INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 | 122A |

PERIODICAL EXAMINATIONS

Statistics-4: Inference

Date: 14.10.68 | Maximum Marks: 50 | Time: $1\frac{1}{2}$ hours

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

1. For a monotone likelihood ratio family show that one sided most powerful tests exist for all sizes α. State your arguments carefully. (You may use the Neyman-Pearson Lemma.) [16]

2. Let X be a real valued random variable with density $f_0(x)$ or $f_1(x)$ where

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty$$

and

$$f_1(x) = \frac{1}{2}\left\{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-1)^2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+1)^2}\right\}$$

Find the M.P. test of $H_0(f = f_0)$ vs. $H_1(f = f_1)$ of size $\alpha = .05$. [16]

3. Let X have the Cauchy density

$$f_\rho(x) = \frac{1}{\pi}\,\frac{1}{1+(x-\theta)^2} \qquad -\infty < x < \infty,\ -\infty < \theta < \infty.$$

Find the locally most powerful test of $H_0(\theta = 1)$ vs. $H_1(\theta > 1)$ and show its power $\to 0$ as $\theta \to \infty$. [16]

4. Identify the minimal sufficient statistic in the following cases:

a) $X_1, \ldots, X_n$ are i.i.d. with common density

$$f_\theta(r) = \theta_i \quad \text{if } r = i,\ i = 1, \ldots, k;$$

here $\theta = (\theta_1, \ldots, \theta_k)$ and $\Omega = \left\{\theta;\ \sum_1^k \theta_i = 1, \theta_i \geq 0\right\}$.

b) $X_1, \ldots, X_n$ are as in (a) but Ω contains only two points $(\frac{1}{k}, \ldots, \frac{1}{k})$ and $(\frac{1}{k} - \frac{1}{k^2}, \frac{1}{k} + \frac{1}{k^2}, \frac{1}{k}, \ldots, \frac{1}{k})$.

5. For neatness [2]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV:1968-69 [1223]

PERIODICAL EXAMINATIONS

Statistics-6: Demography Theory and Practical

Date:14.10.68 Maximum Marks: 50 Time: $1\frac{1}{2}$ hours

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

1. What are the different columns in an abridged life table? How are they [illegible]? Deduce the relationship between ${}_nq_x$ and ${}_nm_x$. Stating clearly the assumption involved. [18]

2. You are given the following data taken from occupational mortality investigation.

Age group	Standard population: population at risk	Standard population: q_x	Occupation X: population	Occupation X: q_x	Occupation Y: population	Occupation Y: q_x
15-24	270,000	.001	4,000	.005	13,000	.00[illegible]
25-34	310,000	.002	16,000	.002	20,000	.001
35-44	350,000	.003	28,000	.002	33,000	.002
45-54	320,000	.008	33,000	.007	29,000	.008
55-64	250,000	.022	28,000	.021	23,000	.025

The mortality experiences of occupations X and Y may be compared by comparing the standardised death rates by 1) direct method (2) by indirect method. Calculate the death rates by each of these methods and state with reasons to what extent you think a reliable comparison between the occupations is obtained. [17]

3. Briefly discuss any three of the following:

a) 'de facto' and 'de jure' population enumeration,

b) methods of identifying members of economically active population,

c) deficiencies in Indian vital registration statistics and suggestions for improving them,

d) errors in census data. [15]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat, Part IV: 1968-69

123A

PERIODICAL EXAMINATIONS

Statistics-7: Econometrics Theory and Practical

Date: 21.10.68 Maximum Marks: 50 Time: $1\frac{1}{2}$ hours

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

1. How and when the question of multi-collinearity becomes important in connection with the estimation of a demand function. Suggest a method to get over this difficult situation. [10]

2. Give a statement of the Cob-Web model of demand and supply, of one commodity only. Examine the identifiability of the equations of the model. Derive the time path of the equilibrium price and give your comments. [15]

3. From the following data draw the concentration curve for the total per capita consumer expenditure, and compute the Lorenz ratio from the same data.

Table (1): Per capita monthly total consumer expenditure (Rs.) by classes of expenditure level, with percent of persons in each class, all-India, Urban, 1953-54.

per capita monthly expenditure class (Rs.)	percentage of persons	per capita monthly total consumer expenditure in (Rs.)
(1)	(2)	(3)
0 - 8	7.52	6.24
8 - 11	12.09	9.36
11 - 13	8.56	11.92
13 - 15	9.29	14.01
15 - 18	11.36	16.27
18 - 21	10.44	18.98
21 - 24	7.79	22.59
24 - 28	8.32	25.64
28 - 34	5.41	30.67
34 - 43	7.85	38.15
43 - 55	4.86	48.70
55 and above	6.51	80.33

[25]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

[123B]

PERIODICAL EXAMINATIONS

Statistics-7: Planning Techniques

Date: 21.10.68 Maximum Marks: 50 Time: $1\frac{1}{2}$ hours

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

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

b) State the basic assumptions of input-output analysis. [18+7]=[25]

2. Discuss the necessary and sufficient conditions for a bill of goods being producible under the Leontief static system. [25]

3. Solve the following linear programming problems graphically and shade the region representing the feasible solutions:

(a)
$$2x_1 + 3x_2 \leq 6$$
$$x_1 + 4x_2 \leq 4$$
$$x_1, x_2 \geq 0$$
$$\text{max. } Z = x_1 + \frac{3}{2} x_2$$

(b)
$$5x_1 + 10x_2 \leq 50$$
$$x_1 + x_2 \geq 1$$
$$x_2 \leq 4$$
$$x_1, x_2 \geq 0$$
$$\text{Min. } Z = 2x_1 + x_2$$

[25]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

[124A]

PERIODICAL EXAMINATIONS

Statistics-4: Probability

Date: 28.10.68 Maximum Marks: 50 Time: $1\frac{1}{2}$ hour

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

1.a) What is a stochastic matrix ? [4]

b) Prove that if A is a stochastic matrix, A^2 is also one. [8]

c) If A^2 is a stochastic matrix, does it follow that A is a stochastic matrix? Give a proof or a counter-example, whichever may be relevant. [8]

d) A is a matrix in which

i) every entry is ≥ 0 , and

ii) the sum of the elements in each row is ≤ 1. A^2 is a stochastic matrix.

Prove that A is also a stochastic matrix. [8]

2.a) P is a stochastic matrix such that P^{10} has only positive entries in the last column. Prove that

$$\lim_n P^n = Q$$

exists and has identical rows. State and prove the needed lemma on weighted averages. [10+4]=[14]

b) Give a stochastic matrix P such that every column in every power of P contains at least one zero and such that $[P^n]$ converges to a limit matrix as (as $n \longrightarrow \infty$). [8]

INDIAN STATISTICAL INSTITUTE [1243]
Research and Training School
B.Stat. Part IV: 1968-69

PERIODICAL EXAMINATIONS

Statistics-4: Sample Surveys Theory
and Practical

Date: 28.10.68 Maximum Marks: 50 Time: $1\frac{1}{2}$ hours

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

1.a) Prove that the probability of selection of a specified unit at a specified draw is equal to $1/N$ for s.r.s. without replacement where N is the population size.

b) Prove that the sample mean is the best linear unbiased estimator of the population mean in the subclass $\sum_{1}^{n} a_r y'_r$ where a_r $(r = 1,2,\ldots, n)$ is the coefficient to be attached to the variate value of the unit appearing at the r-th draw and y'_r is the value of the unit drawn at the r-th draw, for srs without replacement.

c) Assuming that the finite population is a random sample from an infinite, normal super-population, derive exact confidence limits for the population mean. [5+7+5]=[17]

2.a) Derive an exact upper bound for the bias of the classical ratio estimator.

b) Derive an approximately unbiased ratio estimator whose asymptotic bias does not contain terms of order n^{-1} and N^{-1} for srs without replacement.

c) Suppose there are M domains in the population. Derive the estimator of $\sum_{j=1}^{M} \lambda_j \bar{Y}_j$ and derive its variance for s.r.s. without replacement where $\sum_{1}^{M} \lambda_j = 0$ and $_j\bar{Y}$ is the j^{th} domain population mean. [5+8+5]=[18]

3. From a population of size 120 a sample of 10 is drawn with s.r.s. without replacement. The values of y and x of two characteristics measured on each of them are as follows:

Unit No.	1	2	3	4	5	6	7	8	9	10
y	2	4	2	4	3	6	7	5	4	1
x	65	125	54	120	95	78	230	150	118	40

The population is divided into 4 domains. The sample units in those 4 domains are as given below.

Domain:	1	2	3	4
Unit No.	(1,3,10)	(2,5)	(4,6,8,9)	(7)

Given that the population mean of x for domain 3 is 153 estimate by ratio method the population mean of y for that domain. Estimate its efficiency over the usual unbiased estimate. Give rationale for using ratio method here. [10]

4. Assignments. [5]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

[125]

MID-YEAR EXAMINATIONS

Statistics-4: Probability

Date: 18.12.68 Maximum Marks: 100 Time: 3 hours

Note: Answer as much as you can. Marks allotted for each question are given in brackets [ ].

1.a) There are 12 states in a Markov chain. In the stochastic matrix, there are positive entries in the following cells and zero in every other cell. Determine the transient states, ergodic classes and cyclically moving sub-classes in each ergodic class.

(1,4), (1,5), (2,1), (3,2), (3,6), (4,6), (5,3), (6,1), (6,7), (7,4), , , (8,1), (8,9), (9,10), (9,11), (10,11), (11,12), (12,9). [12]

b) Same as above but (1,8) and (9,3) also contain positive entries (in addition to the 17 cells of the previous problem). You are advised to draw a completely new network for this second problem, otherwise there may be confusion. [12]

2.a)

$$P = \begin{matrix} .6 & .3 & .1 \\ .2 & 0 & .8 \\ 0 & .6 & .4 \end{matrix}$$

obtain $\lim_n P^n$. Describe your method; proofs (for the theorems used) need not be given. [12]

b) Determine any nonsingular stochastic matrix P such that

$$\lim_n P^n \text{ is } Q = \begin{matrix} .5 & .2 & .3 \\ .5 & .2 & .3 \\ .5 & .2 & .3 \end{matrix}$$ [12]

c) Determine a nonsingular stochastic matrix P such that

i) $\lim_n P^n$ is the Q above, and

ii) the entry in the (1,1)- cell is .3. [8]

3.a) Prove that every real eigen value of a stochastic matrix lies on the interval [-1, +1]. [10]

b) Show that no stochastic matrix of order 2 X 2 can have the eigen values -1, -1. [12]

c) Give an example of a stochastic matrix whose eigen values are

i) +1, +1,

ii) +1, -1. [5+5]=[10]

d) Form the maximum weighted average of 1, 6 and 4 satisfying the following conditions:

i) As usual, the weights w_1, w_2, w_3 are all nonnegative and $w_1 + w_2 + w_3 = 1$;

and (ii) $w_1 \geq w_2$, $w_1 \geq w_3$.

