INDIAN STATISTICAL INSTITUTE

QUESTION PAPERS

for

STATISTICIAN'S DIPLOMA EXAMINATION

May 1976

Price Rupees Two only

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1976

Paper I (Theoretical): Official Statistics and Descriptive Statistics

Time: 4 hours — Full marks: 100

(a) Figures in the indicate full marks.

(b) Use of calculating machines is not permitted.

GROUP A: Official Statistics (50 marks)

(Answer Question 1 and any two other questions from this group)

1. Give a brief account of the history and activities of the National Sample Survey Organisation, mentioning particularly the following features:

(i) Wings and Divisions; (ii) Method of data collection; (iii) Sampling design; (iv) Data processing system. (12+8)=20

2. Describe the sources of available housing statistics in India. Comment on the limitations of these statistics. Name also the source of the data regarding construction of residential quarters by employers for their workers in the industrial sector. (8+4+2)=14

3. Write short-notes on any two of the following index numbers relating to India, mentioning the compiling agencies, items covered, the base periods, weighting system and method of averaging.

i) Index number of Industrial Production

ii) Index number of Earnings for Mining Workers

iii) Index number of Mineral Production

iv) Index number of Foreign Trade. (7+7)=14

4. Describe briefly the work of any two of the following organisations in the field of statistics:-

i) FAO or UNESCO

ii) Office of the Registrar General of India

iii) The Reserve Bank of India (7+7)=14

5. What are the sources of statistical information on "labour and employment" in India. Write a critical note on the contents of these publications. (7+7)=14

GROUP B: Descriptive Statistics (50 marks)

(Answer Question 6 and any two of the rest from this group)

6. EITHER

a) What is a statistical population? Explain by illustration.

b) Explain the meaning of parameter, statistic and sampling error.

c) Briefly explain the nature of statistical conclusion. (6+6+4)=16

OR

Briefly describe the stages of operation and the nature of organisation required for completion of a survey project after field data have been collected. (16)

Please turn over

7. (a) Define a concentration curve and explain its use.

(b) Obtain the coefficient of concentration for a variable x such that the density function of x is

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\,\frac{1}{x}\,e^{-\frac{1}{2\sigma^2}(\log x - \mu)^2} \quad \text{and } x > 0$$

(6+10)=16

8. (a) Explain the use of mathematical models of probability distribution.

(b) How many modes can a Binomial distribution have? Find out the unique or the multiple modes, as the case may be, of a Binomial distribution.

(c) Prove the relation

$$\mu_{r+1} = p(1-p)\left(\frac{d\mu_r}{dp} + n\,r\,\mu_{r-1}\right) \text{ for a Binomial}$$

distribution where μ_r is the rth central moment and p is the probability of success. (5+5+6)=16

9. (a) Prove that $\beta_2 \geq \beta_1 + 1$ where β_1 and β_2 have usual meaning.

(b) If ρ is correlation coefficient between x and y, then what can be said about the relation between x and y in the general case as well as in the case when x and y follow a bivariate normal distribution, when $\rho = 0$. Give reasons. (8+8)=16

NEATNESS (Groups A and B) (4)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – May 1976

Paper II (Theoretical): Probability Theory and Statistical Methods

Time: 4 hours | Full marks: 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is not permitted.

GROUP A: Probability Theory (50 marks)

(Answer any three questions from this group)

1. (a) Let $(\Omega, \mathcal{A}, P)$ be a probability space and let B be an event lying in $\mathcal{A}$ and that $P(B) > 0$. Show that the function $P(\cdot \mid B)$ defined by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

for $A \in \mathcal{A}$ is, like P, a probability function on $\mathcal{A}$. Explain the concept of statistical independence of a number of events in $\mathcal{A}$, clearly distinguishing between complete independence and pairwise independence.

(b) A box contains 8 pairs of shoes. If 4 shoes are chosen at random, what is the chance that there will be no complete pair among them? (10+6)=16

2. The density function $f(x)$ of a random variable X is given by

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ K \exp\left[-x/3\right] & \text{if } x > 0 \end{cases}$$

(a) Obtain the value of the constant K.

(b) What is the distribution function of the variable X?

(c) Obtain the density function of the random variable

$$Y = 1-\exp\left[-X/3\right].$$

(d) Calculate E(X) and E(Y) (3+3+6+4)=16

3. Let X_1, X_2 be independent random variables having the same discrete distribution given by the following probability mass function :

$$P\left[X_1 = k\right] = P\left[X_2 = k\right] = (1-p)^k p, \quad k = 0, 1, 2, \ldots$$

where $0 \leq p \leq 1$.

(a) Obtain the distribution of $X_1 + X_2$

(b) Obtain the conditional distribution of X_1, given $X_1 + X_2 = x$ (where $x = 0, 1, 2, \ldots$).

(c) Find the mean and variance of $X_1 + X_2$.

4. (a) When do you say that the sequence $\{X_n\}$ of random variables converges in probability to a constant α ?

(b) Prove Chebycheff's inequality:

$$P\left\{ |X - E(X)| \geq \epsilon \right\} \leq \frac{Var(X)}{\epsilon^2}$$

(c) If $\{X_n\}$ is a sequence of independent and identically distributed random variables such that $P\{X_n = 0\} = (1-p)$ and $P\{X_n = 1\} = p$ where $0 \leq p \leq 1$, prove that the sequence of random variables $\{Y_n\}$ where $Y_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$, $n \geq 1$ converges in probability to p. (3+5+8)=16

5. The random variables X_1, X_2 follows a bivariate normal distribution with $E(X_1) = E(X_2) = 0$, $V(X_1) = V(X_2) = \sigma^2$, Cor $(X_1, X_2) = \rho\sigma^2$

(a) Write down the joint density of (X_1, X_2).

(b) Write down the joint density of $(X_1 + X_2, X_1 - X_2)$.

(c) Prove that the random variable $\dfrac{X_1 + X_2}{|X_1 - X_2|}\sqrt{\dfrac{1-\rho}{1+\rho}}$ follows a student's t-distribution with 1 degree of freedom. (2+6+8)=16

GROUP B: Statistical Methods (50 marks)

(Answer any three questions from this group)

6. (a) Describe briefly the methods of moments and the method of maximum likelihood for point estimation.

(b) Suppose $X_1, X_2, \ldots, X_n$ are independent and distributed uniformly on the interval $[0, \theta]$, $0 < \theta < \infty$. Estimate θ by each of the two methods mentioned in (a). (4+4+8)=16

7. (a) Define unbiasedness and consistency as criteria of a good estimator. Give a set of sufficient conditions for consistency.

(b) Let m_r be the rth sample central moment for random samples of size n from a population whose moments upto that of order 2r exist. Show that m_r is consistent for μ_r, the population central moment of order r. (4+4+8)=16

8. Describe One-way classified data. Explain in details how you will analyse such data both for the fixed-effects and random-effects models. (4+12)=16

9. (a) For testing a simple hypothesis against a simple alternative, describe the sequential Probability Ratio Test (SPRT) of strength (α, β). How does one determine the boundary points associated with the SPRT?

(b) Let X be a random variable distributed as follows: $P\{X = 1\} = p$, and $P\{X = 0\} = 1-p$, where $0 < p < 1$. For testing H_0: $p = 0.5$ against H_1: $p = 0.8$ show that the usual sequential probability ratio test can be written in the following form: At the mth stage let m_1 be the number of successes. Then

i) accept H_0 if $m_1 \leq Cm + d_1$,

ii) reject H_0 if $m_1 \geq Cm + d_2$,

iii) continue sampling otherwise,

where c, d_1, d_2 are constants depending on the strength of the test. (5+3+8)=16

10. Write notes on any two of the following:

(a) Cramer-Rao Inequality and its utility.

(b) Likelihood-ratio tests.

(c) Relationship between confidence interval and testing of hypothesis. (8+8)=16

NEATNESS (Groups A and B) (4)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1976

Paper III (Theoretical): Sample Surveys and Design & Analysis of Experiments

Time : 4 hours | Full marks: 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is not permitted.

GROUP A: Sample Surveys (50 marks)

(Answer any three questions from this group)

1. Distinguish between sampling error and non-sampling error in surveys. Describe methods used by statisticians for controlling these errors. (16)

2. What is the ratio method of estimation? Obtain approximations to the mean square error and bias of the ratio-estimator under simple random sampling without replacement. (16)

3. Describe systematic sampling procedure and discuss its advantages over simple random sampling procedure.

From a population of size N = nk, a systematic sample of size n is drawn by random start. Suggest an unbiased estimator of the population total and obtain its sampling variance. Under what condition is systematic sampling more precise than simple random sampling without replacement if one is interested in estimating the population total? (16)

4. Let y_{1j} and y_{2j} respectively denote the values of a character Y for the jth unit (j = 1, 2, ... N) on the first and the second occasions. Consider the following sampling scheme over two occasions.

On the first occasion a simple random sample S of size n is selected without replacement; on the second occasion a sub-sample S_1 of m units is selected without replacement from S and an independent simple random sample S_2 of u units is selected without replacement from the whole population.

Show that the estimator

$$\hat{\bar{Y}}_2 = \alpha\, \hat{\bar{Y}}_{2u} + (1-\alpha)\, \hat{\bar{Y}}_{2m}$$

where $\hat{\bar{Y}}_{2u} = \frac{1}{u} \sum_{S_2} y_{2j}$

$$\hat{\bar{Y}}_{2m} = \frac{1}{m} \sum_{S_1} (y_{2j} - y_{1j}) + \frac{1}{n} \sum_{S} y_{1j}$$

and α is any constant, is an unbiased estimator of the population mean of Y on the second occasion. Find the variance of $\hat{\bar{Y}}_2$. Determine α for which the variance of $\bar{Y}_2$ is minimum. (4+6+6)=16

Please turn over

5. Write short notes on any two of the following:

(a) Uses of random Sampling Numbers.

(b) Cluster Sampling

(c) Difference Estimator (8+8)=16

GROUP B: Design & Analysis of Experiments (50 marks)

(Answer any three questions from this group)

6. (a) Why do you need a design at all? Discuss the requirements of a good experiment.

(b) Set out the important assumptions underlying the analysis and interpretation of data obtained from the standard experimental designs. (3+6+7)=16

7. (a) What is experimental error? Discuss with illustrations the various measures for controlling the experimental error.

(b) How the optimum size and shape of plots can be determined through uniformity trial data? (3+8+5)=16

8. Define a balanced incomplete block design with parameters v, b, r, k and λ, and prove: (i) $\lambda(v-1) = r(k-1)$, and (ii) $b \geqslant v$. Work out the efficiency of the design relative to a randomised block design. (2+2+2+10)=16

9. Two factors A and B each appear at m levels. Explain what do you mean by the interaction of these two factors A and B. Explain the principle of generalised interaction for a 2^m factorial experiment.

Explain the method of obtaining the sum of squares due to partially confounded treatment contrast.

How do you calculate the loss of information on a partially confounded treatment contrast? (2+4+5+5)=16

10. Write short notes on any two of the following :

(a) Youden square designs

(b) Missing plot technique

(c) Simple Lattice designs

(d) Fractional replications. (8+8)=16

NEATNESS (Groups A and B) (4)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – May 1976

Paper IV (Theoretical): Applied Statistics Group Papers

Time: 4 hours
(for two groups)

Full marks: 100

(a) Candidates will be required to answer questions from those two groups of subjects only for which they have already registered their options.

(b) Separate answer-books are to be used for each of the two groups attempted.

(c) Figures in the margin indicate full marks.

(d) Use of calculating machines is not permitted.

GROUP (a): ECONOMIC STATISTICS (Half-paper – 50 marks)

(Answer any three questions from this group)

1. (a) Explain the problem of aggregation in measurements of macroeconomic variables.

   (b) Define a 'true' cost-of-living index. Do you think the commonly used index numbers estimate this true index correctly? (8+8)=16

2. (a) What are the main components of an economic time series? Discuss with examples.

   (b) Suggest a method of estimating seasonal patterns, given monthly prices of fish in a given market for 20 years. You may assume that the seasonal pattern is constant over the period. (8+8)=16

3. Describe a method of estimating the income elasticities of demand from family budget data. What would be the main objections if these elasticities are used for demand projections? (12+4)=16

4. (a) Explain the concept of productivity in the theory of the firm. Distinguish between marginal productivity and returns to scale.

   (b) Describe a method of estimating the Cobb-Douglas production function from cross-section data on firms within an industry. Comment on the nature of the estimates you would obtain for the marginal productivities of labour and capital. (8+8)=16

5. (a) Explain carefully the distinction between "final demand" and "derived demand". Why is it necessary to make this distinction in economics?

   (b) Describe some practical uses of input-output analysis. (8+8)=16

6. (a) Discuss the main sources of data on the size-distribution of income in India.

   (b) How would you use these data for studying intertemporal shifts in the income distribution? What are the main limitations of such studies? (8+8)=16

NEATNESS (2)

GROUP (b): STATISTICAL QUALITY CONTROL (Half-paper - 50 marks)

(Answer <u>any three</u> questions from this group)

1. (a) Explain the role of rational subgrouping in operating a control chart and give justification for taking small samples of size 4 or 5 in case of $\bar{X}$-chart.

(b) Interpret the following situations observed on an $\bar{X}$-R chart:

i) R chart shows state of control but points on $\bar{X}$-chart are found systematically shifting upwards.

ii) Points on both $\bar{X}$ and R charts are too close to the control line

iii) Three-points on R-chart fall beyond upper control limit whereas all the points on $\bar{X}$-chart are within control limits. (7+9)=16

2. (a) Define 'Tolerance limits' and 'Confidence limits'. Describe a method for obtaining two sided tolerance limits for a controlled process producing an item with a measurable quality characteristic. The mean value is known and it can be assumed that the characteristic is normally distributed and other necessary stipulations (tolerance coefficient and confidence coefficient) are given.

(b) Define the operating characteristic function of $\bar{X}$-chart. Explain how you would proceed to determine the OC curve for such a chart if the sample size n is given and the standard deviation is known and remains constant. (10+6)=16

3. (a) Explain the terms (i) AOI (ii) AOQ (iii) ASN. Derive the mathematical expression for (i) the OC curve of a double and (ii) AOQL of a single sampling plan by attributes.

(b) When using Military Standard 105D what are the criteria to be adopted to decide whether to reduce or tighten inspection or to continue normal inspection? (10+6)=16

4. A production unit desiring to develop a new product, is interested in conducting an experiment to determine which method of production would require the minimum processing time.

Six machines and six operators are available for this experimentation. There are also six alternative methods of production.

Suggest a suitable design for the experiment.

With suitable notations, give the estimates of the components of variance. (16)

NEATNESS (2)

GROUP (c) : STATISTICAL METHODS IN GENETICS (Half-paper -50 marks)

(Answer any three questions from this group)

1. In a breeding experiment with F_2 data, derive the amount of information on recombination fraction given by the mating of coupling heterozygotes when there is complete dominance at one locus and incomplete dominance at the other. (Assume that the recombination fraction is the same in both male and female gametogenesis.) (16)

2. For the single back-cross where the two factors segregate in the ratios of 1 : 1 and 3 : 1, work out a suitable test for detecting linkage. Estimate the linkage, if present, by a suitable method and calculate its standard error.

   If the gene-ratios are disturbed for both the factors, what would be the efficient method for estimating linkage. (16)

3. Starting with the population : (p^2 AA + 2pq Aa + q^2 aa) derive the following relationships in case of inbred progenies:-

   (a) Covariance of full-sibs,

   (b) Covariance of parent and offspring,

   (c) Variance of offspring,

   if the genotypic values for AA, Aa and aa are 1,1 and 0 respectively. (16)

4. Discuss 'Hardy-Weinberg' law of equilibrium under random mating.

   In the case of complete positive assortive mating where dominants mate with dominants and recessives with recessives only, and starting from an initial panmictic population (p^2, 2pq, q^2), show that the amount of heterozygosity in the nth generation is given by

   $$y_n = Y_a/(1+nq)$$

   where, y_n = the heterozygosity in the nth generation and y_0 = the same in the intial generation. (16)

5. Write short notes on any two of the following :

   (a) Estimation of the gene-frequencies in O-A-B blood group system.

   (b) Linear discrimination function and gain over straight selection.

   (c) Fisher's Fundamental Theorem of Natural Selection.

   (d) Equilibrium under mutation and selection. (16)

Please turn over

GROUP (d): VITAL STATISTICS AND DEMOGRAPHY (Half-paper – 50 marks)

(Answer any three questions from this group)

1. (a) Explain the difference between a rate and a ratio in Vital Statistics.

   (b) Define the mean population of an area during a given period.

   (c) What do you mean by the force of mortality at age x? (5+5+6)=16

2. How does a current life table differ from a generation life table?

   What are the different measures of mortality in a life table? Show how they are inter-related, stating the necessary assumptions.

   Derive the death rate above age x in a life table population. (4+8+4)=16

3. How do population projections differ from population estimates? Explain the modified G.P. method for obtaining population estimates. Delineate the important steps in the component method for population projection. (4+5+7)=16

4. Write notes on any two of the following :

   (a) errors in census data;

   (b) net reproduction rate;

   (c) morbidity incidence rate and morbidity prevalence rate. (8+8)=16

NEATNESS (2)

GROUP (e): EDUCATIONAL AND PSYCHOLOGICAL STATISTICS (Half-paper –50 marks)

(Answer any three questions from this group)

1. (a) What is a T-scale? How would you derive T-scale equivalents for raw scores? Compare T-scale with other scales and discuss its advantages and disadvantages.

