INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part. IV: 1969-70

[402]

PERIODICAL EXAMINATION

Statistics-4: Probability

Date: 22.9.69 Maximum Marks: 100 Time: 3 hours

Note: The whole paper carries about 130 marks. You may attempt any bit of any question. Marks allotted for each question are given in brackets [ ].

1.a) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and f_n, f real valued measurable functions on Ω. Define the following types of convergence of f_n to f. [5×4]=[20]

(i) almost uniformly
(ii) almost everywhere
(iii) in measure
(iv) in distribution
(v) in mean

b) Prove the following implications among these types of convergence.

(i) <—> (ii) if $\mu(\Omega) < \infty$ [12]

(ii) —> (iii) if $\mu(\Omega) < \infty$ [8]

(v) —> (iii) [6]

(iii) —> (iv) if $\mu(\Omega) < \infty$ [8]

c) Give examples to show the falsity of the following implications when $\mu(\Omega) < \infty$.

(i) —> (v) [3×5]=[15]

(iii) —> (ii)

(iv) —> (iii)

2.a) Define the product space of a sequence of probability spaces $(X_i, \mathcal{B}_i, P_i)$. [6]

b) Define a tail set in the product space. [4]

c) Show that if P is the product measure and E is a tail set, then $P(E) = 0$ or 1. [10]

d) Indicate how you deduce from the result in (c) that if f_n is a sequence of independent random variables, the convergence set of $\sum_{n=1}^{\infty} f_n$ has probability 0 or 1. [6]

3.a) Prove that if A_n is a sequence of events such that

$$\sum_{n=1}^{\infty} P(A_n) < \infty, \text{ then } P(\lim\sup A_n) = 0$$ [6]

b) Prove that if A_n are independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$, then $P(\lim\sup A_n) = 1$. [8]

c) Show that the hypothesis of independence cannot be dropped in (b). [4]

4.a) If $X_1, X_2, \ldots, X_n$ are independent random variables with finite expectations, prove that

$$P[\max_{1 \leq j \leq n} |S_j - ES_j| \geq c] \leq \frac{1}{c^2} \sum_{j=1}^{n} \sigma_j^2 \qquad [10]$$

b) Deduce that if $X_1, X_2, \ldots$ is a sequence of independent identically distributed random variables with finite variance then $\frac{1}{n} S_n \longrightarrow E(X_1)$ in probability. [6]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [403]

PERIODICAL EXAMINATION

Statistics-4: Statistical Inference

Date: 29.9.69 Maximum Marks: 100 Time: 3 hours

Note: Answer as many questions as you can. All questions carry equal marks.

1. Explain clearly what is a statistical decision problem and discuss the various components that enter into the decision procedure.

2. Discuss the concepts of admissibility and unbiasedness in relation to a decision problem. Also discuss briefly the Bayes and minimax procedures of selecting an optimum decision procedure.

3. How do you view the problem of estimation as one of decision theory? Illustrate. What is the definition of unbiasedness as employed in the theory of estimation? Also define consistency of an estimate. Give some examples to show that unbiasedness and consistency do not imply each other.

4. What is meant by 'Fisher's information'

a) in a sample; (b) in a statistic?

What is the relation between the two?

Define a sufficient statistic for a family of dominated probability measures and prove a necessary and sufficient condition for a statistic to be sufficient for a family of probability measures defined on a sample space with finite number of points.

5.a) Let $T(x)$ be a real valued sufficient statistic for a family of dominated probability measures

$$\mathcal{P} : \left\{ P_\theta \ \ \theta \in \Omega \right\} \text{ on } (\mathcal{X}, \mathcal{A}),$$

where Ω is an interval of R'. Obtain explicitly the form of the frequency function stating clearly the assumptions involved.

b) What is a minimal sufficient statistic for a family of probability measures? Describe a method of obtaining the minimal sufficient statistic.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70

[404]

PERIODICAL EXAMINATION

Statistics-8: Educational Statistics Theory and Practical.

Date: 13.10.69 Maximum Marks: 100 Time: 3 hours

Note: Answer Q.1 and any five from the rest.

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

a) Summation Score; (b) Factorial Validity
c) Reliability of the linear combination of scores
d) Group heterogeneity and validity
e) Test discrimination coefficient. [5+5+5+5]=[20]

2. Describe the errors of measurement, substitution and prediction in psychological tests; define them and their standard deviations. State the equations and explain the meaning of symbols used in the equations. [16]

3.a) Describe the computational procedure of obtaining T-score from the raw score.

b) Suppose a test is composed of items with equal difficulty Show that

$$r_{tc} = \frac{\overline{r_{ic}}}{\overline{r_{it}}}$$

where r_{tc} = correlation between test score and criterion

$\overline{r_{ic}}$ = average correlation between items and criterion

$\overline{r_{it}}$ = average correlation between items and total score. [8+8]=[16]

4.a) Show how the method of analysis of variance can be used to estimate the reliability of a test.

b) In a test of 55 items, the standard deviation of the total score was 7.5. The sum of the variances of the items was 9.8327. Estimate the reliability of the total score. [8+8]=[16]

5.a) In a sample of 61 fifteen-year-old high school students, of whom 26 were male and 35 were female, the mean weights in Kilogram were 67.8 and 56.6 respectively. The standard deviation of the weights for the combined group was 13.2. Find the point biserial correlation between sex and body weight for fifteen-year-old high school students.

b) Briefly sketch the proof of the following proposition 'In order to increase the reliability of a test from r to R, the number of items should be multiplied by

$$\frac{(1-r)R}{(1-R)r}\text{'}$$ [8+8]=[16]

6.a) Describe in brief Spearman's two-factor theory and Thurstone's multiple-factor theory.

b) Write a short note on 'correction for guessing'. [8+8]=[16]

7.a) Describe different types of validity of a test. Write a short note on the effect of test length on validity of the test.

7.b) Let x and y be the scores on two parallel tests and

$$V(x-y) = 75.0, \quad V(x+y) = 800.0$$

What is the reliability of $(x+y)$ score. [10+6]=[16]

8.a) A test consisting 4 items was administered on a large number of students. The following table gives the values of p_i = proportion of students answering the ith item correctly.

p_{ij} = proportion of students answering the i^{th} and j^{th} item correctly.

Let X denote the total number of items correctly answered by a student. Compute mean and variance of X

Item	Value of p_{ij} for item 1	2	3	4	p_i
1	-	-	-	-	0.80
2	0.68	-	-	-	0.73
3	0.58	0.48	-	-	0.65
4	0.40	0.38	0.25	-	0.44

b) A test in Geometry has a reliability of .80 and a test in Algebra has reliability of .90. The correlation between tests is .60. Estimate the degree of correlation if

i) the Algebra test alone is made perfectly reliable

ii) the Geometry test alone is made perfectly reliable.

[10+6]=[16]

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

Discuss the two indices (i) difficulty (ii) discrimination that could be obtained for each item.

b) Show that test validity must always be less than the index of reliability. What is meant by 'correction for attenuation'? [10+6]=[16]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70

[405]

Statistics-6: Sample Surveys

Date: 3.11.69 Maximum Marks: 100 Time: 3 hours

Note: Answer all questions.

1. A population of 12 units is divided into 2 strata thus:
Stratum I. U_1, U_3, U_9, U_{10}.

Stratum II. $U_2, U_4, U_5, U_6, U_7, U_8, U_{11}, U_{12}, U_{13}$.

A simple random sample without replacement of size 2 is taken from Stratum I and a sample of size 3 is taken from Stratum II with probabilities proportional to the serial numbers of the units and with replacement.

a) Characterise the sampling design fully.

b) Write down expressions for the conventional estimator, $\hat{Y}_{con}$, of the population total, $V(\hat{Y}_{con})$ and $\hat{V}_{con}(\hat{Y}_{con})$.

c) Write down $\hat{Y}_{HT}$, $V(\hat{Y}_{HT})$ and $\hat{V}_{HT}(\hat{Y}_{HT})$.

2. Prove that for any general stratified sampling strategy

$$\hat{Y}_{HT} = \sum_{i=1}^{k} \hat{Y}_{HT}^{(i)}$$

where k is the number of strata and $Y_{HT}^{(i)}$ denotes the H.T. estimator of the total for the i-th stratum.

3. For question (1) above derive an unbiased estimator, based on the sampling given in that question, of the variance of $\hat{Y}^*_{con}$, the conventional estimator that we would use had we taken a simple random sample of size 5 without replacement from the over all population.

4. For a general sampling design, write down (a) a set of necessary and sufficient conditions (nsc) for the existance of an unbiased estimator $\hat{Y}$ of the total Y and

(b) a set of n.s.c. for the existence of an unbiased estimator of $\bar{V}(\hat{Y})$, and

(c) write down $\hat{V}_G(\hat{Y}_{HT})$ - the Grundy's estimator of $V(Y_{HT})$ and prove that it is unbiased for $V(\hat{Y}_{HT})$.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [illegible]

PERIODICAL EXAMINATION

Statistics-7: Econometrics

Date: 10.11.69 Maximum Marks: 100 Time: 3 hours

(All questions carry equal marks.)

1. Discuss the identification problem in the context of estimating a demand curve from time series data on prices and quantities.

2. How do you go over to a market demand function from the individual demand or to the demand for a group of commodities from the demand functions of the individual commodities? Discuss the limitations of your method, if any.

3. Why is it generally necessary to use cross section data along with time series data while estimating a demand function?

4. Describe the Cobweb model of demand and supply of a commodity and find out the time path of the price of the commodity.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [407]

PERIODICAL EXAMINATION

Statistics-5: Statistical Methods, Theory and Practical

Date: 17.11.69 Maximum Marks: 100 Time: 3 hours

Note: Answer <u>any four</u> questions.

