ECO220Y1Y, Test #2, Prof. Murdock

December 2, 2022, 9:10 – 11:00 am

U of T E-MAIL:	**@MAIL.UTORONTO.CA**
SURNAME (FAMILY 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 5 questions (all 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 statistical table (Standard Normal) and readings, figures, tables, and other materials for *some* test questions. (A question may give all needed background directly.) 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) Suppose birth weights are Normally distributed with a mean of 3,148 grams and a standard deviation of 435 grams.

(a) [3 pts] What is the chance that a randomly selected newborn weighs between 3,000 and 3,400 grams? Answer with a quantitative analysis.

(b) [3 pts] A newborn who weighs 3,975 grams is at which percentile? Answer with a quantitative analysis.

(c) [3 pts] What is the weight in grams of a newborn who is at the 14^{th} percentile? Answer with a quantitative analysis.

(2) See ***Supplement for Question (2)***: *2022 Ontario Public Sector Salary Disclosure.*

(a) [6 pts] What is the probability that a randomly selected employee in the 2022 disclosure has a salary between the 73^{rd} and 88^{th} percentiles? Illustrate this probability with an appropriate graph. Answer with a fully labelled graph.

(b) [8 pts] Is sampling error a *plausible* explanation for the sample median in **Summary #2** being as small as it is? Explain and substantiate your claims with relevant evidence from Summaries #1 to #3. Answer with 2 – 3 sentences.

(c) [8 pts] For which scenario would the 5^{th} percentile value change the most? *Explain*. Answer with 2 – 3 sentences.

Scenario A: In Summary #2, the effect on 101.6699 from changing the sample size (n) from 40 to 400

Scenario B: In Summary #3, the effect on 106.3143 from changing the sample size (n) from 40 to 80

Scenario C: In Summary #3, the effect on 106.3143 from changing the simulation draws (m) from 10,000 to 1,000

(3) See ***Supplement for Question (3)***: *Stats Canada and 2021 Employment Rates by Age and Education.*

(a) [2 pts] Given Table 14-10-0020-01, amongst people whose highest educational attainment is graduating high school, what fraction are aged 25 to 54 years? Answer with a quantitative analysis.

(b) [11 pts] In Canada in 2021, for a random sample of 500 people aged 15 years and up, is it *plausible* that over 45 percent of them are 55 years old and up? Next, *explain* in this *context.* Answer with a quantitative analysis & 2 sentences.

(c) [11 pts] In Canada in 2021, for a random sample of five people that are aged 25 to 54 and have above a bachelor's degree, is it *plausible* that less than four are employed? Next, *explain* in this *context.* Answer with a quantitative analysis & 2 sentences.

(4) See ***Supplement for Question (4)***: *Legacy and Athlete Preferences at Harvard.*

(a) [8 pts] In Table 1, *interpret* the difference between 4.89 and 7.58. It is started for you. Answer with 2 – 3 sentences.

Using the complete Harvard University applications and admissions data for students who would, after four years, have graduated university from 2014 to 2019,

(b) [6 pts] Consider all information in the *Supplement* for Question (4) and carefully read each sentence below. Answer by filling in each blank with the correct percent value, rounding to the nearest integer.

For a random sample of 10 white legacy applicants, the expected percent admitted is ________ with a standard deviation of ________, which reflects how much sampling error may affect the percent for a sample.

For a random sample of 20 admitted students, the expected percent that are recruited athletes is ________ with a standard deviation of ________, which reflects how much sampling error may affect the percent for a sample.

For a random sample of 30 white admitted students, the expected percent that are non-ALDC is ________ with a standard deviation of ________, which reflects how much sampling error may affect the percent for a sample.

(5) See ***Supplement for Question (5)***: *Are Effective Teachers for Students with Disabilities Effective Teachers for All?*

(a) [9 pts] Focus on **Panel A** of Table 3 and define events $\boldsymbol{DQ1}$, $\boldsymbol{DQ2}$, …, $\boldsymbol{DQ5}$ as being in the first through fifth quintiles, respectively, for students with disabilities (SWD). Similarly, define events $\boldsymbol{NQ1}$, $\boldsymbol{NQ2}$, …, $\boldsymbol{NQ5}$ for non-SWD.

- For a teacher-year in the fourth quintile of the math VAM for students without disabilities, what is the chance of being in the second quintile of the math VAM for students with disabilities? Answer with a quantitative analysis writing probabilities in formal notation.

- For a teacher-year in the second quintile of the math VAM for students with disabilities, what is the chance of being in the fourth quintile of the math VAM for students without disabilities? Answer with a quantitative analysis writing probabilities in formal notation.

- In general, is $P(A \mid B) = P(B \mid A)$? Explain what is happening in this context. Answer with 2 sentences.

(b) [3 pts] What would Table 3 look like if, for both math and English, the value-added measure for students with disabilities were *independent* of the value-added measure for students without disabilities? Answer with 1 sentence.

(c) [7 pts] In Panel B of Table 3, find the number 1.4%. *Interpret* it. It is started for you. Answer with 2 – 3 sentences.

