DAWSON COLLEGE
MATHEMATICS DEPARTMENT
FINAL EXAMINATION
201-401-DW STATISTICS FOR SOCIAL SCIENCE, Sections 00001 and 00002
THURSDAY, MAY 19th, 2022, 9:30 - 12:30

Student Name: ______________________________________

Student I.D.: ______________________________________

Instructors: Mehdi Moodi and Ahed Hindawi

TIME: (3hours)

Instructions:

- Print your name and student ID number in the space provided above.
- All questions are to be answered directly on the examination paper.
- Read every question very carefully and write your solutions legibly for what I cannot read will be considered incorrect. Show all your work step by step. Avoid giving equivocal and ambiguous answers.
- Provide exact answers where possible. If an exact answer is not possible, round your answer to the nearest 10000th, i.e. 4 digits after the decimal point.
- Students are only permitted to use the Sharp EL-531** calculator during tests and examinations.
- This examination consists of 15 questions. Please ensure that you have a complete exam booklet before starting.
- This booklet must be returned intact.

Solutions

PART 1 (7 questions with 3 points for each)

1) Given A and B are two mutually exclusive events such that P(A)=0.3 and P(B)=0.4. Then $P(A|B)$ will be equal to

a) 0.7 b) 0.12 c) 0 d) 1-(0.3)(0.4) e) 1

2) Given A and B are two independent events such that P(A)=0.2, P(B)=0.4. Then $P(A \cup B)$ will be equal to

a) 0.08 b) 0 c) 0.6 d) 0.52 e) 0.4

3) Which one of the descriptions represents a binomial random variable?

a) The number of times you roll a die until you get 6
b) The number of defective batteries in a sample of 50 randomly selected batteries coming from a manufacturing process in which 0.01% of all batteries are defective.
c) A box contains 6 red and 4 blue marbles. You draw 3 marbles without replacement and count the number of red marbles drawn.
d) The number of traffic accidents per day in a particular city
e) The weight of a certain species of farmed fish at harvest

3) Two fair dice are rolled at once. The probability that the sum of dots appearing on the top faces of the two dice is greater than or equal to 10 will be.

a) 3/36 b) 4/36 c) 2/36 d) 4/6 e) 1/6

4) Which one of these options is a discrete random variable?

a) The time between customers entering a checkout lane at a retail store.
b) The average amount spent on electricity each July by a randomly selected household in a certain state.
c) The amount of rain recorded at an airport one day.
d) The number of coins that match when three coins are tossed at once.
e) The amount of liquid in a 12-ounce can of soft drink.

5) The null hypothesis for a certain test is rejected at the 0.01 level of confidence. Which of the following is a valid conclusion?

a) The null hypothesis is definitely wrong.
b) The risk of rejecting the null hypothesis by mistake is 99%.
c) The risk of rejecting the null hypothesis by mistake is 1%.
d) Type two error is equal to 0.01.
e) None of the above.

6) Suppose that the test statistic from a one-sample hypothesis test does not fall in the critical region. Which of these statements would be a correct interpretation of this result?

a) The alpha level should be changed.
b) The null hypothesis is false and the alternative hypothesis is supported.
c) The null hypothesis cannot be rejected.
d) The null hypothesis is false and the alternative hypothesis is true.

7) A sample of size 15 drawn from a normally distributed population has sample mean 35 and sample standard deviation 14. In order to construct a 95% confidence interval for the population mean, we use

a) $\bar{x} + z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)$
b) $\bar{x} + t_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$
c) $\bar{x} + z_{\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}})$
d) $\bar{x} + t_{\frac{\alpha}{2}}(\frac{s}{\sqrt{n}})$

PART 2

1) [6 marks] The following two-way contingency table gives the breakdown of the co-owners in a particular residential complex according to their age and opinion on a proposed land-scaping project:

	Under 30	Between 30 and 65	Over 65	
For	18	25	39	82
Against	24	10	7	41
Undecided	13	5	11	29
	55	40	57	152

We randomly select one individual from this population. Given:
$A =$ *The event that the selected person is under* 30
$B =$ *The event that the selected person is against the project*
Calculate $P(A), P(B), P(A \cap B), P(A \cup B), P(A|B)$ and $P(B|A)$.

