ECO220Y1Y, Test #3, Prof. Murdock

January 17, 2020, 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 9 questions (some 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, a statistical table (Standard Normal), readings, figures, tables, and other materials required for some 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 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) [6 pts] See ***Supplement for Question (1)***: *Financial Wellbeing.* On each line, write the probability in **formal notation** using the **definitions of events in the *Supplement***, give its numeric value, and show any work.

(a) The chance of observing high financial wellbeing for someone reporting low financial wellbeing:

_____

(b) The chance someone has high reported financial wellbeing:

_____

(c) The chance someone has neither low reported financial wellbeing nor low observed financial wellbeing:

_____

(2) A continuous index from 0 to 100 measures a person's overall health. Higher values mean better health. Suppose the health index is Normally distributed across people with a mean of 58.53 and a standard deviation of 14.44.

(a) [3 pts] What percent of people have a health index above 50? Answer with a quantitative analysis.

(b) [3 pts] Ninety percent of people have a health index above which value? Answer with a quantitative analysis.

(3) [14 pts] See ***Supplement for Question (3)***: *Decile Income Transition Matrix in Canada.* Consider a random sample of 100 people in the 1st income decile in 2007 (Sample 1) and another random sample of 100 people in 2nd income decile in 2007 (Sample 2). What is the chance there are more upwardly mobile people in Sample 2 than in Sample 1? Before computing it, explain why this probability should be large or small, and explain the key factors affecting its size. Next, compute it. (Hint: Define the random variable X_1 to be the number of upwardly mobile people in Sample 1 and define the random variable X_2 to be the number of upwardly mobile people in Sample 2.) Answer with 2 – 3 sentences & a quantitative analysis that shows your work and reasoning.

(4) [5 pts] Elementary school students are tested for eligibility for gifted programs. In Ontario, it begins with the CCAT-7 test written by all third graders. For the percentile scores on the CCAT-7 test for a random sample of two students $(n = 2)$, what is the sampling distribution of the sample mean? Answer with a fully labelled graph.

(5) See ***Supplement for Question (5)***: *Start-Up Firms in Norway.*

(a) [2 pts] What is the overall fraction of start-ups surviving at 5 years among the last two categories combined (i.e. combining "**Non-University PhD**" and "**University**")? Answer with a quantitative analysis.

(b) [6 pts] For "**Non-University PhD**," what is the 75% confidence interval estimate of the fraction of start-ups surviving at 5 years? Answer with a quantitative analysis. (You are *not* asked for an interpretation.)

(c) [15 pts] Comparing "**Non-University PhD**" and "**University**", what is the point estimate of the difference in the fraction surviving at 5 years? For a 99% confidence level, what are the lower and upper confidence limits of the difference? *Interpret* the results, including discussion of the magnitude of both the point estimate and the width of the CI. Answer with a quantitative analysis, confidence interval & 3 – 4 sentences *interpreting* the results.

(6) [6 pts] Consider the weight-for-age percentile scores for children under five years of age. For a random sample of twenty-two children in the first, second, or third deciles, what is the probability that the mean percentile score is above 14? Answer with a quantitative analysis that shows your work and reasoning.

(7) ***Supplement for Question (7)***: *Ontario Disclosure of University and College Sector Salaries.*

(a) [7 pts] In **Summary #2** note the 75^{th} percentile is 160157.8. For a random sample of 2,000 employees (rather than 25 employees), we should expect the 75^{th} percentile to be larger or smaller than $160,157.8? Why or why not? Include all relevant evidence that directly supports your assessment. Answer with 2 – 3 sentences.

(b) [6 pts] A sample size of 25 is not sufficiently large for the Central Limit Theorem (CLT), and the sampling distribution of the sample mean is positively skewed. If a sample size of 25 were sufficiently large such that the CLT applied, then which value should appear as the 99th percentile in **Summary #3**? Answer with a quantitative analysis.

(c) [7 pts] Is sampling error a plausible explanation for why \$146,968.9 (the mean in **Summary #2**) is more than \$5,000 above \$141,859.8 (the mean in **Summary #1**)? Why or why not? Include all relevant evidence that directly supports your assessment. (Also, identify what \$146,968.9 and \$141,859.8 measure.) Answer with 2 – 3 sentences.

(8) [10 pts] Recall "Sex ratios among Canadian liveborn infants of mothers from different countries" from 2012 in the *Canadian Medical Association Journal* (DOI:10.1503 /cmaj.120165). Recall also that absent interference, 51.2 percent of infants are male. For the third child of moms born in Pakistan, there are 2,081 males and 1,926 females. Defining p as the proportion male, test $H_0: p = 0.512$ versus $H_1: p > 0.512$ by computing the P-value. How strong is the evidence in favor of H_1? What should we conclude about sex selection for this subgroup? Explain the conclusion in this specific context. Answer with a quantitative analysis & 1 – 2 sentences.

