ECO220Y1Y, Test #2, Prof. Murdock

December 3, 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 9 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, statistical table (Standard Normal), 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, a fully labelled graph, and/or sentences. Any answer guide asking for a quantitative analysis *always* automatically means that you must show your work and make your reasoning clear.
    - 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] For parking permits in a city, 33% of households have no permit, 49% have one permit, and 18% have two permits. What is the mean and standard deviation? Answer with a quantitative analysis.

(2) [6 pts] Recall "Toronto is segregated by race and income." Unlike Toronto, City X has no segregation. In City X 19% of the population lives in a high-income neighborhood. In City X 45% of the population is white. What is the joint probability table? To answer: define events, show your work & fill in the table putting *events* in the grey cells.

Joint Probability Table for City X

(3) [4 pts] A fair die has six sides with the values 1, 2, 3, 4, 5, and 6, where each is equally likely. Two scenarios:

- **Scenario 1:** Roll one fair die and get one hundred times the value, which gives the payout in dollars
- **Scenario 2:** Roll one hundred fair dice and get the sum of the values, which gives the payout in dollars

In Scenario 2 the expected value of the payout is _____ (higher, lower, the same) compared to Scenario 1. In Scenario 2 the standard deviation of the payout is _____ (higher, lower, the same) compared to Scenario 1. The payout in Scenario 1 follows a _____ distribution. The payout in Scenario 2 follows a _____ distribution. Fill in the blanks.

(4) There is a 51.2% chance of giving birth to a male infant (if there is no sex selection associated with the pregnancy).

(a) [4 pts] In a hospital with 1,000 infants born, what is the chance that more than 50% are male? Answer with a quantitative analysis.

(b) [3 pts] Imagine redoing the previous part, but for a hospital with 10,000 infants born. In this context and *without* focusing on calculations, *explain why* the answer would be larger, smaller, or the same. Answer with 1 – 2 sentences.

(c) [4 pts] In a hospital with 10,000 infants born, what is the chance that more than 52% are male? Answer with a quantitative analysis.

(d) [3 pts] Imagine redoing the previous part, but for a hospital with 1,000 infants born. In this context and *without* focusing on calculations, *explain why* the answer would be larger, smaller, or the same. Answer with 1 – 2 sentences.

(5) See ***Supplement for Question (5)***: *Return on Student Loans in Canada.*

(a) [10 pts] To add to the *Supplement*, note that for all institutions *outside* the top ten the overall rate of default within the first three years is 7%. Among those borrowers who default within the first three years, what fraction attend top-ten ranked institutions? Use a probability tree. Answer with a probability tree & a quantitative analysis.

(b) [6 pts] Next, *without* doing another whole analysis, *explain why* the answer to the question in the previous part would be higher, lower, or the same if instead of top-ten ranked institutions it asked about unranked institutions. Answer with 2 – 3 precise sentences that integrate the relevant numeric support.

(6) See ***Supplement for Question (6)***: *The Globalization of Postsecondary Education.* Further, define events as:

- **Event H:** International student from China (*Note:* Use "C" for Canada); **Event I:** International student from India; **Event S:** International student from South Korea
- **Event A:** Attends in Australia; **Event C:** Attends in Canada; **Event U:** Attends in US; **Event K:** Attends in UK

The entire Table 1 is conditional on international students from China, India, and South Korea who study in Australia, Canada, the UK, and the US. Of course, students from other countries, like the Ivory Coast, attend in other countries, like France. You do *not* need to condition on there being only three source countries and four destination countries. In other words, for the purposes of all parts below, ignore all other countries not in Table 1.

(a) [3 pts] In 2000, for a randomly selected international student from South Korea, what is the chance that they study in the US? Answer with a probability statement using formal notation & a quantitative analysis.

(b) [3 pts] In 2010, for a randomly selected international student who is attending in Canada, what is the chance they are from India? Answer with a probability statement using formal notation & a quantitative analysis.