   (b) What do you understand by item validity? Describe the item validity indices which involve the slope of the regression/item score on test score. (8+8)=16

2. (a) Illustrate geometrically the facts of intercorrelations among the tests, common factor loadings, the concept of communality. (Consider the case where there are three tests and two common factors.)

   (b) What do you mean by

      (i) specific factor, and

      (ii) group factor? (12+4)=16

Please turn over

3. (a) "To the extent that a test is speeded, an odd-even coefficient of reliability is inflated."
 – Discuss.

 (b) Discuss the relation between reliability and the <u>range of ability</u> in the population. (8+8)=16

4. Discuss the following :

 (a) Relation of total score variance to the difficulty values of the items included in the test.

 (b) Relation of reliability to item intercorrelations.

 (c) Effect of item intercorrelations upon the total score distribution.

 (d) Effect of total score distribution upon the discrimination ability of the test. (4+4+4+4)=16

5. Write short notes on:

 (a) Percentile scale.

 (b) Incidental and explicit selections.

 (c) Cronbach and Warrington index of speededness of a test.

 (d) Differential prediction. (4+4+4+4)=16

NEATNESS (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – May 1976

Paper V (Practical): Methods of Numerical Computation; Descriptive Statistics and Official Statistics

Time: 5 hours | Full marks : 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is permitted.

GROUP A: Methods of Numerical Computation (26 marks)

(Answer any two questions from this group)

1. (a) Solve the following system of equations using any method known to you

$$2x_1 - 2x_2 + 4x_3 = -12$$
$$2x_1 + 3x_2 + 2x_3 = 8$$
$$-x_1 + x_2 - x_3 = 3.5$$

(b) Calculate approximate value of $f(x)$ for $x = 1.75$ by applying approximate interpolation formula to the following data.

1.0	0.84147
1.1	0.89121
1.2	0.93204
1.3	0.96356
1.4	0.98545
1.5	0.99749
1.6	0.99957
1.7	0.99166
1.8	0.97385

(6+7)=13

2. (a) Compute to 3 decimal places, the real root of

$$3x^3 + 5x - 40 = 0$$

(b) Given $Ax = b$

where $A = \begin{pmatrix} 1.22 & -1.32 & 3.96 \\ 2.12 & -3.52 & 1.62 \\ 4.23 & -1.21 & 1.09 \end{pmatrix}$

$b = \begin{pmatrix} 2.12 \\ -1.26 \\ 3.22 \end{pmatrix}$

find x.

(6+7)=13

3. Given the following rounded values of the function

$f(x) = \sqrt{\frac{2}{\pi}}\ e^{-x^2/2}$ for different values of x, find approximately,

(a) the value of x for which $f(x) = 0.6448$

Please turn over

3. (b) (contd.) the value of the integral $\Gamma(1) = \sqrt{\frac{2}{\pi}} \int_0^1 e^{-x^2/2} dx$

x	f(x)
0.000	.7978846
0.125	.7916754
0.250	.7733362
0.375	.7437192
0.500	.7041307
0.625	.6563219
0.750	.6022749
0.875	.5441190
1.000	.4839414

(6+7)=13

GROUP B: Descriptive Statistics (50 marks)

(Answer all questions from this group)

4. The following is the frequency distribution of marks of 240 students in a Public Examination, U column representing the distance of the class midpoint from some arbitrary origin in units of class-width, assumed constant. It is known that the arithmetic mean and standard deviation, is original units are 40.604 and 7.92 respectively. Identify the classes and draw the histogram.

U	Frequency
− 4	3
− 3	15
− 2	45
− 1	57
0	50
1	36
2	25
3	9
Total	240

Find also β_1 and β_2 of the distribution and comment.

(5+5+10)=20

5. (a) A certain car-hire firm has two cars, which it hires out day by day. The number of demands for a car each day is distributed as a Poisson distribution with mean 1.567.

Calculate the proportion of days in which neither car is used and the proportion of days in which some demand is refused.

(b) Test scores in 3 subjects are independently normally distributed with mean 35, 40 and 42 and s.d.'s 5.5, 6.2 and 6.5 respectively. Find the percentage of students securing 60 or more in all the subjects and percentage of those securing less than 30 in at least one of the subjects. (6+6)=12

Please turn over

6. In a partially destroyed laboratory records of an analysis of correlation data the following results were legible :

Variance of x = 0.34

Regression equations: $8.10x - 0.87y + 36.67 = 0$
$40.55x - 18.08y = 214.48$

What were (i) the mean of x and mean of y

(ii) the variance of y

and (iii) correlation coefficient between x and y. (6)

7. Calculate the trend values for the following data by using moving averages of an appropriate period. Represent the trend values and the original data on a graph paper.

year	Number of cars (in thousand)
1930	882.3
1931	714.4
1932	541.9
1933	561.0
1934	593.2
1935	635.8
1936	694.4
1937	724.4
1938	585.7
1939	352.1
1940	699.2
1941	813.3
1942	820.6
1943	816.2
1944	834.8
1945	806.1
1946	795.0

(8+4)=12

GROUP C: Official Statistics (24 marks)

(Answer both the questions from this group)

8. From the official publications placed at your disposal, collect data on <u>any three</u> of the following:

(a) Value of exports of merchandise from West Bengal, Tamil Nadu, Gujarat and Kerala in any two recent years.

(b) Stock of coal and lignite at factories at the end of any four consecutive months.

(c) Value of export of

i) Ayurvedic & Unani Medicine, and

ii) Homeopathic medicine

from India to any two countries in each case for any one month.

Please turn over

8. (d) Number of major bridges on all roads in Assam
(contd.) and Andhra Pradesh (in) any three recent years.

(e) Quantity of food grains procured by or for the Government of Bihar and Haryana in any two years. (4x3)=12

9. From the official publications supplied to you, collect data on number of tourists arrived in India from U.K., Canada, U.S.S.R., and Australia in recent five years, giving the percentage increase over the previous year. Present these data in a neat tabular form and comment on the salient features of the data collected by you. (12)

[Note:

i) Full reference to publications consulted are to be given along with the answers (title of publication, particular issue, table number, page etc.)

ii) The data furnished should be methodically arranged in a neat tabular form.

iii) Aritmetic manipulation of data, if necessary, is permissible. In that case original data collected are also to be shown, side by side.]

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – May 1976

Paper VI (Practical) : Statistical methods; Design & analysis of Experiments and Sample surveys

Time : 5 hours Full marks: 100

(A) Figures in the margin indicate full marks.

(b) Use of calculating machines is permitted.

GROUP A: Statistical Methods (40 marks)

(Answer any two questions from this group)

1. (a) Net income of 162 families in town A and 419 families in town B in a certain year are given below.

Annual net income (Rs.)	No. of families town A	No. of families town B
1500 – 3000	–	4
3000 – 4500	10	45
4500 – 6000	23	95
6000 – 7500	40	120
7500 – 9000	32	67
9000 – 10500	28	51
10500 – 12000	20	17
12000 – 13500	4	9
13500 – 15000	2	5
15000 – 16500	1	3
16500 – 18000	2	2
18000 – 19500	–	1

Calculate for each town the standard error of the mean income, on the assumption that the families included are a random sample of the middle class families in the town.

Hence calculate the standard error of the difference of the two mean incomes and test whether they differ significantly or not.

(b) A firm is advertising that it has been successful in designing a new home automatic clothes washer which is more effective in removing dirt than the most popular washer now in use. And in support of its claim, it is also displaying the following data on the dirt removed (in a suitable unit) by the most popular washer and the new washer for 14 equally sized and equally soiled loads of clothes which were washed with the same soap and for the same length of time, 7 loads being washed by each washer.

Dirt removed by								
	Popular washer	13	10	9	12	11	10	8
	New washer	10	11	12	13	9	14	12

Using median test, see if the firm's claim is genuine.

(12+8)=20

Please turn over

2. (a) The following data give the number of defective bricks found in four consecutive days of manufacture.

11	14	9	16
12	16	12	18
8	21	14	11
13	9	21	13
16	8	9	13
11	18	7	15
8	10	13	11
12	8	16	8
12	14	13	9
10	13	12	17

Test for the equality of variances of the populations.

(b) In a greenhouse experiment on wheat, four fertilizer treatments of the soil and four chemical treatments of the seed were used. Each combination was applied to two plots which were placed at random in the available space. The table gives the yields in some suitable unit. Analyse the data.

Fertilizer	Chemical treatment			
	1	2	3	4
1	21.4, 21.2	20.0, 20.3	19.6, 18.8	17.6, 16.6
2	12.0, 14.2	13.6, 13.3	13.0, 13.7	13.3, 14.0
3	13.5, 11.9	14.0, 15.6	12.7, 12.0	12.4, 13.7
4	12.8, 13.8	14.1, 13.2	14.2 13.6	12.0, 14.6

(8+12)=20

3. (a) A random sample of size 177 from a univariate population gives the following moments:

$$m_2 = 1.10781 \times 5^2$$
$$m_3 = 0.08057 \times 5^3$$
$$m_4 = 5.12575 \times 5^4$$

Test whether the population departs from normality in respect of (a) Skewness, and (b) Kurtosis.

(b) Counts of antibodies, before and after vaccination, yield powers of 2, 2^x being the count before and 2^y the count after vaccination. In an experiment, values of x and y occurred with the frequencies as in the table given below. Test whether the regression of y on x is linear or not.

y \ x	4	5	6	7	8	9
6	2					
7	2	1	2	1		
8		3	6	2	1	
9	2	1	3	5		1
10		2	2	3	1	1
11			1		3	

(5+15)=20

Please turn over

GROUP B: Design and Analysis of Experiments (30 marks)

(Answer all questions from this group)

4. Construct the B.I.B. design $v = b = 13$, $r = k = 4$, $\lambda = 1$. (10)

5. An experiment considered the effect, in terms of breaking strength of cotton fibres, of the level of potash in the soil. Five levels of potash were applied in a randomised blocks pattern with three blocks. The data are presented below. The values from one plot of block 2 and one plot of block 3 are missing.

Strength index of cotton

Potash (lb./acre)	144	108	72	54	36
Replicate					
1	7.46	7.17	7.76	8.14	7.63
2	7.68	7.57	–	8.15	8.00
3	7.21	7.80	7.74	–	7.93

Analyse the data to find the effect of the levels of potash considered. (20)

GROUP C: Sample surveys (30 marks)

(Answer all questions from this group)

6. The number of diseased plants (out of 9) in 40 areas are as in the following table:

sl.no.of area	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
no. of diseased plants	1	4	1	2	5	1	1	1	7	2	3	3	2	2	3	1	2	7	2	6

sl.no.of area	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
no. of diseased plants	3	5	3	4	5	1	4	5	4	1	4	2	6	5	3	4	1	7	3	6

(a) Draw a simple random sample (without replacements) of 10 areas.

(b) Draw a random sample of 5 areas with replacement and with probability proportional to the number of diseased plants.

(5+5)=10

Please turn over

7. EITHER

A stratified sample of 50 units gives the following estimated stratum means and variances:

Stratum number	N_i	n_i	$\bar{y}_i$	s_i^2
1	30	5	35	40
2	50	10	40	55
3	60	15	40	80
4	60	20	55	140

(a) Estimate the variance within the whole population. Verify that the existing allocation is optimal for 4 strata.

(b) Estimate the sampling variance of the estimated population mean for the above allocation and for a random sample of 50 drawn without stratification and comment. (12+8)=20

OR

To determine the yield-rate of paddy in a district six villages were selected at random and three plots were selected at random in each selected village. The yields, in suitable units, are given below:

Plot No.	Village					
	1	2	3	4	5	6
I	16	6	18	13	17	12
II	18	5	3	7	11	11
III	11	16	16	7	17	8

(a) Evaluate the estimate of the standard error of the estimate of mean yield per plot.

(b) Assuming a cost function of the form

$$C = 20.12 + 2.4\, m + 1.3\, m\, n$$

where m and n stand for the number of first-stage and second-stage units respectively, determine the optimum value of n for a given cost of $c_0 = 2{,}000$. (12+8)=20

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1976

Paper VII (Practical): Applied Statistics Group Papers

Time: 5 hours   Full marks: 100

(a) Candidates will be required to answer questions from those two groups of subjects only, for which they have already registered their options.

(b) Separate answer-books are to be used for each of the two groups attempted.

(c) Figures in the margin indicate full marks.

(d) Use of calculating machines is permitted.

GROUP (a): ECONOMIC STATISTICS (Half-paper - 50 marks)

(Answer any two questions from this group)

1. The standard deviation of logarithms of a two-parameter lognormal distribution of income declined from 0.70 to 0.60 over a period of 10 years. Draw Lorenz curves to depict the change in the given distribution. Also, compute the Lorenz measure of inequality in each case. (25)

2. The following table gives official estimates of production of rice and wheat in India from 1957-58 to 1964-65. Investigate if there were any significant trends in the two series.

Estimates of production of selected crops, India (1957-58 to 1964-65)

Year	Production ('000 metric tons)	
	rice	wheat
(1)	(2)	(3)
1957-58	25,718	8,513
-59	32,254	10,627
-60	31,930	11,316
-61	34,807	11,737
1961-62	35,085	12,801
-63	33,518	11,596
-64	37,157	10,519
-65	30,483	13,035

(25)

3. Table below gives the estimates of average monthly per capita consumption expenditures (y_j) on selected items by decile groups of the population based on a ranking in ascending order of per capita aggregate expenditure (x) on all the items, in urban India during 1963-64. Obtain Engel elasticities for the two items using appropriate methods. State your assumptions very clearly.

Please turn over

3. (contd.)

Decile group number	Average per capita agregate expenditure (Rs.) (x)	Average monthly per capita expenditure (Rs.) on cereals (y_1)	clothing (y_2)
(1)	(2)	(3)	(4)
1	12.49	5.94	.15
2	17.81	7.79	.36
3	21.34	8.52	.46
4	25.04	9.34	.85
5	29.54	9.61	1.38
6	35.20	9.86	1.47
7	42.40	10.19	2.36
8	52.88	10.23	3.46
9	69.12	9.77	5.24
10	123.53	10.05	12.74
All groups	35.42	8.89	2.04

GROUP (b): STATISTICAL QUALITY CONTROL (Half-paper 50 marks)

(Answer any two questions from this group)

1. (a) A sample of 25 units is selected at random from a process producing 8 per cent defectives. Compute the probability of getting 4 defectives in the sample using

(i) binomial distribution
(ii) Poisson approximation
(iii) Normal approximation.

(b) The following table gives the results of daily inspection of radio transmitting tubes for a certain quality characteristic.

Sample number	Number inspected	Number rejected	Sample number	Number inspected	Number rejected
(1)	(2)	(3)	(1)	(2)	(3)
1	106	4	11	189	15
2	122	8	12	162	24
3	29	0	13	33	4
4	50	0	14	51	4
5	39	1	15	69	4
6	60	2	16	120	8
7	162	12	17	108	7
8	18	2	18	142	10
9	133	8	19	36	1
10	33	0	20	141	22

i) Use a suitable control chart to analyse the above data.

ii) Is the process under statistical control?

iii) What standard should you recommend for future working?

Please turn over

1\. (c) (contd.) To estimate the overall average machine utilisation for a group of looms in a jute mill with 100 looms in one section, 5 rounds were taken and the utilisation was estimated to be about 80%. In order to obtain overall utilisation for this group of looms with an error of $\pm$ 2% with 95% confidence (use normal distribution) how many snap reading rounds have to be made on this group of looms? (5+12+8)=

2\. A double sampling plan is given by

$N = 1000, \quad n_1 = 50, \; c_1 = 0, \; n_2 = 100, \; c_2 = 2.$

i) Draw the OC curve, calculating at least 7 suitable points (justify any approximation that may be used).

ii) Obtain AOQL for the above plan assuming it as acceptance/ rectification plan.

iii) For this plan, obtain the lot qualities for which the probabilities of acceptance are 0.95 and 0.10 respectively. (12+7+6)=

3\. An experiment was conducted to test the effect of different treatments of warp beams on the warp breakage-rates during weaving.

4 warp beams P, Q, R and S were treated differently and were woven simultaneously on 4 looms over 4 days. At the end of each day, the warp beams were interchanged between the four experimental looms in such a manner as to ensure that after completion of the experiment the warp beam had worked on each of the four looms for one day.

The plan of the experiment and the warp breakage rates are given in the table below.

Warp breakage rate on different warp beams

	Weaving period			
Loom \ Day	1	2	3	4
1	5.52(P)	9.16(S)	5.77(R)	5.07(Q)
2	6.60(S)	5.14(R)	2.01(Q)	6.00(P)
3	2.87(Q)	6.02(P)	6.53(S)	2.83(R)
4	9.78(R)	6.26(Q)	8.00(P)	0.77(S)

(a) Carry out an analysis of variance (ANOVA) test to detect the effects of the warp beam treatment, loom differences and period differences if any; offer your comments.