1. A matrix $G(n \times m)$ is said to be a generalised inverse of $A(m \times n)$ if $AGA = A$.

Show that if G is a g-inverse of A, then

(a) GA is idempotent, $R(GA) = R(A)$.

(b) a general solution of a consistent equation $A\underset{\sim}{x} = \underset{\sim}{y}$ is
$$\underset{\sim}{x} = G\underset{\sim}{y} + (I - GA)\underset{\sim}{z}$$
where $\underset{\sim}{z}$ is arbitrary.

(c) $\underset{\sim}{q}'\underset{\sim}{x}$ has a unique value for all $\underset{\sim}{x}$ obeying $A\underset{\sim}{x} = \underset{\sim}{y}$ if and only if $\underset{\sim}{q}'GA = \underset{\sim}{q}$,

(d) the most general form of a g-inverse of A is
$$G + U - GAUAG$$
where $U(n \times m)$ is an arbitrary matrix,

(e) show that the converse of (a) is also true.

2. Let E_n be the vector space of real n-tuples with the usual inner product definition and $A(n \times m)$ be a matrix with elements from the real number field. Show that for a matrix P to be a projection operator projecting arbitrary vectors in E_n onto $\mathcal{M}(A)$ it is necessary and sufficient that the following four conditions hold:

(a) $P' = P$ (b) $P^2 = P$ (c) P be of the form $P = AD$ for some D (d) $R(P) = R(A)$.

Hence or otherwise show that $P = A(A'A)^{-}A$ where $(A'A)^{-}$ denotes a g-inverse of $A'A$.

3. Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{0}, \Sigma)$. Show that for $\underset{\sim}{X}'A\underset{\sim}{X}$ and $\underset{\sim}{X}'B\underset{\sim}{X}$ to be distributed as independent chisquare variables it is necessary and sufficient that

(a) $\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$ (b) $\Sigma B \Sigma B \Sigma = \Sigma B \Sigma$

(c) $\Sigma A \Sigma B \Sigma = 0$

obtain the degrees of freedom of the two chisquare variables.

4. A steel bar 18 inches long is subjected to a carefully regulated hardening process. The hardening is determined at the extremities of the bar, and at nine positions in between. The following results are obtained.

Distance from one end of the bar (d, in inches)	0	1.8	3.6	5.4	7.2	9.0	10.8	12.6	14.4	16.2	18.0
Vicker's hardening number (h)	250	276	298	335	374	414	454	503	558	604	671

It is required to determine a mathematical function to graduate the change in hardness along the bar. Fit the function $h = A + Bd + Cd^2$ and assuming the true relation to be quadratic in d the standard errors of the estimated co-efficients.

A, C, D, B are four points in order on a straight line. The following table gives the averages of a number of measurements of the length of the segments. Find least squares estimates of the lengths and obtain their standard errors. Clearly mention the assumptions.

Lengths of the segments in cm.

segment	No. of measurements	av. length
AC	3	45.10
AD	2	77.95
CD	2	32.95
CB	3	98.36
DB	4	65.55
AB	2	141.03

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70

MID-YEAR EXAMINATION

Statistics-4: Probability

Date: 19.12.69 Maximum Marks: 100 Time: 3 hours

<u>Note:</u> Answer as many questions or parts thereof as you can. Marks allotted for each question are given in brackets [ ].

1.a) State Kolmogorov's inequalities. [4]

b) Deduce that if $X_1, X_2, \ldots$ are independent, $EX_k = 0$ and $\sum_{k=1}^{\infty} \sigma_k^2 < \infty$, then $S_n = X_1 + \ldots + X_n$ converges a.s. [8]

c) Prove that if $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, then $\frac{1}{n} S_n \to 0$ a.s.. (You may assume the lemmas of Cauchy and Kronecker). [6]

2.a) Define the equivalence (Khintchine) of two sequences of random variables. State and prove the equivalence lemma. [8]

b) If $X_1, X_2, \ldots$ are i.i.d., prove that $\frac{S_n}{n} \to c$ a.s. if and only if $E|X_1| < \infty$. Also then show that $c = E(X_1)$. [12]

c) State Kolmogorov's three series theorem. [4]

3.a) Formulate the central limit problem in the case of random variables centred at expectations and having bounded variances. Call these conditions (C). [4]

b) Prove that if $\psi(u) = \int g(u,x)\,dK(x)$ with usual notation then e^{ψ} is a limit characteristic function under (C). Deduce the uniqueness lemma. [12]

c) State the comparison lemma and the convergence lemma clearly. [8]

d) State the central limit theorem under conditions (C). Prove it assuming the results of (b) and (c). [10]

e) Deduce from (d) that $N(0,1)$ is a limit law under (C) and prove the Lindeberg-Feller theorem. [12]

f) Prove Liapounov's theorem. [6]

g) State the central limit theorem in the case of bounded variances when the variables have non-zero expectations. [4]

4.a) State and prove the Daniel-Kolmogorov consistency theorem. [12]

b) What is a stochastic process? What is the significance of the theorem of (a) in this connection? [6]

5.a) State and prove Markov's inequality. Deduce Tchebichev's inequality. [8]

b) State and prove Holder's inequality. [8]

Neatness and clarity. [3]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70

[409]

MID-YEAR EXAMINATION

Statistics-4: Statistical Inference

Date: 20.12.69 Maximum Marks: 100 Time: 3 hours

Note: You may attempt any 6 questions. Five questions carry full marks. All questions carry equal marks.

1. Discuss briefly the general statistical decision problem bringing out clearly the various factors that enter into it. Also explain how the problems of point estimation and tests of hypotheses can be viewed as decision problems.

2. Let $P(.; \theta)$, $\theta \in \Omega$ be a family of dominated measures on $(X, \mathcal{A})$. Give a construction through which you can obtain the minimal sufficient statistic for the family. Show explicitly that the statistic you obtain is minimal sufficient.

3. Let $P(.; \theta)$ $\theta \in \Omega$ be a family of dominated probability measures on $(X, \mathcal{A})$. Obtain a set of sufficient conditions for the existence of minimum variance unbiased estimates of estimable functions of θ. Is such an estimate unique?

4.a) Explain the concept of a lower bound for the variance of an estimator of a function of the parameter.
Let $P(.; \theta)$, $\theta \in \Omega$ where Ω is an open interval of the real line, be a family of dominated probability measures on $(\mathbb{R}, \mathbb{B})$ and let $f(x, \theta)$ be a determination of the density functions.
Let $A(\theta, \phi) = Var\left[\frac{f(x; \theta)}{f(x; \phi)} \mid \theta\right] < \infty$ for all $\theta, \phi \in \Omega$ $(\theta \neq \phi)$. Let $T(x_n)$ be an unbiased estimate of a function $g(\theta)$, based on a random sample of size n. Obtain a lower bound for the variance of $T(x_n)$.

b) nder any required additional conditions obtain the Cramér-Rao lower bound for the variance of $T(x_n)$.

c) Characterize the family of distributions for which the variance of $T(x_n)$ is equal to the Cramér-Rao lower bound.

5.a) With the same probability set up as in Question 4 derive the Bhattacharya bounds for the variance of $T(x_n)$ stating the regularity conditions you need for the derivation. How do you deduce Cramér-Rao inequality from this?

b) Let $N(\mu, \sigma^2)$, $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$ be a family of normal distributions with unknown parameters μ and σ^2. Obtain a lower bound for the variance of any estimate of μ based on a random sample of size n.

6. Define asymptotic efficiency of an estimate. Show that the method of maximum likelihood gives asymptotically efficient estimates (state the regularity conditions clearly).

7.a) Describe the method of scoring for the maximum likelihood estimation.

b) What is meant by information matrix corresponding to a single observation and corresponding to a random sample of size n? What is the relation between them?

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [410]

MID-YEAR EXAMINATION

Statistics-5: Statistical Methods (Theory)

Date: 22.12.69. Maximum Marks: 100 Time: 3 hours

Note: Answer all questions. All questions carry equal marks.

1.a) Explain the concept of 'estimable parametric function' 'best linear unbiased estimate' in the context of the usual Gauss-Markoff (G-M) linear set up.

b) What are 'normal equations'? Show that the sum of squares of deviations of observations from their expectations under the G-M set up is minimised when one substitutes for the unknown parameters any solution of the normal equations.

c) Show that if a linear function L of the parameters assume a unique value for all solutions of the normal equations, then L is estimable.

d) If the random vector $\underset{\sim}{Y}$ has expectation

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

and

$$X = \begin{pmatrix} 2 & 1 & 1 \\ 1 & -1 & -4 \\ -1 & 2 & 7 \\ 3 & 2 & 3 \end{pmatrix}$$

are all linear functions of the parameters in β estimable? If not, give an example of a non-estimable linear function and establish it to be so.

