STAT C-403
STATISTICAL QUALITY CONTROL

<u>Statistical Quality Control</u>

1) What is quality; discuss briefly various dimensions of quality?

2) What are chance and assignable causes of variability? What part do they play in the operation and interpretation of a Shewhart control chart? When is the process said to be in statistical control?

3) Define and derive the Rejection Control Limits and when these are used?

4) Distinguish between chance causes and assignable causes of variations. Justify the use of 3-sigma limits in the construction of control charts. Find the control limits for $\bar{X}$-chart when standards are not given.

5) Give the classification of control charts w.r.t quality characteristic of the variable.

6) Construct 3σ control limits for R-chart when standards are not given. Construct the control chart showing the three control limits when sample size is less than seven.

7) Discuss the types of control charts used for analyzing or controlling process variability, when standards are unknown.

8) Discuss R-chart and s-chart for controlling process variability. Which chart is more appropriate for measuring variability for larger sample size and why?

9) Give the interpretation of patterns of $\bar{X}$ and R-chart.

10) Discuss the interpretation of $\bar{X}$ and s charts.

11) Discuss the significance of rational sub-grouping in $\bar{X}$ and R-chart.

12) Distinguish between natural tolerance limits and specification limits. Discuss the ideal situation?

13) Discuss the concept of Process Capability Analysis.

14) 25 Samples of size n = 5 each are taken from a manufacturing process every hour. Assuming quality characteristic is normally distributed, $\bar{x}$ and R are computed for each sample. Given,

$$\sum_{i=1}^{25} \bar{x}_i = 662.5, and \sum_{i=1}^{25} R_i = 9.0$$

(i) Find the 3σ control limits for the $\bar{x}$ and R charts.

(ii) If the specifications limits are 26.40 ± 0.50 and mean of the process is 26.40, estimate the fraction non-conforming, assuming that both charts exhibit control.

15) How do revised control limits differ from modified control limits?

16) Distinguish between process control and product control.

17) Distinguish between a defective and a defect; construct suitable control charts to test whether the process is under Statistical control or not with respect to defects per unit.

18) Name the control charts for attributes. How do you deal with problem of variable sample size, suggest the appropriate control chart and obtain their control limits.

19) Describe the single sampling plan for attributes.

20) Discuss the significance of O.C., A.O.Q., and A.T.I. curves in sampling inspection plans.

21) Discuss the lot quality protection approach to decide the sample size and acceptance number in respect to single sampling plan.

22) Discuss the average quality protection approach to decide the sample size and acceptance number in respect to single sampling plan.

23) What is Double Sampling Plan, obtain the ASN and ATI for Double Sampling Plan

24) Distinguish between:

   a. Incoming quality and outgoing quality
   b. Consumer's risk and producer's risk
   c. AOQ and AOQL
   d. Acceptance rejection and acceptance rectification plans
   e. Single sampling plan and double sampling plan
   f. ASN and ATI for single sampling plan,
   g. Flow Charts of (i) single sampling rectification plan and (ii) double sampling rectification plan when a lot is rejected based on first sample.

25) Describe in brief the steps required for construction of Price Index Numbers.

26) Define at least 6 methods of constructing the Price Index Numbers and which Index Number is considered to be an Ideal Index Number.

27) What is a Cost of Living Index Numbers and how these are used for calculating the real wages of the Worker?

28) What is meant by splicing of index numbers? Explain and Illustrate.

29) What is a Chain Index? Show that the chain indices are equal to the corresponding fixed base indices if the formula used satisfies the circularity test.

30) Explain the concept of (i) base shifting and (ii) deflating of index numbers mentioning its importance.