(b) What is the relative advantage of this design over Randomised block design? State the assumptions involved in using this design. (17+8)=25

Please turn over

GROUP (c): STATISTICAL METHODS IN GENETICS (Half-paper - 50 marks)

(Answer all questions from this group)

1. In a family of Antirrhinum majos, obtained by selfing a yellow-flowered plant known to be heterozygous the following segregation for flower colour was observed

Yellow flowered plants	208
Ivory flowered plants	81

Is this in keeping with the 3:1 ratio expected from selfing a plant heterozygous for a single flower colour gene? (10)

2. The data are on the segregation of the two factors G, g and L, l in an F_2 of Primula sincusis. The factors were in the coupling phase and the following segregation was observed.

GL	Gl	gL	gl
977	16	19	360

Estimate the recombination fraction assuming it to be the same in both male and female gametogenesis. Give an estimate of the variance of this estimate. (15)

3. EITHER

The following table shows the numbers of persons in a random sample of size 435 from a particular community belonging to the various blood groups O, A, B and AB.

Blood group	Number of persons
O	176
A	182
B	60
AB	17

Obtain the maximum likelihood estimates for p, q and r which are the gene frequencies of O, A and B respectively. Also calculate the estimates of their variances. (25)

OR

The following data give the frequencies of eggs laid by gall-flies in flower-heads. The count of flower-heads with 'nil' egg is not available.

Number of eggs	1	2	3	4	5	6	7	8	9
Number of flower-heads	22	18	18	11	9	6	3	0	1

Assuming that the data follow Poisson distribution find the maximum likelihood estimate of the average number of eggs laid by a gall-fly and find its standard error. (25)

Please turn over

GROUP (d) :- VITAL STATISTICS & DEMOGRAPHY (Half-paper – 50 marks)

(Answer any two questions from this group)

1. The number of births occurring in a certain community in 1958 is shown here classified according to age of mother, together with the female population in each relevant age-group for a life table constructed for females:

Age years	Female population ('000)	Number of births	Females in life table (radix 100,000)
15 – 19	181.79	12343	468100
20 – 24	170.01	114511	464923
25 – 29	172.66	116734	458000
30 – 34	175.02	110218	452270
35 – 39	175.10	15131	444763
40 – 44	171.02	11422	435848
45 – 49	166.66	193	424873

The total population of the community in 1958 was 12285.8 thousand.

Determine
(a) the crude birth rate,
(b) the gross reproduction rate,
(c) the net reproduction rate (the sex-ratio at birth being 104.5 males to 100 female births),
(d) the approximate length of a female generation,
and (e) the rate of annual growth of the female population. (25)

2. From the following values of the usual life table functions, calculate the expectation of life at age 30:

Age-group	$_5q_x$
25 – 30	.0284
30 – 35	.0352
35 – 40	.0501
40 – 45	.0606
45 – 50	.0872

Take $l_{25} = 87,870$, $T_{45} = 1,685,581$. (25)

3. A population census taken in December, 1961 counted the number of children aged 0–4 to be 147,737. The number of children born in each year from 1957 to 1961 was as follows:

1957	69,208
1958	68,738
1959	66,680
1960	61,609
1961	50,209

Find the extent of under-enumeration in the census population of the age-group 0–4, given that the mortality corresponds to the following life-table stationary population:

Age-group (x to x+1)	L_x
0 – 1	91,076
1 – 2	85,582
2 – 3	83,158
3 – 4	81,617
4 – 5	80,585

($l_0 = 100,000$) (25)

Please turn over

GROUP (e): EDUCATIONAL & PSYCHOLOGICAL STATISTICS (Half-paper -50 marks)

(Answer any three questions from this group)

1. 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 the validities of test A and test B be, if both tests were increased to infinite length? (16)

2. (a) In a certain test, the sum of squared differences between scores on two comparable halves (i.e. halves with equal means and s.d.s) is 285, N is 50 and the standard deviation of the total score is 8.5. Find the coefficient of reliability for the total scores and the standard error of measurement.

(b) In a sample of 61 fifteen-year-old high school students, of whom 26 were male and 35 female, the mean weight in Kg. were 67.8 and 56.6, respectively. The standard deviation of the weights for the combined group was 13.2. Find the point-biserial correlation between sex and body weight for 15-year-old school students. (8+8)=16

3. (a) Ratings of seven individuals by three raters in a particular trait are presented below:

Ratees	Raters		
	A	B	C
1	3	4	5
2	5	5	5
3	3	3	5
4	1	4	1
5	7	9	7
6	3	5	3
7	6	5	7

Compute the reliability of a single rating and that of the average of the three ratings.

(b) In a test of 55 items, the standard deviation of the total scores was 7.5. The sum of the variances of the items was 9.83. Estimate the reliability of the scores. (10+6)=16

4. The following is a 'centroid' factor-loading matrix

Variable	Common factor coefficients		Communality h^2
	C_1	C_2	
1. Height	.830	.396	.846
2. Arm-span	.818	.469	.889
3. Length of fore-arm	.777	.470	.825
4. Length of lower arm	.708	.401	.708
5. Weight	.786	.500	.868
6. Bitrochantric diameter	.672	.458	.661
7. Chest-girth	.504	.444	.550
8. Chest-width	.647	.333	.529

Obtain the rotated factor-loading matrix, after suitable orthogonal rotations of the centroid factor-loading matrix. (16)

NEATNESS (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – May 1976

Paper VIII (Theoretical) : Subjects of First Paper of Specialisation

Time: 4 hours | Full marks: 100

(a) Candidates are required to answer from that group only for which they have already registered their options.

(b) Figures in the margin indicate full marks.

(c) Use of calculating machines is not permitted.

GROUP A : ECONOMIC STATISTICS

Econometrics – Special Paper I

(Answer any five questions from this group)

1. Explain clearly how you would test the following hypotheses (i) $\beta_1 = 0$ (ii) $\beta_1 = 0, \beta_2 = 0$ (iii) $\beta_1 + \beta_2 = 2$ in the classical normal linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$. State clearly the theoretical results needed to derive the tests. (20)

2. What are dummy variables? Explain, with illustrations, the various uses of dummy variables in econometric regression analysis. Discuss, in this connection, the meaning of the 'dummy variable trap'. (2+14+4)=20

3. Explain clearly what is meant by the problem of identification in the context of a simultaneous equation econometric model. Derive the rank condition of identifiability of a given structural equation of a simultaneous equation econometric model. (8+12)=20

4. Give economic interpretations of the parameters in the Constant Elasticity of Substitution (CES) production function and discuss methods of estimating these parameters. (6+14)=20

5. Define the lognormal distribution. How does one graphically test for the lognormality of an empirical distribution? Show how the 'law of proportionate effect' gives rise to an income distribution of the lognormal form. (3+5+12)=20

6. Explain the concept of the income elasticity of demand. Describe how you would obtain the Engel curves and the income elasticities of demand of different commodities from family budget data. (4+16)=20

7. For a two-commodity consumer's utility function of the Cobb-Douglas type

   i) Show that the consumer's demand for either commodity is independent of the price of the other

   ii) Calculate the price and income elasticities of demand for the commodities. (12+8)=20

Please turn over

8. Define National Income and describe in detail any method of measuring it. Distinguish between

i) net and gross national product,
ii) net national product at factor cost and at market prices,
iii) disposable income and personal income (4+9+2+3+2)=20

9. Write short-notes on any two of the following:

i) Fractile Graphical Analysis

ii) Houthakker's derivation of the Cobb-Douglass production junction as an aggregate production function.

iii) The method of instrumental variables.

iv) Estimation of the effect of household composition on the household consumer expenditure pattern. (10x2)=20

GROUP B: TECHNO-COMMERCIAL STATISTICS

Statistical Quality Control - Special Paper I

(Answer any five questions from this group)

1. (a) Explain the concept of rational sub-groups and its use in control chart techniques.

(b) Explain how to construct and set up group control charts and state the necessary conditions for the advantageous use of such charts.

(c) When would you recommend the use of "modified" control limits in an $\overline{X}$-R chart? Explain how such limits can be constructed and used in practice. (6+6+8)=20

2. (a) In what respects, does a cumulative sum control chart for sample means, differ from

i) an ordinary $\overline{X}$ chart,
and ii) a chart for moving averages?

(b) How is a cumulative sum control chart for sample means related to sequential tests for the mean of a normal population?

(c) Clearly outline the steps in using such a chart. (8+6+6)=20

3. (a) Distinguish between specification limits and tolerance limits.

(b) Explain the meaning of the term "Process Capability".

(c) Describe the various situations that may arise when the process capability is compared with product tolerances and state what course of action you would recommend in each situation and why. (4+6+10)=20

4. (a) For attribute inspection acceptance sampling plan, define the following terms -

i) AQL ii) LTPD iii) AOQL

(b) What are the relative advantages and disadvantages of single, double, and multiple sampling plans?

(c) Derive an expression for the average sample number of a double sampling attribute inspection plan under an acceptance/rejection scheme. (4+6+10)=20

Please turn over

5\. (a) Distinguish between the Type A and Type B operating characteristic curves of an acceptance sampling inspection plan, for attribute characteristics.

(b) Explain the basis for the following statement:

"the long term process average outgoing percent defective, may rarely be expected to exceed one-half the AOQL value associated with the inspections plan in use"

(c) For an inspector, the errors of mis-classification have been found to be as follows:

i) the probability of classifying a non-defective item as a defective = p_1

ii) the probability of classifying a defective items as non-defective = p_2 . Show that if π is the proportion of defectives in a lot of submitted product, the effective proportion defective.

$$\pi_{eff} = p_1 + (1 - p_1 - p_2)$$

(d) If $p_1 = 0$, what would be the effect of the inspector's error on the operating characteristic curve of the sampling plan. (5+5+5+5)=20

6\. (a) State the conditions for rotatability of a second order response surface design. Prove that a single equi-radial set of points satisfying all the moment restrictions for second order rotatability, cannot satisfy the non-singularity conditions. Show that by introducing a few centre points you can get over this difficulty.

(b) For a central composite design to be laid out in two blocks, state and prove the condition for orthogonal blocking. Indicate the experimental combinations to be carried out in each block, when the number of factors is five and a half-replicate of a 2^5 design is used. (10+10)=20

7\. Write short notes on any four of the following :

(a) Pareto Analysis.

(b) Vendor quality rating plans.

(c) Control by Narrow-limit guaging.

(d) Lot-plot technique.

(e) Demerit rating for outgoing quality assurance.

(f) Discovery sampling. (5+5+5+5)=20

GROUP C : BIOMETRIC METHODS

Special Paper I

(Answer any five questions from this group)

1\. For treating a disease D, there are T different drugs available in the market A group of N = KT (K is an integer) persons are known to suffer from the disease D.

(a) Design an efficient experiment to compare the effectiveness of the drugs against the disease D.

(b) Outline the analysis of your design giving a model and assumptions. (7+13)=20

Please turn over

2. Let N_1 and N_2 be the sizes of two samples from two p-variate normal populations. Let $\overline{X}_1$ and $\overline{X}_2$ be the sample mean vectors with corresponding ith components $\overline{X}_{1i}$ and $\overline{X}_{2i}$. Let $((S_{(p)}^{ij}))$ be the inverse of the usual pooled estimate of the common dispersion matrix $\Sigma_{(p)}$. Define

$$D_p^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} S_{(p)}^{ij} (\overline{X}_{1i} - \overline{X}_{2i})(\overline{X}_{1j} - \overline{X}_{2j})$$

(a) Show that $D_{p-1}^2 \leq D_p^2$.

(b) Obtain the distribution of D_p^2.

(c) Obtain the distribution of D_p^2 when $p = 1$, $N_2 = 5$ and N_1 tends to infinity. (7+7+6)=20

3. An expensive blood test determines whether or not a person is affected by a certain disease. The disease is present in about 1 out of every 100 individuals in a given population of size N. Every individual in the population is to be tested for the disease. The testing procedure is to pool the blood of n persons $(1 \leq n \leq N)$, where a single test has two possible results : either (i) all n individuals do not have the disease or (ii) atleast one of the individuals has the disease (call this group as G1). If the test results in (ii), a pooled blood of m individuals $(1 \leq m \leq n)$ from G1 is tested. Let G2 denote the group of remining n-m individuals.

(a) Let X denote the number of persons having the disease in G2. If the second test resulted in (ii), find the distribution of X.

(b) Find the distribution of X if the second test resulted in (i).

(c) Using the results of (a) and (b) suggest a procedure to classify all N individuals. (10+4+6)=20

4. Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_m$ and $\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, \ldots, \underset{\sim}{Y}_n$ be two sets of independent p-variate observations with mean vectors $\underset{\sim}{\mu}_X$ and $\underset{\sim}{\mu}_Y$ and covariance matrices Σ_X and Σ_Y respectively. Let the population distributions be p-variate normal.

Describe test procedure(s) to test the null hypothesis $\underset{\sim}{\mu}_X = \underset{\sim}{\mu}_Y$ when $\Sigma_X \neq \Sigma_Y$ and

(a) $m = n$

(b) $m \neq n$ (10+10)=20

Please turn over

5. Let X_{1i} and X_{2i} be two observations on a rat numbered i, (i = 1, 2, ..., N) before and after a treatment. The joint distribution of (X_{1i}, X_{2i}) is bivariate normal with mean (μ_1, μ_2) and covariance matrix $\begin{pmatrix} a^2 & \rho a^2 \\ \rho a^2 & a^2 \end{pmatrix}$. Let X be a similar observation on a new rat. Using the likelihood ratio principle, find whether the new rat has been treated or not when

(a) μ_1, μ_2, a^2 and ρ are known.

(b) Only a^2 is known. (10+10)=20

6. Write notes on <u>any two</u> of the following:

(a) Multiple correlation.

(b) Principal component analysis

(c) Factor analysis

(d) Confounding

(giving definition, computational procedure and applications).

GROUP D: DESIGN AND ANALYSIS OF EXPERIMENTS

<u>Statistical Aspects - Special Paper I</u>

(Answer <u>any five</u> questions from this group)

1. (a) Discuss the role of 'randomisation' in planning of experiments.

(b) How do you randomize in a (i) 5 x 5 Latin Square experiment (ii) BIB experiment with parameters $b = 14, v = 8, r = 7, k = 4, \lambda = 3$?

(c) How do you obtain a random permutation of 14 symbols using (i) a table of random numbers (ii) a table of random permutations of ten digits? (8+6+6)=20

2. (a) When is an experimental design said to be (i) connected (ii) variance balanced (iii) efficiency balanced?

(b) Consider an experiment involving v treatments in b blocks of experimental units of sizes $k_1, k_2, \ldots, k_b$ and its analysis under the usual assumption of additivity of block and treatment effects. Show that here (i) a linear function of treatment effects is estimable only if it is a contrast (ii) all contrasts are estimable if and only if the design is connected.

(c) Further if the design is equireplicate and $k_1 = k_2 = k_b$ then show that it is variance balanced only if it is either a RBD or a BIBD. (6+6+8)=20

3. (a) What is meant by 'recovery of inter-block information' in an incomplete block experiment?

(b) Illustrate this by obtaining combined intra- inter-block estimates of treatment differences in a balanced incomplete block design and their standard errors.

(c) Explain the role of 'resolvability' in this connection.

(4+12+4) = 20

4. (a) Derive the analysis of a two associate partially balanced incomplete block design.

(b) Describe all computational steps using a triangular association scheme for illustration. (13+7) = 20

5. In a split plot experiment involving 3 levels of irrigation i_0, i_1 and i_2 and there varieties v_0, v_1 and v_2 the whole plot treatments (i_0, i_1 and i_2) are assigned to the three whole plots of each of four replication as in a randomised block lay out and so are the subplot treatments to the three subplots within each whole plot. The yield for the combination (i_1, v_0) is missing for replication 1. Obtain the analysis of variance for the available data. (20)

6. Consider a design for fitting a second degree multiple regression equation.

$$Y = \beta_0 + \sum \beta_i x_i + \sum \beta_{ii} x_i^2 + \sum_{h<i} \beta_{hi} x_h x_i$$

$(i, h = 1, 2, \ldots k)$

satisfying the following symmetry and scaling conditions

$$\sum_{j=1}^{N} x_{ji} = \sum_{j=1}^{N} x_{ji}^3 = \sum_{j=1}^{N} x_{ji} x_{jh} = \sum_{j=1}^{N} x_{ji}^2 x_{jh}$$

$$= \sum_{j=1}^{N} x_{ji}^3 x_{jh} = 0$$

$$\sum_{j=1}^{N} x_{ji}^2 = b, \quad \sum_{j=1}^{N} x_{ji}^2 x_{jh}^2 = c, \quad \sum x_{ji}^4 = c+d$$

for each i, h, $i \neq h$. Here (x_{j1}, x_{j2}, ... x_{jk}) are the values specified for the k explanatory variables for the jth observation (j = 1, 2, ... N).

(a) Obtain necessary and sufficient conditions to satisfy

i) orthogonality $\text{Cov}(\hat{\beta}_{ii}, \hat{\beta}_{hh}) = 0, i \neq h$

ii) rotatability

(b) Give an example of a central composite design which is orthogonal and also rotatable. (8+8+4) = 20

Please turn over

7. (a) What are the special problems that need attention when one is analysing long term experiments involving repetitions of similar experiments over space or in time.