2.a) If X and H are matrices of order $n \times m$ and $k \times m$ respectively and $(H')^{\perp}$ is a matrix of maximum rank such that $H(H')^{\perp} = 0$, then

$$R[X(H^{\perp})'] = R\begin{pmatrix} X \\ H \end{pmatrix} - R(H).$$

b) If A and B are matrices of order $n \times p$ and $n \times q$ respectively and the matrix C is such that $\mathcal{M}(C) = \mathcal{M}(A) \cap \mathcal{M}(B)$, then show that C can be expressed as

$$C = AF$$

where $F = W^{\perp}$ and $W = A'B^{\perp}$ with $(\ \ )^{\perp}$ as defined in 2(a).

c) Let Y, X, β, H, c be matrices of order $n \times 1$, $n \times m$, $m \times 1$, $k \times m$ and $k \times 1$ respectively. Show that if $\mathcal{M}(H') \subset \mathcal{M}(X')$ and the equations $H\beta = c$ are constant then

$$\min_{\beta} (Y - X\beta)'(Y - X\beta) - \min_{\beta} (Y - X)'(Y - X\beta)$$

subj to $H\beta = c$

$$= (H\hat{\beta} - c)'[H(X'X)^{-}H']^{-}(H\hat{\beta} - c).$$

3. a) Let $\underset{\sim}{Y} \sim N_n(X\beta, \sigma^2 I_n)$, where X is a known matrix of order $n \times m$ and rank r. Let $Z_1, Z_2, \ldots, Z_{n-r}$ be linear functions of the type $Z_i = \underset{\sim}{c}_i' \underset{\sim}{Y}$ such that $E(Z_i) = 0$ and $D(\underset{\sim}{Z}) = \sigma^2 I_{n-r}$, then show that
β

$$R^2_{o(n)} = \min_{\underset{\sim}{\beta}} (\underset{\sim}{Y} - X\underset{\sim}{\beta})'(\underset{\sim}{Y} - X\underset{\sim}{\beta}) = \sum_{i=1}^{n-r} Z_i^2 .$$

b) Let $\underset{\sim}{Y}'$ and X' be partitioned as $\underset{\sim}{Y}' = (\underset{\sim}{Y}_1', \underset{\sim}{Y}_2')$ and $X' = (X_1' \mid X_2')$, where both $\underset{\sim}{Y}_1'$ and X_1' have n_1 columns.

If $$R^2_{o(n_1)} = \min_{\underset{\sim}{\beta}} (\underset{\sim}{Y}_1 - X_1\underset{\sim}{\beta})'(Y_1 - X_1\underset{\sim}{\beta}).$$

obtain the distribution of $R^2_{o(n)} - R^2_{o(n_1)}$.

4. Two objects W_1 and W_2 were repeatedly weighed on two spring balances and the results are recorded below.

Spring Balance 1		Spring Balance 2	
Weight of	Grams	Weight of	Grams
W_1	210	W_1	200
W_2	485	W_2	480
W_1 and W_2	710	W_1 and W_2	690

Assuming that the balances may have unequal bias obtain the best linear estimates of the individual weights of W_1, W_2, the total weight of W_1 and W_2, the bias of balance 1 and balance 2 separately and also the standard errors of these estimates.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [411].

MID-YEAR EXAMINATION

Statistics-5: Statistical Methods Practical

Date: 2.12.69 Maximum Marks: 100 Time: $3\frac{1}{2}$ hours

Note: Answer all questions. Marks allotted for each question are given in brackets [ ].

EITHER

1.a) Describe a procedure of drawing a random sample from a bivariate normal population with known parameters.

b) Briefly indicate the method of Fractile Graphical Analysis to compare two bivariate populations. [8+12]=[20]

OR

Compute the Moore-Penrose inverse of the following matrix

$$\begin{matrix} 17 & 21 & 13 \\ 7 & 11 & 9 \\ 3 & -1 & -5 \end{matrix}$$

[20]

2. Observations on 12 independent normal variables with expectations given as linear functions of $\theta_1, \theta_2, \theta_3, \theta_4$ and a common variance σ^2 led to the following normal equations

$$\begin{aligned} 2\theta_1 + \theta_2 + \theta_3 + \theta_4 &= 6.8 \\ \theta_1 + 5\theta_2 + 2\theta_3 + 2\theta_4 &= 12.4 \\ \theta_1 + 2\theta_2 + 10\theta_3 + 7\theta_4 &= 17.2 \\ \theta_1 + 2\theta_2 + 7\theta_3 + 5\theta_4 &= 13.6 \end{aligned}$$

Sweepout operations on these equations gave the following computed figures:

A particular solution: $\hat{\theta}_1 = 2.0$, $\hat{\theta}_2 = 1.6$; $\hat{\theta}_3 = 1.2$, $\hat{\theta}_4 = 0$.

A g-inverse of the matrix C of normal equations

$$C^- = \frac{1}{81}\begin{pmatrix} 46 & -8 & -3 & 0 \\ -8 & 19 & -3 & 0 \\ -3 & -3 & 9 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Basis for the vector space generated by rows of matrix C:

$(1, 0, 0, \frac{1}{9})$, $(0, 1, 0, \frac{1}{9})$, $(0, 0, 1, \frac{2}{3})$.

The sum of squares of the 12 observations was computed as 54.44.

a) Examine if the following linear functions of parameters are estimable

i) $\theta_1 + 2\theta_2 + \theta_3 + \theta_4$; (ii) $\theta_2 - \theta_3$. [5]

b) Test the hypothesis $\theta_1 = \theta_2$. [10]

c) An independent observation on yet another normal variable $N(3\theta_1 + 3\theta_2 - \theta_3, \sigma^2)$ is now obtained as 1.3. Compute revised estimates for $\theta_1 - \theta_2$ and its estimated standard error, showing each step of computation. [20]

3. Three important measurements from which the cranial capacity (c) may be predicted are the glabella - occipital length (L), the maximum parietal breadth (B) and the basio-bregmatic height (H). The prediction formula suggested is

$$C = \alpha L^{\beta_1} B^{\beta_2} H^{\beta_3}$$

These can be written as

$$X_0 = A + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

where $X_0 = \log_{10} C$, $X_1 = \log_{10} L$, $X_2 = \log_{10} B$ and $X_3 = \log_{10} H$.

The mean values and the corrected sum of squares and the products of these transformed characteristics computed on the basis of measurements on the 86 male skulls from the Farrington Street series are given below.

Characters	Mean	Corrected sums of squares and products (S.S.S.P.) X_0	X_1	X_2	X_3
X_0	3.1685	.12692	.03030	.04410	.03629
X_1	2.2752		.01875	.00848	.00684
X_2	2.1523			.02904	.00878
X_3	2.1128				.02886

The inverse of the S.S.S.P. matrix for variables X_1, X_2 and X_3 is computed as

$$\begin{pmatrix} 64.21 & -15.57 & -10.49 \\ . & 41.71 & -9.00 \\ & & 39.88 \end{pmatrix}$$

Using these figures

a) Compute the multiple (linear) regression equation of X_0 on X_1, X_2 and X_3. [10]

b) Test if the three regression co-efficients occurring in this expression

i) are significantly different from zero,

ii) are significantly different from one,

iii) are significantly different from one another. [25]

Practical Records. [20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [412]

MID-YEAR EXAMINATION

Statistics-6: Sample Surveys

Date: 24.12.69 Maximum Marks: 100 Time: 3 hours.

Note: Answer all questions. Marks allotted for each question are given in brackets [ ]. 40 marks are reserved for assignments.

1. A sample size n is drawn from a population of size N as follows: One unit is selected with probabilities proportional to the $\mathcal{X}$-values of the units, $(n-1)$ units by simple random sampling without replacement. Let $\mathcal{Y}$ be the variate under study.

a) Prove that the estimator

$$\hat{Y} = N \cdot \frac{\bar{y}}{\bar{x}} \cdot \bar{X}$$

is unbiased for $Y = \sum_{1}^{N} Y_i$. [6]

b) Characterise the sampling design generated. [5]

c) Find expressions for $\hat{Y}_{HT}$, $V(\hat{Y}_{HT})$ and $\hat{V}_G(\hat{Y}_{HT})$ i.e. the Horvitz-Thompson estimator, its variance and Grundy's estimator of this variance. [5+6+6]=[17]

d) Prove that $\hat{V}_G(\hat{Y}_{HT})$ is non-negative for this sampling method. [5]

2. Suppose that the known values of the non-negative auxiliary variate $\mathcal{X}$ and the unknown values of the main variable $\mathcal{Y}$ are connected in a super-population set-up thus:

$$\left.\begin{aligned} E_\theta(Y_i \mid X_i) &= \alpha X_i \\ V_\theta(Y_i \mid X_i) &= \sigma^2 X_i^2 \\ \text{and} \quad Cov_\theta(Y_i, Y_j \mid X_i, X_j) &= 0 \end{aligned}\right\} \quad \text{for } 1 \leq i \neq j \leq N.$$

where the suffix θ denotes that the expectations etc., are taken over the super-population, and α and σ^2 are unknown positive constants.

Prove that in the class $\mathcal{H}(\nu_0)$ of all sampling strategies $H(P,T)$ which satisfy

i) $\nu(P) = \nu_0$

and ii) $T \in L_0^*(P)$

the sub-class that further satisfy

a) $\pi_i(P) = \nu_0 \cdot \frac{X_i}{X}$

b) $\nu(s) = \nu_0$, a.s.(P)

and c) $T = \hat{Y}_{HT}(P)$

are θ-optimum in the sense that $E_\theta V(H)$ is minimised for this sub-class. [18]

3. Write short notes on

a) the unified framework for sample surveys
b) pps sampling
c) systematic sampling. [9]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [413]

MID-YEAR EXAMINATION

Statistics-7: Econometrics Theory and Practical

Date: 26.12.69 Maximum Marks: 100 Time: 4 hours

Note: Answer all questions. Marks allotted for each question are given in brackets [ ].

1. What do you understand by the term 'Engel curve'. What are the points that have to be taken into consideration in the algebraic formulation of an Engel curve. Give examples of different types of formulation with their economic properties. [20]

2. What is quality elasticity of demand? How do you measure it? What is the relation between value, quantity and quality elasticities? [15]

3. How do you take into account the effect of household composition in the formulation of the Engel curve? Give a method of estimation of the curve in this case. [15]

4. The indices of price and quantity of an agricultural commodity sold in different years are given below. Plot suitable scatter diagrams revealing demand and supply curves for the commodity under the assumption that a Cobweb model would apply and that every year the market is cleared but for random fluctuations. Fit the demand and supply curves to those scatters by inspection.