STAT C-402 LINEAR MODELS

B.Sc.(H) Statistics/Linear Models/SemesterIV

1. For the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, (i = 1, 2, …, n), where $\epsilon_i \sim NID(0, \sigma^2)$,
(i) Obtain the least square estimates of β_0 and β_1.
(ii) Verify the bias and variance properties of $\hat{\beta}_0$ and $\hat{\beta}_1$.
(iii) Show that $\text{cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\frac{\sigma^2 \overline{x}}{S_{xx}}$.
2. Prove that, for every estimable function, there exists a unique BLUE.
3. Consider the simple linear regression model:
$Y = \beta_0 + \beta_1 x + \epsilon$ with $E(\epsilon) = 0$, $V(\epsilon) = \sigma^2$, ϵ's are uncorrelated. Show that:
(i) $\text{cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\frac{\sigma^2 \overline{x}}{S_{xx}}$.
(ii) $E(MSR) = \sigma^2 + \beta_1^2 S_{xx}$

4. (a) For the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, (i = 1, 2, …, n), where $\epsilon_i \sim NID(0, \sigma^2)$, Obtain the least square estimates of β_0 and β_1. Show that they are unbiased. Also find their variances. Also, find the unbiased estimator of the error variance.
(b) For a Simple linear regression model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, find the 100 (1- α)% confidence interval for the difference $\beta_1 - \beta_0$.
5. If $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_n)'$ be a vector of n independent standard normal variates. Let $Q_1 = \underset{\sim}{Y}' A_1 \underset{\sim}{Y}$ and $Q_2 = \underset{\sim}{Y}' A_2 \underset{\sim}{Y}$ be distributed as χ^2 with n_1 and n_2 degrees of freedom respectively. Show that the necessary and sufficient condition for Q_1 and Q_2 to be independently distributed is $A_1 A_2 = 0$.
6. For a given model $\underset{\sim}{Y}_{n\times1} = X_{n\times p} \underset{\sim}{\beta}_{p\times1} + \underset{\sim}{\epsilon}_{n\times1}$ with $E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}, V(\underset{\sim}{\epsilon}) = \sigma^2 I$ and $\rho(X) = p < n$, prove that the least squares estimator of $\underset{\sim}{\beta}$ is BLUE. Also obtain an unbiased estimator of σ^2.
7. For a given model $\underset{\sim}{Y}_{n\times1} = X_{n\times p} \underset{\sim}{\beta}_{p\times1} + \underset{\sim}{\epsilon}_{n\times1}$ with $E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}, V(\underset{\sim}{\epsilon}) = \sigma^2 I$ and $\rho(X) = r < p$, prove that, in the class of all unbiased estimates of any estimable function of the parameters in the model, the least squares estimator is BLUE. Also obtain an unbiased estimator of σ^2.
8. Suppose $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_n)'$ to be a vector of n independent standard normal variates then a necessary and sufficient condition for $\underset{\sim}{Y}' A \underset{\sim}{Y}$ to be distributed as chi-square variate with k d.f. is that A is an idempotent matrix of rank k.
9. State and prove Cochran's theorem.
10. What is a parametric function? Derive a necessary and sufficient condition for which a parametric function is estimable.
11. Test the multiple regression model for significance of individual regression coefficients.
12. Explain the uses of residual analysis and coefficient of multiple determination.
13. Explain the concept of extra sum of squares, sequential and partial F-tests.
14. Write a note on the extra sum of squares method, that can be used to test the hypotheses about any subset of regressor variables.
15. Define multiple regression model and polynomial regression model. Explain the role of orthogonal polynomials in fitting polynomial models in one variable.
16. For the general linear model, set the appropriate hypothesis for testing the significance of regression and develop the test for significance of individual regression coefficients, explain why this test is known as marginal test.

