IFoS 2010

000862

B-JGT-K-TUA

STATISTICS

Paper I

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

INSTRUCTIONS

Candidates should attempt questions 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 of the following : *4×10=40*

   (a) Let X and Y be two independent variables with $N(\mu = 2, \sigma^2 = 9)$ and $N(\mu = 3, \sigma^2 = 16)$ distributions respectively. If $144\,Z = 16X^2 + 9Y^2 - 64X - 54Y + 145$, find E(Z) and V(Z). State (without proof) any theorems you have used.

(b) A random sample of n units is divided into k categories such that the i^{th} category has n_i units. Assuming multinomial probabilities, derive $E(n_i)$, $V(n_i)$ and corr.(n_i, n_j).

(c) Find the constant K such that $f(x, \theta) = K e^{-|x|/\theta}$, $-\infty < x < \infty$, $\theta > 0$ is a density function. Obtain the maximum likelihood estimator of θ. Also suggest an estimator for θ based on the method of moments.

(d) With an illustrative example discuss how non-parametric tests are useful in drawing inferences about the parameters, giving all the necessary mathematical details.

(e) Let X have a N(0, 1) distribution under H_0 and a Cauchy distribution $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$, $-\infty < x < \infty$ under the alternative hypothesis H_1. Find a most powerful size α test of H_0 against H_1 and derive its power.

2\. (a) Let t_i, i = 1, 2, ..., k be independent unbiased estimators of θ with $V(t_i) = V_i$. Obtain the best linear combination of t_i which is unbiased for θ and has minimum variance among such combinations. Further show that

$$\sum_{i=1}^{k} (t_i - \bar{t})^2 / k(k-1) \text{ is an unbiased estimator of}$$

$$V(\bar{t}) \text{ where } \bar{t} = \sum_{1}^{k} t_i \Big| k .$$ *10*

(b) State Chebyshev's Weak Law of Large Numbers (WLLN). For the sequence of r.v.'s $\{X_n\}$ such that $\text{Prob.}(X_k = \pm k^{\alpha}) = 1/2$, $\alpha > 0$, verify whether WLLN holds. *10*

(c) Let $x_1, x_2, ..., x_n$ be n independent observations on a r.v. X with density function $f(x, \theta)$. Further, let $g(x_1, x_2, ..., x_n, \theta) = 0$ be an estimating equation for θ. Show that, if $E(f) = 0$, then $[E(g')]^2/V(g) \leq \mathcal{I}$ where $g' = dg(x_1, x_2, ..., x_n, \theta)/d\theta$ and $\mathcal{I}$ is the total information in the sample. What happens when $g = d \text{ Log } L/d\theta$, where L is the likelihood = $f(x_1, \theta)\, f(x_2, \theta) \, ... \, f(x_n, \theta)$? *10*

(d) Explain the concept of 'sufficient statistics' and briefly indicate their role in obtaining minimum variance unbiased estimators, stating clearly the theorems you refer to. *10*

3. (a) Let $\overline{X}$ be the mean of a random sample of size n from $N(\mu, \sigma^2)$, σ^2 known. Suppose that the parameter μ has a prior distribution which is distributed as $N(\mu_0, \sigma_0^2)$. Show that the posterior distribution of $\mu | \overline{X} = \overline{x}$ is again normal. Obtain the posterior mean and posterior variance. *12*

(b) Describe the chi-square test for testing the homogeneity of two multinomial populations and discuss its application with reference to an example. *12*

(c) For testing independence in a 2×2 contingency table, derive the test statistic. Explain how you would apply the continuity correction. Further if cell frequencies are very small, indicate briefly how you would treat the problem. 12

(d) Give an example of any one variance stabilizing transformation, explaining how it is used. 4

4. (a) Let $X_1, X_2, \ldots, \ldots$ be a sequence of i.i.d. $N(\mu, \sigma^2)$ random variables with known σ^2. Construct a Sequential Probability Ratio Test for testing the hypothesis $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu = \mu_1$. 10

(b) Let Prob.$(X = x) = a_x \theta^x / f(\theta)$, $x = 0, 1, 2, \ldots$ where a_x may be zero for some x. Let $x_1, x_2, \ldots, x_n$ be n independent observations from this distribution. Show that the m.l.e. of θ is a root of the equation $\bar{x} = \theta f'(\theta) / f(\theta)$ which is also the same obtained from the method of moments. Calculate $i(\theta)$, the information content in a single observation. 10

(c) Independent random samples of size n from two normal populations with known variances σ_1^2 and σ_2^2 are used to test the null hypothesis $H_0 : \mu_1 - \mu_2 = \eta$ against the alternative $H_1 : \mu_1 - \mu_2 = \eta'$. For specified values of α and β, the probabilities of type I and type II errors respectively, show that the required sample size is $n = \left(\sigma_1^2 + \sigma_2^2\right)\left(Z_{\alpha/2} + Z_{\beta/2}\right)^2 / (\eta - \eta')^2$, where Z is the standard normal point. 10

(d) If log X is distributed as $N(\mu, \sigma^2)$ then X is said to have a log normal distribution. Write down the density function of X and show that the coefficient of variation is defined by

$$\sqrt{V(X)} \Big/ E(X) = \left(e^{\sigma^2} - 1\right)^{1/2}.$$

10

SECTION B

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

(a) Describe Yates' method of computing factorial effect totals in a 2^3 experimental design.