(c) [3 pts] In 2017, for a randomly selected international student, what is the chance they are from South Korea and study in Canada? Answer with a probability statement using formal notation & a quantitative analysis.

(d) [6 pts] Compared to 2000, is Canada's share of international students from China and India larger in 2017? Answer with 2 – 3 precise sentences that integrate the relevant numeric support.

(7) See ***Supplement for Question (7)***: *Why Have College Completion Rates Increased? An Analysis of Rising Grades.*

(a) [8 pts] In a random sample of 10 students in 1988, what is the chance that more than two students have a GPA of 3.0 or higher? Answer with a quantitative analysis.

(b) [4 pts] Imagine the previous part asked about 2002. In this context and *without* focusing on calculations, *explain why* the answer would be larger, smaller, or the same. Answer with 1 – 2 precise sentences.

(8) [12 pts] See ***Supplement for Question (8)***: *Young Adults Living in the Parental Home in Canada, by Census year.* For a random sample of 1,000 young adults in 2016 that are aged 25 to 29, draw a fully labelled graph illustrating the distribution of the number living with their parents. Next, use the Normal approximation to find the probability that exactly 288 live with their parents. Illustrate that probability on the same graph. Answer with a quantitative analysis & a fully labelled graph.

(9) See ***Supplement for Question (9)***: *Disclosed 2019 Salaries of the Ontario Provincial Police (OPP).*

(a) [4 pts] What number should appear next to "Std. Dev." in the STATA summary? Answer with a quantitative analysis & 1 sentence.

(b) [6 pts] If a sample size of 36 were *sufficiently large*, then what *should* be the value reported next to "99%"? Why is there a discrepancy with the reported number? Answer with a quantitative analysis & 2 – 3 sentences.

This *Supplement* contains the aid sheets, statistical table (Standard Normal), 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:** $r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1}$

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 $\bar{X}$:

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

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

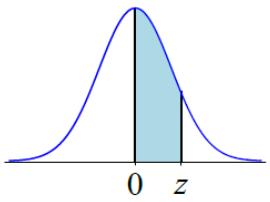
$\sigma_{\bar{X}} = SD[\bar{X}] = \frac{\sigma}{\sqrt{n}}$

Sampling distribution of $\hat{P}$:

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

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

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

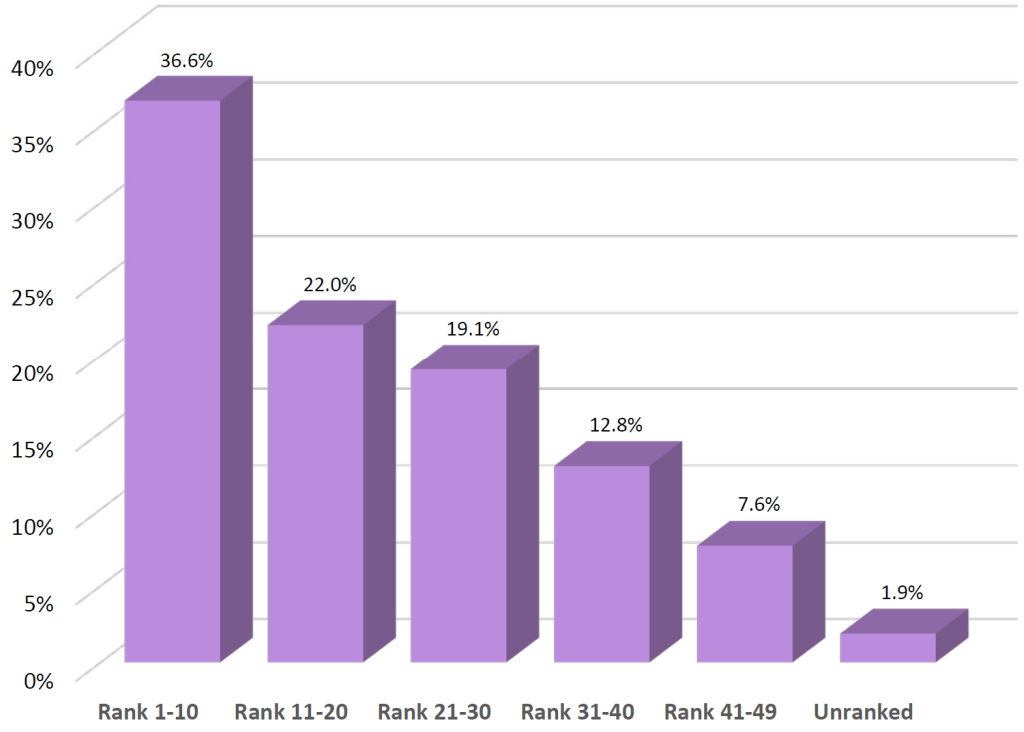


The Standard Normal Distribution:

	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 (5): Consider "Return on Student Loans in Canada" (https://www.nber.org/papers/w29130), a 2021 working paper. It uses administrative data for borrowers from the *Canada Student Loan Program*. Below are excerpts from Figure 1 and Table 2. "Figure 1(c) reports the fraction of borrowers attending universities by their Macleans ranking." (p. 12). **For Table 2, focus on Column (4) of the results: "Default within First 3 Years."**

Figure 1: Descriptive Statistics



(c) Distribution of institutional ranking

Table 2: Average Return, IRR, Early Default, and Early RAP Enrolment by Fields of Study and Institution Ranking

Population Subgroup	Realized Returns Mean	Realized Returns Std. Error	IRR	Default within First 3 Years	RAP within First 3 Years
Field of Study					
Administration/Business	-2.4%	0.3%	2.2%	3%	27%
Agriculture	0.4%	0.7%	2.9%	1%	20%
Arts/Sciences	-8.9%	0.2%	1.0%	8%	36%
Community Services/Educ.	-1.6%	0.4%	2.5%	3%	28%
Engineering	-2.7%	0.3%	2.2%	3%	21%
Health Sciences	0.8%	0.3%	2.9%	2%	14%
Institution Ranking					
Rank 1–10	-4.6%	0.2%	1.7%	4%	25%
Rank 11–20	-3.3%	0.2%	2.1%	5%	28%
Rank 21–30	-7.9%	0.3%	1.2%	7%	37%
Rank 31–40	-6.9%	0.4%	1.4%	8%	34%
Rank 41–49	-8.9%	0.5%	1.2%	8%	37%
Unranked Universities	-11.2%	1.2%	0.8%	10%	38%

Notes: Columns (1) and (2) report the average realized return and its standard error (in %), respectively. Column (3) reports the IRR (in %) on all loans disbursed to the listed subgroup. Columns (4) and (5) report the percent of borrowers who ever defaulted or enrolled in RAP (or IR), respectively, during the first 3 years of repayment.

Supplement for Question (6): Consider "The Globalization of Postsecondary Education: The Role of International Students in the US Higher Education System" (https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.35.1.163), a 2021 article. Below are an excerpt and Table 1. In Table 1 there are 158,409 students in 2000, 551,842 in 2010, and 931,394 in 2017.

> **Excerpt, p. 166:** The increase in international students is not a uniquely US-centric phenomenon. Colleges and universities in Australia, Canada, and the United Kingdom (UK) also experienced a rapid increase in the enrollment of students from China and India since 2000, as shown in Table 1. Although the United States (US) remains the largest destination country for students from these countries, the US higher education system is no longer as dominant as it was 20 years ago.

