Dawson College
Department of Mathematics
Probability and Statistics
201-SN1-RE
Final Examination — Version 1
Fall 2025

Student Name:______________________

Student I.D#:_______________________

Teachers : R.Acteson, O.Cerba, A.Hariton, G.Honnouvo,
A.Jimenez, V.Kalvin, K.Lagota, M.Moodi,
N.Sabetghadam, B.Seamone,

Date: Friday, December 19, 2025,

Time: 1:30 pm to 4:30 pm

Instructions:

- Print your name and ID in the provided space.
- This exam contains 11 pages (including this cover page) and 20 questions for a total of 100 points.
- Check to see if any pages are missing.
- Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page, and please indicate that you have done so.
- Show all steps and justify your answers for questions **11** to **20**.
- You are only permitted to use the **Sharp EL-531**** calculator.
- This examination booklet must be returned intact.
- Good luck!

Question	Points	Score
1	2	
2	2	
3	2	
4	2	
5	2	
6	2	
7	2	
8	2	
9	2	
10	2	
11	8	
12	6	
13	6	
14	8	
15	7	
16	8	
17	10	
18	11	
19	8	
20	8	
Total	100	

1. (2 points) Which sampling technique was used in the following example? To survey employee satisfaction, divide employees into groups based on their department (e.g., marketing, operations, human resources). Randomly select employees from each department. The number selected from each department is proportional to its size in the company to maintain representativeness.

   a. Simple random sampling
   b. Cluster sampling
   c. Stratified sampling
   d. Systematic sampling
   e. Convenience sampling

2. (2 points) A researcher, conducting a survey at a popular fast-food restaurant, is interested in who orders coffee at the restaurant. For each client, he records the random variable G (gender) as 0 = male and 1 = female; the variable S (coffee size) as 0 = small, 1 = medium, 2 = large; and Y (year of birth) as 92 for a person born in 1992. Which of the following are true about the level of measurement?

   a. G is nominal; S and Y are both ordinal
   b. G is nominal; S and Y are both interval
   c. G is nominal; S is ordinal; Y is ratio
   d. G is nominal; S is ordinal; Y is interval
   e. None of the above

3. (2 points) How many ways can the first, second, and third places be chosen from a group of 10 contestants?

   a. 27
   b. 30
   c. 720
   d. 1000
   e. 120

$\boxed{10}\ \boxed{9}\ \boxed{8} = 720$

4. (2 points) A president, secretary, and treasurer are being chosen. The president must be a woman and the other two must be men. How many ways can this be done if there are 3 women and 4 men available?

   a. 3
   b. 9
   c. 12
   d. 18
   e. 36

President (woman) * Secretary (man) * Treasurer (man)

$\boxed{3} * \boxed{4} * \boxed{3} = 36$

5. (2 points) If the coefficient of correlation is -0.80, the percentage of the variation in y that is explained by the variation in x is:

   a. $\boxed{64\%}$
   b. -64%
   c. 80%
   d. -80%
   e. not possible to answer due to insufficient information.

$r = -.80$

$r^2 = .64 \Leftarrow$ coefficient of determination

6. (2 points) A bag contains 14 red and 6 blue marbles. A marble is drawn, with replacement, four times. That is, the marble is replaced each time after it is drawn. Find the probability that at least one blue marble is drawn. Which value is closest to the answer?

   a. 0.01
   b. 0.99
   c. 0.41
   d. $\boxed{0.76}$
   e. 0.24

$$P(\text{at least one blue}) = 1 - P(\text{no blues})$$
$$= 1 - \left(\frac{14}{20}\right)^4$$
$$= 0.7599$$

7. (2 points) The marking scheme for a course is 20% for test 1, 30% for test 2 and 50% for the final exam. Joan got 28/50 on the first test and 43/60 on the second test. What is the minimum percentage score required on the final exam that Joan would need to pass the course? Assume the passing grade is 60%. Which value is closest to the answer?

   a. 60.0%
   b. $\boxed{54.6\%}$
   c. 52.7%
   d. 62.3%
   e. none of the above

$$(\text{test 1})(.20) + (\text{test 2})(.30) + (\text{final})(.50) = 60\%$$
$$\left(\frac{28}{50}\times 100\right)(.20) + \left(\frac{43}{60}\times 100\right)(.30) + x\,(.50) = 60$$
$$x = \frac{60 - \left(\frac{28}{50}\times 100\right)(.20) + \left(\frac{43}{60}\times 100\right)(.30))}{0.5}$$
$$\simeq 54.6$$

