three blanks for number of people. <u>Answer by writing values and labels *inside* the pie chart circles like the *Supplement* AND fill in the three blanks. Show your work below.</u>

Toronto's Segregated Immigrant Population, <u>1981</u>

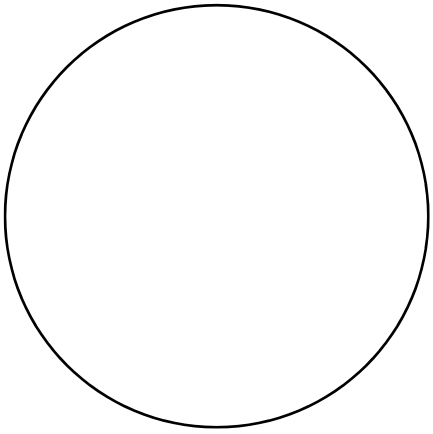
Low Income Neighbourhoods

_____ people



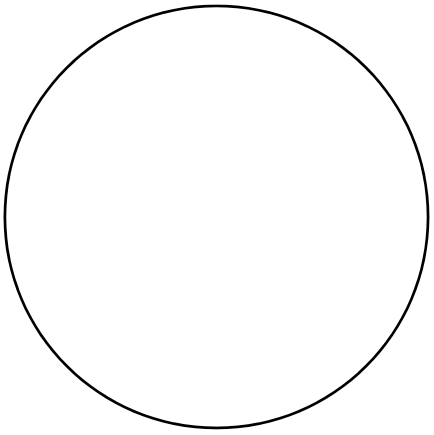
Middle Income Neighbourhoods

_____ people



High Income Neighbourhoods

_____ people



(b) [9 pts] For **2016**, suppose Toronto were *EXACTLY* as in the *Supplement EXCEPT* immigrants were not segregated. Find the 2016 joint probability table with no segregation of immigrants but everything else exactly the same. Answer by filling in the joint probability table below & show your work below.

Joint Probability Table: Toronto's NOT Segregated Immigrant Population, 2016

	L	**M**	**H**
N			
R			
E			

(4) [8 pts] Suppose birth weights are Normally distributed with a mean of 3,147 grams and a standard deviation of 492 grams. What percentage of all newborns either weigh less than 2,400 grams or weigh more than 4,200 grams? Answer with a quantitative analysis & state the answer in percentage terms.

(5) [8 pts] Consider the weight distributions of four types of vehicles: passenger vehicles, light duty trucks, medium duty trucks, and heavy duty trucks. Below are the means and standard deviations (in kilograms) for each type.

	Passenger vehicles	Light duty trucks	Medium duty trucks	Heavy duty trucks
Mean (kg)	2,000	4,500	8,700	16,000
Standard deviation (kg)	800	1,100	1,900	2,500

If there are 17 passenger vehicles, 4 light duty trucks, 2 medium duty trucks, and 3 heavy duty trucks on a bridge, what is the mean and standard deviation of the TOTAL weight on the bridge? (Note: It is reasonable to assume that the weights are unrelated across vehicles on the bridge.) Answer with a quantitative analysis.

(6) See ***Supplement for Question (6)***: *ADHD Diagnoses.*

(a) [16 pts] Comparing the youngest quartile of children (those born in Jun., Jul. or Aug.) with the oldest quartile (those born in Sep., Oct. or Nov.), what is the point estimate of the difference in the rate of ADHD diagnoses? For a 95% confidence level, what are the lower and upper confidence limits of the difference? *Interpret* the results. Answer with a quantitative analysis, a point estimate, confidence interval & 3 – 4 sentences *interpreting* the results.

(b) [12 pts] Comparing children born in March, April or May with those born in December, January, or February yields a 90% confidence interval estimate of [-0.00021, 0.00114], which is based on the rate for March through May minus the rate for December through February. *Interpret* these results. Also, is this 90% CI estimate is surprising given the 95% CI estimate you found in Part (a)? Answer with 3 – 4 sentences. You should **NOT** include a quantitative analysis.