You must prove conclusively that your average is the greatest possible. [10]

4.a) What is a transient state? [2]

b) Prove that if j is a consequent of i and k is a consequent of j, then k is a consequent of ? [7]

c) Prove that every consequent of a nontransient state is

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69
MID-YEAR EXAMINATIONS
Statistics-4: Inference

[126]

Date: 19.12.68 Maximum Marks: 100 Time: 3 hours

Note: Answer questions 7 and 8 and any three other questions. Marks allotted for each question are given in brackets [ ].

1. Derive the UMP unbiased test of size α for independence in a 2×2 contingency table. [16]

2. Let $X_{i1},\ldots,X_{i,n_i}$, $i = 1,2$ be independent normal with mean μ_i and variance σ^2. Find the MP similar test of size α of $H_0(\mu_1 = \mu_2)$ against $H_1(\mu_1 < \mu_2)$. [16]

3. Let $X_1,\ldots,X_n$ be independent, $N(\mu,\sigma^2)$. Show that a MP test of size α exists for $H_0(\sigma^2 = \sigma_0^2)$ vs. $H_1(\sigma^2 = \sigma_1^2)$ if $\sigma_1^2 > \sigma_0^2$ but no such test exists if $\sigma_1^2 < \sigma_0^2$. Does a MP similar test of size α exist for the latter case? [16]

4. Define Fraser-sufficiency. Show how it can be used to get a MP similar test of size α in the following problem. $X_1,\ldots,X_n$ are i.i.d with a continuous distribution function $F(x)$. Test H_0(Median = 0) vs. H_1(median > 0). [16]

5. Let $X_1,\ldots,X_n$ be i.i.d with common density $f_\theta(x) = k(\theta)e^{\theta x}\psi(x)$. Describe the Bayes solutions in a two action problem if
   i) the loss-difference has one sign-change
   ii) the loss-difference has two sign-changes
   (You have to prove all the results you need). [8+8]=[16]

6. a) Discuss briefly the use of bounded completeness in testing.
   b) Show that a sufficient statistic with boundedly complete family of distributions is minimal sufficient. (You may assume the set-up of discrete distributions.) [6+10]=[16]

7. In the following X is a random variable, real or vector valued, with density $f_\theta(x)$ $a \leq \theta \leq b$. The statements below are either true or false. Prove the true statements and provide counter examples in other cases.
   a) If $T(X)$ is sufficient for the family of distributions of X and $\phi(T(X))$ is sufficient for the family of distributions of $T(X)$ then $\phi(T(X))$ is sufficient for the family of distributions of X. [4]
   b) If T_1 is sufficient for θ and T_2 is independent of T_1 under each θ, then the distribution of T_2 is free of θ. [4]
   c) The likelihood ratio test depends on the sufficient statistic only. [4]
   d) The likelihood ratio test is unbiased. [4]

e) If $f_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$ and

$$\phi_{c,d}(x) = 1 \text{ if } x > d \text{ or } < c$$
$$= 0 \text{ otherwise}$$

is an unbiased a test of $H_0(\theta = 0)$ vs. $H_1(\theta \neq 0)$ then $c = -d$. [4]

f) If for some α, and θ_0 there exists a M.I. test of $H_0(\theta = \theta_0)$ vs. $H_1(\theta > \theta_0)$ then the family $\{f_\theta\}$ is an MLR family in x. [5]

g) A necessary and sufficient condition for densities $f_\theta(x)$ to have monotone likelihood ratio, if the mixed second derivative

$$\frac{\partial^2 \log f_\theta(x)}{\partial\theta\partial x}$$

exists, is that this derivative be ≥ 0 for all θ and x. [8]

8. Find the minimal sufficient statistic in each of the following cases and examine whether it is complete

a) $X_1, \ldots, X_n$ are i.i.d and $P_\theta\{X_1 = m\} = \frac{1}{\theta}$,

$$m = 1, 2, \ldots, \theta$$
$$\theta = 1, 2, 3, \ldots$$

b) X is a random variable and the parameter space is the family of all symmetric distributions

$$P_\theta\{X = m\} = P_\theta\{X = -m\}, \quad m = 0, 1, 2, \ldots$$

c) $X_1, \ldots, X_n$ are i.i.d,

$$f_\theta(x_1) = \theta \quad \text{if } x_1 = 1$$
$$= 1 - \theta \text{ if } x_1 = 0$$

and Ω consists of three points $\theta = \frac{1}{2}, \frac{1}{3}$ or $\frac{2}{3}$.

[7+6+6]=[19]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [127]

MID-YEAR EXAMINATIONS

Statistics-5: Statistical Methods Theory and Practical

Date: 20.12.68. Maximum Marks: 100 Time: 3 hours

Note: Answer Group A and Group B in separate answer-scripts. Marks allotted for each question are given in brackets [ ].

Group A

Theory

Maximum Marks: 50 Suggested time: $1\frac{1}{2}$ hours

Answer all questions.

1.a) Define 'tolerance limits'. [3]

b) If $(x_1, x_2, \ldots, x_n)$ represent a sample of size n from $N(\mu, \sigma^2)$, $\bar{x} = \frac{1}{n} \Sigma x_i$, $s^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$ and

$$P(\bar{x}, s) = \int_{\bar{x}-ks}^{\bar{x}+ks} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx$$

show how a value of k could be determined such that the expected coverage by the tolerance limits $(\bar{x}-ks, \bar{x}+ks)$, $EP(\bar{x}, s)$ is equal to a preassigned value α $(0 < \alpha < 1)$. [12]

2. Consider random variables $(Y_1, Y_2, \ldots, Y_n) = \underline{Y}'$ and a linear model under which $E(\underline{Y}) = X\underline{\beta}$, dispersion matrix $D(\underline{Y}) = \sigma^2 I$. The matrix X is known and $\underline{\beta}$ is a vector of unknown parameters. Assume further that the variables are normally distributed. Show that

a) If $R_o^2 = \min_{\beta} (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$ show that R_o^2/σ^2 has a chi-square distribution with $n-r$ d.f. where r = rank of X. [8]

b) Consider a hypothesis: $H\underline{\beta} = \underline{h}$ and let

$$R_H^2 = \min_{\underline{\beta}: H\underline{\beta} = h} (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$$

Show that if each component of $H\underline{\beta}$ is individually estimable

$$R_H^2 - R_o^2 = (H\hat{\underline{\beta}} - \underline{h})'(H(X'X)^- H')^- (H\hat{\underline{\beta}} - \underline{h}) .$$ [10]

What can you say about the distribution of $(R_H^2 - R_o^2)/\sigma^2$ if no component of $H\underline{\beta}$ is individually estimable? [2]

3. Obtain the analysis of covariance for a two-way classified data with equal number of observations per cell taking into account a single concomitant variable which has been observed along with the primary variable of interest. [15]

Group-B
Practical

Maximum Marks: 50 Suggested time: $1\frac{1}{2}$ hours.
Answer all questions.

1. In setting confidence intervals for the median of a continuous distribution using the symmetrically spaced order-statistics $x_{(r)}$ and $x_{(n-r+1)}$ on the basis of a sample of size $n = 28$, find the largest values of r yielding a confidence co-efficient not less than

   a) 0.95
   b) 0.99 [10]

2. The measurements of Nasal Length of individuals, belonging to 4 different ethnic groups, as obtained in an Anthropometric study are given below.

   Apply a non-parametric test to examine whether Nasal Length is useful in distinguishing among these groups.

Group	Measurements in mm.	Sample size
KURUMBA	46,51,48,53,44,45,49,42	8
HAKKIPIKKI	40,46,48,45,47,48	6
MUSLIM	48,44,55,49,54,50,56,47,59	9
SOLIGA	54,50,58,44,49,40,47	7

[20]

3. Thirty persons in the income group Rs.1000-Rs.1500 were asked to supply returns of their monthly incomes for purposes of taxation. But only 20 returns were received till the specified last date, and it has to be decided whether these can be accepted as representative of the 30. There are prior reasons to believe that those with larger incomes have natural reluctance to supply the figures and may delay more than the others.

   The following are the figures in the 20 returns received, in that order:-

   1220,1290,1180,1270,1400,1090,1190,1250,1170,1300,1310,
   1280,1350,1320,1380,1420,1390,1470,1360,1460.

   Examine whether there are sufficient evidences for the above claim. [10]

4. Practical records. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [128]

MID-YEAR EXAMINATIONS

Statistics-7: Planning Techniques

Date: 21.12.68 Maximum Marks: 100 Time: 2 hours

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

1.a) Explain how a linear programming problem involving linear inequality constraints can be converted into one involving constraints in the form of a set of simultaneous linear equations so that there will be a one-to-one correspondence between the feasible solutions to the former and those of the latter. How can you ensure that the optimal values of the objective functions, if there are any, will also be the same in both cases?

b) Explain the procedure for the reduction of any feasible solution to a linear programming problem to a basic feasible solution.

c) Consider the set of equations:

$$5\xi_1 - 4\xi_2 + 3\xi_3 + \xi_4 = 3,$$
$$2\xi_1 + \xi_2 + 5\xi_3 - 3\xi_4 = 0,$$
$$\xi_1 + 6\xi_2 - 4\xi_3 + 2\xi_4 = 15$$

A feasible solution is $\xi_1 = 1, \xi_2 = 2, \xi_3 = 1, \xi_4 = 4$. Reduce the solution to a basic feasible solution.

