CC-11 Sem-5 Statistical Inference-II

1. What is sufficiency? Suppose $X_1, X_2 \sim iid\ Bernoulli(\theta)$. Check whether the statistic $T = X_1 + 2X_2$ is sufficient.
2. Define completeness and bounded completeness. "Completeness imply bounded completeness but the converse may not be true"— discuss.
3. What an unbiased estimator of a parametric function $g(\theta)$? For $Poisson(\lambda)$ distribution, show that $C = \{\alpha\bar{X} + (1-\alpha)s^2; 0 \leq \alpha \leq 1\}$ is a class of unbiased estimators of λ, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
4. Explain the concept of estimation in statistical inference.
5. Consider the Poisson family of distribution and suggest two statistics of which one is sufficient and another is not sufficient for the family.
6. State Rao-Cramer lower bound (RCLB) with the regularity conditions.
7. Give counter examples for the support of the following statements –
   "All the sufficient or minimal sufficient statistics are not complete" & "Unbiased estimator may not be unique".
8. Show that unbiased estimator based on complete statistic is unique. Discuss the use of the above fact in finding the best estimator.
9. Describe the properties of maximum likelihood estimator (MLE).
10. Let $X_i \sim iid\ P(\theta), \theta > 0; i = 1,2, \dots, n$. Find the MLE of θ where it is given that the number of observations equals to zero is n_0. Hence provide the MLE of $\ln\theta$.
11. Define sufficient and minimal sufficient statistic. Consider the family of normal distribution with mean μ and variance σ^2 and show that the distribution belongs to exponential family of distribution. Hence find the sufficient (or minimal sufficient) statistics for (μ, σ^2).
12. Define a parametric function. When a parametric function is said to be estimable? Give an example of a non-estimable parametric function.
13. State and prove Factorization theorem considering discrete distribution.
14. Let $X_1, X_2, \dots, X_n$ be a random sample of from $N(\theta, \theta), \theta > 0$, then find the maximum likelihood estimator (MLE) of θ.

15. Let $X_1, X_2, \dots, X_n$ be a random sample from $R(\theta, \theta+1), \theta \in \mathbb{R}$. Find an unbiased estimator of θ based on sufficient statistic.
16. If T_1 *and* T_2 are two unbiased estimators of $\gamma(\theta), \theta \in \Omega$, and ρ is the correlation coefficient between T_1 *and* T_2 then show that $\sqrt{e_1 e_2} - \sqrt{(1-e_1)(1-e_2)} \leq \rho \leq \sqrt{e_1 e_2} + \sqrt{(1-e_1)(1-e_2)}$. Where $e_i = \frac{Var_\theta(T)}{Var_\theta(T_i)}, i = 1,2,$ T is the UMVUE $\gamma(\theta)$.
17. Let $X \sim Bin(n, \theta), \theta \in (0,1)$. Find UMVUE of θ^2 based on single observation.
18. Show that for $Cauchy(\theta, 1)$ distribution the sample median, not the sample mean, is consistent for $\theta \in \mathbb{R}$.
19. Suppose $X_1, X_2, \dots, X_n \sim iid\ N(\mu, \sigma^2), \mu \in \mathbb{R}$ and $\sigma > 0$. Show that sample mean $(\bar{X}_n)$ and sample median $(\tilde{X}_n)$ both are CAN estimators of μ. Also find their asymptotic relative efficiency and comment.
20. Discuss the use of randomized test with an example.
21. State fundamental Neyman-Pearson lemma. Show that UMP test is necessarily unbiased.
22. Suppose $X_1, X_2, \dots, X_n \sim iid\ f_\theta(x) = \theta x^{\theta-1}; 0 < x < 1,\ \theta > 0$. Find UMP test for testing $H_0: \theta = \theta_0 \quad vs \quad H_1: \theta < \theta_0$, where $\theta_0\ (> 0)$ is known. Check for consistency of the test.
23. What is uniformly most powerful unbiased (UMPU) test? Give an example of a biased size α test.
24. Suppose

$$X_1, X_2, \dots, X_{n_1} \overset{iid}{\sim} N(\mu_{10}, \sigma^2) \ \ and \ \ Y_1, Y_2, \dots, Y_{n_2} \overset{iid}{\sim} N(\mu_{20}, \theta\sigma^2)$$

Where $\mu_{10}, \mu_{20} \in \mathbb{R}$ are known and $\theta, \sigma > 0$ are unknown. The samples are independent. Derive optimum size α test for testing $H_0: \theta = \theta_0 \quad vs. \quad H_1: \theta \neq \theta_0$. Find the power function of the test and hence check the unbiasedness.

