INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

QUESTION PAPERS - CONTENTS

| Sl. No. | Date | Examination Number | Subject |
|---|---|---|---|
| **Periodical Examinations:** | | | |
| 1. | 18. 9.67 | 31 | Statistics-5: Statistical Methods Theory and Practical |
| 2. | 25. 9.67 | 32A | Statistics-6: Sample Surveys |
| 3. | 25. 9.67 | 32B | Statistics-8: Demography |
| 4. | 23.10.67 | 34 | Statistics-7: Economics and Econometrics |
| 5. | 30.10.67 | 33 | Statistics-4: Probability and Inference |
| **Mid-year Examinations:** | | | |
| 6. | 18.12.67 | 87 | Statistics-4: Probability |
| 7. | 19.12.67 | 88 | Statistics-4: Inference |
| 8. | 21.12.67 | 89 | Statistics-5: Statistical Methods Theory |
| 9. | 22.12.67 | 90 | Statistics-5: Statistical Methods Practical |
| 10. | 25.12.67 | 91 | Statistics-6: Sampling Theory |
| 11. | 26.12.67 | 92 | Statistics-6: Sampling Practical |
| 12. | 27.12.67 | 93 | Statistics-7: Econometrics Theory and Practical |
| 13. | 28.12.67 | 94 | Statistics-7: Planning Techniques |
| 14. | 29.12.67 | 95 | Statistics-8: Demography Theory and Practical |
| 15. | 30.12.67 | 96 | Statistics-8: Educational Statistics Theory and Practical |
| **Periodical Examinations:** | | | |
| 16. | 26. 2.68 | 148 | Statistics-5: Statistical Methods Theory and Practical |
| 17. | 4. 3.68 | 149 | Statistics-6: Design of Experiments Theory and Practical |
| 18. | 11. 3.68 | 150 | Statistics-4: Probability and Inference |
| 19. | 25. 3.68 | 152 | Statistics-7: Statistical Quality Control Theory and Practical |
| 20. | 1. 4.68 | 153 | Statistics-8: Genetics |
| 21. | 8. 4.68 | 151 | Statistics-7: Econometrics Theory and Practical |
| **Annual Examinations:** | | | |
| 22. | 20. 5.68 | 208 | Statistics-4: Probability |
| 23. | 22. 5.68 | 209 | Statistics-4: Inference |
| 24. | 24. 5.68 | 210 | Statistics-5: Statistical Methods Theory and Practical |
| 25. | 25. 5.68 | 211 | Statistics-6: Design of Experiments Theory and Practical |
| 26. | 28. 5.68 | 212 | Statistics-7: Econometrics Theory and Practical |
| 27. | 30. 5.68 | 213 | Statistics-7: Industrial Statistics Theory and Practical |
| 28. | 31. 5.68 | 214 | Statistics-8: Genetics Theory and Practical |

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1967-68

31

PERIODICAL EXAMINATION

Statistics-5: Statistical Methods (Theory and Practical)

Date: 18.9.67. Maximum marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

1.(a) Define a generalised inverse of a matrix and state some of its applications. [5]

(b) If $B$ is a g-inverse of $A'A$, show that $ABA'$ is symmetric, idempotent of the same rank as that of $A$. [5]

(c) Let $A : m \times n$ and $B : n \times m$ be given matrices such that

(i) $H = AB$ is idempotent and (ii) rank of $H$ = rank of $A$, then show that $B = A^-$. Prove the converse of this result.

[Hint: You may use the Frobenius theorem on ranks, namely,
rank $(PQR) \geq$ rank $(PQ)$ + rank $(QR)$ - rank $(Q)$]. [5]

2. Let $f(x,y,\rho)$ be a bivariate normal density function with zero means and unit variances. Let $|\rho| < 1$. Then show that

$$f(x,y,\rho) = \sum_{j=0}^{\infty} \frac{\rho^j}{j!} H_j(x)H_j(y)\, \phi(x)\phi(y)$$

where $\phi(x)$ is density function of a standard normal variate and $H_j(x)$ is a Hermite polynomial of degree $j$.

Show that

$$\int_h^{\infty}\int_k^{\infty} f(x,y,\rho)\,dx\,dy = \int_0^{\rho} f(h,k,\rho)\,d\rho + \left(\int_h^{\infty}\phi(x)dx\right)\left(\int_k^{\infty}\phi(y)dy\right).$$

Find the value of $P(x \geq 0, \; y \geq 0)$. Also prove that

$$\int_h^{\infty}\int_k^{\infty} f(x,y,\rho)\,dx\,dy \geq \left(\int_h^{\infty}\phi(x)dx\right)\left(\int_k^{\infty}\phi(y)dy\right) \quad \text{if} \quad \rho \geq 0.$$ [15]

3. Explain clearly the idea of singular multivariate normal distribution. Show that any linear function of normal variates is normally distributed. Is the converse true? Give reasons in your favour.

Moreover, show that linear functions of normal variates are jointly normal. Hence or otherwise, obtain the marginal distribution of a subset of $X' = (x_1\; x_2\; \ldots\; x_p)$ which is normal. [15]

4.(a) If $P$ is an $m \times n$ matrix such that

$$P = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where $A$ is a square matrix and rank of $A$ = rank of

GO ON TO THE NEXT PAGE

then show that

$$Q = \begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

is a reflexive inverse of $P$, the order of $Q$ being $n \times m$. [5]

4.(b) If $A$ is a reflexive inverse of $B$, show that rank of $A$ = rank of $B$. [5]

5. Let $y_i$, $i = 1,2,\ldots, 8$ be independent stochastic variates having a common variance $\sigma^2$ and expectations given by

$E(y_1) = \theta_1+\theta_5$, $E(y_2) = \theta_2+\theta_3$, $E(y_3) = \theta_3+\theta_6$, $E(y_4)=\theta_4+\theta_6$,

$E(y_5) = \theta_1+\theta_7$, $E(y_6) = \theta_3+\theta_7$, $E(y_7) = \theta_2+\theta_8$ and $E(y_8)=\theta_4+\theta_8$.

Show that $\theta_1 - \theta_2$ is estimable and find at least three unbiased estimates of $\theta_1 - \theta_2$. Find also the best linear unbiased estimate (BLUE) and its variance. Establish further that $\theta_1 + \theta_2$ is not estimable. Obtain an unbiased estimate of $\sigma^2$. [14]

6. The following is the inverse of the corrected sums of squares and products of the four variables $X_1, X_2, X_3$ and $X_4$ based on 32 sets of observations.

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| $X_1$ | .00200037 | .00027291 | .00022329 | .00000550 |
| $X_2$ | | .02738289 | .00183149 | -.00009279 |
| $X_3$ | | | .00017112 | -.00001167 |
| $X_4$ | | | | .00000832 |

Also, the corrected sum of products between $Y$ and the X's and also the means are as follows:

$S_{YX_1} = 461.415$ $\quad S_{YX_2} = 334.456$ $\quad S_{YX_3} = -3931.050$

$S_{YX_4} = 16497.822$

$S_{YY} = 3564.070$ $\quad \bar{X}_1 = 39.250$, $\bar{X}_2 = 4.181$, $\bar{X}_3 = 241.500$,

$\bar{X}_4 = 332.094$, $\bar{Y} = 19.659$.

(a) Obtain the lease squares estimates of the parameters of the regression equation

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4.$$ [5]

(b) Estimate the variances of the estimates of parameters. [4]

(c) Test the hypotheses

i) $\beta_3 + \beta_4 = 0$ $\qquad$ ii) $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 2$ [12]

(d) Given $\beta_3 + \beta_4 = 0$ and $\beta_1 + \beta_2 = 1$, estimate the parameters. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

132A

PERIODICAL EXAMINATIONS

Statistics-6: Sample Surveys Theory

Date 25.9.67 Maximum marks: 50 Time: $1\frac{1}{2}$ hours

Note: Answer Q.1 and any two of the remaining questions.

1. Discuss briefly the advantages of a sample survey over the complete enumeration. Also state under what circumstances you would recommend complete enumeration in preference to a sample survey. [14]

2.(a) Show that for simple random sampling without replacement

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2\right] = \sigma_y^2 - V(\bar{y})$$

where $\sigma_y^2$ = Population variance and

$\bar{y}$ = sample mean.

(b) Suggest an estimator based on a simple random sample of n units selected with replacement for estimating the population proportion P. Obtain an unbiased variance estimator for this estimator. [9+9]=[18]

3.(a) Explain the method of Linear and Circular systematic sampling. What are the advantages, if any, of the latter over the former.

(b) Show that the sample mean based on a circular systematic sample, with a random start chosen from 1 to N (N being the total number of units in the population) is unbiased for the population mean for any sampling interval.

(c) Compare the efficiency of circular systematic sampling with that of simple random sampling without replacement for estimating the population mean. [5+6+7]=[18]

4. Write short-notes on:

(a) Estimation based on distinct units for simple random sampling with replacement.

(b) Estimation of variance in systematic sampling.

(c) Centrally located systematic samples. [6+6+6]=[18]

-----------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

[323]

PERIODICAL EXAMINATIONS

Statistics-8: Demography (Theory and Practical)

Date: 25.9.67 Maximum marks: 50 Time: $1\frac{1}{2}$ hours

1. Establish the following relationship between life table functions:

   (a) $e^{o}_{x} = e^{o}_{x:\overline{n}|} + {}_{n}p_x \, e^{o}_{x+n}$

   (b) $${}_{n}q_x = \frac{2n. \; {}_{n}m_x}{2 + n. \; {}_{n}m_x}$$

   (c) $\mu_x = \frac{1}{2} [ \; \mathrm{colog}_e \, p_x + \mathrm{colog}_e \, p_{n-1} ].$ [3 X3]=[9]

2. Given that the complete expectation of life at age 30 and 31 for a particular group are respectively 21.39 and 20.91 years and that the number living at age 30 is 41, 176 find the number that attains the age 31. [8]

3. If $l_x$ (the survivors at age $x$ in a life table) between the ages 1 and 11 be such that it is $l_1 \, [1 - \frac{1}{9} \log_{10} x]$ find

   (a) the complete expectation of life at age 1 during the next 10 years,

   (b) the average age at death of those who die between ages 1 and 11. [9+9]=[18]

4. Write short notes on any <u>three</u> of the following:

   a) 'defacto' and 'dejure' population enumeration

   b) uses of vital statistics

   c) economically active population

   d) development of fertility research in India. [3 X5]=[15]

-------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1967-68

PERIODICAL EXAMINATIONS

Statistics-7: Economics and Econometrics

Date: 23.10.67 Maximum marks: 100 Time: 3 hours

Note: Answer Group A and Group B in separate answerscripts.

## Group A

Answer any two questions out of the following.

1.a) Analyse the procedure of construction of inter-industry tables and show how they can be used to demonstrate the equivalence of the three concepts of national income.

b) Indicate in this connexion the relation between the methods of valuation of interindustry flows and the treatment of distributive services sector. [17+8]=[25]

2.a) Find out the total labour requirements in an economy, given the final bill of goods on the basis of the Leontief input-output system.

b) Show that under long run competitive equilibrium, commodities exchange in ratios given by their labour component in the Leontief model. [12+13]=[25]

3.a) Given labour supply and capacity restrictions in different sections of the economy, how would you obtain the net output possibility schedule in an input-output model.

b) Show how the problem of choice on the demand side can be solved by introducing consumers' valuation of commodities. [13+12]=[25]

## Group B

4. Discuss the scope of Econometrics. [10]

5. State the assumptions underlying the classical least squares and explain briefly their relevance in Economics. [20]

6. What is the problem of auto correlation? State and prove the generalized least squares theorem. If the disturbances in the linear model follow a first order auto regressive process, with $\rho = 1$, show that a simple transformation of the original variables will enable the application of classical least squares. [20]

------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

33

PERIODICAL EXAMINATIONS

Statistics-4: Probability and Statistical Inference

Date: 30.10.67 | Maximum marks: 100 | Time: 3 hours

Note: Answer groups A and B in separate answer-scripts.