Using student- and teacher-level administrative data from the huge Los Angeles Unified School District for the academic years 2007-2008 through 2017–2018,

(d) [7 pts] From Table 3, give one example of two events that are *mutually exclusive* (i.e. disjoint). Next, explain if these same two events are *independent*, making sure to apply the definition of independence. Answer with 2 – 3 sentences.

This *Supplement* contains the aid sheets and statistical table (Standard Normal) and 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:** $s_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{n-1} = \frac{\sum_{i=1}^{n} x_i y_i}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n(n-1)}$

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

<u>Laws of expected value</u>:

$E[c] = c$

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

$E[cX] = cE[X]$

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

<u>Laws of variance</u>:

$V[c] = 0$

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

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

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

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

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

<u>Laws of covariance</u>:

$COV[X,c] = 0$

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

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 $x = 0,1,2,\dots,n$

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

<u>Sampling distribution of $\bar{X}$</u>:

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

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

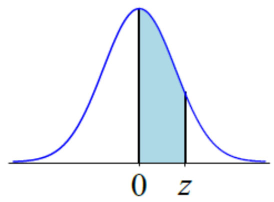
$\sigma_{\bar{X}} = SD[\bar{X}] = \frac{\sigma}{\sqrt{n}}$

<u>Sampling distribution of $\hat{P}$</u>:

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

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

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



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 (2): Recall the 2022 Ontario public disclosure data with *all* public sector employees earning $100,000 CAN or more in the 2021 calendar year. The variable named salary is in $1,000s of Canadian dollars. Below are three different summaries produced by Stata.

Summary #1: This summarizes all employees in the 2022 disclosure.

```
                            salary
-------------------------------------------------------------
      Percentiles      Smallest
 1%    100.1391            100
 5%    101.2492            100
10%    102.2847            100       Obs             244,390
25%    103.3918            100       Sum of Wgt.     244,390

50%    111.2933                      Mean           123.7378
                        Largest      Std. Dev.      35.47401
75%    129.3921           1010
90%    160.1164           1060       Variance       1258.406
95%    188.7571           1520       Skewness       4.610942
99%    266.9365           1630       Kurtosis        57.4611
```

Summary #2: This summarizes a random sample of 40 employees.

```
                            salary
-------------------------------------------------------------
      Percentiles      Smallest
 1%     100.007        100.007
 5%    101.6699       101.1859
10%    102.3109       102.1539       Obs                  40
25%    102.4086       102.2947       Sum of Wgt.          40

50%    108.9492                      Mean           120.7959
                        Largest      Std. Dev.      34.52791
75%    126.7653       154.5566
90%    152.7945       168.6238       Variance       1192.177
95%    175.6801       182.7365       Skewness       3.564894
99%    295.5227       295.5227       Kurtosis       17.69593
```

Summary #3: This summarizes the results of a simulation to obtain the simulated sampling distribution of the sample median for a sample size of 40 ($n = 40$) using 10,000 simulation draws ($m = 10{,}000$).

```
                         sample median
-------------------------------------------------------------
      Percentiles      Smallest
 1%    104.6919       103.3773
 5%    106.3143       103.4144
10%    107.2976       103.5984       Obs              10,000
25%    108.9195       103.6256       Sum of Wgt.      10,000

50%    111.2497                      Mean           111.7129
                        Largest      Std. Dev.      3.782944
75%    114.1313       130.4077
90%    116.6635        131.046       Variance       14.31067
95%    118.5133        131.515       Skewness       .6602635
99%    122.1061       132.4968       Kurtosis       3.639388
```

Supplement for Question (3): The most recent release (January 2022) from Stats Canada describes the population shares and the employment rates in 2021 by educational attainment and age.

Table 14-10-0020-01. Educational attainment and age shares of those aged 15 years and up, Canada, 2021 and Employment rate by educational attainment and age

	Age (share)				Employment rate			
	Total, 15 yrs. and up	15 to 24 yrs.	25 to 54 yrs.	55 yrs. and up	Total, 15 yrs. and up	15 to 24 yrs.	25 to 54 yrs.	55 yrs. and up
Highest educational attainment (share)								
Total, all education levels	1.000	0.142	0.476	0.382	0.602	0.552	0.823	0.345
0 to 8 years	0.039	0.004	0.006	0.029	0.192	0.276	0.501	0.115
Some high school	0.100	0.040	0.020	0.040	0.360	0.351	0.628	0.239
High school graduate	0.196	0.035	0.075	0.086	0.537	0.625	0.747	0.318
Some postsecondary	0.060	0.028	0.018	0.014	0.556	0.536	0.727	0.368
Postsecondary certificate or diploma	0.321	0.023	0.171	0.127	0.655	0.754	0.841	0.385
Bachelor’s degree	0.195	0.012	0.127	0.056	0.732	0.737	0.869	0.422
Above bachelor’s degree	0.090	0.001	0.059	0.029	0.754	0.755	0.894	0.469

Supplement for Question (4): Consider “Legacy and Athlete Preferences at Harvard” published in 2021 in the *Journal of Labor Economics*. Below are excerpts and Table 1. “Legacy” means that at least one parent of the applicant attended Harvard (i.e. is an alumnus). The “Dean’s Interest List” includes applicants whose parents or relatives either have donated money to Harvard or demonstrate potential to donate to Harvard.