25. Suppose $X_1, X_2, \dots, X_n \overset{iid}{\sim} \exp(mean = \theta)\,;\ \theta > 0$. Consider the problem of testing

$$H_0: \theta = \theta_0 \quad vs. \quad H_1: \theta > \theta_0$$

Rewrite the testing problem with the help of the probability $P(X > \theta_0)$. Derive size α likelihood ratio test for testing problem described by the probability, where, instead of the original observations, the total number of observations exceeding θ_0 is given. Find the power function of the test and hence show that it is unbiased. Also check its consistency.

26. What is UMA confidence interval? Derive UMA confidence interval from UMP test with a suitable example.
27. Derive the exact standard error of sample variance (m_2).
28. Discuss Yates' correction for continuity in a 2×2 contingency table.
29. Let $X_1, X_2, \dots, X_n$ be independent and identically distributed random variables with $E(X_1) = \mu(\neq 0), Var(X_1) = \sigma^2(0 < \sigma < \infty)$. Is it possible to find real sequences $\{a_n\}$ and $\{b_n\}$ such that $a_n\left(\overline{X_n}^3 - b_n\right)$ converges in Distribution to a non-degenerate random variable, where $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i, n \geq 1$. Justify your answer.
30. State the Weak Law of Large Numbers (WLLN). If WLLN holds for a sequence $\{X_n\}$, then $E\left[\frac{\bar{X}_n^{\,2}}{1+\bar{X}_n^{\,2}}\right] \to 0 \; as \; n \to \infty$.
31. Let $X_1, \dots, X_n$; $n \geq 1$ be a sequence of iid random variables with common p.d.f.

$$f(x) = \begin{cases} e^{-(x-\theta)} & if \;\; x \geq \theta \\ 0 & if \;\; x < \theta \end{cases}$$

Show that (i) $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to 1 + \theta$ and (ii) $\min\{X_1, X_2, \dots, X_n\} \to \theta$ in probability.

32. State and prove DeMoivre-Laplace Central Limit theorem. Hence or otherwise, show that,

$$e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} \to \frac{1}{2} \;\; as \;\; n \to \infty.$$

33. State and prove the result related to delta method.
34. Derive the concept of variance stabilizing transformation and discuss its advantages. Hence provide test for the equality of correlation coefficients of K independent bivariate normal distributions.

35. Define convergence in law. Let $\{X_n\}$ be a sequence of random variables and X be another random variable defined on the same probability space $(\Omega, \mathcal{F}, P)$ such that $X_n \xrightarrow{L} X$. Let $g(\cdot)$ be a continuous function then show that $g(X_n) \xrightarrow{L} g(X)$.
36. State WLLN for a sequence of random variable. For a sequence of independent random variables $\{X_n\}$ with $E(X_k) = \mu_k$ and $Var(X_k) = \sigma_k^2 \in (0, \infty), k \geq 1,$ show that if $\sum_{n\geq 1} \sigma_n^2/n^2 < \infty$ then WLLN holds for the sequence.
37. Find the asymptotic variance of the bivariate moment $m_{r,s}$ when the observations are drawn from a bivariate distribution. Hence or otherwise find the same for correlation coefficient r and find the value when the parent distribution is bivariate normal.
38. Define convergence in distribution and convergence in probability. Does convergence in distribution imply convergence in probability?
39. . Suppose $X_n \sim iid\ N(0,1), n \in \mathbb{N}$. Define $U_n = \frac{X_1}{X_2} + \frac{X_3}{X_4} + \cdots + \frac{X_{2n-1}}{X_{2n}}$ and $V_n = X_1^2 + X_2^2 + \cdots + X_n^2$. Find the asymptotic distribution of $Z_n = \frac{U_n}{V_n}$.
40. Find the large sample distribution of $b_1\ and\ b_2$.
41. Describe the large sample test for Binomial proportions p, where r.v. $X \sim B(n,p)$ considering all possible alternative hypotheses.
42. Let there be two independent rvs X following $B(m, p_1)$ and Y following $B(n, p_2)$. Describe large sample test for $H_0: p_1 = p_2$ against $H_0: p_1 \neq p_2$.
43. Describe the large sample test for $H_0: m = m_0$ against all possible alternatives where rv $X \sim Poisson(m)$.
44. Describe the large sample test for $H_0: m_1 = m_2 = \cdots = m_k$ against $H_1: not\ H_0$, where independent Poisson variables $X_i \sim Poisson(m_i),\ i = 1,2, \dots, k$.
45. For a random sample taken from a $BN(\alpha, \beta, \sigma_1, \sigma_2, \rho)$, where ρ differs widely from zero, establish a test for $H_0: \rho = \rho_0$ and also set confidence interval for ρ.
46. For two independent random samples taken from $BN(\alpha_1, \beta_1, \sigma_{11}, \sigma_{12}, \rho_1)$ and $BN(\alpha_2, \beta_2, \sigma_{21}, \sigma_{22}, \rho_2)$, establish a test for $H_0: \rho_1 - \rho_2 = 0$ against $H_1: \rho_1 - \rho_2 \neq 0$ when each ρ_i differs widely from zero and also set a confidence interval for $(\rho_1 - \rho_2)$.