Table 1

International Students Enrolled in Post-Secondary Institutions by Destination Country

Year	*2000*	*2010*	*2017*
Panel A. Students from China			
Australia	5,008	87,588	128,498
Canada	4,701	26,298	66,161
United Kingdom	6,158	55,496	96,543
United States	50,281	126,498	321,625
Panel B. Students from India			
Australia	4,578	20,429	51,976
Canada	969	5,868	32,616
United Kingdom	3,962	38,205	16,421
United States	39,084	103,968	142,618
Panel C. Students from South Korea			
Australia	2,361	7,311	8,316
Canada	1,116	4,320	5,277
United Kingdom	2,165	4,347	5,157
United States	38,026	71,514	56,186

Source: UNESCO Institute for Statistics 2018.
Note: Numbers depict total number of international students enrolled across all degree statuses (undergraduate and graduate) in 2000, 2010, and 2017.

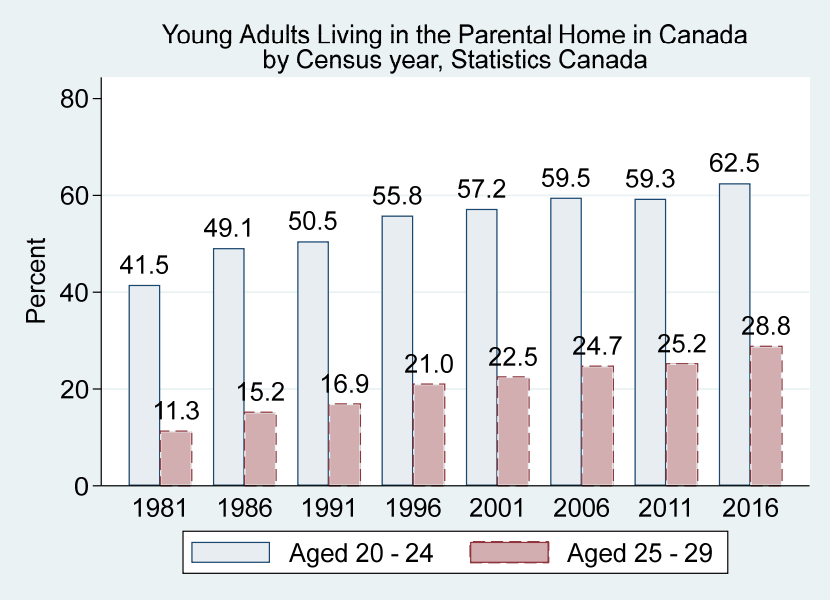
Supplement for Question (7): Consider "Why Have College Completion Rates Increased? An Analysis of Rising Grades" (https://www.nber.org/papers/w28710), a 2021 working paper. Below, Table A7 contains a description of the cumulative distribution of GPA, by year, for undergraduate students in the United States.

Table A7: Description of the Cumulative Distribution of GPA, by Year

	GPA Thresholds							
Year	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
1988	0.07	0.14	0.26	0.47	0.70	0.89	0.99	1.00
2002	0.06	0.12	0.20	0.36	0.56	0.81	0.96	1.00

Note: For each year (1988 and 2002), this shows the fraction of students with a GPA below each of the thresholds separately.

Supplement for Question (8): Recall the figure below.



Supplement for Question (9): Consider the 2020 Ontario disclosure data with all public sector employees earning $100,000 CAN or more in the 2019 calendar year. A total of 5,062 employees of the Ontario Provincial Police (OPP) are in the disclosure, with a mean salary of $124,638.3 and a standard deviation of $17,959.62. For the OPP, the STATA summary below shows the results of a simulation to find the sampling distribution of the sample mean for a sample size of 36 $(n = 36)$ using 10,000 simulation draws $(m = 10{,}000)$. Four numbers in the STATA summary have been intentionally hidden with grey boxes.

Mean salary (X-bar)

Percentiles		Smallest		
1%	118308	115677.5		
5%	119987.7	115842.9		
10%	120918.6	116067.4	Obs	10,000
25%	122593.5	116076.4	Sum of Wgt.	10,000
50%	124571.1		Mean	124683.7
		Largest	Std. Dev.	
75%	126616	136221.1		
90%	128632.6	136500.2	Variance	
95%	129829.9	137143.5	Skewness	
99%	131991.6	137452.9	Kurtosis	