(b) A varietal experiment on wheat carried out in four randomised blocks of four plots each is repeated on three consecutive years at the same site (using the same layout each year). Describe a suitable analysis for this experiment. (10+10)=20

8. Write short notes on <u>any two</u> of the following topics

(a) Quasi Latin Squares

(b) Variance Component analysis

(c) Cross-over Designs. (10+10)=20

GROUP E : SAMPLE SURVEYS

<u>Theoretical Aspects: Special Paper I</u>

(Attempt <u>any four</u> questions from this group)

1. (a) Show how to compute the standard error of the mean of n uncorrelated unbiassed estimates of a parameter.

(b) If a simple random sample of size n be drawn with replacement from a population of N units then discuss the efficiencies of

i) the mean of all the sample values and

ii) the mean of the values of the distinct units in the sample only relative to the sample mean of n units selected at random without replacement. (9+16)=25

2. (a) Explain the principle of stratification of a finite population. Discuss how you will determine the best points of division of a population into k strata (k being determined in advance) for proportional allocation,

i) when the stratification variable is identical with the variable under study in the survey and

ii) when the stratification variable is linearly related with the estimation variable. You may assume the population to be an infinite one with probability density $f(y)$, $a < y < b$.

(b) If the values on a character y be recorded on independent sample random samples of sizes n_h $(h = 1, \ldots, k)$ drawn without replacement separately from k strata of respective sizes N_h $(h = 1, \ldots, k)$ of a population having $N = \sum_{h=1}^{k} N_h$ units, then on the basis of the information so collecte obtain an unbiased estimate of variance of the mean that might be based on a simple random sample of size $n = \sum_{h=1}^{k} n_h$ from the entire population disregarding strata taken without replacement. (4+7+7+7)=25

<u>Please turn over</u>

3. (a) Suppose in a two stage sampling scheme n primary sampling units (PSU) out of N are selected with varying probabilities without replacement with inclusion-probabilities of PSU's and pairs of PSU's as π_i and π_{ij} $(i, j = 1, \ldots N;\ i \neq j)$ respectively and each selected PSU is subsampled in a certain known manner. If T_i be an unbiased estimator of the ith PSU total Y_i based on subsampling at the second stage with σ_i^2 as its variance such that an unbiased estimator $\hat{\sigma}_i^2$ is available for σ_i^2, then show that the sample total of $\frac{T_i}{\pi_i}$ is an unbiased estimator for the population total of Y_i's and obtain expressions for its variance and for an unbiased estimator of the variance.

(b) Let $R = \frac{Y}{X}$ be the ratio of population totals of two variables y and x having ρ as the correlation coefficient If a simple random sample of size n be drawn without replacement from this population of size N, then show that

$$\hat{R} = \bar{r} + \frac{(N-1)\, n\, (\bar{y} - \bar{r}\,\bar{x})}{N\,(n-1)\,\bar{X}}$$

is an unbiased estimator for R.

Assuming n to be sufficiently large show that $V(\hat{R})$ can be closely approximated by

$$(S_y^2 + \bar{R}^2 S_x^2 - 2\bar{R}\rho S_x S_y)/n\bar{X}^2$$

where $\bar{R} = E(Y_i/x_i)$, $\bar{r}$ = sample mean of $\frac{y}{x}$ values, and other notations are standard. (13+12)=25

4. Derive the expression for the variance of the mean of a systematic sample (population size N being an integral multiple of the sample size n) from a finite discrete sequence in terms of the intrasystematic sample correlation coefficient. Investigate the relative precision of systematic, stratified random and simple random sampling procedures in the case where the population consists only of a linear trend.

If N is not an exact multiple of K so that $N = nk + f$ where k is an integer and $0 < f < k$, how can you draw a systematic sample so that the sample mean becomes an unbiased estimate of the population mean? In this case, if 'every kth' systematic sample be chosen at random, what will be your unbiased estimator for the population total ? (6+10+5+4)=25

5. (a) A population consists of N clusters of M elements each. Consider a simple random sample of n complete clusters. Obtain the variance of the estimated mean per element and hence show that for a given bulk of sample elements, if the intra-cluster correlation $\rho > 0$, the cluster as a unit of sampling is less precise than a simple random sample of elements. Explain the differences among cluster sampling, stratified sampling and two-stage sampling.

Please turn over

5. (b) (contd.) A population consists of N primary units, the ith primary unit containing M_i sub-units. A sample of n primary units is selected with equal probability and with replacement. If the ith primary unit is selected, a simple random sample of m_i sub-units is drawn. Give an unbiased estimate for the population mean per sub-unit. Derive the expression for variance of the unbiased estimate and an estimate for this variance. $\overline{(7+4+2+6+6)}$=25

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

(a) PPS systematic sampling.

(b) Lahiri's method of sampling.

(c) Interpenetrating networks of sub-samples.

(d) Unbiased ratio and ratio-type estimators.

(e) Symmetrized Des Raj estimators.

GROUP F : TECHNIQUES OF COMPUTATION

Numerical Analysis - Special Paper I

(Answer any five questions from this group)

1. Find the root of smallest magnitude of the equation

$$x^2 + 0.4002\,x + 0.8.10^{-4} = 0$$

using four significant digit floating point arithmetic. Assuming that the numbers 0.4002 and 0.8 x 10^{-4} are approximate and the true values lie within the intervals (0.4001, 0.4003) and (0.7 x 10^{-4}, 0.9 x 10^{-4}) respectively, find the interval in which the root of smallest magnitude lies. Use four significant digit floating point arithmetic. (20)

2. Define the operators Δ, δ and μ. Show that

$$\mu = \frac{2+\Delta}{2\sqrt{1+\Delta}} = \sqrt{1+\frac{1}{4}\delta^2}$$

If $p_n(x)$ is a polynomial of degree n with leading coefficient a_n, x_0 is an arbitrary point and h is the increment of x, then show that

$$\Delta^n p_n\,(x_0) = a_n\, n!\, h^n$$

Determine the spacing h in a table of equally spaced values of the function $f(x) = \sqrt{x}$ between 1 and 2 so that interpolation with a second degree polynomial in this table will yield accuracy to 5 places of decimals. (20)

3.

Please turn over

3. Derive Simpson's rule for numerical integration and obtain its error term. For which polynomial is Simpson's rule exact ? Construct a rule of the form

$$f(x)\,dx = A_0\,f\left(-\tfrac{1}{2}\right) + A_1\,f(0) + A_2\,f\left(\tfrac{1}{2}\right)$$

which is exact for all polynomials of degree less than or equal to two.

By a double application of Simpson's rule, derive the formula

$$\int_{x_0}^{x_2}\int_{y_0}^{y_2} f(x,y)\,dx\,dy \simeq \frac{hk}{9}\Big[(f_{00}+f_{02}+f_{20}+f_{22}) + 4\,(f_{01}+f_{10}+f_{12}+f_{21}) + 16\,f_{11}\Big]$$

where $x_r = x_0 + rh$, $Y_s = Y_0 + sk$ and $f_{rs} \equiv f(x_r, y_s)$. (20)

4. Derive Newton-Raphson iteration formula for the solution of an equation $f(x) = 0$. Show that under certain conditions, Newton-Raphson method converges quadratically. Show that the Newton-Raphson iteration, as applied to $f(x) = x^n - a$ for the determination of $\alpha = a^{\frac{1}{n}}$ is of the form

$$Z_{k+1} = \frac{1}{n}\left[(n-1)Z_k + \frac{a}{Z_k^{\,n-1}}\right]$$

and that if $\epsilon_k = \alpha - Z_k$, then

$$\epsilon_{k+1} \simeq -\frac{n-1}{2\alpha}\,\epsilon_k^2 \quad \text{when} \quad Z_k \simeq \alpha \,.$$ (20)

5. Describe a method, giving relevant derivations, to compute the inverse of a non-singular square matrix. State the limitations of the method. Calculate the inverse of the matrix,

$$\begin{bmatrix} 2 & 3 & -1 \\ 4 & 4 & -3 \\ -2 & 3 & -1 \end{bmatrix}$$

using the method you have described. (20)

6. Write notes on any <u>two</u> of the following :

(a) Determination of Stationary values,

(b) Solution of a system of linear equations,

(c) Euler-Maclaurin sum formula (20)

Please turn over

GROUP G : STATISTICAL INFERENCE

General Theory - Special Paper I

(Answer any five questions from this group)

1. (a) Explain clearly the concepts of unbiasedness and consistency in the context of estimation of parameters.

(b) If an estimator $T_n(X_1, \ldots, X_n)$ based on a sample of size n is unbiased for the parametric function $g(\theta)$ for all n, does it follow that the sequence $T_n(X_1, \ldots, X_n)$ is consistent for $g(\theta)$? If $T_n(X_1, \ldots, X_n)$ is a consistent estimator of $g(\theta)$ is it also unbiased? Illustrate your answers by an example in each case.

(c) Derive a set of sufficient conditions for a sequence of estimators $T_n(X_1, \ldots X_n)$ to be consistent for a parametric function $g(\theta)$. (5+10+5)=20

2. (a) Let $x_1, \ldots, x_n$ be n independent observations from a population with density function $f(x, \underset{\sim}{\theta})$, where $\underset{\sim}{\theta}$ is unknown parameter vector with components $\theta_1, \ldots, \theta_k$. Under suitable regularity conditions to be stated precisely, derive a lower bound for the generalized variance of any set of unbiased estimators of $\theta_1, \ldots, \theta_k$.

(b) Deduce from (a) that the variance of the estimator of θ_i is not less than a quantity which is defined independently of any method of estimation.

(c) Illustrate your answer by taking

$$f(x, \underset{\sim}{\theta}) = \frac{1}{\sqrt{2\pi\theta_2}} \exp \frac{-1}{2\theta_2}(x-\theta_1)^2$$

(10+2+8)=20

3. (a) Let $L\left(g(\theta), \delta(x)\right)$ denote the loss function and $R(\theta, \delta) = E_\theta\left\{L\left(g(\theta), \delta(X)\right)\right\}$ denote the risk function. Let $\lambda(\theta)$ denote a prior distribution on the parameter space (H). Suppose

$$\int_{(H)} R(\theta, \delta_\lambda)\, d\lambda(\theta) = \sup_\theta R(\theta, \delta_\lambda).$$

Show that (i) δ_λ is minimax

(ii) If δ_λ is the unique Bayes solution with respect to λ, it is unique minimax procedure

and (iii) $\lambda(\theta)$ is least favourable distribution on (H).

(b) Suppose X has Binomial distribution $B(n, p)$ and the loss function is given by $L(p, \delta) = (p - \delta)^2$. A natural estimator of p is $\frac{X}{n}$. Is it minimax? (12+8)=20.

Please turn over

4. (a) Let $f(x, \theta)$ be the probability density of a random variable X, where θ is an unknown parameter. For testing the simple hypothesis $H_0 : \theta = \theta_0$ against the simple alternative $H_1 : \theta = \theta_1$,

i) Show that there exists a test function $\phi(X)$ and a constant K such that

$$(1)\ E_{\theta_0}\{\phi(X)\} = \alpha \quad (0 < \alpha < 1)$$

and (2) $$\phi(x) = \begin{cases} 1 \text{ when } f(x, \theta_1) > K\ f(x, \theta_0) \\ 0 \text{ when } f(x, \theta_1) < K\ f(x, \theta_0) \end{cases}$$

ii) Show that if a test satisfies the conditions (1) and (2) above for some K, then it is most powerful for testing H_0 against H_1 at level .

iii) Show that if a test $\phi(x)$ is most powerful at level α for testing H_0 against H_1, then for some K it satisfies condition (2).

(b) Deduce that the power β of the most powerful level α test $(0 < \alpha < 1)$ for testing H_0 against H_1 is greater than α unless $f(x, \theta_0) = f(x, \theta_1)$ for all x.

(c) Let X be the number of successes in n independent trials with probability p of success. Derive the most powerful test of level α for testing $H_0 : p = p_0$ against $H_1 : p = p_1 (> p_0)$. Show that this test is uniformly most powerful for all alternatives $p > p_0$. (10+2+8)=20

5. (a) Define unbiased confidence set of a parameter θ.

(b) Show that unbiased families of tests lead to unbiased confidence sets and uniformly most powerful unbiased family of tests lead to uniformly most accurate unbiased confidence sets.

(c) Let $x_1, \ldots x_n$ be a sample from $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown. Obtain the uniformly most accurate unbiased confidence interval for σ^2. (3+10+7)=20

6. (a) Let $f(x, \theta)$ be the density function of a random variable X, where θ is an unknown vector of parameters $\theta_1, \ldots, \theta_k$. Explain the concept of Likelihood Ratio Test for testing the simple hypothesis $\theta_1 = \theta_1^0, \ldots, \theta_k = \theta_k^0$ based on a sample of size n. Obtain the asymptotic distribution of a suitable transform of the test criterion.

(b) Let $x_1, \ldots, x_n$ be a sample of size n from $N(\mu, \sigma^2)$. Derive the likelihood ratio test for testing the composite hypothesis $H_0 : \mu = \mu_0$, σ^2 being unspecified. (12+8)=20

7. Write short notes on :

i) Complete class and essentially complete class of decision functions.

ii) Completeness and bounded completeness of a family of daitributions.

iii) Sufficient statistics and Fisher Information. (20)

GROUP H : PROBABILITY THEORY

Basic Probability – Special Paper I

(Answer any six questions from this group)

1. (a) Suppose X is a set and $\underline{S}$ is a class of subsets of X such that 1. $\phi \in \underline{S}$, $X \in \underline{S}$ 2. $A, B \in \underline{S} \Longrightarrow A \cap B \in \underline{S}$ 3. $A \in \underline{S} \Longrightarrow A^c$ is a finite disjoint union of sets from $\underline{S}$. Show that the Class F of finite disjoint unions of sets from S is a field of subsets of X.

(b) Let X = [0, 1) show that the class F of finite disjoint unions of intervals of the form [a, b), $0 \leq a \leq b \leq 1$ is a field of subsets of X.

(c) Suppose on the above field $\underline{F}$ you define P by

$P(A) = 1$ if for some $a < 1$, $[a, 1) \subset A$

$= 0$ otherwise

Is P countably additive? (5+3+8)=16

2. (a) Suppose f is a real valued function on a measurable space such that |f| is measurable. Then should f be measurable?

(b) Suppose a sequence $\{f_n\}$ of measurable functions on a probability space (X, A, P) converge pointwise to a function f. Let $\in > 0$ show that there is a set of measure less than $\in$ such that on the complement of that set f_n converges to f uniformly. (5+11)=16

3. (a) Define the term "product of two measure spaces".

(b) Suppose $(\Omega, \underline{A}, P)$ is a probability space and $(R, \underline{B}, \lambda)$ be real line with Borel σ field and Lebesque measure. Suppose f is a +ve integrable function on Ω. Let $G = \{(x,y) \in \Omega \times R: f(x) > y\}$. Show that the product measure, that is, $P \times \lambda$ measure of the set G is $\int f \, dP$. (6+10)=16

4. (a) For a sequence of random variables on a probability space, show that almost everywhere convergence implies convergence in probability.

(b) Give an example (and prove that it is an example) to show that converse of the above statement is not always correct. (8+8)=16

5. (a) State and prove Borel-Cantelli Lemma.

(b) Suppose f_n is a sequence of independent random variables on a probability space such that $\underset{n}{\text{Limit}} \frac{1}{n} f_n(x) = 0$ a.e.

Show that $\sum_{n=1}^{\infty} P\{x: |f_n(x)| > n\} < \infty$

Here P is the underlying probability. (8+8)=16

Please turn over

6. (a) Define the notion of "independence" of two random variables and of a sequence of random variables.

(b) Let $X = [0, 1)$ with Lebesgue measure. Define $f_n(x) = +1$ or -1 according as the integer i for which $\frac{i-1}{2^n} \leq x < \frac{i}{2^n}$ is odd or even.

Show that f_n is a sequence of independent random variables.

(6+10)=16

7. (a) When do you say that a sequence F_n of distribution functions on the real line converges weakly to a distribution F.

(b) Suppose $F_n(x) = 0$ if $x < 1/n$

$= 1$ if $x \geq 1/n$

Does this sequence of distribution functions converge?

(c) Suppose F_n is a sequence of distribution functions on the real line such that for every rational number x, $F_n(x) \longrightarrow F(x)$ where F is a distribution function. Show that $F_n \rightarrow F$ weakly.

(3+6+7)=16

8. (a) Let Ω be the unit square, λ be Lebesgue measure and f be the random variable $f(x,y) = x+y$. Calculate the distribution function of the random variable f.

(b) Let Ω be the unit square and λ be Lebesgue measure. Let P be the measure defined on the Boreal σ field of Ω by

$$P(A) = 4\lambda\left\{A \cap \left[0, \tfrac{1}{2}\right] \times \left[0, \tfrac{1}{2}\right]\right\}$$

Show $P \ll \lambda$. What is the Radon Nikodym derivative of P w.r.t λ? Is $\lambda \ll P$?

(8+8)=16

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – May 1976

Paper IX (Theoretical): Subjects of Second Paper of Specialisation

Time: 4 hours Full marks: 100

(a) Candidates are required to answer questions from that group only for which they have registered their options.

(b) Figures in the margin indicate full marks.

(c) Use of Calculating Machine is not permitted.

GROUP A: ECONOMIC STATISTICS

Special Paper II: Indian Economics and Economics of Planning

Section I: Indian Economics (50 marks)

(Answer question No.5 and any two of the rest)

1. What do you mean by the "Structure of the Indian economy"? What are the criteria used for specifying the structure in broad terms? How are they related to the procedures of national income accounting? Comment on the significance of "Structural changes" for "economic development" generally. What light does this throw on Indian economic experience since independence? (5+3+2+5+5)=20

2. What, in your opinion, are the principal changes in Indian agriculture that have taken place in recent years (say, since 1960)? Do you think that it is possible to associate any definite trends in the volume of agricultural production, or its structure, with these changes? Discuss. (8+12)=20

3. Briefly trace the evolution of industrial policy in India since 1956 in terms of its objectives and instruments. Do you think that there have been any definite shifts in the policy over these years? (15+5)=20

4. Give a short description of the nature and extent of inflation experienced recently in India. (Please mention and justify(briefly) the dating of the inflationary period under reference). What, in your opinion, were the principal factors behind the inflation? Briefly comment on the policies adopted for combatting inflation. (8+7+5)=20

5. Write a short note on any one of the following:

(a) The machinery for settling industrial disputes in India.

(b) Agricultural taxation in India.

(c) The pattern of credit allocation of commercial banks in India. (10)

Section II: Economics of Planning (50 marks)

(Answer question No.5 and any two of the remaining questions)

6. Let Y_0 and S_0 be the levels of national income and savings in a country at time, t = 0. It is assumed that all savings are invested and that there is a constant marginal (incremental) capital–output ratio, 'K'. Let r_t be the rate of growth of income ($\Delta Y_t/Y_t$) at time t.

Please turn over

6. (a) (contd.) What is the value of r_o ?

(b) Let 's' be the marginal (incremental) savings - income ratio $(\Delta S_y / \Delta Y_t)$. For what values of 's' does r_t increase over time? Does r_t approach any constant value in this case? If so, what value?

(c) Derive the expression for the rate of growth of per-capita income, given a constant rate of growth of population. Suppose that a certain target value for the former is printed, say ρ . Deduce the required investment-income ratio. Suppose this ratio is smaller than (S_o/Y_o), and the gap is financed by drawing upon "foreign exchange". Let F_t be the quantum of foreign exchange drawn upon at time t. What condition on 's' is required for F_t to decline over time? Assuming the condition to be satisfied, can you find out the time, t*, when $F_{t*} = 0$?

(d) Briefly comment on the significances of a "foreign exchange constant" and the "burden of interest-payment on foreign debt" in the light of your answers above. (2+5+8+5)=20

7. (a) Give a brief account of the developments of input-output analysis and linear programming. For what sort of problems can these be used as tools of planning? Do you think that it is feasible to construct a model of planning for an entire economy on the technical basis provided by those tools?

(b) Indicate briefly the uses of the above tools for planning in India, if there are any that you know of . (15+5)=20

8. (a) Give a connected account of the financing of Indian plans, beginning with the Third Five Year Plan, indicating the broad division of sources of revenue relevant for this purpose. What role has taxation in particular played in this? And "foreign aid" ?

(b) Review the necessity of evolving a definite tax-policy in India in this context. (15+5)=20

9. Give an outline of the draft Fifth Five Year Plan of India. Does it seek to tackle any "new" problems that were left out, or very inadequately dealt with, in the earlier plans? Is there anything "new" in its approach to older problems? Discuss fully. (20)

10. Write a short-note on any one of the following:

(a) The concept and use of "capital-output ratio" at the sectoral and national levels of planning.

(b) The significance of price-policy of the public sector for mobilisation of resources for planning.

(c) The rationale and critique of a policy of investment in the "heavy industries" at an early stage of development of an economy. (10)

Please turn over

GROUP B : TECHNO-COMMERCIAL STATISTICS

(Special Paper II)

Section I : Operations Research	(70 marks)
Section I (Alternative): Elements of Book-keeping and Accountancy	(70 marks)
Section II : Statistical Methods in Business	(30 marks)

Section I: Operations Research (70 marks)

(a) Use a separate answer-book for this Section.

(b) Attempt any four questions from this Section.

1. A company has two factories each of which can make four products A, B C and D with following costs and capacities

	Cost in Rs./lb. of product made A	B	C	D	Maximum capacity of all products in lbs./week
Factory 1 : Normal time	1.3	1.4	1.4	1.2	70,000
" : Overtime	1.4	1.5	1.6	1.4	14,050
Factory 2 : Normal time	1.3	1.4	1.5	1.4	90,000
" : Overtime	1.5	1.5	1.7	1.7	18,000

Find out the amounts which each factory should produce by regular production and by overtime production so as to meet the demand at minimum total cost. (17)

2. (a) State Bellman's principle of optimality and explain as to how it can be used to solve multistage decision problems.

(b) There are three marketing areas and the profit obtained in any area depends on the number of salesmen allotted to that area. The following information is available:

Number of salesmen allotted to the area	Profit obtained in thousands of rupees Area 1	Area 2	Area 3
0	38	40	60
1	41	42	64
2	48	50	68
3	58	60	78
4	66	66	90
5	72	75	102
6	83	82	109
7	96	88	119
8	102	95	124

Assuming that we have a total sales force of 8 salesmen, allot the salesmen to areas using Dynamic Programming to maximise the profit. (4+13)=17

Please turn over

3. (a) For an M/G/1 queue, show that $E(w) = \frac{\rho(1+\mu^2\sigma^2)}{2(1-\rho)} E(v)$,

where $E(w)$ = expected waiting time of a customer, not including the service time,

$E(v)$ = expected service time of a customer,

λ = mean number of customers arriving in unit time,

μ = mean number of customers served in unit time,

$\rho = (\frac{\lambda}{\mu})$ = traffic intensity.

and σ_v^2 = variance of service time distribution.

(b) There are N lamps on a life test and the testing is terminated as soon as the rth failure occurs. Failed lamps are not replaced. Assume that life (X) of a lamp has the probability density function

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad x \geq 0, \quad \theta > 0.$$

Show that the expected waiting time to the rth failure is

$$\theta \sum_{i=1}^{r} \left(\frac{1}{N-i+1} \right).$$

(10+7)=17

4. Show, under suitable assumptions (which you must state), that the optimal values of order quantity (Q) and re-order level (r) satisfy the following relations:

(1) $Q = \sqrt{\frac{2D\,[A + b\,\eta(r)]}{ac}}$,

(2) $\int_r^\infty f(x)\,dx = \frac{acQ}{bD}$,

where D = yearly demand, A = cost of ordering,

a = cost of carrying inventory,

c = unit cost of the item,

b = cost of one unit short,

$f(x)$ = frequency function of lead time demand

and $\eta(r) = \int_r^\infty (x-r)\,f(x)\,dx$.

Describe an iterative procedure for obtaining the optimal values of Q and r. (12+5)=17

5. Following is a linear programming problem:

Maximise $Z = X_1 + X_2 + X_4$

subject to
$$X_1 + X_2 + X_3 + X_4 = 4$$
$$X_1 + 2X_2 + X_3 + X_5 = 4$$
$$X_1 + 2X_2 + X_3 = 4 \qquad X_i \geq 0 \quad i = 1, 2, 3, 4, 5$$

i) Find out all the optimal basic feasible solutions by using the Simplex method.

ii) Write down the general form of an optimum solution. (12+5)=17

6. (a) Write, briefly, on variance reduction techniques for Monte-Carlo method.

(b) Write short notes on : (i) Newspaper-boy problem
(ii) Machine-repair problem
(iii) Warehousing problem. (5+3x4)=17

NEATNESS (2)

Section I (Alternative): Elements of Book-keeping and Accountancy
(70 marks)

(a) Use separate answer book for this section.

(b) Attempt Question No.1 and any other three from this section.

1. M/s. Brown & Philip are equal partners in a business, trading under the style of "The Celebrated Cycle Co." On 31st December, 1975, their book-keeper extracted the following balances from the books:-

Cash at Bank Rs.3,990. Plant & Machinery Rs.6,880.
Stock of finished goods on 1.1.75 Rs.7,858. Material on hand on 1.1.75 Rs.2,730. Sales during the year Rs.1,88,680.
Agents Commission Rs.4,890. Brown's Capital Rs.5,358.
Philip's Capital Rs.4,000. Workmen's Wages Rs.26,798.
Business Premises Rs.2,500. Bills Payable Rs.3,840.
Cash Creditor Rs.4,000. Material Purchased Rs.1,41,040.
Sundry Creditors Rs.4,462. General Expenses Rs.720.
Travellers' Salary Rs.3,520. Brown's drawings Rs.900.
Bad Debt written off Rs.958. Rent, Rates and Taxes Rs.1,900.
Sundry Debtors Rs.6,556.

The following adjustments are necessary:

5 per cent depreciation off Plant & Machinery; one year's interest due to cash creditor at 5 per cent annually;
Bad Debts estimated at Rs.320; Commission to Agents Rs.790;
Stock on hand at 31.12.75 - Materials Rs.1,860 and Finished Goods Rs.6,250.

Prepare Trading, Profit & Loss A/c and a Balance Sheet as at 31st December, 1975. (6+10+9)=25

2. (a) What purpose does a Journal serve in modern Book-keeping?

(b) Give the Journal entries necessary to rectify the following errors detected in the books:

i) A payment of Rs.2,500 for purchase of a Typewriting Machine for office use has been debited to the Purchases Account from Cash Book.

ii) A credit sale of Rs.150.50 to Sri D Banerjee has been posted to the debit of Sri D Guha's Account from the Sales Day Book.

iii) A payment of Rs.966 for whitewashing the office rooms has been charged to Building Account. (3+12)=15

Please turn over

3. Bombay Manufacturing Co. Ltd. bought machinery on 1st January, 1972 for Rs.20,000 and began to depreciate it @ 10 per cent per annum on the Diminishing Balance Method.

On 31st December, 1975, the Company decided to change the method of depreciation from Diminishing Balance Method to Straight Line Method with effect from 1st January, 1972 and to adjust the difference in depreciation up to 31st December, 1975 through Profit and Loss Account of 1975. The life of the machinery is estimated to be 10 years.

Show the Machinery Account during these years. (15)

4. On 1st July 1975, Export Agency Ltd. received from Jansen Co. Ltd. two Bills of Exchange for Rs.7,000 and Rs.5,000 for 3 months and 2 months respectively.

On 1st August, 1975, the first Bill was endorsed in favour of a creditor. On the same date the second bill was discounted through Bank at 5 per cent. On due dates both the bills were dishonoured.

Show the journal entries necessary to give effect to the above transactions in the books of Export Agency Ltd. (15)

5. What do you understand by a Triple Column Cash Book? Give the ruling of such a Cash Book and enter five specimen entries therein. (5+10)=15

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

(a) Trade Discount and Cash Discount.

(b) Opening Entries and Closing Entries.

(c) Bank Reconciliation Statement.

(d) Impersonal Accounts.

(e) Income earned but not received. (5x3)=15

<u>Section II: Statistical Methods in Business</u> (30 marks)

(a) Use a separate answer-book for this section.

(b) Attempt <u>any two</u> questions from this section.

1. (a) Compare the following methods of determining relative job values: i) ranking ii) classification iii) factor comparison iv) point system.

(b) Discuss the importance of job analysis in a job evaluation programme. (10+5)=15

2. (a) What components of demand are involved in a forecast?

(b) Why do the moving averages discount random effects?

(c) Compare the results of simple moving average forecasts and exponential moving average forecasts. (4+4+7)=15

3. A firm has a certain budget C for the advertisement of n items it sells. Its profits from sales per unit investment in the n items are C_1, C_2, ..., C_n, respectively. It cannot invest more than a_1, a_2, ..., a_n for advertisement on those commodities respectively. Show how it should invest its funds on the various items? Discuss the case for n = 2. (15)

<u>Please turn over</u>

GROUP C : BIOMETRIC METHODS

Special Paper II : Statistical Methods in Genetics and Bio-assays.

Section I : Statistical Methods in Genetics (50 marks)

(Attempt any two questions from this section)

1. Describe available statistical methods for detection and estimation of linkage between two factors A and B on the basis of progeny of crosses of the types Aa Bb x aabb and Aa Bb x Aa Bb, explaining clearly how results from several such crosses can be tested for homogeneity and combined when justified. (25)

2. Describe the Sib method for estimating the ratio in single-factor segregation in human genetics, explaining carefully how the effect of the method of ascertainment is taken care of in the analysis. (25)

3. Describe the mechanism of inheritance of the O-A-B blood-group system and explain its significance in resolving medico-legal cases. Obtain the likelihood equation for estimation of the frequencies of the genes O, A and B in a panmictic population from which a random sample giving phenotypic frequencies of O, A, B and AB blood-groups is available. Describe the method of scoring for calculating the estimates. (25)

Section II: Bio-assays (50 marks)

(Attempt any two questions from this section)

4. Define the terms tolerance distribution and median effective dose in connection with a biological assay with quantal response. Working out the necessary statistical theory, develop a method for estimating the median effective dose in a quantal assay and also its standard error. (25)

5. Examine critically the implications of the assumption that the tolerance distribution is –

   i) logistic;

   ii) log-normal;

   iii) angle.

   Work out the appropriate dose or response meta-meter in each case. How do you examine which assumption is valid in any given case? (25)

6. Write short-notes on any three :

   i) Fieller's theorem.

   ii) Parallel-line assays and Slope-ratio assays.

   iii) Twin Cross-over Assay.

   iv) Spearman-Karber method of estimation of ED50.

   v) Correction for natural response in quantal assays. (25)

Please turn over

GROUP D : DESIGN & ANALYSIS OF EXPERIMENTS

Special Paper II: Combinatorial Aspects

(Answer *any five* questions)

1. (a) Write down the sum and product tables for GF(5).

   (b) Define 'primitive element' of a finite field. Find all primitive elements of GF(5).

   (c) Write down a latin-square L of order 5 which has the property that the four latin-squares obtained by the cyclic permutations of the last four rows of L while keeping the first row fixed are mutually orthogonal. (5+5+10)=20

2. (a) Define a Balanced Incomplete Block Design (BIBD) with parameters v, b, r, k and λ ; derive the relations between these parameters.

   (b) If v is even show that a necessary condition for the existence of a symmetric BIBD is that $K-\lambda$ must be a perfect square.

   (c) Show that there does not exist a symmetric BIBD with $v = 22$, $k = 7$ and $\lambda = 2$. (5+10+5)=20

3. Consider the Galois Field of residue classes mod p, where p is an (odd) prime.

   (a) Define 'quadratic residues'.

   (b) Show that there are exactly $(p-1)/2$ quadratic residues and they are given by :

$$\left\{ (i^2) \; : \; i = 1, 2, \ldots (p-1)/2 \right\}$$

   (c) If $p = 3 \bmod 4$, show that the quadratic residues form a difference set with respect to addition. Write down the parameters of this difference set.

   (d) Write down the quadratic residues of GF(7) and construct a symmetric BIBD with $v = 7$, $k = 3$ and $\lambda = 1$. (5+5+5+5)=20

4. (a) Give the axiomatic definition of a projective plane of order n.

   (b) Given $GF(p^m)$ discuss the construction of a projective plane of order p^m (p is a prime and m is a positive integer).

   (c) State, <u>without proof</u>, a theorem on the non-existence of projective planes of certain orders. (6+7+7)=20

5. (a) Discuss the principle of confounding in a factorial experiment, giving appropriate examples.

   (b) Give a plan for a 3^4 – experiment in blocks of 9 confounding ABD, ACD^2, AB^2C^2 and BC^2D^2 where A, B, C, D are four factors each at three levels. (10+10)=20

Please turn over

6. (a) Discuss the concept of fractional replication in factorial experiments.

(b) Give a plan for a $\frac{1}{4}$ – replicate of a 2^5 experiment based on the defining relation

$$I = -ABC = -ADE = BCDE.$$

(c) In a 2^6 – experiment identify all the confounded interactions if the key-block is

(1)	(ab)	(acde)	(bcde)
(acdf)	(bcdf)	(ef)	(abef)

(6+7+7)=20

7. Write short notes on any two of the following :

(a) Orthogonal arrays and their relation to mutually orthogonal latin-squares.

(b) Hadamard matrices and their use in the construction of fractional factorials.

(c) Partially balanced incomplete block designs. (10+10)=20

GROUP E : SAMPLE SURVEYS

Special Paper II : Organisational Aspects

(Answer all questions)

1. Draw up a precise plan of work and give an outline of the schedule or questionnaire that you would suggest for ascertaining, through a sample enquiry, the number of persons above the age of 25 years in a municipal town with a population of 25 thousand.

If the result of sample enquiry is to be tested against the complete count of all persons above the age of 25 years, what additional steps would you suggest and what precautions would you recommend? (12+8)=20

2. A study is proposed to be carried out for ascertaining the number of persons likely to retire from the State Govt. service on super-annuation at age 58 years, in each year of the next 5 years. Give a critical assessment of the relative advantages and disadvantages of the following approaches for the study in view:

(a) a census of all employees and ascertaining the age of each employee by interrogation.