47. a) What is a Pearsonian Chi Square statistic? Describe its important uses in statistical inference.

b) Show that Pearsonian Chi square statistic follows Chi-Square distribution with $(k-1)$ degrees of freedom if there be k mutually exclusive classes in the population.

48. Describe the test for goodness of fit when

i) Population proportions of different mutually exclusive classes constituting the population are known

ii) Population proportions of different mutually exclusive classes constituting the population are unknown

49. Describe the test for homogeneity of l similarly classified populations using Pearsonian Chi-Square statistic.

50. Describe the procedure of testing independence of two attributes A & B having levels $A_1, A_2, A_3, \dots A_k$ & $B_1, B_2, B_3, \dots B_l$ respectively constituting a population consisting of $k \times l$ mutually exclusive classes (A_i, B_j); $i = 1,2, \dots k$ and $j = 1,2, \dots l$, by using Pearsonian Chi Square statistic.

51. Establish the Snedecor-Brandt formula for a population classified in a $k \times 2$ $(or,$ $2 \times k)$ contingency table for testing independence of two attributes having k and 2 $(or,$ 2 and $k)$ levels respectively or for testing 2 populations both similarly classified into k classes using Pearsonian Chi Square statistic.

52. Find the simplified formula for Pearsonian Chi Square statistic in a2×2 contingency table.

53. What is Yates correction for continuity? Why this correction is required and how this correction is to be achieved?

CC-12 Sem-5 Linear Models and Regression

1. Describe how, by analysis of covariance, one may use data on concomitant variables to increase the precision of estimates.
2. State the nature of problems tackled by analysis of variance. State the underlying assumptions needed in testing a hypothesis in such a problem.
3. Discuss the situation where a fixed effect model is justified in a two way classified data with one observation per cell.
4. Draw up the blank analysis of covariance table (with one auxiliary variable) for a randomized block experiment comparing r treatments in r blocks. Deduce the expression for the average sampling variance of an estimated treatment difference and hence the gain in precision due to the use of the auxiliary variable.
5. What do you mean by orthogonal splitting of total sum of squares?
6. Starting from an appropriate additive fixed effect model prepare the analysis of variance table for a set of two-way classified data with r (>1) observations per cell.
7. Derive the expectations of various mean squares in the above table. When are these equal to the error variance σ^2?
8. Let $Y_i \sim N(\alpha_0 + \alpha_1 x_i, \sigma^2)$, $i = 1, 2, \ldots, p$ independently. Apply the theory of linear hypothesis to test $\alpha_1 = 0$. Show that the test statistic is increasing in r^2, the square of the sample correlation coefficient.
9. State a linear hypothesis in a linear model. How will you test such a hypothesis?
10. Consider the following multiple regression model

$y_\alpha = \beta_0 + \beta_1 x_{1\alpha} + \ldots \ldots + \beta_p x_{p\alpha} + e_\alpha$, $\alpha = 1, 2, \ldots, n$

where e_α's are random errors which follow independent $N(0, \sigma^2)$. Obtain a test statistic for testing $H_0: \beta_1 = \ldots = \beta_p = 0$. Express this statistic in terms of the sample multiple correlation coefficient and examine the nature of this function. Also obtain a $100(1 - \alpha)$ % confidence interval to the function $\beta_i - \beta_j$.