17. Examine the problem of testing the hypothesis on regression coefficient in multiple linear regression model.
18. Obtain the confidence interval on mean response at a particular point under general linear model.
19. Suppose the postulated model is $E(Y) = \beta_0+\beta_1x_1$ but the true model is $E(Y) = \beta_0+\beta_1x_1 +\beta_2x_2$. Show that both $\hat{\beta}_0$ and $\hat{\beta}_1$ are biased by an amount that depends the values of x's.
20. Suppose the postulated model is $E(Y) = \beta_1x$ but the true model is $E(Y) = \beta_0+\beta_1x$. Show that $\hat{\beta}_1$ is biased by an amount that depends on β_0 and the values of x's.
21. Suppose we postulate the model $E(Y) = \beta_0+ \beta_1x$ but the model $E(Y) = \beta_0+ \beta_1x +\beta_{11}x^2$ is actually the true response function, unknown to us. If we use observations of Y at $x = -1, 0, 1$ to estimate β_0 and β_1 in the postulated model, what biases will be introduced.
22. Suppose we postulate the model $E(Y) = \beta_0+ \beta_1x$ but the model $E(Y) = \beta_0+ \beta_1x +\beta_{11}x^2+\beta_{111}x^3$is actually the true response function, unknown to us. If we use observations of Y at $x = -3, -2, -1, 0, 1, 2, 3$ to estimate β_0 and β_1 in the postulated model, what biases will be introduced?
23. For a simple linear regression model, develop a test for lack of fit.
24. Discuss the problem of testing for lack of fit in simple linear regression model.
25. Derive the analysis of covariance for a single factor experimental design with one covariate. Also obtain the standard error of the difference between any two adjusted treatment means.
26. Derive the analysis of variance for two way classified data with one observation per cell under fixed effect model. Also obtain the variance of the estimated parameters.
27. Derive the expectations due to all different sources of variation for an ANOVA two way classification with one observation per cell under fixed effects model. Under what conditions are these expectations equal?
28. Derive the analysis of variance for two-way classified data under fixed effects model for m observations per cell.
29. Write notes on any two of the following:

(i) General linear model,
(ii) Role of orthogonal polynomials in fitting polynomial models in one variable
(iii) Bias in regression estimates
(iv) Orthogonal columns in X matrix
(v) Stepwise regression method
(vi) Partial F-test
(vii) Coefficient of determination
(viii) Residual analysis
(ix) Linear Parametric Function

30. Suppose X_i, Y_i, Z_i, $i = 1, 2, \ldots, n$ are 3n independent observations with common variance σ^2 and expectations $E(X_i) = \theta_1$, $E(Y_i) = \theta_2$, $E(Z_i) = \theta_1 - \theta_2$, $i = 1, 2, \ldots, n$. Find the BLUEs of θ_1, θ_2, $\text{cov}\left(\hat{\theta}_1, \hat{\theta}_2\right)$ and compute the residual sum of squares. Also find BLUE of $\theta_1 + \theta_2$.
31. (a)Consider the model $E(Y_1) = 2\beta_1-\beta_2-\beta_3$, $E(Y_2) = \beta_2-\beta_4$, $E(Y_3) = \beta_2 + \beta_3- 2\beta_4$ with usual assumptions. Find the estimable functions.

(b) Consider three independent random variables Y_1, Y_2, Y_3having common variance $\sigma^2$and expectations as given below

$E(Y_1) = \beta_1 + \beta_3$, $E(Y_2) =\beta_1 + \beta_2$, $E(Y_3) = \beta_1 + \beta_3$ Determine the condition of estimability of the parametric function . Also determine the sum of squares due to error .

Consider the model $E(Y_1) = 2\beta_1 + \beta_2$, $E(Y_2) = \beta_1-\beta_2$, $E(Y_3) = \beta_1-\beta_3$ with usual assumptions. Obtain the BLUE of $\beta_1+ \beta_2$ and its variance.

(c) Consider $y_{ij}= \mu + \alpha_i + \epsilon_{ij}$ $i=1,2$, $j= 1,2,3$. Write the normal equations. Are $\mu + \alpha_1$, $\alpha_1 + \alpha_2$, $\alpha_1 - \alpha_2$, $\mu + \alpha_1 + \alpha_2$ estimable and why ?

32. Consider the model $E(Y_1) = \beta_1 + \beta_2$, $E(Y_2) = 2\beta_1$, $E(Y_3) = \beta_1 - \beta_2$ with usual assumptions. Find sum of squares due to error.

33. Consider the model $E(Y_{ij}) = \alpha_i + \beta_j$, $i = 1, 2$; $j = 1, 2$. Find the condition under which $\ell_1\alpha_1 + \ell_2\alpha_2 + m_1\beta_1 + m_2\beta_2$ is an estimable function.