(b) a complete enumeration of service records of all employees and finding out their age from the office records lying with the appointing authorities.

(c) a sample survey of a few of the departments and ascertaining the age of a sample of employees in the selected departments by interviewing the employees.

Is there any other approach (one or more) that can be used with advantage for the study in view? If so, give an indication of the approach and give your assessment of the relative advantages and disadvantages. (5+5+5+5)=20

Please turn over

3. The production of fish is required to be estimated in a State where fish is usually consumed without being transacted in the usual markets. How would you approach the problem? Give a precise outline of the steps to be taken. Suggest the control records and checks that the field supervisors should maintain. (20)

4. Write critical notes on the role of the following in relation to the large-scale sample surveys:

(a) Pilot Survey.

(b) Pretesting of Schedules.

(c) Consistency Checks of Data Collected.

(d) Variance Function.

(e) Double Sampling. (4x5)=20

5. On receipt of a report of sudden flood in a certain district in a State, it has been decided to conduct a quick survey to make an objective assessment of the effect of the flood on the standing crops and on house-property. Suppose you are in charge of a team of field workers who are posted in different parts of the district and you are asked to undertake the job. Draw up a suitable set of guidelines and instructions for the field staff containing, inter alia, the following

(a) the purpose of survey

(b) the basic approach of the survey

(c) the frame and the sampling unit

(d) the rates of work and the programme

(e) the precautions to be taken

(f) the arrangement for reporting progress of field work and the suggested form. (3+3+3+3+3+5)=20

GROUP F : TECHNIQUES OF COMPUTATION

Special Paper II: Numerical Computation

Time: 5 hours (Practical with Desk Calculators) Full marks: 100

(Answer any four questions)

1. (a) Find the number of correct digits in the quotient $u = 25.7/3.3$ assuming the dividend and divisor are correct to the last digit given.

(b) Determine with what relative error and with how many correct digits we can find the side x of a square if its area s = 12.34 to the nearest hundredth.

(c) Youngs modulus is determined from the deflection of a rod of rectangular cross-section by the formula

$$E = \frac{1}{4} \frac{l^3 p}{a^3 bs}$$

where l is the length of the rod, a and b are the dimensions of the cross-section, s is the bending deflection, and p is the load. Compute the relative error in a determination of Young's modulus E if p = 20 Kg, $\delta p = 0.1\%$, a = 3mm, $\delta a = 1\%$, b = 44 mm, $\delta b = 1\%$, l = 50 cm, $\delta_l = 1\%$, s = 2.5 cm and $\delta_s = 1\%$.

(5+8+12)=25

Please turn over

2. (a) Find all roots of the equation $5x^3 + 2x^2 - 15x - 6 = 0$ by Graeffe's Root Squaring Method.

(b) Find a real root of the system (correct to 3 place of decimals each)

$$2x^3 - y^2 - 1 = 0$$

$$\text{and } xy^3 - y - 4 = 0$$

in the neighbourhood of $x = 1.2$ and $y = 1.7$ (15+10)=25

3. (a) The following table gives the values of the function $f(x)$ for different value of x.

x	f(x)
2	0.12500
3	0.03703
4	0.01562
5	0.00800
6	0.00463
7	0.00292

Form the difference table and hence obtain the values of $f(1.9)$ and $f(6.5)$.

(b) Given a table of values of the function $y = f(x)$

x	y
321.0	2.50651
322.8	2.50893
324.2	2.51081
325.0	2.51188

Using a suitable formula find the value of f (323.5)

(12+13)=25

4. (a) Evaluate $I = \int_0^1 \frac{dx}{1+x^2}$ correct upto 5 places of decimals and hence obtain the value of π. Compare the result with the true value of π.

(b) Evaluate $I = \int_0^1 \sqrt{1+2x}\, dx$ using the Gaussian formula with 3 ordinates. (13+12)=25

5. Tabulate by Picard's method of successive approximation the numerical solution of $\frac{dy}{dx} = x + y$ correct upto 4 places of decimals with $x_0 = 0$, $y_0 = 1$ between $x = 0.0$ and $x=0.5$ with interval 0.1 (25)

6. Find the characteristic roots and the corresponding characteristic vectors of the matrix

$$\begin{pmatrix} 3 & 10 & 5 \\ -2 & -3 & -4 \\ 3 & 5 & 7 \end{pmatrix}$$

(25)

GROUP G : STATISTICAL INFERENCE

Special Paper II: Special Topics

(Answer any five questions)

1\. (a) Discuss the importance of Non-parametric methods in Statistical Inference. Mention a few practical situations where such methods are usually used.

(b) Obtain the sign test for the hypothesis that the populations from which two samples of size n are drawn, are identical.

(c) If a sequence of integers in given, how will you test, using method of run that the sequence is random. (8+8+4)=20

2\. (a) Explain the term "Statistical tolerance limit". Calculate the probability that the maximum and minimum in a sample of size n from a continuous population will cover at least a portion x of the population.

(b) Calculate the expectation and variance of the proportion covered by the sample range. (10+5+5)=20

3\. (a) Describe the sequential probability ratio test (SPRT) procedure for testing a simple hypothesis against a simple alternative.

(b) Obtain a suitable approximation to the OC curve of the SPRT in such a case.

(c) What are the good features of SPRT? (5+10+5)=20

4\. (a) Describe Stein's two stage sampling for finding an interval estimate of given length of mean of a normal population (univariate) – when the variance is unknown.

(b) Describe how you would tackle the problem in actual sampling case.

(c) Why the usual method of single stage sampling with t statistic fails in this case. (10+5+5)=20

5\. (a) Two multivariate Normal populations are known to have identical dispersion matrices. Write down the form of D^2 – statistic to test the hypothesis that means of the two populations are same.

(b) Show that D^2 – statistic remains invariant under the same linear transformation of variables and that its sampling distribution involves only the population value of D^2 as a parameter. (6+7+7)=20

6\. Let $\underset{\sim}{X}$ (p x 1) be a random p vector. Prove that

(a) A necessary and sufficient condition that $\underset{\sim}{X}$ (p x 1) is $N_p(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$ is that for every $\underset{\sim}{t}' = (t_1, \ldots, t_p)$, $\underset{\sim}{t}'\underset{\sim}{X}$ is $N_1(\underset{\sim}{t}'\underset{\sim}{\mu}, \underset{\sim}{t}'\underset{\sim}{\Sigma}\underset{\sim}{t})$ N_p stands for p variate normal.

(b) Use this result or otherwise prove that $\underset{\sim}{\alpha}' S \underset{\sim}{\alpha} / \underset{\sim}{\alpha}'\Sigma\underset{\sim}{\alpha}$ is central chi-square with n-1 d.f. where $\underset{\sim}{S}$ (p x p) is $W_p(n-1, p, \underset{\sim}{\Sigma})$. $W_r(n-1, p, \underset{\sim}{\Sigma})$ means p dimensional Wishart matrix based on sample size n from p variate normal population. (12+8)=20

Please turn over

7. (a) Define principal components. Given the covariance matrix $\underset{\sim}{\Sigma}$, describe and justify the iterative procedure for obtaining the first principal component.

(b) Define canonical correlations and canonical variates. Show that the multiple correlation coefficient can be regarded as the first canonical correlation.

(c) Discuss the use of principal components in factor analysis.

(3+6+3+4+4)=20

8. Write notes on any two of the following :

(a) Discriminant function.

(b) Chi-square test of Goodness of fit.

(c) Contingency tables and its applications.

(d) Likelihood ratio test.

(e) Hotelling's Generalized T^2 statistic. (10+10)=20

GROUP B : PROBABILITY

Special Paper II : Limit Distributions

(Answer any five questions)

1. (a) Let X, X_n, n = 1, 2, ... be real-valued random variables. What is meant by the statement: "$\{X_n\}$ converges to X in distribution (in law)" ?

(b) Show that $\{X_n\}$ converges in distribution to (the constant real-valued random variable) μ if, and only if, $X_n \to \mu$ in probability.

(c) Is there any situation where, from the convergence in distribution of a sequence $\{S_n\}$, one can infer the almost sure (almost everywhere) convergence of that sequence? Give a complete description, no proofs.

(3+12+5)=20

2. (a) State the inversion formula for obtaining the distribution function from the characteristic function.

(b) EITHER

Deduce the uniqueness theorem concerning characteristic functions.

OR

Deduce that if f is a characteristic function such that $\int_{-\infty}^{\infty} |f(t)|\, dt$ exists finitely, then the corresponding distribution function has a continuous version of its probability density function given by

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)\, e^{-itx}\, dt.$$

(6+14)=20

Please turn over

3. (a) Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed real-valued random variables with finite expectation μ and finite variance σ^2. By using suitable expansions for their common characteristic function, or otherwise, show that, if $S_n = X_1 + \cdots + X_n$, then

i) $S_n/n \to \mu$ in distribution,

ii) $(S_n - n\mu)/\sigma\sqrt{n} \to \Phi$ in distribution, where Φ is the standard normal distribution.

(b) What relation, if any exists between the statement in (a-i) and the "weak law of large numbers"? (8+8+4)=20

4. (a) Let F be a distribution function on the real line, which is purely discrete and has jumps only at the integer-values. If f is the characteristic function of F, check that $f(2\pi)=1$.

(b) Check that the characteristic functions of the (standard) Binomial and Poisson distribution functions satisfy the above relation.

(c) Show that if a characteristic function f is such that $f(2\pi)=1$, then the corresponding distribution function F is of the kind described in Part (a).

(Hint: $1 - \cos\theta \geq 0$ for all θ real). (4+4+4+8)=20

5. Let $\{X_{nj}\}$, $j=1, \ldots, k_n$; $n=1, 2, \ldots$ be a double sequence of real-valued random variables satisfying the condition of uniform asymptotic negligibility (UAN), namely,

(*) $\max_j P[|X_{nj}| > \epsilon] \to 0$ as $n \to \infty$, for every $\epsilon > 0$.

Show that each of the following conditions is equivalent to

(a) $\max_j \int \frac{x^2}{1+x^2}\, dF_{nj}(x) \to 0$ as $n \to \infty$; and

(b) $\max_{|t| \leq T; j} |f_{nj}(t) - 1| \to 0$ as $n \to \infty$, for every $T > 0$.

(10+10)=20

6. (a) State the Lévy-Cramér continuity theorem concerning the convergence in distribution of a sequence of random variables and the convergence behaviour of the sequence of their characteristic functions.

(b) If g_j, $j = 0, 1, \ldots, n$, are characteristic functions, show that $(\sum_0^n a_j g_j)/(\sum_0^n a_j)$ is also a characteristic function, if the a_j are all positive. (Identify the distribution function which will have the above as its characteristic function, in terms of the a_j and the G_j, G_j being the d.f. corresponding to g_j).

Please turn over

6. (c) (contd.) Deduce that, for any fixed characteristic function $\underline{f}$ and positive numbers a_j, the function

$$\left(\sum_{0}^{n} a_j f^j\right) / \left(\sum_{0}^{n} a_j\right)$$

is a characteristic function.

(d) Show, using (a) and (c), or otherwise, that, for an arbitrary characteristic function $\underline{f}$, the function $\exp[\lambda(f-1)]$ is also a characteristic function, for any $\lambda > 0$.

(e) Check that the new characteristic function in (d) above is infinitely divisible. (4+2+2+10+2)=20

7. State each of the following theorems, providing a complete description of the assumptions and notations :

(a) Liapounov's form of the Central Limit Theorem;

(b) The Lindeberg-Feller form of the Central Limit Theorem; how does (a) follow from this?

(c) The Berry-Esseen theorem giving bounds for the "error" in the normal approximation given by (a). (6+8+6)=20

8. (a) Define an infinitely divisible characteristic function (IDCF).

(b) Show that an IDCF never vanishes.

(c) Write down (no proofs) any representation theorem you know for the logarithm of an IDCF. (3+12+5)=20

9. (a) Explain how stable laws arise in the context of limit theorems -- as limit distributions of what kinds of sequences of random variables? Provide as complete a background as you can.

(b) State two familiar classes of probability laws which are stable laws.

(c) Are there any stable laws with finite variance? Name them. Are they the only such laws?

(d) Assume that a "stable" characteristic function $\underline{f}$ satisfies the condition that, for any $b_1, b_2 > 0$, there exist some real $\underline{a}$ and positive $\underline{b}$ such that

$$f(b_1 t) \cdot f(b_2 t) = e^{iat} \cdot f(bt).$$

Show that any $\underline{f}$ satisfying such a condition is necessarily infinitely divisible. (5+4+3+8)=20

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – May 1976

Paper X (Practical) : Subjects of Third Paper of Specialisation

Time: 5 hours   Full marks : 100

(a) Candidates are required to answer questions from that Group only for which they have registered their options.

(b) Figures in the margin indicate full marks.

(c) Use of calculating machine is permitted.

GROUP A : ECONOMIC STATISTICS

Special Paper III – Practical

(Answer Question No.5 and any two from the remaining questions)

1. The following shows the distribution of sample households by level of per capita income according to an enquiry in Kanpur city carried out in 1958–59:

per capita monthly income (Rs.) :	0–10	10–20	20–30	30–40	40–60	60–100	100–150	150–
no. of sample households:	7	93	131	107	121	95	51	38

i) Examine graphically whether the underlying size distribution is approximately two-parameter lognormal.

ii) Do you think a three-parameter lognormal distribution will fit the data much better than a two-parameter lognormal distribution? Examine briefly.

iii) Estimating the parameters of the two-parameter lognormal distribution by any appropriate method, compute the expected number of sample households in the interval 30 – 40 (Rs.) (30)

2. The following shows the values of three variables y, x_1 and x_2 for 14 consecutive years, where y = common logarithm of production in billion dollars at 1929 prices, x_1 = logarithm of employment in million, and x_2 = logarithm of stock of capital in billion dollars at 1929 prices, all figures relating to the US economy:

year	y	x_1	x_2	year	y	x_1	x_2
1940	2.0052	1.6955	2.4930	1947	2.1160	1.7731	2.5555
1941	2.0542	1.7332	2.5151	1948	2.1204	1.7706	2.5625
1942	2.0326	1.7716	2.5302	1949	2.1100	1.7686	2.5601
1943	2.0220	1.8122	2.5405	1950	2.1607	1.7782	2.5725
1944	2.0293	1.8105	2.5434	1951	2.1821	1.8048	2.5866
1945	2.0360	1.8090	2.5491	1952	2.1884	1.8122	2.5982
1946	2.1186	1.7701	2.5556	1953	2.2038	1.8195	2.6107

Suppose that the following model holds:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t, \quad t = 1, 2, \ldots, 14$$

where the ϵ_t's are independently distributed as $N(0, \sigma^2)$.

Utilize the above data to compute

i) best linear unbiased estimates of $\beta_0, \beta_1, \beta_2$,

ii) an unbiased estimate of σ^2 and

iii) 95% confidence interval for β_2.

Also test the null hypothesis that $\beta_1 + \beta_2 = 1$ against the alternative $\beta_1 + \beta_2 > 1$. (30)

3. It is known that the yield response of wheat to the application of nitrogen is of the form $y = \alpha + \beta N + \gamma N^2 + u$ where y is yield per hectare ~~in~~ quintals, N is the dose of nitrogen applied per hectare in kilogrammes, u a random distribance term with mean zero, and α, β, γ are parameters. The following data were obtained in an experiment:

Dose of Nitrogen	Yield			
	Replication Number			
	1	2	3	4
0	11.3	7.8	15.5	14.5
40	26.1	24.3	28.7	31.2
80	34.6	26.9	33.9	35.9
120	38.5	29.8	39.6	39.2
160	42.1	32.3	44.7	41.6
200	41.3	33.2	46.4	42.5

i) Estimate the response function by least squares.

ii) Estimate the dose of nitrogen that will maximize the expected yield per hectare.

iii) Given that the price of wheat is Rs.105 per quintal, price of nitrogen is Rs.2 per kilogram, and that other ~~costs~~ costs of cultivation are Rs.0.20 per kilogram of grain harvested, estimate the dose of nitrogen that will maximize the expected profit per hectare. (30)

4. $Y_{1t} = \beta_{14} Y_{4t} + \gamma_{11} X_{1t} + u_{1t}$ is the first equation in a system of four simultaneous equations that includes two other endogenous variables (Y_{2t} and Y_{3t}) and two other exogenous variables (X_{2t} and X_{3t}). Given the following data estimate this equation by two stage least squares.

