

ECO220Y1Y, Test #2, Prof. Murdock

December 7, 2023, 6:30 pm – 8:20 pm

- Keep ALL pages closed and face up on your desk until we announce the start, and only then you may detach the *Supplement*, which has the aid sheets and readings, figures, tables, and other materials for some test questions.
- There are 8 test pages with 7 questions, with varying numbers of parts, worth a total of 95 points.
- You have 110 minutes. You must stay for a minimum of 60 minutes.

Instructions:

- For each question referencing the *Supplement*, carefully review *all* materials. The ***Supplement is NOT collected:*** write your answers on the test papers. At the end, hand in your test papers (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. Focus on those expectations.
 - 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. We give more blank space than is needed for each answer (with typical handwriting) worth full marks. ***Follow the answer guides and avoid excessively long answers.***

(1) See ***Supplement for Question (1): Canada labour force status by educational achievement.***

(a) [6 pts] Among people aged 25 to 54 years in Canada in 2022, _____ % of those who are employed are a high school graduate (with no further education). Among those with no degree, certificate, or diploma, _____ % are employed. Answer with quantitative analysis using formal notation & by filling in the blanks.

(b) [3 pts] For a randomly selected person aged 25 to 54 years in Canada in 2022, what is the chance that they have a university degree or are unemployed? Answer with quantitative analysis using formal notation.

(2) See ***Supplement for Question (2): Legacy and Athlete Preferences at Harvard.***

(a) [4 pts] Write the values 5.5% and 10% in the excerpts as formal probability statements using events as defined in the ***Supplement*** for this question. Answer with two probability statements using formal notation.

(b) [10 pts] Using Harvard applications and admissions data for students who would, after four years, have graduated university from 2014 to 2019, assess whether each statement is true or false. *Explain.* Answer with 4 – 5 sentences.

- **Statement A:** Table 1 shows that a small fraction of students admitted to Harvard are typical (non-ALDC).
- **Statement B:** Table 1 shows that among students admitted to Harvard as a recruited athlete, the fraction who are Hispanic is higher than the fraction who are Asian American.

(3) [4 pts] The mileage for a fleet of rental cars is Normally distributed with a mean of 45,982 km and a s.d. of 10,881 km. What fraction of cars have mileage above 40,000 km? Answer with a quantitative analysis using formal notation.

(4) Consider income for a random sample of households. For very skewed distributions we often use percentiles. For the income distribution of a population, recall that the chance that a randomly selected household is in any given income decile is 10% and that, similarly, for income quartiles, the chance of being in any given income quartile is 25%.

(a) [7 pts] For a sample of 60 households, what is the chance that more than 2 households are in the highest decile of the income distribution of the population? Answer with a quantitative analysis using formal notation.

(b) [12 pts] For a sample of 600 households, what is the chance that less than 22.5% are in the bottom quartile of the income distribution of the population? Illustrate the answer with a fully labelled graph that uses the same units as the question you are asked. Answer with a quantitative analysis & a fully labelled graph.

(5) See *Supplement for Question (5): OECD, Labor Force Participation Rates of Women*.

(a) [6 pts] Compute the mean and standard deviation of the *difference* between `p12_25_54` and `p22_25_54`. (Note: Usually we find the difference for the more recent year relative to an earlier year.) Answer with a quantitative analysis.

(b) [10 pts] *Fully interpret* the mean and standard deviation of `diff_older`. For a complete interpretation, include all available information that is relevant. Answer with 4 – 5 sentences.

(6) See *Supplement for Question (6): School Accountability, Long-Run Criminal Activity, and Self-Sufficiency*.

(a) [14 pts] Compute the 99% confidence interval estimate of the difference between Columns **(2)** and **(6)** for the row labeled **Food Stamps Participation**. Next, *interpret* this interval. Answer with a quantitative analysis & 2 sentence.

(b) [6 pts] Focus on the row **Proficient in Math – 8th Grade**. Comparing Columns (3) and (5), the point estimate of the difference is _____ [# with units, absolute value], and for an 85% confidence interval estimate, the tabular z value is ____ [#]. For Column (1), and noting the excerpt about missing values, the relevant sample size is closest to _____ [54,352 / 106,929 / 116,122 / 161,281] and the standard error of the point estimate is about _____ [0.0001 / 0.0014 / 0.0208]. Fill in the blanks to answer.