8. (2 points) Suppose two events A and B are independent events. If $P(A) = 0.3$ and $P(A \cup B) = 0.44$, what is $P(B)$?

   a. 0.14
   b. 0.132
   c. 0.31
   d. $\boxed{0.20}$
   e. 0.69

A and B are indep $\Rightarrow P(A \cap B) = P(A)\,P(B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$0.44 = 0.3 + P(B) - P(A)\,P(B)$$
$$0.44 - 0.3 = P(B) - 0.3\,P(B)$$
$$0.14 = (1 - 0.3)\,P(B)$$
$$P(B) = \frac{0.14}{0.7} = 0.2$$

9. (2 points) Two fair dice are rolled until the sum of the two numbers is 7. Let X be the number of rolls required until the first time that the sum of the two numbers is 7. What is the sample space for X?

   a. S = {1,2, 3, . . . , 36}
   b. $\boxed{\text{S} = \{1, 2, 3, \dots\}}$
   c. S = $(2, 36)$
   d. S = $(1, \infty)$
   e. S = {2, 3, . . . , 12}

10. (2 points) A mathematics game uses a set of 44 cards. There are four cards for each of the numbers 0-9 and 4 wild cards. If the cards are shuffled, what is the probability that the top card is a wild card?

   a. $\frac{1}{44}$
   b. $\frac{1}{22}$
   c. $\boxed{\frac{1}{11}}$
   d. $\frac{1}{4}$
   e. None of the above

0 × 4
1 × 4
⋮
9 × 4
W × 4
total 44 cards

$P(\text{Top card is wild card}) = \frac{4}{44} = \frac{1}{11}$

11. (8 points) The table below shows the gender and chosen field of study for six hundred students randomly sampled from a particular university.

Major	Arts	Business	Science	Total
Male	135	81	54	270
Female	165	99	66	330
Total	300	180	120	600

   (a) (6 points) Determine the probability that a randomly chosen student is

      i. Male or Science (or both) $P(M \cup S) = (270 + 120 - 54)/600 = 336/600 = \boxed{0.56}$

      ii. Female and Business $P(F \cap B) = 99/600 = \boxed{0.165}$

      iii. Arts given that she is Female $P(A \mid F) = 165/330 = \boxed{0.5}$

   (b) (2 points) Are the events of being male and studying Science independent? Support your answer.

If independent then $P(M \cap S) = P(M)\,P(S)$

$$P(M \cap S) = \frac{54}{600} = 0.09$$

$$P(M)\,P(S) = \left(\frac{270}{600}\right)\left(\frac{120}{600}\right) = 0.09$$

SAME ∴ events are $\boxed{\text{independent}}$

12. (6 points) The city council recently commissioned a study of park users in their community. Data collected on the age (X) of each person and the amount of time (Y), in hours, spent in the park. Here is the summation of the data values.

$$\sum x = 163 \qquad \sum x^2 = 5465 \qquad \sum xy = 618.1$$
$$\sum y = 26.1 \qquad \sum y^2 = 129.25 \qquad n = 6$$

(a) (4 points) Find the least-squares line.

$$SS_{xy} = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n} = 618.1 - \frac{163(26.1)}{6} = -90.95$$

$$SS_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 5465 - \frac{163^2}{6} = 1036.83$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_x} = \frac{-90.95}{1036.33} = -0.08772$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{26.1}{6} - (-0.08772)\left(\frac{163}{6}\right) = 6.733$$

least squares line:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\boxed{\hat{y} = 6.733 - 0.08772x}$$

(b) [2 Marks] If a person is 5 years old, what amount of time do you predict they spend in the park?

$$\hat{y} = 6.733 - 0.08772(5)$$
$$\approx \boxed{6.29 \text{ hours}}$$

13. (6 points) A small coffee shop owner keeps a daily record of the number X of customers who buy something during the first hour the shop is open. Based on these records, X has the following probability distribution.

(a) (3 points) Find the expected number of customers who make a purchase during the first hour that the shop is open.

X	P(x)	x P(x)	$x^2 P(x)$
0	.1	0	0
1	.1	.1	.1
2	.2	.4	.8
3	.2	.6	1.8
4	.4	1.6	6.4
	TOTAL	2.7 $\Sigma x P(x)$	9.1 $\Sigma x^2 P(x)$

$\mu = E(x) = \Sigma x P(x)$

$= \boxed{2.7 \text{ customers}}$

(b) (3 points) Find the standard deviation of the random variable X.

$\sigma^2 = E(x^2) - \mu^2$

$= 9.1 - (2.7)^2$

$= 1.81$

$\sigma = \sqrt{1.81}$

$\approx \boxed{1.345 \text{ customers}}$

14. (8 points) Suppose it is believed that the probability that a patient will recover from a disease following medication is 0.8. In a group of twenty such patients, $n=20$, $X \sim Bin(n=20, p=0.8)$

   (a) (2 points) find the expected number of patients who recover from the disease.

   $$\mu = np = 20(0.8) = \boxed{16 \text{ patients}}$$

   (b) (2 points) find the variance of the number of patients who recover from the disease.

   $$\sigma^2 = npq = 20(.8)(.2) = \boxed{3.2}$$

   (c) (2 points) find the probability that at least 15 patients recover.

   $$P(x \geq 15) = 1 - P(x \leq 14) = 1 - 0.1958 = \boxed{0.8042}$$

   (d) (2 points) find the probability that five or more patients do not recover from the disease.

   $$X \sim Bin(n=20, p=0.2); \quad P(x \geq 5) = 1 - P(x \leq 4) = 1 - 0.6296 = \boxed{0.3704}$$

15. (7 points) A study asserts that 75% of children under 13 in a certain region have been vaccinated for chicken pox. Assuming this claim is accurate, use the normal approximation with continuity correction to estimate the probability that at least 480 out of 650 randomly selected children were vaccinated.

$X \sim Bin(n=650, p=0.75)$ Approximate $P(X \geq 480)$

check:

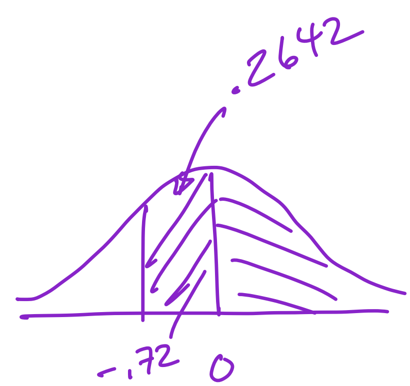
$np = 650(.75) > 5$
$nq = 650(.25) > 5$ } true.

$\therefore$ its is appropriate to use Normal dist to approximate Bin dist

$\mu = np = 650(.75) = 487.5$

$\sigma = \sqrt{npq} = \sqrt{650(.75)(.25)} \approx 11.04$

$$P(X_{Bin} \geq 480) \approx P(X_{Norm} \geq 479.5)$$
$$\approx P\left(Z \geq \frac{479.5 - 487.5}{11.04}\right)$$
$$\approx P(Z \geq -0.7247)$$
$$\approx P(Z \geq -0.72)$$
$$\approx .2642 + .5$$
$$\approx \boxed{0.7642}$$



16. (8 points) The length of human pregnancies from conception to birth varies according to a distribution that is approximately Normal with mean of 266 days and standard deviation of 16 days. Calculate the probability that

$X \sim N(\mu = 266, \sigma = 16)$

(a) (4 points) a randomly selected pregnant woman will have a pregnancy length exceeding 270 days.

$$P(X > 270) = P\left(Z > \frac{270-266}{16} = 0.25\right) \approx 0.5 - 0.0987 \approx \boxed{0.4013}$$

(b) (4 points) $\bar{x}$, the average pregnancy length, for 6 randomly chosen women exceeds 270 days.

$$P(\bar{X} > 270) = P\left(Z > \frac{270-266}{16/\sqrt{6}} = 0.61\right) \approx 0.5 - 0.2291 \approx \boxed{0.2709}$$

17. (10 points) From a random sample of 40 bags of candy, the average weight recorded was 59.2 grams, with a standard deviation of 3.5 grams.

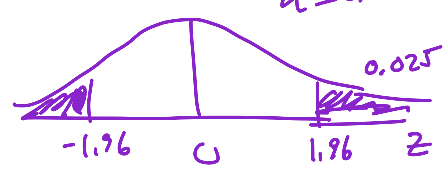
(a) (4 points) Find a 95% confidence interval that approximates the true mean weight of the entire population of these candy bags.

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$59.2 \pm 1.96 \frac{(3.5)}{\sqrt{40}}$$

95% C.I. for μ

$$\boxed{(58.1, 60.3)}$$

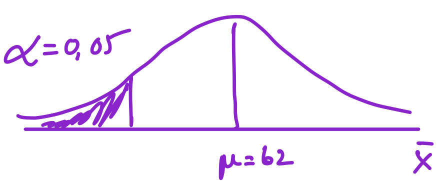


(b) (6 points) The label states that each bag contains 62 grams of candy, but we suspect the true mean weight is lower. Does the sample of 40 bags provide evidence that the average weight is less than 62 grams? Perform the appropriate hypothesis test at the 0.05 significance level.

i. State the null and alternative hypotheses.

ii. Identify the rejection region(s).

iii. What conclusion is reached?

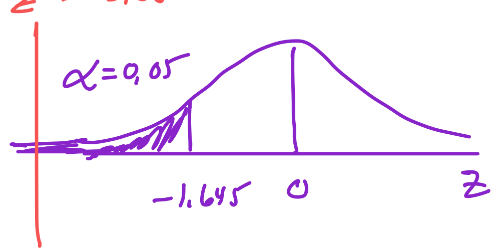


$H_0: \mu = 62$

$H_a: \mu < 62$

test statistic

$$Z^* = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{59.2 - 62}{3.5/\sqrt{40}} = -5.06$$



Since $Z^* < -1.645$ then there is sufficient evidence to reject H_0.

18. (11 points) A recent survey found that 80% of kids spend more than 2 hours per day playing games on their phones or online. A disbeliever took a separate random sample of 500 kids in the attempt to show that the true percentage of kids who spend more than 2 hours per day playing games on their phones or online is less than 80%.

(a) (5 points) State the null and alternative hypotheses. Find the p-value if 384 of the surveyed kids said that they spend more than 2 hours per day playing games on their phones or online.

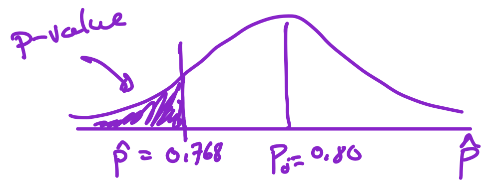
$H_0: p = 0.80$

$H_a: p < 0.80$

$n = 500$

$x = 384$

$\hat{p} = \frac{x}{n} = \frac{384}{500} = 0.768$



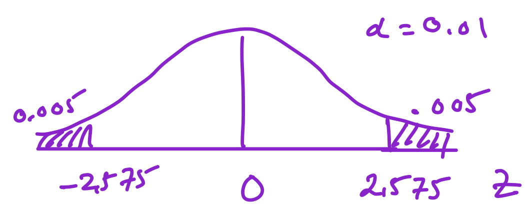
$$z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.768 - 0.80}{\sqrt{\frac{(.8)(.2)}{500}}} = -1.79$$

$$\text{p-value} = P(Z < z^*) = P(Z < -1.79) \approx 0.5 - 0.4633 \approx \boxed{0.0367}$$

(b) (2 points) Suppose the level of significance is $\alpha = 0.05$, is there sufficient evidence to reject the null hypothesis, based on the p-value found in part a?

Since p-value $< \alpha$ then there is sufficient evidence to reject H_0.

(c) (4 points) Determine the sample size required to estimate the true proportion of kids who play more than 2 hours per day of games on their phone or online if you want the estimate to be within 1% with 99% confidence? Use a preliminary value of $\hat{p} = 0.80$ which was taken from a small sample.



$E = 1\% = 0.01$

using $\hat{p} = 0.8$ from part a).

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{E^2} = \frac{(2.575)^2(.8)(.2)}{(0.01)^2} = \boxed{10609}$$

19. (8 points) Researchers at a paint manufacturer claim that the addition of a new accelerator will reduce the drying time by more than 5%. Five test samples were conducted, with the following percentage decreases in drying time:

5.2 6.4 6.8 4.7 4.2

Assume that the percentage decrease in drying time for all paint are normally distributed. Conduct the appropriate test of hypothesis, at a level of significance of 0.05. Does the data support the claim?

i. State the null and alternative hypotheses.

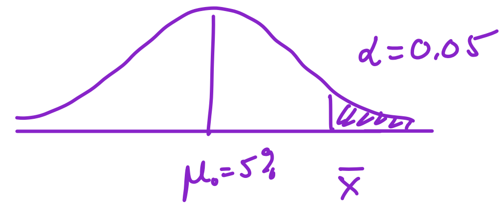
ii. Identify the rejection region(s).

iii. What conclusion is reached?

sample $\begin{cases} \bar{x} = 5.46\% \\ s = 1.108\% \\ n = 5 \end{cases}$

$H_0: \mu = 5\%$

$H_a: \mu > 5\%$ (Claim)

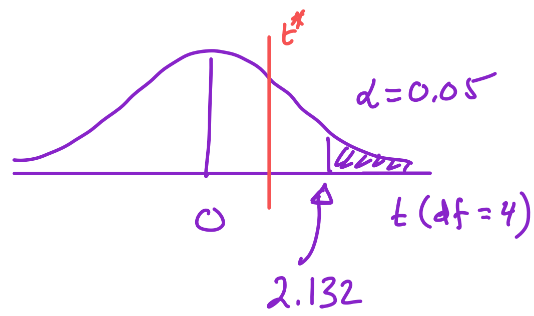


t-dist ($df = n-1 = 4$)

test statistic

$$t^* = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{5.46 - 5}{1.108/\sqrt{5}} = 0.928$$



Since $t^* < 2.132$ then there is insufficient evidence to reject H_0 and to support the claim that the new addition of the accelerator will reduce the drying time by more than 5%.

20. (8 points) A survey conducted among randomly selected travelers who visited service-station restrooms along the 401 yielded the following findings:

	Quality of Restrooms			
Gender	Above average	Average	Below average	Total
Female	7 $E_{11}=8.85$	24 $E_{12}=29.5$	28 $E_{13}=20.65$	59
Male	8 $E_{21}=6.15$	26 $E_{22}=20.5$	7 $E_{23}=14.35$	41
Total	15	50	35	100

At the 0.05 significance level, does the sample offer sufficient evidence to conclude that the ***quality of facilities*** is not independent of ***respondent gender***?

<u>Test for independence</u>

H_0: Quality of facilities & Gender are independent
H_a: " " " " " dependent.

If independent then prob(Female and Above average) = $P(F)\,P(\text{Above}) = \left(\frac{59}{100}\right)\left(\frac{15}{100}\right)$
And the expected value of (F and Above) $= n p_{11} = 100\left(\frac{59}{100}\right)\left(\frac{15}{100}\right) = \frac{59(15)}{100} = 8.85$

Similarly, $E(\text{Male and Below average}) = \frac{41(35)}{100} = 14.35$

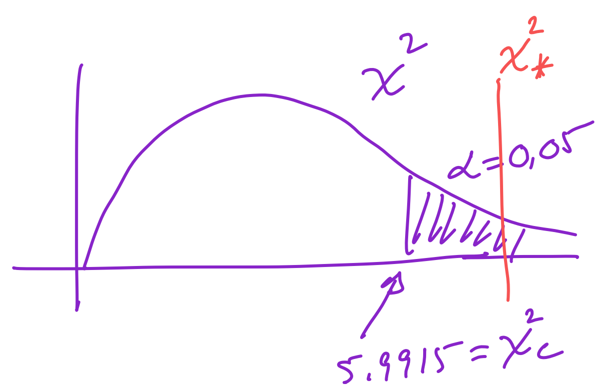
<u>test statistic</u>

$$\chi^2_* = \sum_{ij} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$= \frac{(7-8.85)^2}{8.85} + \frac{(8-6.15)^2}{6.15} + \frac{(24-29.5)^2}{29.5} + \frac{(26-20.5)^2}{20.5} + \frac{(28-20.65)^2}{20.65} + \frac{(7-14.35)^2}{14.35}$$

$$= 9.825$$

$$\chi^2_c = 5.9915$$



Since $\chi^2_* > \chi^2_c$ then there is sufficient evidence to reject H_0 and that Quality of facilities and Gender are <u>not</u> independent

$df = (r-1)(c-1)$
$= (2-1)(3-1) = 2$