5/11

2404E368

Candidate's Seat No :_____

IM.Sc. (AIML) Sem.-4 Examination

CC-212

Statistical Inference Theory

Time : 2-30 Hours] **April-2025** **[Max. Marks : 70**

Instructions: All questions are compulsory. Use of non-programmable scientific calculator is allowed.

Q.1 **(a)** The mean annual cost of automobile insurance is \$939 (CNBC, February 23, 2006). Assume that the standard deviation is $\sigma = \$245$. **(07)**

a. What is the probability that a simple random sample of automobile insurance policies will have a sample mean within \$25 of the population mean for each of the following sample sizes: 30, 50, 100, and 400?

b. What is the advantage of a larger sample size when attempting to estimate the population mean?

(b) Suppose a simple random sample of size 50 is selected from a population with σ_{10}. Find the value of the standard error of the mean in each of the following cases (use the finite population correction factor if appropriate). **(07)**

a. The population size is infinite.

b. The population size is $N_{50,000}$, N_{5000}, N_{500}.

OR

(a) The average score for male golfers is 95 and the average score for female golfers is 106. Use these values as the population means for men and women and assume that the population standard deviation is $\sigma = 14$ strokes for both. A simple random sample of 30 male golfers and another simple random sample of 45 female golfers will be taken. **(07)**

a. What is the probability that the sample mean is within 3 strokes of the population mean for the sample of male golfers?

b. What is the probability that the sample mean is within 3 strokes of the population mean for the sample of female golfers?

c. In which case, part (a) or part (b), is the probability of obtaining a sample mean within 3 strokes of the population mean higher? Why?

(b) A population has a mean of 200 and a standard deviation of 50. Suppose a simple random sample of size 100 is selected and is used to estimate μ. **(07)**

a. What is the probability that the sample mean will be within ± 5 of the population mean?

b. What is the probability that the sample mean will be within ± 10 of the population mean?

Q.2 **(a)** The Employment and Training Administration reported that the U.S. mean unemployment insurance benefit was \$238 per week. A researcher in the state of Virginia anticipated that sample data would show evidence that the mean weekly unemployment insurance benefit in Virginia was below the national average. **(07)**

a. Develop appropriate hypotheses such that rejection of H_0 will support the researcher's contention.

(P.T.O)

b. For a sample of 100 individuals, the sample mean weekly unemployment insurance benefit was \$231 with a sample standard deviation of \$80.
c. What is the p-value?
d. At $\alpha = 0.05$, what is your conclusion?

(b) Fowley Marketing Research, Inc., bases charges to a client on the assumption that telephone surveys can be completed within 15 minutes or less. If more time is required, a premium rate is charged. With a sample of 35 surveys, a population standard deviation of 4 minutes, and a level of significance of .01, the sample mean will be used to test the null hypothesis H0: $\mu \leq 15$. **(07)**

a. What is your interpretation of the Type II error for this problem? What is its impact on the firm?
b. What is the probability of making a Type II error when the actual mean time is $\mu = 17$ minutes?
c. What is the probability of making a Type II error when the actual mean time is $\mu = 18$ minutes?

OR

(a) Consider the following hypothesis test: **(07)**

$$H_0: P \geq 0.75$$
$$H_1: P < 0.75$$

A sample of 300 items was selected. Compute the p-value and state your conclusion for each of the following sample results. Use $\alpha = 0.05$.

a. $\bar{p} = 0.68$
b. $\bar{p} = 0.70$
c. $\bar{p} = 0.72$
d. $\bar{p} = 0.77$

(b) The Coca-Cola Company reported that the mean per capita annual sales of its beverages in the United States as 423 eight-ounce servings. Suppose you are curious whether the consumption of Coca-Cola beverages is higher in Atlanta, Georgia, the location of Coca-Cola's corporate headquarters. A sample of 36 individuals from the Atlanta area showed a sample mean annual consumption of 460.4 eight-ounce servings with a standard deviation of $s = 101.9$ ounces. Using $\alpha = 0.05$, do the sample results support the conclusion that mean annual consumption of Coca-Cola beverage products is higher in Atlanta? **(07)**

Q.3 **(a)** The following results come from two independent random samples taken of two populations. **(07)**

Sample 1	Sample 2
$n_1 = 50$	$n_2 = 35$
$x_1 = 13.6$	$x_2 = 11.6$
$\sigma_1 = 2.2$	$\sigma_2 = 3.0$

a. What is the point estimate of the difference between the two population means?
b. Provide a 90% confidence interval for the difference between the two population means.

c. Provide a 95% confidence interval for the difference between the two population means.