$P(A) = \frac{55}{152}$ $P(B) = \frac{41}{152}$ $P(A \cap B) = \frac{24}{152}$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{72}{152}$

$P(A|B) = \frac{24}{41}$ $P(B|A) = \frac{24}{55}$

2) [4+2+2 marks] One thousand lottery tickets are sold for \$3 each. There will be one prize of \$1000 and two prizes of \$500 and five prizes of \$100. Let X denote the net gain from the purchase of a random ticket.

a) Construct the probability distribution of X.

X				
P				

b) Compute the expected value $E(X)$.

c) Compute the standard deviation σ of X.

x	997	497	97	−3
$P(x)$	0.001	0.002	0.005	0.992
$x\,P(x)$	0.997	0.994	0.485	−2.976
$x^2P(x)$	994.009	494.018	47.045	8.928

$\mu = \sum x P(x) = -0.5$

$\sum x^2 P(x) = 1544$

$\sigma = \sqrt{1544 - (-0.5)^2} = \sqrt{1543.75} \approx 39.29058$

3) [4+4 marks] A salesman makes a sale on 25% of his calls on regular customers. He makes **five** sales calls each day.

a) Construct the probability distribution of X, the number of sales made each day.

X	0	1	2	3	4	5
P	0.2373	0.3955	0.2637	0.879	0.0146	0.00098

b) Assuming that the salesman makes 250 sales calls per year, find the mean and standard deviation of the number of sales made per year.

$X \sim Bi(n=250,\ p=0.25)$

$\mu = np = 62.5$

$\sigma = \sqrt{np(1-p)} = \sqrt{(250)(0.25)(0.75)} = 6.846532$

4) [3+3 marks] The lengths of time taken by students on an exam (if not forced to stop before completing it) are normally distributed with mean 68 minutes and standard deviation 4.6 minutes.

a) Find the proportion of students who will finish the exam if a 75-minute time limit is set.

b) What time limit should be set so that 98% of the students finish the exam?

$X \sim N(\mu = 68, \sigma = 4.6)$

a) $P(X \leq 75) = P(Z \leq 1.52) \simeq 0.9357$

b) $z^* = 2.055$ $\quad x^* = \mu + \sigma z^* = 77.453$

5) [5 marks] Suppose that in a certain region of the country the mean duration of a marriage that ends in divorce is 8.2 years with standard deviation 2.4 years. Find the probability that in a sample of 35 divorces, the mean age of the marriages is at most 9 years.

$\bar{X} \sim N\left(\mu_{\bar{X}} = 8.2, \sigma_{\bar{X}} = \frac{2.4}{\sqrt{35}}\right)$

$P(\bar{X} \leq 9) = P(Z \leq 1.97) \simeq 0.9756$

6) [7 marks] In a study of dummy foal syndrome, the average time between birth and onset of noticeable symptoms in a sample of six foals was 20.2 hours, with sample standard deviation 1.8 hours. Assuming that the time to onset of symptoms in all foals is normally distributed, construct a 95% confidence interval for the mean time between birth and onset of noticeable symptoms.

$n=6 \quad \bar{x}=20.2 \quad s=1.8 \quad t_{0.025,5}=2.571$

$20.2 \pm 2.571 \frac{1.8}{\sqrt{6}} = 20.2 \pm 1.8893$

$[18.3107, 22.0893]$

7) [5 marks] An economist wishes to estimate, to within 3 minutes, the mean time that employed persons spend commuting each day, with 98% confidence. On the assumption that the standard deviation of commuting times is 10 minutes, estimate the minimum size sample required.

$E=3 \quad \alpha=0.02 \quad \sigma=10 \qquad n \geq \left(\frac{(2.33)(10)}{3}\right)^2 = 60.32$

$n=61$

8) [7 marks] A magazine publisher tells potential advertisers that the mean household income of its regular readership is $61,500. An advertising agency wishes to test this claim against the alternative that the mean is different. A sample of 40 randomly selected regular readers yields mean income $59,800 with standard deviation $5,250. Perform the relevant test at the 5% level of significance. Answer either by constructing rejection regions (critical value approach) or by estimating the p-value of the test and comparing it to α.