## Group A

1. In the joint distribution of X and Y, the whole probability is situated inside the triangle whose vertices are (0,0), (2,1) and (2,0). The joint density function is K.xy, where K is constant. Determine

   (a) the value of K, [5]

   (b) the distribution of X, [7]

   (c) the conditional dictribution of Y given that $X = \alpha$, $0 < \alpha < 2$, [9]

   (d) the distribution of XY, [11]

   (e) E(X Y). [7]

2. X takes the values 1, 2 with probabilities $\frac{1}{3}$, $\frac{2}{3}$ respectively. Y takes the values -1, +1 with probabilities $\frac{1}{2}$, $\frac{1}{2}$. Correlation coefficient of X and Y is $+\frac{1}{10}$. Find $E(X^2Y)$. [11]

## Group B

3.a) Prove the sufficiency part of Neyman-Pearson's lemma. [9]

b) Let $X_1, \ldots, X_n$ be i.i.d. with

$$f(x_1, \theta) = \frac{1}{2\theta} \quad \text{if} \quad -\theta \leq x \leq \theta,\ \theta > 0$$
$$= 0 \quad \text{elsewhere.}$$

Find the M.P. test of $H_0(\theta = 2)$ against $H_1(\theta = 3)$ of size $\alpha$. [8]

c) Is the test in (b) M.P. against $\theta = 4$? Against $\theta = 1$? [8]

EITHER

4.a) Define sufficiency and minimal sufficiency. [4]

b) Prove Neyman's factorisation theorem for discrete random variables. [10]

c) Prove that if T is a real valued boundedly complete sufficient statistic then T is minimal sufficient. [6]

d) Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, 1)$ where $\theta = 0, \pm 1, \pm 2 \ldots$ Find the minimal sufficient statistic. [5]

OR

5.a) State and prove the Rao-Blackwell theorem. [10]

b) Give an example of (i) a minimum variance unbiased estimator not attaining the Cramer-Rao lower bound; (ii) a non-estimable parametric function; (iii) an estimable parametric function which does not have a minimum variance unbiased estimator. [15]

-------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

87

MID-YEAR EXAMINATIONS

Statistics-4: Probability

Date: 18.12.67 Maximum marks: 100 Time: 3 hours

1.a) $X_1, X_2, \ldots$ are random variables taking the values $1, 2, \ldots, N$ only. When do we say that they form a Markov chain? [5]

b) $X_1, X_2, X_3$ are random variables, each taking the values 1 and 2 only. Their joint distribution is as follows:

$$P(X_1 = 1, X_2 = 2, X_3 = 1) = \frac{1}{12},$$

$$P(X_1 = 2, X_2 = 1, X_3 = 1) = \frac{1}{6},$$

$$P(X_1 = 1, X_2 = 2, X_3 = 2) = \frac{1}{4},$$

$$P(X_1 = 2, X_2 = 1, X_3 = 2) = \frac{1}{2}.$$ [10]

Do $X_1, X_2, X_3$ form a Markov chain? Give reasons.

2. $X_1, X_2, \ldots$ form a Markov chain and each random variable takes the values $1, 2, \ldots, N$ only.

$$p_i = P(X_1 = i), \; {}_m p_{ij} = P(X_{m+1} = j \mid X_m = i)$$

All the Markov chains considered in this question will satisfy the following condition:

For arbitrary positive integral m and arbitrary values $i_1, \ldots, i_m$ from among $1, 2, \ldots, N$, $P(X_1 = i_1, \ldots, X_m = i_m) > 0$.

a) Prove the chain rule

$$P(X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n)$$

$$= p_{i_1} \cdot {}_1p_{i_1 i_2} \cdots {}_{n-1}p_{i_{n-1}, i_n}.$$

b) Suppose it so happens that ${}_m p_{ij}$ does not depend on $i$; that is, ${}_m p_{ij} = {}_m q_j$ for $i = 1, 2, \ldots, N$, and arbitrary $m$ and $j$.

What can you say about the random variables? What is the distribution of the random variable $X_\lambda$ ($\lambda$ is any positive integer $\geq 2$)?

c) Prove that $P(X_5 = i_5 \mid X_3 = i_3, X_2 = i_2, X_1 = i_1)$
$= P(X_5 = i_5 \mid X_3 = i_3)$.

d) Prove that $P(X_1 = i_1 \mid X_2 = i_2, X_3 = i_3, X_4 = i_4)$
$= P(X_1 = i_1 \mid X_2 = i_2)$.

e) ${}_mP$ is the matrix in whose i-th row and j-th column we have ${}_m p_{ij}$.

Prove that the element in the i-th row and j-th column of ${}_1P \cdot {}_2P$ is exactly $P(X_3 = j \mid X_1 = i)$.

In the proof, do you make use of the Markovian property?

[5 × 8] = [40]

3. $X_1, X_2, X_3$ form a Markov chain; each random variable takes only the values 1 and 2.

$$_1p_{ij} = {}_2p_{ij} \quad \text{for all } i \text{ and } j.$$

The joint distribution of $X_1$ and $X_2$ is as follows:

$P(X_1 = 1, X_2 = 1) = \frac{3}{10}$, $P(X_1 = 1, X_2 = 2) = \frac{1}{10}$,

$P(X_1 = 2, X_2 = 1) = \frac{2}{5}$, $P(X_1 = 2, X_2 = 2) = \frac{1}{5}$.

Obtain the distribution of the random variable $X_3$. [15]

4. X takes the values 1, 2, 3 with probabilities 1/4, 1/4, 1/2, respectively, Y tables the values 1, 2 with probabilities 2/3, 1/3 respectively. Determine that joint distribution of X and Y which will maximize the correlation coefficient of X and Y. What is this maximum value? [15]

5. In the joint distribution of X and Y, probability is uniformly distributed over the triangle whose vertices are (0, 0), (2, 0) and (0, 1). Find the fr. f. of the random variable (X - Y). [15]

-------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

MID-YEAR EXAMINATIONS

Statistics-4: Inference

188

Date: 19.12.67 Maximum marks: 100 Time: 3 hours

Answer Q.7 and two other questions from each group.

## Group A

1. Find the UMP unbiased test of size $\alpha$ of $H_0(\theta = \theta_0)$ against $H_1(\theta \neq \theta_0)$ in the following cases.

   a) $X$ has density $f_\theta(x) = \theta e^{-\theta x}$ $x > 0$. [11]
   $\theta_0 = 1$

   b) $X_1, \ldots, X_n$ are i.i.d. with common density
   $f_\theta(x_i) = \theta^{x_i}(1-\theta)^{1-x_i}$ $x_i = 0, 1$, $\theta_0 = 1/2$. [11]

2.a) Let $X_1, \ldots, X_n$ be i.i.d. with common density
$f_\theta(x) = K(\theta)e^{\theta x}\,\gamma(x)$. Find the most powerful test, of size $\alpha$, of $H_0(\theta = \theta_0)$ against $H_1(\theta > \theta_0)$ and show that its power is a monotonic function of $\theta$.

b) Let $X$ have density $f_\theta(x) = \frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2}$ , $-\infty < x < \infty$.
Find the locally most powerful tests of $H_0(\theta = 0)$ against $H_1(\theta > 0)$ and against $H_1(\theta < 0)$. Show that the power of the former test -> 0 as $\theta \to \infty$. [11]

3.a) Briefly describe how you may use a least favourable distribution to construct the most powerful test of a composite null hypothesis against a simple alternative hypothesis. Apply your method to an example where the least favourable distribution is not degenerate. [15]

b) $X_1, \ldots, X_n$ are i.i.d. $N(\theta, 1)$. You have to test $H_0(\theta = 0)$ against $H_1(\theta = 1)$. Assuming $\theta$ has a priori probability 1/2 of being 0 or 1, find the Bayes test. [7]

## Group B

4.a) Prove the Cramer-Rao inequality, stating your assumptions clearly. [11]

b) Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, 1)$. Show that only for $\theta$ the Cramer-Rao lower bound is attained. Characterise the parametric functions for which one of the Bhattacharya lower bounds is attained. [11]

5.a) If $X_1, \ldots, X_n$ are i.i.d. with common density $f_\theta(x_i) = K(\theta) e^{\theta x_i} \gamma(x_i), -\infty < \theta < \infty$, then show that $\frac{1}{n} \Sigma X_i$ is an admissible estimator for $h(\theta) = E_\theta(X_i)$. [17]

b) Hence show that if $X_1, \ldots, X_n$ are i.i.d. and

$$f_\lambda(x_i) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \quad x_i = 0, 1, 2, \ldots, 0 < \lambda < \infty,$$

then $\frac{\Sigma X_i}{n}$ is an admissible estimator for $\lambda$. [5]

6.a) If $X_1, \ldots, X_n$ are i.i.d. with common density $f_\theta(x_i) =$

$$= \frac{1}{\theta^m \Gamma m} e^{-x_i/\theta} x_i^{m-1}, \quad x_i > 0, \quad \theta > 0,$$

find the improper Bayes estimator for the prior $p(\theta) = \frac{1}{\theta^3}$ $\theta > 0$. Show that it has smaller risk than the best unbiased estimator of $\theta$. [11]

b) Show that if a proper Bayes estimator is unbiased then its average risk is zero. [11]

7. Let $X$ have the density $f_\theta(x) = {}^nC_x \theta^x (1-\theta)^{n-x}$ $0 < \theta < 1$, $x = 0, 1, \ldots, n$. Show that $X$ is an admissible estimator of $\theta$ without using the Cramer-Rao inequality. [12]

-----

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

89

MID-YEAR EXAMINATIONS

Statistics-5: Statistical Methods Theory

Date: 21.12.67 Maximum marks: 100 Time: 3 hours

Note: Attempt any five questions. All questions carry equal marks.

1.a) Let $x'Ax$ and $x'Bx$ be distributed as Chi-squares and $x'Cx \geq 0$ for all $x$. Then show that if $x'Ax = x'Bx + x'Cx$ and $x \sim N(0,I)$, then $x'Cx$ and $x'Bx$ are independently distributed as Chi-squares. Generalise this result.

b) Let $x \sim N(0,I)$. Obtain necessary and sufficient condition for $x'Ax$ to be distributed as Chi-square. Is the same result true even when $x \sim N(\mu,I)$, $\mu \neq 0$? Give modifications if necessary.

2. Let $\Delta_p^2$ and $\Delta_q^2$ denote the Mahalanobis distances between two populations based on p characters and sub-set of q characters respectively. Obtain the test procedures for testing the following hypotheses:

i) $H_0(\Delta_p^2 = 0)$ against $H(\Delta_p^2 > 0)$

ii) $H_0(\Delta_p^2 = \Delta_q^2)$ against $H(\Delta_p^2 > \Delta_q^2)$ when $\Delta_q^2 \neq 0$.

iii) $H_0(\Delta_p^2 = 0)$ against $H(\Delta_p^2 > 0)$ when $\Delta_q^2 = 0$.

3.a) Let S be distributed as $W(p,n,I)$. If $S = TT'$ where T is a lower triangular matrix with positive diagonal elements, then find the distribution of the elements of T. Give interpretations of the elements of T. Hence or otherwise show that $a'Sa$ and

$$[a'S^{-1}a]^{-1}$$

are distributed as $\chi^2$ with n and n-p+1 degrees of freedom respectively if $a'a = 1$.

b) If $S_i$, $i = 1,2,\ldots, k$ are independent Wisharts $(p, n_i, I)$, then show that

$$\sum_{i=1}^{k} S_i$$

is distributed as Wishart $(p, n, I)$, $n = \sum_{i=1}^{k} n_i$.

4. Let $u_1,\ldots,u_k$ and $v_1,\ldots,v_m$ be two sets containing respectively k and m one-dimensional random variables. Their joint dispersion matrix is given by

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Some or all of $u_i$ may be a subset of the variables $v_i$.

a) Let $\sigma_i^2 = V(u_i) - \text{cov}(u_i, \Sigma b_j v_j)/V(\Sigma b_j v_j)$

where $b_1,\ldots, b_k$ are non-zero elements. Find the values of $b_1,\ldots, b_k$ which minimizes

$$\sum_{i=1}^{k} w_i^2 \sigma_i^2 .$$

GO ON TO THE NEXT PAGE

4.b) Show that the best $r$ linear functions of $v_1, \dots, v_m$ for predicting $u_1, \dots, u_k$ in the sense of minimizing $\Sigma \sigma_i^2$ correspond to first $r$ eigenvectors that arise out of the determinantal equation $| \Sigma_{21}\Sigma_{12} - \lambda \Sigma_{22}| = 0$. Interpret the results when $u_1, u_2, \dots, u_k$ is a subset of $v_1, \dots, v_m$.

