STATISTICS

PAPER – I

SECTION I

(Probability)

1. (a) A coin is tossed until there are either two consecutive heads or two consecutive tails or the number of tosses becomes five. Describe the sample space along with the probability associated with each sample point, if every sequence of it tosses has probability 2^{-n}.

   (b) A die is of the shape of a regular tetrahedron whose faces bear the numbers 111, 112, 121, 122. A_1, A_2, A_3 are respectively the events that the first two, the last two and the extreme two digits are the same, when the die is tossed at random. Find whether or not the events A_1, A_2, A_3, are (i) pairwise independent, (ii) mutually (i.e. completely) independent. Determine $P(A_1 \mid A_2 A_3)$ and explain its value by argument.

2. Obtaining the mgf of the Poisson variate X with mean λ, find the limit as $\lambda \to \infty$, of the mgf of $(X-\lambda)/\sqrt{\lambda}$ and interpret the result in the context of the central limit theorem. Also prove that

$$\lim_{\lambda \to \infty} \sum_{j=\lambda+\alpha\sqrt{\lambda}}^{\lambda+\beta\sqrt{\lambda}} \frac{e^{-\lambda} \lambda^{j}}{j!} = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{1}{2}u^2} du$$

   Show that 2X is not a Poisson variate. Give a set of conditions under which X+Y too is a Poisson variate.

3. (a) State and prove Kolmogorov's inequality in generalization of Chebyshev's inequality and give one application of the same.

   (b) X and Y are standard normal variates with coefficient of correlation ρ. Write down the density of their joint distribution and obtain its characteristic function. Hence or otherwise show that the regression of Y on X is linear.

4. Prove that χ^2 has the reproductive property. Obtain the distribution of the ratio of two independent chi-square variates. Deduce from it the t-distribution and the Cauchy distribution.

SECTION B

(Statistical Inference)

5. (a) The random variable X has a uniform distribution in the range (α, β). Find the maximum likelihood estimates of α and β on the basis of a random sample

   $(x_1, x_2, \ldots, x_n)$.

   (b) Prove that the sample mean is a sufficient and unbiased estimator of the mean of a normal population whether the variance be known or not. What can you say about the sample variance as a sufficient estimator of the population variance ?

   (c) If a sufficient estimator exists, prove that it is a function of the maximum likelihood estimator.

6. (a) Show that with the exponential distribution

   $dF(x) = \theta e^{-\theta x} dx,\ x \geq 0$

Central confidence limits for large samples of size n and 95% confidence coefficient are

$$\left\{1 \pm 1.96/\sqrt{n}\right\}\bar{x}$$

where $\bar{x}$ is the mean of the sample observations $x_1, x_2, \ldots, x_n$ drawn randomly from the exponential population.

(b) Describe Wald's SPRT. How to obtain the lines of rejection and acceptance of the null hypothesis ?

7. (a) Explain the rationale of likelihood ratio test procedures. Employing the LR criterion, obtain the two-tail t-test for testing, on the basis of a random sample $x_1, \ldots, x_n$ from a normal population whose parameters are unknown, the hypothesis that the mean of the population is μ_0:

(b) A discrete variate X has the following distributions under the null hypothesis H_0 and the alternative hypothesis H_1:

i	1	2	3	4	5
$P(X=i/H_0)$	·1	·2	·3	·3	·1
$P(X=i/H_1)$	·2	·1	·4	·1	·2

For $\alpha = 0.5$ find the most powerful test for H_0 against H_1. Discuss the result in the light of the Neyman-Pearson lemma.

8. (a) Justify the use of chi-square as test of goodness of fit. A set of 6 similar coins was tossed together 640 times and the frequency f_x of the tosses giving x heads was found to be as follows:

x :	0	1	2	3	4	5	6
f :	13	62	145	194	153	64	9

Fitting a suitable distribution to the data, carry out the test of goodness of fit as far as you can without the use of am statistical table and state how you would proceed further.

(b) Describe the run test, stating the null hypothesis which it tests. Develop the test when Type I error is α.

SECTION B

(Linear Inference and Multivariate Analysis)

9. (a) Describe the model $y_i = \alpha + \beta x_i + \varepsilon_i$ and show how it includes the regression model and the functional relationship model with a measurement error. Find the maximum like lihood estimators $\hat{\alpha}, \hat{\beta}, \hat{\varepsilon}^2$ of α, β, var ε_i. Deduce whatever properties of the estimators, you can from the result:

$$E\left[\exp\left\{u\frac{\hat{\alpha}-\alpha}{\sigma}+v\frac{\hat{\beta}-\beta}{\sigma}+w\frac{\hat{n}\sigma^2}{\sigma^2}\right\}\right]$$

$$= (1-2w)^{1-n/2}\exp\left[\left(u^2\sum x^2/n-2\bar{x}uv+v^2\right)\left\{\frac{1}{2}\Sigma(x_i-\bar{x})^2\right\}\right]$$

10. (a) For a random sample of size n drawn from a bivariate normal population, the sample correlation coefficient is r. Show that the statistic $r\sqrt{\{(n-2)/(1-r^2)\}}$ follows the t-

distribution with n–2 d.f. when ρ=0. How will you use this statistic to test for the population hypothesis ρ=0?

(b) Write a short note on orthogonal polynomials.

11. (a) If $\underset{\sim}{X}$ is distributed as N (μ, Σ), prove that the marginal distribution of any set of components of $\underset{\sim}{X}$ is multivariate normal.