Year	0	1	2	3	4	5	6
Index of price	102	132	112	125	116	121	118
Index of quantity:	124	93	106	98	104	100	102

[20]

5. The following table gives the estimated percentage of population along with estimated per capita monthly total expenditure and expenditure on cereals for urban areas of India during 1957-58 by per capita monthly total expenditure classes.

Fit a constant elasticity Engel curve to the data and find out whether cereal is a luxury or necessary item of expenditure.

Monthly per capita Exp. classes (in Rupees)	P.C. of population	Per capita monthly total Exp. (Rs.)	Per capita monthly Exp. on cereals (Rs.)
(0)	(1)	(2)	(3)
0 - 8	2.09	6.37	3.67
8 - 11	5.56	9.73	4.68
11 - 13	7.26	11.95	5.17
13 - 15	7.11	13.96	5.72
15 - 18	10.42	16.52	6.30
18 - 21	11.38	19.41	6.49
21 - 24	9.79	22.35	7.37
24 - 28	11.09	25.94	7.46
28 - 34	9.27	30.74	7.79
34 - 43	9.31	37.91	7.67
43 - 55	7.27	48.65	7.70
55 and above	9.45	86.54	8.47

[30]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [414]

MID-YEAR EXAMINATION

Statistics-8: Demography - Theory and Practical

Date: 27.12.69 Maximum Marks: 100 Time: 3 hours

Note: Answer all questions. Marks allotted for each question are given in brackets [ ].

1. Discuss briefly the method adopted by the census Actuary for the construction of Indian life tables (1941-50) based on two successive census counts. [25]

2. Write short notes on:
   i) crude death rate
   ii) force of mortality
   iii) infant mortality rate and its adjustment
   iv) complete expectation of life. [25]

3. The following table shows the growth of a bacterial colony as observed in a certain locality in square centimeters. Fit a logistic curve to the data.

Age of colony (in days)	Area in square centimeters
0	0.24
1	2.78
2	13.53
3	36.30
4	47.50
5	49.50

[25]

4.a) Define Vital Statistics and briefly discuss the uses of vital statistics to a public health administrator.

b) State some of the recommendations of the expert bodies set up in recent years to examine the question of improving the collection, registration and compilation of vital statistics. [10+15]=[25]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B-Stat. Part IV: 1969-70

PERIODICAL EXAMINATIONS

Statistics 4: Probability

Date: 23.2.70 Maximum marks: 100 Time: 3 hours

Note: The whole paper carries 110 marks. Answer as many questions as you can. Marks allotted for question are given in brackets ( ).

1. (a) Define a Markov chain and the n-step transition probabilities associated with it. (8)

(b) State and prove the Chapman-Kolmogorov equation. (8)

(c) What is the relation between the matrices $P=((p_{ij}))$ and $P_n=((p_{ij}^n))$ when the transition probabilities are stationary? (4)

2. (a) If $P_{ij}(s)$ and $F_{ij}(s)$ are the generating functions of the sequences p_{ij}^n and f_{ij}^n, then prove that

$$P_{ij}(s) = F_{ij}(s) \cdot P_{jj}(s) + \partial_{ij} \qquad (12)$$

where ∂_{ij} is the Kronecker delta.

(b) Prove that if state E_j is transient, then $\sum_n p_{ij}^n < \infty$. (6)

(c) Prove that $\sum_n f_{ii}^n = 1$ if and only if $\sum_n p_{ii}^n = \infty$. (10)

3. (a) If E_i is recurrent and E_j is accessible from E_i then prove that E_i and E_j commute and that E_j is recurrent. (10)

(b) Also then show that $\sum_n f_{ij}^n = 1$ and $Q_{ij} = 1$ where $Q_{ij} =$ $P(E_j$ occurs i.o. $/E_i$ occurs at time $0)$. (14)

4. (a) Define the period of a state E_i Show that if E_i and E_j commute then their periods are equal. (10)

(b) If E_i and E_j commute show that there exists an integer r_j such that $p_{ij}^n > 0$ implies $n \equiv r_j \pmod{d(E_i)}$. Further there exists an $N(j)$ such that $n \geq N(j)$ implies that $p_{ij}^{n\,d(E_i)+r_j} > 0$. (12)

(c) If E_i is recurrent and $f_{ii}^1 > 0$, then prove that

$$\lim p_{ii}^n = \frac{1}{\sum_n n\, f_{ii}^n} \qquad (16)$$

INDIAN STATISTICAL INSTITUTE
Research and Training School

M.Stat. Part IV: 1969-70

PERIODICAL EXAMINATION

Statistics-4: Inference

Date: 2.3.70 Maximum Marks: 100 Time: 3 hours

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

1. Let $\{P_{\underset{\sim}{\theta}},\ \underset{\sim}{\theta} \in \Theta\}$ be a family of dominated probability measures on $(\mathcal{X}, \mathcal{A})$ and Θ be a subset of a k-dimensional Euclidean space. Let $\{f(x, \underset{\sim}{\theta}),\ \underset{\sim}{\theta} \in \Theta\}$ be a determination of the density functions.

   i) Define the information matrix (a) due to a single observation, (b) due to a random sample of size n. How are they related?

   ii) What is meant by a minimal set of sufficient statistics for the above family of measures?

   iii) What is the relation between the information matrix due to a minimal set of sufficient statistics based on a random sample of size n and that due to the random sample itself.

2. With the same set up as in question 1, suppose it is required to estimate parametric functions $g_1(\underset{\sim}{\theta}), \ldots, g_r(\underset{\sim}{\theta})$

   i) Obtain a lower bound for the generalised variance of estimators of $g_1(\underset{\sim}{\theta}), \ldots, g_r(\underset{\sim}{\theta})$ stating clearly the regularity conditions.

   ii) Show that the generalised variance of unbiased estimators of $g_1(\underset{\sim}{\theta}), \ldots, g_r(\underset{\sim}{\theta})$ based on sufficient statistics is not greater than the generalised variance of any estimators of the same based on the sample.

3. Let x be a normal variate with mean μ and variance σ^2. Obtain the lower bound for the generalised variance of estimators of μ and σ^2 based on a sample of size n. Can you obtain estimator of μ whose variance is equal to the lower bound you obtain above? What will be the correlation of your estimator with a given estimator, say $\hat{\sigma}^2$ of σ^2? How do you explain the observed correlation?

4.a) Let $f(x, \theta)$, $\theta \in \Theta$ be a family of density functions of probability measures P_θ, $\theta \in \Theta$ on $(\mathcal{X}, \mathcal{A})$. Obtain a set of necessary and sufficient conditions that a test $\phi(x_1, \ldots, x_n)$ for the simple hypothesis $H_0: \theta = \theta_0$ against the simple alternative $H_1: \theta = \theta_1$ based on a random sample of size n be a MPT of size α.

b) A sample of observations is drawn from a normal distribution with mean 2 and unknown variance σ^2. Derive the MPT of size α for the hypothesis $H_0: \sigma^2 = 9$ against the alternative $H_1: \sigma^2 = 16$, based on a sample of size 10 from the population.

c) Define a monotone likelihood ratio family of distributions with a single real valued parameter θ. Obtain the UMPT of size α for the hypothesis $H_0: \theta \leq \theta_0$ against the alternative $\theta > \theta_0$.

5.b) A machine is known to produce defective items. From a lot of sufficiently large size N, a sample of size n is drawn and each item is inspected and m items are detected to be defective. Derive a test for the hypothesis that the number of defectives in the lot is $\leq D_0$ against the alternative that it is $> D_0$. Is your test UMP?

6. The probability of success in an experiment is an unknown constant p. The experiment is continued until a specified number m of successes have been obtained. Based on this it is desired to test the hypothesis that $p \leq p_1$ or $p \geq p_2$ against the alternative $p_1 < p < p_2$ where p_1 and p_2 are known values. Derive an optimum size α test for this hypothesis and discuss its optimum properties.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70

PERIODICAL EXAMINATION

Statistics-8: Statistical Methods

Date: 9.3.70 Maximum Marks: 100 Time: 3 hours

Note: Answer all questions. Marks allotted for each question are given in brackets [ ].

1.a) Define a central Wishart distribution $W_p(n, \Sigma)$.

b) Show that if $S_1 \sim W_p(n_1, \Sigma)$, $S_2 \sim W_p(n_2, \Sigma)$ and S_1 is independently distributed of S_2 then $S_1 + S_2 \sim W_p(n_1 + n_2, \Sigma)$.

c) Show that if

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \sim W_{p+q}\left(n, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

where Σ_{22} is p.d. of order q, then

$$S_{11} - S_{12} S_{22}^{-1} S_{21} \sim W_p(n - q, \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}).$$

d) Obtain the characteristic function of the Wishart distribution $W_p(n, \Sigma)$. [25]

2.a) Obtain the null distribution of Mahalanobis D^2.

b) Discuss some applications of Mahalanobis D^2 in multivariate analysis. [25]

3. The means and the dispersion matrix based on measurements of three biometrical characters (Y_1, Y_2, Y_3) of $n = 112$ male desert locusts in an intermediate phase are given below:

	mean	Dispersion matrix Y_1	Y_2	Y_3
Y_1	24.92	2.68960	0.20090	0.66746
Y_2	6.60		0.12250	0.10613
Y_3	9.76			0.32490

Is there any evidence to conclude that the population means of the three characteristics Y_1, Y_2, Y_3 are 25, 8, and 10 respectively? [35]

4. The correlation coefficients between the scores in two halves of psychological test applied on three groups of 30, 20 and 25 students were 0.63, 0.48 and 0.71 respectively. Examine if the groups are different in respect of the correlation coefficient. [15]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [411]

PERIODICAL EXAMINATION

Statistics-7

Date: 16.3.70 Maximum Marks: 100 Time: 3 hours

Note: Answer Groups A and B in separate answerscripts. Marks allotted for each question are given in brackets [ ].

