2021

STATISTICS — HONOURS

Paper : CC-9

(Statistical Inference-I and Sampling Distributions)

Full Marks : 50

The figures in the margin indicate full marks.

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

Answer ***any five*** questions from question nos. **1-8**. 2×5

1. Let R_1 and R_2 be independent and identically distributed Uniform $(0, \theta)$, $\theta > 0$. Then can we use the statistics $T(R_1, R_2) = \max(R_1, R_2)/\min(R_1, R_2)$ to estimate θ?

2. Let X be a single observation from $N(\theta, 1)$, $\theta \in \Theta = \{-5, 5\}$. Suggest a test for testing $H : \theta = -5$ against $K : \theta = 5$ with both type I and type II errors ≈ 0.

3. Let X and Y be independent random variables with probability mass functions $P(X = -1) = 1/4$, $P(X = 1) = 3/4$ and $P(Y = -1) = 3/4$, $P(Y = 1) = 1/4$ respectively. What is the distribution of $X^2 - Y^2$?

4. Let $\{X_1, ..., X_{10}\}$ be a random sample from $N(\theta, 1)$, $\theta \in R$. Consider testing problem $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$. Define P-value of the test based on T = Number of X_i's such that $|X_i - \theta| < c$, c is some known positive number (only definition).

5. Suppose we are lucky enough to repeat a random experiment 100 times under identical experimental conditions. Each time we observe a random sample of size 10 and compute the value t of a statistics $T(X_1, ..., X_{10})$. The values of the statistics are $t_1, ..., t_{100}$. Based on these values is it feasible to find the approximate sampling distribution of T and its standard error?

6. Based on a random sample $\{X_1, ..., X_5\}$ from $N(0, 1)$, construct $T(X_1, ..., X_5)$ that follows Snedecor's F distribution with d.f. (2, 3).

7. Let $X_1, ..., X_n$ be a random sample from $N(0, 1)$. Define sample range $R(X_1, ..., X_n)$. Are the sampling distributions of $R(X_1, ..., X_n)$ and $R(-X_1, ..., -X_n)$ identical?

8. Distinguish between Student's t- and Fisher t-distributions with examples.

Answer ***any two*** questions from question nos. **9-11**. 5×2

9. Using a suitable example, explain the concept of P-value for a left-tailed test.

10. Let $X_1, ..., X_n$ be independent random variables having an identical continuous distribution $F(x)$. Find the distribution of $U = \dfrac{F(X_{(n)}) - F(X_{(2)})}{F(X_{(n)}) - F(X_{(1)})}$, where $X_{(i)}$'s are order statistics. Prove all intermediate steps in detail.

11. Suppose we consider samples of individuals of sizes n_1 and n_2 from two different areas A and B of a city. Further, suppose readings on a continuous marker Y are taken for these two samples. If an individual with the marker value greater than a threshold K (known) is considered to be a diseased person, then based on Y –values perform exact test for H_0 : P_A(= P(An individual is diseased in Area A)) = P_B(=P(An individual is diseased in Area B)) against H_1: $P_A > P_B$. State the underlying assumptions, if any, clearly.

Answer ***any three*** questions from question nos. **12-16**.

12. For a bivariate sample $\{(Y_i, X_i); i = 1,..., n\}$, consider the regression model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1,...,n$, x_i is the observed value of X_i and ε_i's are independent and identically distributed as $N(0, \sigma^2)$. Derive the exact sampling distribution of the least square estimate of the slope β_1. Hence, find a confidence interval of the slope with confidence coefficient $1-\alpha$, $\alpha \in (0, 1)$. Is it possible to comment on acceptance or rejection for the testing problem $H_0 : \beta_1 = \beta_{1.0}$ against $H_0 : \beta_1 \neq \beta_{1.0}$ ($\beta_{1.0}$ is a known specified value) from above confidence interval? If so, explain. 5+3+2

13. (a) Let $\{(X_1,..., X_n\}$ be a random sample from $N(\mu,\sigma^2)$, $\mu \in R, \sigma > 0$. Find the sampling distribution of

$$\frac{(X_n - \bar{X}_n)^2}{[(n-1)S_n^2 - \{n/(n-1)\}(X_n - \bar{X}_n)^2]},$$

where $\bar{X} = n^{-1}\sum_1^n X_i$ and $S_n^2 = (n-1)^{-1}\sum_1^n (X_i - \bar{X}_n)^2$.

(b) Let $\{X_1,..., X_n\}$ be a random sample from a continuous distribution with distribution function $G(.)$. Then for positive integer $m < n$, find the distribution of

$$\frac{\sum_1^m ln\,\{G(X_i)\}}{\sum_1^n ln\,\{G(X_i)\}}$$

5+5

14. (a) Based on two independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with all parameters unknown, describe a size α test procedure for testing $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$. Hence, comment on testing the null hypothesis $H_{01} : \sigma_1^2 = \sigma_2^2 = 1$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$?

(b) Find the mean and variance of Student's t–distribution. Show that its density approaches to standard normal distribution as degrees of freedom increases. 6+4

15. (a) Let $X_1,..., X_n$ be a random sample from exponential (λ) distribution with $\lambda > 0$. Show that correlation $(X_{(1)}, X_{(n)}) = n^{-1}\{\sum_1^n i^{-2}\}^{-1/2}$.

(b) Find the moment generating function of Chi-Square distribution with d.f. n. Hence or otherwise prove the additive property of Chi-square distributions. 5+5

16. Let (X, Y) follow bivariate normal distribution $BN(0, 0, 1, 1, \rho = 0)$.. If R denote the sample correlation coefficient based on random sample of size n drawn from the above distribution, show that $\frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$ follows t–distribution with $(n - 2)$ d.f. Discuss its use for testing $H : \rho = 0$ in bivariate normal population $BN(\mu_x, \mu_y, \sigma_x{}^2, \sigma_y{}^2, \rho)$. 7+3