(7) See **Supplement for Question (7): 2023 Ontario public disclosure: Hospitals & Boards of Public Health Sector.**

(a) [4 pts] Of all employees in the Hospitals & Boards of Public Health sector with their 2022 salary disclosed, over _____ % [50 / 60 / 70 / 80 / 90] have a salary within one standard deviation of the mean. Further, roughly _____ % [1 / 5 / 10 / 25 / 50] have a salary within five thousand dollars of the disclosure cut-off of \$100,000. Finally, _____ % [#] have a salary above \$750,000, which is _____ [#] employees per 10,000. Fill in the blanks to answer.

(b) [4 pts] Which values should you expect to see on the two blank lines in Summary #2? Answer with a quantitative analysis using formal notation.

(c) [5 pts] If the Central Limit Theorem applied, which number should appear in the spot currently occupied by 143.3268 in Summary #3? Answer with a quantitative analysis using formal notation.

This **Supplement** will *NOT* be collected or graded: write your answers on the test papers. **Supplement: Page 1 of 6**

Supplement for Question (1): Consider the joint probability table below, retrieved from the Statistics Canada website. Refer to events using the one letter in parentheses given for each.

Canada, 2022, Both Sexes, Aged 25 to 54 Years				
	<i>Labour Force Status</i>			
<i>Highest level of educational achievement</i>	Employed (E)	Unemployed (U)	Not in labor force (N)	Total
No degree, certificate, or diploma (L1)	0.0371	0.0036	0.0172	0.0579
High school graduate (L2)	0.1204	0.0075	0.0265	0.1544
High school graduate, some post-secondary (L3)	0.0272	0.0017	0.0053	0.0341
Postsecondary certificate or diploma (L4)	0.3099	0.0124	0.0326	0.3549
University degree (L5)	0.3527	0.0132	0.0327	0.3986
Total	0.8473	0.0384	0.1143	1.0000

Supplement for Question (2): 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. Two values are in boldface for easy reference.

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) (1)	Recruited Athlete (2)	Legacy (3)	Dean’s Interest List (4)	Faculty/Staff (5)
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.

Define events as:

- **Event A:** A randomly selected applicant is admitted;
- **Event T:** A randomly selected applicant is typical (in other words, not ALDC);
- **Event R:** A randomly selected applicant is a recruited athlete.

Supplement for Question (3): N/A (all information given with the question)

Supplement for Question (4): N/A (all information given with the question)

Supplement for Question (5): Recall the OECD data on the percent of women participating in the labor force. For each member nation in the OECD, we have four variables named p12_25_54, p22_25_54, p12_55_64, p22_55_64. The first two variables record the labor force participation rates for women aged 25 to 54 years in 2012 and 2022, respectively. The last two variables record the labor force participation rates for women aged 55 to 64 years in 2012 and 2022, respectively.

.	summarize	p12_25_54	p22_25_54	p12_55_64	p22_55_64;
		Variable	Obs	Mean	Std. Dev.
					Min
					Max
		-----+-----			
		p12_25_54	37	77.9194	10.21291
		p22_25_54	37	81.13087	8.951574
		p12_55_64	37	48.77017	14.05247
		p22_55_64	37	60.55905	13.81927
		-----+-----			
.	correlate	p12_25_54	p22_25_54	p12_55_64	p22_55_64;
(obs=37)			p12_25_54	p22_25_54	p12_55_64
			-----+-----		
		p12_25_54	1.0000		
		p22_25_54	0.9613	1.0000	
		p12_55_64	0.4441	0.4066	1.0000
		p22_55_64	0.7132	0.7084	0.8266
		-----+-----			

Next, consider creating two more variables to measure the difference from 2012 to 2022 for middle-aged women (aged 25 to 54 years) and the difference from 2012 to 2022 for older women (aged 55 to 64 years). Some values are intentionally erased and replaced with a blank line.

.	summarize	diff_miage	diff_older;		
		Variable	Obs	Mean	Std. Dev.
					Min
					Max
		-----+-----			
		diff_miage	37	_____	_____
		diff_older	37	11.78888	8.209536
		-----+-----			

This space is deliberately left blank.