$$\begin{cases} H_0: \mu = 61500 \\ H_a: \mu \neq 61500 \end{cases} \qquad \bar{X} = 59800 \quad S = 5250$$

$$Z = \frac{59800 - 61500}{5250/\sqrt{40}} = -2.05$$

P-value $= P(\cancel{P}\ Z < -2.05) = 2(0.0202) = 0.0404$

since P-value $< \alpha$ we reject H_0

9) [1+4+1 marks] The city wishes to investigate the effectiveness of a campaign for composting. Historically 15% of all residents in the city regularly compost their bio-waste. In a survey commissioned by the city, 270 of 1,500 randomly selected residents stated that they compost regularly.

a) Find the sample proportion.
b) Find the probability that, when a sample of size 1,500 is drawn from a population in which the true proportion is 0.15, the sample proportion will be at least as large as the value you computed in part (a). You may assume that the normal distribution applies.
c) Give an interpretation of the result in part (b). How strong is the evidence that the campaign to promote composting has been effective?

a) $\hat{p} = \frac{270}{1500} = 0.18$

$$\hat{p} \sim N\left(\mu_{\hat{p}} = p,\ \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\right)$$

$$\hat{p} \sim N\left(\mu_{\hat{p}} = 0.15,\ \sigma_{\hat{p}} = \sqrt{\frac{(0.15)(0.85)}{1500}}\right)$$

b) $P(\hat{p} \geq 0.18) = P(Z \geq 3.25) \approx 0.0006$

c) There is strong evidence for the effectiveness of the campaign. The probability of observing 3% increase in the proportion of people who compost by sheer chance is close to zero.

10) [7 marks] The average number of days to complete recovery from a particular type of knee operation is 145 days. From his experience a physician suspects that use of a topical pain medication might be **shortening** the recovery time. He randomly selects the records of seven knee surgery patients who used the topical medication and calculates a sample mean of 142 days and a sample standard deviation 3 days. Assuming a normal distribution of recovery times, perform the relevant test of hypotheses at the 10% level of significance.

$$\begin{cases} H_0: \mu = 145 \\ H_a: \mu < 145 \end{cases} \qquad t = \frac{142-145}{3/\sqrt{7}} = -2.65 \qquad t_{0.1,6} = 1.44$$

since $-2.65 < 1.44$ we reject H_0

There is statistically significant evidence for the effectivenen of the topical pain medication at this level of significance.

11) [7 marks] Records of 40 used passenger cars and 40 used pickup trucks (none used commercially) were randomly selected to investigate whether there was any difference in the mean time in years that they were kept by the original owner before being sold. For cars the mean was 5.3 years with standard deviation 2.2 years. For pickup trucks the mean was 7.1 years with standard deviation 3.0 years. Based on these data, test the hypothesis that there is a difference in the means against the null hypothesis that there is no difference. Use the 1% level of significance.

$$\begin{cases} H_0: \mu_t - \mu_c = 0 \\ H_a: \mu_t - \mu_c \neq 0 \end{cases} \qquad Z = \frac{(7.1-5.3)-0}{\sqrt{\frac{2.2^2}{40} + \frac{3^2}{40}}} = 3.06$$

$$Z_{0.005} = 2.575$$

since $3.06 > 2.575$ we reject H_0

12) [7 marks] A genetic engineering company claims that it has developed a genetically modified tomato plant that yields on average more tomatoes than other varieties. A farmer wants to test the claim on a small scale before committing to a full-scale planting. Ten genetically modified tomato plants are grown from seeds along with ten other tomato plants. At the season's end, the resulting yields in pound are recorded as below. Construct the 99% confidence interval for the difference in the population means based on these data.

	$\bar{X}$	S
GM tomatoes	22.1	3.2
None GM tomatoes	20.8	2.8

$$S_p^2 = \frac{(9)(3.2)^2 + 9(2.8)^2}{18} = 9.04 \qquad t_{0.005,18} = 2.878$$

$$E = (2.878)(0.7755) = 2.2318$$

$$22.1 - 20.8 = 1.3$$

$$-0.9318 = 1.3 - 2.2318 \leq \mu_{GM} - \mu_{NGM} \leq 1.3 + 2.2318 = 3.5318$$

$$[-0.9318, 3.5318]$$