(7) See ***Supplement for Question (7)***: *Monte Carlo Simulation and the 2016 Adult Age Distribution in Canada.*

(a) [7 pts] In the STATA summary, what does the value **29.7** mean? Answer with 1 sentence that *interprets* it & a probability statement in formal notation.

(b) [7 pts] Which of the three histograms shows what you would expect to see if you constructed a histogram to describe a random sample of 100 Canadian adults $(n = 100)$ in 2016? *Explain* the correct *and* incorrect choices. Answer with a histogram choice & 2 – 3 sentences.

This *Supplement* contains the aid sheets (formulas and Normal table) and readings, figures, tables, and other materials required 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:** $r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1}$

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** $V[X] = np(1-p)$

If *X* is Uniform $(X \sim U[a,b])$ **then** $f(x) = \frac{1}{b-a}$ **and** $E[X] = \frac{a+b}{2}$ **and** $V[X] = \frac{(b-a)^2}{12}$

Sampling distribution of $\bar{X}$:

$\mu_{\bar{X}} = E[\bar{X}] = \mu$

$\sigma_{\bar{X}}^2 = V[\bar{X}] = \frac{\sigma^2}{n}$

$\sigma_{\bar{X}} = SD[\bar{X}] = \frac{\sigma}{\sqrt{n}}$

Sampling distribution of $\hat{P}$:

$\mu_{\hat{P}} = E[\hat{P}] = p$

$\sigma_{\hat{P}}^2 = V[\hat{P}] = \frac{p(1-p)}{n}$

$\sigma_{\hat{P}} = SD[\hat{P}] = \sqrt{\frac{p(1-p)}{n}}$

Sampling distribution of $(\hat{P}_2 - \hat{P}_1)$:

$\mu_{\hat{P}_2 - \hat{P}_1} = E[\hat{P}_2 - \hat{P}_1] = p_2 - p_1$

$\sigma_{\hat{P}_2 - \hat{P}_1}^2 = V[\hat{P}_2 - \hat{P}_1] = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}$

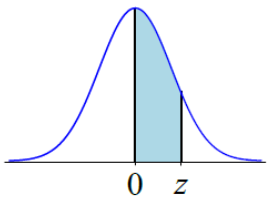
$\sigma_{\hat{P}_2 - \hat{P}_1} = SD[\hat{P}_2 - \hat{P}_1] = \sqrt{\frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}}$

Inference about a population proportion:

***z* test statistic:** $z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ **CI estimator:** $\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$

Inference about comparing two population proportions:

CI estimator: $(\hat{P}_2 - \hat{P}_1) \pm z_{\alpha/2}\sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{n_2} + \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}}$



The Standard Normal Distribution:

	Second decimal place in z									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999

Supplement for Question (1): A November 13, 2018 article by Josh Dehaas posted on CTV News is titled "Majority of Canadians oppose StatCan's plan to access bank data: Nanos survey" (https://www.ctvnews.ca/politics/majority-of-canadians-oppose-statcan-s-plan-to-access-bank-data-nanos-survey-1.4175791).

> **Excerpt from CTV News:** New polling finds that about three-quarters of those surveyed aren't on board with Statistics Canada's temporarily-shelved plan to access the financial transaction data and personal details of 500,000 Canadians without their permission.
>
> The Nanos Research survey of 1,000 adult Canadians conducted between Nov. 3 and Nov. 7 found that 55 per cent of those surveyed oppose the idea, 19 per cent somewhat oppose it, nine per cent are supportive, 14 per cent are somewhat supportive and three per cent are unsure.

Nanos Research publically posts full reports on its website (http://www.nanos.co/our-insight/).