5. State and prove a general theorem on the distribution of the Chisquare statistic of the goodness of fit. Give some of its applications.

6.a) Obtain an exact test for independence of two characters in a contingency table. Also, derive a large sample test procedure.

b) Given p multinomial populations. Obtain a large sample test for testing homogeneity of these populations. Suppose in one population, two cell frequencies are mixed up. Give the modifications in your test procedure.

7.a) Consider a sample of n values from an absolutely continuous distribution. Define the sample quantile of the p-th order and obtain its asymptotic distribution.

b) Let $m_r$ be the r-th central moment in a sample of size n. Find cov $(m_r, m_s)$ upto order $1/n$.

c) Let $\xi$ be distributed as

$$P(\xi = x) = a_x \theta^x / f(\theta) \quad \text{for} \quad x = 0, 1, 2, \dots$$

where $a_x \geq 0$, $0 < \theta < R$ and $f(\theta) = \Sigma a_x \theta^x$. If $a_x = 1/(x!)$. Find the estimate of $\theta$ based on n observations on $\xi$ and its standard error.

-----------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part II: 1967-68 [90]

MID-YEAR EXAMINATIONS

Statistics-5: Statistical Methods
Practical

Date: 22.12.67 Maximum marks: 100 Time: 3 hours

1. The following table gives the head length and head breadth of the first and second sons of 25 families.

| Head length first son $x_1$ | Head breadth first son $x_2$ | Head length second son $x_3$ | Head breadth second son $x_4$ |
|---|---|---|---|
| 191 | 155 | 179 | 152 |
| 195 | 149 | 201 | 152 |
| 181 | 148 | 185 | 149 |
| 183 | 153 | 188 | 149 |
| 176 | 144 | 171 | 142 |
| 208 | 157 | 192 | 152 |
| 189 | 150 | 190 | 149 |
| 197 | 159 | 189 | 152 |
| 188 | 152 | 197 | 159 |
| 192 | 150 | 187 | 151 |
| 179 | 158 | 186 | 148 |
| 183 | 147 | 174 | 147 |
| 174 | 150 | 185 | 152 |
| 190 | 159 | 195 | 157 |
| 188 | 151 | 187 | 158 |
| 163 | 137 | 161 | 130 |
| 195 | 155 | 183 | 158 |
| 186 | 153 | 173 | 148 |
| 181 | 145 | 182 | 146 |
| 175 | 140 | 165 | 137 |

a) Estimate the mean vector $\mu$ and the variance-covariance matrix $\Sigma$. [25]

b) Find the sample multiple correlation coefficient between $x_4$ and $(x_1, x_2, x_3)$ [8]

c) Test the hypothesis:

'The multiple correlation coefficient between $x_4$ and $(x_1, x_2, x_3)$ is zero'. [7]

d) Test the hypothesis (at 10 percent level of significance)

'The mean vector of the population is

$\bar{X} = (185.0, 150.0, 185.0, 150.0)$'. [20]

2. Three objects are weighed six times in a chemical balance by putting some objects on the left pan and some on the right pan and balancing the pan by putting standard weights. Each weighing involves an error which is distributed as $N(0, \sigma^2)$, $\sigma^2$ being unknown. The results are as follows.

| objects on the left pan | objects on the right pan | standard weight in left pan | standard weight in right pan |
|---|---|---|---|
| A, B | C | 0.9 | - |
| A, C | B | - | 16.9 |
| [illegible] | B, C | 24.9 | - |
| [illegible] | - | - | 19.7 |
| - | B, C | 35.3 | - |
| A, C | - | - | 28.6 |

It is not known whether the balance is unbiased or not.

It was later known that the objects A and B resulted from the break-up of a fourth object D into two pieces and the weight of D was the same as the weight of C.

a) Write down the model and the observational equations. [8]

b) Find the best linear unbiased estimate of the bias of the balance and the actual weights of A and B. [15]

c) Find the standard errors of the estimates in (b). [7]

3. Records. [10]

------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

91

MID-YEAR EXAMINATIONS

Statistics-6: Sampling Theory

Date: 25.12.67. Maximum marks: 100 Time: 3 hours

Note: Answer question 5 and any three of the other questions

1. Suppose a sample of $n$ units is to be drawn from a finite population of N units using circular systematic sampling.

   i) State, giving reasons, whether you prefer the sampling interval $[N/n]$ or $[N/n] + 1$ for estimating the population mean $\bar{Y}$.

   ii) Can the interval be fixed arbitrarily, and if so, what is its effect on the unbiased nature of the estimator and on its variance, when the units have been arranged in ascending or descending order of a size-measure.

   iii) If, instead of using one fixed interval I, a sequence of prespecified intervals $\{I_j\}$, $j = 1, 2, \ldots, n-1$, is used, $I_j$ being the interval used for selecting the (j+1) unit, derive an unbiased estimator for $\bar{Y}$.

   iv) If the sampling interval is taken as the integer nearest to (N/n), find the condition under which there would not be any need to repeat or cross the random start to achieve the required sample size.

[5+8+7+5]=[25]

2.a) What are the advantages of stratified sampling?

b) In stratified sampling where $n_i$ units are selected using simple random sampling without replacement from the $N_i$ units in the i-th stratum, $i=1, 2, \ldots, k$, derive an unbiased estimator of the overall population proportion P of units possessing a particular characteristic. Also derive the sampling variance of the estimator of P.

c) Suppose the objective is to estimate the difference between the rates of incidence of a particular disease in two villages, one a model village having $N_1$ persons and the other a neighbouring village having $N_2$ persons. Let $C_1$ and $C_2$ be the average costs of medically examining a person in the two villages. Assuming the cost to be fixed at C', determine the optimum allocation of the total sample size to the two villages when simple random sampling with replacement is used in each village.

[5+10+10]=[25]

3.a) In two-stage sampling, show that the variance of an estimator t of $\theta$ can be expressed as

$$V(t) = V_1 E_{2/1}(t) + E_1 V_{2/1}(t),$$

where $E_1$ and $V_1$ are the unconditional expected value and variance over the first stage sampling and $E_{2/1}$ and $V_{2/1}$ are the conditional expected value and variance over the second stage for a given sample at the first stage.

3.b) For estimating the total number of cultivators (Y), a sample of n villages is selected with replacement from the N villages in the population with probability proportional to the current number of households ($M_i$), and from each sample village m households are selected circular systematically. The number of households in each of the nm sample households is determined. Suggest an unbiased estimator of Y and obtain an unbiased variance estimator for it.

[10+15]=[25]

4.a) If $\hat{Y}$ and $\hat{X}$ are unbiased estimators of Y and X based on any probability sampling design, derive the approximate expressions for the bias variance of the ratio estimator

$$\hat{R}(= \hat{Y} / \hat{X}),$$

stating clearly the assumptions involved.

b) What do you understand by ratio method of estimation?

c) Under what circumstances is the ratio estimator more efficient than the conventional unbiased estimator?

[15+5+5]=[25]

5. Write brief notes on any three of the following

i) cluster sampling
ii) Lahiri's method of pps selection
iii) regression method of estimation
iv) self-weighing design. [25]

-----------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

92

MID-YEAR EXAMINATIONS

Statistics-6: Sampling Practical

Date: 26.12.67 | Maximum marks: 100 | Time: 3 hours

Answer any three questions. Each question carries 28 marks and 16 marks are for practical record.

1. A population of 112 villages has been divided into 3 strata having 51, 37 and 24 villages respectively on the basis of the type of available auxiliary information. We have a sample of 6 villages selected with simple random sampling without replacement from the first stratum, a sample of 5 villages selected with replacement with probability proportional to size measure x (geographical area, in acres) from the second stratum and two linear systematic samples (of four villages each) selected without replacement from the third stratum. For each selected village, the total area under wheat(y) is observed. The observed values and other relevant information are given in the following table.

| sample village | stratum 1 | stratum 2 | | stratum 3 (y) | |
|---|---|---|---|---|---|
| | y | x | y | sample 1 | sample 2 |
| 1 | 75 | 729 | 247 | 427 | 335 |
| 2 | 101 | 617 | 238 | 326 | 412 |
| 3 | 5 | 870 | 359 | 481 | 503 |
| 4 | 78 | 305 | 129 | 445 | 348 |
| 5 | 78 | 569 | 223 | - | - |
| 6 | 45 | - | - | - | - |

Total of x in stratum 2 = 26,912 acres.

Estimate thd population total of y and obtain an estimate of its variance.

2. To estimate the total number of words (Y) in an English dictionary, 10 out of 26 alphabets were selected with replacement and with probability proportional to the number of pages devoted to an alphabet and for each selected alphabet, two pages were drawn with equal probability without replacement. The relevant sample data are given in the table given below.

i) Estimate unbiasedly Y and obtain an estimate of its relative standard error.

ii) Estimate also the efficiency of the above method of sampling compared to that of drawing 20 pages from the dictionary with equal probability and with replacement.

| sl. no. | sample alphabet | no. of pages devoted | no. of words in sample page | |
|---|---|---|---|---|
| | | | 1 | 2 |
| 1 | S | 131 | 34 | 27 |
| 2 | C | 97 | 27 | 26 |
| 3 | N | 21 | 44 | 38 |
| 4 | S | 131 | 24 | 29 |
| 5 | F | 43 | 25 | 32 |
| 6 | J | 7 | 42 | 48 |
| 7 | U | 18 | 24 | 21 |
| 8 | P | 85 | 53 | 24 |
| 9 | A | 49 | 47 | 55 |
| 10 | D | 54 | 38 | 57 |

(Total number of pages in the dictionary: 980).

3. For estimating the percentage of absentees in the 325 factories situated in a district, a sample of 20 factories was drawn with equal probability without replacement. Utilising the data given in the following table, estimate the percentage of absentees (R) and its relative standard error. Also obtain an approximate estimate of the bias of the ratio estimate.

| sr. no. of factory | no. of workers | no. of absentees | sr.no. of factory | no.of workers | no. of absentees |
|---|---|---|---|---|---|
| 1 | 95 | 9 | 11 | 148 | 16 |
| 2 | 79 | 7 | 12 | 89 | 4 |
| 3 | 30 | 3 | 13 | 57 | 5 |
| 4 | 45 | 2 | 14 | 132 | 13 |
| 5 | 28 | 3 | 15 | 47 | 4 |
| 6 | 142 | 8 | 16 | 43 | 9 |
| 7 | 125 | 9 | 17 | 116 | 12 |
| 8 | 81 | 10 | 18 | 65 | 8 |
| 9 | 43 | 6 | 19 | 103 | 9 |
| 10 | 53 | 2 | 20 | 52 | 8 |

4. For estimating the total agricultural population (Y) in a region, a sample of villages was selected from each stratum with probability proportional to previous census population with replacement, and a sample of households was selected from each sample village linear systematically. The sampling intervals used in the sample villages were so specified that the sampling design was self-weighting with 250 as the constant inflation factor for each sample household. Using the data given in the table given below, estimate Y unbiasedly and obtain an estimate of its relative standard error.

| stratum | no. of sample villages | total agricultural population in sample households of sample village | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 7 | 57 | 48 | 72 | 63 | 71 | 54 | 62 |
| 2 | 5 | 48 | 35 | 76 | 54 | 30 | - | - |
| 3 | 6 | 25 | 22 | 34 | 45 | 68 | 55 | - |

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

[93]

MID-YEAR EXAMINATIONS

Statistics-7: Econometrics Theory and Practical

Date: 27.12.67 Maximum marks: 100 Time: 3 hours

Note: Answer Questions 1 and 2 and any two of the rest.