[20+25+15]=[50]

2. Given a basic feasible solution $X_B = B^{-1}b$ to the set of constraints $AX = b$ for a linear programming problem, with the value of the objective function for this solution being $Z = C_B X_B$. If for any column a_j in A but not in B, the condition $Z_j - Y_j < 0$ holds, and if at least one $\xi_{ij} > 0$, show that it is possible to obtain a new basic feasible solution by replacing one of the columns of B by a_j, and the new value of the objective function $\bar{Z}$ satisfies $\bar{Z} \geq Z$. If, however, $Z_j - Y_j \geq 0$ for every column a_j in A, show that the corresponding value of the objective function will be the maximum that it can attain.

[50]

3.a) Explain how you would find an initial basic feasible solution to a linear programming problem.

b) Solve the following linear programming problem by the simplex method:

$$\xi_1 + 3\xi_2 + \xi_4 \leq 4,$$
$$2\xi_1 + \xi_2 \leq 3,$$
$$\xi_2 + 4\xi_3 + \xi_4 \leq 3,$$
$$\xi_1, \xi_2, \xi_3, \xi_4 \geq 0$$
$$\max\ Z = 2\xi_1 + 4\xi_2 + \xi_3 + \xi_4.$$

[25+25]=[50]

-----

[129]

INDIAN STATISTICAL INSTITUTE
Research and Training School
M.Stat. Part IV: 1968-69

MID-YEAR EXAMINATIONS

Statistics-7: Econometrics Theory and Practical

Date: 23.12.68 | Maximum Marks: 100 | Time: 3 hours

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

1. Discuss, how you would formulate the aggregate demand function for a particular commodity starting from the theory of consumer behaviour. Consider both linear and non-linear formulations for this purpose. [15]

2.a) If $X_1, X_2, \ldots, X_n$ are independently and identically distributed as $\Lambda(\mu, \sigma^2)$ then, give the distribution function of $\prod_{i=1}^{n} X_i$. [6]

b) Derive the equation of the Lorenz curve for a random variable distributed as $\Lambda(\mu, \sigma^2)$ and examine the properties of this Lorenz curve. [14]

3.a) What are family budget data? What are the major uses of these data? Briefly describe the method of collection and the main source of such data in India. [15]

b) Examine the role of household size in the formulation of the Engel curve. [10]

4. The following table gives the distribution of Income as obtained from the Indian income Tax Returns (1955). Fit a Pareto distribution to the appropriate number of income classes, by estimating the parameter from the Lorenz-ratio of the distribution.

Table: Statistics of Individual Salaries assessed, 1955.

range of income in rupees (annual) (1)	number of incomes assessed (2)	total income assessed (Rs.) (3)
below - 4200	16361	42266049
4201 - 6000	43498	204483698
6001 - 8400	60032	383320767
8401 - 10000	19835	179451981
10001 - 15000	17583	217018779
15001 - 25000	8550	161809868
25001 - 40000	3779	117902100
40001 - 55000	1178	54294343
55001 - 70000	312	19250598
70001 - 85000	127	9757201
85001 -100000	60	5423397
100001 -150000	89	10623869
150001 -200000	34	6323521
200001 and above	12	2929444

[40]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

[130]

MID-YEAR EXAMINATIONS

Statistics-6: Sample Surveys Theory

Date: 25.12.68 Maximum Marks: 100 Time: 3 hours

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

1. a) Give a <u>general</u> unbiased linear estimator of the population total Y for multistage designs and derive its variance in terms of the primary variances V_1 based on second and higher stage units.

b) Derive a general rule for unbiased variance estimation in multi-stage designs using (a).

c) Consider a stratified two-stage design where both primaries and secondaries within each stratum are drawn by srs without replacement (srswor). Spell out the formulae for variance of the estimator and unbiased estimator of variance using either (b) otherwise. [5+8+7]=[20]

2. a) Suppose there are two strata. From stratum 1 a sample of n_1 units is selected from the N_1 units in that stratum by srswor. In stratum 2 there are N_2 primaries with M_2 secondaries in each primary. A sample of n_2 primaries and m_2 secondaries from each selected primary is selected by srswor. Suppose the cost function is $C = n_1c_1 + n_2c_2 + n_2m_2c_3$ where c_1 = cost per unit in stratum 1, c_2 = cost per primary in stratum 2 and c_3 = cost per secondary in stratum 2. Derive an unbiased estimator of the population total Y and then derive the optimum value of m_2 which minimises the variance of the estimator for fixed C.

b) Consider a four stage design in which $n > 1$ primaries are drawn with srs with replacement (srswr). Each time a primary is selected, a sample of secondaries is drawn with unequal probabilities without replacement. Within each selected secondary, a sample of third stage units is selected by srswor and finally from each selected third stage unit a systematic sample is selected circular systematically. Suppose we have unbiased estimators of the primary totals selected in the sample. Derive an unbiased estimator of Y and its unbiased variance estimator using the unbiased estimators of sampled primary totals. [13+7]=[20]

3. a) Prove that for Lahiri's method the probability of selection of i-th unit is proportional to its size X_i.

b) Suppose that in a sample of n units the first unit is selected with p.p.s. of X_i's and the remaining (n-1) units with srswor from the remaining (N-1) units in the population. Derive the probability of inclusion of i-th unit in the sample in terms of X_i and prove that the probability of selection of the sample, P(s), is proportional to ΣX_i of the units in the sample.

c) Suppose there are N clusters in the population and the size of i-th cluster is M_i. Suppose n clusters are selected with probabilities proportional to M_i and with replacement. Derive an unbiased estimator of Y, its variance and unbiased variance estimator. [5+7+8]=[20]

4.a) Derive the optimum allocation for stratified srswor using the cost function $C = c_0 + \Sigma c_i n_i$.

b) Suppose there are M domains in the population and the above stratified srswor sample is drawn. Derive an unbiased estimator of i-th domain total and its variance explicitly.

c) Suppose a srswor sample of n units is drawn and the post-stratified estimator

$$\hat{\bar{Y}} = \Sigma N_i \bar{y}_i / N$$

is used, where $\bar{y}_i$ = mean of the $n_i > 0$ units falling in stratum i and $\bar{y}_i = 0$ if $n_i = 0$. Derive the exact bias of $\hat{\bar{Y}}$.

[7+6+7]=[20]

5.a) Describe circular systematic sampling of n units and prove that the sample mean is unbiased for this scheme.

b) Assuming $N = nk$, n units are selected systematically after a random start from 1 to k. Prove that

$$V(\bar{y}) = \frac{\sigma^2}{n} [1 + (n-1)\rho] \quad \text{where } \sigma^2 =$$

population variance and ρ = intraclass correlation between pairs of sample units.

c) From a population of N units, a srswor sample of n' units is selected and their x-values measured. Then a srswor of sample n units from those n' units is selected and their y-values measured. Prove that

$$\hat{\bar{Y}} = \bar{y} + k(\bar{x}' - \bar{x})$$

is an unbiased estimator of $\bar{Y}$, where $\bar{y}$ and $\bar{x}$ are sample means of n units, $\bar{x}'$ is sample mean of n' units and k is a constant. Derive the variance of $\hat{\bar{Y}}$.

[5+6+9]=[20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

[131]

MID-YEAR EXAMINATIONS

Statistics-6: Sample Surveys Practical

Date: 26.12.68 . Maximum Marks: 100 . . Time: 3 hours

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

1. A survey is to be conducted for estimating the total number of literates in a town having three communities, some particulars of which are given in table 1 based on the results of a pilot study.

Table 1. A rough idea of the total number of persons and proportions of literates in 3 communities.

Community	Total number of persons	Percentage of literates
1	60,000	40
2	10,000	80
3	30,000	60

i) Treating the communities as strata and assuming srswr in each stratum, allocate a total sample size of 2000 persons to the strata in an optimum manner for estimating the overall proportion of literates in the town, using the data in Table 1.

ii) Estimate the efficiency of stratification as compared to unstratified sampling. [25]

2. For studying the living conditions of the working class population residing in an industrial area, a stratified two-stage sampling design is proposed, in which from each stratum a sample of factories is to be drawn systematically with pps*, size being the number of workers in an earlier period (x), and a sample of workers is to be selected from each sample factory linear systematically with a random start using the current pay-roll. The relevant data are given in Table 2.

i) Determine the constant weight to be used in a self-weighting design for ensuring a total sample size of about 1000 workers.

ii) Specify the sampling interval to be used in each sample factory for achieving a self weighting design, using the constant weight determined in (i).

iii) Also find the approximate number of workers expected to be selected from each sample factory, thereby determining approximately the total sample size.

Table 2. Number of workers in sample factories.

Stratum	Sample factory	Number of workers past	Number of workers current	Stratum	Sample factory	Number of workers past	Number of workers current
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	1	99	183	3	1	2697	2839
	2	523	465		2	4667	6255
	3	110	64		3	1423	1158
	4	741	829		4	1064	1150
2	1	4200	3504	4	1	90	91
	2	3187	2527		2	618	416
	3	2215	2186		3	150	131
	4	5322	5286		4	266	282

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

MID-YEAR EXAMINATIONS

Statistics-II: Sample Surveys Practical

[illegible]

Date: 26.12.68 Maximum Marks: 100 Time: 3 hours

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

1. A survey is to be conducted for estimating the total number of literates in a town having three communities, some particulars of which are given in table 1 based on the results of a pilot study.

Table 1. A rough idea of the total number of persons and proportions of literates in 3 communities.

Community	Total number of persons	Percentage of literates
1	60,000	40
2	10,000	80
3	30,000	60

i) Treating the communities as strata and assuming srswr in each stratum, allocate a total sample size of 2000 persons to the strata in an optimum manner for estimating the overall proportion of literates in the town, using the data in Table 1.

ii) Estimate the efficiency of stratification as compared to unstratified sampling. [25]

2. For studying the living conditions of the working class population residing in an industrial area, a stratified two-stage sampling design is proposed, in which from each stratum a sample of factories is to be drawn systematically with pps*, size being the number of workers in an earlier period (x), and a sample of workers is to be selected from each sample factory linear systematically with a random start using the current pay-roll. The relevant data are given in Table 2.

i) Determine the constant weight to be used in a self-weighting design for ensuring a total sample size of about 1000 workers.

ii) Specify the sampling interval to be used in each sample factory for achieving a self weighting design, using the constant weight determined in (i).

iii) Also find the approximate number of workers expected to be selected from each sample factory, thereby determining approximately the total sample size.