The relevant report, posted on November 14, 2018, is "Canadians choose protecting data privacy over Statistics Canada better understanding consumer behavior and trends" (Submission 2018-1326, http://www.nanos.co/wp-content/uploads/2018/11/2018-1326-StatsCan-and-Privacy-Populated-Report-with-Tabs.pdf).

The following is an excerpt from page 4 of the "STAT SHEET" attached to that report.

Support for access of Statistics Canada to personal financial data

Question	Answer	%
Do you support, somewhat support, somewhat oppose, or oppose Statistics Canada being able to get financial transaction data along with the personal details of Canadians without their permission from financial institutions?	Support	9.1
	Somewhat support	14.4
	Somewhat oppose	19.2
	Oppose	54.7
	Unsure	2.6

Source: Nanos Research, random sample of 1,000 Canadians, 18 years or older, November 3rd to 7th, 2018.

Supplement for Question (2): A table titled "HINC-01. Selected Characteristics of Households, by Total Money Income in 2017" by the United States Census Bureau reports detailed information about the distribution of household income (https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-01.html).

A summary of the 2017 U.S. household income distribution: the median is \$60,000, the 10th percentile is \$15,000, the 25th percentile is \$30,000, the 75th percentile is \$110,000, the 90th percentile is \$180,000, and 8 percent of households make \$200,000 and over.

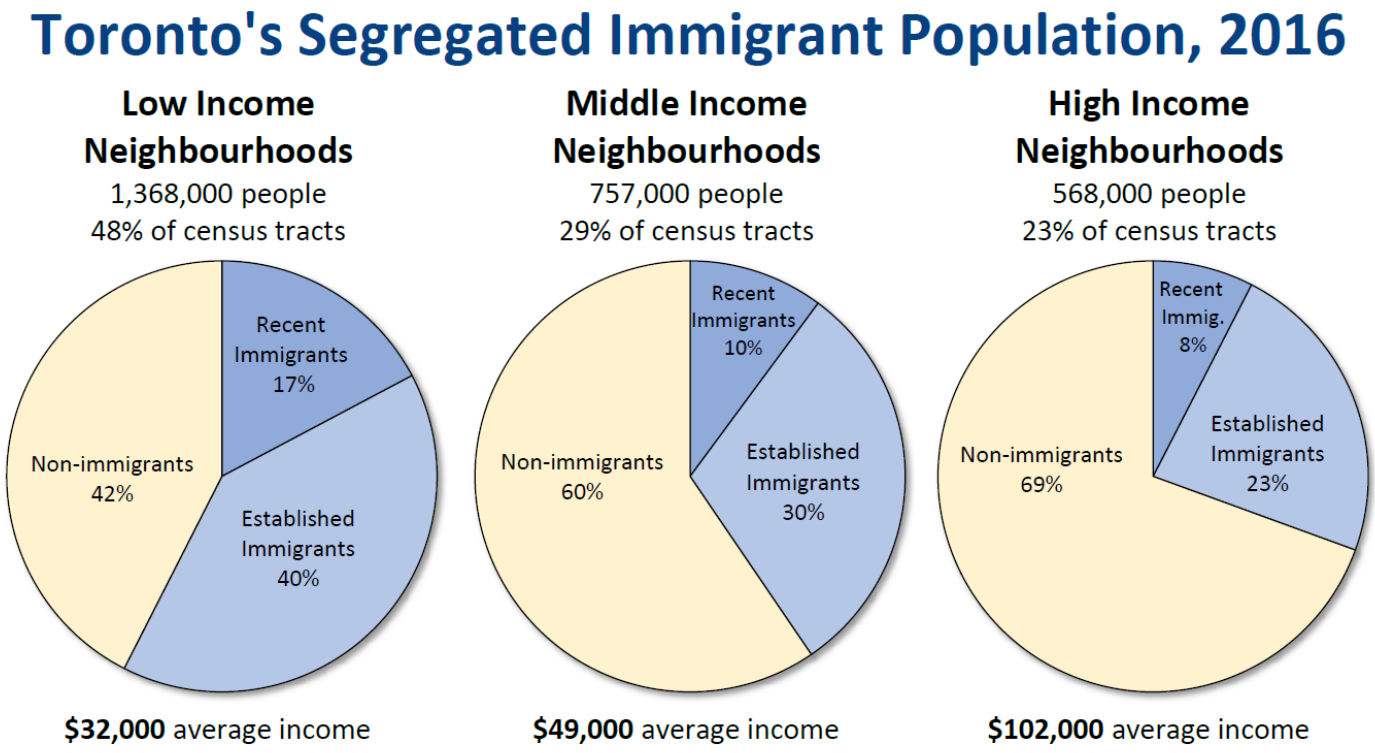
True/False and *Explain*: "For a randomly selected U.S. household in 2017, the chance household income falls between the 10th and 15th household income percentiles *equals* the chance household income falls between the 90th and 95th household income percentiles. However, the chance that household income falls between \$15,000 and \$20,000 *is greater than* the chance household income falls between \$180,000 and \$185,000."