1. Consider a single equation model

$Y = X\beta + \epsilon$ where $Y$ is a vector of observations on the regressand $y$,
$X$ is a matrix of observations on $k$ regressors,
$\beta$ is a vector of coefficients,
$\epsilon$ is a vector of disturbances.

Examine the following statements and indicate whether they are correct or not. If they are incorrect, specify the correct form:-

i) If the disturbance term is not serially independent, the classical least squares estimates of $\beta$ are biased but consistent.

ii) When the regressors are not independent of the disturbance term, the classical least squares estimates of $\beta$ are unbiased but have a low degree of precision.

iii) If some of the regressors are highly inter-correlated, the classical least squares estimates of $\beta$ are consistent.

iv) If the regressors include among them the regressand with a lag and if the disturbance term follows a normal distribution, the classical least squares estimates of $\beta$ are biased but have the large sample properties of consistency and efficiency. [15]

2. The following data give the output per acre (Y) labour input per acre ($x_1$) and capital investment ($x_2$) per acre, according to size classes. Estimate the relationship $Y = Ax_1^{\alpha} x_2^{\beta}$, making the necessary assumptions by the method of least squares.

Test the significance of the coefficients $\alpha$ and $\beta$ and find out the '$R^2$'. Briefly comment on the estimated regression equation.

| Size class (in acres) | $x_1$ labour (days) | $x_2$ capital (Rs.) | Y output (Rs.) |
|---|---|---|---|
| 0 - 5 | 43 | 684 | 107 |
| 5 - 10 | 47 | 681 | 131 |
| 10 - 15 | 28 | 412 | 68 |
| 15 - 20 | 21 | 512 | 56 |
| 20 - 25 | 22 | 309 | 55 |
| 25 - 30 | 21 | 271 | 57 |
| 30 - 50 | 21 | 329 | 63 |

[35]

3. What is the identification problem in simultaneous linear economic equation systems?

Examine the identifiability of the equations in the following model. What is the reduced form?

$$C_t = \alpha_1 + \alpha_2 Y_t + u_{1t},$$
$$I_t = \beta_1 + \beta_2 Y_t + \beta_3 I_{t-1} + u_{2t},$$
$$Y_t = C_t + I_t + Z_t,$$

where $C_t$ = consumption Expenditure in time period t,
$I_t$ = gross fixed capital formation in time period t,
$Y_t$ = gross domestic product in time period t,
$Z_t$ = Exogenous factor in time period t (gross investment in stocks). [25]

4. What is the problem of errors in variables in regression analysis? Suppose that

$$X = \chi + u$$
$$Y = \varphi + u$$
and $$\varphi = a + b\chi$$

where X and Y are the observed values, $\chi$ and $\varphi$ are true values and u and v are the errors of observation normally distributed with zero mean and $\sigma_u^2$, $\sigma_v^2$ as variances. Assuming further that $\sigma_u^2 / \sigma_v^2 = \lambda$, find out the maximum likelihood estimate of b and show that it is identical with the least squares estimate if $\sigma_u^2 = 0$. What are the properties of least squares estimates in this case? [25]

5. Write short notes on the following (any two):

1) Auto-correlation.
2) Multicollinearity.
3) Use of Instrumental variables when there are errors in observations.
4) Qualitative variables and Dummy variables. [25]

---------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

94

MID-YEAR EXAMINATIONS

Statistics-7: Planning Techniques

Date: 28.12.67 Maximum marks: 100 Time: 3 hours

Answer any four questions out of the following.

1.a) Define the following concepts: primal and dual of a linear programme, feasible and optimal solutions of a linear programme. [9]

b) Show that the solutions for quantities and prices of commodities in the Leontief static system can be interpreted as optimal solutions, if it is formulated in terms of linear programming. [16]

2. Use the basic assumptions of the Leontief dynamic system to derive the schedule of net-output possibilities in a one period programme, given the initial stocks. Define clearly in this connection the concept of efficiency locus and explain briefly its relevance to optimal plan. [25]

3. Show that the assumption of equality of supply and demand in the case of both flow and stock relations in the Leontief dynamic system may result in economically meaningless solutions. State and justify any procedure to make the system meaningful. [25]

4. Define the concept of balanced growth. Show that the Leontief dynamic system (with no excess capacity) has a maximal rate of balanced growth out of all balanced growth paths (with or without excess capacity).
(Rigorous proof is not required). [25]

5. Solve the following problem by simplex method:

Maximise $z = 60x_1 + 80x_2 + 90x_3 + 90x_4$

subject to

$$100x_1 + 100x_2 + 100x_3 + 100x_4 \leq 1500$$
$$7x_1 + 5x_2 + 3x_3 + 2x_4 \leq 100$$
$$3x_1 + 5x_2 + 10x_3 + 15x_4 \leq 100$$
$$x_1, x_2, x_3, x_4 \geq 0$$

Formulate the dual of the above problem and give its solutions. [25]

--------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68



MID-YEAR EXAMINATIONS

Statistics-8: Demography (Theory and Practical)

Date: 29.12.67 Maximum marks: 100 Time: 3 hours

1. Obtain the differential equation of the growth of population and derive the logistic stating clearly the conditions under which it is applicable. Describe the properties of such a growth. Discuss the law of growth of population with reference to India, given the following data:

| year | 1891 | 1901 | 1911 | 1921 | 1931 | 1941 | 1951 | 1961 |
|---|---|---|---|---|---|---|---|---|
| population in millions | 236 | 235 | 249 | 248 | 276 | 313 | 357 | 389 |

[10+5+5]=[25]

2.a) Enumerate clearly the essential features of an abridged life table.

b) The frequency distribution of the length of life of male insects belonging to a certain species S in an environment 'E' is given by

$$.25Ne^{-.25x}$$

(x being measured in days and N being the total number of insects at the start). The total number of male insects of the same species in a colony brought up in the same environment is 555 at present. Estimate the number of surviving insects in this colony 3 days hence. [15+15]=[30]

3. The proportion of ever married women observed in successive 5 year age groups are shown in the table below. The average number of female children born on completion of reproductive periods to women (including those who were widowed or divorced before the end of their reproductive periods) married at various ages are also shown in the same table. Calculate the Gross Reproduction Rate for the population. Explain the method of your calculation.

| age group in years | percentage of ever married women | average number of female children born to ever married women (completed fertility) |
|---|---|---|
| 15 - 19 | 4 | 1.868 |
| 20 - 24 | 43 | 1.297 |
| 25 - 29 | 73 | 0.933 |
| 30 - 34 | 80 | 0.633 |
| 35 - 39 | 81 | 0.300 |
| 40 - 44 | 82 | 0.069 |
| 45 - 49 | 83 | 0.002 |

[20]

4.a) Define 'underlying cause of death'. Briefly indicate the procedure of its selection and its significance from the public health point of view.

b) What types of errors occur in census data? How do they arise?

c) Explain briefly:

incidence and prevalence rates;
criteria for classifying acute and chronic illnesses;
duration of illness. [10+10+10]=[30]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

MID-YEAR EXAMINATIONS

Sh-5: Educational Statistics

Date: 30.12.67 Maximum marks: 100 Time: 3 hours

Answer question 1 and any five of the rest.

1. Write short notes on any five of the following:

a) Parallel tests
b) Incidental and explicit selections
c) Linen score
d) Item discrimination
e) Standard error of measurement
f) Percentile norm. [4+4+4+4+4]=[20]

2. Describe the errors of measurement, substitution and prediction in psychological tests, giving the equations defining them and their standard deviations. Explain the meaning of symbols used in the equations. [16]

3.a) Write a short note on the effect of test length on the coefficient of validity and reliability.

b) Explain how the method of analysis of variance can be used to estimate the reliability of a test. [8+8]=[16]

4.a) What function of test reliability (and test length) is invariant with respect to changes in test length? Carefully state the assumptions under which the above function is obtained.

b) There are two tests x and y. What is likely to be their correlation when both are doubled in length? Only the following information is available on the two tests.

$r_{xy} = .50 \quad r_{xx'} = .60, \quad r_{yy'} = .70.$ [8+8]=[16]

5.a) What is the effect of increasing test length p times on the variance of the observed scores and the standard error of measurement.

b) Assume a set of test scores each of which has been divided into comparable halves for purpose of obtaining a split-half reliability. Designate these halves by $x_a$ and $x_b$. Let $d = x_a - x_b$ (the difference between a person's score on part a and part b). The total score on the test is the sum of their halves ($S = x_a + x_b$). Assume that the halves are comparable so that their means and standard deviations are identical.

i) Write $r_{x_a x_b}$ in terms of the standard deviations of s and d.

ii) Write the reliability of the total test in terms of the standard deviations of s and d. [6+5+5]=[16]

6.a) Prove that the correlation between true and observed scores for a test of double length is $\sqrt{2r/(1+r)}$, where r is the reliability of the original test. List the assumptions used in making this derivation.

6.b) Obtain the reasonable limits for the difference of true scores $(t_i - t_j)$ when the corresponding observed scores difference is $(x_i - x_j)$ and $x_i$ and $x_j$ are the observed scores for the i-th and j-th persons. It is given that $\sigma_{so}$ is the standard error of measurement.

c) Write briefly what you mean by speed and power tests.

[6+6+4]=[16]

7.a) What is 'correction for guessing'? How is item difficulty value corrected for guessing? Discuss the commonly used formulae giving the underlying assumptions.

b) Show that the relationship between the reliability of the explicit selection variable and its validity for predicting the incidental selection variable is given by

$$\frac{r_{xy}\sqrt{1 - r_{xx'}}}{\sqrt{1 - r_{xy}^2}} \text{ equals a constant.}$$

[8+8]=[16]

8.a) What is meant by item analysis of a test?

b) A test in mathematics has reliability of .75 and a test in English has reliability of .70. The intercorrelation of the two tests is .45.

Estimate the degree of intercorrelation if

i) the mathematics test alone is made perfectly reliable

ii) the English test above is made perfectly reliable.

[8+4+4]=[16]

-----

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

| 148 |

PERIODICAL EXAMINATIONS

Statistics-5: Statistical Methods Theory and Practical

Date: 26.2.68 Maximum marks: 100 Time: 3 hours

Note: Answer any two question from Group A and all questions from Group B.

## Group A

1. Derive the standard error of sample $\sqrt{\beta_1}$. Describe a large sample test of normality based on sample $\sqrt{\beta_1}$. Why is it not possible to test for significance of the sample $\beta_1$ directly without taking the square root? [25]

2. Derive the limiting chisquare distribution for the sample good of fit (chisquare) statistic stating clearly the assumptions involved. How would the test procedure be affected if the sample frequencies represent the pooled frequencies in the different classes obtained by stratified sampling. [25]

3. The sample quartiles in a sample of size 100 from a certain population were - 0.63, 0.08 and 0.72 respectively. Test if those could be regarded as sufficient evidence to descard the hypothesis that the population sampled is standard normal. [25]

## Group B

4. In an experiment to determine whether five makes of automobile average the same number of miles per gallon, three cars of each make were selected at random in each of three cities and given a test run on one gallon of a standard gasoline. The table gives the number of miles travelled. Make an analysis of variance and determine whether there is a significant effect (a) of makes, (b) of cities.

| Make | Los Angeles | San Francisco | Portland |
|---|---|---|---|
| A | 20.3, 19.8, 21.4 | 21.6, 22.4, 21.3 | 19.8, 18.6, 21.0 |
| B | 19.5, 18.6, 18.9 | 20.1, 19.9, 20.5 | 19.6, 18.3, 19.8 |
| C | 22.1, 23.0, 22.4 | 20.1, 21.0, 19.8 | 22.3, 22.0, 21.6 |
| D | 17.6, 18.3, 18.2 | 19.5, 19.2, 20.3 | 19.4, 18.5, 19.1 |
| E | 23.6, 24.5, 25.1 | 17.6, 18.3, 18.1 | 22.1, 24.3, 23.8 |

[25]

5. An investigator was asked to take 10 independent measurements on the maximum internal diameter of a pot. The standard deviations of these measurements was .0345 mm. The experiment was repeated after a few days when the investigator had had sufficient practice and the standard deviation of the ten measurements was .0126 mm. Does it mean that with practice the investigator has become more consistent? State the assumptions underlying the test. Also state the null hypothesis and the alternative considered. Indicate the critical point(s) of the test at 5 % level of significance? [15]

The coefficient of correlation between two characters estimated by four investigators from samples of sizes 20, 25, 200 and 250 were 0.318, 0.253, 0.278 and 0.289 respectively. Would you regard the difference as due to fluctuations of sampling? [10]

------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

PERIODICAL EXAMINATIONS
Statistics-6: Design of Experiments

Date: 4.3.68 Maximum marks: 100 Time: 3 hours

Note: Answer Groups A and B in separate answerscripts.