11. Explain how valid error is determined in testing any hypothesis in analysis of variance. Show, with reference to a two-way layout with m (> 1) observations per cell, that it depends on the model but not on the hypothesis to be tested.

12. It is known from past experience that the fertility gradient of a given land varies in two perpendicular directions. It is desired to test differential effect of *p* varieties of wheat. How will you use this land to have a well-designed experiment? Give detailed analysis of the data obtained from the experiment.

13. Explain how error control is achieved through the technique of analysis of covariance. Test whether the inclusion of the concomitant variable is worthwhile or not.

14. Consider a two-way classified data with r (>1) observations per cell. Define different effects and analyse the data to test the significance of the effects. How will you interpret when the hypothesis of no interaction effect is rejected but the hypotheses of absence of the main effects are accepted?

15. Assuming linear regression of the study variable y on p independent variables x_1, x_2, ..., x_p, test for the significance of the regression coefficients. Express the test statistics in terms of multiple correlation coefficient. Also test for the equality of any two regression coefficients.

16. Stating clearly all the necessary assumptions, analyse in detail the one way classified data (fixed effects model). Show that if the number of observations to be taken is fixed, the average variance of the estimators of all elementary contrasts of the effects due to the levels of the factor is minimum if the number of observations on each level is the same. Also state the changes in the analysis, if any, for the corresponding random effects model.

17. A fixed set of p methods of teaching was applied to r students chosen randomly from each of q schools which form a random sample of all schools in a country. Describe procedures for analysing the scores of the students to examine the existence of differences among the methods, among the schools and the significance of interactions. Write the assumptions clearly.

18. For forecasting the rainfall in a region, the average rainfall together with the average temperature and the average humidity of the place for the last n years were noted. On the basis of these data find a suitable forecasting formula based on linear regression. Write down the distributions of the estimators of the regression coefficients. Hence or otherwise describe the procedure for testing whether temperature can be dropped from the equation.

19. Explain with examples what you mean by regression of Y on X when X is (a) stochastic and (b) non-stochastic.

Briefly describe how, on the basis of observed data set (x_i, y_i), i = 1(1) n you will proceed step by step to test

(i) presence of regression

(ii) linearity of regression

(iii) whether the regression is a polynomial of a given degree.

Also indicate the advantage of using orthogonal polynomials in (iii).

20. What is a linear model in analysis of variance? Distinguish between fixed-effects model and random effects model.

21. Starting from an appropriate random effects model in balanced one-way classified data, show that the expected value of the F-statistic cannot be less than one.

22. Consider two variables y and x where y is the study variable and x is the independent variable. Suppose that one can approximate the regression of y on x by a polynomial. Describe a procedure for determining the degree of the polynomial.

23. State how the concomitant information helps to control the error of an experiment.

24. Consider the following analysis of covariance model in a CRD:

$$y_{ij} = \mu + \tau_i + \gamma x_{ij} + e_{ij} \qquad i = 1, 2, \ldots, I;\ j = 1, 2, \ldots, J$$

when y_{ij}, x_{ij}'s are the observations corresponding to the study variable and the concomitant variable respectively; μ, τ_i, γ stand respectively for the general mean effect, the i-th treatment effect and the regression coefficient and e_{ij}'s are random errors which follow $N(0, \sigma^2)$ distribution.

(i) Find a $100(1 - \alpha)$ % confidence interval for γ.

(ii) Obtain an unbiased estimator of $\tau_1 - \tau_2$ and the variance of the estimator.

(iii) Suggest a test for H_0: $\tau_1 + \tau_3 = 2\tau_2$.

25. Describe a procedure for testing the presence of any regression of y on x.

DSE-B1-1 Sem-5 Operations Research

1. Define a convex set.

2. Define an extreme point of a convex set.

3. Define a slack variable in a Linear Programming Problem (LPP).

4. Define a surplus variable in the LPP.

5. Indicate the criteria for optimum solution in a Simplex Tableau for maximization LPP.

6. Indicate the criteria for unbounded solution in a Simplex Tableau.

7. Indicate the criteria for infinite number of optimum solutions in a Simplex Tableau.

8. Define duality in connection with the LPP.

9. Is it possible to develop a South-East corner rule to generate the initial solution in a Transportation problem?

10. State the criteria to check for optimality of a transportation problem through the modified difference (MODI) method.