34. (a)Suppose $\underset{\sim}{Y} \sim N_3(\underset{\sim}{0}, I)$ and let

$$A = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}, C = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(i) Are $\underset{\sim}{Y}'A\underset{\sim}{Y}$ and $2y_1 + y_2$ independent?

(ii) Are $\underset{\sim}{Y}'A\underset{\sim}{Y}$ and $B\underset{\sim}{Y}$ independent?

(iii) Are $\underset{\sim}{Y}'A\underset{\sim}{Y}$ and $\underset{\sim}{Y}'C\underset{\sim}{Y}$ independent?

(iv) Are $\underset{\sim}{Y}'A\underset{\sim}{Y}$ and $\underset{\sim}{Y}'D\underset{\sim}{Y}$ independent?

(v) Are $\underset{\sim}{Y}'C\underset{\sim}{Y}$ and $\underset{\sim}{Y}'D\underset{\sim}{Y}$ independent?

(b) Suppose $\underset{\sim}{Y} \sim N_3(\underset{\sim}{0}, I)$ and let

$$A = \frac{1}{3}\begin{bmatrix} 2 & 0 & -\sqrt{2} \\ 0 & 3 & 0 \\ -\sqrt{2} & 0 & 1 \end{bmatrix}$$

Find the distribution of $\underset{\sim}{Y}'A\underset{\sim}{Y}$,stating the appropriate theorem to be used and also find the distribution of $\underset{\sim}{Y}'D\underset{\sim}{Y}$, where $D = I - A$. Are they independent?

35. Suppose $\underset{\sim}{Y} \sim N_3(\underset{\sim}{0}, I)$ and let $A = \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$. Are $\underset{\sim}{Y}'A\underset{\sim}{Y}$ and $y_1+y_2+y_3$ independent?

36. Four objects A, B, C, D are involved in a weighing experiment. Put together they weighed Y_1 grams; when A and C are put in the left pan of the balance and B and D are put in the right pan, a weight of Y_2 grams were necessary in the right pan for the balance. With A and B in the left pan and C and D in the right pan, Y_3 grams were needed in the right pan. Finally A and D in the left pan and B and C in the right pan, Y_4 grams were needed in the right pan. If the observations Y_1, Y_2, Y_3, Y_4 are all subject to uncorrelated errors with common variance σ^2, obtain the BLUEs of individual weights and the total weight of the four objects, and variance of the estimate of the total weight of four objects.

37. Stating clearly the underlying assumptions of the simple linear regression model through the origin, obtain the least squares estimate of the regression parameter along with its variance.

38. Develop a prediction interval for the future observation y_0 corresponding to a specified level x_0 of the regressor variable x in the simple linear regression model.

39. Suppose y_i ($i = 1, 2, \ldots, n$) is a random sample from a standard normal distribution. Show that $\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n}(y_i - \bar{y})^2$ are independently distributed.

40. Suppose the hypothesis of homogeneity of k-treatment means is rejected in ANOVA testing for one way classification under fixed effect model, how would you proceed to test the hypothesis of equality of two specific treatment means?

41. Suppose that we have fit the straight-line model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$ but the response is affected by a second variable x_2 such that the true regression function is $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

(i) Is the least-squares estimator of the slope in the original simple linear regression model unbiased?

(ii) Show the bias in $\hat{\beta}_1$.

42. Consider the simple linear regression model $Y = \beta_0 + \beta_1 x + \in$ with $E(\in) = 0$, $V(\in) = \sigma^2$ and $\in$'s are uncorrelated. Show that:

(i) $E(MSE) = \sigma^2$, and

(ii) $E(MSR) = \sigma^2 + \beta_1^2 S_{xx}$.

43. Write the simple linear regression model in matrix notation. Hence obtain the least squares estimators of the unknown parameters and their variances.

44. Suppose that we are fitting a straight line and wish to make the standard error of the slope as small as possible. Suppose that the "region of interest" for x is $-1 \le x \le 1$. Where should the observations x_1, x_2, …, x_n be taken? Discuss the practical aspects of this data collection plan.

