U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

February 2025 Semester End Main Examinations

Programme: B.E. **Semester: IV**

Branch: Artificial Intelligence and Machine learning **Duration: 3 hrs.**

Course Code: 24AM4PCIST **Max Marks: 100**

Course: Inferential Statistics

Instructions: 1.Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT - I	*CO*	*PO*	Marks
1	a)	Define the following: i. Population ii. Sample iii. Parameter iv. Statistic v. Estimate and Estimator.	*CO1*	*PO1*	**06**
	b)	Show that sample mean $\bar{x}$ is an unbiased and consistent estimator of p, when the sample is drawn from Bernoulli distribution B (1, p), where p is the probability of success.	*CO2*	*PO2*	**07**
	c)	How does a parametric test differ from a non-parametric test? Let $X_1, X_2, \ldots, X_n$ be a random sample from N(Θ,1) distribution. Find the sufficient estimator of Θ.	*CO2*	*PO2*	**07**
		OR			
2	a)	Suppose $X_1, X_2, \ldots X_8$ is a random sample drawn from a population with mean μ and variance σ^2. Consider the estimates $Y = \frac{X_1+X_2+\ldots+X_8}{8}$ and $Z = \frac{X_1+2X_2+\ldots+8X_8}{64}$. Discuss whether they are unbiased for μ. Also obtain the relative efficiency of Y over Z.	*CO2*	*PO2*	**10**
	b)	Consider the problem of estimating the average weight of students in a school. i. Discuss why interval estimation is preferred over point estimation in this scenario. ii. Explain the role of sample size in improving the precision of the interval estimate.	*CO2*	*PO2*	**10**
		UNIT - II			
3	a)	Highlight the key differences between probability sampling and non-probability sampling methods by considering examples for each.	*CO1*	*PO1*	**06**
	b)	In a population of 4 units whose values are 4, 7, 8 and 1. Calculate the sample mean for all possible simple random samples of size 2	*CO1*	*PO2*	**08**

			with replacement (WR) and verify that the sample mean is an unbiased estimate of the population mean.			
		c)	In a city, there are 25 steel suppliers who provide critical raw material for a production unit. Each supplier is assessed based on their reliability in delivering steel on time and rated on a scale from 1 to 5. The ratings of the suppliers are as follows: 4, 3, 5, 2, 4, 3, 5, 1, 3, 4, 5, 2, 3, 4, 5, 1, 3, 4, 5, 2, 3, 4, 5, 1, 3. i) Select all possible systematic samples of size 5. ii) Show that the sample mean is an unbiased estimator of the population mean.	*CO1*	*PO2*	**06**
			OR			
	4	a)	i. A population is divided into four strata with the following sizes: **Stratum** | 1 | 2 | 3 | 4 **Number of individuals** | 50 | 100 | 150 | 200 select a total sample of 50 individuals using equal and proportional allocation. ii. In the context of a health survey, define stratified sampling and cluster sampling. Explain when cluster sampling would be more appropriate than stratified sampling.	*CO1*	*PO2*	**10**
		b)	Explain the differences between cluster sampling and stratified sampling with an example.	*CO1*	*PO1*	**05**
		c)	Justify the use of simple random sampling with replacement over simple random sampling without replacement, despite the higher variance. Provide an example to support your explanation.	*CO1*	*PO1*	**05**
			UNIT-III			
	5	a)	Distinguish between one tailed and two-tailed tests in the context of testing of hypothesis with suitable examples.	*CO1*	*PO1*	**06**
		b)	Let p be the probability that a coin will fall head in single toss. In order to test the hypothesis Ho: P = 0.5 the coin is tossed 6 times and Ho is rejected if more than 4 heads are obtained. Find the probability of i) Type I error. ii) Type II error if H_1: P=1/3	*CO1*	*PO2*	**08**
		c)	In a certain factory there are two independent processes, manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 kg with a standard deviation of 12 kg while the corresponding figures in a sample of 400 items from the other process are 124 and 14 respectively. Find the 99% confidence limits for the difference in the average weights of items produced by the two processes respectively.	*CO1*	*PO2*	**06**
			OR			
	6	a)	Discuss the different types of hypotheses used in statistical analysis and research. Explain the significance of each type and provide examples to illustrate their applications in real-world scenarios	*CO3*	*PO1*	**05**