(9) [5 pts] A researcher wants to prove a majority (over half) of British citizens favour Brexit. In a sample of 1,000, 520 favor Brexit yielding a P-value of 0.1030. Fill in the blanks with "smaller than," "larger than," or "equal to."

(a) For a sample of 100 people where 52 favor Brexit, the P-value will be _____ 0.1030.

(b) For a sample of 1,000 people where 521 favor Brexit, the P-value will be _____ 0.1030.

(c) For a sample of 10,000 people where 5,200 favor Brexit, the P-value will be _____ 0.1030.

(d) To prove $H_1: p \neq 0.5$ where 520 (of 1,000) favor Brexit, the P-value will be _____ 0.1030.

(e) To prove $H_1: p < 0.5$ where 480 (of 1,000) favor Brexit, the P-value will be _____ 0.1030.

This *Supplement* contains the aid sheets, a statistical table (Standard Normal), readings, figures, tables, and other materials for some 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:	**Laws of variance**:	**Laws of covariance**:
$E[c] = c$	$V[c] = 0$	$COV[X,c] = 0$
$E[X+c] = E[X] + c$	$V[X+c] = V[X]$	$COV[a+bX, c+dY] = bd * COV[X,Y]$
$E[cX] = cE[X]$	$V[cX] = c^2 V[X]$	
$E[a+bX+cY] = a + bE[X] + cE[Y]$	$V[a+bX+cY] = b^2V[X] + c^2V[Y] + 2bc * COV[X,Y]$	
	$V[a+bX+cY] = b^2V[X] + c^2V[Y] + 2bc * SD(X) * SD(Y) * \rho$ where $\rho = CORRELATION[X,Y]$	

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** $V[X] = \frac{(b-a)^2}{12}$

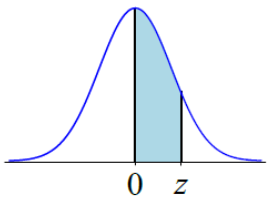
Sampling distribution of $\bar{X}$:	**Sampling distribution of $\hat{P}$**:	**Sampling distribution of $(\hat{P}_2 - \hat{P}_1)$**:
$\mu_{\bar{X}} = E[\bar{X}] = \mu$	$\mu_{\hat{P}} = E[\hat{P}] = p$	$\mu_{\hat{P}_2-\hat{P}_1} = E[\hat{P}_2 - \hat{P}_1] = p_2 - p_1$
$\sigma_{\bar{X}}^2 = V[\bar{X}] = \frac{\sigma^2}{n}$	$\sigma_{\hat{P}}^2 = V[\hat{P}] = \frac{p(1-p)}{n}$	$\sigma_{\hat{P}_2-\hat{P}_1}^2 = V[\hat{P}_2 - \hat{P}_1] = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}$
$\sigma_{\bar{X}} = SD[\bar{X}] = \frac{\sigma}{\sqrt{n}}$	$\sigma_{\hat{P}} = SD[\hat{P}] = \sqrt{\frac{p(1-p)}{n}}$	$\sigma_{\hat{P}_2-\hat{P}_1} = SD[\hat{P}_2 - \hat{P}_1] = \sqrt{\frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}}$

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



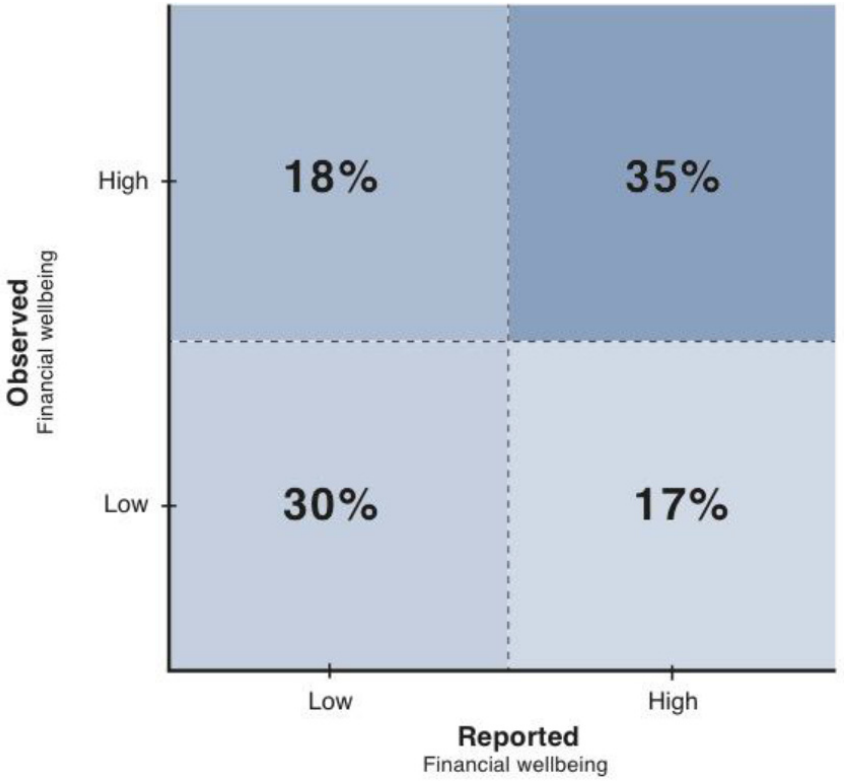
The Standard Normal Distribution:

	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): The figure to the right is from a March 2018 technical report: "Using Survey and Banking Data to Measure Financial Wellbeing" prepared by the Commonwealth Bank of Australia and the Melbourne Institute. On the y-axis is objectively observed financial wellbeing based on banking data. On the x-axis is a person's subjective sense of financial wellbeing based on survey data.

