2025

STATISTICS — HONOURS

Paper : DSCC-4

(Statistical Inference - I)

Full Marks : 75

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

1. Answer ***any five*** questions : 2×5

(a) In a study to estimate the average marks of students in a college, a random sample of 50 students is selected. Identify the population, sample, parameter and statistic.

(b) Distinguish between a population distribution and a sampling distribution.

(c) The heights of all students in a college follow a normal distribution with mean 165 cm and standard deviation 10 cm. State the mean and standard deviation of the sampling distribution of the sample mean for samples of size 25.

(d) Distinguish between an estimator and an estimate.

(e) What is meant by the efficiency of an estimator?

(f) Give an example of a composite null hypothesis.

(g) A medical screening test rejects the null hypothesis that a patient is disease-free even when the patient is actually healthy with probability 0.02. Interpret this statement.

(h) A researcher claims that the average weight of a product is more than 500 gms. A 95% confidence interval is found to be (495, 510). Comment on the validity of the claim.

2. Answer ***any four*** questions : 5×4

(a) Let $X_1, X_2, \ldots, X_n$ denote a random sample from a population with mean μ and variance σ^2. Define an estimator.

$$T = \frac{1}{n}\sum_{i=1}^{n}\left(X_i + c\right),$$

where c is a real number.

Find the bias and mean square error (MSE) of T as an estimator of μ. For what value of c is the MSE minimized?

(b) Let X_1, X_2,..., X_n be i.i.d. random variables taking values in {0, 1, 2} with

$$P(X=0)=\frac{\theta}{3}, P(X=1)=\frac{\theta}{6}, P(X=2)=1-\frac{\theta}{2}, \text{ where } 0 \le \theta \le 2.$$

Find the maximum likelihood estimator (MLE) of θ.

(c) Let T_1 and T_2 be unbiased estimators of $g(\theta)$ with equal variance. Find the value of $a \in (0, 1)$ that minimizes $Var\{aT_1 + (1 - a)T_2\}$.

(d) A hospital is testing a new drug. The null hypothesis H_0 is that the drug has no effect and the alternative H_1 is that the drug is effective. Explain the concept of Type I error and Type II error in this context. Discuss the trade-off between Type I and Type II errors in hypothesis testing.

(e) A mobile network provider claims that at least 90% of its customers experience uninterrupted service during peak hours. A random sample of customers is surveyed and each customer reports whether they experienced uninterrupted service or not. Describe an appropriate statistical test to assess the claim of the company.

(f) Two confidence intervals are constructed for the same population mean using the same data : one at 90% confidence level and the other at 95% confidence level. Which interval is wider? Explain why. Discuss the trade-off between confidence level and precision in interval estimation.

3. Answer ***any three*** questions :

(a) (i) Let $X_1, X_2, ..., X_n$ denote a random sample from a $N(\mu, \sigma^2)$ population. Consider a parametric function $\theta = 2\mu + \sigma^2$. Explain the meaning of a parametric function in this context. Suggest suitable point estimators for μ, σ^2 and the parametric function θ. Comment on whether the estimator of θ is unbiased.

(ii) Let X_1 and X_2 be a random sample from a Poisson distribution with mean λ. Suggest two unbiased estimators of λ^2. Compute the variance of each estimator. Which estimator is more preferable and why? 7+8

(b) (i) Give an example where the maximum likelihood estimator (MLE) and the method of moments estimator (MME) of a parameter are different.

(ii) Let X have a normal distribution with mean θ and variance 1. Consider the estimator $T = e^{X+c}$. Find the value of c for which T is an unbiased estimator of e^{θ}. 8+7

(c) (i) Explain the Critical Region approach and the p-value approach used in hypothesis testing. Compare the two approaches and discuss their relative merits and demerits.

(ii) A biologist who studies spiders was interested in comparing the lengths of female and male green lynx spiders. Assume that the length X of the male spider is approximately $N\left(\mu_X, \sigma_X^2\right)$ and the length Y of the female spider is approximately $N\left(\mu_Y, \sigma_Y^2\right)$. If $\sigma_X^2 / \sigma_Y^2 = d$, a known constant, find a $100(1 - \alpha)\%$ confidence interval for $\mu_X - \mu_Y$. (6+3)+6

(d) Assume that IQ scores for a certain population are approximately $N(\mu, 100)$. To test $H_0 : \mu = 110$ against the one-sided alternative hypothesis $H_1 : \mu > 110$, we take a random sample of size $n = 16$ from this population and observe the sample mean $\bar{x}$.

(i) Suggest a suitable test statistic for testing H_0 against H_1. Identify the distribution of the test statistic used to test H_0, and explain why this distribution is appropriate.

(ii) Describe the critical region of the test at significance level α.

(iii) Explain what is meant by the power of the test in this setting and discuss how it depends on μ.

(iv) Describe what would change in the testing procedure if the population variance were unknown. 4+2+3+6

(e) (i) A medical researcher wants to examine whether there is an association between treatment (drug/placebo) and recovery (yes/no), when the sample size is small. Describe how Fisher's exact test can be used to analyze this data.

(ii) The number of calls received by two call centres in a fixed time period are modelled as independent Poisson random variables with means λ_1 and λ_2. Describe a statistical test that can be used to assess whether the two call centres receive calls at the same rate. 8+7