Excerpt (Abstract): We use public documents from the *Students for Fair Admissions v. Harvard University* lawsuit to examine admissions preferences for recruited athletes (A), legacies (L), those on the Dean’s Interest List (D), and children of faculty and staff (C), together abbreviated ALDCs. About 43% of white admits are ALDC; the share for African American, Asian American, and Hispanics is about 16%.

Excerpts: Our data consist of 166,727 completed applications of people who would have graduated from Harvard from 2014 to 2019. ALDC applicants are admitted at a rate above 30% and typical applicants are admitted at a rate of just 5.5%. Table 1 breaks down these aggregate admissions rates by ALDC category and race. One might suppose that recruited athletes are a small share of students admitted to Harvard. They are not, representing 10% of admits.

Table 1. Admission Rates by Race and ALDC Status

	Typical (non-ALDC)	Recruited Athlete	Legacy	Dean’s Interest List	Faculty/Staff
White	4.89	87.94	37.07	41.96	45.78
African American	7.58	86.11	28.57	32.53	20.00
Hispanic	6.16	88.52	35.63	41.79	42.11
Asian American	5.13	87.07	35.14	47.83	47.56

Notes: All numbers are percentages. ALDCs = Recruited Athletes, Legacies, those on the Dean’s Interest List, and Children of Faculty and Staff.

Supplement for Question (5): Consider the excerpts and Table 3 below for a 2022 *NBER Working Paper* "Are Effective Teachers for Students with Disabilities Effective Teachers for All?" *Notes:* The value-added measures are continuous variables to measure teacher effectiveness. The word "plurality" means the highest fraction, but not a majority.

Excerpt (Abstract): The success of many students with disabilities (SWDs) depends on access to high-quality general education teachers. Yet, most measures of teacher value-added measures (VAM) fail to distinguish between a teacher's effectiveness in educating students with and without disabilities. We create two VAM measures: one focusing on teachers' effectiveness in improving outcomes for SWDs, and one for non-SWDs. We find top-performing teachers for non-SWDs often have relatively lower VAMs for SWDs. Overall, SWD-specific VAMs may be more suitable for identifying which teachers have a history of effectiveness with SWDs.

Excerpts: We use data from Los Angeles Unified School District (LAUSD), the second-largest school system in the United States. In 2021, LAUSD enrolled over 520,000 students in grades K-12 and employed nearly 24,000 K-12 teachers. The LAUSD administrative data have detailed information on student and teacher characteristics. This study uses student- and teacher-level administrative data from the academic years 2007-2008 to 2017-2018.

Excerpts: Table 3 describes the distribution of VAMs across each group by quintile and subject. One finding is that teachers with the highest (or lowest) VAM for one student group tend to also have the highest (or lowest) VAM for the other student group. For example, Panel A of Table 3 shows that the plurality (9.8%) of math teachers fall into the top quintile (i.e., quintile 5) for SWD VAMs and the top quintile for non-SWD VAMs, with the next largest group (9.0%) of teachers falling into the bottom quintile (i.e., quintile 1) for both VAMs. This pattern persists for English language arts teachers, though to a lesser extent. However, for teachers within the middle quintiles, a large majority fall into a bin off the diagonal (i.e., the values in bold) across both subjects.

Table 3: Teacher Distribution of Value-Added Measures (VAM) for Students with Disabilities (SWD) and Value-Added Measures (VAM) for Students without Disabilities (Non-SWD) by Quintile

Panel A: Math		Non-SWD VAM Quintiles				
		1	2	3	4	5
	1	**9.0%**	5.2%	3.2%	1.7%	0.9%
	2	5.0%	**5.5%**	4.6%	3.2%	1.7%
SWD VAM Quintiles	3	3.3%	4.5%	**4.9%**	4.4%	2.9%
	4	2.0%	3.3%	4.5%	**5.6%**	4.7%
	5	0.7%	1.5%	2.9%	5.1%	**9.8%**
Panel B: English Language Arts		Non-SWD VAM Quintiles				
		1	2	3	4	5
	1	**7.4%**	5.0%	3.8%	2.6%	1.3%
	2	5.0%	**4.8%**	4.3%	3.6%	2.3%
SWD VAM Quintiles	3	3.6%	4.6%	**4.4%**	4.0%	3.4%
	4	2.6%	3.4%	4.2%	**5.0%**	4.7%
	5	1.4%	2.3%	3.3%	4.9%	**8.2%**

Notes: Each teacher-year observation is assigned to a quintile by its relative rank in the value-added measure (VAM) distribution. The lowest quintile of VAMs is 1, and the highest quintile of VAMs is 5. Observations are at the teacher-year level. 34,373 total Math observations. 35,197 total English Language Arts observations. SWD (student with disability). VAM (value-added measure).