0741

D-VSF-L-FGA

STATISTICS

Paper I

Time Allowed : Three Hours | *Maximum Marks : 200*

INSTRUCTIONS

Candidates should attempt Questions No. 1 and 5 which are compulsory, and any THREE of the remaining questions, selecting at least ONE question from each Section.

All questions carry equal marks.

Marks allotted to parts of a question are indicated against each.

Answers must be written in ENGLISH only.

Assume suitable data, if considered necessary, and indicate the same clearly.

(Notations and symbols are as usual)

SECTION A

1. Answer any ***four*** parts : *4×10=40*

(a) Box I contains 4 white and 2 red balls, and Box II contains 3 white and 5 red balls. Two balls are chosen at random from Box I without observing their colours and are put in Box II. A ball is then picked from Box II. What is the probability that it is white ?

(b) A certain manufacturing plant uses a specific bulk product. The amount of product used in a day can be modelled by an exponential distribution with parameter 4 (measured in tons). What is the probability that the plant will use more than 4 tons on a given day ?

(c) People either like dogs or dislike them. If a statistician wants to estimate the probability p that a person likes dogs, how many people must be included in the sample ? Assume that the statistician will be satisfied if the error of estimation is less than 0·04 with probability equal to 0·90. Assume also that the statistician expects p to lie in the neighbourhood of 0·06.

[You may need some values from the ones given below :

$z_{0\cdot025} = 1\cdot96$, $z_{0\cdot05} = 1\cdot65$]

(d) Let $X_1, \ldots, X_{10}$ be a sample from a Bernoulli distribution with $P(X_i = 1) = p$ for some $p \in (0, 1)$. Find the uniformly minimum variance unbiased estimator (UMVUE) of p^2.

(e) We have 5 independent normal distributions with unknown means $\mu_1, \ldots, \mu_5$ respectively and an unknown common variance σ^2. From each of these distributions a sample of size 10 is drawn; thus, $X_{i,1} \ldots X_{i,10}$ denotes the sample of size 10 from $N(\mu_i, \sigma^2)$. Obtain an unbiased estimator of σ^2. What is the distribution of this estimator ?

2. (a) Let A, B and C be three independent events with $P(A) = 0{\cdot}2$, $P(A^c \cap B^c \cap C^c) = 0{\cdot}42$, $P(A \cap B \cap C) = 0{\cdot}015$.
Let $p = P(C \cap A^c \cap B^c)$. Show that p equals either 0·14 or 0·18.
[Here A^c, B^c and C^c denote the complements of the events A, B, and C respectively.] *10*

(b) What is the characteristic function of a random variable X whose probability density fucnction is

$$f(x) = \frac{1}{2} e^{-|x|} \text{ for } -\infty < x < \infty \text{ ?}$$ *10*

(c) Let (X, Y) be a random vector with joint density given by

$$f(x, y) = \begin{cases} c & \text{if } x^2 + y^2 \le 1, \\ 0 & \text{if } x^2 + y^2 > 1. \end{cases}$$

Evaluate c. Show that X and Y are uncorrelated. Are X and Y independent ? *10*

(d) Let $X_1, X_2, \ldots$ be random variables. Give an example for each of the following :

(i) X_n converges in probability to X, but X_n does not converge almost surely to X.

(ii) X_n converges in probability to X, but $E(|X_n - X|^p)$ does not converge for any $p > 0$. *10*

3. (a) Let X be a Poisson random variable with unknown mean λ. Find a function of λ for which the amount of information in a sample of size n is independent of λ. *10*

(b) Let $Y_1, Y_2 ..., Y_n$ be a random sample with Y_i having a density function

$$f(y_i) = \begin{cases} \frac{1}{\alpha} e^{-y_i/\alpha} & \text{for } 0 < y_i < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\overline{Y}$, the mean of the sample, is a sufficient statistic for α. State clearly the theorem you use. *10*

(c) Let $\hat{\theta}$ be a statistic that is normally distributed with an expected value and a variance equal to θ and σ_{θ}^2 respectively. Explain how to obtain a $(1 - \alpha)$ 100% confidence interval for θ. *10*

(d) Let μ_1 and μ_2 be the average lifespan of a rhinoceros in captivity and in the wild respectively. The data of the lifespan of 9 rhinoceros in captivity and 9 rhinoceros in the wild were recorded as $y_{11}, y_{12}, ..., y_{19}$ and as $y_{21}, y_{22}, ..., y_{29}$ respectively. The data showed

Captive rhinoceros lifespan	Wild rhinoceros lifespan
sample size = 9	sample size = 9
sample mean $\overline{y}_1 = 35{\cdot}22$ years	sample mean $\overline{y}_2 = 31{\cdot}56$ years
$\sum_{i=1}^{9} (y_{1i} - \overline{y}_1)^2 = 195{\cdot}56$	$\sum_{i=1}^{9} (y_{2i} - \overline{y}_2)^2 = 160{\cdot}22$

At 0·05 level of significance decide whether the lifespans of the rhinoceros in captivity and in the wild are same or not. *10*

[You may need some values from the ones given below :

at degrees of freedom 16, $t_{0\cdot025} = 2\cdot120$,
$t_{0\cdot05} = 1\cdot746$

at degrees of freedom 18, $t_{0\cdot025} = 2\cdot101$,
$t_{0\cdot05} = 1\cdot734$

$z_{0\cdot025} = 1\cdot96$, $z_{0\cdot05} = 1\cdot65$]

4. (a) Suppose that $y_{ij} = \alpha_i + \beta t_{ij} + \varepsilon_{ij}$, $i = 1, \ldots n$, $j = 1, \ldots m$, where α_i and β are unknown parameters, t_{ij} are known constants, and ε_{ij} are i.i.d. random variables with 0 mean. Find explicit forms for the least square estimators (LSE) of β, α_i, $i = 1, \ldots n$. *10*

(b) What is a randomized block design ? Explain how to obtain the confidence interval for the difference between a pair of means in a randomized block design. *10*

(c) Use the method of least squares to fit a straight line to the data given below.

x	y
−2	0
−1	0
0	1
1	1
2	3

Sketch the line.

What are the variances of the estimated slope and intercept ? Are they correlated ? *10*

(d) Let p be the probability that a randomly chosen Indian is a vegetarian. We want to test $H_0 : p = 0{\cdot}2$ against $H_1 : p = 0{\cdot}6$ on the basis of a sample of size 10. Let Y be the number of people in our sample who are vegetarians. If we have a rejection region $\{y \geq 4\}$, then

(i) what is the Type I error ?

(ii) what is the Type II error ? *10*

SECTION B

5. Answer any ***four*** parts : *4×10=40*

(a) Let X be a random variable with mean 11 and variance 9. Using Chebychev's theorem find the following :

(i) a lower bound for $P(6 < X < 16)$,

(ii) the value of c such that

$$P(|X - 11| \geq c) \leq 0{\cdot}09.$$

(b) Suppose X_1 and X_2 are two normal random variables both with mean 0, but with variances 1 and 16 respectively. On the same graph sketch both the probability density functions and on another graph sketch both the probability distribution functions, clearly labelling the curves. If a and b are such that $P(X_1 \leq a) = P(X_2 \leq b) = 0{\cdot}49$, then which of the three is correct : $a < b$, $a > b$ or $a = b$?

(c) Two independent random samples of sizes m and n are taken from two independent normally distributed populations with unknown variances σ_1^2 and σ_2^2. Explain how you would construct a 90% confidence interval for the ratio σ_1^2 / σ_2^2.

(d) The number of students in four courses in two different universities A and B are given in the table below.

Course	A	B
I	27	32
II	31	29
III	26	35
IV	25	28

With this data explain how to test the null hypothesis (at a level of significance α) that the population relative frequency distributions in the courses at universities A and B are identical against the alternate hypothesis that the population relative frequency distributions are shifted in respect to their relative locations.

(e) Consider the randomized block design with v treatments, each replicated r times. Let t_i be the treatment effect of the i-th treatment. Find

$\text{Cov}\left(\sum l_i \hat{t}_i, \sum m_i \hat{t}_i\right)$ where $\sum l_i \hat{t}_i$ and

$\sum m_i \hat{t}_i$ are the best linear unbiased estimators of $\sum l_i t_i$ and $\sum m_i t_i$ respectively and $\sum l_i = \sum m_i = \sum l_i m_i = 0$.

6. (a) Let X be a random variable with $P(X = i) = \frac{1}{10}$ for i = 0, 1, ..., 9 and let Y be a random variable (independent of X) with a uniform distribution on [0, 1).

(i) What is the distribution function of X + Y ?

(ii) What is the moment generating function of X + Y ?

(iii) What is the correlation function between X and X + Y ? *10*

(b) Let $X_1, X_2, \ldots$ be a Markov chain with state space {1, 2, 3} and transition probability matrix

$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0\cdot4 & 0\cdot3 & 0\cdot3 \\ 0\cdot2 & 0\cdot4 & 0\cdot4 \\ 0\cdot6 & 0\cdot1 & 0\cdot3 \end{pmatrix} \end{array}$$

What is the invariant distribution of this Markov chain ? *10*

(c) (i) Obtain the distribution of the sum of 30 independent Poisson random variables, each with mean 1/3.

(ii) Obtain the distribution of the sum of 30 independent Bernoulli random variables, each with mean 1/3. *10*

(d) State the Borel-Cantelli lemma and its converse. *10*

7. (a) From a finite population we obtain a sample of size n.

   (i) Obtain the variance of the sample mean, when the sampling is done *with* replacement.

   (ii) Obtain the variance of the sample mean, when the sampling is done *without* replacement.

   (iii) If the population has mean μ and variance σ^2, then what can you say about the asymptotic distribution of the sample mean ? *10*

(b) A bag has 5 balls some red in colour and the remaining blue. You pick out 3 balls without replacement and observe that 2 of them are red and 1 of them is blue. What is the maximum likelihood estimate of the number of red balls in the bag ? *10*

(c) Explain the notions of the power of a test and the P-value of a test. *10*

(d) What is a $h \times k$ contingency table and how do you test the hypothesis that the two classifications producing the $h \times k$ contingency table are independent of each other ? *10*

8. (a) Given the points $(x_1, y_1), ..., (x_n, y_n)$ of a scatter diagram, let $y = a + bx$ be the fitted regression line. Describe a measure of scatter about the regression line and its relation to the sample variance of y and the sample correlation coefficient of x and y. *10*

(b) Let X_1, X_2, ... be independent random variables with a common probability density function f. Compare the Neyman-Pearson and the sequential probability ratio test methods of testing $H_0 : f = f_0$ versus $H_1 : f = f_1$. What are the errors of acceptance and rejection in both these methods ? *10*

(c) The blood group classifications of human beings are O, A, B and AB with relative frequencies proportional to r^2, $(p^2 + 2pr)$, $(q^2 + 2qr)$ and 2pq respectively, where $p + q + r = 1$. In a large sample of N persons tested, the observed frequencies in the four types were found to be n_1, n_2, n_3 and n_4 respectively. Find the estimates of the proportions p, q and r and also the large sample variances of the estimates. *10*

(d) Suppose $X_1, X_2, ..., X_{20}$ are independent random variables, each having a normal distribution with known mean μ and unknown variance σ^2. Obtain a lower bound for the variance of any unbiased estimator of σ^2. State any theorem you use. *10*