This *Supplement* will *NOT* be collected or graded: write your answers on the test papers. ***Supplement: Page 3 of 6***

Supplement for Question (6): Consider a 2023 NBER Working Paper “School Accountability, Long-Run Criminal Activity, and Self-Sufficiency.” The authors use data from the southern part of the United States. Below are excerpts and Table 1.

Excerpts: Table 1 presents the descriptive statistics for a total of more than 160,000 students from 194 unique high schools. We show tabulations for the full sample, as well as by schools’ performance ratings. As displayed in Panel A, black and white students comprise 41 and 55 percent of all students, respectively, and the percentage of black students is decreasing along accountability ratings. Similarly, there is a negative relationship between the fraction of free-lunch eligible students and schools’ ratings. The opposite pattern is displayed between the fraction of students who were proficient in eight grade subject tests. Eighth-grade standardized test scores were missing for an overwhelming majority of the analysis sample. As a result, we use discrete achievement indicators (e.g., proficient; advanced), which are available for more than 70 percent of the analysis sample, to proxy for subject-specific eighth grade achievement level in math and English Language Arts (ELA).

As shown in Panel B, 24 percent of students, who attended a high school that was rated as Unsatisfactory, were arrested as an adult. About 8 percent were incarcerated (Column 2). The gap between the arrest and incarceration rates of students in schools rated as Unsatisfactory and Average is slightly more than 2 percentage points (Columns 2 and 4). Driving under influence, disorderly conduct, possession of drugs and shoplifting are the most common types of arrests in the data.

The first column of Table 1 reveals that 51 percent of students in our sample used food stamps [which is a social safety net program to help people in poverty buy food] as an adult. The fraction of individuals receiving government assistance are disproportionately associated with low performing schools. For example, while 37 percent of individuals who attended schools with an Excellent rating received food stamps as an adult, the rate is about 74 percent among individuals who attended Unsatisfactory schools.

Table 1: Summary Statistics by Accountability Ratings

	All Schools (1)	School Rating				
		Unsatisfactory (2)	Below Average (3)	Average (4)	Good (5)	Excellent (6)
Panel A: Student Characteristics						
Black	0.414	0.792	0.671	0.554	0.363	0.258
White	0.554	0.191	0.297	0.417	0.608	0.698
Female	0.481	0.471	0.465	0.479	0.48	0.485
Free/Reduced Lunch	0.406	0.677	0.595	0.503	0.404	0.254
Proficient in Math - 8 th Grade	0.337	0.302	0.274	0.247	0.351	0.379
Proficient in ELA - 8 th Grade	0.374	0.325	0.290	0.282	0.381	0.432
Panel B: Adult Outcomes						
Arrest	0.199	0.239	0.238	0.216	0.204	0.167
Incarceration	0.052	0.081	0.077	0.059	0.052	0.037
Food Stamps Participation	0.513	0.738	0.667	0.596	0.526	0.370
Welfare Participation	0.515	0.740	0.672	0.597	0.527	0.372
Sample Size	161,281	13,365	13,932	19,074	62,445	52,465

Notes: The sample is three cohorts of first-time ninth graders in public high schools in the academic years of 2000/2001, 2001/2002, and 2002/2003. A student performing at or above the proficient level on the state’s eighth grade subject-specific assessment is labeled as proficient. English Language Arts is abbreviated ELA.

Supplement for Question (7): Recall the 2023 Ontario public disclosure data with *all* public sector employees earning \$100,000 CAN or more in the 2022 calendar year. The variable named salary is in \$1,000s of Canadian dollars. Below are three summaries produced by Stata where all are only for employees in the “Hospitals & Boards of Public Health” sector.

