Dawson College
Mathematics Department
Final Examination

Fall 2022: Thursday, Dec. 15-th (9:30 am to 12:30 pm)

201-BZS-05 (Sec. 00001), Probability and Statistics (Science)

Examiner: S. Shahabi.

Student's Full Name: ______________________ **ID:** ______________

- Print your name and student ID number in the space provided above;
- All questions are to be answered directly on the examination paper in the space provided. If you need more space for your answer, use the back of the page;
- No book, notes, graphing/programmable calculator or cellphones are permitted. You are only permitted to use the Sharp EL-531XG calculator during the examination;
- A Formula Sheet and the relevant Stat Tables are provided by the examiner;
- You must show all your work and justify all your answers;
- This examination booklet, all the Statistics Tables and the Formula Sheets must be returned intact.

Question	# Marks	student's scores
1	8	
2	5	
3	8	
4	5	
5	6	
6	8	
7	4	
8	8	
9	8	
10	8	
11	8	
12	8	
13	8	
14	8	
Total	100	

THIS EXAMINATION BOOKLET CONTAINS 8 PAGES
(INCLUDING THIS COVER PAGE)

1. **[1+1+1+5 pts.]** Let $D = \{1, 2, 3, ..., 8\}$, and answer the following:

(i) How many five-digit numbers from D are there? Five positions $N = \underline{a}\ \underline{b}\ \underline{c}\ \underline{d}\ \underline{e}$ to be filled, & repetition allowed $\Rightarrow$ the answer $= 8\times8\times8\times8\times8 = \boxed{8^5}$

(ii) How many of these numbers have distinct digits?

no repetition $\Rightarrow$ the anwer $= \boxed{8\cdot7\cdot6\cdot5\cdot4} = \frac{8!}{3!} = P(8,5)$

(iii) How many of those from **(ii)** are less than 60000?

$N < 60000 \Leftrightarrow$ the left digit $= 1,2,3,4,5$ $\Rightarrow$ answer $= \boxed{5\times7\times6\times5\times4}$

(iv) How many of those from **(ii)** have 3, 5 and 8 in them such that 3 occurs after 5 and 5 occurs after 8?

First we need to choose two more digits (from the five remaining digits). This can be done in $\binom{5}{2}$ ways. Now Suppose we have $a, b, 8, 5, 3$ to form a 5-digit number.

In $2\times1 + 2\times1 + 2\times1$ of them 8 & 5 & 3 are adjacent. (e.g.: $\underline{a}\,853\,\underline{b}$)

" $2\times2 + 2\times1$ " " only 5 & 3 " " . (e.g: $8\,\underline{a}\,53\,\underline{b}$)

" $2\times1 + 2\times2$ " " only 8 & 5 " " . (e.g: $\underline{a}\,85\,\underline{b}\,3$)

And in 2×1 of them 8 & 5, as well as 5 & 3, are separated. ($8\,\underline{a}\,5\,\underline{b}\,3$)

$\Rightarrow$ the answer $= \left((6) + (6) + (6) + (2)\right)\cdot\binom{5}{2} = 200$

2. **[5 pts.]** A jar contains four balls: {white, green, blue, black}. We draw three balls, *one ball at a time with replacement*, at random. What is the probability of observing **exactly two colours**?

$|S| = \underline{4} \times \underline{4} \times \underline{4} = 4^3$. ($S$: all possibilities)

E = the desired event. To find $|E|$, we 1st choose 2 colours out of 4: $\binom{4}{2}$. Denote these colours by C_1, C_2. We could have any of the following configurations: $\left\{\begin{matrix} C_1C_1C_2, & C_1C_2C_1, & C_2C_1C_1 \\ C_1C_2C_2, & C_2C_1C_2, & C_2C_2C_1 \end{matrix}\right\}$

$\Rightarrow |E| = \binom{4}{2}\times 6 \Rightarrow P(E) = \frac{\binom{4}{2}\times 6}{4^3} \left(= \frac{9}{16}\right)$

3. **[3+5 pts.]** A manufacturer employs 3 analytical plans for the design and development of a particular product. Plan 1, Plan 2, & Plan 3 are used for 30%, 20% and 50% of the products, respectively. The defective rates of the three plans are 0.01, 0.03, and 0.02 respectively.

(i) What is the probability that a randomly selected product is defective?

$$P(D) = P(\text{Plan }1)\cdot P(D\mid \text{Plan }1) + P(\text{Plan}2)\cdot P(D\mid \text{Plan}2)$$
$$+ P(\text{Plan}3)\cdot P(D\mid \text{Plan }3)$$
$$= \left(\frac{3}{10}\right)\left(\frac{1}{100}\right)+\left(\frac{2}{10}\right)\left(\frac{3}{100}\right)+\left(\frac{5}{10}\right)\left(\frac{2}{100}\right) = \frac{19}{1000} = 0.019$$

(Total Probability Formula) (or 1.9%)

(ii) If the selected product was indeed defective, which plan was most likely used and thus responsible?

$$P(\text{plan}1 \mid D) = \frac{P(\text{Plan}1)\cdot P(D\mid\text{plan}1)}{P(D)} = \frac{\frac{3}{1000}}{\frac{19}{1000}} = \frac{3}{19}$$

$$P(\text{plan}2 \mid D) = \text{- - - - - - - -} = \frac{6/1000}{19/1000} = \frac{6}{19}$$

$$P(\text{Plan}3 \mid D) = \text{- - - - - - - - - -} = \frac{10/1000}{19/1000} = \boxed{10/19}$$

$\therefore$ Most likely, Plan 3 was responsible!

(Baye's Formula)

4. **[5 pts.]** Assuming that there are, on the average, 9 potholes on every 150 meters of Maisonneuve Street in Montreal, what is the probability of observing only 5 potholes between two locations (on Maisonneuve) with a distance of 1500 meters?

$$\lambda = \left(\frac{1500}{150}\right)(9) = 90 \quad (\text{a Poisson dist. with } \lambda = 90)$$

$$P(E) = P(X=5) = \frac{e^{-90}\cdot 90^5}{5!} \quad (\approx 4.032068762\times 10^{-32})$$

(The chance is almost zero!)

5. **[3+3 pts.]** The length of time (in hours) required to unload trucks at a depot is exponentially distributed with $\lambda = 4$. Now answer the following:

(i) What proportion of the trucks can be unloaded in less than 2 hours?

$$f(x) = \frac{1}{4} e^{-x/4} \quad (x \geq 0).$$

$$P(X<2) = \int_0^2 \frac{1}{4} e^{-x/4}\, dx = \left[-e^{-x/4}\right]_0^2 = 1 - e^{-1/2} = \frac{\sqrt{e}-1}{\sqrt{e}} \approx \boxed{0.3935}$$

(or 39.35%)

(ii) How long would it take to unload 60% of the trucks?

$$P(X \leq \overset{?}{K}) = 0.6 \Rightarrow \int_0^K \frac{1}{4} e^{-x/4}\, dx = 0.6 \Rightarrow 1 - e^{-K/4} = 0.6$$

$$\Rightarrow e^{-K/4} = 0.4 \Rightarrow e^{K/4} = 2.5 \Rightarrow K = 4\ln(2.5) \approx \boxed{3.66 \text{ (hours)}}$$

6. **[3+5 pts.]** In flipping a pair of *fair coins* repeatedly, we stop once we observe TT ***for the second time.***

(i) If X is the r.v. whose value represents the number of flips when we stop, give a formula for the prob. dist. fun. $f(x) = \mathbf{P}(X = x)$, for $x = 2, 3, 4, 5, \ldots$

$$p = P(\text{TT in any flipping}) = 1/4 \qquad q = 3/4$$

$$f(n) = \binom{n-1}{1} \cdot (p)^2 (q)^{n-2} = \frac{3^{n-2} \cdot (n-1)}{4^n} \quad (n = 2, 3, 4, \ldots)$$

(ii) *Verify* that $\sum_x f(x) = 1$. **(Hint:** A useful formula: $\underbrace{1 + 2x + 3x^2 + 4x^3 + \cdots}_{\sum_{n=2}^{\infty}(n-1)x^{n-2}} \overset{(*)}{=} \frac{1}{(1-x)^2}$)

$$\sum_x f(x) = \sum_{n=2}^{\infty} f(n) = \sum_{n=2}^{\infty} \frac{3^{n-2}(n-1)}{4^n}$$

$$= \frac{1}{4^2} \sum_{n=2}^{\infty} \left(\frac{3}{4}\right)^{n-2} \cdot (n-1) \overset{(*)}{=} \frac{1}{4^2} \cdot \frac{1}{(1-\frac{3}{4})^2} = \frac{1}{4^2} \cdot (4^2) = 1 \checkmark$$

$x = \frac{3}{4}$ in $(*)$

7. **[4 pts.]** Assume that the following table is the probability distribution function for a discrete r.v. X:

Find a and b if $E[X] = \frac{529}{180}$

x	0	1	2	3	4	5
$f(x)$	$\frac{1}{12}$	$\frac{1}{12}$	a	b	$\frac{1}{9}$	$\frac{2}{9}$

$\sum_x f(x) = 1 \Rightarrow \frac{1}{12} + \frac{1}{12} + a + b + \frac{1}{9} + \frac{2}{9} = 1 \Rightarrow a + b = \frac{1}{2}$ ①

$\sum x f(x) = E[X] \Rightarrow \frac{1}{12} + 2a + 3b + \frac{4}{9} + \frac{10}{9} = \frac{529}{180} \Rightarrow 2a + 3b = \frac{13}{10}$ ②

$$\begin{cases} a+b = \frac{1}{2} & (1) \\ 2a+3b = \frac{13}{10} & (2) \end{cases} \qquad \begin{cases} 2a+2b = 1 \\ 2a+3b = \frac{13}{10} \end{cases} \xrightarrow{\text{subtract}} \boxed{b = \frac{3}{10}}$$

$\longrightarrow a = \frac{1}{2} - b = \frac{1}{2} - \frac{3}{10} \Rightarrow \boxed{a = 1/5}$

8. **[8 pts.]** A research journalist claims that NHL players average at least 180 pounds in weight. A r.s. of 49 players' weights gave $\bar{X}_0 = 176.8$ lbs, with $S = 6.2$ lbs. Does this support the journalist's claim? Use $\alpha = 0.02$.

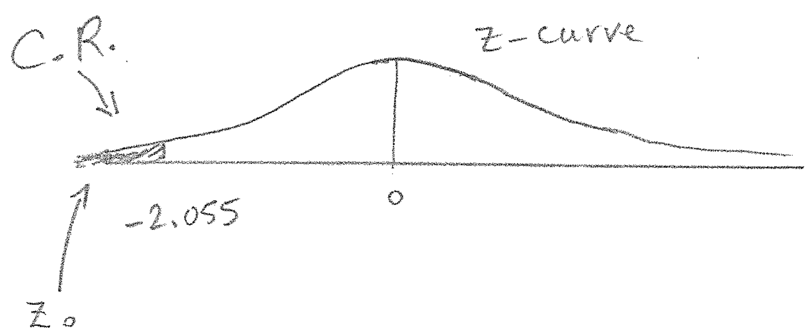
$$\begin{cases} H_0 : \mu \geq 180 \\ H_a : \mu < 180 \end{cases}$$

$n = 49$ (one-tailed test) $\quad z_\alpha = 2.055$

$$z_0 = \frac{\bar{X}_0 - \mu_0}{S/\sqrt{n}} = \frac{176.8 - 180}{6.2/\sqrt{49}} \approx -3.61 \in \text{Critical Region}$$

$\Rightarrow$ we may reject H_0

$\Rightarrow$ The sample does not the journalist, so, he/she is probably wrong.



