STATISTICS

PAPER – I

Time Allowed: 3 Hours *Maximum Marks: 300*

SECTION A

(Probability)

1. (a) An urn contains 5 white and 3 black balls, a second urn contains 6 white and 5 black balls. A ball is randomly transferred from the first urn to the second and then from the second to first urn. If a ball is now selected randomly from the first urn, find the probability that it is white.

(b) Describe a priori probabilities, likelihoods and posterior probabilities in defining the Bayes theorem. Explain why, in spite of its easy deductibility from the postulates of probability, it has been the subject of such extensive controversy.

(c) A sportsman's chance of shooting an animal at a distance $s(>r)$ is $\frac{r^2}{s^2}$. He fires when s = 2r and if he misses he reloads and flies when s = 3r. 4r,..... . If he misses at distance nr, the animal escapes. Calculate the odds against the sportsman.

(d) (i) For n events A_1, A_2,……, A_n show that

$$P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$$

(ii) If p be the probability that none of the n events occurs, then

$$p \le e^{-\sum_{i=1}^{n} pi}$$

15+10+15+20

2. (a) (i) What is the relationship between standard deviation arid mean deviation of the data 11, 15, 19 and 23?

(ii) if X and Y are symmetric show that

$$E\left(\frac{X}{X+Y}\right) = \frac{1}{2}.$$

(b) Let X and Y be jointly distributed with p.d.f.

$$f(x,y) = \begin{cases} x+y, & 0 < x < 1, 0 < y < 1 \\ 0, & otherwise \end{cases}$$

Find marginals conditional distributions and correlation coefficient.

(c) (i) Discuss discrete and continuous random variable with examples. Indicate the distributions for the flowing moment generating functions

$$M_1(t) = \frac{(1+e^t)^5}{32}, M_2(t) = e^{\frac{1}{4}(e^t-1)},$$

$$M_3(t) = e^{2t+\frac{1}{2}t^2}$$

(ii) If the probability of hitting a target is $\frac{1}{3}$ and if 6 shots are fired, what is the conditional probability of the target being hit at least twice assuming that at least one hit is already scored?

10+20+(20+ 10)

3. (a) State and prove Kolmogorov's inequalities and explain its uses.

(b) Verify the folloviing:

(i) $X_n \xrightarrow{a.s} C \Rightarrow X_n \xrightarrow{p} C$

(ii) $X_n \xrightarrow{q.m} C \Rightarrow X_n \xrightarrow{a.s} C$

(c) State and prove the Borel-Cantelli lemma and express its significance.

20+20+20

4. (a) Do the following functions define distribution functions of random variables?

(i) $$F(x) = \begin{cases} 0, & x < -a \\ \frac{1}{2}\left(\frac{x}{a} + 2\right) & -a \le x \le a \\ 1, & x > a \end{cases}$$

(ii) $$F(x) = \begin{cases} 0, & if \;\; x \le 1 \\ \frac{k}{4}(x-1)^3 & if \;\; 1 < x \le 3 \\ 1, & if \;\; x > 3 \end{cases}$$

Justify your answer in each case.

(b) Define negative binomial distribution and state its deductions. What will be the limiting distribution?

(c) Explain central limit theorem and weak law of large numbers for a sequence $\{X_n\}$ of i.i.d random variables. State Lindeberg-Levy theorem.

(10+10)+20+20

SECTION B

5. (a) (i) Define unbiasedness and consistency of the estimator. Obtain unbiased estimate of a from N (μ, σ^2) if μ is known.

(ii) Does a sufficient statistics for θ exists in

$$f(x,\theta) = \frac{1}{2}e^{|x-\theta|}, -\infty < x < \infty.$$

(b) Let M_0 is an MVUE of $\gamma(\Theta)$ and M_1 is any other unbiased estimator of $\gamma(\Theta)$ with efficiency e < 1. Can any unbiased linear combination of M_0 and M_1 be an MVUE of $\gamma(\Theta)$? Justify your answer,

(c) (i) State the conditions under which maximum likelihood estimators are identical with those given by method of moments.