Summary #1: All employees in the “Hospitals & Boards of Public Health” sector in the 2023 disclosure.

salary				
Percentiles		Smallest		
1%	100.1499	100.0008		
5%	100.8495	100.0013		
10%	101.8666	100.0016	Obs	40,951
25%	104.9297	100.0032	Sum of Wgt.	40,951
50%	111.3786		Mean	121.7868
		Largest	Std. Dev.	38.43549
75%	123.4785	736.1647		
90%	144.4412	773.824	Variance	1477.287
95%	170.001	780.0134	Skewness	5.897737
99%	319.1237	844.9925	Kurtosis	53.50201

Summary #2: The results of a simulation to obtain the simulated sampling distribution of the sample mean for a sample size of 10 using 100,000 simulation draws. Notice the two blank lines (and another number is erased).

sample mean				
Percentiles		Smallest		
1%	106.8972	102.4656		
5%	109.157	102.5241		
10%	110.6331	102.6273	Obs	100,000
25%	113.6462	102.8235	Sum of Wgt.	100,000
50%	118.3273		Mean	_____
		Largest	Std. Dev.	_____
75%	126.1433	225.7138		
90%	137.8608	227.6509	Variance	
95%	146.0097	239.1076	Skewness	1.877735
99%	165.0742	242.8165	Kurtosis	8.214477

Summary #3: The results of a simulation to obtain the simulated sampling distribution of the sample mean for a sample size of 30 using 100,000 simulation draws. One value is in boldface for easy reference.

sample mean				
Percentiles		Smallest		
1%	110.7988	106.8698		
5%	112.8264	107.1197		
10%	114.0764	107.1317	Obs	100,000
25%	116.6332	107.3403	Sum of Wgt.	100,000
50%	120.4771		Mean	121.7831
		Largest	Std. Dev.	7.030083
75%	125.6103	166.6694		
90%	131.2345	167.1835	Variance	49.42206
95%	135.1099	169.2172	Skewness	1.07523
99%	143.3268	170.0756	Kurtosis	4.680963

This *Supplement* will *NOT* be collected or graded: write your answers on the test papers. **Supplement: Page 5 of 6**

Sample mean: $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$ **Sample variance:** $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2}{n-1} - \frac{(\sum_{i=1}^n x_i)^2}{n(n-1)}$ **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: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ **Conditional probability:** $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$

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

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

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

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

Laws of expected value:

$$\begin{aligned} E[c] &= c \\ E[X + c] &= E[X] + c \\ E[cX] &= cE[X] \\ E[a + bX + cY] &= a + bE[X] + cE[Y] \end{aligned}$$

Laws of variance:

$$\begin{aligned} V[c] &= 0 \\ V[X + c] &= V[X] \\ V[cX] &= c^2 V[X] \\ V[a + bX + cY] &= b^2 V[X] + c^2 V[Y] + 2bc * COV[X, Y] \\ V[a + bX + cY] &= b^2 V[X] + c^2 V[Y] + 2bc * SD(X) * SD(Y) * \rho \end{aligned}$$

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

Laws of covariance:

$$\begin{aligned} COV[X, c] &= 0 \\ COV[a + bX, c + dY] &= bd * COV[X, Y] \end{aligned}$$

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

Sampling distribution of \bar{X} :

$$\begin{aligned} \mu_{\bar{X}} &= E[\bar{X}] = \mu \\ \sigma_{\bar{X}}^2 &= V[\bar{X}] = \frac{\sigma^2}{n} \\ \sigma_{\bar{X}} &= SD[\bar{X}] = \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Sampling distribution of \hat{P} :

$$\begin{aligned} \mu_{\hat{P}} &= E[\hat{P}] = p \\ \sigma_{\hat{P}}^2 &= V[\hat{P}] = \frac{p(1-p)}{n} \\ \sigma_{\hat{P}} &= SD[\hat{P}] = \sqrt{\frac{p(1-p)}{n}} \end{aligned}$$

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

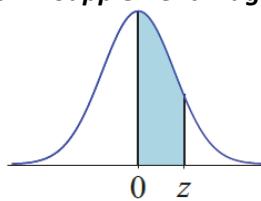
$$\begin{aligned} \mu_{\hat{P}_2 - \hat{P}_1} &= E[\hat{P}_2 - \hat{P}_1] = p_2 - p_1 \\ \sigma_{\hat{P}_2 - \hat{P}_1}^2 &= V[\hat{P}_2 - \hat{P}_1] = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1} \\ \sigma_{\hat{P}_2 - \hat{P}_1} &= SD[\hat{P}_2 - \hat{P}_1] = \sqrt{\frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}} \end{aligned}$$

Inference about a population proportion:

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: