INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73
Statistics-6: Official Statistics
and
M. Stat. Part I: 1971-72
Applied Statistics-2
ANNUAL EXAMINATION

[401]

Date: 14.8.1972 Maximum Marks: 100 Time: 3 hours

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

1. Describe the organisation and activities of any one of the statistical organisations which you have visited during your stay at the C.S.O. Mention the broad contents of one of the important publications of that organisation.

2. Describe the statistical system of India or any other country.

3. Enumerate the improvements made in the collection of data during the 1971 population census operations. How is the Census data utilised by a statistician, politician and the sociologist?

4. Describe briefly the nature of official statistics collected and compiled mentioning inter-alia the agency responsible as also the principal publication in any two of the following fields:-

   i) Agricultural Statistics, (ii) Foreign Trade Statistics,
   iii) Housing and Construction Statistics.

5. Define the Labour Force. Describe the purpose, methodology and content of Labour Force Surveys.

6. Write short notes on any two mentioning the scope and content of data available:-

   i) Road Transport Statistics, (ii) Road Statistics,
   iii) Railway Statistics.

7. Write short notes on any two mentioning the scope and content of data available:-

   i) Income Tax Statistics, (ii) Financial Statistics,
   iii) Company finance statistics.

8. Give an account of any one of the following official index numbers mentioning the agency compiling the index, the base period, item coverage, weighting diagram etc.

   i) index number of industrial production
   ii) consumer price index number for non-manual, non-agricultural employees.

9. Describe the methods of national income estimation explaining the methods adopted for preparation of estimates of national product in any one sector.

10. Describe the organisation responsible, methods of data collecting and content of information collected in the Annual Survey of Industries. How is the data utilised for planning purposes?

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

PERIODICAL EXAMINATION

Subject: Statistics-5: Statistical Methods (Theory and Practical)

Date: 11.12.72. Maximum Marks: 100 Time: 3 hours

Note: Answer any two from Group A and all the questions from Group B.
Marks allotted for each question are given in brackets [ ].

Group A

1. Suppose you are given two random samples drawn independently from two normal populations. Describe the common small-sample procedure for setting up confidence limits for the difference between the population means. State all the underlying assumptions carefully, and obtain the necessary sampling distribution in any manner you like. [4+4+12]=[20]

2.a) State the Lindeberg-Levy form of the multivariate central limit theorem.

b) Hence show that the joint distribution of sample raw moments (m_1', m_2' ... m_r') is asymptotically normal as $n \longrightarrow \infty$.

c) How do you find from here the a.d. of the sample central moment m_r? (Prove the necessary theorem.)

d) Obtain the following expression for asymptotic variance

$$a.v(m_2) = \frac{2\sigma^4}{n}$$

where σ denotes s.d., stating underlying assumptions. [4+4+9+3]=[20]

3.a) Define fractiles of a population (ξ_p) and fractile statistics ($\hat{\xi}_p$) of a sample.

b) Show that, under suitable conditions, the a.d. of $\hat{\xi}_p$ is normal.

c) What do you know about the a.d. of the sample inter-decile range $D_9 - D_1$, where D_1 and D_9 are respectively the first and the ninth deciles of the sample? [4+10+6]=[20]

Group B

In every case state the underlying assumptions

4. Consider the following test procedure: Reject the null hypothesis that the mean of a Poisson variable is $\lambda = 1$, if the sample mean based on 100 observations exceeds 1.2 or falls short of 0.8. Find the level of significance of the test and also its power against an alternative hypothesis H_1 : $\lambda = 1.1$. [6+6]=[12]

5. If it costs a rupee to draw one individual in a sample, how much would it cost, in sampling from a population with mean 200 and s.d. 25, to take a sample large enough to ensure that the sample mean lies within 0.1 % of the true value in 95 % of the cases? [8]

Please Turn Over

6. A random sample of 1000 men from the North of England showed their mean wage to be £ 2 7s per week with a s.d. of £ 1 8s. Another sample of 1500 men from the South of England gave a mean of £ 2 9s and a s.d. of £ 2. Discuss the suggestion that the mean rate of wages was higher in the southern region. [10]

7. EITHER

The following shows the wing length and tongue length, both in millimeters, of 10 bees:

Wing:	9.7	9.8	9.6	9.7	9.8	9.6	9.6	9.6	9.2	9.1	9.7
Tongue :	6.5	6.7	6.7	6.7	6.7	6.6	6.6	6.6	6.4	6.2	6.6

Test whether the two characteristics are significantly correlated and set up 95 % confidence limits for the true correlation coefficient. [4+4+6]=[14]

OR

Below are given the correlation coefficients (r) between intelligence test scores found in a study of the relative importance of hereditary and environmental factors:

	reared apart		living together	
	sample size	r	sample size	r
two brothers	50	0.235	40	0.342
twins	45	0.451	55	0.513

Comment on the figures using appropriate tests of significance. [14]

8. The following are the rates of diffusion of CO_2 through samples of two types of soil of different porosity : Through a fine soil (F) : 20, 31, 18, 23, 23, 28, 23, 26, 27, 26, 12, 17, 25; through a coarse soil (C): 19, 30, 32, 28, 15, 26, 35, 18, 25, 27, 35, 34.

Compare the means and the variabilities of the two series of measurements by means of suitable statistical tests. [9+7]=[16]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

403

PERIODICAL EXAMINATION

Statistics-7: Industrial Statistics (Theory and Practical)

Date: 18.12.72 Maximum Marks: 100 Time: 3 hours

Note: The paper carries 110 marks. You may answer any part of any question. The maximum you can score is 100. Marks allotted for each question are given in brackets [ ].

1. A housewife asks a butcher to grind up several cuts of beef to form a blend of equal parts of proteins and fats (50 % of each). The butcher being conscientious, wishes to do this at the least cost per pound of meat purchased.

	Chuck	Flank	Porter-house	Rib roast	Round	Rump	Sir-loin
% protein	19	20	16	17	19	16	17
% fat	16	18	25	23	11	28	20
cost 1 lb. in (Rs.)	15	16	2	3	7	10	8

Formulate the butcher's problem as a linear programming problem. [10]

2. What is a convex set? Show that the set of feasible solutions, and the set of optimal feasible solution to a linear programming problem are convex sets. [10]

3. Solve by simplex method:

$$3x_1 + 4x_2 + x_4 \leq 7$$
$$6x_1 + x_2 + 5x_4 \leq 3$$
$$4x_1 + 5x_2 + 2x_4 = 8$$
$$x_j \geq 0$$

$$\text{maximum } z = 3x_1 + 4x_2 + x_3 .$$ [20]

4. Show that if a linear programming problem has an optimal solution, then the dual problem also has an optimal solution and the optimal values of the objective function for the primal and its dual are the same.

Write down the dual of problem 3. [20]

5. Determine if there are $x_j \geq 0$ satisfying

$$4x_1 + 5x_2 + x_3 = 10$$
$$2x_1 + 3x_2 - 4x_3 = 7$$
$$x_1 + 4x_2 - 2x_3 = 6$$

(Introduce artificial variables and use Phase I of L.P.) [20]

p.t.o.

6. Show that if for some $Z_j - C_j > 0$, for a minimisation problem, all $y_j = B^{-1} a_j \leq 0$ then the problem has an unbounded solution.(The symbols have their usual meanings.) [10]

7.a) When is a basic feasible solution to a linear programming problem, said to be degenerate?

b) Suppose in an optimal solution to the dual programme, $u_1 > 0$. What can we say about the optimal solution to the primal?

c) Suppose the primal problem has multiple optimal solutions. What can we say about the dual problem?

d) Consider the problem maximise $c'X$, subject to $AX \leq b$ and $X \geq 0$. If $b_i = 0 \; \forall_i = 1,2,\ldots, m$ then show that that if an optimal solution exists, then the optimal value of $c'X = 0$.

e) For the problem in (d) suppose an optimal solution does not exist. What can you say about the dual problem? [5 X4]=[20

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

[404]

PERIODICAL EXAMINATION

Statistics-7: Planning Techniques

Date: 1.1.1973 Maximum Marks: 100 Time: 3 hours

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

1.a) Solve the following linear programming problem by the simplex method [8]

$$\text{Maximise} \quad Z = 4x_1 + x_2 + 3x_3 + 5x_4$$

$$\text{subject to} \quad \begin{aligned} -4x_1 + 6x_2 + 5x_3 - 4x_4 &\leq 20 \\ 3x_1 - 2x_2 + 4x_3 + x_4 &\leq 10 \\ 8x_1 - 3x_2 + 3x_3 + 2x_4 &\leq 20 \end{aligned}$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0.$$

b) While solving the above what characteristics of the problem do you observe? Explain your answer. [2]

c) Show, using the simplex algebra, how and under what conditions the optimal solution becomes unbounded. [10]

2.a) Solve the following linear programming problem by the simplex method [10]

$$\text{Maximise} \quad Z = 3x_1 + 2x_2$$

$$\text{subject to} \quad \begin{aligned} 8x_1 + 2x_2 &\leq 18 \\ x_1 &\leq 4 \\ x_2 &\leq 6 \end{aligned}$$

$$x_1 \geq 0, \quad x_2 \geq 0$$

b) What is the simplex criterion for the uniqueness of the optimal solution ? Is the optimal solution for the problem in 2(a) unique ? If not, find another solution. [8]

3. For the linear programming problem of a firm

$$\text{Maximise} \quad Z = c'x$$

$$\text{subject to} \quad AX \leq b$$

$$x \geq 0$$

where X is an n-vector of activity levels, A is an $m \times n$ matrix of technological coefficients, b is an m-vector of the endowment of fixed resources and c is an n-vector of profit coefficients, write down the dual and give economic interpretations of the dual objective function, variables and constraints. State properties of optimal solutions to this dual and its primal and give economic interpretations of these properties. [15 +10]=[25]

Give a description of the Leontief static open input-output model stating clearly the technological assumptions and their economic implications. [20]

Please Turn Over

5.a) Write down the dual of the following linear programming problem

$$\begin{aligned} \text{Minimise} \quad & Z = x_1 + x_2 + x_3 \\ \text{subject to} \quad & x_1 - 3x_2 + 4x_3 = 5 \\ & x_1 - 2x_2 \leq 3 \\ & 2x_2 - x_3 \geq 4 \end{aligned}$$

$x_1 \geq 0$, $x_2 \geq 0$, x_3 unrestricted in sign.

b) Suppose that a manufacturer has m factories and supplies n localities. Suppose also that the costs c_{ij} $(i = 1, \ldots, m;\ j = 1, \ldots, n)$ of shipping a ton of product from factory i to locality j are given, that the capacities k_i $(i = 1, \ldots, m)$ of the factories are known and that the minimum number of tons r_j $(j = 1, \ldots, n)$ to be supplied to each locality is fixed. The problem for the manufacturer is to find a pattern of shipments that involves the least possible total transportation cost.

The linear programming formulation of the above problem is as follows

$$\begin{aligned} \text{Minimise} \quad & T = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\ \text{subject to} \quad & \sum_{j=1}^{n} x_{ij} \leq k_i \quad i = 1, \ldots, m \\ & \sum_{i=1}^{m} x_{ij} \geq r_j \quad j = 1, \ldots, n \\ & x_{ij} \geq 0 \quad (i=1, \ldots, m;\ j=1, \ldots, n) \end{aligned}$$

where x_{ij} denotes the number of tons shipped from factory i to locality j.

Write down the dual of the above linear programming problem. Suggest economic interpretations of the variables, constraints and the objective function of the dual. [5+7]=[12]

WISH YOU A HAPPY NEW YEAR

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

PERIODICAL EXAMINATION

Statistics-6: Sample Surveys (Theory and Practical)

Date: 8.1.73 Maximum Marks: 100 Time: 3 hours

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

1. EITHER

When a simple random sample of size n is drawn without replacement from a population of size N, give an unbiased estimator of the population mean of a characteristic y and derive its variance and an unbiased estimator for the variance. Hence suggest methods of estimating a population proportion and a sub-population total.

OR

When sampling in all strata is simple random sampling without replacement, prove that under Neyman's optimum allocation of a fixed total sample size to the different strata the variance of the stratified sample estimator of the population mean of a character y is never greater than the variance under proportional allocation of the sample size which in turn is, under certain conditions, not greater than the variance of an unstratified simple random sample estimator. Mention clearly due to what factors one procedure gains precision over another procedure.

2. EITHER

Explain how to draw a sample of size n from a population of size N by (a) linear systematic sampling and (b) circular systematic sampling. Give unbiased estimators of the population mean of a character y in both these sampling procedures. State whether you can estimate unbiasedly the variance of these estimators on the basis of the sample observations.

If instead of a single systematic (linear or circular) sample, we draw m independent samples each of size n, give an estimate of the population mean based on all the samples put together. Also derive an unbiased estimator of the variance of this pooled estimator.

OR

i) Obtain a comparison of the variances of a sample mean based on a simple random sample without replacement and a linear systematic sample (assuming population size N is an integral multiple of sample size n) of equal size in terms of the 'intra-class correlation' between pairs of sample units.

ii) Show that in a population where for the character y the values of the units are in arithmetical progression, the efficiency of systematic sampling is approximately n times that of simple random sampling without replacement in estimating the population mean $\bar{Y}$ on the basis of a sample of size n.

p.t.o.

3. EITHER

Describe briefly the procedures of selecting one unit with probability proportional to the size (x) of the unit by (i) cumulative total method and (ii) Lahiri's method and discuss their relative merits.

Give an unbiased estimator for the population total of a character y on the basis of a sample of size n selected with probabilities proportional to the sizes (x) of the units with replacement and derive an unbiased estimator of the variance of that estimator.

OR

Give an unbiased estimator for the population total of a character y as proposed by Horvitz and Thompson for a general sampling scheme without replacement and derive its variance and unbiased estimators of the variance as proposed by (i) Horvitz and Thompson and (ii) Yates and Grundy. State whether these variance estimators may assume negative values or not.

4. A population consisting of 112 villages is stratified (by size of the agricultural area of the villages) into 3 strata, as shown in Col. (2) of the table below. The number N_i of villages in different strata is given in Col. (3). The number of villages n_i sampled from the i-th stratum is given in Col. (4). The sampling scheme used within each stratum is specified in Col. (5). In Col. (6) the yield of jute is given for the selected villages, and in Col. (7) auxiliary information, where available, is given.

Stratum	Size of the village (in acres of agricultural area)	No. of villages in each stratum (N_i)	No. of villages selected from the i-th stratum (n_i)	Sampling scheme	Yield of Jute (in tonnes) (y_{ij})	Auxiliary information on agricultural area
(1)	(2)	(3)	(4)	(5)	(6)	(7)
I	≤ 40	51	6	Probability proportional to agricultural area, with replacement	75 101 5 78 79 45	25 39 2 26 26 16
II	40 - 80	38	4	Simple random sampling without replacement	247 238 359 125	
III	> 80	23	6 (3 in each sub-sample)	circular systematic sampling	Sub-sample 1 427 326 481 Sub-sample 2 401 348 498	

It is also known that the total agricultural area for Stratum I is 1237 acres.

i) Estimate the total yield of jute for the population

ii) Obtain an estimate of the variance of the above estimate.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

[406]

PERIODICAL EXAMINATION

Statistics - 8: Demography (Theory and Practical)

Date: 15.1.73 Maximum Marks: 50 Time: $1\frac{1}{2}$ hours

Note: Answer any three questions; Question No.4 is compulsory. Marks allotted for each question are given in brackets [].

1.a) Enumerate the basic reasons for the deficiency in the Indian vital registration data.

b) Define 'underlying cause of death'. Examine briefly its importance in public health administration. [10+5]=[15]

2. Explain why mortality situations of two places cannot usually be compared on the basis of conventional I.M.R..

Describe the construction of adjusted I.M.R. for the purpose by approximate matching of births and deaths and also by method of exact matching. [3+12]=[15]

3. What is called a replacement index? What does it aim to measure? How is it related to N.R.R.?

Give the concept of μ_x, as the instantaneous force of mortality.

Derive the mathematical relationship between a complete and a curtate expectation of life. [5+5+5]=[15]

4. The table below gives the age-specific fertility rates and age-sp. mortality rates of the current and L.T. population as also the age-distribution of the current population and a column for life table population L_x, taken as the Standard population.

Compute the values of any three: (a) Net reproduction rate, (b) Standardised mortality rate, (c) replacement index and (d) Vital index, for the current population.

Table

Age-group	Sp.fert.rate (considering fem. children	Age-sp. mort. rate/1000 fem. popul. only L. T.	Current	Current fem. popul.	L. T. fem. population as standard
(1)	(2)	(3)	(4)	(5)	(6)
0 - 4		81.0	74.1	32656	410
5 - 9		4.3	4.1	31557	378
10 - 14		3.1	2.7	22994	369
15 - 19	.128	4.0	4.2	17255	360
20 - 24	.143	6.5	5.5	19085	350
25 - 29	.129	6.8	7.1	18024	339
30 - 34	.097	7.1	6.4	14852	319
35 - 39	.069	7.2	7.3	11841	291
40 - 44	.029	9.0	8.5	10754	260
45 - 49	.009	10.8	11.0	6307	229
50 +		67.1	60.9	25008	745
Total				210313	4050

[20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1972-73 [407]

PERIODICAL EXAMINATION

Statistics-4: Probability

Date: 22.1.73 Maximum Marks: 100 Time: 3 hours

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

1.a) Prove that if $I_1, I_2, \ldots$ are all intervals such that $(I_1 \cup I_2 \cup \ldots)$ is a finite interval J, then

$$\lambda(J) = \sum_1^{\infty} \lambda(I_n).$$

Here λ stands for length. [10]

b) $F(x)$ is a continuous function defined on $(-\infty, \infty)$. It is strictly increasing in the sense that $x_1 > x_2$ implies that $F(x_1) > F(x_2)$. If I is any finite interval, put $\mu(I) = F(b) - F(a)$ where a and b are respectively the left-hand and right-hand end-points of I. If I is a degenerate interval, put $\mu(I) = 0$.

Prove that if $I_1, I_2, \ldots$ are all intervals such that $(I_1 \cup I_2 \cup \ldots)$ is a finite interval J, then

$$\mu(J) = \sum_1^{\infty} \mu(I_n).$$ [15]

2.a) Determine a complex number $z = (a + ib)$ such that $z^2 = (1 + 2i)$. You must show how you get your answer. [12]

b) With the complex number $z = (a + ib)$, we associate the matrix

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = M(z).$$

i) Prove that for any two complex numbers z_1 and z_2,

$$M(z_1 z_2) = M(z_1) \cdot M(z_2).$$ [5]

ii) Show that for every complex number z, $|z|^2$ = determinant of $M(z)$. Hence deduce, using a theorem from matrix theory, that $|z_1 z_2| = |z_1| \cdot |z_2|$ always. [12]

3.a) When do we say that a set of real numbers has Lebesgue measure zero? Prove that the union of two sets of Lebesgue measure zero is a set of Lebesgue measure zero. [10]

b) What is a probability distribution P on $(-\infty, \infty)$? [6]

c) Let P be a probability distribution on $(-\infty, \infty)$. Put $F(x) = P(-\infty, x]$. Prove rigorously that $F(x) \to 1$ as $x \to \infty$. [15]

d) Let P be a probability distribution on $(-\infty, \infty)$. The spectrum S of P is defined as follows:

The number α is put in the set S if and only if for every $\partial > 0$, $P(\alpha - \partial, \alpha + \partial) > 0$.

The spectrum is thus a set of real numbers.

Prove that the spectrum is a closed set. [15]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73 [408]

PERIODICAL EXAMINATION

Statistics-8: Educational Statistics (Theory and Practical)

Date: 29.1.73 Maximum Marks: 100 Time: 3 hours

Note: Answer Q. 1 and Q. 2 and any three from the rest. Marks allotted for each question are given in brackets [ ].

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

a) Test score as a function of ability,
b) Wrong score and its use,
c) Estimation of reliability in case of speed test,
d) Errors of measurement and errors of prediction. [8+8+8]=[24]

2. EITHER

Consider the case in which the validities of the two tests are different. Suppose test A has a validity coefficient of 0.66 and a reliability of 0.72 and test B has a validity of 0.70 and a reliability of 0.90. If each test has unit length, at what length will the tests have equal validity? What would validities of test A and test B if both tests were increased to infinite length? [16]

OR

Write brief notes including equation on the estimation of factor loadings by the Centroid method. Show that the number of common factors is equal to the rank of the correlation matrix.

3.a) Let ρ_{XY} be the validity of measurement X with respect to measurement Y, and $\rho_{XX'}$ be the reliability of measurement X. Suppose that measurement X is lengthened and its new (increased) reliability is denoted by $\rho'_{XX'}$. The validity of X with respect to Y will also be increased. Denote this increased validity by ρ'_{XY}. (Thus we have $\rho'_{XX'} > \rho_{XX'}$, $\rho'_{XY} > \rho_{XY}$). Show that

$$\rho_{XY} / \rho'_{XY} = \sqrt{\rho_{XX'} / \rho'_{XX'}}$$

Further show that the ratio of the validity coefficient to the index of reliability does not change with the increase of test length.

b) Briefly describe Spearman's two-factor theory. [14+6]=[20]

4.a) Write short notes on 'inference of true gain from the observed gain in score'.

b) Show that the reliability of ρ_k of a test at length k in terms of its reliability $\rho_{k'}$ at length k' is

$$\rho_k = \frac{k\rho_{k'}}{k' + (k - k')\,\rho_{k'}}.$$ [12+8]=[20]

p.t.o.

5.a) Let Y_1 and Y_2 be measurements with true scores T_1 and T_2 and let $X = Y_1 + Y_2$ be a composite measurement with true score T. Then prove that

$$\rho^2_{XT} \geq 2 \left[1 - \frac{\sigma^2(Y_1) + \sigma^2(Y_2)}{\sigma^2(X)} \right]$$

where

ρ_{XT} = the correlation between X and T

$\sigma(Y_i)$ = the standard deviation of Y_i, $i = 1,2$

$\sigma(X)$ = the standard deviation of X.

b) Discuss: Reliability of essay type of examination. [13+7=[20]

6. Write short notes on the following:

a) Index of speededness of a test as suggested by Cronbach and Warrington .

b) Item analysis

c) Norms. [8+6+6]=[20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

[409]

MID-YEAR EXAMINATION

Statistics-7: Planning Techniques (Theory and Practical)

Date: 19.2.73. Maximum Marks: 100 Time: 4 hours

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

1.a) Solve the following problem by the simplex method

$$\text{Maximise} \quad z = 2x_1 + 3x_2$$

$$\text{subject to} \quad \begin{aligned} 2x_1 + x_2 &\leq 80 \\ x_1 + x_2 &\leq 50 \\ x_1 + 2x_2 &\leq 80 \\ x_1 \geq 0,\ x_2 &\geq 0 \end{aligned}$$ [7]

b) Write down the dual of the problem in part (a) and obtain the optimal solution of this dual from your simplex computations of the primal. [5]

c) Show algebrically how the optimal solution of the dual for the kind of problem as in (a) could be obtained from the simplex computations of the primal without actually solving the dual problem. [8]

2. State a linear programming problem with the open static input-output system as constraints. Using this linear programming formulation and its dual discuss the different properties of the system and also give economic interpretations of the primal and dual problems. [23]

3. State and prove the non-substitution theorem for the open Static Leontief input-output system. Give an intuitive economic justification of the result. [20]

4. Give a description of the Dorfman, Samuelson and Solow version of the open dynamic Leontief system stating clearly the different assumptions made. For this system derive the algebraic expression of the efficiency locus and show the nature of the locus graphically for the two-industry case. In what sense points on this locus are efficient? [22]

5. Given below in (a) - (e) are five simplex tableau (using standard notation) for five linear programming problems where the problems in (a) and (b) are maximisation and those in (c) - (e) are minimization problems. Examine carefully and give your observations about what these tableau tell us. Justify your answer in each case. [5 ×3]=[15]

(a)

Basis	c	P_0	2	4	1	1	0	0	0
			P_1	P_2	P_3	P_4	P_5	P_6	P_7
P_2	4	1	0	1	0	6/15	6/15	-1/5	0
P_1	2	1	1	0	0	-1/5	-1/5	3/5	0
P_3	1	1/2	0	0	1	9/60	-6/60	1/20	1/4
$z_j - c_j$		$6\frac{1}{2}$	0	0	0	7/20	12/20	9/20	1/4

PLEASE TURN OVER

5. (contd.)

b)

Basis	c	P_0	4 P_1	1 P_2	3 P_3	5 P_4	0 P_5	0 P_6	0 P_7
P_5	0	20	-4	6	5	-4	1	0	0
P_6	0	10	3	-2	4	1	0	1	0
P_7	0	20	8	-3	3	2	0	0	1
z_j-c_j		0	-4	-1	-3	-5	0	0	0

c)

Basis	c	P_0	0 P_1	1 P_2	-3 P_3	0 P_4	2 P_5	0 P_6
P_2	1	4	2/5	1	0	1/10	4/5	0
P_3	-3	5	1/5	0	1	3/10	2/5	0
P_6	0	11	1	0	0	-1/2	10	1
z_j-c_j		-11	-1/5	0	0	-8/10	-12/5	0

d)

Basis	c	P_0	0 P_1	0 P_2	0 P_3	-28 P_4	-1 P_5	-2 P_6
P_6	-2	7	3	0	0	14	1	1
P_2	0	19	6	1	0	44	5/2	0
P_3	0	0	0	0	1	3	0	0
z_j-c_j		-14	-6	0	0	0	-1	0

e)

Basis	c	P_0	0 P_1	0 P_2	0 P_3	-107 P_4	-1 P_5	-2 P_6
P_1	0	7/3	1	0	0	14/3	1/3	-1/3
P_2	0	5	0	1	0	16	1/2	-2
P_3	0	0	0	0	1	3	0	0
z_j-c_j		0	0	0	0	107	1	2

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

[410]

MID-YEAR EXAMINATION

Statistics-7: Industrial Statistics (Theory and Practical)

Date: 20.2.73 Maximum Marks: 100 Time: 4 hours

Note: The paper carries 110 marks. Answer as many questions as you can. Maximum you can score is 100. Marks allotted for each question are given in brackets [ ].

1.a) Explain what is meant by the statement that a process is out of statistical control.

b) Comment on the statement: 'If a process is under statistical control then all items produced are satisfactory.'

c) Comment on the statement: 'In the ideal situation specification limits and control limits should be identical'. [3×4]=[12]

2. a) Explain the justification for using in control charts the 3σ-limits without regard to the actual probability distribution of the quality characteristic plotted.

b) What is the statistical hypothesis tested when a p-chart is used. What is the testing procedure? [4+4]=[8]

3. What is process capability? Depending on how the process capability is related to the specification limits what are the various possible conclusions and actions one can take? [10]

4. The following data give the results of inspection of enamel plates of same size for spots.

Plate Number	Number of spots	Plate Number	Number of spots	Plate Number	Number of spots
1	8	9	10	17	35
2	7	10	10	18	10
3	9	11	18	19	23
4	11	12	19	20	11
5	12	13	10	21	13
6	8	14	5	22	16
7	10	15	28	23	14
8	18	16	24	24	13

a) Examine whether the process is under statistical control and set standards for the average number of spots per plate

b) Obtain the U.C.L. at 1% level of significance for further production.

c) Suppose the average number of spots per plate jumps to 20. What is the probability that this shift will be detected by the next two samples. (A shift in the average value is suspected when an observation falls outside the control limits.) [12+4+4]=[20]

PLEASE TURN OVER

5.a) A sample of 30 units selected at random from a process producing a fraction defective. .05 contains 4 defectives compute the exact probability of this event and its poisson and Normal approximations. [1

6.a) What is operating characteristic curve of a sampling plan.

b) Explain the double sampling procedure.

c) Explain how Dodge Romig's single sampling plans with given values of AOQL are constructed. [4+6+10]=[

7. What Dodge Romig AOQL sampling plan would you use under the following conditions?

a)

Single/Double	AOQL ./. defective	lot size	process average
Single	2.0 ./.	300	1.25 ./.
Double	1.5 ./.	850	.046 ./.
Single	2.0 ./.	508	.85 ./.
Double	1.5 ./.	200	.38 ./.

What are the LTPD's of the above plans. [2

b) Explain the following terms :

i) AOQL
ii) Producers' risk
iii) Consumers' risk
iv) Average amount of Inspection (AOI)
v) LTPD . [5 X2]=[1

INDIAN STATISTICAL INSTITUTE [411]
Research and Training School
B.Stat. Part IV: 1972-73

MID-YEAR EXAMINATION

Statistics-5: Statistical Methods Theory

Date: 21.2.73. Maximum Marks: 100 Time: 3 hours

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

1. Let $\pi_1, \pi_2, \ldots, \pi_k$ be the probabilities of the k classes of a multinomial distribution and $n_1, n_2, \ldots, n_k$ be the observed frequencies of these classes in a sample of $n = \Sigma n_i$ observations.

   Starting from the basic results on the asymptotic distribution of linear and quadratic functions of the class frequencies, find the asymptotic distribution of
   $$\chi^2 = \sum_i (n_i - n\pi_i)^2/n\pi_i .$$ [20]

2. When does the conditional joint distribution of a set of independent Poisson variables have the multinomial form? Starting from this result, obtain the asymptotic distribution of the chi-square measure of goodness of fit proposed in Q. 1 above. [6+14]=[20]

3. What is meant by the test of independence in a contingency table? Describe the chi-square test of independence and derive the asymptotic distribution of the test statistic. (You need not prove every result.) [3+3+14]=[20]

4. Obtain the z-transformation of the correlation coefficient as a variance stabilizing transformation. What are the other advantages of this transformation? Outline the various uses of the z-transformation in statistical inference. [4+4+12]=[20]

5. Give an account of the Kolmogorov (one-sample) test of goodness of fit, mentioning its advantages and dis-advantages compared with the chi-square test. Mention in particular the problem of constructing a confidence region for the unknown distribution function. [8+6+6]=[20]

6. You are given n_i observations on a r.v. y corresponding to the value x_i of a non-stochastic variable $x (i = 1, 2, \ldots, k)$. Stating all necessary assumptions and theorems, obtain the test procedures for testing the under-mentioned hypotheses: (a) $E(y|x = x_i) = \text{constant}, \forall i$, (b) $E(y|x = x_i) = \alpha + \beta x_i$, $i = 1, \ldots, K$, where α and β are constants. [6+6+8]=[20]

7. Write short notes on any two:
   (i) The importance of the normal distribution in large sample theory.
   (ii) The P_λ test for combination of probabilities.
   (iii) Uses of orthogonal polynomials in curvilinear regression analysis. [2×10]=[20]

INDIAN STATISTICAL INSTITUTE [412]
Research and Training School
B.Stat. Part IV: 1972-73

MID-YEAR EXAMINATION

Statistics-5: Statistical Methods Practical

Date: 22.2.73 Maximum Marks: 100 Time: 3 hours

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

In every case, state your assumptions carefully

1. A statistician rejects the hypothesis that the mean μ of a normal population (with variance = 4) is 1 if the mean of a random sample of 25 observations does not lie between 0.2 and 1.8. Compute the power of the test against the alternatives $\mu = 0$ and 1.5, and also the level of significance of the test. [15]

2. The additional hours of sleep gained by 8 patients after using each of 2 drugs are shown below:

patient no.	1	2	3	4	5	6	7	8
addl. hours:								
drug 1	2.1	0.2	0.9	3.8	3.5	-0.2	-1.3	-0.3
drug 2	3.6	4.7	1.8	5.5	4.6	-0.3	-0.4	1.9

Test whether the second drug gives, on the average, at least one hour more of sleep than the first drug. [15]

3. Below is given the frequency distribution of 196 lots of manufactured articles according to the number of defectives found in them. The expected frequencies are based on a Poisson distribution fitted to the data.

no. of defectives:	0	1	2	3	4	5	6	7	8	9	10
obs. frequency:	7	33	54	37	34	16	8	5	1	1	0
exp. frequency:	10.78	31.28	45.35	43.84	31.78	18.43	8.91	3.69	1.34	0.43	0.17

Test the goodness of fit of the Poisson distribution to the data. [12]

4. The following data were collected in an experiment on the therapeutic value of a drug on guinea-pigs suffering from certain disease:

	no. of guinea-pigs	
	recovered	died
treated	9	6
not treated	1	7

Test whether the beneficial effects of the drug are significant by using the χ^2-test and also the exact probability test. [7+8]=[15]

PLEASE TURN OVER

5. The following shows the estimated variances of a certain measurement taken on samples of individuals drawn from four different caste-groups living in a region:

caste-group :	1	2	3	4
sample size :	35	40	62	37
variance (cm^2):	9.42	10.89	8.32	11.35

Test whether the true variances are equal, making suitable assumptions. [1

6. The following sums were obtained in an analysis of 11 pairs of observations on the no. of licensed vehicles (x) and the number of road casualties (y):

$$\Sigma x = 5711, \quad \Sigma y = 2396, \quad \Sigma x^2 = 3134543, \quad \Sigma xy = 1296836$$
$$\Sigma y^2 = 539512$$

i) Set up 95 % confidence limits for the correlation coefficient between x and y.

ii) Assuming that the regression of y on x is linear, test whether the intercept is significantly different from zero. [10+10]=[2

7. Practical Records. [1

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73 [415]

MID-YEAR EXAMINATION

Statistics-6: Sample Surveys-Theory

Date: 23.2.73 Maximum Marks: 100 Time: 3 hours

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

1. A simple random sample of size 2 is drawn without replacement from a population (Y_1, Y_2, Y_3) of 3 units. Corresponding to the three possible samples, $s_1 = (y_1, y_2)$, $s_2 = (y_1, y_3)$ and $s_3 = (y_2, y_3)$, let a linear estimator e(s) for estimating the population mean be defined as follows:

$$e(s_1) = \frac{2}{3} y_1 + \frac{1}{2} y_2$$
$$e(s_2) = \frac{1}{3} y_1 + \frac{1}{2} y_3$$
$$e(s_3) = \frac{1}{2} y_2 + \frac{1}{2} y_3$$

a) Show that the estimator e(s) is unbiased for the population mean for all values of Y_1, Y_2 and Y_3.

b) Find the variance of the estimator e(s) and its value when $Y_1 = 1$, $Y_2 = 2$, and $Y_3 = 3$ and comment on the following statement:

'In simple random sampling without replacement the sample mean is the uniformly minimum variance unbiased estimator in the class of all linear unbiased estimators of the population mean'.

2. i) A finite population of N units having the values Y_i, $i = 1,2,\ldots, N$; are presumed to be drawn at random from a superpopulation in which $E(Y_i) = \mu$, $V(Y_i) = \sigma_i^2$ and $Cov(Y_i, Y_j) = 0$, $(i \neq j)$. Obtain a comparison of the expected variances of the sample mean based on samples of size n selected from this finite population (a) systematically (assume N is a multiple of n), and (b) with simple random sampling without replacement, and comment on your result.

ii) A 10 % systematic sample is selected from a list of persons, in which families have been arranged by age. The average size of a family may be taken as five members. For which of the following characteristics do you expect this sample to be more efficient than one drawn with simple random sampling without replacement and why ?

(a) proportion of persons belonging to different religions;

(b) average age of males; and

(c) proportion of children going to school.

A population of N units is divided at random into n groups and one unit is selected with probability proportional to size from within each of the random groups. Give an unbiased estimator for the population total (Y) of a character y under this scheme for which

(a) the variance is minimum if the groups are of equal size, and

PLEASE TURN OVER

(b) the minimum variance is less than the variance of an unbiased estimator of Y based on a sample of size n drawn from the entire population with probability proportional to size and with replacement.

4. Describe the Midzuno-Sen sampling scheme using the information on an auxiliary character x and show that with this scheme

(a) Yates-Grundy estimator for the variance of Horvitz-Thompson estimator of the population total of a character y never admits negative values; and

(b) the ratio of sample totals of y and x is an unbiased estimator for the ratio of population totals of y and x.

5. A population consists of N primary units, each containing M secondary units. To estimate the population mean ($\bar{Y}$) of a character y the following three sampling schemes are considered:

i) n primary units to be selected with equal probability and without replacement;

ii) $K = \frac{nm}{M}$ (assumed to be an integer) primary units to be drawn with equal probability and with replacement and, within each selected primary unit, all the M secondary units to be enumerated; and

iii) nm secondary units to be drawn directly from the NM secondary units with equal probability and with replacement.

Give unbiased estimators of $\bar{Y}$ in the three schemes and obtain a comparison of their variances and comment on your results.

6. A population consists of N primary units, and the i^{th} primary unit contains M_i secondary units, $(i = 1,2,\ldots, N)$. suppose a sample n primary units is drawn with probabilities proportional to size and with replacement, and in the i^{th} selected primary unit m_i secondary units are selected from the M_i secondary units with simple random sampling without replacement, $(i = 1,2,\ldots, n)$.

a) Under the above set up, give an unbiased estimator for the population total of a character y, and derive its variance and an unbiased estimator of the variance.

b) Keeping the selection of primary units as given above, whatever be the sampling design adopted for the selection of secondary units within the selected primary units how do you estimate the variance of an estimator of the population total without estimating the variation within primary units?

7. In a given two-stage sample design let n be the number of primary units selected out of N primary units, and m_i be the number of secondary units selected within the i^{th} selected primary unit $(i = 1,2,\ldots, n)$, and the variance function of an estimator $\hat{Y}$ of the population total of a character y be of the form:

$$V(\hat{Y}) = \frac{1}{n} [A_1 + \sum_{i=1}^{N} \frac{A_{2i}}{m_i}] + A_3,$$

where A_1, A_{2i} and A_3 are population parameters independent of the sample sizes n and m_i.

a) Obtain an optimum allocation of the total sample size in terms of secondary units to the selected primary units such that the average number of sample secondary units per selected primary unit is a given number m.

b) For the optimum allocation achieved at (a), assuming a suitable cost function and a fixed total cost, derive the optimum value of m and n and the corresponding optimum variance of the estimator $\hat{Y}$.

8.a) In a general stratified sampling design propose the 'combined' and 'separate' ratio estimators for the ratio of population totals of two characters y and x and discuss their relative merits from the point of view of bias and precision.

b) Under what circumstances there is need to use ratio estimators which are less biased (than the usual ratio estimators) or completely free from bias? When the total sample is drawn in terms of m independent sub-samples each of same size and selected according to the same sample design, which are the possible biased ratio estimators that may be computed? Suggest adjustments by which those biased ratio estimators may be made less biased or completely freed from bias.

INDIAN STATISTICAL INSTITUTE [414]
Research and Training School
B.Stat. Part IV 1972-73

MID-YEAR EXAMINATION

Statistics-6 : Sample Surveys (Practical)

Date: 24.2.1973 Maximum Marks: 100 Time: 3 hours

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

1. Given below is an artificial population of size 4 for which the values (y_i) of a variate y and the corresponding initial probabilities of selection (P_i) are noted. By drawing all the six possible samples of size two from this population, compute the exact variances of the following unbiased estimators of the population total of y in the specified sampling schemes and comment on your results.

(a) varying probabilities of selection with replacement and the usual unbiased estimator in the scheme.

(b) varying probabilities of selection without replacement and
  (i) Des Raj's ordered unbiased estimator,
  (ii) Murty's unordered unbiased estimator,
and (iii) Hovitz-Thompson estimator.

Artificial population of size 4

Sl.no.of the unit	P_i	y_i
1	0.1	0.8
2	0.2	1.4
3	0.3	1.8
4	0.4	2.0

[40]

2. From a population of 151 farms, a sample of 21 farms has been selected with equal probability and without replacement. For each selected farm the total area under crops (x) and the area under oats (y) have been recorded. The total area under crops for the 151 farms is 31197 acres.

p. t. o.

2. (Contd...)

Sample data on total area under crops (x) and area under oats (y) for 21 farms.

Sl.no.of sample farm	Total area under crops in acres(x)	Area under oats in acres (y)
(1)	(2)	(3)
1	218	22
2	230	25
3	222	32
4	150	34
5	271	38
6	218	39
7	174	40
8	195	45
9	180	49
10	208	51
11	238	52
12	228	56
13	255	56
14	216	61
15	270	62
16	197	67
17	235	71
18	155	76
19	281	76
20	243	92
21	276	94

(a) Compute estimates of total area under oats for the 151 farms

(i) without using information on total area under crops,

(ii) using information on total area under crops and a ratio estimator, and

(iii) using information on total area under crops and a regression estimator.

(b) Obtain estimates of variance for the above estimates and comment on your results.

[40]

3. A population consists of 69 farms (N) and each farm has a number of fields (M_i) growing wheat. To estimate the total area under wheat in the population, a sample of 13 farms (n) has been chosen with equal probability and without replacement, and within each selected farm a number (m_i) of fields growing wheat have been selected again with equal probability and without replacement. The data on this two-stage sample are presented in the following table.

3. (Contd...)

Area under wheat (y) for the sampled fields and other data.

Sl. no. of selected farm	Total number of fields growing wheat in the selected farm (M_i)	Number of selected fields growing wheat (m_i)	Area under wheat for the selected fields in acres (y)
(1)	(2)	(3)	(4)
1	2	2	17,20
2	2	2	14,26
3	2	2	13,59
4	4	2	14,14
5	4	2	18,20
6	4	2	20,21
7	4	2	16,25
8	4	2	20,22
9	5	2	4,12
10	6	2	13,16
11	6	2	12,27
12	7	3	16,17,23
13	7	3	16,20,25

(a) Obtain an unbiased estimate for the total area under wheat in the 69 farms.

(b) Compute an unbiased estimate of the variance of the estimate obtained above.

[40]

4. Practical Records

[20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

[415]

MID-YEAR EXAMINATION

Statistics-4: Probability

Date: 26.2.73 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: 25

Answer all the questions.

1. Let $\underline{\underline{A}}$ be a σ-field of subsets of a set X and let P be a probability on $\underline{\underline{A}}$. If $A_1, A_2, \ldots$ is a sequence of sets from $\underline{\underline{A}}$.

   a) Prove that there exists a sequence $B_1, B_2, \ldots$ of sets from $\underline{\underline{A}}$ such that $B_i \subset A_i$, B_i's are pairwise disjoint and $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$.

   b) Prove that $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$. [12]

2. Let $\underline{\underline{A}}$ be a σ-field of subsets of a set X and let P be a probability on $\underline{\underline{A}}$. Explain what is meant by expectation of a positive measurable function. (Give the full development starting from the expectation of simple functions). [13]

Group B

Maximum Marks: 75

Answer all the questions.

3. Write down the four conditions that the function $F(x)$ should satisfy in order to be the distribution function of a probability distribution. Is this set of conditions

   i) necessary,
   ii) sufficient?

   Proofs are not needed. [4]

4. $F(x)$ is the distribution function of a probability distribution. We construct four new functions, as follows:

   i) $G_1(x) = \{F(x)\}^2$ for all x;
   ii) $G_2(x) = \sqrt{F(x)}$ for all x;
   iii) $G_3(x) = F(x+1)$ for all x;
   iv) $G_4(x) = \frac{1}{2}\{F(x+3) + F(x-2)\}$ for all x.

   Which of the four functions G_1, G_2, G_3 and G_4 are distribution functions (of probability distributions)? <u>Give reasons</u>. [28]

5.a) P is a probability distribution. E is the set of points α such that $P\{\alpha\} > 0$. Prove that E is finite or countably infinite. [10]

PLEASE TURN OVER

5.b) Construct a probability distribution Q which satisfies the following conditions:

i) Its spectrum is an infinite set, and at the same time.
ii) Its distribution function F(x) takes only rational values. [10]

c) The distribution function F(x) of a probability distribution Q takes only the values 0 and 1. What can you say about the nature of Q? Prove your result rigorously. [10]

6.a) When do we say that a set F (of real numbers) has Jordan content zero ? [5]

b) f(x) is a bounded function defined on the closed interval [a, b]. It is right-continuous at a and left-continuous at b. The set D of points of discontinuity of f(x) has Jordan content zero. Prove directly from the definition of the Riemann integral that f(x) is R-integrable on [a, b].

So you are not allowed to use any of the standard theorems on R-integrability. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV 1972-73 [416]

MID-YEAR EXAMINATION

Statistics-8 : Demography (Theory and Practical)

Date: 27.2.1973 Maximum Marks: 50 Time: 2 hours

Note: Answer question no.4 and any two from the rest. Marks allotted for each question are given in brackets [ ].

1. Discuss briefly any three of the following :-

   a) Errors in census age returns and how they arise.

   b) Underlying cause of death.

   c) Secondary attack rate.

   d) Broad items of information usually collected in a general health survey.

   [15]

2. What is the basic difference between a logistic law of population growth and an intrinsic rate of natural increase in a stable population.

   Derive Verhulst's logistic equation and indicate some of the important properties of the curve.

   [15]

3. a) Define a cumulative fertility function. What are the specific informations available from such a curve

   b) What are the methods available for estimating a current population

   Give a mathematical summary for a component method of population projection.

   c) If ${}_4f_1$ indicates the mean number of females 1-4 years old, ${}_4d_1$, the number of females deceased at those ages and l_1, the number of females surviving first year of age, then give the formula used for estimating the number of survivors at age 5.

   5 + 5 + 5 = [15]

4. The following table gives the number of population by age-groups and also the number of deaths in the same age-groups. In the same calander year, number of deaths under one year against the number of births 1758, as recorded in vital registration department was found to be 347.

   You are asked to compute the number of survivors l_x at different ages for x = 0, 1, 5, 10 and 20.

Age-groups	No. of population in the group	No. of deaths in the same group
1 - 4	3872	126
5 - 9	4571	69
10 - 19	7433	46
20 - 29	5894	63

[20]

[417]

INDIAN STATISTICAL INSTITUTE
Research and Training School
PERIODICAL EXAMINATION
B.Stat. Part IV: 1972-73

Statistics-4: Inference

Date: 7.5.73 Maximum Marks: 100 Time: 3 hours

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

1. $\{(x_i, y_i),\ i = 1, \ldots, n\}$ is a random sample of size n from a Bivariate normal population $N_2(m_1, m_2, \sigma_1, \sigma_2, \rho)$, $-\infty < m_1, m_2 < \infty$, $0 < \sigma_1, \sigma_2 < \infty$, $-1 < \rho < 1$. Obtain either the distribution of

$$t = \frac{(b - \beta)\sqrt{S_{11}}}{\sqrt{(S_{22} - S_{12}^2/S_{11})/n-2}}$$

(where the notations have usual significance) or the distribution of the sample correlation coefficient r (involving integral representation). [15]

2.a) $X_1, \ldots, X_{2m+1}$ are i.i.d. r.v.'s. each having the distribution $N(\theta, 1)$, $-\infty < \theta < \infty$. Show that the sample median $X_{(m+1)}$ is an UE of θ. Is it consistent? [4+1]=[5]

b) $X \sim B(n, \frac{1}{2})$. Z is a $(0, \pm 1)$ r.v. such that
$Pr(Z = 0 \mid X = r) = \alpha$, $Pr(Z = +1 \mid X = r) = (1 - \frac{r}{n})(1-\alpha)$,
$Pr(Z = -1 \mid X = r) = \frac{r}{n}(1 - \alpha)$, $0 < \alpha < 1$, $r = 0, 1, \ldots, n$.
Prove that $X' = X + Z \sim B(n, \frac{1}{2})$. [7]

c) $X_1, \ldots, X_n$ are i.i.d. r.v.'s, each having the distribution $N(\theta, 1)$, $\theta > 0$.
Show that $\sqrt{n}\left[\int_{\sqrt{n}\,\bar{x}}^{\infty} e^{-t^2/2}\, dt \Big/ e^{-\frac{n\bar{x}^2}{2}}\right]$ is an UE of $\frac{1}{\theta}$. [5]

d) A statistic T_n (based on n observations) is consistent for θ and var $(T_n) \to 0$ as $n \to \infty$. Show that $E(T_n) \to \theta$ as $n \to \infty$. [5]

e) X is a discrete r.v. having a uniform distribution over $1, \ldots, N$ i.e. $Pr(X = i \mid N) = 1/N$, $i = 1, \ldots, N$. The parameter set is $\Omega = \{N : N = 1, 2, \ldots\}$. $X_1, X_2, \ldots, X_n$ are n i.i.d. r.v.'s, each having the same distribution as X. Consider $X_{(n)} = \max_{1 \le i \le n} (X_i)$ as an estimator of N.
Show that $X_{(n)}$ is complete sufficient, consistent but biased. [2+4+2+2]=[10]

3.a) Define a sufficient statistic. State and prove the factorization theorem in the discrete case. [3+7]=[10]

b) What do you mean by a statistic being complete? Clearly explain, with an illustration, how the existence of a complete sufficient statistic helps in determining m.v.u. estimators of estimable parametric functions. [3+7]=[10]

p.t.o.

3.c) $X_1, \ldots, X_k$ are k i.i.d. $B(n, p)$ variables, $0 < p < 1$.
Obtain the m.v.u.e. of pq. Is it consistent? (Prove completeness of the relevant statistic.) [5+1]{

d) $X_1, \ldots X_n$ are n i.i.d. $N(0, \sigma^2)$ variables, $0 < \sigma < \infty$.
Obtain the m.v.u.e. of σ. Is it consistent? [

e) In case (d), obtain the m.v.u.e. of the proportion

$$P(\sigma) = \int_a^b \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\, dx$$ where a and b $(-\infty < a < b < \infty)$

are two fixed constants. [

4. In each of the following, give example of an estimate satisfying the stated requirements.

a) U but not C (b) C but not U (c) S but not C
d) S but not U (e) C but not S (f) U but neither C nor
g) S but neither U nor C (h) C but neither U nor S
i) none of the properties (U, S, C)
j) all the properties (U, S, C) [1

[U = unbiased, C = consistent, S = sufficient]

Clarity and neatness - 4.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1972-73

[418]

PERIODICAL EXAMINATION

Statistics-8: Complex Variables

Date: 14.5.73 Maximum Marks: 50 Time: $1\frac{1}{2}$ hours

Note: The paper carries 80 marks. Answer as many questions as you. Maximum you can score is 50. Marks allotted for each question are given in brackets [].

1. Suppose that a and b are two vertices of a square. Find the two other vertices in all possible cases. [10]

2. Show that an analytic function can not have a constant value without reducing to a constant. [10]

3. Show that $x^2 + iy$ is not an analytic function at any point in the complex plane. [10]

4. Prove that the functions $f(z)$ and $\overline{f(\bar{z})}$ are simultaneously analytic. [10]

5. What is the radius of convergence of the series $\Sigma e^n z^{n^2}$. [10]

6. If $f(z) = \Sigma a_n z^n$ what is $\Sigma n^3 a_n z^n$? [10]

7. What is Euler's formula? Derive it from the definitions of the sine and cosine functions.

   Prove that $(\cos \theta + i \sin \theta)^n = \cos n\theta + i \sin n\theta$. [10]

8. Do $\cos \theta$ and $\sin \theta$ have a period? If so, what is it? Prove your statement. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

[419]

PERIODICAL EXAMINATION

Statistics-6: Design of Experiments (Theory and Practical)

Date: 21.5.73 Maximum Marks: 100 Time: 3 hours

Note: Answer Q. 1 and Q.2 and any other two from the rest. Marks allotted for each question are given in brackets [ ].

1. A varietal trial with 13 varieties was carried out in a balanced incomplete block design. The plan and yields in pounds per plot are given below. Analyse the data and draw your conclusion. (Treatment numbers are written in parentheses (.) .) [30]

Block				
1	(3) 25.3	(6) 19.9	(9) 29.0	(11) 24.6
2	(3) 23.0	(4) 19.8	(8) 33.3	(12) 22.7
3	(10) 16.2	(11) 19.3	(12) 31.7	(13) 26.6
4	(2) 27.3	(5) 27.0	(8) 35.6	(11) 17.4
5	(7) 23.4	(8) 30.5	(9) 30.8	(10) 32.4
6	(4) 30.6	(5) 32.4	(6) 27.2	(10) 32.8
7	(1) 34.7	(5) 31.1	(9) 25.7	(12) 30.5
8	(3) 34.4	(5) 32.4	(7) 33.3	(13) 36.9
9	(1) 38.2	(2) 32.9	(3) 37.3	(10) 31.3
10	(2) 28.7	(4) 30.7	(9) 26.9	(13) 35.3
11	(1) 36.6	(4) 31.1	(7) 31.1	(11) 28.4
12	(1) 31.8	(6) 33.7	(8) 27.2	(13) 41.1
13	(2) 30.3	(6) 31.5	(7) 39.3	(12) 26.7

2. Develop in details the intrablock analysis of block designs. [30]

3. Define a latin square design and give its statistical analysis. [20]

4. Give the two definitions of the connectedness of a block design and establish their equivalence. [20]

5. State and prove the necessary and sufficient condition in terms of the latent roots of the C - matrix for a block design to be balanced. [20]

6. Describe a method of construction of balanced lattice. Define a B.I.B.D. and develop its intrablock analysis. [20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73 [420]

PERIODICAL EXAMINATION

Statistics-5: Statistical Methods (Theory and Practical)

Date: 28.5.73. Maximum Marks: 100 Time: 4 hours

Note: Answer Q. 5 and any three questions from Qs. 1 to 4. Marks allotted for each question are given in brackets [ ].

1. Regression equations of the form $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ are proposed to be set up on the basis of independently drawn samples from three different populations. How would you test whether the coefficients β_1, β_2 and β_3 are the same for all the populations? Obtain the procedure, stating necessary assumptions. [20]

2. Give an account of the S-method of multiple comparisons proving the following theorem:
If $\underset{\sim}{y} \sim N(X\beta, \sigma^2 I_n)$ with X nonstochastic and rank $(X)=r$, then the probability is $1-\alpha$ that simultaneously for all estimable linear parametric functions ψ in a q-dimensional space L of estimable functions

$$\hat{\psi} - S\hat{\sigma}_{\hat{\psi}} \le \psi \le \hat{\psi} + S\hat{\sigma}_{\hat{\psi}}$$

where $S = (q\, F_{q,n-r}(\alpha))^{1/2}$. [20]

3. Consider the observations $\{y_{ijk}\}$ in the random effects model for the complete two-way layout. Assume that $y_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk}$ where the $\{a_i\}, \{b_j\}, \{c_{ij}\}$ and $\{e_{ijk}\}$ are independently normal, with zero means and respective variances $\sigma^2_A, \sigma^2_B, \sigma^2_{AB}$ and σ^2_e. Now justify the procedures usually followed for testing (i) $H_A : \sigma^2_A = 0$, (ii) $H_B : \sigma^2_B = 0$ and (iii) $H_{AB} : \sigma^2_{AB} = 0$.

(The symbols have their usual significance) [20]

4. Write short notes on the following :

(i) Partial correlation - measurement and statistical significance ;

(ii) Forecasts from multiple regression equations and their standard errors. (10+10)=[20]

5. Either : In a study of 89 firms the regressand was total cost (x_1) and the regressors, rate of output (x_2) and rate of absenteeism (x_3). The means were : $\bar{x}_1=5.8$, $\bar{x}_2 = 2.9$, $\bar{x}_3 = 3.9$ and the matrix showing sums of squares and products adjusted for means is

Please turn over

	x_1	x_2	x_3
x_1	113.6	36.8	30.1
x_2		50.5	-66.2
x_3			967.1

Estimate the relationship between x_1 and the other two variables : $x_1 = \beta_o + \beta_2 x_2 + \beta_3 x_3$). Test the overall correlation of the equation and the partial effect of x_3, given x_2. Also set up the ANOVA table to show the effect of adding the regressor x_3 when x_2 is already present.

Obtain 99 percent confidence limits for x_1 regression coefficient β_2. [40]

5. OR The following data relate to an experiment on productivity with three machines and two methods of working :

method	machine no. 1	machine no. 2	machine no. 3
I	$n = 5$ $\Sigma x = 25.3$ $\Sigma x^2 = 145.0$	$n = 7$ $\Sigma x = 39.8$ $\Sigma x^2 = 262.3$	$n = 12$ $\Sigma x = 32.6$ $\Sigma x^2 = 105.6$
II	$n = 8$ $\Sigma x = 56.2$ $\Sigma x^2 = 435.5$	$n = 4$ $\Sigma x = 32.8$ $\Sigma x^2 = 280.4$	$n = 10$ $\Sigma x = 31.9$ $\Sigma x^2 = 120.3$

Here x denotes output and n the number of observations taken for each combination of machine and method of working. Analyse the data and comment on your findings. [40]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

[421]

PERIODICAL EXAMINATION

Statistics-8: Genetics

Date: 4.6.73. Maximum Marks: 75 Time: 3 hours

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

1. Explain the two most important differences from genetical view point between MITOTIC and MEIOTIC cell divisions. Explain what useful purpose is served by the REDUCTION DIVISION. Explain and derive Mendelian segregation ratios for two pairs of alleles segregating independently at two loci. Assume that one of the alleles at each locus is completely dominant over the other. [4+3+5]=[12]

2. Give your comments on Mendel's Laws of Inheritance in the light of present knowledge in genetics.

   Explain the following terms with illustrations (any six only): Alleles, Genotype, Phenotype, Heterozygote, Dominant gene, Linkage, Mutation and Selection. [6+6]=[12]

3. Give the genetics of M-N Blood groups system. Prove that the Gene-counting method and the Maximum Likelihood Method give identical gene frequency estimates for this system of blood groups. Derive variances of the gene frequency estimates. [3+8+2]=[13]

4. Two populations $P_1(0.3, 0.6, 0.1)$ and $P_2(0.1, 0.4, 0.5)$ mix up in equal numbers. Find the genotypic composition of the mixture. What will be the genotypic composition after one generation of random mating? Give the gene frequencies of the mixture before random mating and of the population resulting after one generation of random mating. Are they different? Give your comments. Examine if the two original populations P_1 and P_2 were in Hardy-Weinberg equilibrium. [2+1+2+2+2+4]=[13]

5. Give in brief Bridge's chromosome balance theory of Sex-determination in *Drosophila melanogaster*. What will be sex in each case of the chromosome combinations, (2A + 2X), (4A + 4X), (3A + 2X), (2A + 3X), (3A + XO), (3A + XY), (3X + XXY) and (2A + XY), and also give your remarks on their fertility.

   Both man and *Drosophila melan.* follow in a general way the XX - XY formula for normal sex determination, so far as the sex chromosomes are concerned. But what are the important points of difference between the two species?

   How are the embryonic developments of the gonads and the Wolffian and the Mullerian Ducts controlled by the XX and the XY chromosome combinations in man? [4+4+8+9]=[25]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73 [422]

PERIODICAL EXAMINATION

Statistics-7: Econometrics (Theory and Practical)

Date: 11.6.73 Maximum Marks: 100 Time: 3 hours

Note: Answer Q.5 and any three of the rest.
Marks allotted for each question are given in brackets [].

1.a) 'The production function is a technical engineering relation between inputs and output'. Is the purity of this relation always retained in the empirical estimation of the production function? Give reasons for your answer.

b) Discuss the important properties of the Cobb-Douglas production function. [10+10]=[20]

2.a) How does economic theory suggest the algebraic form of demand relationship to be derived from time series samples?

b) How do you propose to solve the problem of aggregation of individual demands from time series samples? [10+10]=[20]

3. Discuss the significance of the pooling of time series and cross-section samples in dealing with the problems of (i) multicollinearity, (ii) identification and (iii) least squares bias in demand analysis. [20]

4. Derive expressions for the Lorenz concentration coefficients for the Pareto and log-normal distributions. [20]

5.a) The following gives the average expenditure on cigarettes by families in four different levels of income:

Range of income (Rs.):	100-200	200-300	300-400	400-
Averages of income (Rs.):	140	245	347	502
Expenditure on cigarettes (Rs.):	2.2	5.8	8.3	12.3
Family size :	3.4	3.9	4.3	5.2

Find the Engel elasticity of demand of cigarettes, assuming constant elasticity. (You may give equal weights to the four income groups).

b) The following gives the labour input, the capital input and output in suitable units of each of seven firms in an industry:

Labout input (L):	5.2	6.8	9.2	3.4	12.6	15.8	5.7
Capital input (K):	12.7	15.3	20.7	10.5	23.6	32.9	10.8
Output (Q):	2.4	3.4	4.9	2.1	6.8	8.6	3.2

Estimate the parameters of the Cobb-Douglas production function

$$Q = AL^{\alpha} K^{\beta}.$$ [40]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1972-73

[423]

PERIODICAL EXAMINATION

Statistics-4: Probability

Date: 18.6.73. 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

Note: The paper carries 55 marks. Answer as many questions as you can. Maximum you can score is 50.

1.a) $F(x)$ is the distribution function of a probability distribution P on $(-\infty, \infty)$. Show that $\{F(x)\}^k$ is also a distribution function (of a probability distribution Q). k is a positive integer. [9]

b) Let P and Q be as above. Let $F(x) < 1$ for all x. Prove that
$$\lim_{x \to \infty} \frac{Q(x, \infty)}{P(x, \infty)} \text{ exists.}$$
What is the value of the limit? [9]

2.a) The distribution function $F(x)$ of a probability distribution P takes every value on the open interval $(0, 1)$. $F(x)$ may or may not take the values 0 and 1.
Prove that P is a continuous distribution. [9]

b) The distribution function $F(x)$ takes every value on the closed interval $[0, 1]$. What can we say about the nature of the distribution? [9]
The answer should consist of one sentence only.

3.a) If a and b are real numbers, how do you define $e^{(a+ib)}$? [4]

b) Prove that $e^{z_1+z_2} = e^{z_1} \cdot e^{z_2}$; z_1 and z_2 are arbitrary complex numbers. [6]

c) Prove that the characteristic function of a bounded distribution is uniformly continuous on $-\infty < t < \infty$.

Group B: Maximum Marks: 50

Note: Throughout $X_0, X_1, X_2, \ldots$ is a Markov Chain with a countable state-space $\mathcal{S}$ and having stationary transition probabilities.
Answer all the questions.

4.a) For $s, s' \in \mathcal{S}$, obtain the relation
$$p^{(k+\ell)}(s, s') = \sum_{r \in \mathcal{S}} p^{(k)}(s, r)\, p^{(\ell)}(r, s')$$

b) Define period of a state. Show that all states within the same class have a common period. [5+7]=[12]

p.t.o.

5.a) Define an essential state. If s is an essential state and s' is an inessential state show that s cannot lead to s'.

b) When is a state s and to be recurrent. Show that a recurrent state is necessarily essential.
[You may assume the following result:

$$g(s, s') = \begin{cases} f^*(s, s') & \text{if } s' \text{ is recurrent]} \\ 0 & \text{otherwise} \end{cases}$$

[5+7]=[12]

6. Prove that if $g(s, s') = 1$ and $f^*(s, s') > 0$, then $g(s, s') = 1$. Hence show that a recurrent state can lead only to recurrent states. [12]

7. In an instance $\mathcal{S}$ consists of 5 elements $s_1, \ldots, s_5$ and the transition probability matrix is

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

[The (k, ℓ)th entry in the matrix P is the probability $p(s_k, s_\ell)$.]

(a) Decompose $\mathcal{S}$ into classes.

(b) Pick out the essential classes, if any.

(c) Decide which classes are recurrent. [4+4+6]=[14]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. (Hons.) Part IV: 1972-73 [424]

ANNUAL EXAMINATION

Statistics-4: Inference

Date: 9.7.73 Maximum Marks: 100 Time: 3 hours

Note: The paper carries 110 marks. Answer as much as you can. Maximum you can score is 100. Marks allotted for each question are given in brackets [ ].

1.a) State and prove the Cramer-Rao inequality. [2+10] =[12]

b) $X \sim P(\lambda)$. Prove that for estimating λ^2 there does not exist any unbiased estimator whose variance attains the Cramer-Rao lower bound but that there exists one whose variance attains a Bhattacharya bound. Hence obtain the m.v.u.e. of λ^2. [8]

2.a) If T is the m.v.u.e. of θ, prove that any polynomial g(T) in T is the m.v.u.e. of $E_\theta\{g(T)\}$. [10]

b) $X^{n\times 1} \sim N(A'\beta, \sigma^2 I_n)$. Consider an estimable linear parametric function $\lambda'\beta$ and let $\xi'X$ be the BLUE of it. Prove that $\xi'X$ is the m.v.u.e. of $\lambda'\beta$. [5]

c) Prove directly that a complete sufficient statistic has zero covariance with every unbiased estimator of zero. [5]

3. Consider a probability density function $f_\theta(x)$ depending on k parameters $\underline{\theta} = (\theta_1, \ldots, \theta_k) \in \Omega_k \subset R^k$. Let $X_1, \ldots, X_n$ be independent replicates of X of which the density is $f_\theta(x)$.

a) Consider the problem of estimating the parametric vector $g(\underline{\theta}) = (g_1(\underline{\theta}), \ldots, g_\ell(\underline{\theta}))$, $\ell \leq K$, $g_i(\theta) : R^k \to R'$, $i = 1, \ldots, \ell$. Define a regular unbiased estimator (rue) of $g(\underline{\theta})$. Prove that if a complete sufficient statistic $\underline{T} = (T_1, \ldots, T_k)$ exists for $\underline{\theta}$ then there always exists a unique rue for every estimable parametric vector $g(\underline{\theta})$. [4+8]=[12]

b) Prove that if $\underline{T}$ is the most efficient estimator of $\underline{\theta}$ (in the sense of being regular unbiased), then T_i is the m.v.u.e. of θ_i, $i = 1, \ldots, k$. [8]

c) $X \sim N(m, \sigma^2)$ and $X_1, \ldots, X_n$ are n independent replicates of X. Obtain rue estimator of (m^2, σ^2). [5]

4.a) Prove the consistency of a maximum likelihood estimator. [20]

b) $y_1, \ldots, y_n$ are independent normal variates with $E(y_r) = r\theta$, $var(y_r) = r^3\sigma^2$, $r = 1, \ldots, n$. Show that the ml estimator of θ is exactly normally distributed and is unbiased. [10]

c) Consider the discrete probability distribution

$$Pr(N = n/a, \pi) = \binom{n-1}{a-1}\pi^a(1-\pi)^{n-a}, \quad n = a, a+1, a+2, \ldots$$

Assuming a known, show that the mle of π is biased and that its asymptotic variance is $\frac{\pi^2(1-\pi)}{a}$. [10]

p.t.o.

4.d) $Y = \log X \sim N(\mu, \sigma^2)$. Show that the mle of

$E(X) = \exp\{\mu + \frac{1}{2}\sigma^2\}$ is $\exp\{\bar{y} + \frac{1}{2}S_y^2\}$ where $\bar{y}$ and S_y^2

are the sample mean variance of y on the basis of n independent observations $y_1, \ldots, y_n$. Prove that this estimator is biased upwards but is asymptotically unbiased. [10]

e) $X_1, \ldots, X_n$ are n independent observations on X having the density $f(x, \theta) = \frac{1}{\theta}$, $0 \leq x \leq \theta$, $\theta > 0$. Prove that the mle of θ is biased. [5]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. (Hons.) Part IV: 1972-73

[225]

ANNUAL EXAMINATION

Statistics-8: Complex Variables

Date: 10.7.73 Maximum Marks: 100 Time: 3 hours

Note: The paper carries 110 marks. Answer as much as you can. Maximum marks you can score is 100. Marks allotted for each question are given in brackets [ ].

1. What do you understand by a line integral? Show that the line integral $\int_\gamma p\,dx + q\,dy$ defined in a region Ω depends only on the end points of γ if and only if there is a function $U(x, y)$ in Ω so that

$$\frac{\partial U}{\partial x} = p, \quad \text{and} \quad \frac{\partial U}{\partial y} = q.$$ [20]

2. State and prove Cauchy's theorem for a rectangle. [25]

3. Prove that if the piecewise differentiable closed curve γ does not pass through the point a, then

$$\frac{1}{2\pi i} \int_\gamma \frac{dz}{z-a}$$

is an integer. Hence define winding numbers. [15]

4. State and prove Morera's theorem. [15]

5. Show that a polynomial with complex coefficients has a root in $\mathbb{C}$, the complex plane. [10]

6. Define zeros and poles of analytic functions. Determine the zeros and poles of the analytic function

$$\frac{1}{z^2 - 1}.$$ [10]

7. Compute the integrals $\int_\gamma x\,dz$ where γ is the directed line segment from 0 to 1+i and the integral

$$\int_{|z|=2} \frac{dz}{z^2 - 1}.$$ [15]

INDIAN STATISTICAL INSTITUTE [426]
Research and Training School
B.Stat.(Hons.) Part IV: 1972-73

ANNUAL EXAMINATION

Statistics-4: Probability

Date: 12.7.73 Maximum Marks: 100 Time: 3 hours

Note: The paper carries 110 marks. Answer as much as you can. Maximum one can score is 100. Marks allotted for each question are given in brackets [ ].

1.a) Write down the four necessary and sufficient conditions that a function $F(x)$ should satisfy in order that it may be the distribution function of a probability distribution. Proofs are not needed. [7]

b) Q is a nonnegative finitely additive interval function such that $Q(-\infty, \infty) = 1$. Show that its distribution function will satisfy one of the four conditions mentioned above. [7]

c) Give an example of a nonnegative finitely additive interval function P whose distribution function fails to satisfy each of the remaining three conditions. [10]

2.a) Prove that the spectrum of every probability distribution is nonempty. [10]

b) The probability distributions P, Q and D have distribution functions $F(x)$, $G(x)$ and $F(x)$. $G(x)$ respectively. Show that the spectrum of D is contained in Spec (P) U Spec (Q). [10]

c) In connection with the above question, give an example where spec (D) is a proper subset of spec (P) U Spec (Q). [10]

3.a) What is meant by the oscillation of a bounded function at a point? [6]

b) $f(x)$ is a bounded function and is discontinuous at $x = a$. P is a probability distribution such that $P\{a\} > 0$. Prove that the Riemann-Stieltjes integral

$$\int_{-\infty}^{\infty} f(x)dP \text{ does not exist.}$$ [10]

4.a) Prove that if the characteristic function of a probability distribution is differentiable at $t = 0$, then $\phi'(0)$ is 0 or a purely imaginary number. [10]

b) $\phi(t)$ and $\psi(t)$ are real-valued characteristic functions. $\phi'(0)$ does not exist. Prove that $\phi(t).\psi(t)$ is not differentiable at $t = 0$. [10]

c) P is a probability distribution having a first moment. Prove that $\phi'(t)$, the derivative of the characteristic function, is uniformly continuous on $-\infty < t < \infty$. [10]

5. $F_n(x)$ is the distribution function of P_n, $n = 1,2,3,$ and $F(x)$ is that of P. $F_n(x) \to F(x)$ on a dense set D of real numbers. Show that $P_n \to P$ weakly. [10]

INDIAN STATISTICAL INSTITUTE [427]
Research and Training School
B.Stat. (Hons.) Part IV: 1972-73

ANNUAL EXAMINATION

Statistics-6: Design of Experiments
(Theory and Practical)

Date: 13.7.73 Maximum Marks: 100 Time: 4 hours

Note: Answer Q.1 and any three from the rest. Marks allotted for each question are given in brackets [ ].