Table 2. Number of workers in sample factories.

Stratum	Sample factory	Number of workers: past	Number of workers: current	Stratum	Sample factory	Number of workers: past	Number of workers: current
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	1	99	163	3	1	2697	2839
	2	523	455		2	4667	6258
	3	110	64		3	1423	1168
	4	741	839		4	1064	1150
2	1	4200	3504	4	1	90	91
	2	3187	2527		2	618	416
	3	2215	2186		3	150	131
	4	5322	5285		4	266	282

Tablo 4: Sizo of 49 Largo Unitod Statos citios (in 1000's) in, 1920 (x_1) and 1930 (y_1).

x_1	y_1	x_1	y_1	x_1	y_1
76	80	2	50	242	291
138	143	507	634	87	105
67	67	179	260	30	110
29	50	121	113	71	79
381	464	50	64	256	288
23	48	44	58	43	61
37	63	77	89	25	57
120	115	64	63	94	85
61	69	64	77	43	50
387	459	56	142	298	317
93	104	40	60	36	46
172	183	40	64	161	232
78	106	38	52	74	93
66	86	136	139	45	53
60	57	116	130	36	54
46	65	46	53	40	58
				48	75

5. Noatnoss and clarity. [5]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [132]

MID-YEAR EXAMINATIONS

Statistics-8: Demography (Theory and Practical)

Date: 28.12.68 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 from Census returns (1941-50) with special reference to the age groups 0-5 and 60+. [25]

2. Explain clearly the different stages which led to the formulation of the law of population growth

$$P_t = \frac{L}{1 + e^{(\beta - t)/\alpha}}.$$

where P_t, L, β, α and t have their usual significance:

Discuss the law of growth of population with reference to India from the following data.

Year	1891	1901	1911	1921	1931	1941	1951	1961
Population in millions	236	235	249	248	276	313	357	436

[20]

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 clearly the assumptions you make.

Age group in years	Percentage of ever married women	Average no. 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

[25]

4. Briefly discuss any three of the following:

a) Hospital records as a source of morbidity data:

b) Incidence and prevalence rates.

c) Uses of vital statistics.

d) Underlying cause of death and the procedure for its selection.

[10+10+10]=[30]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

MID-YEAR EXAMINATIONS

Statistics-8: Educational Statistics Theory and Practical

Date: 30.12.68 Maximum Marks: 100 Time: 3 hours

Note: Answer question 8 and any five from the rest. Marks allotted for each question are given in brackets [ ].

1. Briefly describe the computational procedure and properties of any two of the following methods of standardising the raw scores:

a) Percentile scores (b) Stanine grades (c) z-scores. [8+8]=[16]

2.a) What do you understand by 'biserial correlation coefficient? Derive the formula with the help of which you can estimate 'normal' the correlation coefficient using the biserial correlation coefficient.

b) Test X is reported as having a mean of 68.1, a standard deviation 24.8, a reliability of 0.97 and a validity of 0.75 with the criterion. The mean value of the criterion scores is 117.8 and the corresponding standard deviation is 20.1. If we screen a group using scores on test X and obtain a selected subgroup with standard deviation 15.0 on test X, what will the validity be for this test on this selected group? [10+6]=[16]

3. Describe what you understand by (a) factor loadings (b) factor scores (c) communality (d) uniqueness (e) tetrad difference (f) multiple-factor theory. [3+3+2+2+3+3]=[16]

4.a) Discuss the effects of the following on the variance of the total score.

i) item difficulty (ii) item variance
iii) item validity (iv) item inter-correlation.

b) Determine the reliability of the difference score obtained from the following pair of tests X and Y where

reliability of test X = .95
reliability of test Y = .85
correlation between test X and Y = .40

Assume that the variances of tests X and Y are both equal to 1. [10+6]=[16]

5.a) Show, after stating the underlying assumptions, that the reliability of a test is

$$r_{xx'} = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} p_i q_i}{s_x^2}\right) \text{ where}$$

n = the number of items in the test.
p_i = the difficulty value of the i-th item; $q_i = 1-p_i$
s_x^2 = the variance of the total score.

b) Why should the odd-even method of estimating test reliability be used for 'power tests' only? [10+6]=[16]

6.a) What do you understand by 'item validity'? What are the item validity indices which involve the slope of the regression of item on test?

b) From the equations showing the relationships of test length to validity and to reliability, determine the relationship between test reliability and validity as the length of the test is increased while the criterion remains unchanged i.e. write

$$f(r_{xx'}, r_{xy}, R_{kk'}, R_{ky}) = 0 \quad \text{where}$$

$r_{xx'}$ = the reliability of the original test.

r_{xy} = the validity of the original test.

$R_{kk'}$ = the reliability of the test when the test length is increased k times.

R_{ky} = the validity of the test when the test length is increased k times. [8+8]=[16]

7.a) What are parallel tests? What are the hypotheses H_{mvc}, H_{vc} and H_m?

b) Prove that if a test of n items is a sub-test of a test with m items (n < m) the correlation r_{nm} is

$$r_{nm} = \sqrt{\frac{\frac{1-r}{m} + r}{\frac{1-r}{n} + r}}$$ where r is the reliability of a unit test. [6+10]=[16]

8. Write short notes on any <u>four</u> of the following:

a) Group heterogeneity and Reliability of a test

b) Correction for guessing

c) Coefficient of discrimination of a test

d) Ø-coefficient

e) Rank correlation. [5+5+5+5]=[20]

INDIAN STATISTICAL INSTITUTE [134]
Research and Training School
B.Stat. Part IV: 1968-69

PERIODICAL EXAMINATIONS

Statistics-4

Date: 24.2.69 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

Probability

Maximum Marks: 50 Suggested time: $1\frac{1}{2}$ hours

Answer all questions.

1.a) S is the smallest set of positive integers satisfying the following conditions:

i) S is closed under addition,

and ii) S contains 20, 24, 28.

Find the smallest positive integer k such that 4k, 4(k+1), 4(k+2), ... all belong to S. [8]

b) α and β are positive integers such that their G.C.F. is 1. Is the following statement true?

'There is a positive integer m and a negative integer n $\ni (m\alpha + n\beta) = 1$'. If your answer is YES, prove it. If your answer is NO, give a counterexample. [9]

c) α, β are positive integers with G.C.F.I. Prove that there are integers m, n such that $m\alpha + n\beta = 1$. [8]

2. In a finite Markov chain, each state is a consequent of every state. S is the set of positive integers n such that $p_{11}^{(n)} > 0$. d = G.C.F.S.

a) Prove that if $p_{1j}^{(\ell)} > 0$, $p_{1j}^{(n)} > 0$, then $(\ell - n)$ is a multiple of d. [8]

b) How are the cyclically moving subclasses $c_1, c_2, \ldots, c_d$ defined? [8]

c) Prove that if states x and y lie in the same subclass,

$$p_{xy}^{(kd)} > 0 \text{ for all large } k.$$ [9]

GROUP B

Inference

Maximum Marks: 50   Suggested time: $1\frac{1}{2}$ hours

.. Answer all questions.

1.a) Show that if a complete sufficient statistic exists then every estimable parametric function has a best unbiased estimator. [8]

b) Let X_1, X_2 be independent $N(\mu, \sigma^2)$. Let

$\psi(\mu, \sigma^2) = P_{\mu,\sigma^2}(X_1 > 0)$. Find the best unbiased estimator of ψ. [10]

2.a) $X_1, \ldots, X_n; Y_1, \ldots, Y_n$ are i.i.d., normal with mean μ. Variance of $X_1 = \sigma_1^2$ and that of Y_1 is σ_2^2. If $0 < \sigma_1, \sigma_2 < \infty$, show that no best unbiased estimator for μ exists. Determine the minimal sufficient statistic and show that it is not complete. [12]

b) Assuming above that $\sigma_1 = \sigma_2 = 1$, find the b.u.e. of μ^2. Calculate its variance and show it is strictly greater than the Cramer-Rao lower bound. [6]

3.a) Let X be $N(\mu, 1)$. Let $H_0(\mu = 1)$ be tested against $H_1(\mu = -1)$. Let $gx = -x$. Does g leave the model invariant? Does g leave the testing problem invariant? [2]

b) Let $P_\theta\{X = r\} = n_{c_r}\, \theta^r (1-\theta)^{n-r}$, $\theta = \frac{1}{4}$ or $\frac{3}{4}$ or $\frac{1}{2}$

$r = 0, 1, 2, \ldots, n.$

Let $H_0(\theta = \frac{1}{2})$ be tested against $H_1(\theta \neq \frac{1}{2})$. Let $gx = n-x$. Does g leave the testing problem invariant? [2]

c) Let X_1, X_2 be i.i.d. $N(\mu, \sigma^2)$, $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$. Let $H_0(\mu = 0)$ be tested against

$$H_1\left(\frac{|\mu|}{\sigma} = 1\right).$$

Let $g_a x = a.x$. Does $G = \{g_a;\ -\infty < a < \infty,\ a \neq 0\}$ leave the testing problem invariant? Find a maximal invariant statistic. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [135]

PERIODICAL EXAMINATIONS

Statistics-5: Statistical Methods
Theory and Practical

Date: 3.3.1969 Maximum Marks: 100 Time: 3 hours

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

1.a) Show that under certain conditions to be stated in full, the sample third central moment has an asymptotic normal distribution in a sense to be made precise by you. State without proof, all results you may need in this connection.

b) Compute the standard error of the sample third central moment.

2.a) Show that if $O_1, O_2, \ldots, O_k$ are the observed frequencies in the k classes of a multinomial distribution with respective class probabilities

$$\pi_1, \pi_2, \ldots, \pi_k \ (\sum_1^k \pi_i = 1),$$

then the variance of a linear function $\Sigma a_i O_i$ is given by the expression

$$V(\Sigma a_i O_i) = n[\Sigma a_i^2 \pi_i - (\Sigma a_i \pi_i)^2].$$

b) Obtain the limiting distribution of Pearson's goodness of fit chisquare statistic, as computed from a single sample of size n drawn from a multinomial distribution.

3.a) What is a 'U-statistic'?

b) State the conditions under which a U-statistic is asymptotically normally distributed.

c) Compute the standard error of Spearman's rank correlation coefficient when the variates concerned are independently distributed.