45. Eight experiments are to be done at the coded levels (±1, ±1) of two predictor variables X1 and X2. Two experimenters A and B suggest the following designs:

A: Take one observation at each of (X1, X2) = (-1, -1) and (1,1) and take three observations at each of
(-1, 1) and (1, -1).

B: Take two observations at each of the four sites.

If a model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ is to be fitted by the least squares but it is feared there may be
Some additional quadratic curvature expressed by the extra $\beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2$, evaluate the
anticipated biases in the estimated coefficients b_0, b_1, b_2 for each design.

46. Values of a response Y are observed at six locations of a predictor variable X coded as -5, -3, -1, 1, 3, and 5. A model
$Y = \beta_0 + \beta_1 X + \varepsilon$ is to be fitted but there is fear that bias in the data, arising from second order (quadratic) effect
$\beta_2 X^2$, might occur. How would the presence of β_2 bias the estimates b_0 *and* b_1?

47. Show that the square of the multiple correlation coeff. R^2 is equal to the square of the correlation between Y and $\hat{Y}$?

48. We fit a straight line model to a set of data using the formulas $b = (X'X)^{-1} X'Y, \hat{Y} = Xb$ with the usual definitions.
We define $H = X(X'X)^{-1} X'$. Show that

$$SS(\text{due to regression}) = Y'HY = \hat{Y}'\hat{Y} = \hat{Y}' H^3 \hat{Y}$$

49. Show that $X'e = 0$.

50. Show that, for any linear model (where p is the number of parameters)

$$\sum_{i=1}^{n} V(\hat{Y}_i)/n = trace \frac{\{X(X'X)^{-1} X'\}\sigma^2}{n} = p\sigma^2/n.$$

51. Suppose $Y = X\beta + \varepsilon$ is a regression model containing a β_0 term in the first position, and $1 = (1 \quad 1 \quad 1 \ldots 1 \quad 1)'$ is
an nX1 vector of ones. Show that $(X'X)^{-1} X'1 = (1 \quad 0 \cdots 0 \quad 0)'$ and hence that $1'X(X'X)^{-1} X'1 = n$.

52. By noting that $X_0 = (1 \quad \overline{X_1} \quad \overline{X_2} \ldots)'$ can be written as $X'1/n$, and applying the result in the previous problem ,
show that $V(\hat{Y})$ at the point $(\bar{X}_1 \quad \overline{X_2} \quad \overline{X_3} \ldots)'$ is $\frac{\sigma^2}{n}$.

53. Verify the following properties of the residual vector $\hat{\varepsilon}$

(i) $E(\hat{\varepsilon}) = 0$
(ii) $cov(\hat{\varepsilon}) = \sigma^2(I - H)$
(iii) $cov(\hat{\varepsilon}, y) = \sigma^2(I - H)$
(iv) $cov(\widehat{\varepsilon}, \hat{y}) = O$
(v) $(\hat{\varepsilon}' y) = y'(I - H)y$
(vi) $(\hat{\varepsilon}' \hat{y}) = 0$
(vii) $(\hat{\varepsilon}' X) = 0'$

STAT C-401
STATISTICAL INFERENCE

1. (a) Let $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ be the order statistics of a random sample of size 5 from the uniform distribution $U(0,\theta)$. Show that $2Y_3$ is an unbiased statistic for θ. Determine the joint probability density function of Y_3 and the sufficient statistic Y_5 for θ. Find the conditional expectation $E(2Y_3 | Y_5)$ and compare its variance with that of $2Y_3$.

(b) State and prove Factorisation theorem on sufficiency. (5, 5)

2. (a) Define completeness of a statistic. Show that the family of $N(\theta,\sigma^2)$ distributions, where σ^2 is known, is complete. Hence obtain MVU estimator of θ.

(b) Show that the most general form of the distribution for which the sample arithmetic mean is the m. l. estimator of θ has the p.d.f. :

$$f(x, \theta) = \exp[(x-\theta) A'(\theta) + A(\theta) + B(x)],$$

where $A(\theta)$ and $B(x)$ are arbitrary functions of θ and x respectively. (5, 5)

3. (a) State optimum properties of maximum likelihood estimators.