(ii) $$f(x,\theta) = e^{-(x-\theta)}, \theta \le x < \infty,$$
$$-\infty < \theta < \infty, then$$

$$P\left[X_{(1)} - \frac{1}{n}\log\alpha \leq \theta <\leq X_{(1)}\right] = 1-\alpha$$

(15+5)+20+(10+10)

6. (a) Let (x_1, x_2) be the two independent observations from

$$f(x,\theta) = \frac{1}{\theta}e^{-x/\theta}, 0 < x < \infty, > 0.$$

For testing H_0: Θ = 2, against the alternative H_1: Θ = 4, the H_0 is accepted when $x_1 + x_2 < 9.448$. Determine UMP test.

$$\left\{P\left[\chi_4^2 \geq 9.448\right] = 0.05, P\left[\chi_4^2 \geq 0.711\right] = 0.95\right\}$$

(b) Define MP and UMP region. Compare them and show that a MP test is necessary unbiased.

(c) Describe real valued function T(x) and the families with MLR. Obtain UMP test for $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$

25+15+20

7. (a) State the properties of L-R test and suggest the conditions where - 2 $\log_e$ X has the asymptotic chi-square distribution.

(b) Let $(x_1, x_2, \ldots\ldots, x_n)$ be a random sample from $N(2, \sigma^2)$. Determine the L-R test of $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$

Examine, whether the test is unbiased.

(c) Consider a sequence of i.i.d variable with pdf $f(x,\theta) = \theta e^{-\theta x}, x \geq 0, \theta > 0.$ Derive SPRT of strength (α, β) for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. Also obtain its OC function.

25+15+20

8. (a) Explain the main difference between the parametric and non-parametric approaches to the theory of statistical inference. Specify the underlying assumptions and the null hypothesis in median, sign and U tests.

(b) Describe oh-square and kolmogorov test for goodness of fit and compare.

(c) Discuss complete sufficient statistics and suggest the method to find MVUE for $e^{-\lambda}$ P(λ).

25+15+20

SECTION C

(Linear Inference and Multivariate Analysis)

9. (a) Discuss least square estimate and show that least-square estimates are the same as the best linear esfimates in linear estimation.

(b) Let $y_i = m + bx_i + e_i$ (i = 1, 2, ……, n), flod estimate of m and b and express Var $(\hat{b})$.

(c) In the case of two-y classified data with unequal number of observations per cell derive the expectation of the MS components of the ANOVA table.

20 + 20 + 20

10. (a) Define contrasts, orthogonal contrasts and show that SS due ton observations is equal to the total SS due to the orthogonal contrasts.

(b) For the model $y = X\beta + e$, we have $\hat{\beta} = (X'X)^{-1} X'y$, $E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \sigma^2 (X'X)^{-1}$, then test statistic $F = \frac{Q_2(\beta_0)}{p} \bigg/ \frac{Q_1}{n-p}$ for $H_0 : \beta = \beta_0$, where $Q_1 = (y - X\hat{\beta})'(y - X\hat{\beta})$, $Q_2 = (X'e)'(X'X)^{-1}(X'e)$, $e = (y - X\beta_0)$, find the distribution of F and state its properties.

(c) In the multiple linear regression model with normality assumption obtain 100(1 - α)% confidence interval to find out whether the regression on the regressor variables is significant or not

20+25+15

11. (a) (i) If $X \sim N(\zeta\sigma^2)$ then $a'X \sim N(a'\zeta, a'Ba)$ find $B = \tilde{T}\tilde{T}'$ when $\tilde{T}y = (X - \zeta)$.

(ii) If the probability density function of a random vector X is $k \exp\left\{-\frac{1}{2}(X - \zeta)'\Sigma^{-1}(X - \zeta)\right\}$ determine E(X) and V(X).

(b) Consider the sets of random variable

$X_1, X_2, \ldots, X_q$ *and* $X_{q+1}, X_{q+2}, \ldots, X_o$ with vectors

$$X^{(1)} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_q \end{pmatrix}, X^{(2)} = \begin{pmatrix} X_{q+1} \\ X_{q+2} \\ \vdots \\ X_p \end{pmatrix} \text{ and } \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \Sigma \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$$

The marginal distribution of the set is multivariate normal with means, variances and covariances obtained by taking the proper components of μ and Σ respectively.