4. Consider as in Q.2, a sample of size n from a multinomial distribution and let the class probabilities be known functions of a single unknown parameter θ. Write

$$c(\theta) = \sum_{i=1}^k (O_i - n\pi_i(\theta))^2 / n\pi_i(\theta).$$

Show that if $\hat{\theta}$ is a root of the equation

$$\frac{dc(\theta)}{d\theta} = 0,$$

then, under certain conditions.

(a) $\hat{\theta}$ is Fisher consistent for θ.

(b) $\sqrt{nI(\theta_0)}\,(\hat{\theta} - \theta_0)$ is asymptotically a standard normal variable, where θ_0 is the true value of θ and

$$I(\theta_0) = \sum_{i=1}^k \left(\frac{d \log \pi_i(\theta)}{d\theta}\right)_0^2 \pi_i(\theta_0).$$

(c) $c(\hat{\theta})$ is in the limit distributed as χ^2 on k-2 d.f.

5.a) Show that to terms of order $O(\frac{1}{n})$, the variance of $\tanh^{-1} r$ is independent of the population parameters where r is the sample correlation coefficient in a sample of size n from the bivariate normal population.

b) Correlations between two variables X and Y were computed from three different samples as follows

Sample No.	Sample size	Value of r
1	215	0.415
2	137	0.531
3	70	0.497

Examine if there are significant differences in values of r. If not, compute the pooled estimate of r.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [1??]

PERIODICAL EXAMINATIONS

Design of Experiments (Theory and Practical)

Date: 10.3.69 Maximum Marks: 100 Time: 3 hours

Note: Answer as many questions as you can. Marks allotted for each question are given in brackets [ ].

1. Describe the layout and the model of a Randomised Block design.

Explain how you would obtain the analysis of variance table and make use of it to test

i) equality of all the treatment effects,

ii) equality of two specified treatment effects.

Obtain the expression for the efficiency of the randomised block design compared with that of Completely randomised design. [5+20+5]=[30]

2. Explain with illustration the role of randomisation in the conduct of experiments. Why an experimenter introduces replication while conducting experiments? [16+9]=[25]

3. Describe the use of Analysis of Covariance in increasing the precision of an experiment. Give the model and the method of analysis when a single concomitant variable is used on a Randomised Block design.

a) Can you test the usefulness of the concomitant variable on the basis of the analysis of covariance table obtained above?

b) How will you use the analysis of covariance table obtained above) to test the equality of all the treatment effects when the introduction of the concomitant variable has been found to be useful in (a)? [3+10+12]=[25]

4. Consider the results given in the following table for an experiment involving six treatments in 4 randomised blocks. Suppose that the yield for treatment 1 in block 1 is missing. Analyse the data.

Find an estimate of the variance of the difference between the mean of the treatment with a missing value and that of any other treatment.

Table

Yield for a Randomised-blocks Experiment.
(The treatments are indicated by numbers within parentheses)

Block	Treatment and yield					
1	(1)	(3)	(2)	(4)	(5)	(6)
	-	27.7	20.6	16.2	16.2	24.9
2	(3)	(2)	(1)	(4)	(6)	(5)
	22.7	28.8	27.3	15.0	22.5	17.0
3	(6)	(4)	(1)	(3)	(2)	(5)
	26.3	19.6	38.5	36.8	39.5	15.4
4	(5)	(2)	(1)	(4)	(3)	(6)
	17.7	31.0	28.5	14.1	34.9	22.6

[25+5]=[30]

5.a) Draw up a randomised blocks layout for an experiment involving 6 treatments A, B, C, D, E, F in 5 blocks.

b) Ten test animals were divided at random into two groups of five each. One group was given a particular type of food A, and the other group a different food B. Table below gives the increase in weight in pounds of the animals in a six-week period.

Table showing the increase in weight of animals given two different types of food

Type of food	Increase in weight (lb.)				
A	1.2	2.4	1.3	1.3	0.0
B	1.0	1.8	0.8	2.6	1.4

Examine whether there is any difference in the effects of the two types of food (i) by use of the t-test, (ii) by use of the F-test.

[6+7+7]=[20]

------

[137]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

PERIODICAL EXAMINATIONS
Statistics-7: Industrial Statistics Theory
... ptional

Date: 17.3.69 · Maximum Marks: 100 · Time: 3 hours.

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

1. Explain the role of rational sub-grouping in controlling the quality during a production process.

Following are the summarised data on a measurable characteristic.

Sample	X	R	Sample	X	R	Sample	X	R
1	0.7540	.0011	6	.7539	.0009	11	0.7541	.0017
2	0.[illegible]42	.0014	7	.7541	.0013	12	0.7542	.0011
3	0.[illegible]	.0009	8	.7543	.0011	13	0.7548	.0011
4	0.[illegible]3	.0010	9	.7547	.0007	14	0.7540	.0009
5	0.[illegible]0	.0008	10	.7549	.0015	15	0.7551	.0012

Plot an $\bar{X}$ - R chart and examine the data for control using sample size as 4. How would you modify the control chart if the specified limits for the above measurable characteristic are 0.7528-0.7571? Plot the modified chart. [5+20]=[25]

2. Give two examples of industrial situations where Poisson law is applicable.
Following data were collected to establish standard for the defects on an electronic component.

Sample No.	No. inspected	Total No. of defects	Sample	No. inspected	Total No. of defects
1	4	17	11	8	31
2	7	23	12	6	39
3	5	24	13	3	29
4	7	27	14	8	30
5	7	32	15	9	31
6	7	33	16	6	21
7	6	18	17	5	26
8	7	28	18	7	20
9	7	29	19	3	24
10	6	31	20	6	29

Analyse by means of control chart technique and recommend standard for average number of defects per item. [5+20]=[25]

3.a) The specification limits on the gross weight of an ink bottle are 110 ± 3 gms. It was found that the weight of an empty bottle has a mean of 53.9 gms. and a s.d. of 0.7 gms. Empty bottles are fed into the filling machine in a random order. Find out what is maximum allowable standard deviation at the filling stage (i.e. s.d. of the weight of ink filled in a bottle) so that the final product meets the specification limits on the gross weight.

3.b) Using Dodge-Romig sampling inspection tables select plan to satisfy the following:

i) Single and double sampling plans for lot size 5000, LTPD = 3.5 per cent, process average 1.10 what are AOQL for these plans.

ii) Single sampling plan for lot size 1200 LTPD = 3.0 per cent, process average 0.5 per cent. Find out AOQL for this plan. [10+10+5]=[25]

4.a) Explain the following terms

(i) AOI, (ii) AOQ (iii) AOQL

and also derive mathematical expression for AOQL of a single sampling and O.C of a double sampling plans by attributes.

b) For following double sampling plan

$N = 5000$, $n_1 = 100$, $c_1 = 0$, $n_2 = 100$, $c_2 = 1$.

i) Compute the probability of acceptance of 2 per cent defective lots.

ii) What will be the AOQ if the lots submitted contain 2 per cent defective items and the rejected lots are screened.

iii) What would be the average number of articles inspected per lot in (ii). [10+15]=[25]

5. Write short notes on following:

a) A process in a state of 'Statistical Quality Control'

b) Process capability and specification

c) Group control charts

d) Screening under sampling plans by attributes

e) 3-sigma limits and Probability limits.

[25]

----

[133]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69
PERIODICAL EXAMINATIONS

Statistics-7

Date: 24.3.69 Maximum Marks: 100 Time: 3 hours

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

GROUP A

Econometrics

Maximum Marks: 70 Suggested time: 2 hours

Answer all questions.

1. Explain the concept of 'Economies of scale' in the context of family budget analysis.

   While estimating an Engel Curve from family budget data how would you take into account the possible effect of this phenomena. [20]

2. Describe a method of estimating Engel elasticities from the concentration curves, giving the necessary proofs. [20]

3. Derive the general form of the total cost function under a given production function. How do you think the shape of total cost curve would look like? Give reasons for your answer. [20]

4. Write short note on any one of the following:

   a) Cobb-Douglas Production function

   b) Unit consumer scale. [10]

GROUP B

Planning Techniques

Maximum Marks: 30 Suggested time: 1 hour

EITHER

1. Explain Charnes' perturbation method for resolving the degeneracy problem. [30]

OR

Solve by the simplex method:

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

subject to

$$100\,\xi_1 + 100\,\xi_2 + 100\,\xi_3 + 100\,\xi_4 \leq 1500$$

$$7\,\xi_1 + 5\,\xi_2 + 3\,\xi_3 + 2\,\xi_4 \leq 100$$

$$3\,\xi_1 + 5\,\xi_2 + 10\,\xi_3 + 15\,\xi_4 \leq 100$$

$$\xi_1, \xi_2, \xi_3, \xi_4 \leq 0$$

Also formulate its dual and give its solutions. [30]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [139]

PERIODICAL EXAMINATIONS

Statistics-8: Genetics Theory and Practical

Date: 31.3.69 Maximum Marks: 100 Time: 3 hours

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

1.a) What is genetical linkage? Define cross over ratio. Briefly describe the construction of a chromosome map.

b) In corn, white endosperm (w) is recessive to purple (W) and Shrunken endosperm (s) is recessive to full (S). A pure purple shrunken is crossed to a pure white full. The F_1 is then crossed to a white shrunken, and the offspring are as follows:

Purple shrunken	3149
Purple full	120
White shrunken	115
White full	3334

Test for the presence of linkage. Estimate recombination fraction and the variance of the estimate, if linkage is present. [10+15]=[25]

2.a) What is meant by sex linked inheritance? Describe with suitable examples.

b) In Drosophila, the mutant gene for light eye colour known as 'vermillion' (v) is in the X chromosome and is recessive to red (V). A vermillion female is crossed to a red-eyed male. Give the eye colour of the F_1 (together with their sex) and of the F_2 (when the F_1 are interbred).

c) In humans, red-green colour blindness is controlled by a sex-linked recessive gene (c). Thalassemia is a type of anemia common in Mediterranean populations, but is relatively rare in other peoples. This anemia is inherited autosomally and is expressed in two degrees, severe cases that are usually fatal in childhood (called thalassemia major) and mild cases (called thalassimia minor). Use symbols T T for thalassemia major, Tt for thalassemia minor, and tt for normal individuals.