(b) Show that in sampling from the distribution with p.d.f. :

$$f(x, \theta) = \theta e^{-\theta x}, \quad 0 < x < \infty.$$

$\frac{1}{\bar{x}}$ is the m.l. estimator of the parameter θ. Also show that $E\left[\frac{n-1}{n\bar{x}}\right] = \theta$ and compare the variance of the m.l. estimator with that of variance of unbiased estimator. (5, 5)

4. (a) State and prove Rao-Blackwell theorem and explain its significance in point estimation.

(b) State & prove invariance property of Consistent estimator. Hence obtain a Consistent estimator of $\theta^2 + \theta - \sqrt{\theta}$, w[illegible] x follows poisson Distn with parameter θ. (5, 5)

5. (a) Describe the procedure of obtaining estimators by the method of minimum Chi-square.

(b) A random variable X takes the values 0, 1, 2, with respective probabilities

$$\frac{\theta}{4N}+\frac{1}{2}\left(1-\frac{\theta}{N}\right), \quad \frac{\theta}{2N}+\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right), \quad \frac{\theta}{4N}+\frac{1-\alpha}{2}\left(1-\frac{\theta}{N}\right),$$

where N is a known number and α and θ are unknown parameters. Estimate θ and α by the method of moments based on a random sample of size n. (5, 5) selecting *three* from each section.

Section I

1. (*a*) If

$$X_1, X_2, \ldots\ldots, X_n$$

is a random sample of size n from $N(\mu, \sigma^2)$, where μ is known and if :

$$T = \sum_{i=1}^{n} |X_i - \mu|.$$

Examine, if T is unbiased for σ. If not, obtain an unbiased estimator of σ.

P.T.O.

(*b*) If

$$X_1, X_2, \ldots\ldots, X_n$$

is a random sample from

$$f(x : \theta) = \frac{1}{\theta}; \quad 0 \leq x < \infty$$

show that $\left(\frac{n+1}{n}\right)Y_n$ and $2\bar{X}$ are consistent and unbiased estimators for θ. Further find the efficiency of $2\bar{X}$ relative to $\left(\frac{n+1}{n}\right)Y_n$, where

$$Y_n = \max\{X_1, X_2, \ldots\ldots, X_n\}.$$

(*c*) Does there exist MVB estimator for the parameter θ in case of sample of size *n* is drawn from

$$f(x, \theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}; \quad 0 < x < \infty \ ?$$

If so, find its variance. 4,5,3½

2. (*a*) Show that MVU estimator is unique.

(*b*) If X_1, X_2 is a random sample of size 2 from a distribution having pdf

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

zero elsewhere, find the joint pdf of a statistic

$$Y_1 = X_1 + X_2$$

for θ and $Y_2 = X_2$. Show that Y_2 is an unbiased estimator of θ with variance θ^2. Find

$$E[Y_2 | y_1] = \psi(y_1)$$

and the variance of $\psi(Y_1)$. Interpret the result. 5,7½

3. (*a*) If X has a Poisson distribution and the prior distribution of its parameter λ is a Gamma distribution with the

parameters α and β, show that :

(*i*) The posterior distribution of λ given $X = x$ is a gamma distribution with the parameters

$$(\alpha + x) \text{ and } (\beta + 1)$$

(*ii*) The mean of the posterior distribution of λ is

$$\mu = \frac{(\alpha + x)}{\beta + 1}.$$

(*b*) If

$$X_1, X_2, \ldots\ldots, X_n$$

is a random sample from

$$f(x, \theta) = \frac{\theta^{k+1}}{\Gamma(k+1)} e^{-x\theta} x^k; \quad x > 0, \theta > 0$$

where k is known constant. Obtain MLE of θ. Show that it is consistent but not unbiased for θ. 5,7½

4. (*a*) Describe method of minimum chi-square. When do you use modified minimum chi-square ?

(*b*) If

$$X_1, X_2, \ldots\ldots, X_n$$

is a random sample from

$$f(x, \theta) = \theta(1-\theta)^{x-1}, x = 1, 2, \ldots\ldots$$

estimate θ by the :

(*i*) method of moments

(*ii*) method of MLE. 6½,6

Section II

5. (*a*) Define MPCR. Show that every MPCR is necessarily unbiased.