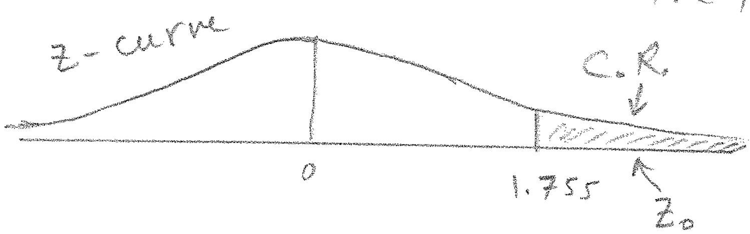
9. **[8 pts.]** A political party claims that more than 2/3 of all citizens support their platform. A survey of 4050 citizens included 2754 who supported their platform. Does this survey agree with the claim? Conduct a Test of Hypothesis with $\alpha = 0.04$.

$n = 4050 \quad X = 2754 \quad \hat{p} = \frac{2754}{4050} = 0.68 \quad \hat{q} = 0.32$

$$\begin{cases} H_0: p \leq \frac{2}{3} \\ H_a: p > \frac{2}{3} \text{ (one-tailed test)} \end{cases}$$

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.68 - \frac{2}{3}}{\sqrt{\frac{(\frac{2}{3})(\frac{1}{3})}{4050}}} \approx \frac{0.01\bar{3}}{0.0074} \approx 1.80 \in C.R.$$

$Z_\alpha = 1.755$ $\Rightarrow$ We may reject H_0, and thus support the polical party's claim.



10. **[8 pts.]** A candidate for political party polls 1000 men and 1000 women voters. If 459 of the men and 388 of the women said that they would vote for the candidate, use $\alpha = 0.05$ to decide whether or not the proportion of men who would vote for the candidate is greater than the proportion of women.

$n_1 = 1000$ men $\quad n_2 = 1000$ women $\quad \hat{p}_1 = \frac{459}{1000} \quad \hat{p}_2 = \frac{388}{1000}$

$X_1 = 459$ vote $\quad X_2 = 388$ vote

$$\begin{cases} H_0: p_1 - p_2 \leq 0 \\ H_a: p_1 - p_2 > 0 \end{cases}$$

$\hat{p} = \frac{459 + 388}{1000 + 1000} = 0.4235 \quad \hat{q} = 0.5765$

$Z_\alpha = 1.645$

$$Z_0 = \frac{(\hat{p}_1 - \hat{p}_2) - (0)}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.459 - 0.388}{\sqrt{(0.4235)(0.5765)(\frac{2}{1000})}} \approx 3.21 \in C.R.$$

$\Rightarrow$ We may reject H_0 and support the idea that the proportion of men who would vote for the candidate is (probably) higher than the proportion of women.

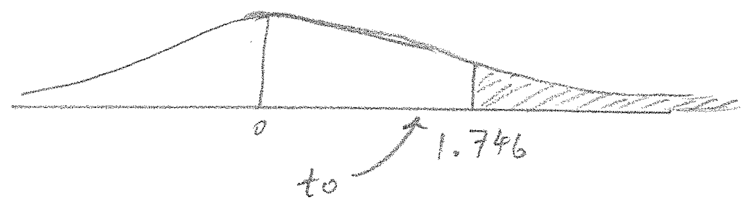
11. **[8 pts.]** A medical journal asserts that vegetarians live longer, on the average, than non-vegetarians. Is this assertion supported by the data below? Test using $\alpha = 0.05$, and assume that the two populations have equal standard deviations.

	n_i	$\overline{X}_i$	S_i^2
vegetarians	9	80 yrs	34
non-vegetarians	9	76 yrs	30

veg. non-veg.

$$\begin{cases} H_0: \mu_1 \le \mu_2 \\ H_a: \mu_1 > \mu_2 \end{cases}$$

a t-curve



$$S_p = \sqrt{\frac{34(9-1)+30(9-1)}{9+9-2}} \approx 5.6569$$

$$t_0 = \frac{(80-76)-(0)}{5.6569\sqrt{\frac{1}{9}+\frac{1}{9}}} \approx 1.5 \notin C.R. \Rightarrow \text{We may not reject } H_0.$$

$\Rightarrow$ We cannot support the claim that vegetarians live longer than non-vegetarians.

