DAWSON COLLEGE

Mathematics Department

FINAL EXAMINATION

Probability and Statistics (201-BZS-05) Fall 2024

Instructor:
Noushin Sabetghadam

Student Name: ______________________________ **Student ID. #:** __________________________

Instructions:

- You are required to show your work on each problem, in a reasonably neat and coherent way. Work scattered all over the page without a clear ordering could receive very little credit.
- Round your answers to the at least four decimal places.
- For the Test of Hypothesis questions, you need to write the exact H_0 and H_α, you need also clearly write the final result and decision.
- You are only allowed to use a calculator Sharp EL531.
- The formula sheet and the tables are at the end of the examination booklet and must be returned intact.

This examination consists of 20 questions. Please ensure that you have a complete examination booklet before starting.

Question	# Marks	student's scores
1-5	10	
6	6	
7	6	
8	6	
9	6	
10	6	
11	6	
12	6	
13	6	
14	6	
15	6	
16	6	
17	6	
18	6	
19	6	
20	6	
Total	100	

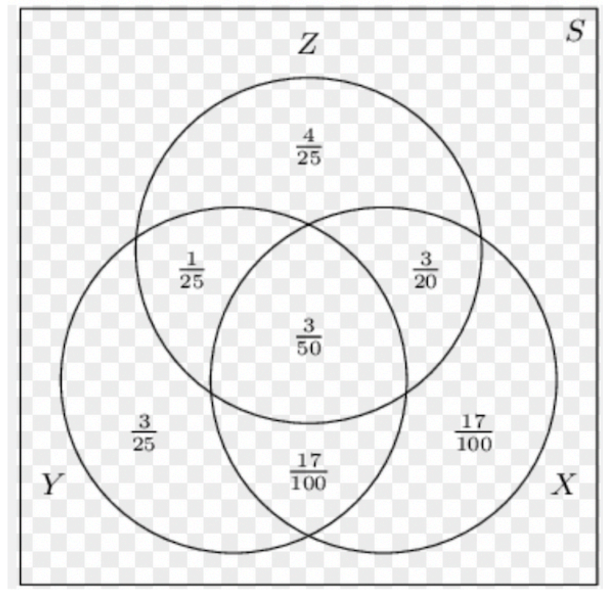
(Each Question 2 marks) For the questions 1-5, write only the final answers in the given boxes. For these problems, the marks are only considered for the correct final answers.

(1) Find the level of measurement of the following variables.

Temperature in Fahrenheit

Level of Education

(2) In the given Venn diagram there are three events X, Y and Z. The numbers indicate the probability of different parts. Find the conditional probability $P(X|Z)$.



Find the probability $P(Y^c \cap Z^c)$.

(3) From a deck of 52 cards, how many different 4-card hands are possible if we want to have one from each suit?

(4) Suppose that hat A contains 3 red and 4 black poker chips and hat B contains 2 red and 3 black poker chips. You pick one chip of hat A and without seeing its colour, you put it in hat B. Then your friend picks a chip from hat B. What is the probability that your friend pick a black chip?

(5) In a small community college there are 400 students; 270 of them do Computer Science, 300 do English and 50 do Business studies. All those doing Computer Science do English, 20 take Computer Science and Business studies and 35 take English and Business studies. Using a Venn diagram, calculate the probability that a pupil drawn at random will take English or Business studies but not Computer Science.

For the following questions (Questions 6-20) show all your work with the details.

(6) **(3 marks)** Complete the table below of the final exam grades of 675 students in Calculus I in Fall 2020.

Class	Class Limits	Class Marks	frequency	rel.. frequency.	cum. frequency
1	1-20		24		
2	21-				81
3	-		72		
4	-				594
5	-			0.12	

(3 marks) Use the data in the table to estimate $\bar{x}$ and S.

(7) Tim's Hardware store manager believes that there is a linear relationship between the income from sales of goods (Y, in thousands of dollars) and the amount spent on advertising (X, in thousands of dollars).

The table gives us :

$\sum x = 43.2, \quad \sum x^2 = 195.34$

$\sum y = 185 \quad \sum y^2 = 3601 \quad \sum xy = 835$

Advertising Spending	Income from Sales
5.3	21
3.8	16
3.1	13
2.9	12
4.4	23
4.9	20
5.1	23
5.4	24
3.2	14
5.1	19

(a) (4 marks) Find the least squares regression line.

(b) (1 marks) Use the answer in part (a) to Predict that how much is predicted to be the store's income if they spend 4000$ on advertising.

(c) (1 marks) Calculate the coefficient of determination and comment on the accuracy of the estimate obtained in part (b).

(8) Five boys and three girls take an elevator in the ground floor of a building to go up, and all of them get out of the elevator before the sixth floor which means that in the sixth floor when the elevator door gets opened, the elevator is empty.

(a) (2 marks) In how many ways can they get out of the elevator?

(b)(2 marks) In how many ways can they get off if nobody gets out in the first floor, and in each other floor exactly two of them getting out of the elevator?

(c)(2 marks) In how many ways can they get out of the elevator such that at least a boy gets out in each floor?

(9) Four men and 3 women are ranked according to their scores on an exam. Assume that no two scores are alike. Let X denotes the highest ranking achieved by a woman (so X=1 indicates that a woman achieved the highest score on the exam).

(a) (3 marks) What is the value of $P(X = 2)$?

(b) (3 marks) What is the highest value for X and what is the probability for that?

(10) **(6 marks)** It is estimated that 50% of emails are spam. A certain software has been applied to filter these spam emails before they reach your inbox. It is claimed that this software can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 6%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

(11) **(6 marks)** In a fishing event, a small lake is populated with some Coho Salmons, Rainbow trouts and Largemouth basses, 30% of these fishes are tagged. Each participant is allowed to capture 4 fish during the day. For each tagged fish, the participant wins 20$, and for each non-tagged fish, the participant wins only 5$. However, it costs 40$ to participate in this event. Set up a **probability table** and find **the expected value of a participant's net earnings** if she/he catches 4 fish at the end of the day.

(12) On the average, a certain computer part lasts ten years, where the lifespan of these parts is modelled by an exponential distribution; $f(t;\lambda) = \begin{cases} \frac{1}{\lambda}e^{(-t/\lambda)} & 0 \leq t, \\ 0 & \text{otherwise} \end{cases}$

(a)(3 marks) What is the probability that a computer part lasts more than 7 years?

(b) (3 marks) If your computer contains 5 such parts, then find the probability that no part is broken during the first 7 years.

(13) **(a)(3 marks)** Determine the value of k such that $f(t)$ forms a probability density function.

$$f(t) = k\sqrt{t}(1-t), \quad t \in [0,\ 1].$$

(b) (3 marks) Use k from part **(a)** to find the cumulative density function $F(x)$.

(14) According to data from Transport Canada's National Collision Database, using a phone or another electronic device, contributed to an estimated 22.5% of fatal collisions.

(a) (3 marks) What is the probability that, for the next 15 fatal car crashes, more than 2 of them will be caused by using a phone or another electronic device?

(b) (3 marks) What is the probability that, for the next 100 car crashes in Canada, at least 20 and at most 40 (both inclusive) of them will be caused by using a phone or another electronic device.? Use an approximation method, and verify first that it is valid to use such an approximation.

(15) In 2018, when inspecting a section of the Champlain bridge, there were two kinds of cracks: Structural cracks, and Expansion cracks. Suppose the number of Structural cracks followed a Poisson distribution with an average of 3 cracks per 10-meter section, and the number of Expansion cracks followed a Poisson distribution with an average of 4 cracks per 5-meter section.

(a) (3 marks) What was the probability that an inspector could find exactly 10 Structural cracks in a 30-meter section?

(b) (3 marks) What was the probability to find neither Structural nor Expansion cracks in a 10-meter section assuming that the number of Structural cracks and the number of Expansion cracks were independent.

(16) **(a) (4 marks)** The average number of extension requests in a Calculus class at a university in Montreal used to be about thirty per semester. It seems that this number dropped significantly after the new program implemented in 2015. A sample of 14 semesters shows a mean number of extension requests equal to 22.3 with sample standard deviation of 5.8 requests. Test the hypothesis at 1% level of significance and clearly state a conclusion in the context of the problem.

(b) (2 marks) Find a 95% confidence interval for the average number of extension requests in Calculus classes in the new program using the given sample.

(17) **(6 marks)** The blood sugar readings for 10 people with diabetes were recorded before and after a medication regime:

before	128	145	125	156	129	147	130	127	125	145
after	117	133	108	132	111	129	122	124	113	129

Can we conclude that the medication lowers blood sugar by more than 10 points? Conduct a Test of Hypothesis using $\alpha = 0.05$ and write all the details and the conclusion.

(18) **(6 marks)** Conduct a Test of Hypothesis and use **p-value approach**, with $\alpha = 0.05$ to determine if the proportion of Chat GPT users in English Cegeps and French Cegeps are different. Use the following table of two samples of these Cegeps:

	sample size	Number of GPT users
English Cegeps	480	283
French Cegeps	386	212

(19) **(6 marks)** A researcher for the financial services company collected the following records of each day of the week employees called in sick to work. Can the researcher conclude that proportion of employees who call in sick is not the same for each day of the week? Design and conduct a hypothesis test at the 1% significance level.

Days of Week	Frequency
Monday	94
Tuesday	65
Wednesday	60
Thursday	78
Friday	91
Total	388

(20) There are three genotypes and their possible effects on Alzheimer's disease are given in the following two-way table:

	No Disease	Alzheimer's disease	Total
Genotype A	268	807	1075
Genotype B	199	759	958
Genotype C	42	184	226
Total	509	1750	2259

(6 marks) Does this data indicate a significant relationship between the genotypes and Alzheimer's disease? Use a $\alpha = 5\%$.