A woman who is red-green colour blind and afflicted with thalassemia minor marries a man with normal vision but who is afflicted with thalassemia minor.

i) Show the genotypes of these two individuals.

ii) Show the possible genotypes and phenotypes and the proportion of each for all children from this marriage. Designate the sons and daughters.

iii) What proportion of the children would be expected to succumb to thalassemia? [7+8+10]=[25]

3. Calculate the amount of information per individual provided by F_2 coupling and repulsion data regarding linkage between two factors A- a and B- b when

i) classification is complete and

ii) both the factors show complete dominance.

Compare the different results. [25]

4. For two factors which are individually Mendelian and show dominance,

a) Write down the expected frequencies for the phenotypic classes of F_2.

b) Describe the procedure for the detection of linkage in this case, clearly stating the algebraic expressions to be used.

c) Find the variance of the maximum likelihood estimator of cross over ratio, assuming this to be same for both the parents.

d) Comment on the efficiency of the estimator obtained by equating the linkage function to its expection.

[4+7+7+7]=[25]

5.a) Describe any procedure for detection of linkage from F_2 data when single factor segregations are disturbed.

b) When single factor segregations are subject to viability disturbances, suppose backcross data on both coupling and repulsion are available. Give a suitable method for estimation of recombination fraction and find the large sample variance of the estimator. [10+15]=[25]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [140]

ANNUAL EXAMINATIONS

Statistics-4: Probability

Date: 19.5.69. 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 whole paper carries 110 marks. You may attempt all questions. The maximum you can score is 100.

WE CONSIDER MARKOV CHAINS WITH A FINITE NUMBER OF STATES. $P = \{p_{ij}\}$ IS THE TRANSITION MATRIX.

Group A

1.a) i and j lie in the same ergodic class.

$$S = E\left\{n:\ p_{ii}^{(n)} > 0\right\},\quad T = E\left\{n:\ p_{jj}^{(n)} > 0\right\}.$$

Prove that S and T have the same greatest common divisor d. [Do not assume the existence of cyclically moving subclasses. If you require any lemma from elementary number theory, state it.] [8]

b) If there are no transient states and only one ergodic class, and $d = 1$ for this class, prove that $p_{ij}^{(n)} > 0$ for all large n. [8]

c) P is such that $Q = \lim_n P^n$ exists. Prove that if there is only one ergodic class, the rows of Q are all the same. (There may be transient states). [12]

2.a) P is such that $Q = \lim_n P^n$ exists. Write down two properties of Q and prove them. [7]

b) Prove that there is no stochastic matrix P such that $\lim_n P^n$ is

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$ [10]

c) P is a stochastic matrix and $\lim_n P^n$

$$= \begin{bmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{6} & \alpha & \beta \end{bmatrix}$$

Show that α and β are uniquely determined. What are their values? [10]

Group B

3.a)

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

Determine the matrix

$$Q = \lim_{n} \frac{P + P^2 + P^3 + \ldots + P^n}{n}$$

Hint: First obtain the limit of a convergent subsequence of the sequence $\{P^n\}$.

b) Consider a Markov chain whose transition matrix is P above. Determine the absorption probabilities (into ergodic classes) by solving the usual linear equations and also by considering the various paths from the transient states and into the ergodic classes. Show that the answers you got by the two methods agree.

c) The transition matrix of a Markov chain is P and that of another Markov chain is P^k, k being a positive integer. Prove that a state is transient in the second chain if and only if it is transient in the first chain.

d) If $\lim_{n} P^n$ is the identity matrix, prove that the Markov chain has no transient states.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

[141]

ANNUAL EXAMINATIONS

Statistics-4: Inference

Date: 21.5.69 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 any two questions.

1.a) Define Fraser and Hajec sufficiency. Explain their use in testing hypotheses. [13]

b) Let X be a discrete random variable with $f_\theta(x) = P_\theta\{X = x\}$, $\theta \in \Omega$. Supposing each $f_\theta(x)$ is symmetrical about the origin show that $|X|$ is sufficient for θ. Generalising this result show that if $G = (g_1, \ldots, g_m)$ is a finite group such that $f_\theta(gx) = f_\theta(x)\ \forall\ \theta, g, x$, then the maximal invariant under G is sufficient for θ. [12]

2.a) State and prove Stein's theorem on invariance and sufficiency and discuss briefly its role in reducing data through invariance and sufficiency. [10+3]=[13]

b) Obtain the one-sided student's t-test as a M.P. invariant test. (You have to prove the MLR property of t-distributions.) [12]

3.a) Let $X_1, \ldots, X_n$ be random variables with joint density $f_\theta(x_1, \ldots, x_n) = f_0(x_1 - \theta, \ldots, x_n - \theta)$ $-\infty < \theta < \infty$. Let $E_0(X_1^2) < \infty$. Find the best translation invariant estimator under the squared error loss. Show it is unbiased. [12+3]=[15]

b) X_1 and X_2 are i.i.d. with common density $f_\theta(x) = f_0(x - \theta)$ specified below. Find the best translation invariant estimator.

i) $f_0(x) = 1$ if $0 \leq x \leq 1$
$= 0$ otherwise.

ii) $f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ $-\infty < x < \infty$. [6+4]=[10]

Group B

Maximum Marks: 50

Answer any two questions.

1.a) Derive Barankin's lower bound to the variance of an unbiased estimator and show how the Cramer-Rao bound can be obtained from it. [15]

b) Let $P_\theta\{X = -1\} = \theta$, $P_\theta\{X = n\} = (1-\theta)^2\theta^n$

$$n = 0, 1, 2, \ldots, \qquad 0 < \theta < 1.$$

Find the locally best unbiased estimators of θ.

2.a) Prove the Fundamental Identity for a SPRT assuming $E(Z_1^2) > 0$. [10]

b) Let X_i be i.i.d. with common density

$f_\theta(x) = \frac{1-\theta^2}{2} \exp(-|x| + \theta x)$ $\theta = \pm\frac{1}{2}$. Construct the SPRT of H_0 $(\theta = -\frac{1}{2})$ vs. $H_1(\theta = \frac{1}{2})$ with boundaries $B < 1 < A$ and find its error of first kind exactly. [15]

3.a) Discuss briefly the concept of Pitman efficiency. Let $\{X_i\}$ be i.i.d., normal with mean θ and variance unity. Let ϕ be a test of size α, of $H_0(\theta = 0)$ vs. $H_1(\theta > 0)$, which rejects H_0 for large values of

$$\bar{X}_n = \sum_1^n X_i \mid n;$$

let ϕ_{2n} be a test of size α which rejects H_0 for large values of Median $(X_1, \ldots, X_n)$. Calculate the Pitman efficiency of ϕ_{1n} w.r.t. ϕ_{2n}. [9+3]=[12]

b) Consider two independent samples of size $n = m$ from continuous populations F and G where $G(x) = F(x-\Delta)$ $\Delta \geq 0$ and F has a density. You have to test $H_0(F = G)$ vs. $H_1(G(x) = F(x - \Delta), \Delta > 0)$. Obtain the lower bound to the Pitman efficiency of the two sample Fisher-Yates test with respect to the usual t-test. [13]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [142]

ANNUAL EXAMINATIONS

Statistics-5: Statistical Methods Theory

Date: 23.5.69 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 any two questions.

1.a) Obtain the standard error of the sample correlation coefficient r, in a sample of size n, from a bivariate normal population. [17]

b) Derive the $\tanh^{-1}$ transformation as a variance stabilising transformation of r. [8]

2.a) Let $(Y_{n1}, Y_{n2}, \ldots, Y_{nn})$ be a random permutation of integers $1, 2, \ldots, n$, and $c_n(i, j)$, $i, j = 1(1)n$ be n^2 real numbers. Consider

$$S_n = \sum_{i=1}^{n} c_n(i, Y_{ni}).$$

State sufficient conditions on a sequence $\{c_n\}$, $n = 1, 2, \ldots$ for the asymptotic normality of S_n. [8]

b) Let $O_1, O_2, \ldots, O_k$ and $E_1, E_2, \ldots, E_k$ be the observed and the corresponding expected frequencies in k mutually exclusive and exhaustive classes, as recorded in a sample of size n drawn without replacement from a finite population of size N. Use 2(a) to obtain sufficient conditions underwhich

$$\frac{N}{N-n} \sum_{i=1}^{k} (O_i - E_i)^2 / E_i$$

is asymptotically distributed as chisquare on $k-1$ d.f. [17]

3.a) Show that the null distribution of the Wilk's Λ criterion is the same as that of the product of several independent beta variables. [10]

b) Hence compute the t-th moment of Λ. [10]

c) State without proof a large sample approximation for the null distribution of Wilk's Λ. [5]

Group B

Maximum [illegible]

Answer any two questions.