Y_1	Y_4	X_1	X_2	X_3
122	127	1	1	143
127	134	1	2	152
132	141	1	3	162
148	160	1	4	187
173	191	1	5	219
184	198	1	6	229
212	230	1	7	262
263	275	1	8	309
262	280	1	9	319
287	309	1	10	353
306	334	1	11	384
330	353	1	12	410

(30)

Please turn over

5. A shipping company has three types of ships. Each ship of type 1 can carry cargo whose total weight does not exceed 1000 tonnes and which does not occupy more than 20,000 cubic metres of space. The corresponding weight and space constraints on ships of types 2 and 3 are respectively 2000 tonnes and 30,000 cubic metres, and 1500 tonnes and 40,000 cubic metres. The company has 4 ships of type 1 and two ships each of types 2 and 3. Each ship regardless of type can carry either steel or cotton or a combination of the two commodities subject only to the relevant weight and space constraints. A tonne of steel occupies 11 (eleven) cubic metres of space while a tonne of cotton occupies 110 (one hundred and ten) cubic metres of space. Each tonne of steel transported fetches Rs.100 to the company while a tonne of cotton fetches Rs.600. At the port 11,000 tonnes of steel and 2000 tonnes of cotton are awaiting shipment. Derive the amount of each commodity that should be loaded in each ship of each of the three types so as to maximize the freight earnings of the company. (40)

GROUP B: TECHNO-COMMERCIAL STATISTICS

(Special Paper III : Practical)

Section I : Statistical Quality Control (50 mark

(a) Use separate answer book for this section

(b) Attempt any two questions from this section.

1. An automatic process produces a certain type of components at the rate of 20 per minute. The outer diameter of these components must not exceed a specified value of 85 units. The components which have an outer diameter greater than 85 units are classified as defectives and have to be scrapped, as cost of rework is very high. Due to the deterioration of some of the parts of the process the outer diameter of the components produced increases with the passage of time. Samples of four components were taken from the process every fifteen minutes. When reset, the process produces components, initially, with a mean value of $\bar{x}$ = 50 units. The average and range values of these measurements for 25 samples are given below:

Sample number	Average	Range	Sample number	Average	Range
(1)	(2)	(3)	(1)	(2)	(3)
1	55.5	15	14	47.3	65
2	54.5	7	15	66.0	31
3	54.8	20	16	69.3	15
4	53.0	5	17	71.3	18
5	48.3	25	18	63.8	15
6	45.5	37	19	70.0	17
7	55.8	20	20	75.0	13
8	56.3	9	21	70.5	18
9	58.3	7	22	73.0	18
10	57.3	27	23	74.0	4
11	63.3	5	24	81.8	31
12	58.8	3	25	70.8	49
13	54.3	30			

1. (a) ~~Test the~~ linearity of the regression trend line.
(contd.) (b) Work out a suitable control chart scheme and recommend the period after which the process is to be reset ~~so that~~ no defective is produced. (25

2. A double sampling plan is given by

$$N = 2000, \quad n_1 = 100, \quad n_2 = 150, \quad c_1 = 1 \text{ and } c_2 = 4$$

The lots rejected by the plan are 100 per cent inspected and all the defectives found are replaced by good ones.

(a) Draw OC curve for this plan, calculating at least 6 points and stating the approximations used, if any.

(b) Draw the ASN curve for the above plan.

(c) Find the average amount of inspections per lot for lots having 3 per cent defectives. (25)

3. A study was undertaken of the factors affecting variation in the length of steel bars the lengths of four bars from each of four screw machines and two heat treatments were measured with the following results (the data are given in suitably transformed units).

Machine	Heat Treatment W	Heat Treatment L	Machine	Heat Treatment W	Heat Treatment L
A	6	4	C	1	− 1
	9	6		2	0
	1	0		0	0
	3	1		4	1
B	7	6	D	6	4
	9	5		6	5
	5	3		7	5
	5	4		3	4

Assume that the length of steel bar X_{ij} when machine i and heat treatment j are used is a normally distributed random variable, X_{ij}'s are independent and varianc of X_{ij} is σ^2. Formulate the following problems precisely and solve.

i) Is there interaction between the machines and the heat treatments?

ii) Are the expected lengths of steel bars different for the two heat treatments ?

iii) Are the expected lengths of steel bars different for the different machines? (25)

Please turn over

GROUP B: Section II: Operations Research (30 marks

(a) Use separate answer book for this section

(b) Attempt any two questions

1. There are 3 jobs to be assigned, one each to 5 machines and the associated cost matrix is as follows:

machines	jobs 1	2	3	4	5
1	11	17	8	16	20
2	9	7	12	6	15
3	13	16	15	12	16
4	21	24	17	28	26
5	14	10	12	11	15

(a) Find an approximate solution to the assignment problem given above, treating it as a transportation problem.

(b) Solve the given assignment problem using the Hungarian method. (15)

2. The following table gives the sales/year, rate of production, inventory carrying cost and set up cost for 5 items manufactured by M/s. Bolts and Nuts Ltd., using the same equipment.

product	sales/ year	production/ day	inventory cost/ unit/year (Rs.)	set up cost (Rs.)
1	5,00,000	10,000	0.020	8.00
2	13,000	1,000	0.005	10.00
3	2,00,000	10,000	0.008	17.00
4	1,80,000	2,000	0.022	5.00
5	92,000	1,000	0.015	6.80

Obtain the optimal number of cycles in a year assuming that there are 300 working days available in a year. (15)

3. A ready-made garment manufacturer has to process 7 items through the two stages of production, viz. cutting and sewing. The time taken for each of these items at the different stages are given below in appropriate units.

Item	processing time: cutting	sewing
1	5	2
2	7	6
3	3	7
4	4	5
5	6	9
6	7	5
7	12	8

(a) Find an order in which these items are to be processed through these stages so as to minimize the total processing time.

3. (b) (contd.) Suppose a third stage of production is added viz., pressing and packing with processing time for these items as follows:

item	processing time pressing and packing
1	10
2	12
3	11
4	13
5	12
6	10
7	11

~~Find~~ an order in which these seven items are to be processed so as to minimize the time taken to process all the items through all the three stages.

(c) Is the order of items obtained in (a) as good as the order of items obtained in (b) ? (15)

Section II (Alternative): Elements of Book-keeping and Accountancy - Practical (30 marks)

(a) Use separate answer-book for this section.

(b) Attempt any two questions from this group, Question No.1 being compulsory.

1. From the following particulars, prepare a Bank Reconciliation Statement as at 30th June, 1975:

		Rs.
(a)	Overdraft balance on 30th June, 1975 as per Bank Statement	13,095
(b)	Cheque deposited in bank not recorded in Cash Book	500
(c)	Cheque received and recorded in the bank column but not sent to bank for collection	1,085
(d)	Several Cheques were drawn in the last week of June, 1975 totaling Rs.16,050; of these cheques totalling Rs.8,050 were cashed. Similarly several cheques totalling Rs.10,400 were sent for collection. Of these Cheques of the value of Rs.2,500 were credited on 3rd July, 1975 and the remaining Cheques were credited before 30th June, 1975.	
(e)	On 12th June, 1975 the credit side of the bank column of the Cash Book was cast Rs.1,500 short and on 15th June, 1975 the credit balance of Rs.2,400 was brought forward on 16th June, 1975 as Debit balance of Rs.2,400.	
(f)	Interest of Rs.1,800 was charged by the bank but was not recorded in Cash Book.	

(20)

2. ~~What do~~ you understand by Double Entry System of Book-keeping? What books of account are usually maintained under this system by a large business undertaking? (10)

3. What is depreciation? Why and how is it charged to accounts. (10)

GROUP B: Section III: Statistical Methods in Business (20 marks)

(a) Use separate answer-book for this section.

(b) Attempt all questions.

1. Analyse the following transition matrix and determine the long-run equilibrium market share for each firm.

From	To A	To B	To C
A	1.00	0.00	0.00
B	0.10	0.75	0.15
C	0.10	0.05	0.85

Given that the initial market share is (0, 0.5, 0.5) for A, B and C, obtain the market share for Y after two transitions. (10)

2. A manufacturer of electronic calculators has the following sales for his product:

Month	Year 1969	1970	1971
January	30	112	115
February	115	125	165
March	23	160	32
April	88	186	135
May	152	220	128
June	94	180	69
July	170	60	263
August	82	53	84
September	40	40	45
October	41	40	40
November	120	21	110
December	180	40	165

Develop a forecasting plan that would help him in production scheduling. (10)

Please turn over

GROUP C : BIOMETRIC METHODS

Special Paper III - Practical

(Answer <u>any three</u> questions)

1. (a) The effect of apple mosaic on growth was measured on 2-year old seedling unpruned trees. propagated from infected and uninfected buds from the same mother tree. From the following data on the product of the height and cross-sectional area of the stem of each tree (in cubic cms.), determine the effect of the disease upon both the mean and the uniformity of growth :

VIRUS PRESENT			HEALTHY	
893	660	444	1384	1324
325	202	742	1325	1065
~~701~~	991	170	1870	1652
741	16	184	1515	1446
408	1121		1759	

(b) The following statistics are based on 54 specimens of *Neesiella* species A and 23 specimens of *Neesiella* species B. Here

X_1 : length of leaf (cm)

X_2 : length of inflorescence (cm)

X_3 : length of fruit (cm)

X_4 : length of calyx in fruit (cm)

Mean vector of species A: (4.28, 1.99, 1.07, 0.89)

Mean vector of species B: (5.66, 4.83, 1.04, 0.86)

Pooled variance vector : (1.9255, 3589, 0.0141, 0.0425)

Pooled correlation matrix

$$C = \begin{bmatrix} 1 & 0.3498 & 0.1757 & 0.4736 \\ & 1 & -0.0978 & 0.2860 \\ & & 1 & 0.5227 \\ & & & 1 \end{bmatrix}$$

$$C^{-1} = \begin{bmatrix} 1.3786 & -0.3140 & 0.0291 & 0.5783 \\ & 1.2703 & 0.4008 & -0.4239 \\ & & 1.5142 & -0.9193 \\ & & & 1.8753 \end{bmatrix}$$

i) Examine if the two species can be considered to be different with respect to the mean vector of these characters.

ii) Work out a formula to identify a new specimen of *Neesiella* into one of the two species A or B on the basis of a linear combination of observations on these four characters.

iii) Estimate the proportion of *Neesiella* species A that your method will misclassify.

(10+8+7+7)=32

Please turn over

2-(a) Microcephaly in human being is characterized by a distinctly small skull and brain, an underdeveloped body, and some newrophatic symptoms. The inheritance of this trait is not established. It has been suggested that the partial deletion of chromosome 18 may cause microcephaly. It has been found that the exposure of the pregnant mother to X - radiation can cause microcephaly in the foitus. Komai and others collected data from 77 families, each with two normal parents and in each of which there was at least one child with microcephaly. It has been suggested that microcephaly may be inherited as a simple autosomal recessive. Let θ be the segregation probability. Using the following data, test the hypothesis that microcephaly is simple recessive by testing for a suitable value of θ :

Family size	No. of families	No. of affected children γ: 1	2	3	4	5
		No. of families with γ affected children				
2	6	6	-	-	-	-
3	10	5	1	4	-	-
4	14	8	4	2	-	-
5	14	6	5	3	-	-
6	13	5	5	1	1	1
7	10	4	3	2	-	1
8	5	3	-	2	-	-
10	4	2	1	-	-	1
12	1	-	-	-	1	-

(b) Let π_1 and π_2 be two populations corresponding to African Negroes and Whites respectively. Also let $p_o = 0.630$ and $p_o^1 = 0.028$ be the frequencies in π_1 and π_2 respectively of the R^o allele in the Rhesus blood group system. Assume that the hybrid population π (American Negroes) is formed by gene flow from π_2 to π_1 (but not vice versa) at the constant rate $\alpha = 0.036$ per generation. Find the gene frequency in π of R^o for the fifth generation. (22+10)=32

3. An extract of adrenal tissue, of unknown pressor amine content, was assayed against a standard pressor amine by injecting into a cat and recording the transient rise of blood pressure. A 2 x 2 assay was used, with a high/low dose ratio of 3 for both standard (s) and unknown (U). Five observations were made at each dose of S and U, and the order of injections was randomized. From the data given below, calculate the potency of the extract (and its 95 per cent confidence limits using Fieller's Theorem or otherwise) in terms of the standard. Data are peak blood pressure increases in millimetres of mercury.

S_1 5μg	S_2 15μg	U_1 0.1 ml	U_2 0.3 ml
15	41	19	51
17	47	25	42
17	35	20	47
13	50	16	56
18	33	23	54

(32)

4. (a) The following table gives the results of a randomized block experiment showing the percentage of unsaleable ears of corn (out of 36 ears of corn in each plot), the treatments being a control A, and three mechanical methods of protecting against damage by corn earworm larvae. Analyse the data and write your conclusions.

Treatments	Block					
	1	2	3	4	5	6
A	42.4	34.3	24.1	39.5	55.5	49.1
B	33.3	33.3	5.0	26.3	30.2	28.6
C	8.5	21.9	6.2	16.0	13.5	15.4
D	16.6	19.3	16.6	2.1	11.1	11.1

(b) A number of experiments have indicated that electrical stimulation may be helpful in preventing the wasting away of muscles that are denervated. A factorial experiment on rat was conducted in order to learn something about the best type of current and the most effective method of treatment. The factors and levels were:

m : no. of treatment periods daily — 1 and 3

n : length of treatments (minutes) — 1 and 2

p : type of current — Galvanic and Feradic

The weights of denervated muscle in units of 0.01 grams for two groups of eight subjects each are given below (in usual notations for treatments):

Treatments	Group I	Group II
(1)	32	43
m	47	41
n	26	36
mn	61	76
p	29	39
mp	51	34
np	36	31
mnp	76	65

Each group is homogeneous with respect to normal muscle weight.

Analyse the data to find out the best type of current and the best method of treatment. (14+ 18)=32

NEATNESS (4)

Please turn over

GROUP D : DESIGN AND ANALYSIS OF EXPERIMENTS

Special Paper III - Practical

(Answer any three questions)

1. The plan and yields of a varietal trial is given below. Variety numbers are given in parentheses.

Block	Variety and yield			
1	(15) 24	(9) 25	(1) 26	(13) 20
2	(5) 27	(7) 28	([illegible]) 24	(1) 27
3	(10) 26	(1) 28	(14) 24	(2) 22
4	(15) 34	(11) 31	([illegible]) 31	(3) 38
5	(6) 41	(15) 33	(4) 33	(7) 29
6	(12) 34	(4) 32	(3) 28	(1) 30
7	(12) 32	(14) 25	(15) 24	(8) 26
8	(5) 23	(3) 23	(1[illegible]) 24	(5) 27
9	(5) 28	(4) 28	([illegible]) 26	(13) 25
10	(10) 25	(12) 27	(13) 28	(6) 26
11	(9) 26	(7) 26	(10) 23	(3) 24
12	(8) 27	(6) 27	(2) 25	(9) 26
13	(5) 30	(9) 36	(11) 32	(12) 32
14	(7) 30	(13) 28	(14) 24	(11) 25
15	(10) 24	(4) 25	(8) 32	(11) 31

(a) Identify the design and obtain its parameters.

(b) Carry out its intra-block analysis. (12+20)=3

2. A manurial experiment involving 3 levels of each of Nitrogen (N), Phosphorous (P) and Pottasium (K) fertilizers was conducted at Hiriyur, Mysore on sugarcane crop. The levels used for each of the fertilizers are coded by 0, 1 and 2. A 3^3 confounded design in two replications using blocks of size 9 plots was used for the experiment. The layout plan and the yield figures in Kg/plot of size 9 m x 7 m are given below:

Replication I

Block I N.P.K.	Yield	Block II	Yield	Block III	Yield
202	896	102	809	220	670
221	890	012	746	122	657
011	900	200	870	212	611
101	825	020	697	111	725
022	900	110	972	021	550
112	715	211	807	201	1187
210	570	121	1175	010	1010
120	840	222	610	002	850
000	748	001	643	100	625

Please turn over

2.
(Contd.)

Replication II

Block I		Block II		Block III	
102	845	100	428	101	656
011	465	012	807	010	550
110	480	111	795	112	801
201	505	202	480	200	523
022	440	020	336	021	940
212	621	210	340	211	1000
121	410	122	1050	120	915
220	905	221	1640	222	661
001	890	001	835	002	555

Find the interactions confounded. Analyse the data and interpret the results. Find the estimates of the linear and quadratic effects of N along with their S.E. (4+20+8)=32

3. In order to study the effect of Nitrogen (N) at three levels, deferment and no deferment of clipping in fall (F) and deferment and no deferment of clipping in Spring (S) on hay from alfalfa grass, an experiment was conducted. The levels of Nitrogen were coded as 0, 1 and 2 and those of each of F and S were coded by 0 and 1. The design adopted was a confounded asymmetrical factorial with 6 plot blocks in three replications. The yield figures in Kg/plot are shown below along with the treatment combinations.