Group A: Econometrics

Maximum Marks: 60 Suggested time: 2 hours
Answer all questions

1. Derive a general formulation of cost function of a particular firm, starting from the first principle. What would be the general shape of this curve? In our real life observation which part of the curve generally observed? [15]

2. Recast the inter-industry flow relations in terms of production function terminology. Bring out the differences of this production function with the Cobb-Douglas production function. [15]

3. Draw the concentration curve of per capita monthly total expenditure and per capita monthly expenditure on milk and milk products and compute the related concentration ratios from the following data. [30]

Cons. Exp. in Rs./person/30 days on all items (x_j) and milk and milk products (y_j)

exp class	p_j	$\bar{x}_j$	$\bar{y}_j$
0 - 8	9.03	6.18	0.08
8 - 11	14.73	9.49	0.21
11 - 13	11.33	12.01	0.46
13 - 15	9.87	13.98	0.54
15 - 18	13.05	16.44	0.93
18 - 21	9.48	19.53	1.32
21 - 24	8.49	22.46	1.73
24 - 28	6.96	25.66	2.23
28 - 34	6.17	30.34	2.86
34 - 43	5.34	38.26	4.24
43 - 55	2.75	48.23	5.66
55 -	2.80	89.45	7.01
all class	100.00	20.13	1.43

p_j is the percentage of persons falling in the j^{th} group.

Group B: Planning Techniques

Maximum Marks: 40 Suggested time: 1 hour
Attempt any two questions.

1. Examine some necessary and sufficient conditions for a bill of goods being produced in static input-output analysis. [20]

2. Explain how a linear programming problem involving inequalities in the constraints can be converted into one involving only equalities so that there is a one-to-one correspondence between the feasible solutions of the original problem and those of the reformulated problem. Further show that the optimal solutions in the two cases are the same. [20]

3. Solve the following linear programming problems graphically:

(a)

$$x_1 + x_2 \geq 1$$
$$x_2 - 5x_1 \leq 0$$
$$5x_2 - x_1 \geq 0$$
$$x_1 - x_2 \geq -1$$
$$x_1 + x_2 \leq 6$$
$$x_1 \leq 3$$
$$x_1, x_2 \geq 0$$
$$\text{Max. } Z = 3x_1 + 2x_2$$

(b)

$$x_1 + x_2 \leq 1$$
$$-0.5x_1 - 5x_2 \leq -10$$
$$x_1, x_2 \geq 0$$
$$\text{Max. } Z = -5x_2$$

[20]

[Graph Paper to be supplied]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70

[419]

PERIODICAL EXAMINATIONS

Subject: Statistics-7: Statistical Quality Control
Theory and Practical

Date: 30.3.70 Maximum Marks: 100 Time: 3 hours

Note: Answer Q. No.1 and any other four questions from the rest. Marks allotted for each question are given in brackets [ ].

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 establish standards.

b) A minimum of 8 per cent copper in any sheet is the market specification. Excess of 0.2 per cent on an average results in a loss of Rs.17000 per annum to the factory. Estimate how much saving can be effected by maintaining statistical control at a proper level so as to satisfy the market specification. [15 +10]=[25]

2.a) Distinguish between specification limits and control limits. Describe different situations for a process under control with reference to the corresponding two sided specification limits.

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.0 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]

3.a) What is a group control chart? Explain how would you construct a group control chart for a measurable characteristic of an item being produced on a group of 5 machines. Indicate your assumptions.

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]

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

5.a) Explain the following terms

i) AQL (ii) ASN (iii) AOQL.

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

Consumer's Risk = 0.10. [6+12]=[18]

6. Write short notes on following:

i) Rational subgrouping

ii) Modified control limits

iii) Group control charts

iv) Process in a state of statistical control. [18]

Neatness : 3.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [400]

PERIODICAL EXAMINATION

Statistics-8: Genetics

Date: 6.4.70 Maximum Marks: 100 Time: 3 hours

Note: Answer any four questions. Marks allotted for each question are given in brackets [].

1.a) State and explain the Mendelian Theory of Inheritance, bringing out the important features in it. [8]

b) What considerations led Mendel to select Pisum sativum for his famous experiments on heredity. [4]

c) Do you see some special reasons for Mendel's success in discovering the laws of inheritance which escaped the notice of generations of breeders? Give your critical comments. [5]

d) On the basis of Mendel's hypothesis, work out the results of the following experiments: (i) Yellow-wrinkled (GGww) X Green-round (ggWW); (ii) The progenies of (i) selfed; (iii) Progenies of (i) crossed with the original Yellow-wrinkled. In each cross, show the gametes produced, the genotypes and phenotypes of the offsprings and their relative proportions. [2+3+3]=[8]

2.a) What is Epistasis? Two loci are such that homozygous recessive (bb) at one locus suppresses the expression of the genotypes AA, Aa or aa at another locus not on the same chromosome. If AaBb X AaBb matings are effected in this organism, how many phenotypes will be expected and in what proportions? (A and B are dominant over their respective recessive alleles a and b). Illustrate your answer by a diagram. [4+6]=[10]

b) Distinguish between a recessive gene and a hypostatic gene? [2]

c) In one of Bateson and Punnet's experiments with sweet peas, they crossed 'white flower' plants of two varieties. Now both these 'white' breeds were pure and in each variety, the white was recessive to red. But the progenies produced by the mating were all with red flowers (the dominant type). The latter 'red flower' plants produced both red and white types in the ratio (9 red : 7 white). Explain these observations. [10]

d) Why these observations are regarded as a landmark in the development of Genetics? [3]

3.a) Explain Gynandromorphism and its occurrence in Drosophila melanogaster. Can this phenomenon occur in Man? Give reasons. [6+4]=[10]

b) Explain Bridge's concept of 'Balanced sex determination' in Drosophila melanogaster. What will be the sex in (XX+ 3A), (XXX+ 2A), (XO+ 3A). (XX+ 2A) and (XY+ 2A).

What is the function of the Y-chromosome in male Drosophila? [5+5]=[10]

c) How is an Intersex differentiated from a true Hermaphrodite? Give an example of each kind. [5]

4.a) Give Bernstein's theory of inheritance of ABO blood groups. How many gene loci are involved? [7]

b) In an Iowa study the following type frequencies were obtained:

O	A	B	AB	Total
743	672	147	61	1623

4. (continued)

Estimate the gene-frequencies in the population (Bernstein's improved formulae), calculate the expected phenotype numbers, and test if the sample indicated Hardy-Weingberg equilibrium in the population. [11]

c) In the ABO system of blood groups, do you know the genotype of an individual from his phenotype? If not, why? What other information may be of use in ascertaining the genotypes when they are not obvious from the respective phenotypes? Illustrate by examples. [2+3]=[5]

d) Do you think the discovery of an additional antiserum might be helpful in the matter? What should be the characteristic of that serum? [2].

5.a) State Hardy-Weinberg Rule of equilibrium. Starting with a population with the genotype proportions, AA (x), Aa (2y) and aa (z), which is random mating, establish the rule for an autosomal locus. [11]

b) Show how is the equilibrium established in a population for a completely sex-linked locus (alleles A and a), starting with the male population gene frequency (recessive, q_x) and the female population gene frequency (recessive, q_{xx}). What will be the equilibrium gene frequency in males and females? [14]

6.a) What do you know about the genetics of ABH secretion in saliva in man? How many loci and alleles there are involved? [7]

Is there any interaction between the Secretor locus and the ABO blood group locus? [1]

In a population sample of 400, there were 280 secretors. Find out the secretor/non-secretor gene-frequencies in the sample. [4]

b) In the M-N system of blood groups, prove that a population in Hardy-Weinberg equilibrium cannot possess more that 50 % of the phenotypes as MN theoretically. [4]

c) A woman, whose mother was group O, herself belonged to group B in the ABO blood group system. She was of group MN in the M-N system, and she married a man of group AB, MN. What is the chance that their first child would be (i) B, M; (ii) AB, MN; (iii) A, N? [6]

d) Give the generalised form of Hardy-Weinberg Rule for n alleles at an autosomal locus. [3]

INDIAN STATISTICAL INSTITUTE
Research and Training School

B.Stat. Part IV: 1969-70

ANNUAL EXAMINATION

Statistics-4: Probability.

Date: 18.5.70 Maximum Marks: 100 Time: 3 hours

Note: Answer as much as you can. Marks allotted for each question are given in brackets []. State clearly the results you use without proving. In questions 3,5,6 the transition probabilities are stationary. The maximum marks you can score in each group is 50.