(c) Define partial correlation coefficients in multivariate analysis Show that

$$\hat{\Sigma}_{112} = \hat{\Sigma}_{11} - \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$$

(5+10)25+20

12. (a) T^2 statistics is the generalization of t-statistic. Explain us relationship with D^2-statistic. Explain the uses of T^2 statistic in test of hypothesis.

(b) Let $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$ and linear discriminant function $L(U) = (\mu_2 - \mu_1)'\Sigma^{-1}U$ is sufficient for the class of normal densities $N_p(\alpha\mu_1 + \beta\mu_2, \Sigma), \alpha + \beta = 1$, prove it. Discuss the methods used in it.

30 + 30

STATISTICS

PAPER - II

Time Allowed: 3 Hours *Maximum Marks: 300*

SECTION A

(Sampling Theory and Design of Experiments)

1. (a) Suppose one university is interested in determining the attitude of students towards a proposed change in the method of examination in its affiliated colleges and teaching departments. How would you proceed to plan a sample survey for this purpose?

   (b) Suppose N_1 of the N units in a population are known to have the value zero and a simple random sample of n units is selected from the $N — N_1$ non-zero units with replacement. Suggest an unbiased estimator of the overall population mean and obtain its efficiency compared to simple random sampling of n units directly from all the N units with replacement.

   (c) Suppose a stratified sampling design is adopted for estimating the overall proportion P of units having a specified characteristic by dividing the population of units into K strata and selecting n units from N_i units in the i^{th} stratum, (i = 1, 2,... k), with equal probability and without replacement. Suggest an unbiased estimator of P and obtain its sampling variance.

   20. 20, 20

2. (a) Explain the ratio method of estimation. Derive the conditions, if any, under which a ratio estimator of the population means $\bar{Y}$ using the information on an auxiliary variable X, is more efficient than the sample mean $\bar{y}$ in the case of simple random sampling without replacement, staling clealy the assumptions involved.

   (b) Indicating what is a' regression estimator, state under what circumstances you would use this estimator in preference to a ratio estimator.

   (c) Explain clearly the concept of PPS sampling. What are its advantages and disadvantages?

   20, 20, 20

3. (a) Explain uniformity trials. Discuss their use in picking up a proper design.

   (b) Explain Fisher's missing plot technique and indicate how you will carry out the exact analysis of variance for a randomized block design with one missing observation.

   (c) Write a note on fractional replication.

   15, 30, 15

4. (a) Comment on the necessity, advantages and disadvantages of factorial experiments and explain the "odd and even rule" for writing down the expressions for main effects and tolerations in 2n factorial experiments.

   (b) Give an outline of the analysis of balanced incomplete block design and obtain its efficiency compared to a randomized block design.

   (c) Write an explanatory note on split-clot experiments.

   20, 25, 15

SECTION B

(Engineering Statistics)

5. (a) Describe the importance of normal distribution in the theory of quality control.

   (b) Distinguish between 'Process Control' and 'Product Control' Examine the basis of the p-chart and explain the approximations and assumptions involved and how they are met in actual plant conditions.

   (c) Distinguish between chance and assignable causes of variation.

15, 30, 15

6. (a) Explain single and double sampling inspection plans for attributes. What are the statistical advantages of double sampling plan?

   (b) Given lot size (N), LTPD (p_t), consumer's risk (P_c) and process average (T), derive the most economical single sampling plan for acceptance purposes.

   (c) Define AOQL and describe how to obtain the same for the double sampling plan.

15, 30, 15

7. (a) Write a note on sequential sampling inspection plans for attributes.

   (b) Discuss the role of censored samples in life testing experiments.

   (c) Define the terms hazard rate and reliability function with their important properties. Obtain them in the case then the life-time distribution is assumed one-parameter exponential.

20, 20, 20

8. (a) Discuss the estimation of reliability characteristics in Weibull distribution with censored sample.

   (b) Obtain the MLE and UMVUE of the parameter of a one-parameter exponential distribution when first r failure times are recorded. Which of the two will you prefer?

   (c) What do you understand by hot and cold use of redundancy in reliability improvement? Discuss.

20, 20, 20

SECTION C

(Operational Research)

9. (a) Explain the concept, scope and tools of OR as applicable to business and industry.

   (b) What are the basic steps in constructing a model? Describe how to decide whether a model could optimize profits, sates, costs or other trends. Illustrate your answer with suitable examples.