## Group A

1. Explain the roles of randomisation and local control in planned experiments. Illustrate these by discussing some experiments. [15]

2. Explain the lay-out for an experiment to compare five treatments using a Latin Square design. How would you estimate the efficiency of this design relative to a design using the rows as blocks?

   Write down four mutually orthogonal latin squares of order five. [15]

3. Write short notes on any two.
   a) Analysis of covariance
   b) Missing plot technique
   c) Cross-over designs. [20]

## Group B

1.a) The following table gives the yield of wheat per plot in a manurial experiment carried out in a 4 × 4 Latin Square. The four manurial treatments are denoted by the numbers 1,2,3,4 in parantheses.

Yields in a 4 × 4 Latin Square Experiment.

| Row \ Column | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|
| 1 | (2) 425 | (3) 442 | (4) 540 | (1) 340 | 1747 |
| 2 | (4) 384 | (1) 512 | (2) 490 | (3) 408 | 1794 |
| 3 | (3) 506 | (4) 508 | (1) 536 | (2) 600 | 2150 |
| 4 | (1) 451 | (2) 568 | (3) 499 | (4) 347 | 1865 |
| Total | 1766 | 2030 | 2065 | 1695 | 7556 |

Total S.S.(corrected) = 87,863

Test whether the treatments are significantly different.

b) Carry out the analysis when the observation in the first row and fourth column is not available. [15+10]=[25]

GO ON TO THE NEXT PAGE

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV:1967-68

150

PERIODICAL EXAMINATIONS

Statistics-4

Date:11.3.68 Maximum marks: 100 Time: 3 hours

Note: Answer Groups A and B in separate answerscripts.

Group A: Inference

EITHER

1.a) Let $X_1,\ldots,X_n$ have joint density $e^{\theta_1 T_1(x)+\theta_2 T_2(x)}$, $a_i < \theta_i < b_i$, $i = 1,2$. Describe briefly how you can obtain the UMP unbiased similar test of $H_0$ $(\theta_1 = \theta_1^0)$ against $H_1$ $(\theta_1 \neq \theta_1^0)$. [20]

b) Let $X_1,\ldots,X_n$ be i.i.d. $N(\mu, \sigma^2)$. Find the UMP unbiased similar test of $H_0(\mu = \mu_0)$ against $H_1(\mu \neq \mu_0)$. [20]

OR

2.a) Show that if a confidence interval for $\theta$ is shortest in some class of confidence intervals in the sense of Neyman and has finite average length under all $\theta$ then in that class it minimises the average length under all $\theta$. [20]

b) Let $X$ be a real valued random variable with density $f_\theta(x)$ $(-\infty < \theta < \infty)$ which has monotone likelihood ratio in $x$. Suppose that UMP unbiased tests of the hypotheses $H_0(\theta = \theta_0)$ vs. $H_1(\theta \neq \theta_0)$ exist for all $\theta_0$ and are given by

$$c_1(\theta_0) \leq x \leq c_2(\theta_0).$$

Suppose also that $c_1$, $c_2$ are continuous functions of $\theta_0$. Show that a UMP unbiased confidence interval exists. [20]

3. Let $X$ and $Y$ be independently distributed according to one-parameter exponential families so that the joint distribution is given by

$$c(\theta_1)e^{\theta_1 T(x)}\psi(x)k(\theta_2)e^{\theta_2 U(y)}h(y).$$

i) Find the UMP unbiased test of $H_0(\theta_1 = a,\ \theta_2 = b)$ against $H_1(\theta_1 \neq a,\ \theta_2 = b)$. Show that it is unbiased against $H_1'(\theta_1 \neq a$ or $\theta_2 \neq b)$. [5]

ii) Show that there does not exist a UMP unbiased test of $H_0(\theta_1 = a,\ \theta_2 = b)$ against $H_1'(\theta_1 \neq a$ or $\theta_2 \neq b)$. [5]

GO ON TO THE NEXT PAGE

## Group B: Probability

4 The minimal state space of a homogeneous Markov chain consists of 1, 2, 3 and 4. The joint distribution of $X_1$, $X_2$ is as follows:

| Point | Probability at the point |
|---|---|
| (1, 1) | 1/12 |
| (1, 2) | 1/12 |
| (2, 1) | 1/6 |
| (2, 2) | 1/6 |
| (3, 4) | 1/6 |
| (4, 3) | 1/3 |

a) What is the transition probability matrix P ? [5]

b) Obtain the distribution of the random variable $X_{100}$. [6]

c) Obtain the joint distribution of the random variables $X_{100}$ and $X_{101}$. [5]

5.a) State Markov's basic theorem on stochastic matrices having at least one zero-free column. [4]

b) P is a stochastic matrix containing at least one zero - free column. Prove that $-1 < |P| < +1$.

$|P|$ means determinant of P. [5]

c) If S is any stochastic matrix, prove that

$-1 \leq |S| \leq +1$. [4]

d) Give numerical examples of stochastic matrices A and B (of any convenient order) such that $|A| = +1$, $|B| = -1$. [2+2]=[4]

6.a) Find the eigen values of the transition probability matrix

$$\begin{pmatrix} 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \\ 1/4 & 1/4 & 1/2 & 0 \\ 0 & 1/4 & 3/4 & 0 \end{pmatrix}$$ [8]

b) Explain the statements:

i) j is a consequent of i,
ii) transient state (or integer) [2+2]=[4]

c) Prove that if i is nontransient and j is a consequent of i, j is nontransient. [5]

-----------------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV:1967-68

152.

PERIODICAL EXAMINATIONS

Statistics-7: S.Q.C. Theory and Practical

Date:25.3.68 Maximum marks: 100 Time: 3 hours

Note: Answer Q. No. 1 and any other four questions from the rest.

1. Following table gives the averages and ranges in sample of size 4 of test records of copper content in commercial brass sheets.

| Sample | $\bar{X}$ | R | Sample | $\bar{X}$ | R |
|---|---|---|---|---|---|
| 1 | 11.10 | 0.6 | 16 | 11.45 | 1.3 |
| 2 | 11.70 | 1.2 | 17 | 11.55 | 1.6 |
| 3 | 11.35 | 1.0 | 18 | 9.98 | 0.4 |
| 4 | 11.25 | 1.0 | 19 | 10.78 | 1.2 |
| 5 | 11.40 | 2.0 | 20 | 11.23 | 0.7 |
| 6 | 11.00 | 0.6 | 21 | 10.93 | 1.7 |
| 7 | 11.20 | 1.0 | 22 | 11.50 | 2.7 |
| 8 | 11.35 | 1.2 | 23 | 10.78 | 0.7 |
| 9 | 11.50 | 2.0 | 24 | 10.95 | 1.1 |
| 10 | 10.88 | 1.1 | 25 | 11.48 | 2.9 |
| 11 | 10.85 | 1.0 | 26 | 11.80 | 0.4 |
| 12 | 11.53 | 1.2 | 27 | 12.20 | 2.0 |
| 13 | 11.15 | 0.8 | 28 | 11.88 | 1.5 |
| 14 | 11.28 | 1.0 | 29 | 11.23 | 0.8 |
| 15 | 11.00 | 0.8 | 30 | 11.30 | 0.6 |

a) Draw a control chart and test whether the process is under statistical control.

b) A minimum of 9 % copper in any sheet is the market specification. Excess of 0.1 % on an average results in a loss of Rs.8000 per annum to the factory. Estimate how much saving can be affected by maintaining statistical control at a proper level so as to satisfy market specification. [15+10]=[25]

2. What is average run length (ARL) of a control chart ? Derive the general expression $(1-p^{\lambda})/p^{\lambda}(1-p)$ for ARL of $\bar{X}$-chart where $\lambda$ successive points beyond a control limit indicate lack of control and p is the probability of a point falling beyond a control limit. [3+15]=[18]

3.a) Distinguish between specification limit and control limit. Describe different situations for a process under control with reference to the corresponding specification limit.

b) The tolerance specified for the inside diameter of a component is 0.005" ± 0.001". The components below lower specification would be reworked at the cost of Rs.C each and those above upper specification would be scrapped incurring a loss of Rs.5C each. The process s.d. is given to be 0.0006". Where should process be centered so that the total cost of rework and scrap is minimum? [4+6+8]=[18]

4.a). What is a group control chart? Explain how would you proceed to construct a group control chart for a group of 6 machines producing same item. Indicate your assumptions.

4.b) Briefly describe how to construct (i) control chart for number defectives and (ii) control chart for defect per unit when sample size varies. Give a brief interpretation for these charts. [6+12]=[18]

5.a) Explain any three

i) AQL (ii) AOI (iii) AOQL (iv) Indifference quality.

b) Derive general expression for OC and AOI of a double sampling plan. Construct a single sampling plan for AQL = 0.03 Producer's Risk = 0.05, LTPD = 0.08 and Consumers' Risk = 0.10. [6+12]=[18]

6.a) From Dodge and Romig Inspection tables construct a single and double sampling plan for

i) lot size = 900, LTPD = 5 %. Process average = 0.4 %.

ii) lot size =6000, AOQL = 2 %. Process average = 0.6 %.

b) In a inspection plan lot quality is measured by m, the average number of defects per item in the lot. Derive the acceptance and rejection numbers for the r-th sequential stage and outline graphical acceptance sampling procedure. Setup an item-by-item sequential sampling procedure for

AQL ($m_1$) = 0.8 Producers' Risk ($\alpha$) = 0.05

LTPD ($m_2$) = 2.4 Consumers' Risk ($\beta$) = 0.10. [6+12]=[18]

Neatness. [3]

---------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

[153]

PERIODICAL EXAMINATIONS

Statistics-8: Genetics

Date:1.4.68 Maximum marks: 100 Time: 3 hours

Note: Answer any four questions.

1.a) Briefly describe with examples Mendel's First and Second Principles of inheritance. Discuss their importance and applicability in the study of heredity and comment on their universality, if any.

b) Explain what is meant by cross over ratio and indicate in bare outline how this can be used to prepare autosomal maps, clearly stating the assumptions needed. [16+9]=[25]

2.a) What is meant by linkage? When is a character said to be sex-linked? In some Drosophila families it is found that sex ratio is approximately 2 female: 1 male. How can you explain the phenomenon?

b) A colourblind man has a normal brother and a colourblind sister. Give the genotypes of the parents.

c) Suppose one has to find homozygous individuals (for a single gene) by using a test cross to a recessive individual. What should be the size of the progeny raised so that in less than one per cent of the cases a heterozygous may pass off as a homozygous? [10+7+8]=[25]

3. For the estimation of linkage, calculate the amount of information per individual in a (i) Backcross and (ii) $F_2$ data for genes showing (a) complete dominance and (b) incomplete dominance. Comment on the loss of information because of incomplete classification. [In case of $F_2$ assume the recombination fraction to be the same for both the male and female gametogenesis]. [25]

4. You are given the following estimating equation

$$n_1 n_4 / n_2 n_3 = (2P + P^2)/(1 - 2P + P^2)$$

for $F_2$ data where $n_1$, $n_2$, $n_3$ and $n_4$ are the observed frequencies for the phenotypes AB, Ab, aB and ab respectively, and P stands for $(1-p)^2$, p being linkage ratio for both the male and female gametogenesis.