Group A

1. State and prove Minkowsky's inequality. [8]

2.a) State Kolmogorov's inequality. Deduce that if $X_1, X_2, \ldots$ are independent with $E(X_k) = 0$ for all k and $\Sigma \operatorname{Var}(X_n) < \infty$, then $S_n = X_1 + \ldots + X_n$ converges almost surely. State the version of strong law of large numbers that follows from this. [14]

b) State a necessary and sufficient condition for almost sure convergence of S_n/n when $X_1, X_2, \ldots$ are i.i.d. [4]

3.a) When is a state of a Markov Chain called transient, positive or null? [6]

Show that if two states mutually communicate then both are simultaneously transient or simultaneously positive or simultaneously null. [6]

Show that if i is a recurrent state and j is accessible from i then i is accessible from j. [6]

b) State the fundamental limit theorem (on the limit of $p_{ij}^{(n)}$) for all pairs of states. [6]

c) Show that in a finite Markov Chain there is no null state and not all states are transient. [6]

d) Show that in a finite Markov Chain, a state j is transient if and only if there exists a state k such that k is accessible from j but j is not accessible from k. (Hint: Use (c) and the last part of (a).) [6]

e) If P is the transition probability matrix of a finite irreducible aperiodic Markov Chain, show that for some n, all elements of P^n are positive. [4]

Group B

4. State the Lindeberg-Feller central limit theorem and deduce from it the De Moivre-Laplace theorem. Can the condition $\max \frac{\sigma_k^2}{s_n^2} \to 0$ be dropped from the former? (Hint: consider the [illegible] distribution with mean 0 and variance $2(2^k - 1)$.) [4+6+6]=[16]

5.a) Suppose that in a Markov Chain, x_j denotes the probability of absorption into the recurrent class C when the initial state j is transient. Show that

$$x_j = \sum_{\lambda \in T} p_{j\lambda} x_\lambda + \sum_{k \in C} p_{jk} \quad \ldots (1)$$

where T is the set of transient states. [6]

b) If y_j denotes the probability of staying for ever in T when the initial state is j, show that

$$y_j = \sum_{\ell \epsilon T} p_{j\ell} \, y_\ell \qquad \ldots \quad (2)$$ [6]

c) Show that (1) has a unique solution if and only if the only bounded (non negative) solution of (2) is $y_j = 0$ for all j in T. Show that this is the case when the Markov Chain is finite. [6+6]=[12]

d) Find the probability of ruin of A when the initial fortune of A is j units, the fortune of B is infinite and A, B play a series of games as follows. At time n if A is not already ruined then A and B play a game in which A gains unit money with probability p and loses unit money with probability $q = 1-p$. Assume $q \geq p$. [6]

6. For a Markov chain with state-space 0,1,2,... and transition probabilities given by $p_{i0} = q_i$ and $p_{i,i+1} = 1 - q_i$, $0 < q_i < 1$, $i = 0,1,\ldots$, show that the states are transient if and only if $\Sigma q_i < \infty$, [6]

If $q_i = q$ for all i, show that the Markov Chain is positive and find the stationary distribution. [4+6]=[10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [422]

ANNUAL EXAMINATIONS

Statistics-4: Inference

Date: 19.5.70 Maximum Marks: 100 Time: 3 hours

Note: Answer groups A and B in separate answerscripts. Marks allotted for each question are given in brackets [ ].

Group A

Answer as many questions or parts thereof as you can. The maximum you can score in this group is 50.

1.a) What are similar tests? What are tests having Neyman structure? How are they both related? [8]

b) In a multiparameter family of exponential distributions with density function

$$f(X, \theta, \eta) = C(\theta\ \eta)\ \exp\ \theta\, U(X) + \sum_{j=1}^{k} \eta_j T_j(X)$$

Obtain the uniformly most powerful similar test for $H_0: \theta = \theta_0$ against the alternative $\theta > \theta_0$, with η unspecified. [10]

2.a) Describe locally most powerful tests and locally most powerful unbiased tests. Give methods of obtaining such tests. [12]

b) Obtain a locally most powerful unbiased test for testing $\mu = \mu_0$ in $N(\mu, 1)$. How does this compare with the UMPU test for the same hypothesis? Why? [6]

3.a) Explain clearly the place of OC and ASN functions in sequential testing. [10]

b) Define SPRT of strength (α, β) for testing a simple hypothesis against a simple alternative and indicate how you determine the boundaries. Also discuss broadly the consequences of the approximations you make. [8]

Group B

Answer as many questions or parts thereof as you can. The maximum you can score is 50.

1.a) Discuss the place of bounded completeness in testing composite hypotheses. [9]

b) Show that the usual t-test employed for testing the hypothesis $\xi = 0$ in a normal distribution with unknown mean ξ and unknown variance σ^2 is a uniformly most powerful unbiased test against the alternative $\xi > 0$. [9]

2. Let $f(X, \theta)$ be the frequency function of a random variable X and let $\theta' = (\theta_1, \ldots, \theta_k)$ be the unknown parameter ranging over $\mathbb{R}^k$. Derive a large sample test for the simple hypothesis $\theta' = \theta'_0 = (\theta_1^0, \ldots, \theta_k^0)$. State clearly the assumptions and the main results you make use of. [18]

3. In an acceptance sampling inspection of manufactured goods, it is required to test the hypothesis that the fraction of defectives in the manufactured goods is smaller than a preassigned quantity, say p_o. Devise a suitable sequential test for this purpose and derive its OC and ASN functions. [18]

4. Viva Voce. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70

ANNUAL EXAMINATIONS

Statistics-5: Statistical Methods Theory

Date: 21.5.70 Maximum Marks: 100 Time: 3 hours

Note: Answer groups A and B in separate answerscripts. Marks allotted for each question are given in brackets [ ]. Answer all the questions.

Group A

1. $\underline{X}' = (X_1, X_2, \ldots, X_p)$ is said to follow a nonsingular p-variate normal distribution $N_p(\underline{\mu}, \Sigma)$ if it has the following density function in $E_p(R)$.

$$\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp -\frac{1}{2}(\underline{X} - \underline{\mu})' \Sigma^{-1}(\underline{X} - \underline{\mu})$$

where Σ is a p.d. matrix.

(a) Show that $\underline{X}'$ has a $N_p(\underline{\mu}, \Sigma)$ distribution iff $\underline{\lambda}'\underline{X}$ is univariate normal $N(\underline{\lambda}'\underline{\mu}, \underline{\lambda}'\Sigma\underline{\lambda})$ for every vector $\underline{\lambda}$.

(b) Give an example to show that any arbitrary but fixed finite number of linear functions of $X_1, X_2, \ldots, X_p$ could have univariate normal distribution though the distribution of $X' = (X_1, X_2, \ldots, X_p)$ need not be p-variate normal.

(c) Show that if $\underline{X}$ has a $N_p(\underline{\mu}, \Sigma)$ distribution then $\underline{\mu}$ and Σ are the mean vector and dispersion matrix of $\underline{X}$. [25]

2. a) Define a matrix variate beta distribution $B_p(\frac{n_1}{2}, \frac{n_2}{2})$.

b) Show that if $U_1 \sim B_p(\frac{n_1}{2}, \frac{n_2}{2})$ and $U_2 \sim B_p(\frac{n_1+n_2}{2}, \frac{n_3}{2})$ where U_1 and U_2 are independently distributed, then

$$(U_2)^{1/2} U_1 [(U_2)^{1/2}]' \sim B_p(\frac{n_1}{2}, \frac{n_2+n_3}{2}).$$

c) Show that if $U \sim B_p(\frac{n_1}{2}, \frac{n_2}{2})$ then for each fixed vector a such that $a'a = 1$, $a'Ua \sim B(\frac{n_1}{2}, \frac{n_2}{2})$. [25]

Group B

1. Let $U' = (u_{ij})$ be a $p \times n$ matrix of random variables, the columns of which are independently distributed as p-variate normal with the same dispersion matrix Σ. Further

$$E(U) = X\Theta$$

where X is a known matrix of order $n \times m$ and Θ is a matrix of unknown parameters of order $m \times p$.

Consider the hypothesis $H\Theta = 0$ (a null matrix of order $k \times p$) where $\mathcal{M}(H') \subset \mathcal{M}(X')$. Derive a testing procedure for testing the above hypothesis justifying the steps involved in the derivation. [25]

2. Let $U' = (u_{ij})_{p \times n}$ be a matrix the columns of which are independently and identically distributed as $N_p(\underline{\mu}, \Sigma)$.

(a) Describe the Hotelling T^2 test for the hypothesis $\underline{\mu} = \underline{0}$.

(b) Show how one could improve upon this test if $\Sigma = (\sigma_{ij})$ is known to have the following structure $\sigma_{ii} = \sigma^2$ for each i, $\sigma_{ij} = \rho\,\sigma^2$ for each $i, j,\ i \neq j$.

(c) Describe a practical situation where the model of 4(b) would be appropriate. [25]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70

ANNUAL EXAMINATIONS

Statistics-5: Statistical Methods Practical

Date: 22.5.70 Maximum Marks: 100 Time: 3 hours

Note: Marks allotted for each question are given in brackets [ ].

Group A

Maximum Marks: 80
Answer any part of any question.

1. An objective of the Scientific Career Study is the determination of personality dimension associated with different post college career decisions of college science majors. Over 200 freshmen were selected from 6 different colleges in eastern Massachusetts and were administered the study of values test and other personality instruments.

After 3 years of followup three interesting groups were found when the post college plans of a sample of seniors were examined. The groups are as follows:

(1) Research Group; (2) Applied Science Group;
(3) Nonscience Group.

The mean scores of the three groups in the tests Theoretical (X_1), Economic (X_2), Aesthetic (X_3) and Social (X_4) and estimates of dispersion matrices are given below.

Table 1.1: Study of Values: Means.

Values	Research Group $n_1 = 33$	Applied Science Group $n_2 = 25$	Non-science Group $n_3 = 38$	Total Sample $n = 96$
1. Theoretical	56.73	51.56	48.95	52.30
2. Economic	33.24	41.52	38.21	37.36
3. Aesthetic	41.55	36.60	34.95	38.17
4. Social	34.58	29.84	36.58	34.14

Table 1.2: Estimated Dispersion Matrix based on the total sample.