(b) $\underset{\sim}{X} \sim N\left(\underset{\sim}{O}\,\Sigma\right)$ is composed of the sub-vectors $\underset{\sim}{X}^{(1)}$ and $\underset{\sim}{X}^{(2)}$. Find the regression function of $\underset{\sim}{X}^{(1)}$ on $\underset{\sim}{X}^{(2)}$. State (without proof) the optimum properties of this function.

12. (a) Describe how the T^2-statistic can be used to test the hypothesis that in a normal population—

(i) the mean vector is a given vector

(ii) the mean vector is equal to that of another normal population whose convariance matrix is the same as that of the former but unknown.

(b) Formulating the problem of discriminar analysis state the procedure for using the statistics

(i) $\underset{\sim}{x}'\Sigma^{-1}\left(\mu^{(1)} - \mu^{(2)}\right)$

(ii) $\underset{\sim}{x}'S^{-1}\left(\overline{x}^{(1)} - \overline{x}^{(2)}\right)$

STATISTICS

PAPER - II

SECTION I

(Sampling Theory and Design of experiments)

1. (a) What is stratified sampling ? Why is stratified sampling generally adopted in sample surveys?

   (b) Ignoring the finite population correction factor, prove in the usual notation, that

   $V_{opt} < V_{prop} < V_{ran}$

   When can one or the other of these inequalities get reversed?

2. You are asked to plan a sample survey to find out the incidence of educated unemployment in your State.

   Give details of your plan including an outline of a suitable questionnaire for this purpose.

3. What is meant by the design of an experiment? Describe the situations in which the three basic designs, namely, Completely Randomized Design, Randomized Blocks Design and Latin Square Design, are appropriate, and state the rule or rules by which treatments are allocated to units in these designs. Show that under linear additive model each of these designs is orthogonal.

4. What do you mean by a split-plot design? Discuss its advantages and disadvantages.

   A split-plot design with p whole plot and q subplot treatments is arranged in a randomized block with r replications. Give the complete analysis of this design.

SECTION II

(Engineering Statistics)

5. Explain clearly the use of control charts in industry, stating conditions under which different types of control charts should be used.

   Deduce the OC of $\overline{X}$ and R charts.

6. Describe single, double and multiple sampling inspection plans. Give a general outline of methods for determining the constants involved in single and double sampling plans.

7. (a) Discuss the following concepts in connection with sampling inspection plans:

   Consumer's risk, Produce's risk, AOQL and LTPD.

   (b) Give a brief critical account of a sequential sampling plan for attributes.

8. (a) Define reliability and availability. Give brief account of reliability of series and parallel systems.

   (b) What is redundancy? Give different types of redundancy and explain the use of redundancy in reliability Improvement.

SECTION III

(Operational Research)

9. Elucidate the statement, 'Model-building is the essence of Operational Research approach' giving concrete examples of application choosing one each from the fields of business, industry and national planning, respectively.

10. (a) State three different types of inventory models, and explain them briefly.

(b) In a certain manufacturing situation the production is instantaneous and the demand per year is R. Show that the optimal order quantity q per run which minimizes the total cost is given by

$$q = \sqrt{\frac{2RC_3(C_1 + C_2)}{C_1C_2}}$$

where C_1 = storage cost per unit per year

C_2 = shortage cost per unit per year

C_3 = setup cost per run.

11. (a) Investigate the M/M/I Queueing system.

(b) Give some more important applications of Queueing theory.

12. Write short notes on any two of the following:

(a) Replacement Models.

(b) Duality in linear programming.

(c) Input and output formats.

(d) Assignment problems.

(e) Subroutines.

SECTION IV

(Quantitative Economics)

13. (a) What purpose is served by an index number? Show that the factor reversal test and the time reversal test are not satisfied by Laspeyres' and Paasche's index numbers. Further, show that both these tests are satisfied by Fisher's deal index number.

(b) Discuss how you will proceed for constructing a cost of living index number for a given expenditure group in a city.

14. (a) Enumerate the different components of a time cries. What purpose is served by analyzing a time series?

(b) Discuss different method of determining trend a time series. What are their relative merits and demerits?

15. (a) Explain the meaning of elasticity of demand with respect to price. Given the demand for a commodity and the corresponding price, how will you calculate the elasticity of demand with respect to price ?

(b) Let d_1 and d_2 represent the demands of a commodity for two strata of a population. If η_1 and η_2 be the elasticities of demand with respect to national income for the two strata, show that the corresponding elasticity η for the two strata combined would be given by

$$\eta = \frac{\eta_1 d_1 + \eta_2 d_2}{d_1 + d_2}$$

16. (a) Describe the indirect least squares and two-stage least squares procedures-of estimation.

(b) Discuss the importance of short-term economic forecasting.

SECTION V

(Demography and Psychometry)

17. (a) Define reproduction rates. Explain how far they may be looked upon as indices of population growth. What is meant by saying that the Net Reproduction Rate (NRR) for a country is 1.127 ? Show that for any community the net-reproduction rate is -necessarily less than the gross reproduction rate.

(b) Explain why the mortality situations of two places cannot usually be compared on the basis of crude death rates. Describe the construction of standardised death rates for this purpose.

18. Describe the structure of & complete life table. Explain how the different columns of a life table may be computed on the basis of observed age-specific mortality rates. How does an abridged life table differ from a complete life table? Describe a method of constructing an abridged table.

19. Starting from a suitable assumption regarding the relative growth rate of population, derive the logistic curve. Give a method of fitting this curve. Also, discuss the importance of this curve.

20. (a) What is the problem of measurement in Education and Psychology? Explain clearly the terms scaling, reliability and validity as used in problems of measurement in Education and Psychology.

(b) What are intelligence tests and how are they used in measuring intelligence? Explain clearly the terms "mental age" and IQ in this connection.