Calculate [on the large sample assumption] the variance of the estimate and show that the estimate is efficient. How does it compare with other rival estimates when single factor segregations are disturbed? [25]

5. Suppose the data for several families segregating for a single factor are available. Describe the methods you can apply for testing the heterogeneity of the families under different conceivable situations. Indicate how you can tackle the problem in case of hierarchial classifications. Give the expressions of $\chi^2$ which you will use for these problems, following Brandt and Snedecor's formula. [25]

6. For families of size $n$ segregating into $k$ classes with expected frequencies $m_i$ and observed frequencies $n_i$ $(i = 1, 2, 3 \ldots k)$ show that if $x = \sum_{i=1}^{k} \lambda_i n_i$, $x' = \sum_{i=1}^{k} \lambda_i' n_i$ ; $E(x) = E(x') = 0$.

then (i) $\frac{1}{n}$ Variance $(x) = \sum_{i=1}^{k} m_i \lambda_i^2$

and (ii) $\frac{1}{n}$ Covariance $(x, x') = \sum_{i=1}^{k} m_i \lambda_i \lambda_i'$.

For $F_2$ data involving two factors which are individually Mendelian, explain how you would use the above to set up tests by which linkage could be detected and Mendelian hypothesis for each character verified, all three being done independently. [25]

--------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV:1967-68

151

PERIODICAL EXAMINATIONS

Statistics-7: Econometrics Theory and Practical

Date:8.4.68 Maximum marks: 100 Time: 3 hours

1.(a) The following gives the percentage distribution of persons and the average percapita monthly total consumer expenditure by monthly per capita expenditure classes in rural India for 1960-61.

| monthly per capita expenditure class (Rs.) (0) | percentage of persons (1) | average per capita monthly total expenditure (Rs.0.00) (2) |
|---|---|---|
| 0 - 8 | 6.44 | 6.66 |
| 8 - 11 | 11.84 | 9.57 |
| 11 - 13 | 9.93 | 12.05 |
| 13 - 15 | 9.90 | 14.03 |
| 15 - 18 | 13.77 | 16.42 |
| 18 - 21 | 11.62 | 19.61 |
| 21 - 24 | 9.20 | 21.45 |
| 24 - 28 | 7.56 | 25.75 |
| 28 - 34 | 7.62 | 30.99 |
| 34 - 43 | 5.87 | 37.96 |
| 43 - 55 | 3.06 | 47.46 |
| 55 and above | 3.19 | 84.49 |

i) Draw the lorenz curve for monthly percapita total expenditure (X), showing clearly the necessary computations involved.

ii) Assuming that X is distributed as $\Lambda(\mu, \sigma^2)$, estimate $\sigma$ from the observed lorenz ratio of the given distribution. [25]

(b) Following are observations on quantities sold and prices of an agricultural commodity for six consecutive time periods, in a certain market.

| time t | quantity sold $x_t$('000 metric ton) | price $p_t$ (Rs./quintal) |
|---|---|---|
| 1 | 100.0 | 130.0 |
| 2 | 115.0 | 112.0 |
| 3 | 107.5 | 123.0 |
| 4 | 112.8 | 116.5 |
| 5 | 109.0 | 121.0 |
| 6 | 110.5 | 118.0 |

Assuming a Cobweb scheme for demand and supply and that market is cleared in every time period, plot the scatter of points both for the demand and supply curves. Pass straight lines through these scatters by inspection and estimate the price elasticities of supply and demand at the point of equilibrium. [20]

2. i) Define Lorenz curve for any positive variate. [3]

ii) State with proof the general properties of a Lorenz curve. [10]

iii) If in particular the distribution be $\Lambda(\mu, \sigma^2)$, what can you say about the Lorenz curve? [12]

3. EITHER

(a) What are the criteria that should be taken into consideration in the algebraic formulation of the Engel curve? [10]

OR

(b) What is the homogeneity hypothesis in relation to the Engel curve? State the consequent modification in the formulation of the Engel curve. [10]

4. EITHER

(a) How does economy of scale arise in household consumption? How would you formulate the Engel curve in this case? Indicate any method that you may know for estimating the parameters in this case. [20]

OR

(b) Criticise the approach of fitting a function

$q_t = f(p_t)$ to a time series data, where

$q_t =$ per capita quantity demanded (of a certain commodity) in the period $t$

$p_t =$ price of the commodity at time $t$. [20]

-----------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

208

ANNUAL EXAMINATIONS

Statistics-4: Probability

Date: 20.5.68 Maximum Marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

Answer groups A and B in separate answerscripts.

Note: The whole paper carries 130 marks. You may answer as much as you can from either group without necessarily restricting the total to 100. The maximum you can score is 100.

ALL MARKOV CHAINS CONSIDERED HAVE STATIONARY TRANSITION PROBABILITIES

Group A

1.a) P is a square matrix with non-negative entries and such that the sum of the elements in each row is <1. Prove that the sequence $P, P^2, P^3, \ldots$ converges towards the zero matrix. [9]

b) What is the minimal state space of a finite Markov chain? Prove that if the minimal state space is (1,2,..., N) and $1 \leq i \leq N$, $Pr(X_m = i) > 0$ for at least one value $m \leq N$. [9]

c) In a Markov chain, the transition matrix P is $\begin{pmatrix} a & 1-a \\ 1-a & a \end{pmatrix}$, where $0 < a < 1$. The first random variable $X_1$ takes the values 1,2 with probabilities $\alpha, \beta$ respectively. Let $P_n$ be the distribution of $X_n$. Obtain $\lim_{n\to\infty} P_n$. [9]

2.a) P is the transition matrix of a finite Markov chain in which every state is a consequent of every state; 1 is the greatest common factor of the set of positive integers n such that $p_{11}^{(n)} > 0$.

Prove that if i and j are any two states, $p_{ij}^{(n)} > 0$ for all sufficiently large n. [9]

b) MC-I is a finite Markov chain with transition matrix P, the states are 1,2,..., N. MC-II is another Markov chain with states 1',2',..., N' and transition matrix $\pi = P^n$ (n is some positive integer).

Prove that i is a transient state if and only if i' is transient. [10]

c) In a certain finite Markov chain, there is only one ergodic class and the transition matrix P satisfies the equation $P = P^2$.

Prove that it is not possible to divide the ergodic class into cyclically moving subclasses. Also, show that all the rows of P are identical. [8]

d) Prove that if A is a symmetric matrix (that is $a_{ij} = a_{ji}$ always), every power of A is symmetric. [3]

e) Prove that if the transition matrix P of a finite Markov chain is symmetric, the chain contains no transient states. [8]

Group B

3. In the case of a stochastic matrix P of order 14 × 14, the following cells contain positive entries and the other cells contain zero entries:

(1, 4), (1,6), (1,7); (2,7), (2,12), (2,14);
(3,4), (3,7); (4,9); (5,11); (6,13); (7,9);
(8,11), (8,14); (9,1); (10,4), (10,6); (11,8);
(12,2); (12,14); (13,1), (13,3), (13,10); (14,5).

(a) Determine the transient states, divide the non-transient states into ergodic classes, and partition each ergodic class into cyclically moving subclasses. [10]

(b) Give a positive integer $\alpha$ such that $P^{\alpha}, P^{2\alpha}, P^{3\alpha}, \ldots$ is convergent; try to make $\alpha$ as small as possible. [5]

4. In the case of a Markov chain with 7 states, the transition matrix is

$$P = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix}$$

(a) Determine the transient states, ergodic classes and cyclically moving subclasses. [5]

(b) Determine (approximately) the matrix $P^{1,000,000}$. [12]

(c) Determine (approximately) the numbers $p_{C2}^{(1,000,001)}$ and $p_{74}^{(1,000,001)}$ [6]

5. a) In a finite Markov chain, there are 3 transient states, i, j and k; E is an ergodic class. Prove that the absorption probabilities $\mathcal{G}_i(E)$, $\mathcal{G}_j(E)$ and $\mathcal{G}_k(E)$ are the solutions of the standard system of linear equations in which the coefficients are obtained from the 'south-east' portion of P (with some modifications).

Prove that the matrix of coefficients, in the standard system of linear equations, is nonsingular. [7+7]=[14]

b) If the transition matrix P is

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{4} & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \end{bmatrix}$$

determine each absorption probability $\mathcal{G}_i(E)$; here i runs through the various transient states and E runs through the various ergodic classes. Determine these probabilities by setting up the standard linear equations and solving them. [7]

c) In the foregoing example, determine the various absorption probabilities $g_i(E)$ by considering the set of all paths from i to E and adding their probabilities. Verify that the answers you get by this method agree with the answers obtained in the preceding subquestion. [6]

-----

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68 [209]

ANNUAL EXAMINATIONS

Statistics-4: Inference

Date: 22.5.68. Maximum Marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

Answer groups A and B in separate answerscripts.

Answer any two questions from each group.

## Group A

1.a) Show that all similar tests have Neyman-structure iff there exists a boundedly complete sufficient statistic. [5]

b) State and prove Basu's theorem on independence of statistics. [5]

c) $X_1,\ldots,X_n$ are i.i.d. $N(\mu_1,\sigma^2)$, $Y_1,\ldots,Y_n$ are i.i.d. $N(\mu_2,\sigma^2)$ and X's and Y's are independent. Find the UMP unbiased test of $H_0(\mu_1=\mu_2)$ against $H_1(\mu_1\neq\mu_2)$. [15]

2.a) Let $X_1,\ldots,X_n$ be independent and normal with same variance $\sigma^2$. Let $E(X_i)=\theta_i$, $i=1,\ldots,p$.

$= 0, i = p+1,\ldots,n$.

Show that no UMP unbiased test of

$H_0(\theta_1=\ldots\theta_k=0)$, $2\leq k\leq p$, exists. [10]

b) Show that the usual F-test for the above problem is the UMP invariant test. Hence show that the F-test is unbiased. [15]

3.a) If G is a finite group leaving a testing problem invariant and there exists a minimax test of size $\alpha$ then show that there exists an invariant minimax test of size $\alpha$. [10]

b) If G is a group leaving a testing problem invariant and $\phi_1$ is the unique UMP unbiased test then show that $\phi_1$ is almost invariant under G. [7]

c) Give an example where the UMP invariant test of a composite hypothesis coincides with that obtained by the likelihood ratio principle or give an example where the two principles lead to different tests. [8]

## Group B

4.a) Neglecting the excess over boundaries prove the optimum property of the SPRT. [12]

b) Let $X_i$ be i.i.d. $N(\theta,1)$. Construct an SPRT of $H_0(\theta=0)$ vs. $H_1(\theta=1)$ with $\alpha=\beta=0.1$. Calculate its OC and ASN for $\theta=0, .5, 1$. [13]

5. Two random samples of equal size are drawn from two populations with continuous distribution function F and G. Characterise the class of all similar tests of $H_0(F=G)$ and briefly discuss the importance of the rank tests. [5]

5. (contd.)

Show that the one-sided Wilcoxon's U-test can be obtained as a locally most powerful test against a class of alternatives. Obtain the greatest lower bound to its Pitman efficiency as compared with the standard t-test for location shift alternatives. [10+10]=[20]

6.a) Let $X_1, \ldots, X_n$ be i.i.d. with common density $f(x,\theta)$. Assuming the consistency of the maximum likelihood estimator for the unknown parameter $\theta$, prove its asymptotic normality under suitable conditions. [10]

b) Briefly discuss the relation between shortness of confidence intervals as defined by Wilks, shortness as defined by Neyman and UMP tests.

Let $X$ be $N(\theta, 1)$. You are to find a confidence interval for $\theta$. Find a group $G$ which leaves this problem invariant and under which the shortest invariant confidence interval is $X \pm t_\alpha$ where $1-\alpha$ is the confidence coefficient

$$\int_{-t_\alpha}^{t_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = 1-\alpha. \qquad [8+7]=[15]$$

--------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68 [210]

ANNUAL EXAMINATIONS

Statistics-5: Statistical Methods Theory and Practical

Date: 24.5.68 Maximum Marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

Answer groups A and B in separate answerscripts.

## Group A

Note: Answer any three questions from this group.