		b)	Compare and contrast Type I and Type II errors in hypothesis testing. i. Discuss which error is more critical and why. ii. Explain why the Type I error rate is fixed in statistical inference, and iii. Power of a test and its significance.	*CO3*	*PO1*	**07**
		c)	A factory claims the average lifetime of its light bulbs is 1000 hours. A sample of 15 light bulbs is selected, and the following lifetimes (in hours) are recorded: 980, 1025, 990, 1055, 1010, 1005, 1025, 990, 985, 1000, 1015, 1040, 1015, 1030, 998 The critical region for the test is defined as any sample median outside the range of 970 to 1020 hours. i. State the null and alternative hypotheses. ii. Calculate the sample median of the lifetimes. iii. Determine whether to reject the null hypothesis based on the sample median and the critical region. iv. Observe the change in the decision when sample mean is outside the range of 970 to 1020 hours is used as critical region	*CO2*	*PO2*	**08**
			UNIT - IV			
	7	a)	Provide the test statistic for the following testing i. Single proportion mean for small sample ii. Difference of two population proportion for large sample iii. Difference of two population means for large sample	*CO3*	*PO1*	**06**
		b)	Suppose we are conducting hypothesis tests i. $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ ii. $H_0: \mu = \mu_0$ versus $H_2: \mu > \mu_0$ Calculate the P- value for the following observed values of the test statistic Z_0: 2.25 and Z_0: 0.10 for both the cases.	*CO3*	*PO2*	**06**
		c)	Two groups A and B consists of 100 people suffering from a disease. When a serum is given to group A but not to group B, it is found that 75 and 65 people in group A and B respectively recovered from the disease. Test the hypothesis that the serum helps to cure the disease.	*CO3*	*PO2*	**08**
			OR			
	8	a)	Outline the large sample procedure for testing the equality of two population means.	*CO3*	*PO1*	**06**
		b)	Globally the long-term proportion of newborns who are male is 51.46%. A researcher believes that the proportion changes under severe economic conditions. To test this belief, a sample of 5000 birth records from a period of economic recession was examined, and it was found that 52.55% of the newborns were boys. Is there sufficient evidence at the 5% level of significance to support the researcher's belief that the proportion of boys born during an economic recession differs from the global proportion of 51.46%?	*CO3*	*PO2*	**08**
		c)	The nicotine content in milligrams (mg) in cigarettes of a certain brand is normally distributed with mean μ and standard deviation σ =0.1mg. The brand advertises that the mean nicotine content of its cigarettes is 1.5 mg, but measurements on a random sample of	*CO3*	*PO2*	**06**

			100 cigarettes of this brand gave mean $\bar{x}$ =1.53 mg. Is this evidence that the mean nicotine content is actually higher than advertised? To answer this question, we test the hypothesis H_o: μ=1.5 H_1: μ > 1.5 Using the 5% level of significance, which of the following is the correct conclusion from the test			
			UNIT - V			
9	a)		Discuss the key assumptions of the t-distribution and explain why each assumption is important. What are the potential consequences if these assumptions are not satisfied?	*CO3*	*PO1*	**05**
	b)		The average breaking strength of the steel rods is specified to be 18.5 thousand kg. For this, a sample of 14 rods were tested. The mean and standard deviation obtained were 17.85 and 1.955, respectively. Solve and test the significance of deviation at α=5% level.	*CO3*	*PO2*	**05**
	c)		The sales data of an item in six shops before and after a special promotional campaign are as follows: **Shops** | **A** | **B** | **C** | **D** | **E** | **F** **Before campaign** | 53 | 28 | 31 | 48 | 50 | 42 **After campaign** | 58 | 29 | 30 | 55 | 56 | 45 Can the campaign be judged as success at 5% level of significance?	*CO3*	*PO2*	**10**
			OR			
10	a)		A chemical engineer is investigating the inherent variability of two types of test equipment that can be used to monitor the output of a production process. He suspects that the old equipment, type 1, has a larger variance than the new one. Two random samples of $n_1 = 12$ and $n_2 = 10$ observations are taken, and the sample variances are $S_1^2 = 14.5$ and $S_2^2 = 10.8$.	*CO3*	*PO2*	**05**
	b)		A study was conducted to evaluate whether a new sequence of operation elements reduces the mean cycle time of a production procedure. Data from 5 cycles under the new sequence were recorded as follows: 12.37, 11.97, 12.15, 12.08, 12.31 minutes. The current mean cycle time is 12.5 minutes. At a significance level of 5%, should the new sequence be adopted based on the observed data?	*CO3*	*PO2*	**05**
	c)		A market analyst wished to see whether consumers have any preference among 6 flavors of a new fruit soda. A sample of 156 people provided these data. **Flavor** | **Cherry** | **Strawberry** | **Orange** | **Lime** | **Grape** | **Ginger** **No. of people** | 42 | 38 | 26 | 14 | 10 | 26 Is there enough evidence to reject the null hypothesis that there is no preference in the selection of fruit soda flavors at 0.05 significance level?	*CO3*	*PO2*	**10**

******