11. What do you mean by pay of matrix?

12. What do you mean by a zero-sum game?

13. Define a saddle point in a pay-off matrix.

14. Explain the role of an artificial variable in solving the LPP.

15. Show that if the i^{th} primal constraint is an equation, then the i^{th} dual variable is unrestricted in sign.

16. Express the transportation problem as LPP. Find the rank of the coefficient matrix of an $m \times n$ transportation problem and interpret the result.

17. Describe the Vogel's method to find the initial basic feasible solution of a transportation problem.

18. Define an assignment problem. Is it possible to express such a problem as LPP? — Justify.

19. Explain with suitable examples how the theory of dominance may be used to reduce the order of a matrix game.

20. Derive the minimum ratio exit criterion for LPP.

21. Describe the dual simplex method to solve LPP.

22. Distinguish between pure and mixed strategies in a two-person zero-sum game. Derive the mixed strategies in a 2 × 2 zero-sum game, without any saddle point in the pay-off matrix.

23. A firm produces two products. These products are processed on three different machines. The time required to manufacture one unit of each of the two products and the daily capacity of the three machines are given in the following table (a blank in the table indicates that the corresponding product does not require processing in that particular machine).

	Time per unit (in minutes)		Machine capacity (minutes per day)
Machine	Product A	Product B	
M-1	--	3	441
M-2	4	--	472
M-3	2	5	730

It is required to determine the daily number of units to be manufactured for each product. The profits per unit for product A and B (in thousand Rs) are respectively 4 and 3. It is assumed that all the amounts produced are consumed in the market. The objective is to maximize the daily profit. Formulate the problem as an LPP and solve it.

24. Consider the following 3 persons and 4 Jobs assignment problem where the elements in the matrix are respective net returns of assignment, in suitable units. Find the optimum assignment schedule such that the total net return is maximized.

Person	Job			
	J1	J2	J3	J4
A	20	25	22	28
B	15	18	23	17
C	19	17	21	24

25. Solve the following two-person zero-sum game.

		Player B			
Player A		B1	B2	B3	B4
	A1	2	2	3	-1
	A2	4	3	2	6

26. Write the dual of the following problem (dual must contain one unrestricted variable):

$Minimize\colon\ Z = x_1 + x_2 + x_3,$

subject to

$x_1 - 3x_2 + 4x_3 = 5$

$x_1 - 2x_2 \leq 3$

$2x_2 - x_3 \geq 4$; $x_1, x_2 \geq 0$ and x_3 is unrestricted in sign.

27. There are five jobs to be assigned, one each to five machines and the associated cost matrix is as follows:

		Machine 1	2	3	4	5
Job	A	11	17	8	16	20
	B	9	7	12	6	15
	C	13	16	15	12	16
	D	21	24	17	28	26
	E	14	10	12	11	15

Find the assignment of machines to jobs that will minimize the total cost.

28. Write the algorithm to solve LPP using Graphical method for maximization of profit.

29. What do you mean by optimum solution? How multiple optimal solutions are recognized when using the simplex algorithm?

30. Check whether the following system of linear equations has degenerate solutions: $2x_1 + x_2 - x_3 = 2$, $3x_1 + 2x_2 + x_3 = 3$. If yes, find all the degenerate basic feasible solutions.

31. What is the difference between Assignment Problem and Transportation Problem?

32. Find all the basic solutions of the following system :

$$4x_1 + 2x_2 + x_3 = 4, \quad 2x_1 + x_2 + 5x_3 = 5.$$

33. Write the steps of the algorithm for solving LPP by the Simplex method.

34. Write the LPP model of the following Transportation problems:

		Destination				
		I	II	III	IV	Supply
Source	A	40	25	22	33	100
	B	44	35	30	30	30
	C	38	38	28	30	70
Demand		40	20	60	30	

35. Solve the following LPP by the two-phase method and give your conclusions about the solution:

$$Minimize: \quad z = 4x_1 + 3x_2$$

$$subject\ to$$

$$2x_1 + x_2 \geq 10,$$

$$-3x_1 + 2x_2 \leq 6,$$

$$x_1 + x_2 \geq 6,$$

$$x_1, x_2 \geq 0.$$

36. What do you mean by basic feasible solution? Find the initial basic feasible solution for the following transportation problem and test whether this solution is optimal or not. If not, find the optimal solution.

		Market				
		A	B	C	D	Available
Plant	I	14	9	18	6	11
	II	10	11	7	16	13
	III	25	20	11	34	19
Requirements		6	10	12	15	