   (c) What do you understand by Markov chains? Explain how it can be used for predicting sales-force needs.

20, 20, 20

10. (a) Discuss (M/M/1) : (∞/FCFS) queueing model and find the expected line length in the system.

    (b) State the general linear programming problem Given its mathematical formulation and explain the terms:

    (i) Basic solution

    (ii) Feasible solution

    (iii) Basic feasible solution, and

(iv) Optimal solution.

(e) Show that the optimal solution of a linear programming - problem can be found amongst its feasible solutions.

20, 20, 20

11. (a) Discuss An inventors' model with probabilistic demand.

(b) Define assignment problem and give its mathematical formulation. Also give an algorithm to solve an assignment problem.

(c) Explain any one method for solving a transportation problem. Would you recommend this method to solve an assignment problem?

15, 25, 20

12. (a) Describe the inter-relation amongst machine language, FORTRAN compiler, the FORTRAN source program and object code.

(b) Write a note on library functions available in FORTRAN IV.

(c) Write a FORTRAN program to evaluate the following sum:

$$S\sum_{i=1}^{10}\frac{(-1)^n X^{n/2}}{n(n-1)}$$

20, 20, 20

SECTION D

(Quantitative Economics)

13. (a) Why is the multiplicative model the most commonly used assumption, as compared to additive model, in time series analysis? Discuss the role of the method of moving averages in time series analysis

(b) Explain clearly with the help of an illustration how seasonal index is useful in planning sales or production for specific periods. Are there any limitations of seasonal index?

(c) Elucidate the role of random components in a time series. Discuss the variety difference method in this connection.

20, 20, 20

14. (a) Slate the properties of an 'ideal' index number and discuss four important index numbers in respect of them.

(b) Explain the meaning and uses of cost of living index number. How are weights assigned to various groups and sub-groups?

(c) Explain what is meant by 'bias' in an index number. Given an example of one index number having an 'upward bias' and one index number having a 'downward bias'

21, 20, 15

15. (a) What do you mean by price elasticity of demand? Row will you obtain this measure on the basis of time series data? Discuss the various difficulties that are likely to occur in this connection.

(b) Show that the least squares estimate of the slope coefficient in the simple linear regression equation is inconsistent when both the variables are subject to errors of measurement.

(c) Given the two-equation model:

$$C_t = \alpha + \beta Y_t + u_t$$
$$Y_t = C_t + I_t$$

where C_t and Y_t are endogenous variables and I_t is an exogenous variable, derive the two-stage least square estimate of β and prove that it is consistent under certain conditions (to be stated).

20, 20, 20

16. (a) Describe how you will test the presence of multicollinearity among the explanatory variables ma linear model.

(b) Explain the concept of identification for a simultaneous equation model. Discuss with an example, how the specification of disturbances can lead to identifying relations I a simultaneous equations model.

(c) Write a note on techniques and models used in demand projections.

20, 20, 20

SECTION E

(Demography and Psychometry)

17. (a) Explain why the modality situation of two places cannot usually be compared on the basis of crude death rates. Describe the construction of standardized death rates for this purpose.

(b) What factors govern the growth of population? How do you find out whether the population of a country is increasing or decreasing or remaining stationary?

(c) Write a note on the importance of vital statistics for demographic purposes

20, 20, 20

18. (a) What is meant by force of mortality at age x? Derive Makeham's formula starting from suitable assumptions

(b) Explain gross and net reproduction rates. What interpretation will you give if the net reproduction rate is just 1, less than 1 or greater than 1?

(c) Describe the structure of a complete life table.

20, 20, 20

19. (a) Describe a suitable method of constructing an abridged life table.

(b) Describe briefly the component method for population projection mentioning the demographic data which will be required for this purpose.

(c) Given a brief account of the National Sample Sample Survery in the field of demography and show how far the finding have been useful in preparation of Five Year Plans in this respect.

20, 20, 20

20. (a) Why is it considered desirable to convert raw scores to some standard scores? Define' standardized scores' and 'normalized scores' and describe how they are derived.

(b) Explain the term Intelligence Quotient (IQ). Describe any method of measuring it.

(c) Discuss the various methods to determine the reliability of test scores.

20, 20, 20