1. Let $Y_1, Y_2, \ldots, Y_n$ be n independent normal variates with the same variance $\sigma^2$ and mean values given by

$$E(\underset{\sim}{Y}) = X\underset{\sim}{\beta}$$

where X is a known matrix of rank r, and $\underset{\sim}{\beta}' = (\beta_1, \beta_2, \ldots, \beta_m)$, a vector of unknown parameters. Consider a hypothesis $H_o$ that $\beta$ satisfies a set of consistent linear equations $H\underset{\sim}{\beta} = \underset{\sim}{\xi}$ ($\underset{\sim}{\xi}$ known) where each element of $H\underset{\sim}{\beta}$ is an estimable linear function of $\beta_1, \beta_2, \ldots, \beta_m$ and rank of H is s. Write

$$R_o^2 = \min_{\underset{\sim}{\beta}} (\underset{\sim}{Y} - X\underset{\sim}{\beta})'(\underset{\sim}{Y} - X\underset{\sim}{\beta})$$

$$R_{H_o}^2 = \min_{\underset{\sim}{\beta},\text{ subject to } H_o} (\underset{\sim}{Y} - X\underset{\sim}{\beta})'(\underset{\sim}{Y} - X\underset{\sim}{\beta})$$

Show that

a) $R_o^2/\sigma^2$ is distributed as a chisquare with n d.f. [2]

b) If $H_o$ is true, $(R_H^2 - R_o^2)/\sigma^2$ is distributed as a central chisquare with s d.f. independently of $R_o^2$. [10]

2. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with variance covariance matrix $\Sigma$ and mean values as in question 1. Let Z ($n \times \overline{n-r}$) be a matrix of rank n-r such that $X'Z = 0$, and $\hat{\underset{\sim}{\beta}}$ a possible choice of $\underset{\sim}{\beta}$ for which $(\underset{\sim}{Y} - X\underset{\sim}{\beta})'(\underset{\sim}{Y} - X\underset{\sim}{\beta})$ is minimum. Show that:

(a) for $\underset{\sim}{\lambda}'\hat{\underset{\sim}{\beta}}$ to provide the best linear unbiased estimate of $\underset{\sim}{\lambda}'\underset{\sim}{\beta}$ for each estimable $\underset{\sim}{\lambda}'\underset{\sim}{\beta}$, it is necessary and sufficient that $\Sigma$ be of the form $\Sigma = XAX' + ZBZ'$ where A and B are arbitrary p.s.d. matrices. [8]

(b) The condition on $\Sigma$ given in (a) is equivalent to any one of the following conditions:

b1) if $\mathcal{M}(X)$ denotes the linear manifold generated by the columns of X and $\underset{\sim}{x} \in \mathcal{M}(X)$ then $\Sigma\underset{\sim}{x} \in \mathcal{M}(X)$. [5]

b2) the columns of X are spanned by r linearly independent eigen vectors of $\Sigma$. [5]

3.a) Show that, under the appropriate null hypothesis the distribution of the Wilk's $\Lambda$ criterion is the same as the product of several independent Beta variables. [10]

GO ON TO THE NEXT PAGE

3.b) Obtain an expression for the $t^{th}$ moment of $\Lambda$ when the d.f. of the sum of product matrix representing deviation from hypothesis is q, that for the sum of product matrix due to error is n-q and the order of each matrix is $p \times p$. [4]

c) Show that when $q = 2$, $\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \frac{n-p-1}{p}$ is distributed as a variance ratio with 2p and 2(n-p-1) as the d.f. for numerator and denominator respectively. [4]

4. Let $(X_1, X_2, \ldots, X_p)$ be distributed as p-variate normal with mean values zero and variance covariance matrix $\Sigma$ (possibly singular).

(a) Show that for $\underset{\sim}{X}' A \underset{\sim}{X}$ to be distributed as chisquare it is necessary and sufficient that $A$ satisfies

$$\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$$

What is the d.f. of this chisquare?

Give an example of an A satisfying this condition, and obtain the d.f. of the corresponding chisquare.

[7+1+2]=[10]

(b) Let A and B be two symmetric matrices satisfying the condition in (a). Show that for $\underset{\sim}{X}' A \underset{\sim}{X}$ and $\underset{\sim}{X}' B \underset{\sim}{X}$ to be distributed as independent chisquares it is necessary and sufficient that

$$\Sigma A \Sigma B \Sigma = 0.$$

[3]

5.a) Let $x_{(p)}$ and $y_{(p')}$ denote respectively the pth quantile of x and the p'th quantile of y based on n pairs of observations on dependent variables x and y. Show that (in a sense you are required to make precise and under conditions you should state in full) the joint distribution of $x_{(p)}$ and $y_{(p')}$ tends to a bivariate normal distribution as $n \to \infty$.

[Hint: You may assume the Bahadur representation of sample quantiles] [8]

b) Determine the asymptotic variance covariance matrix of the approximating bivariate normal distribution. [5]

c) In a sample of size 100 from a bivariate normal population with unknown mean $\mu_x$ of x and $\mu_y$ of y and variance covariance matrix $\begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}$ the sample medians of x and y were computed respectively as 2.5 and 1.8. Test for the hypothesis $\mu_x = \mu_y = 2$. [5]

[Hint: $\int_0^\infty \int_0^\infty \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)} \, du \, dv = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho.$]

GO ON TO THE NEXT PAGE

## Group B

Note: Answer either question 6 or question 7 from this group.

6. In an experiment to determine whether five makes of automobile average the same number of miles per gallon three cars of each of five makes were selected at random in each of three cities and given a test run on one gallon of standard gasoline. The following table presents the data. Regard the problem as one arising from a mixed model with the makes of car as fixed and cities chosen at random. Obtain an analysis of variance of the data and determine whether there is significant effect of (a) makes (b) cities (c) interactions makes X cities.

| Make | cities | | |
|---|---|---|---|
| | Los Angeles | San Francisco | Portland |
| A | 20.3, 19.8, 21.4 | 21.6, 22.4, 21.3 | 19.8, 18.6, 21.0 |
| B | 19.5, 18.6, 18.9 | 20.1, 19.9, 20.5 | 19.6, 18.3, 19.8 |
| C | 22.1, 23.0, 22.1 | 20.1, 21.0, 19.8 | 22.3, 22.0, 21.6 |
| D | 17.6, 18.3, 18.2 | 19.5, 19.2, 20.3 | 19.4, 18.5, 19.1 |
| E | 23.6, 24.5, 25.1 | 17.6, 18.3, 18.1 | 22.1, 24.3, 23.8 |

[ Sum of 45 observations = 923.2

Sum of square of 45 observations = 19107.06 [26]

7. The mean values of three biometrical characters and the matrix of pooled variances and covariances were obtained from two groups of female desert locusts - one in the phase gregaria, and the other in an intermediate phase between gregaria and solitaria. They are given below:

| Phase | n | $\overline{X}_1$ | $\overline{X}_2$ | $\overline{X}_3$ |
|---|---|---|---|---|
| gregaria | 20 | 25.80 | 7.81 | 10.77 |
| Intermediate | 72 | 28.35 | 7.41 | 10.75 |

Matrix of pooled variances and covariances

| | $X_1$ | $X_2$ | $X_3$ |
|---|---|---|---|
| $X_1$ | 4.7350 | 0.5622 | 1.4685 |
| $X_2$ | | 0.1431 | 0.2174 |
| $X_3$ | | | 0.5702 |

Test if the biometrical measurements $X_1$, $X_2$ and $X_3$ are useful for discriminating between the two phases. [26]

8. Practical records. [10]

9. Viva Voce. [10]

---------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

[211]

ANNUAL EXAMINATIONS

Statistics-6: Design of Experiments Theory and Practical

Date: 25.5.68. Maximum Marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

Answer groups A and B in separate answerscripts.

## Group A

Note: Answer any two questions from this group.

1.a) Explain what is meant by a balanced incomplete block (BIB) design. State and prove some necessary conditions on the parameters of a BIB design. [10]

b) Show that in a BIB design (i) the least squares estimates of orthogonal treatment contrasts are uncorrelated (ii) the least squares estimate of any normalised treatment contrast has the same variance.

What is the average variance for estimates of treatment comparisons of type $\tau_i - \tau_j$? [10]

2. Let $\Phi_d(x)$ denote cyclotomic polynomial of order d. Use the relation $x^h - 1 = \prod_{d/h} \Phi_d(x)$ to find $\Phi_{15}(x)$. Verify that $x^4+x+1$ is an irreducible factor of $\Phi_{15}(x)$. Hence or otherwise give a procedure for obtaining elements of GF(16). (You need not write out the elements explicitly). Express $\alpha^6 + \alpha^3$ as a power of $\alpha$ for some primitive element $\alpha$ of the GF(16) obtained by you.

State a procedure for obtaining 15 mutually orthogonal Latin squares (m.o.l.s.) of order 16.

Show that one can not construct more than (s-1) m.o.l.s. of order s. [4+4+2+4+6]=[20]

3. Write short notes on any two:

(1) Fractional replications
(2) Use of concomitant information
(3) Role of randomisation in planned experiments. [10+10]=[20]

## Group B

Note: Answer any two from questions 4, 5 and 6 of this group

4. Explain what is meant by (i) confounding (ii) partial confounding. Give a scheme of partial confounding for a $2^5$ experiment using five replications in blocks of 8 plots each such that (i) no main effect or two factor interaction is confounded, and (ii) no effect is confounded in more than one replication. (Write down the complete lay-out for any one replication and only the key-block for the remaining ones).

How would you compute the sum of squares due to a partially confounded effect? [4+12+4]=[20]

5. The following results on yield from two replications in an experiment designed to compare 16 barley varieties numbered 25 to 40 (variety number being given in parenthesis).

| Repl. I Block | | | | | Repl. II Block | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | (25) 2320 | (26) 2470 | (27) 2220 | (28) 2390 | 1 | (25) 1200 | (29) 1550 | (33) 2080 | (37) 2440 |
| 2 | (29) 2950 | (30) 3560 | (31) 3240 | ([illegible]) 2680 | 2 | (26) 1540 | (30) 2880 | (34) 2390 | (38) 3220 |
| 3 | (33) 3180 | (34) 3020 | (35) 2690 | (36) 2790 | 3 | (27) 1540 | (31) 2310 | (35) 3520 | (39) 3780 |
| 4 | (37) 2380 | (38) 2810 | (39) 2450 | (40) 2730 | 4 | (28) 1660 | (32) 2320 | (36) 1470 | (40) 2770 |

(a) Analyse the data giving your comments on the results. [15]

(b) Obtain standard errors for the different paired comparisons between varieties. Also give an expression for the average variance for all pairs of varietal differences. [5]

6. In a varietal and manurial experiment on oats two levels of nitrogen (0 and 0.2 cent per acre designated by $n_0$ and $n_1$ respectively) were applied to each of three varieties. First a latin square lay-out was used for allocation of plots to the varieties. The different levels of nitrogen were applied by splitting each plot into two sub-plots and by choosing one at random for $n_0$ (the other received $n_1$).

| Row \ Col. | 1 | | | 2 | | | 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $n_0$ | $n_1$ | | $n_0$ | $n_1$ | | $n_0$ | $n_1$ |
| 1 | $V_3$ | 63 | 70 | $V_2$ | 60 | 102 | $V_1$ | 53 | 74 |
| | | $n_0$ | $n_1$ | | $n_1$ | $n_0$ | | $n_1$ | $n_0$ |
| 2 | $V_2$ | 80 | 82 | $V_1$ | 64 | 68 | $V_3$ | 99 | 97 |
| | | $n_1$ | $n_0$ | | $n_0$ | $n_1$ | | $n_0$ | $n_1$ |
| 3 | $V_1$ | 90 | 62 | $V_3$ | 89 | 129 | $V_2$ | 89 | 82 |

Analyse the data. Obtain standard errors for differences between pairs of treatment combinations. [15+5]=[20]

7. Practical record [10]

8. Viva Voce [10]

---------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

[212]

ANNUAL EXAMINATIONS

Statistics-7: Econometrics Theory and Practical

Date: 28.5.68 Maximum Marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

Answer groups A and B in separate answerscripts.

Group A Max. marks: 40

Note: Answer Q.3 and any one from Q. Nos. 1 and 2.