1.a) Show that if the variables $Y_1, Y_2, \ldots, Y_p$ have mean zero and a finite dispersion matrix Σ and if $\underset{\sim}{Y}' = (Y_1, Y_2, \ldots Y_p)$ then with probability one $\underset{\sim}{Y} \in \mathcal{M}(\Sigma)$; that is, the equations $\underset{\sim}{Y} = \Sigma \underset{\sim}{X}$ are consistent for $\underset{\sim}{X}$; that is

$$\underset{\sim}{\ell}' \Sigma = \underset{\sim}{0}' \Rightarrow \underset{\sim}{\ell}' \underset{\sim}{Y} = 0.$$ [5]

1.b) $\underset{\sim}{Y}$ is said to have a p-variate normal distribution if for every $\underset{\sim}{\ell}$, $\underset{\sim}{\ell}'\underset{\sim}{Y}$ is univariate normal (with possibly a zero variance). Use this definition to establish the following propositions

i) If $\underset{\sim}{Y}$ is p-variate normal, then $\underset{\sim}{Y}$ has finite mean vector $\underset{\sim}{\mu}$ and a finite dispersion matrix Σ. [5]

ii) The characteristic function of $\underset{\sim}{Y}$ is given by

$$E\, e^{i\, \underset{\sim}{t}'\underset{\sim}{Y}} = e^{i\, \underset{\sim}{t}'\underset{\sim}{\mu} - \underset{\sim}{t}'\Sigma\underset{\sim}{t}/2}$$

(We denote the distribution of $\underset{\sim}{Y}$ by $N_p(\underset{\sim}{\mu}, \Sigma)$). [5]

c) Suppose Rank $\Sigma = r < p$ and let B $(p \times r)$ be such that $\Sigma = B B'$. Write $C = (B'B)^{-1} B'$ and observe $CB = I(r)$. Define $\underset{\sim}{G} = C\underset{\sim}{Y}$. Then show that if $\underset{\sim}{Y} \sim N_p(\underset{\sim}{0}, \Sigma)$ then (i) $\underset{\sim}{G} = N_r(\underset{\sim}{0}, I)$; (ii) With probability one, $\underset{\sim}{Y} = B\underset{\sim}{G}$. [10]

2.a) Define a central Wishart distribution $W_p(n, \Sigma)$. [3]

b) Show that if $S \sim W_p(n, \Sigma)$, and $\underset{\sim}{\ell}$ is a fixed vector such that $\underset{\sim}{\ell}'\Sigma\underset{\sim}{\ell} = 0$, then $\underset{\sim}{\ell}'S\underset{\sim}{\ell} = 0$ with probability one. If $\underset{\sim}{\ell}'\Sigma\underset{\sim}{\ell} = 1$ then $\underset{\sim}{\ell}'S\underset{\sim}{\ell} \sim \chi^2_n$. [2+5]=[7]

c) If a matrix T is such that $\underset{\sim}{\ell}'\Sigma\underset{\sim}{\ell} = 0 \Rightarrow \underset{\sim}{\ell}'T\underset{\sim}{\ell} = 0$ with probability 1 and $\underset{\sim}{\ell}'\Sigma\underset{\sim}{\ell} = 1 \Rightarrow \underset{\sim}{\ell}'T\underset{\sim}{\ell} \sim \chi^2_n$ where Σ is some fixed non-negative definite matrix, is it necessarily true that $T \sim W_p(n, \Sigma)$? If not, give a counter-example. [2+7]=[9]

d) Obtain the characteristic function of $W_p(n, \Sigma)$ when Σ is possibly singular. [6]

3. Let

$$\underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{Y}_2 \end{pmatrix} \begin{matrix} p \\ q \end{matrix}$$

have a $p + q$ variate normal distribution

$$N_{p+q}\left(\begin{pmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

and let Σ_{11} $(p \times p)$ be nonsingular.

Consider n independent observations on $\underset{\sim}{Y}$ and let T^2_{p+q} and T^2_p be the Hotelling's T^2 based on observations on all the $p+q$ variables and the first p variables respectively for testing the hypothesis that the mean values are all equal to zero. Show that if

$$\underset{\sim}{\mu}_2 = \Sigma_{21}\Sigma_{11}^{-1}\underset{\sim}{\mu}_1 \quad \text{then}$$

$$\frac{n-p-q}{q}\left(\frac{T^2_{p+q} - T^2_p}{n-1+T^2_p}\right)$$

has a variance ratio distribution on d.f. q for numerator and $n-p-q$ for denominator. [25]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [143]

ANNUAL EXAMINATIONS

Statistics-5: Statistical Methods Practical

Date: 24.5.69 Maximum Marks: 100 Time: 3 hours

<u>Note:</u> Answer any two questions from questions 1 to 3. Marks allotted for each question are given in brackets [ ].

1.a) The quartiles in a sample of size 401 from a certain population, were computed as

$$Q_1 = 3.15, \quad Q_2 = 7.30, \quad Q_3 = 14.56$$

Examine if this could be considered as sufficient evidence to discard the hypothesis that the population distribution is exponential with density

$$\frac{1}{10} e^{-x/10}, \quad 0 \leq x < \infty .$$

The population quartiles for this distribution are 2.88, 6.93 and 13.86.

b) Assuming that the population distribution to be exponential with density

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

use the sample quartiles to estimate θ. [40]

2. A study was made of the relationship between oxygen tension and respiratory rate of bacteria. Three Y measurements were obtained for each value of X.

Oxygen tension in mm Hg(X)	Respiratory rate (Y)		
1	9	11	8
3	58	57	64
5	76	76	78
7	82	84	82
9	88	86	90
11	91	89	91
13	93	92	92
15	95	96	95

(a) Fit a linear equation to the regression of Y on X and test for the adequacy of a linear fit.

(b) If the linear fit is inadequate, use a table of orthogonal polynomials to determine the degree and the equation of the best fitting polynomial regression. [40]

3. In an experiment on the effect of radiation on weight loss, total weight losses (in gms.) of a series of twenty four rats were recorded at 1, 3, 6 and 7 days after radiation. Twelve of the rats had been subjected to 500r whole body radiation while twelve others had been subjected to 600r. Given below are the pooled dispersion matrix based on 22 d.f. and the differences in the average weight loss ('higher' minus 'lower' levels of radiation)

Weight loss to i-th day

	1	3	6	7
1	6.00	3.71	3.33	3.08
3		25.71	7.35	3.60
6			21.42	17.27
7				24.08
Difference in average weight loss:	0.50	2.92	-2.58	-3.50

Examine if there is significant difference between levels of radiation with respect to weight loss of rats. [40]

4. Viva Voce. [10]

5. Practical Records [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [144]

ANNUAL EXAMINATIONS

Statistics-6: Design of Experiments

Date: 26.5.69 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 any two questions from Qns. 1 to 3

1.a) Describe a balanced incomplete block design (BIBD) with parameters b, v, r, k, λ. [2]

b) Establish the following relations among the parameters of a BIBD

i) $bk = rv$

ii) $\lambda(v-1) = r(k-1)$

iii) $b \geq v$ [2+3+7]=[12]

c) In usual notations write down the expression for the estimate of a treatment effect in a BIB experiment in terms of the Q's. (No derivation is necessary). Show that the estimates of the treatment comparisons of the form $\tau_i - \tau_j$ $(i \neq j)$ have the same variance and show that this variance is higher than the variance of the estimate of $\tau_i - \tau_j$ $(i \neq j)$ in a randomised block experiment with the same number of replicates, assuming σ^2 to be same in both cases. [1+2+3]=[6]

2. Explain the roles of randomization and local control in planned experiments. Illustrate these by discussing some examples. [12+8]=[20]

3. Obtain a plan for conducting a 2^5 Factorial experiment in blocks of 2^3 units each without confounding any main effect or 2-factor interaction. [20]

4. Viva Voce. [10]

Group B

Maximum Marks: 50

Answer all questions

1. EITHER

a) Explain the lay-out for an experiment to compare 5 treatments using a Latin Square design.

b) When is a pair of Latin Squares of order s said to be mutually orthogonal?

Write down FOUR mutually orthogonal Latin Squares of order 5. [3+3+14]=[20]

OR

a) Write 'true' or 'false' to each of the following statements (Need not give reasons):

i) 'Randomisation is a method by which every experimental unit has an equal chance of receiving a treatment'.

ii) In a particular experimental design suppose $t_1, t_2, \ldots, t_m$ denote the treatment effects. Two contrasts

$$\sum_{i=1}^{m} a_i t_i \quad \text{and} \quad \sum_{i=1}^{m} b_i t_i \quad (\text{i.e., } \Sigma a_i = 0 = \Sigma b_i)$$

and said to be mutually orthogonal if $\sum_{i=1}^{m} a_i b_i = 1$.

iii) In a randomised block experiment with r blocks and t treatments where one yield is missing, the variance of the usual estimate of the treatment effect with the missing value is given by

$$\frac{\sigma^2}{r}\left[1 + \frac{r}{(r-1)(t-1)}\right]$$

where σ^2 denotes the variance of the experimental error.

iv) Suppose $Z_m = \{0, 1, \ldots, m-1\}$ is the residue class modulo m. Z_m is, under usual optations modulo m, a field if and only if m is a prime number.

v) If F is a field with characteristic zero, it must have finite number of elements.

vi) For every integer $n \geq 1$ there exists a polynomial of degree n irreducible over the Galois field GF(p) where p is a prime number.

vii) There exists a Graeco-Latin Square (a pair of mutually orthogonal Latin Squares) of order 6.

viii) There does not exist a Graeco-Latin square of order 10.

ix) In a 2^4 Factorial experiment conducted with 4 plots in a block if the pencils (1,1,1,0) and (0,1,1,1) are confounded then so also the pencil (1,1,1,1).

x) In a Factorial experiment conducted in randomised blocks an effect or interaction is said to be confounded when the effect or interaction gets inextricably mixed up with block difference. [10]

b) The following table gives the plan of a 2^5 Factorial experiment in blocks of size 4.

Block				
1	(1)	abcde	abc	de
2	a	bcde	bc	ade
3	b	acde	ac	bde
4	c	abde	ab	cde
5	d	abce	abcd	e
6	ad	bce	bcd	ae
7	abd	ce	cd	abe
8	acd	be	bd	ace

Identify the confounded effects and interactions. [10]

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

Yields in a 4 X 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. Write down the estimates of each of the treatment effects in the descending order. Find the variance of the estimates. [14+4+2]=[20]

3. Practical Records [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

[145]

ANNUAL EXAMINATIONS

Statistics-7: Planning Techniques

Date: 27.5.69. Maximum Marks: 100 . Time: [illegible] hours

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

1. Formulate the open static input-output model in terms of linear programming and give an interpretation of its dual. Show that the Leontief solutions for quantities and prices are the optimal solutions of the programming problem. [33]

2. 'The Dynamic Leontief system (with zero final demand) has a unique positive rate of balanced growth with no excess capacity and positive initial stocks'. Discuss the statement.

   Prove that in the set of all balanced growth paths in the Leontief dynamic system (with or without excess capacity), one with no excess capacity is maximal. [33]

3. 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. [33]

4. Show that if the primal has an optimal solution, its dual also has an optimal solution. Prove that the correspon-ding values of the objective functions of the two problems are equal. [33]

5. Solve the following problem by the simplex method:

   Maximize $Z = 8\xi_1 + 19\xi_2 + 7\xi_3$

   subject to

$$3\xi_1 + 4\xi_2 + \xi_3 \leq 25$$
$$\xi_1 + 3\xi_2 + 3\xi_3 \leq 50$$

   Formulate the dual of this problem and solve it. [33]

One mark is allotted for neatness in the answers.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

[146]

ANNUAL EXAMINATIONS

Statistics-7: Econometrics

Date: 28.5.69. 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 any three questions. 2 marks are alloted for neatness in the answers.