(b) Consider the following hypothesis test. **(07)**

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 \neq 0$$

The following results are for two independent samples taken from the two populations.

Sample 1	Sample 2
$n_1 = 80$	$n_2 = 70$
$x_1 = 104$	$x_2 = 106$
$\sigma_1 = 8.4$	$\sigma_2 = 7.6$

a. What is the value of the test statistic?
b. What is the probabilistic value?
c. With significance level 0.05, what is your testing conclusion?

OR

(a) Consider the following hypothesis test. **(07)**

$$H_0: \mu_1 - \mu_2 \leq 0$$
$$H_1: \mu_1 - \mu_2 > 0$$

The following results are for two independent samples taken from the two populations.

Sample 1	Sample 2
$n_1 = 40$	$n_2 = 50$
$x_1 = 25.2$	$x_2 = 22.8$
$\sigma_1 = 5.2$	$\sigma_2 = 60$

a. What is the value of the test statistic?
b. What is the probabilistic value?
c. With significance level 0.05, what is your testing conclusion?

(b) The following results come from two independent random samples taken of two populations. **(07)**

Sample 1	Sample 2
$n_1 = 12$	$n_2 = 15$
$x_1 = 4.5$	$x_2 = 12.6$
$\sigma_1 = 1.5$	$\sigma_2 = 4.5$

a. What is the point estimate of the difference between the two population means?
b. Provide a 90% confidence interval for the difference between the two population means.
c. Provide a 95% confidence interval for the difference between the two population means.

Q.4 **(a)** Use the Kruskal-Wallis test to determine whether there is a significant difference in the following groups. Use significance level of 0.05. **(07)**

Group 1	19	21	29	22	37	42	
Group 2	30	38	35	24	29		
Group 3	39	32	41	44	30	27	33

(P.T.O)

(b) Use the Wilcoxon matched pairs signed rank test and for significance level 0.05 to analyse the before-and-after measurements given. Assume the underlying distributions are symmetrical. **(07)**

Before	After
49	43
41	29
47	30
39	38
53	40
51	43
51	46
49	40
38	42
54	50
46	47
50	47
44	39
49	49
45	47

OR

(a) Use the Mann-Whitney U test and the following data to determine whether there is a significant difference between the values of group 1 and group 2. Let $\alpha = 0.05$. **(07)**

Group 1	Group 2
15	23
17	14
26	24
11	13
18	22
21	23
13	18
29	21

(b) Test the following sequence of observations by using the runs test and $\alpha = 0.05$ to determine whether the process produced random results. **(07)**

M M N N N N N M M M M M M N N M M M M M N M M N N N N N N N N N N N N N M M M M M M M M M M M

Q.5 Attempt any **SEVEN** out of **TWELVE:** **(14)**

(1) What is the difference between cluster sampling and convenience?

(2) A statistician calculates a 95% confidence interval for μ when σ is known. The confidence interval is Rs. 18000 to Rs. 22000, what is the amount of the sample mean?

(3) If we have normal populations with known population standard deviations σ1 and σ2, the confidence interval estimate for the difference between two population means $(\mu_1 - \mu_2)$ is _____.

(4) By increasing the sample size, the precision of confidence interval is _____.

(5) Define point estimator and interval estimator.

(6) The region of acceptance of H_0 is called _____.

(7) Write the Condition for applying the Central Limit Theorem (CLT) which approximates the sampling distribution of the mean with a normal distribution.

(8) Consider a set of 18 samples from a standard normal distribution. We square each sample and sum all the squares then what will be the number of degrees of freedom for a Chi Square distribution?

(9) What is the range of chi square statistics?

(10) For $r \times c$ contingency table, what will be the degree of freedom for chi square test?

(11) Give the name of a non-parametric test that will be used to test the randomness.

(12) Three brands of coffee are rated for taste on a scale of 1 to 10. Six persons are asked to rate each brand so that there is a total of 18 observations. The appropriate test to determine if three brands taste equally good is_____.

_____ **** _____ **** _____