$t_\alpha = 1.746$

$d.f. = 16$

12. **[6+2 pts.]** Consider the test scores for 10 students before and after a tutoring session:

Before	75	62	67	70	55	59	60	64	72	59
After	77	65	68	72	62	61	60	67	75	68
After-Before	2	3	1	2	7	2	0	3	3	9

(i) Is there a significant improvement? Conduct a T. of H., using $\alpha = 0.05$;

(A: After B: Before)

$\bar{d} = \frac{32}{10} = 3.2$

$S_d = 2.74$

$$\begin{cases} H_0: \mu_A \le \mu_B \\ H_a: \mu_A > \mu_B \end{cases} \quad \text{or} \quad \begin{matrix} \mu_d \le 0 \\ \mu_d > 0 \end{matrix}$$

$$t_0 = \frac{\bar{d} - \mu_d}{S_d/\sqrt{n}} = \frac{3.2 - 0}{2.74/\sqrt{10}} \approx 3.69 \in C.R. \Rightarrow \text{We may reject } H$$

$\Rightarrow$ probably a significant improvement.

$t_\alpha = 1.833$

$(d.f. = 9)$

(ii) Construct a 90% C.I.E. for the mean improvement in session.

$t_{\alpha/2} = 1.833$

$$\bar{d} - t_{\alpha/2}\frac{S_d}{\sqrt{n}} < \mu_A - \mu_B < \bar{d} + t_{\alpha/2}\frac{S_d}{\sqrt{n}}$$

$$3.2 - (1.833)\frac{2.74}{\sqrt{10}} < \quad < 3.2 + (1.833)\frac{2.74}{\sqrt{10}}$$

or $I_{90\%} = [1.612, 4.79]$

13. **[3+5 pts.]** A population is made up of four groups: A, B, C, and D. The composition of the population is as follows:

group	A	B	C	D
percentage	40%	30%	20%	10%

A high council of 200 people is to serve the society. The composition of *the existing Council* is:

group	A	B	C	D
# of members in the Council	72	68	42	18

(i) If the Council was a *random sample* drawn from the population, what would be the expected number of council members from each of the groups A, B, C, D?

$E(A) = (200)(0.4) = 80$ $\quad E(C) = (200)(0.2) = 40$

$E(B) = (200)(0.3) = 60$ $\quad E(D) = (200)(0.1) = 20$

(ii) Test the hypothesis, using $\alpha = 0.05$, that the proportions in the Council match the proportions in the whole population.

$\begin{cases} H_0: \text{good fit} \\ H_a: \text{poor fit} \end{cases}$ $\quad \chi^2_\alpha = 7.815$ (d.f. = 3)

$$\chi_0^2 = \frac{(72-80)^2}{80} + \frac{(68-60)^2}{60} + \frac{(42-40)^2}{40} + \frac{(18-20)^2}{20} = \frac{65}{30} \approx 2.167 < \chi_\alpha$$

$\chi_0^2 \notin C.R. \Rightarrow$ We may not reject H_0

$\Rightarrow$ The proportions in the Council match the proportions in the whole society.

14. **[8 pts.]** Consider the results below from a consumer study of the performance of 4 competing brands of toothpaste:

	Brand A	Brand B	Brand C	Brand D	
0 cavities	9, 10	13, 12	17, 15	11, 13	50
1-5 cavities	63, 60	70, 72	85, 90	82, 78	300
> 5 cavities	28, 30	37, 36	48, 45	37, 39	150
	100	120	150	130	500

Test the hypothesis that the incidence of cavities is independent of the brand of toothpaste. Use $\alpha = 0.05$.

$\begin{cases} H_0: \text{independent} \\ H_a: \text{not-independent} \end{cases}$ $\quad \chi^2_\alpha = 12.592$ (d.f. = (4−1)(3−1) = 6)

$$\chi_0^2 = \frac{(9-10)^2}{10} + \frac{(13-12)^2}{12} + \cdots + \frac{(37-39)^2}{39} \approx 1.91 < \chi_\alpha$$

$\Rightarrow$ We may not H_0 $\Rightarrow$ the incidence of cavities is (prob.) independent of the brands.