You **MUST USE** the following definitions of events in your answers:

- **OL**: Observed financial wellbeing is low
- **OH**: Observed financial wellbeing is high
- **RL**: Reported financial wellbeing is low
- **RH**: Reported financial wellbeing is high



Supplement for Question (3): 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.

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): Consider Table 1A (shown below) from the 2018 paper "University Innovation and the Professor's Privilege" by Hvide and Jones in the *American Economic Review* (https://doi.org/10.1257/aer.20160284).

Excerpt from page 1,868: Table 1A provides summary statistics for start-up firms in Norway between 2000 and 2007. In total there were 48,844 start-ups and 128 of these were started up by individuals with PhDs employed at a university. We define a university start-up as a newly incorporated company where at least one of the initial owners is a fulltime university employee with a PhD. By comparison, there were 452 start-ups by individuals with PhDs who were not employed at universities.

TABLE 1A—SUMMARY STATISTICS FOR START-UP FIRMS IN NORWAY, 2000–2007

		All	Non-University PhD	University
Number of start-ups		48,844	452	128
Fraction surviving at 5 years	Mean	0.74	0.83	0.87
Sales at 5 years	Mean	5,160	2,308	2,659
	(Standard deviation)	(13,282)	(4,777)	(9,934)
	Median	1,751	628	183
	75th percentile	4,834	2,210	1,550
	95th percentile	20,769	10,815	9,374
Employees at 5 years	Mean	3.31	1.68	1.22
	(Standard deviation)	(7.77)	(3.16)	(2.89)
	Median	1	1	0
	75th percentile	4	2	1
	95th percentile	13	7	5
Profits at 5 years	Mean	198	220	100
	(Standard deviation)	(554)	(599)	(600)
	Median	43.1	41.2	−6.50
	75th percentile	283	296	215
	95th percentile	1,358	1,555	1,555

Notes: Sales, employees, and profits are conditional on survival at year 5. Profits and sales are measured in 1,000 NOK.

Supplement for Question (7): Recall the 2017 Ontario public disclosure data that includes *all* public sector employees earning $100,000 CAN or more in the 2016 calendar year (https://www.ontario.ca/page/public-sector-salary-disclosure-2016-all-sectors-and-seconded-employees). Consider the university and college sectors, which includes 23,319 employees. See the STATA summaries below.

Summary #1: This summarizes all employees in the university and college sectors.

Salary

	Percentiles	Smallest		
1%	100324.5	100000		
5%	102333.8	100000		
10%	104646.1	100000.1	Obs	23,319
25%	109556.6	100003.5	Sum of Wgt.	23,319
50%	131769.3		Mean	141859.8
		Largest	Std. Dev.	41435.85
75%	161838	504907		
90%	190319.1	512215.6	Variance	1.72e+09
95%	213515.1	514665.3	Skewness	2.479115
99%	288628.9	1045583	Kurtosis	20.26363

Summary #2: This summarizes a random sample of 25 employees in the university and college sectors.

Salary

	Percentiles	Smallest		
1%	100639.2	100639.2		
5%	101760.1	101760.1		
10%	102622.5	102622.5	Obs	25
25%	105859.9	103199.9	Sum of Wgt.	25
50%	138586.3		Mean	146968.9
		Largest	Std. Dev.	57032.7
75%	160157.8	194648		
90%	200131.5	200131.5	Variance	3.25e+09
95%	213557.5	213557.5	Skewness	2.300223
99%	363821	363821	Kurtosis	9.351374

Summary #3: This summarizes the results of a simulation to obtain the simulated sampling distribution of the sample mean for a sample size of 25 ($n = 25$) using 50,000 simulation draws ($m = 50{,}000$).

Mean salary (X-bar)

	Percentiles	Smallest		
1%	125257.8	116134.1		
5%	129360.3	117010.4		
10%	131725.5	117088.1	Obs	50,000
25%	136101	117321.4	Sum of Wgt.	50,000
50%	141328.7		Mean	141909.8
		Largest	Std. Dev.	8323.163
75%	147051.5	189569.9		
90%	152769	190254.4	Variance	6.93e+07
95%	156395.8	191217	Skewness	.5074134
99%	163964.3	198280.2	Kurtosis	3.726959