Supplement for Question (3): Recall *one* of the figures appearing in a September 30th, 2018 article in *The Star* titled "Toronto is segregated by race and income. And the numbers are ugly." The figure titled "Toronto's Segregated Immigrant Population, 2016" below shows a **2016** analysis. The ***next page*** shows an analysis for **1981**.

You must use these **definitions of events** in your answers:

- **N** : The event that a randomly selected Torontonian is a **non-immigrant**
- **R** : The event that a randomly selected Torontonian is a **recent immigrant**
- **E** : The event that a randomly selected Torontonian is an **established immigrant**
- **L** : The event that a randomly selected Torontonian lives in a **low income** neighborhood
- **M** : The event that a randomly selected Torontonian lives in a **middle income** neighborhood
- **H** : The event that a randomly selected Torontonian lives in a **high income** neighborhood

Note: Some population numbers are rounded to the nearest 1,000 and others to the nearest 100. This is why they do not sum to the same exact total. This tiny discrepancy – caused by rounding – does not affect the meaning of the figure.



Supplement for Question (3), continues on the next page >>>>>

Supplement for Question (3), cont'd: Below is the joint probability table for the six events (**N**, **R**, **E**, **L**, **M**, and **H**) for **1981**. The population of Toronto was 2.1 million people in 1981. (*Source:* September 2018 presentation by J. David Hulchanski and Richard Maaranen "Neighbourhood Socio-Economic Polarization & Segregation in Toronto Trends and Processes since 1970" http://neighbourhoodchange.ca/documents/2018/09/hulchanski-2018-toronto-segregation-presentation.pdf.)

Joint Probability Table: Toronto's Segregated Immigrant Population, 1981

	L	M	H
N	0.124	0.387	0.081
R	0.053	0.081	0.008
E	0.076	0.165	0.025

Supplement for Question (6): In the article "Attention Deficit-Hyperactivity Disorder and Month of School Enrollment," Layton et al. (2018) study the rate of claims-based ADHD diagnosis (insurance claims). Below are excerpts.

Excerpt from PowerPoint slides that accompany the article:

- In U.S. states that have an age cutoff at September 1 for entry to kindergarten, the rates of diagnosis and treatment of ADHD were higher among children born in August than among those born in September, which suggests that the age within a class cohort influences diagnosis.
- Data came from a large insurance database.

Excerpt from p. 2,122 of the article: Younger children in a school grade cohort may be more likely to receive a diagnosis of attention deficit-hyperactivity disorder (ADHD) than their older peers because of age-based variation in behavior that may be attributed to ADHD rather than to the younger age of the children. Most U.S. states have arbitrary age cutoffs for entry into public school. Therefore, within the same grade, children with birthdays close to the cutoff date can differ in age by nearly 1 year.