	X_1	X_2	X_3	X_4
X_1	36.03	- 3.65	10.54	- 9.71
X_2		54.80	-24.50	-27.22
X_3			86.02	-12.14
X_4				56.37

i).a) Are the scores in the different tests useful in distinguishing the three groups? [35]

b) Is there any evidence to conclude that the pair (X_1, X_2) is distributed independently of the pair (X_3, X_4)? [10]

ii).a) Find the best linear discriminant function between the Research group and Applied Science group. [15]

b) Can we here regard $2X_1 - X_2 + X_3 + 2X_4$ as the true discriminant function between the Research group and Applied Science group? [15]

c) To which of the two groups Research and Applied Science will you classify a student with scores 52, 35, 43 and 36 in Theoretical, Economic, Aesthetic and Social tests respectively. [10]

2. The following gives the heart weight in grams of 12 female and 15 male cats.

<u>7.1</u>, <u>7.2</u>, <u>7.3</u>, <u>7.4</u>, <u>7.6</u>, 7.6, 8.0, 8.3, <u>8.7</u>, <u>9.0</u>, 9.1, 9.3
9.4, <u>9.5</u>, <u>9.5</u>, 9.6, <u>10.1</u>, <u>10.1</u>, <u>10.2</u>, 10.7,
11.2, 11.7, 12.7, 12.8, 13.6, 14.9, 15.6

where the figures for female cats are underlined. Is there any evidence to regard the samples as coming from the same population? [15]

Group B

3. Practical Records [10]

Group C

4. Viva Voce [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70

ANNUAL EXAMINATIONS

Statistics-7: Planning Techniques

Date: 23.5.70 Maximum Marks: 50 Time: 2 hours

Note: Answer any three questions. Marks allotted for each question are given in brackets [ ].

1. Examine the necessary and sufficient conditions for a bill of goods being producible in Leontief's Static system. [16]

2. Use the basic assumptions of the Leontief dynamic system to obtain the schedule of net output possibilities for a two-commodity economy with a one-period production plan and given initial stocks. Explain in this connection the concept of efficiency locus and show how the derivation of such a locus can be viewed as a problem in linear programming. [16]

3. Briefly explain the simplex method for solving linear programming problems. [16]

4. If X is a feasible solution of the primal and W is a feasible solution of the dual problem such that the values of the objective functions in the two cases are equal, prove that X and W are optimal solutions to to their respective problems.

If a primal linear programming problem has an optimal solution, show that the dual also has an optimal solution and the corresponding values of the objective functions to the two problems are equal. [16]

5. Solve the following problem by the Simplex method:

Maximize $Z = 60\xi_1 + 60\xi_2 + 90\xi_3 + 90\xi_4$

subject to

$$100\xi_1 + 100\xi_2 + 100\xi_3 + 100\xi_4 \leq 1500$$
$$63\xi_1 + 45\xi_2 + 27\xi_3 + 18\xi_4 \leq 900$$
$$3\xi_1 - 5\xi_2 - 10\xi_3 - 15\xi_4 \geq -100$$
$$\xi_1, \xi_2, \xi_3, \xi_4 \geq 0$$

Also formulate its dual and give its solutions. [16]

Two marks are reserved for neatness.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70 [426]

ANNUAL EXAMINATIONS

Statistics-6: Design of Experiments Theory and Practical

Date: 25.5.70 Maximum Marks: 100 Time: 3 hours

Note: Answer Groups A and B in separate answerscripts. Marks allotted for each question are given in brackets [ ].

Group A

Answer question No.1 and any two questions from the rest.

1.a) 'Randomisation is a method by which every experimental unit has an equal chance of receiving a treatment'. Discuss this statement in the context of different designs. [6]

b) Clearly state the restrictions which are being imposed on the number of treatments and the number of replications of a treatment as we pass from a completely randomised design to a randomised block design and then to a latin square design. Also state how at the sacrifice of flexibility greater control over error is achieved in the above designs. [8]

2. In a garden, r of the N plants were chosen at random for which a treatment T_1 is given and for the remaining a different treatment T_2 is given. The yields are observed after a period of two months. Obtain an unbiased estimate of variance for estimating the difference of the treatment effects by their observed mean yields assuming a suitable randomisation model. [13]

3.a) Define a balanced incomplete block design (BIBD). [2]

b) For a BIBD, prove that $b \geq t$ where the symbols are used in their usual sense. [8]

c) When is a BIB design called resovable? Can you improve upon the inequality in b above when the design is resolvable? [3]

4.a) Discuss relative merits of factorial experiments and one-factor-at-a-time experiments. [6]

b) Show that for conducting a 2^4 experiment in block of 2^2 units each, there cannot be any design such that no main effects and no two factor interactions are confounded. [7]

5. Answer any two:

a) 'Missing plot technique is a scientific way of getting back the lost observation'. Comment on this statement.

b) What is the principle of analysis of covariance? What are the conditions on the choice of concomitant variables?

c) Explain the role of transformations on experimental data.

d) What is a split-plot design? Why is it said that this design confounds main effects?

$[6\frac{1}{2}+6\frac{1}{2}]=[13]$

Group B

Answer any two questions.

1.a) What design do you suggest and why, given the following information? Eight roasts can be cut from each of 4 animals. The experimenter wishes to study the effect of freezing, length of freezing, storage temperatures and length of storage upon tenderness of roasts and to use the following in all combinations:

storage temperatures	:	10°c	and	15°c
lengths of storage	:	20	and	40 days
freezing temperatures	:	0°c	and	-10°c
lengths of freezing	:	5	and	10 days

Give a plan of the design and the appropriate break down of the total degrees of freedom. [2+2+5+3]=[12]

b) Four surface treatments each at 2 different intensities of application are to be studied in all combinations with respect to the durability of surface of motor car tyres. It is decided to consider each tyre to be of homogeneous material and different tyres to be heterogeneous. On four consecutive parts on the surface of a tyre any four of the 2^4 treatment combinations can be applied. You are allowed to use as many automobiles as you need. All main effects and interactions of all types are of interest. Write down any plan for conducting the experiment with the smallest number of cars which consists of driving the car for 100 miles on a standard road and noting down the difference in resistences of the surface before and after the experiment. [8]

2. The following table gives the yield (Y) of tea plants as observed in a randomised block experiment carried in four blocks of four plots each. Shown in parentheses are the preliminary yields (X) recorded on the same plants.

Block	Treatments			
	A	B	C	D
1	91 (85)	88 (81)	88 (90)	102 (93)
2	118 (121)	94 (93)	110 (106)	109 (114)
3	109 (114)	105 (106)	115 (111)	94 (93)
4	102 (107)	91 (92)	96 (102)	88 (92)

Examine if the treatment effects are significantly different, correcting for difference in preliminary yields. [Total (corrected) sum of squares and products:

$S_{YY} = 1526.00$, $S_{XX} = 2040.00$, $S_{XY} = 1612.00$] [20]

3. The following is a 5×5 Latin square for data taken from a manurial experiment with sugarcane. The five treatments were as follows:

A : No manure,

B : an inorganic manure,

C, D and E : three levels of a farm-yard manure.

Plan and Yield of sugarcane (in a suitable unit) per plot.

Row	Column I	II	III	IV	V	Total
I	A 52.5	E 46.3	D 44.1	C 48.1	B 40.9	231.9
II	D 44.2	B 42.9	A 51.3	E 49.3	C 32.6	220.3
III	B 49.1	A 47.3	C 38.1	D 41.0	E 47.2	222.7
IV	C 43.2	D 42.5	E 67.2	B 55.1	A 45.3	253.3
V	E 47.0	C 43.2	B 46.7	A 46.0	D 43.2	226.1
Total	236.0	222.2	247.4	239.5	209.2	1154.3

Total (crude) S.S. = 54273.51

a) Estimate the treatment effects and their standard errors. [4+2]=[6]

b) Examine whether these effects differ significantly from one another. [14]

Group C

Viva Voce [10]

Group D

Practical Record Note [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School

[427]

B.Stat. Part IV: 1969-70

ANNUAL EXAMINATIONS

Statistics-7: Industrial Statistics Theory and Practical

Date: 27.5.70 Maximum Marks: 100 Time: $3\frac{1}{2}$ hours

Note: Answer groups A and B in separate answerscripts. Marks allotted for each question are given in brackets [].

Group A: Maximum Marks: 40

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

1.a) What is an extreme point of a convex set? Show that the objective function of a linear programming problem attains the optimal value, if it exists, at an extreme point of the convex set of its feasible solutions.

b) Show that by every iteration of the simplex procedure the net decrease in the value of objective function is $\theta\,(Z_k - C_k)$ where $Z_k = C_B B^{-1} a_k$. Hence derive the condition for optimality. What conclusion will you draw if all $y_{ik} \leq 0$?

c) State the rules of transformation to get the next simplex tableau. [7+6+3]=[16]

2.a) Explain the meaning of the following terms:

(i) Artificial variables, (ii) Optimal basic feasible solution, (iii) Non-degenerate basic feasible solution.

b) Show that if an optimal solution to a linear programming problem exists, such a solution contains at most m positive variables where m denotes the number of basic variables. [3+9]=[12]

3.a) Show that if the primal problem has an optimal solution then the dual also has an optimal solution. Explain clearly how you would obtain a solution to the dual if the primal is solved by the revised simplex method form I.

b) Consider the following linear programming problem:

maximise $Z = 3x_1 + 2x_2$

subject to
$$x_1 + x_2 \geq 1$$
$$x_1 + 2x_2 \leq 7$$
$$x_1 + 2x_2 \geq 10$$
$$x_1,\ x_2 \geq 0$$

i) Write the dual problem.

ii) Solve the dual and hence solve the primal. [4+8]=[12]

4. A company has two types of aircrafts X and Y with equal loading capacities and the following operating costs when loaded

Types of aircraft	Cost per mile (n.m.)
X	2.00
Y	3.00

The present location of the four aircrafts is as follows:

Location	Type of aircraft
J	X
K	Y
L	Y
M	X

Four customers of the company located at A,B,C and D want to transport about the same size of load to their final destinations at a distance of 600, 300, 1000 and 500 miles respectively. The total costs of flying empty aircrafts from the location points (J,K,L and M) to the loading points (A,B,C and D) are as follows.