1. A 5 × 5 latin square experiment was carried out to compare the effects of five different manurial treatments A, B, C, D, E on the yield of wheat. The crop on one of the plots was lost accidentally; the plan and the yields (in certain units) of the other plots are given below. Analyse the data and draw conclusions.

Yields in manurial experiment with one plot missing

B .	D 6.4	E 3.3	A 9.5	C 11.8
C 9.3	A 4.0	B 6.2	E 5.1	D 8.4
D 7.6	C 15.4	A 6.5	B 6.0	E 4.6
E 5.3	B 7.6	C 13.2	D 8.6	A 4.9
A 9.3	E 6.3	D 11.8	C 15.9	B 7.6

[25]

2. Show that a connected block design is balanced if and only if its C-matrix is of the form $\alpha I_v + \beta E_{v,v}$ ($\alpha, \beta > 0$) where I_v is unit matrix of order v and $E_{v,v}$ is $v \times v$ all-one matrix. Hence or otherwise prove that a proper, equi-replicate, binary block design is balanced if and only if every pair of treatments occurs in a constant number of blocks of the design. [25]

3.a) Prove that for a resolvable BIBD (v, b, r, k, λ), we have $b \geq v+r-1$. Also show that a resolvable BIBD is affine resolvable if and only if $b = v+r-1$. Give an example of an affine resolvable BIBD.

b) Describe a method of construction of a series of affine resolvable BIBD. [4+10+4+7]=[25]

4. A randomised block design was conducted for v treatments in v blocks. Due to some accidents, after the completion of the experiment it was found that v observations could not be obtained for analysis. They were found to correspond to treatment i in block i, $i = 1,2,\ldots, v$. Develop, from first principles, suitable statistical analysis for such incomplete data. You are to consider the aspects of testing and estimating the treatment effects and connectedness and balancing of the resulting design. [25]

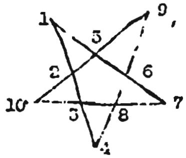
p.t.o.

5.a) If x is a primitive root of $GF(v)$ where $v = 4t+3$ $(t > 0)$ is a prime or prime power, then (i) show that the initial block

$$(x^0, x^2, x^4, \ldots, x^{4t})$$

has symmetrically repeated differences, and hence (ii) it gives a BIBD. Obtain the parameters of the resulting BIBD.

b) Consider the following configuration:



Show that when lines are treated as blocks and points (numbered 1 to 10) as treatments, the above configuration yields a PBIBD(2). Obtain the parameters of the PBIBD(2).

[15+10]=[25]

---- ---

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat.(Hons.) Part IV: 1972-73 [428]

ANNUAL EXAMINATION

Statistics-5: Statistical Methods Theory

Date: 16.7.73 Maximum Marks: 100 Time: 3 hours

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

1. EITHER

Obtain the expression for the asymptotic covariance between two central moments computed from a random sample. Show, in particular, that cov $(\bar{x}, s^2) = \mu_3/n$. (Prove the necessary formula for the asymptotic variance of a function of several statistics.)

OR

State and prove the theorem on the distribution of the goodness of fit chi-square when the null hypothesis is composite and the parameters have been estimated from the sample.

2. Regressions equations of the form $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ are being estimated on the basis of p independently obtained sets of observations. Obtain the procedures for testing whether the true regression equations are (i) parallel, (ii) identical, for the different sets of observations. State your assumptions carefully.

3. Define the multiple correlation coefficient.
Suppose we consider families each having (n+1) brothers. Show that if r be the correlation coefficient between the heights of any pair of brothers and R_n the multiple correlation coefficient of the height of one brother on those of n other brothers, then

$$R_n = r\sqrt{\frac{n}{1+(n-1)r}}$$

If the joint distribution of $x_1, \ldots, x_p$ is multinormal, what is the conditional distribution of x_1, given $x_2, \ldots, x_p$? How is the conditional variance related to the multiple correlation coefficient $R_{1.23\ldots p}$?

4. Prove that if $S \sim W_p(n, \Sigma)$ and Σ^{-1} exists, then $\frac{L'\Sigma^{-1}L}{L'S^{-1}L} \sim \chi^2(n-p+1)$ for every fixed vector L. Hence find the distribution of the generalized T^2-statistic and show how it can help in inferences about the mean vector of a multinormal population.

6. a) Define the multivariate normal distribution.

b) Obtain the maximum likelihood estimators of the mean vector μ and the dispersion matrix Σ.

c) Show that the estimator of μ is statistically independent of the estimator of Σ. (You need not prove every result that is necessary.)

p.t.o.

7. Answer any two of the following:

i) What is the S-method of multiple comparisons? What is its advantage over the usual method of pair-wise comparisons? Give the underlying theory of the S-method.

ii) Mention some common test-procedures, including the Scheffè procedure, for the Fisher-Behrens problem.

iii) n variates are such that each pair has correlation ρ. Show that the k-th partial correlation between any pair, i.e. the correlation when k other variates are held constant, ρ_k, obeys the difference relation

$$\rho_k - \rho_{k-1} = - \rho_k \rho_{k-1}$$

and hence that

$$\rho_k = \frac{\rho_o}{1 + k\rho_o} .$$

iv) Let Δ_p^2 and Δ_q^2 denote the Mahalanobis distances between two populations based on p characters and a subset of q characters respectively. Show that $\Delta_p^2 \geq \Delta_q^2$ in general.

INDIAN STATISTICAL INSTITUTE [429]
Research and Training School
B. Stat.(Hons.) Part IV: 1972-73

ANNUAL EXAMINATION

Statistics-5: Statistical Methods (Practical)

Date: 17.7.73 Maximum Marks: 100 Time: 3 hours

Note: Answer any two question from Q. Nos. 1 to 3 and all the rest. Marks allotted for each question are given in brackets [ ].

1. The following are the sample means and the pooled estimate of the dispersion matrix in a study based on 50 observations on each of two species:

Character	Means		Dispersion matrix		
	Iris Versicolar	Iris Setosa	1	2	3
Sepal length	5.94	5.01	0.1953	0.0922	0.0996
Sepal width	2.77	3.43		0.1211	0.0472
Petal length	4.26	1.46			0.1255

Test the significance of the difference between the true mean vectors. [30]

2. The following is the matrix of intercorrelations, based on a sample of 140 students, between arithmetic speed (x_1), arithmetic power (x_2), memory for words (x_3) and memory for meaningful symbols (x_4):

$$\begin{pmatrix} 1.00 & 0.42 & 0.04 & 0.02 \\ & 1.00 & 0.15 & 0.25 \\ & & 1.00 & 0.67 \\ & & & 1.00 \end{pmatrix}$$

i) Test whether the matrix given is internally consistent.

ii) Find the multiple correlation coefficient $R_{2.134}$ and test its statistical significance. [30]

3. The coded data below are the flow rates of a fuel through three types of nozzle as measured by four different operators, each of whom made three determinations on each nozzle.

Nozzle	Operator 1	Operator 2	Operator 3	Operator 4
A	6, 6, -15	26, 12, 5	11, 4, 4	21, 14, 7
B	13, 6, 13	4, 4, 11	17, 10, 17	-5, 2, -5
C	10, 10, -11	-35, 0, -14	11, -10, -17	12, -2, -16

Analyze the data according to the mixed model, assuming that the operators were drawn at random from a larger population. [30]

p.t.o.

4. Suppose a sample of size 20 yields a correlation coefficient $r = 0.65$. Set up 95 % confidence limits for the true correlation coefficient. [15]

5. A sample of 47 female cats gave the following results:

$\Sigma x = 110.9$, $\Sigma y = 432.5$, $\Sigma x^2 = 265.13$, $\Sigma y^2 = 4064.71$

and $\Sigma xy = 1029.62$.

Here x is body weight in kg. and y the heart weight in grams.

Test whether the regression line of y on x passes through the origin. [15]

6. Practical Record [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat.(Hons.) Part IV: 1972-73 [430]

ANNUAL EXAMINATION

Statistics-7: Econometrics (Theory and Practical)

Date: 18.7.73 Maximum Marks: 100 Time: 4 hours

Note: Answer Q. 6 and any three from the rest. Marks allotted for each question are given in brackets [ ].

1. Assuming the disturbances to follow a first order autoregressive scheme, examine the problem of linear estimation of parameters. Suggest some modifications in the procedure that may make the application of ordinary least squares a good approximation. How will sampling variances of the parameters be affected if you make a direct application of ordinary least squares? [18]

2. Give an account of the Pareto law of income distribution and discuss the arguments for and against its use. Derive an expression for the Lorenz ratio when income distribution follows this law. [18]

3. Discuss how the theoretical properties of Cobb-Douglas production function accord with Douglas's findings from empirical research. What are the main points of criticism raised against Douglas's work? [18]

4. Explain the economic situations that require the application of generalised least squares? Show that it results in the best linear unbiased estimates in such cases. Discuss how heteroscedastic disturbances can be tackled and indicate some method of detecting their presence. [18]

5. What should be the dependent variable and the determining variable of the Engel curve? Give reasons for your answer. Examine the criteria for an algebraic formulation of the Engel curve. [18]

6. Distribution of assessed income in India is given below for the years 1951 and 1960. Examine if the inequality of distribution of income has increased.

Annual income	1951 no. of assesses (thousand)	1951 assessed income (Rs. crores)	1960 no. of assesses (thousand)	1960 assessed income (Rs. crores)
below 10000	371.2	174.2	643.0	358.7
10000 - 20000	61.8	84.1	155.2	212.7
20000 - 70000	32.7	108.2	79.1	270.9
70000 -100000	2.5	20.5	6.5	53.9
100000 -200000	2.4	33.3	4.6	61.3
over 200000	1.8	154.4	2.6	234.5

[36]

Practical Records. [10]