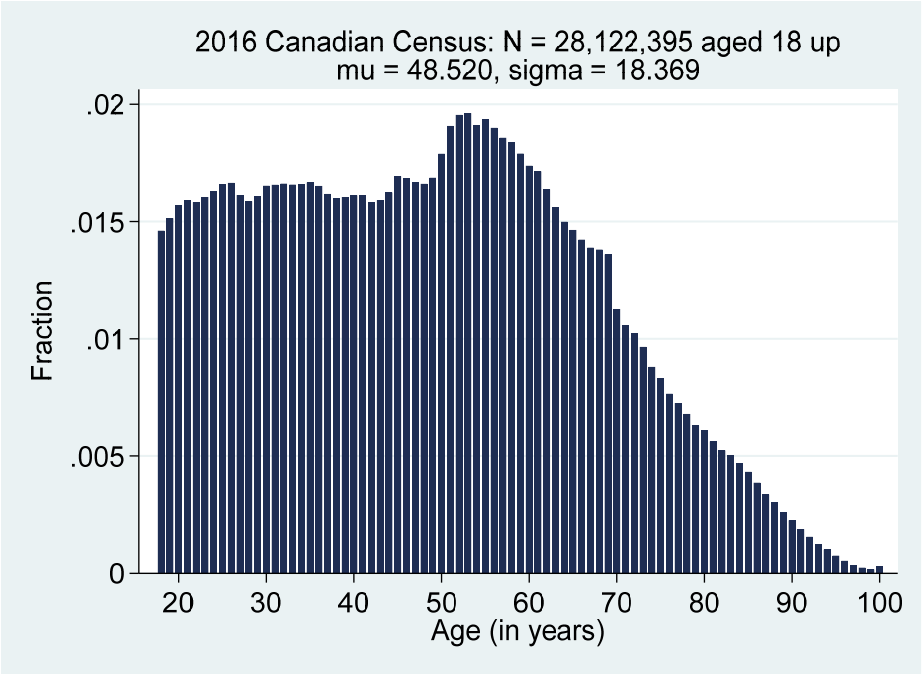
We used data from 2007 through 2015 from a large insurance database to compare the rate of ADHD diagnosis among children born in August with that among children born in September.

The study included 407,846 children in U.S. states [that maintained a September 1 cutoff in 2010 through 2014] who were born in the period from 2007 through 2009 and were followed through December 2015.

Figure 1. Differences in Diagnosis Rates of Attention Deficit-Hyperactivity Disorder (ADHD) According to Month of Birth

	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Total number of children	32,690	31,238	34,405	34,565	34,977	34,415	36,577	36,319	35,353	34,405	31,285	31,617
Number of children with ADHD	265	280	307	312	287	317	320	309	225	240	232	243
Rate per 10,000 children	81.1	89.6	89.2	90.3	82.1	92.1	87.5	85.1	63.6	69.8	74.2	76.9

Supplement for Question (7): Using Statistics Canada, 2016 Census of Population, Catalogue no. 98-400-X2016008, the figure to the right shows the age distribution of adults in the Canadian population.



Below is a STATA summary of Monte Carlo simulation results for the sampling distribution of the sample mean for a sample size of 10. It uses 10,000 simulation draws to obtain the simulated sampling distribution of the sample mean. One value below is in boldface to make it easier to find the specific value that one part asks about.

At the bottom are some histogram choices for the last part of the question.

STATA summary of the Monte Carlo simulation results for X-bar for n = 10:

X-bar

Percentiles		Smallest		
1%	35.5	29.2		
5%	38.9	**29.7**		
10%	41	29.9	Obs	10,000
25%	44.5	29.9	Sum of Wgt.	10,000
50%	48.5		Mean	48.50512
		Largest	Std. Dev.	5.821651
75%	52.5	67.6		
90%	56	67.7	Variance	33.89162
95%	58	68.2	Skewness	.0370604
99%	62.1	70.6	Kurtosis	2.848827

Three histogram choices for Part (b) of Question (7):