Location	Loading point			
	A	B	C	D
J	300	200	400	100
K	450	150	450	300
L	600	150	150	750
M	200	200	400	200

Determine the allocation which minimizes the total cost of transportation. [12]

Group B: Maximum Marks: 40

Answer Q.5 and any other two from the rest.

5.a) Write down clearly any five practical situations where waiting line problems arise.

b) Under Poisson process, show that the generating function of $p_n(t)$ is given by $P(Z,t) = e^{\lambda(Z-1)t}$ where $p_n(t)$ is the probability of occurrence of n events in a time interval of length t and λ is the average rate of occurrence of the events. Hence show that

$$p_r(t) = \frac{(\lambda t)^r}{r!} e^{-\lambda t} .$$

c) If the customers arrive in Poisson fashion at an average of λ customers per unit of time and the service time follows the negative exponential distribution with mean $1/\mu$, obtain the distribution of V, the total time spent in the system by a random arrival assuming statistical equilibrium. Hence obtain E(V).

d) At a public telephone booth, arrivals are considered to follow the Poisson law with an average inter-arrival time of 15 minutes. The length of a telephone call is assumed to be distributed negative exponentially with an average of 5 minutes.

i) What percentage of arrivals will have to wait for service?

ii) What is the average waiting time?

iii) What is the average waiting time for an individual given that one has to wait before getting the phone? [16]

6. In the case of the M/M/C waiting line model with an average of λ arrivals per unit time and average service time of $1/\mu$ under steady state:

a) Derive the expressions for the state probabilities p_n.

b) Show that the average queue length is

[8+4]=[12]

7. For a queuing model with random input and general service time distribution (M/G/1) show that when the system is in a state of statistical equilibrium the average number of customers in the system is

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

and the average waiting time is $E(w) = \frac{\rho^2 + \lambda^2 \sigma_v^2}{2\lambda(1-\rho)}$ where $\rho = \lambda/\mu$, λ = mean arrival rate, μ: mean number of customers serviced in unit time and σ_v^2 is the variance of the service time of a customer. [12]

8.a) There are N identical machines in a workshop. An average of λ machines break down per unit of time and a breakdown requires on an average $1/\mu$ time units for its repair by a service engineer. A crew of $C < N$ service engineers is employed for the shop. Assume Poisson input and negative exponential service time distribution and show that under steady state

$$p_n = p_0 \rho^n \binom{N}{n} \qquad 1 \leq n < c$$

$$= p_0 \frac{\rho^n \, n!}{c! \, c^{n-c}} \binom{N}{n} \qquad c \leq n \leq N$$

where $\rho = \lambda/\mu$.

b) For a group of 7 identical machines it was found that the machines will operate for an average of 80 hours and then require an average of 40 hours repair. If the profit per running machine hour is Rs.15/- and the cost per serviceman per hour is Rs.4/- how many serviceman should be engaged to maximize the net profit? [12]

Group C

Practical records. [10]

Group D

Viva Voce. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1969-70

Statistics-7: Econometrics Theory and Practical

Date: 29.5.70 Maximum Marks: 100 Time: 3 hours

Note: Answer groups A and B in separate answerscripts. Marks allotted for each question are given in brackets [ ]. The maximum you can score in groups A and B is 80.

Group A: Maximum Marks: 56

Answer question 4 and any other two of the rest.

1. Assuming that $X \cap \Lambda(\mu, \sigma^2)$ derive the expression for for Lorenz concentration coefficient - L. If $\mu = 3.0$ and $\sigma = 5.0$ find out the value of L. [12]

2. Discuss the concept of the production function. Specify the mathematical form of the Cobb-Douglas production function. What was the striking feature in the estimates obtained by Cobb and Douglas? [12]

3. What is a concentration curve? Discuss its possible uses in econometrics. [12]

4. The Engel curve for cereals was estimated to be

$y = 0.44x^{0.53}$, where y = per capita monthly expenditure on cereals
x = per capita monthly total expenditure

in the year 1956-57 in India.

The following table gives the percentage distribution of persons by per capita monthly total expenditure classes as also average per capita monthly total expenditure in each class.

Percentage of persons and per capita monthly total expenditure by per capita expenditure class, 1956-57, India

expenditure class (Rs.) (1)	p.c. of persons (2)	per capita total expenditure (Rs.0.00) (3)
0 - 11	23.76	8.23
11 - 15	21.20	12.93
15 - 21	22.53	17.74
21 - 28	15.45	23.90
28 - 43	11.51	34.01
43 -	5.55	69.03

Assuming that during the 1957-58 the Engel curve remains same and that average total per capita expenditure rises by 10 p.c. over 1956-57 level as also concentration of total per capita expenditure declines by 10 p.c. over the level of 1956-57 and that the distribution of total expenditure was log-normal in 1956-57 and remain of the same form in 1957-58, estimate the change in total demand for cereals. [32]

Group B: Maximum Marks: 24

Answer any two questions.

1. What assumptions are made about the error term in fitting a linear regression by least squares method? Are these assumptions justified in the case of economic data?

What procedure do you adopt in fitting a linear regression when the error term is autocorrelated? [12]

2. What do you mean by multicollinearity in regression analysis? What difficulties does it create in estimating the parameters and their variances? How will you remove these difficulties? [12]

3. Formulate a two-equation linear model, and test for the identifiability of the equations. Show that if the parameters of one of the relationships is estimated by least squares method (disregarding the other relation), the estimates will not be consistent. What method will you adopt for getting a consistent estimate? [12]

Group C

Practical Record [10]

Group D

Viva-Voce [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1969-70

[429]

ANNUAL EXAMINATIONS

Statistics-8: Genetics

Date: 30.5.70 Maximum Marks: 100 Time: 3 hours

Note: Answer groups A and B in separate answerscripts. Marks allotted for each question are given in brackets [].

Group A: Maximum Marks: 50

Answer question 4 and any other two of the rest.

1. Explain with illustrations, one for each, what you understand by

(a) Sex-linked, (b) Sex-limited, and
(c) Sex-influenced characters. [5+7+8]=[20]

2. A random mating population is initially

$$p^2AA + 2pq\,Aa + q^2aa$$

and the relative 'fitness' of AA, Aa and aa are respectively $(1-s_1)$, 1, and $(1-s_2)$.

(a) Prove that an equilibrium is reached when

$$s_1 p = s_2 q, \quad p = \frac{s_2}{s_1 + s_2} \quad \text{and} \quad q = \frac{s_1}{s_1 + s_2}.$$

(b) In a population with heterozygote advantage, with selection coefficients, .005 and .01 against the homozygous dominants and the recessives respectively, find out the gene frequencies after the polymorphism is balanced. [15+5]=[20]

3.a) The population, (rAA + 2sAa + taa), starts practicing complete positive assortative mating. (A is completely dominant over a). Prove that,

$$H_n = \frac{2p\,H_0}{2p + nH_0}, \quad \text{where } H_0 = \text{Initial heterozygote}$$

frequency, H_n = heterozygote frequency after n generations of the assortative mating.

b) Given a population, (.10, .40, .50), calculate the heterozygote frequencies for the next three generations on complete possitive assortative mating. [14+6]=[20]

4. Estimate the gene frequencies and test if the given data indicate Hardy-Weinberg equilibrium at the ABO blood group locus:

Blood groups ->	O	A	B	AB	Total
obs. no.	59	64	74	28	225
obs. %.	26.22	28.45	32.89	12.44	100.00

[10]

Group B: Maximum Marks: 50

Answer Q.4 and any other two of the rest.

1.a) Prove that selfing in a population does not affect the gene frequencies, and that as

$$n \to \infty, \quad D_n \to (D_o + \tfrac{1}{2} H_o), \text{ and } R_n \to (R_o + \tfrac{1}{2} H_o),$$

where n = the number of generations of selfing, the initial population being (D_o, H_o, R_o) and the population after n generations of selfing being (D_n, H_n, R_n).
[D-homozygous dominants, R-homozygous recessives, H-heterozygotes].

b) Prove that in successive generations of sib-mating the heterozygote frequency ($\bar{H}$) declines according to the relation,

$$\bar{H}_{n+2} = \tfrac{1}{2}\bar{H}_{n+1} + \tfrac{1}{4}\bar{H}_n$$ (n, n+1, n+2 represent numbers of generations) [6+14]=[20]

2.a) Explain clearly the difference between the two types of twins and justify the adjucts 'Identical' and 'Fraternal' for them. What are 'Siamese' or conjoined twins?

b) How can you interpret the observed ratios, 1:1:1 in the caucasoid and 2:1:2 in the mongoloid for MM:MF:FF sex-combinations in the twin birth statistics. [10+10]=[20]

3.a) In the U.S. 'coloured' population, the following birth statistics were noted in a certain census period:

Single births (thousand)		Twins (thousand)	
Male	= 1,803	MM	= 16.3
Female	= 1,709	MF	= 17.9
		FF	= 16.1

Estimate the numbers of monozygotic and dizygotic twin pairs.

b) Early baldness (30-40 years) was observed in 7.75 % of the females and in 42.96 % of the males in a population in hardy-Weinberg equilibrium. Estimate the frequencies of the two alleles controlling early baldness. [10+10]=[20]

4. The following M-N blood groups data were obtained in a genetic survey of the Bado Gadaba tribe. Estimate the gene frequencies and test if the data are consistent with Hardy-Weinberg equilibrium:

M-N blood groups in Bado-Gadaba

	M	MN	N	Total
Obs. no.	122	78	12	212

[10]