(*b*) If

$$X_1, X_2, \ldots\ldots, X_n$$

is a random sample of size n from $N(\mu, \sigma_i^2)$ where σ_i^2 is known. Develop UMP test for testing simple

$$H_0 : \mu = \mu_0$$

against one sided alternatives. 5½,7

6. (a) Describe LR test for testing

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta \neq \theta_0$$

when a sample of size n is drawn from $N(\theta, \sigma^2)$ where σ^2 is known.

(b) State Neyman Pearson lemma and explain its significance.

Let X_1, X_2 be a random sample from

$$f(x, \theta) = \theta x^{\theta-1} ; \quad 0 < x < 1, \theta > 0$$

It is given that the critical region for testing

$$H_0 : \theta = 1 \text{ against } H_1 : \theta = 2 \text{ is}$$

$$C = \{(X_1, X_2) : x_1 x_2 \geq \frac{3}{4}\}$$

Find level of significance and power of the test. 7,5½

7. (*a*) Obtain $100(1 - \alpha)\%$ confidence interval for the Binomial Proportion p.

(*b*) Develop a general method for constructing confidence intervals. Consider a random sample of size n from the exponential distribution with pdf

$$f(x, \theta) = e^{-(x-\theta)}; \quad \theta \leq x < \infty, \; -\infty < \theta < \infty.$$

Show that :

$$P\left\{X_1 + \frac{1}{n}\log \alpha \leq \theta \leq X_1\right\} = 1 - \alpha$$

where symbols have their usual meanings. 6,6½

8. Describe any *three* of the following : 4,4,4½

(*a*) Optimum properties of ML estimators

(*b*) LR test and its properties

(*c*) Invariance property of consistent estimator

(*d*) Method of moments.

SECTION I

(a) Explain the concepts of consistency and unbiasedness. Give an example of an estimator

(i) which is consistent but not unbiased.

(ii) which is unbiased but not consistent.

(b) State and prove invariance property of consistent estimators. Hence obtain a consistent estimator of $\theta^2+\theta-\sqrt{\theta}$, when X follows Poisson distribution with parameter θ. (6,6½)

(a) Define MVB and MVU estimators. Obtain MVB estimator of θ in case of a random sample of size n from $N(\mu, \theta)$, where μ is known.

(b) Let T_0 be MVU estimator and T_1 be an unbiased estimator with efficiency e. Show that no unbiased linear combination of T_0 and T_1 can be a MVU estimator. (7½,5)

(a) State and prove Cramer-Rao inequality and explain its significance.

(b) Obtain an unbiased estimator for θ^2 in case of a random sample of size drawn from the binomial distribution with p.m.f. :

$$p(x)=\binom{m}{x}\theta^x(1-\theta)^{m-x};\quad x=0, 1, 2, \text{------}, m.$$

(7½,

P.T.

SECTION – II

4. (a) Let $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ be the order statistics of a random sample of size 5 from the uniform distribution $U(0,\theta)$. Show that $2Y_3$ is an unbiased statistic for θ. Determine the joint probability density function of Y_3 and the sufficient statistic Y_5 for θ. Find the conditional expectation $E(2Y_3 \mid Y_5)$ and compare its variance with that of $2Y_3$.

(b) State and prove Factorisation theorem on sufficiency. (6,6½)

5. (a) Define completeness of a statistic. Show that the family of $N(\theta,\sigma^2)$ distributions, where σ^2 is known, is complete. Hence obtain MVU estimator of θ.

(b) Show that the most general form of the distribution for which the sample arithmetic mean is the m. l. estimator of θ has the p.d.f. :

$$f(x, \theta) = \exp[(x-\theta) A'(\theta) + A(\theta) + B(x)],$$

where $A(\theta)$ and $B(x)$ are arbitrary functions of θ and x respectively. (6½,6)

6. (a) State optimum properties of maximum likelihood estimators.