Some Formulae

$$\bar{x} = \frac{\sum x}{n}, \qquad s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

$$SS(x) = \sum (x-\bar{x})^2 = \sum x^2 - \frac{\left(\sum x\right)^2}{n}, \qquad SS(xy) = \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}$$

$$\bar{x} = \frac{\sum (f \cdot x)}{n}, \qquad s^2 = \frac{SS(x)}{n-1} = \frac{\sum (f \cdot x^2) - \frac{(\sum (f \cdot x))^2}{n}}{n-1}$$

$$z = \frac{x-\bar{x}}{S} \quad \text{or} \quad z = \frac{x-\mu}{\sigma}.$$

$$\widehat{y} = a + bx \quad \text{where} \quad b = \frac{SS(xy)}{SS(x)}, \quad a = \bar{y} - b\bar{x}, \quad \text{and} \quad r = \frac{SS(xy)}{\sqrt{SS(x)SS(y)}}$$

$$P(n,r) = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!} \quad \text{and} \quad C(n,r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

$$P(E \cup F) = \Pr(E) + \Pr(F) - \Pr(E \cap F), \qquad \Pr(E') = 1 - \Pr(E),$$

$$\Pr(E \cup F \cup G) = \Pr(E) + \Pr(F) + \Pr(G) - \Pr(E \cap F) - \Pr(E \cap G) - \Pr(F \cap G) + \Pr(E \cap F \cap G)$$

$$\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)}, \qquad \text{Independence:} \qquad \Pr(E \cap F) = \Pr(E)\Pr(F)$$

If $B_1,\ B_2, \ldots,\ B_k$ constitute a partition of S, then

$$\Pr(E) = \Pr(B_1)\Pr(E|B_1) + \Pr(B_2)\Pr(E|B_2) + \cdots + \Pr(B_k)\Pr(E|B_k),$$

and

$$\Pr(B_j|E) = \frac{\Pr(B_j)\Pr(E|B_j)}{\Pr(E)} = \frac{\Pr(B_j)\Pr(E|B_j)}{\Pr(B_1)\Pr(E|B_1) + \cdots + \Pr(B_k)\Pr(E|B_k)}$$

Discrete : $\sum_x f(x) = 1, \quad \Pr(X=x) = f(x), \quad \mu = E[X] = \sum_x x f(x), \quad \sigma^2 = \mathrm{Var}(X) = \sum_x x^2 f(x) - \mu^2,$

$$b(x;n,p) = \binom{n}{x} p^x q^{n-x}, \qquad \mu = np, \qquad \sigma^2 = npq,$$

$$h(x;N,k,n) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad \mu = n \cdot \frac{k}{N}, \qquad \sigma^2 = n \cdot \frac{k}{N} \cdot \left(1 - \frac{k}{N}\right) \cdot \frac{N-n}{N-1},$$

$$g(x;p) = pq^{x-1}, \qquad b^*(x;r,p) = \binom{x-1}{r-1} p^r q^{x-r}, \quad \mu = \frac{r}{p} \quad \sigma^2 = \frac{r(1-p)}{p^2},$$

$$P(x;\lambda) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, \qquad \mu = \sigma^2 = \lambda,$$

Continuous : $\int_{-\infty}^{\infty} f(x)\,dx = 1, \quad \Pr(a < X < b) = \int_a^b f(x)\,dx, \quad \mu = \int_{-\infty}^{\infty} x f(x)\,dx, \quad \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

$$f(x,\lambda) = \frac{1}{\lambda} e^{-x/\lambda} \ (x > 0), \qquad \mu = \sigma = \lambda,$$

$$\mu_{\bar{X}} = \mu, \qquad ,\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}, \qquad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \qquad \text{or} \qquad E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \qquad (\text{with} \quad \text{d.f.} = n-1)$$

$$\hat{p} = \frac{X}{n}, \qquad \mu_{\hat{p}} = p, \quad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}, \qquad Z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \qquad \text{or} \qquad t_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \qquad (\text{with} \quad \text{d.f.} = n-1)$$

$$d = x_1 - x_2, \qquad \mu_{\bar{d}} = \mu_d, \quad \sigma_{\bar{d}} = \frac{\sigma_d}{\sqrt{n}}$$

$$E = t_{\alpha/2} \cdot \frac{S_d}{\sqrt{n}}, \qquad t_0 = \frac{\bar{d} - d_0}{S_d/\sqrt{n}} \qquad (\text{with} \quad \text{d.f.} = n-1)$$

$$\hat{p}_1 = \frac{x_1}{n_1}, \quad \hat{p}_2 = \frac{x_2}{n_2}, \quad \mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2, \quad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$z_0 = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \qquad E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \qquad (\text{with} \quad \text{d.f.} = k-1)$$

$$\chi_0^2 = \sum_i \sum_j \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}, \qquad (\text{with} \quad \text{d.f.} = (r-1)(c-1))$$

$$\left(\text{Expected Value} = \frac{(\text{row total}) \times (\text{column total})}{\text{grand total}}\right)$$

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2, \qquad \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \qquad z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$E = z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad \text{with} \quad \text{d.f.} = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}$$

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \text{where} \quad S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

$$E = t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$