1. Would you suggest least square estimation if both the dependent and independent variables are subject to errors of observation. Justify your answer. Give without proof Wald's and Bartlets methods of estimation in this connection. [16]

2. Show how a lagged dependent variable can appear as an explanatory variable with reference to the following:
   i) Koyck's model of distributed lag
   ii) Expectational and Adjustment models. [16]

3. State with proof the identifiability conditions in a simultaneous econometric model. [16]

4. Write short notes on any two of the following:
   i) Dummy variables,
   ii) Measurement of quality variation from the family budget data,
   iii) Supply function. [16]

Group B

Maximum Marks: 50

1. The following table gives the percentage distribution of persons by monthly per capita total expenditure classes along with average monthly per capita total expenditure as also average per capita expenditure on cereals and cereal substitutes for rural India during July 1959-June 1960.

exp. class	p.c. of population	per capita exp. on cereal and cereal substitutes (Rs.)	per capita total exp. (Rs.)
(1)	(2)	(3)	(4)
0 - 8	7.15	4.10	6.58
8 - 11	13.83	5.68	9.95
11 - 13	11.14	6.72	12.15
13 - 15	10.45	7.36	13.59
15 - 18	14.04	8.48	16.46
18 - 21	11.24	9.00	19.68
21 - 24	7.66	9.34	21.74
24 - 28	6.65	10.44	25.52
28 - 34	8.07	11.01	31.26
34 - 43	4.87	11.68	37.90
43 - 55	2.60	13.78	48.56
55 - above	2.30	16.40	86.01

Draw the concentration curve for per capita total expenditure and the specific concentration curve for the expenditure on cereals and cereal substitutes. Classify this group of commodities as necessary or luxury by a visual inspection of the curves.

Assuming that the Engel Curve (against total expenditure) for cereals and cereal substitutes is of constant elasticity form and that per capita total expenditure follows a log-normal distribution, find the constant expenditure elasticity of cereals and cereal substitutes. Comment if this estimate of elasticity corroborates your earlier classification of this group of commodities. [30]

2. Practical Records. [10]

3. Viva Voce. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69 [147]

ANNUAL. EXAMINATIONS

Statistics-7: Industrial Statistics Theory and Practical

Date: 29.5.69 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 1 and any other two questions from 2 to 4 in this group.

1.a) Explain the meaning of the following terms

(i) artificial variables (ii) extreme points (iii) basic feasible solution (iv) non-degenerate basic feasible solution.

b) State and prove the optimality condition to obtain the solution for a maximization linear programming problem using simplex method.

c) Show that an optimal solution to a linear programming problem occurs at an extreme point of the convex set of its feasible solutions. Derive the rules for transforming the simplex tableau to improve basic feasible solution while trying to reach the above extreme point at which the optimal solution occurs. [4+4+8]=[16]

2.a) Show that if $a_1, a_2, \ldots, a_m$ is a basis of the vector space V and

$$\beta = \sum_{j=1}^{m} c_j a_j$$

where c_j's are scalars such that $c_m \neq 0$ then $a_1, a_2, \ldots, a_{m-1}, \beta$ form a new basis of V. State the generalised form of the above theorem and indicate its role in solving a linear programming problem.

b) If the vector a_k which is not in the basis matrix B of a linear programming replaces the vector a_r $(r < m)$ in the basis matrix, show that the net decrease affected in the objective function is $\theta(z_k - c_k)$ where $z_k = C_B B^{-1} a_k$. What would you conclude when all $y_{ik} \leq 0$? [9+3]=[12]

3.a) If you are using two phase method, what conclusions can you draw under the following situations at the end of phase I?

i) Max $Z^* < 0$; one or more artificial vectors appear in the basis at positive level

ii) Max $Z^* = 0$; no artificial vectors appear in the basis

iii) Max $Z^* = 0$; one or more artificial vectors appear in the basis at a zero level

where Z^* denotes the objective function of phase I.

3.b) Show that if the primal problem has an optimal solution then the dual problem also has an optimal solution. Obtain solution to the following problem by solving its dual problem

$$\text{Max} \quad Z = x_1 + 2x_2$$
$$\text{subject to} \quad 3x_1 + 2x_2 \geq 6$$
$$x_1 + 6x_2 \geq 3$$
$$x_1, x_2 \geq 0$$

[3+9]=[12]

4. A company has four territories and four salesmen available for assignment. It is estimated that a typical salesman operating in each territory would bring in the following annual sales.

Territory	Annual sales
I	60,000
II	50,000
III	40,000
IV	30,000

The four salesmen are considered to differ in ability. It is estimated that working under the same conditions their yearly sales would be proportionately as 7:5:5:4 for the salesmen A,B,C and D respectively. If the criteria is the maximum expected total sales, obtain an optimal assignment. [12]

5. Practical Records and Viva Voce. [10]

Group B

Maximum Marks: 50

Answer Question 1 and any other two from 2 to 4 in this group.

1.a) Obtain the steady state probabilities for an unlimited queue served in order of arrival and a single service channel, assuming Poisson arrivals and exponential service. Find also:

i) average number of persons in the queue

ii) probability distribution for the waiting time of an arrival

iii) average waiting time of an arrival. [4+2+4+2]=[12]

b) At what average rate must a clerk at a supermarket counter work inorder to insure a probability of 0.90 that a customer will not have to wait longer than 12 minutes? it is assumed that there is only one counter to which customers arrive in a poisson fashion at an average rate of 15 per hour. The length of the service by clerk has an exponential distribution. [8]

In a machine shop there are N machines and stoppages occur completely randomly in running time, at the same rate λ for all machines, and independently for all machines. There is one operator always available for restarting stopped machines. The frequency distribution of service time is exponential with mean $1/\mu$.

Show that the number of machines running follows a truncated Poisson distribution of parameter ρ^{-1}, where ρ the servicing factor $= \lambda/\mu$. Derive expressions for operative utilisation and machine availability. [15]

3.a) Consider a machine shop where there are 8 automatic machines in operation. From past experience it is known that each machine will operate for an average period of 60 hours and then require an average of 40 hours of maintenance. Machine stoppages are known to occur in a poisson fashion and service is exponential. One hour's work on the machine produces a profit of Rs.10/-. If a service engineer is employed he should be paid wages of Rs.3/-, per hour. Find out how many service engineers should be employed?

b) Suppose it is possible to introduce some sort of preventive maintenance at a fixed cost of Rs.10/-, per hour for all machines. The effect of preventive maintenance is to increase the average running time of machine to 80 hours and decrease the repair time to 20 hours. In this situation how many service engineers are to be appointed? Decide whether it is worthwhile to have preventive maintenance. [9+6]=[15]

4. For a queueing model with random input and general service time, show that when the system is in a state of statistical equilibrium

Average queue length $E(n)=\rho + \dfrac{\rho^2 + \lambda^2 \sigma_v^2}{2(1-\rho)}$ and

Average waiting time $E(w)= \dfrac{\rho^2 + \lambda^2\sigma_v^2}{2\lambda(1-\rho)}$ where

$\rho = \lambda/\mu$; λ = mean arrival rate; μ = mean number of customers serviced in unit time and σ_v^2 is variance of the service time of a customer. [10+5]=[15]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1968-69

ANNUAL EXAMINATIONS

Statistics-8: Genetics

Date: 31.5.69 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: 46

Answer any two questions.

1.a) Describe with examples Mendel's first and second principles in the theory of Genetics and comment on their uses and applicability.

b) In an equilibrium population with regard to blood groups in O-A-B series, let the frequencies of genes O, A, B be respectively 0.72, 0.25 and 0.03. What is the probability that an individual known to be of type A is homozygous? What is the probability that an offspring resulting from the mating of two type A individuals is homozygous?

[12+11]=[23]

2.a) Show that under random mating any population will attain equilibrium with regard to one autosomal locus with two alleles in one generation.

b) Prove that for an autosomal locus with m alleles $A_1, A_2, \ldots, A_m$, a population with the following genotypic distribution

$$\sum_i p_i^2 A_i A_i + 2 \sum_{i<j} p_i p_j A_i A_j \qquad \text{is in}$$

equilibrium under random mating. [12+11]=[23]

3.a) Define 'Mendelian Population' and give a brief outline of the genetic theory of organic evolution.

b) Consider the case of a simple Mendelian character in which the genotypic array in the present generation is

$$p^2 AA + 2pqAa + q^2 aa, \qquad p + q = 1.$$

Suppose that the recessives aa are completely eliminated or sterilised, the mating being random otherwise in the present and all subsequent generations. Find the frequency of 'a' gene after n generation.

c) Suppose a recessive trait occurs in 1 in 1000 of a random mating population. How many generations of complete selection against the recessive individuals would be necessary to reduce the proportion to 1 in 1,000,000?

[6+12+5]=[23]

Group B

Maximum Marks: 44

Note: Answer question 2 and any one from the rest.

1.a) Write down the genotypes, phenotypes and phenotypic frequencies in terms of the gene frequencies for O-A-B blood group system under random mating.

b) Show that under repeated sib-mating the heterozygotic frequency in a population with two autosomal alleles 'A' and 'a' at one locus goes to zero at the rate

$$[(\frac{1+\sqrt{5}}{4})^{n+1} - (\frac{1-\sqrt{5}}{4})^{n+1}].$$

[4+18]=[22]

2.a) For M-N blood group system write down the different genotypic mating types, their frequencies under random mating and their segregation ratios (i.e., probabilities of producing different types of offsprings).

b) The following are the results of MN testing for two populations, Brahmin and Kaystha. Estimate the gene frequencies and test whether the two populations are homogeneous with respect to M-N blood groups

Genotypes	Observed Frequency	
	Brahmin	Kaystha
MM	79	39
MN	109	67
NN	42	34
Total	230	140

[6+16]=[22]

3.a) State and prove the general equilibrium law in a population with two autosomal alleles 'A' and 'a' at one locus.

b) Show that

$$m + (m-2)F = 0$$

where m denotes the correlation coefficient between the two parents and F, the overall inbreeding coefficient of the population.

c) Outline a few major factors due to which the original distributions are not reproduced in genetic surveys.

[11+7+4]=[22]

Viva Voce. [10]