(b) Show that in sampling from the distribution with p.d.f. :

$$f(x, \theta) = \theta e^{-\theta x}, \quad 0 < x < \infty.$$

$\frac{1}{\bar{x}}$ is the m.l. estimator of the parameter θ. Also show that $E\left[\frac{n-1}{n\bar{x}}\right] = \theta$ and compare the variance of the m.l. estimator with that of variance of unbiased estimator. (6½,6)

7. (a) State and prove Rao-Blackwell theorem and explain its significance in point estimation.

(b) Obtain $100(1-\alpha)\%$ confidence interval for difference of binomial proportions when $X_1, X_2, \text{------}, X_m$ is a random sample from $B(m, p_1)$ and $Y_1, Y_2, \text{------}, Y_n$ is a random sample from $B(n, p_2)$, for large m and n. (6½,6)

8. (a) Describe the procedure of obtaining estimators by the method of minimum Chi-square.

(b) A random variable X takes the values 0, 1, 2, with respective probabilities

$$\frac{\theta}{4N}+\frac{1}{2}\left(1-\frac{\theta}{N}\right),\ \frac{\theta}{2N}+\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right),\ \frac{\theta}{4N}+\frac{1-\alpha}{2}\left(1-\frac{\theta}{N}\right),$$

where N is a known number and α and θ are unknown parameters. Estimate θ and α by the method of moments based on a random sample of size n. (6½,6)

(600)

GENERIC ELECTIVE:IV
APPLIED STATISTICS

Applied Statistics (GE-IV)

1. Index numbers are economic barometers. Elucidate.
2. Laspeyre's price index tends to overestimate price changes. Substantiate this statement.
3. Paasche's price index tends to underestimate price changes. Substantiate this statement..
4. State the important steps in the construction of cost of living index number
5. Define Pearl's vital index. How is it related to CBR and CDR?
6. SDR's are not of much utility for overall comparison of mortality conditions prevailing in two regions. Substantiate this statement with the help of an example.
7. Distinguish between Short term and long term fluctuations.
8. Show that $m_x = 2q_x/(2 - q_x)$, symbols having usual meaning.
9. How does NRR indicate the growth of population?
10. Distinguish between additive and multiplicative models of a time series analysis.
11. Name the characteristic movement of time series with which you will mainly associate the following.
    a. Fall in production of rice due to floods
    b. An era of recession
    c. Increase in literacy rate in a developing country
    d. Increase in sale of umbrella during rainy season
    e. Decrease in death rate in a developing country
    f. New launches & Phase out of gadgets from market
12. What is quality, discuss various dimensions of quality?
13. Explain process control in respect to statistical quality control?
14. What do you mean by control charts for variables?
15. What are different criteria for the choice of base period? Discuss different types of averages used in the construction of index number with relative merits and demerits.
16. Describe aggregate expenditure method to calculate cost of living index. What are its different uses?
17. What is an index number? Describe the problems that are involved in the construction of an index number of prices.
18. What is a time series? Indicate the objectives of its analysis. Explain its importance with respect to economics and business studies.
19. Discuss different components of time series with illustrations.
20. Describe the method of 'Ratio to trend' for measuring the seasonal indices in time series data, stating clearly the assumptions made in this method.
21. Define age specific death rate and discuss its merits over CDR.

22. Explain the purpose for standardizing death rates. Describe direct method of standardized death rate.

23. Define a life table. What are different assumptions used in the construction of the life table. Define central mortality rate and instantaneous force of mortality.

24. Distinguish between rates and ratios. Define GFR and discuss its usefulness over CBR.

25. What are different mathematical tests to measure the formula error in the construction of index number. Describe any two tests.

26. Describe the method of weighted relatives for the construction of cost of living index. State three uses of cost of living index number.

27. What is meant by process control in industrial statistics? Explain how a control chart helps to control the quality of a manufactured product. Discuss major parts of control charts.

28. Discuss the interpretation of X-bar and R-charts.

29. What do understand by control chart for a fraction defective? Explain its construction. Give the theoretical distribution on which the control limits are based.

30. Distinguish between defect and defective. Give some examples of defects for which c-chart is applicable. How do you calculate control limits for a c-chart? Discuss the assumptions and approximations involved in the calculations.