Replication I				Replication II			
Block I N.F.S.	Yield	Block II	Yield	Block I	Yield	Block II	Yie[illegible]
001	294	000	249	000	232	001	26[illegible]
010	226	011	340	011	254	010	23[illegible]
100	403	101	523	101	523	100	41[illegible]
111	520	110	404	110	404	111	45[illegible]
200	370	201	457	200	411	201	485
211	487	210	481	211	436	210	399

Replication III

000	308	001	386
011	288	010	310
100	617	101	550
111	730	110	562
201	565	200	630
210	639	211	830

Find the confounded interactions. Analyse the data and interpret the results. Find the loss of information of the confounded interactions. (6+20+6)= 32

4. (a) Construct the BIB Design with parameters

$$v = b = 15 \qquad r = k = 7 \qquad \lambda = 3.$$

(b) Construct two mutually orthogonal latin squares of order 12. (16+16)=32

(4)

NEATNESS

Please turn over

GROUP E : SAMPLE SURVEYS

Special Paper III - Practical

(Answer Questions 1 and 5 and any two of the rest)

1. A list of thirty villages in a region and their area in square miles (to one decimal place) are given below.

Serial no.	Area	Serial no.	Area	Serial no.	Area
1	6.9	11	7.7	21	7.4
2	1.5	12	2.1	22	2.6
3	3.5	13	6.0	23	3.0
4	9.9	14	11.3	24	2.5
5	8.4	15	6.9	25	4.0
6	9.7	16	7.8	26	3.7
7	6.1	17	1.6	27	2.0
8	12.0	18	3.0	28	6.4
9	9.4	19	1.3	29	13.3
10	1.2	20	9.0	30	9.3

i) Draw a sample of five village with probability proportional to area with replacement by Lahiri's method of selection.

ii) Write down clearly the steps that you adopt for selection. Give reference to random number table used (title, page and commencing row etc.)

iii) Is the method adopted by you the least time consuming? If 'yes', give reasons. If 'no', which other step will be least time consuming and why?

(10+10+5)=2!

2. A sample of five villages is chosen with equal probability with replacement from a list of seventy villages of a region. In each of the five sampled villages, a list of households was prepared. In each sample village, a sample of four households was selected from the list with equal probability without replacement. From each sampled household, the area cultivated (in acres) by the household in 1974 and in 1975 were ascertained. Using the data given below,

i) estimate the area cultivated in 1975 by the ratio method of estimation;

ii) estimate the root mean square error of the estimate in (i).

iii) state the conditions for this ratio estimate to be more efficient than the unbiased estimate from the two stage sampling design adopted for this survey. Has this condition been satisfied in this survey?

Please turn over

2. (contd.)

Area cultivated (in acres)

Sample village number	Total no. of households	Sample household number 1: Area cultivated in 1974	1: 1975	2: Area cultivated in 1974	2: 1975	3: Area cultivated in 1974	3: 1975	4: Area cultivated in 1974	4: 1975
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	757	2.1	1.9	0.4	0.5	0.9	0.6	1.3	1.1
2	78	4.3	5.0	1.3	1.0	0.7	0.7	2.6	1.3
3	548	1.1	1.2	1.1	0.8	1.7	1.1	0.4	0.9
4	1064	0.2	0.6	0.7	0.5	0.5	0.9	0.4	0.7
5	875	1.3	0.3	6.3	5.0	8.5	10.1	1.8	2.8

Total cultivated area in the region in 1974 = 78000 acres.

(10+9+6)=25

3. It is proposed to conduct a survey of urban households in a city in 1976 for estimating the number of poor households (according to a specified definition of poverty). The city is devided into four strata. Each stratum is a group of 1971 census blocks. From each stratum a sample of four census blocks will be chosen with probability proportional to 1971 census population with replacement. A list of households will be prepared in each selected block. From each selected block, a sample of households will be chosen linear systematically with a random start, for detailed enquiry. The sampling design should be made self weighting. Making use of the data given below, work out the following :

i) Determine the raising factor with this self weighting design on an anticipated total sample size of 500 households.

ii) Determine the sampling interval to be used in each sample block with the self weighting raising factor arrived t in (i).

iii) What will be the expected number of households with the sampling interval determined in (ii).

Stratum	Sample block	Population in 1971	Number of households in 1976	Stratum	Sample block	Population in 1971	Number of households in 1976
1	1	3552	796	3	1	2554	549
(292000)	2	420	82	(345000)	2	506	102
	3	2819	576		3	7409	1528
	4	4892	1117		4	4873	1043
2	1	4040	917	4	1	3802	796
(471000)	2	3644	801	(475000)	2	977	208
	3	2305	568		3	2252	467
	4	5082	1034		4	5542	1119

Total population of strata in 1971 are shown in brackets below the stratum.

To estimate the total number of households in each stratum in 1976, assume that the population in 1976 will be more than in 1971 by 5 per cent on an average and that the average household size is 4.2.

(9+8+8)=25

Please turn over

4. A list of 20 factories in a region and data on two characteristics - number of workers (x) and output in thousand rupees (y) - are given below.

serial no.	x	y	serial no	x	y
1	51	1350	11	198	5562
2	67	3416	12	242	5582
3	74	3481	13	335	6315
4	80	3740	14	387	6719
5	85	3601	15	443	6660
6	97	3821	16	528	7295
7	116	4216	17	644	7288
8	127	4530	18	750	7610
9	144	5097	19	824	8063
10	173	5330	20	1095	9250

i) Draw five independent samples of four exclusive units each such that the probability of selecting any sample of four units is proportional to their total size, by Medzuno's method of selection. The 'size' of a factory is measured by the number of workers (x).

ii) From each of the five samples, work out the usual unbiased estimate of total output (Y) making use of the values of x and y for the sample units and of X (= population total of the character x = total of x for the 20 factories).

iii) Work out a single estimate of Y by combining suitably the five sample estimates of Y.

iv) Estimate the variance of the estimate in (iii) by the method of interprenetrating sub-samples. (6+6+5+8)=25

5. It is proposed to conduct a survey in the city of Calcutta to study conditions of the poor in the city.

i) Suggest six important tables to be generated from the survey data. Blank proforma of tables with proper title and headings for rows and clumns as necessary should be given.

ii) What items of data should be collected to build these tables? Give definition of terms used (which are not obvious).

iii) What reference period would you adopt for the different items and why?

iv) Prepare a schedule for enquiry by housing the items for data collection and all the necessary identification particulars in suitable places of schedule.
(8+5+4+8)=25

Please turn over

GROUP F : TECHNIQUES OF COMPUTATION

Special Paper III - Practical based on UR Machines.

(Attempt all questions from this group)

In the operation to be carried out you have to deal with two decks of cards (i) Village Card (Master) (ii) Households Card (Detail). There is one village card (Master) corresponding to a group of household cards (Detail). The information punched on these cards and the columns for each of the items of information is given on the card design given below

Card Designs

Village Card (Master)

Sr. no. (1)	item (2)	No. of cols. (3)	Card cols. (4)	Remarks (5)
1.	Card design index	4	1 - 4	Punch 1111
2.	Zone State Region	3	5 - 7	As coded
3.	Stratum	2	8 - 9	,,
4.	Sub Round	1	10	,,
5.	Sub Sample	1	11	,,
6.	Sample Village	1	12	,,
7.	Price of land per acre	6	13 - 18	Rs. only

Household Card (Detail)

Sr. no.	item	No. of cols.	Card cols.	Remarks
1.	Card design index	4	1 - 4	Punch 2222
2.	Zone State Region	3	5 - 7	As coded
3.	Stratum	2	8 - 9	,,
4.	Sub Round	1	10	,,
5.	Sub Sample	1	11	,,
6.	Sample village	1	12	,,
7.	Sample household no.	2	13 - 14	,,
8.	Household size	2	15 - 16	No.
9.	Land owned	5	17 - 21	0000.0

Step 1

You are given a deck each of (i) Household Cards (CDI 2222), (ii) Master Cards (CDI 1111). Sort in Zone State Region X Stratum X Sub Round X Sub Sample X Sample Village. Order separately for two decks. Write down the card counts for both the decks in the answer sheet. (15

Step 2

Take these two decks of Household Cards (CDI 2222) and Village Cards (CDI 1111) as sorted by you and match and merge on a Card Collator, the two decks into one deck taking into consideration the sorting order. Detail Cards should follow the Village Cards (Master) belonging to the same village . Select unmatched cards from both the decks. Mark selected cards as 01 and 02 respectively for Master Cards and Detail Cards. Write your Roll Number on these cards and hand it over to the examiner.

Please turn over

On the answer sheet write down the CDI, Zone State Region, Stratum, Sub Round, Sub Sample, Sample Village of the selected cards and also the count of merged cards.

Prepare wiring chart for the collator and hand it over to the examiner. (25)

Step 3

Take the merged deck of cards for previous step and make use of the Calculating Punch to obtain for each household, total value of land = Price of land per acre X land owned and punch the result on column 22 to 33 (Rs. Ps.). After the end of the operation, separate Master Cards from Detail Cards without disturbing the order of the sets.

Mark the Master Card as 03 and Detail Cards as 04 and hand over the master cards to the examiner with your roll number written on the front card.

On the answer sheet write down the cards counts of both the sets.

Prepare a wiring chart for the Calculating Punch and hand it over to the examiner.

Step 4

Take the set of Household Cards (Detail) from previous step (i.e. deck 04) and with the help of accounting machine, prepare a statement showing the following information as per layout below.

Zone	State	Region	Total household size	Total value of land
			(CC 15 - 16)	(CC 22 - 33)
1	1	1		
1	1	2		
Sub-total by Zone State				
1	2	1		
1	2	2		
Sub-total				

(25)

List the various types of operations, tabulations and/or calculations - which each of the major machines of ICL, IBM or Power-samas, machine system can perform.

(Time: half-hour) (15)

Please turn over

GROUP G : STATISTICAL INFERENCE

Special Paper III - Practical

(Answer all questions)

1. The standing height in inches of 15 university students are

68, 65, 66, 67, 68, 64, 69, 66, 70, 71, 67, 74, 72, 65, 7(.

In the general population the variation in standing height is 2.55 inches and the distribution is normal. Is the variation in student's height different from the variation in the general population? Find the alternative at which the power of the test that you use will be .8.
(6+10)=16

2. We have data on the weight of mice 120-130 days old. The data available consists of 10 litters, each litter containing 1 male and 1 female mouse. The weight of brothers and sisters is given below in grammes.

Brother's weight	7.5	12	10.5	14	9.7	14.3	16	6.4	11.3	12.4
Sister's weight	10	8	9	9.5	10.2	9.3	8	8.4	11.8	11.9

Is there any difference in the weight of brothers and sisters at 120-130 days old? Assuming normal distribution, state your null hypothesis carefully. Find the alternatives for which the power of your test is .7, .8, .9. (6+10)=16

3. The following quantities are computed from data available on the length of the right wings of queens, drones and workers of Vespa Vulgares, all caught in the same locality. The measurements are in 0.01 mm.

129 Queens	Mean length of right wing	=	920.5
	Standard deviation	=	9.64
136 Drones	Mean length of right wing	=	848.1
	Standard deviation	=	19.52
183 Workers	Mean length of right wing	=	725.4
	Standard deviation	=	17.04

Are the differences in mean length due to chance only? Draw the 95 per cent confidence ellipsoid of the difference of mean length of queens from that of drones and workers. (10+15)=25

Please turn over

4. $\{X_i\}$ is a sequence of independent identically distributed random variables; the common distribution is normal with unknown mean μ and variance $\sigma^2 = 4$.

(a) Write down the UMP test of size $\alpha = .01$ of H_o $(\mu = 0)$ Vs H_1 $(\mu > 0)$ based on a sample of size 100. Find the values μ_1, μ_2 of μ at which the power is .9 and .5.

(b) Let $Z_i = 1$ if $X_i > 0$
$= 0$ if $X_i \leq 0$

If your observations consist of n values of Z_i's what will be your UMP test of size $\alpha = .01$ of $H_o(\mu = 0)$ Vs H_1 $(\mu > 0)$? Choose n such that the power of your test is .9 at $\mu = \mu_1$. After choosing n in this way calculate the power of your test at $\mu = \mu_2$.

(c) Choose a sequential probability ratio test of H_o $(\mu = u)$ Vs. H_1 $(\mu = \mu_1)$ such that $\alpha = .01$ and $\beta = .1$. Calculate the power of this test for $\mu = \mu_2$. Calculate the ASN for $\mu = 0, \mu_1, \mu_2$.

(d) $X_1, \ldots X_{10}$ were observed and the variance was found to be 4.6. How many more observations are needed to get a 95 per cent confidence interval for μ of length .5 ? (5+15+18+5)=43

GROUP H : PROBABILITY THEORY

Special Paper III - Special Topics

(Answer any five questions)

1. (a) Define a stationary stochastic process with independent increments.

(b) An orbiting satellite expends W_i watts of energy in transmitting data, essentially instantaneously, at time t_i when it passes over its control station for the ith time. The infinite sequence 0 t_1 t_2 ... of transmission times is fixed in advance. Let

$S_t = \sum_{t_i \leq t} W_i$ Assume $W_1, W_2, \ldots$ are independent random variables.

i) Does the process S_y have independent increments?

ii) Does S_t have stationary increments ?

iii) Does S_t have stationary increments, if we assume that $t_i = i$ for $i = 1, 2, \ldots$ and $W_1, W_2, \ldots$ are identically distributed?

Please turn over

1. (c) (Contd.) Let $X(t)$ be a process with stationary, independent increments. Let $\mathcal{B}$ denote the Borel σ-algebra on R (real line). Show that for $B \in \mathcal{B}$

$$P\,(X\,(t+\tau) \in B \,/\, X(S),\ S \leq t) = P(X\,(t+\tau) \in B/X(t))\text{a.}$$

(2+12+6)=20

2. (a) Let $X(t)$ be a normalized Brownian motion with $0 \leq t < \infty$. Show that, for any finite collection T_0 of points $0 = t_0 < t_1 < \ldots < t_n = \tau$ and any $x > 0$,

i) $\text{Prob}(\max_{t \in T_0} X(t) > x) \leq 2\ \text{Prob}\ (X(\tau) > x)$

ii) $\text{Prob}\ (\max_{t \in T_0} |X(t)| > x) \leq 2\ \text{Prob}\ (|X(\tau)| > x)$

(b) Let $X(t)$ be a normalized Brownian motion with $0 \leq t < \infty$. Prove that, as $t \to \infty$,

$$\frac{X(t)}{t} \to 0 \quad \text{a.s.}$$

(8+12)=20

3. (a) Let $X_0, X_1, X_2, \ldots$ be a countable Markov chain with state space the set of positive integers. Suppose the process starts at j. Let $G = \bigcup_{n \geq 1} \{X_n = j\}$.

Show

i) $P\,(G) = 1 \Rightarrow \begin{cases} P(X_n = j \text{ infinitely often}) = 1 \\ \sum_{n \geq 1} P(X_n = j) = \infty \end{cases}$

ii) $P\,(G) < 1 \Rightarrow \begin{cases} P\,(X_n = j, \text{ infinitely often}) = 0 \\ \sum_{n \geq 1} P(X_n = j) < \infty \end{cases}$

(b) Use (a) to show that, if $\{Z_n\}$ are the successive fortunes in a fair coin tossing game,

$$P\,(Z_n = 0 \text{ infinitely often}) = 1$$

(A fair coin is a coin with P (Head) = P (Tail)= 1/2)

(12+8)=20

4. (a) A message consists of a sequence of letters A, B and C, whose probabilities of occurrence do not depend on the preceeding combination of letters and they are given by P (A) = 0.7, P(B) = 0.2 and P(C) = 0.1. Perform the coding by the method of Shannon-Fano for separate letters and blocks consisting of two letter combinations and find their respective efficiencies.

Please turn over

4. (b) (Contd.) There are 12 coins of equal value; however one coin is counterfeit. The counterfeit coin is known to weighless than the others by 0.25 gms. Show that, by using a chemical balance with no weights, at least 3 weighings are necessary to identify the counterfeit coin.

(Give an information theoretic argument.)

(c) The probabilities for a signal to be received or not received are 2/3 and 1/3. As a result of noise, a signal entering through the receiver can be recorded at its output with probability 3/4 and not recorded with probability 1/4. In the absence of a signal at the input, it can be recorded at the output with probability 1/2 and not recorded with probability 1/2. Determine the quantity of information about the presence of the signal at the input one could obtain by observing at the output.

(7+6+7)=20

5. A gambler with initial capital of Rs.12 decides to play against an adversary with a capital of Rs.1000 until one of them is ruined. Assume that the probability of the gambler winning a bet is 0.52 in which case he gains Re.1. If he does not win a bet he loses Re.1.

i) What is the probability of the gamblers ultimate victory?

ii) Determine the gambler's ultimate gain or loss.

iii) What is the expected duration of the game? (6+7+7)=20

6. State and Prove Shannon's first coding theorem for the ergodic source and a stationery channel with finite memory. (You may assume Mcmillan's theorem and Feinsteins fundamental lemma.) (20)

7. (a) Calculate the capacity of a symmetric binary channel whose channel matrix is

$$\begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix}$$

(b) Let A and B be two binary symmetric channels, with capacities C_A and C_B respectively. If C is the capacity of the cascade of A and B, show that $C \leq \min(C_A, C_B)$

(c) Let [A, B, M] be an information channel with input alphabets $A = \{a_1, \dots, a_r\}$, output alphabets $B = \{b_1, \dots b_s\}$ and channel matrix $M = ((m_{r,s}))$. Let $P = (p_1, \dots, p_r)$ and $Q = (q_1, \dots q_r)$ be two input probabilities and $R = \lambda P + (1-\lambda) Q$ for some $0 < \lambda < 1$. Show

$$\lambda I_P(A; B) + (1-\lambda) I_Q(A; B) \leq I_R(A; B)$$

(6+6+8)=20