1. Discuss the effect of household composition in the formulation of the Engel curve and give a formulation taking into consideration this effect. Suggest any method, you know, for estimating the Engel function. [15]

2. What is Solow's concept of neutral technological change? Derive the technological change function in this case and comment briefly on the probability implications of the model. What data do you require to estimate the function? [15]

3. The following table gives the number of persons percent, per capita total expenditure, and per capita expenditure on cereals, according to per capita expenditure classes. Compute the Engel elasticities from concentration curves, after specifying the assumptions you make. Estimate the increase in per capita expenditure on cereals when per capita total expenditure increases by 10 per cent, other things remaining the same. [25]

| per capita expenditure classes (Rs.) | percentage distribution of persons | per capita total expenditure (Rs.) | per capita expenditure on cereals (Rs.) |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| 0 - 8 | 9.03 | 6.18 | 3.68 |
| 8 - 11 | 14.73 | 0.49 | 5.57 |
| 11 - 13 | 11.33 | 12.01 | 6.86 |
| 13 - 15 | 9.87 | 13.98 | 7.63 |
| 15 - 18 | 13.05 | 16.44 | 8.51 |
| 18 - 21 | 9.48 | 19.55 | 9.21 |
| 21 - 24 | 8.49 | 22.46 | 10.41 |
| 24 - 28 | 6.96 | 25.66 | 10.37 |
| 28 - 34 | 6.17 | 30.34 | 11.49 |
| 34 - 43 | 5.34 | 38.26 | 12.82 |
| 43 - 55 | 2.75 | 48.23 | 14.32 |
| 55 and above | 2.95 | 89.45 | 16.49 |

Group B Max. marks: 60

Note: Answer Q.6 and any one from Q.Nos. 4 and 5.

Establish the relationship between the Gini mean difference and the Lorenz ratio. [15]

GO ON TO THE NEXT PAGE

5. Show, stating necessary assumptions that the exponents of a Cobb-Douglas production function equal the shares of the factors of production. Derive the cost function from the Cobb-Douglas relation. [15]

6. In a study to fit a production function for an economy, data for 21 years were collected on output ($X_1$) measured at constant prices, total workers ($X_2$) and total stock of capital ($X_3$) measured at constant prices. Defining $Y_i = \log X_i$, $i = 1,2,3$, the variance-covariance matrix of the above variables together with time ($Y_4$) as another variable was

| | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ |
|---|---|---|---|---|
| $Y_1$ | .00562850 | .00187085 | .00045005 | .25020000 |
| $Y_2$ | | .00093135 | .00000045 | .05515500 |
| $Y_3$ | | | .00030905 | .00057500 |
| $Y_4$ | | | | 38.50000000 |

On the assumption that the production function can be represented by a linear regression of $Y_1$ on $Y_2$, $Y_3$ and $Y_4$, estimate the elasticities of production with respect to capital and labour. Comment on the results, performing any tests you consider desirable. Comment also on the significance of the time variable $Y_4$. [25]

7. Practical Records. [10]

8. Viva Voce. [10]

---------------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

213

ANNUAL EXAMINATIONS

Statistics-7: Industrial Statistics Theory and Practical

Date: 30.5.68 Maximum Marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

Answer groups A and B in separate answerscripts.

Group A

Answer question 1 and any two questions of the rest from this group

1.a) If $X_B = B^{-1} b$ is a basic feasible solution of the problem Maximise $Z = CX$ subject to $AX = b$, $X \geq 0$, and $P_k$ a vector of $A$ not in $B$ and $Y_k = B^{-1} P_k$, show that if we introduce $P_k$ into the basis, the vector to be removed is determined by finding

$$\theta = \frac{x_{B_r}}{Y_{rk}} = \underset{i}{\text{Min}} \left\{ \frac{x_{B_i}}{Y_{ik}} , y_{ik} > 0 \right\}$$ [6]

b) Prove that the net decrease in the value of objective function is $\theta(z_k - c_k)$ where $z_k = c_B B^{-1} P_k$. Hence derive the conditions for optimality. What conclusion you will draw if all $y_{ik} \leq 0$? [6]

c) State the rules of transformation to get the next tableau. [4]

2.a) Explain the meaning of the following terms - i) slack variable, ii) feasible solution, iii) Basic feasible solution and iv) non-degenerate basic feasible solution. [3]

b) Show that any basic feasible solution of a set of linear equations is an extreme point of the convex set of feasible solutions and vice versa. Is the correspondence always unique either way? [9]

3.a) Show that evaluations of primal slack variables in the final tableau give the optimal values of the dual structural variables and the evaluations of primal structural variables give the optimal values of dual slack variables. [7]

b) What conclusion you can draw about the dual problem when the primal is unbounded? [3]

c) Write down the dual of the following problem
Maximise $Z = 3x_1 + x_2$ subject to

$$x_1 + x_2 - x_3 \leq 1$$
$$2x_1 + 3x_2 - x_3 = 2$$
$$x_1 + x_2 + x_3 \geq 3$$
$$x_j \geq 0$$ [2]

4.a) Explain the need for artificial basis in Linear programming. Compare the M method with the two phase method. [4]

GO ON TO THE NEXT PAGE

4.b) If you are using the M method, what conclusions you can draw under the following situations when the optimality criteria is satisfied.

i) No artificial variable appears in the basis

ii) One or more artificial variables appear in basis at zero level

iii) One or more artificial variables appear in basis at a positive level. [4]

c) In the two phase method, if one or more artificial variables appear in the basis at zero level at the end of phase I, indicate what modification in the rule to determine the vector to be removed has to be made. What happens if we do not use the modification? [4]

Group B

Answer question 5 and any two questions of the rest from this group.

5.a) Formulate a Linear Programming problem in the form of revised simplex and show that the inverse of the basis $B_1$ of the revised simplex form is given by

$$B_1^{-1} = \begin{bmatrix} 1 & C_B B^{-1} \\ 0 & B^{-1} \end{bmatrix}$$

where B is the corresponding basis in simplex. [6]

b) Solve the following problem by the revised simplex.
Maximise $x_1 + 2x_2$ subject to

$$x_1 + x_2 \leq 3$$
$$x_1 + 2x_2 \leq 5$$
$$3x_1 + x_2 \leq 6$$
$$x_j \geq 0$$ [6]

c) Compare the revised simplex with simplex and bring out the salient points of difference. What are the advantages of revised simplex over simplex? [4]

6. A supermarket has two girls ringing up the sales at the counters. If the service time for each customer is exponential with mean 4 minutes and if people arrive in Poisson process at the rate of 10 per hour

a) What is the probability that the customer has to wait for service? [8]

b) What is the expected percentage idle time for each girl? [4]

7.a) What are the characteristics required to describe a queuing system? [3]

b) In the single server, Poisson arrival and exponential service time queue system, the arrivals actually consist of pairs of customers. If the average arrival rate is $\lambda$ pairs per unit time and the average service rate is $\mu$ customers per unit time, show that the steady state probabilities $P_n$ of n customers in the system satisfy the following equations.

$$\lambda P_0 = \mu P_1$$
$$(\lambda+\mu) P_1 = \mu P_2$$
$$(\lambda+\mu)P_n = \mu P_{n+1} + \lambda P_{n-2} \quad (n \geq 2).$$ [5]

GO ON TO THE NEXT PAGE

7.c) Show that the generating function $P(z)$ of $P_n$ is given by

$$P(z) = \frac{2(1-\rho)(1-z)}{\rho z^3 - (2+\rho)z + 2} \quad \text{where} \quad \rho = \frac{2\lambda}{\mu}.$$ [4]

8.a) Discuss briefly the meaning of steady state conditions in a congestion system and the interpretation of steady state probabilities. [3]

b) Show that for the Queue system with Poisson arrival, general independent service time distribution and single server, the average number in the system and the average waiting time are given by

$$E(n) = \rho + \frac{\rho^2 + \lambda^2\sigma^2}{2(1-\rho)}$$

$$\lambda E(w) = \frac{\rho^2 + \lambda^2\sigma^2}{2(1-\rho)}$$

where $\lambda$ is the mean arrival rate, $\mu$ is the mean service rate, $\rho = \lambda/\mu$ and $\sigma^2$ is the variance of the service time. [5]

c) Show that the average waiting time for exponential distribution is twice the average waiting time for a regular service distribution both with the same mean service rate. [4]

9. Practical records. [10]

10. Viva Voce. [10]

------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1967-68

214

ANNUAL EXAMINATIONS

Date: 31.5.68 Maximum Marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

Answer groups A and B in separate answerscripts.

Note: Answer any two questions from group A and both the questions from group B.

## Group A

1.a) Discuss the physical as well as genetical basis of the classification of human beings into the blood groups, viz., O, A, B and AB. Explain how this classification can be utilised in blood transfusion and draw up a donor recipient table.

b) Describe Bernstein's method of estimating the gene frequencies with reference to O-A-B blood group system and comment on its merits, if any.

c) In a case of disputed parentage two babies were of type MN and N respectively. Their mothers were also of type MN and N, but it was uncertain to which mother either baby belonged. The husband of woman MN was of type N; the husband of woman N was of type M. To what mother did the type N baby belong? [10+6+4]=[20]

2. Suppose the initial population at birth is in equilibrium with respect to an autosomal character involving two alleles. Let the fertilities of the recessive, heterozygous and dominant genotypes be in the ratio 1 - S: 1 :1-s, the mating being random otherwise.

i) Find for this case the equilibrium value of the frequency of the mutant gene.

ii) Evaluate the progress of the population and calculate the frequency of the mutant gene after n generations, in the particular case when S=1 and s=0 and the same selection continues all throughout the generations. What is the practical implication of your result? Explain with the help of an example. [10+10]=[20]

3. Hogben states the following proposition in regard to random mating as applied to single gene substitutions which do not involve the X chromosome.

'Equilibrium is attained in a single generation of random mating so that if anything occurs to upset the pre-existing equilibrium a new equilibrium is reached after mating has once occurred'.

i) Prove the proposition.

ii) Prove that the same is not true for a sex-linked locus.

iii) Show that if random mating continues for all the generations, ultimately equilibrium for a sex-linked locus is arrived at, the stable genotypic frequencies being given as

| AA | Aa | aa |
|---|---|---|
| $p^2$ | $2p(1-p)$ | $(1-p)^2$ |

for the homogametic sex and

GO ON TO THE NEXT PAGE

| A | a |
|---|---|
| p | 1 - p |

for the heterogametic sex, where p is equal to $(p_0 + 2p_1)/3$ and $p_0$ and $p_1$ are frequencies of A in heterogametic sex in generations 0 and 1 respectively.
[8+3+9]=[20]

## Group B

4.a) Assume that the genotypic frequencies in a population are given as

| Genotype: | AA | Aa | aa |
|---|---|---|---|
| Frequency: | $p^2$ | $2pq$ | $q^2$ |

, $p + q = 1$.

Given that a man is of genotype Aa, show that the probability that his brother is of the same genotype is $(1 + pq)/2$.

b) The following data were obtained in an $F_2$ population, concerning segregation for green and yellow plant colour, and purple and yellow aleurone colour in maize.

| | | Aleurone colour | |
|---|---|---|---|
| | | Purple | White |
| Plant colour | Green | 127 | 67 |
| | Yellow | 19 | 44 |

Test if there is any linkage between the genes. [10+15]=[25]

5. One thousand people in a particular community were classified according to sex and according to whether or not they were colourblind as follows:

| | Male | Female |
|---|---|---|
| Normal | 442 | 514 |
| Colourblind | 38 | 6 |

According to a genetic model these numbers should have relative frequencies

| | Male | Female |
|---|---|---|
| Normal | $p/2$ | $p^2/2 + pq$ |
| Colourblind | $q/2$ | $q^2/2$ |

where $q = 1 - p$ is the proportion of defective genes in the population.

i) State the assumptions clearly under which the above model is obtained

ii) Obtain the maximum likelihood estimate of q and the estimate of its standard error.

iii) Test whether the data are consistent with this model. [4+10+11]=[25]

6. Viva Voce. [10]

-------