(b) What are the sources of non-sampling errors during the field operations stage in a large scale sample survey and how do you control them ?

(c) Let y be the response variable and $X_1, X_2, ..., X_p$ be p predictor variables. Let $g(X_1, X_2, ..., X_p)$ be a predictor of y. Show that $E(y - g(X_1, X_2, ..., X_p))^2$ is minimum when $g(X_1, X_2, ..., X_p)$ is the conditional expectation

$M(X_1, X_2, ..., X_p) = E(y \mid X_1, X_2, ..., X_p)$. Further prove that $\rho(y, M) \geq \mid \rho(y, g) \mid$ for any function g. When is the lower bound attained ?

(d) (i) A field researcher has to select one out of the five plantations in a forest at random for further study. Having no access to random numbers, how does she make the required selection using a fair coin ? *5*

(ii) A statistics student was asked to select a random sample of size two from a normal population with mean $\mu = 50$ and variance $\sigma^2 = 9$. How should he proceed assuming that he was given random number tables ? *5*

(e) Write down the density function of a p-variate r.v. distributed as $N_p(\mu, \Sigma)$ with mean vector μ and variance covariance matrix Σ. Specify the distributions along with parameters of (i) $Y = l'X$, where l is a $p \times 1$ vector $(l_1, l_2, ..., l_p)$ of constants and (ii) $Z = CX$, where C is a $q \times p$ matrix representing q linear functions. (No proofs required).

6. (a) Describe the ratio method of estimation for estimating the population total Y of a study variable y when auxiliary information on a related variable x is available. Denote this estimator by $\hat{Y}_R$ and derive its Bias and Mean Squared Error. Find a condition under which $\hat{Y}_R$ would fare better than $\hat{Y}$, the unbiased estimator obtained without using information on x. *15*

(b) Obtain the variance expression for the Horvitz Thompson estimator $\hat{Y}_{HT}$ of the population total Y of a study variable y for fixed sample size given by Horvitz and Thompson. Also cast this in the form given by Yates and Grundy, namely

$$V = \sum_{i \neq j}^{N}\sum^{N} \alpha_{ij}\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2$$ and hence write

down an unbiased estimator $\hat{V}$ of V [α_{ij} has to be clearly specified by you]. Comment on the non negativity of $\hat{V}$. *15*

(c) Discuss how uniformity trials in experimental designs and pilot surveys in large scale sample surveys are conducted mentioning their usefulness. 10

7. (a) Describe the technique of Analysis of Dispersion as a generalization of ANOVA in the univariate case wherein the decomposition is given by 'Residual Sum of Squares' and 'Deviation from the hypothesis'. Give a practical illustration where this technique could be applied. 12

(b) Define Mahalanobis-D^2 statistic. How is it related to Hotelling's-T^2 statistic ? Explain clearly how D^2 is treated as a measure of distance mentioning an application. 12

(c) If the matrix $S \sim$ Wishart $W_p(k, \Sigma)$ and $|\Sigma| \neq 0$, show that $|S| / |\Sigma|$ is distributed as the product of p independent central χ^2 variables. What are their degrees of freedom ? (State clearly any results that you have assumed). 9

(d) What do you understand by the term 'multiple correlation coefficient' of Y on the variables $X_1, X_2, ..., X_p$ denoted by $\rho_{0 \cdot 12 ... p}$? What does this measure ? What range of values can it take ? 7

8. (a) Explain what is meant by a split-plot design. Suppose that factor A has p levels which are arranged in a Randomized Block Design having r blocks. Let factor B have q levels which are applied to the plots of a block after subdividing each plot into q sub plots. Write down the model clearly explaining the notations and present a blank ANOVA table describing sources of variation and the corresponding degrees of freedom. *10*

(b) How is the F-test in ANOVA based on normality assumption ? When this assumption does not hold good even after transformation of the variable, explain how a non parametric test based on ranks could be used for testing that the distribution functions of the continuous populations under study are the same. *10*

(c) Describe the analysis for a Latin Square Design giving the model, Blank ANOVA table and test statistics. Also mention the advantages of LSD. What are 'Orthogonal Latin Squares' ? *10*

(d) What do you understand by 'randomized response technique' ? Illustrate this technique with reference to Warner's unrelated question model. How efficient is this method (w.r.t. variances) compared to Direct Response Conventional method ? *10*