INDIAN STATISTICAL INSTITUTE 10

QUESTION PAPERS

for

STATISTICIAN'S DIPLOMA EXAMINATION

November 1977

203 BARRACKPORE TRUNK ROAD
CALCUTTA 700035

Price : Rupees Two only

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1977

Paper I (Theoretical) : Official Statistics and Descriptive Statistics

Time : 4 hours — Full marks : 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is not permitted.

GROUP A: Official Statistics (50 marks)

(From this Group, answer Question 1 and any two questions from the rest)

1. Describe the statistical system in India, bringing out the role of the Central Statistical Organisation. (18)

2. Write an explanatory note on the present availability of agricultural statistics in India, mentioning clearly the major gaps and the steps being taken for filling these. Comment on the reliability of the available statistics. (16)

3. Describe the salient features of the 1971 Population Census in India. How do you regard it as an improvement over the 1961 Census ? Mention the various uses of the census data. (16)

4. State briefly the official data collected and compiled, mentioning inter alia the agency responsible and a principal publication, its scope and periodicity, for any two of the following :

   i) Foreign trade statistics

   ii) Road transport statistics

   iii) Health statistics

   iv) Labour statistics (8+8)=16

5. Describe the functions and activities of the National Sample Survey Organisation with a brief note on the development of the sampling design and the use of statistical data supplied by this organisation. (16)

GROUP B: Descriptive Statistics (50 marks)

(Answer any three questions from this group)

6. (a) Discuss the various methods that may be used for collecting data in a statistical enquiry.

   (b) Write notes on ratio chart and pie chart. (8+8)=16

7. (a) What is a weighted average ? Write a short note on the use of this type of average in the construction of price indices.

   (b) A certain variable x takes the values 1, 2, ..., k with certain frequencies $f_1, f_2, \ldots, f_k$. If $F_i' = \sum_{j=i}^{k} f_j$ (cumulative frequency of the greater than type for the value i), show that $\bar{x} = \sum_{i=1}^{k} F_i'/n$, where $n = \sum_{i=1}^{k} f_i$. Also, obtain the corresponding formula for s^2. (5+5+6)=16

Please turn over

8. (a) What is meant by the dispersion of a frequency distribution? Compare the range and the standard deviation as measures of dispersion.

(b) What is Gini's mean difference and what is its merit as a measure of dispersion? Show that the standard deviation may be expressed in a comparable form and has the same merit.

(2+6+4+4)=16

9. (a) What are theoretical distributions and what is their utility ?

(b) Show that the Poisson distribution may be looked upon as an approximation to a binomial distribution. Also, obtain the Poisson distribution more directly by starting from a suitable probability model. Hence cite some real-life situations where this distribution would be appropriate.

(4+5+5+2)=16

10. (a) Explain fully why the correlation coefficient may be looked upon as the proper measure of association for two jointly normally distributed variables.

(b) Show, by deriving the necessary formulae, how the true correlation coefficient between two variables x and y may be affected by

i) heterogeneity of data,

ii) observational errors in the values of x and y. (6+5+5)=16

11. (a) What is meant by a multiple regression equation and what is its utility ? Write down the linear multiple regression equation of x_1 on $x_2, x_3, \ldots, x_p$ in terms of the means, standard deviations and correlation coefficients of these (jointly distributed) variables.

(b) Define the multiple correlation coefficient $r_{1.23 \ldots p}$ and the partial correlation coefficient $r_{12.34 \ldots p}$.

Show that

$$r^2_{1.23 \ldots p} = 1 - \text{var}(x_{1.23 \ldots p})/\text{var}(x_1)$$

and

$$r^2_{1p.23 \ldots (p-1)} = 1 - \text{var}(x_{1.23 \ldots p})/\text{var}(x_{1.23 \ldots (p-1)})$$

and interpret these results.

(3+3+8+2)=16

NEATNESS (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1977

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) Define a discrete random variable and its probability mass function.

   (b) Consider the following distribution function of a random variable :

$$F(x) = \begin{cases} 0 & , \ x < 0 \\ 1/4 & , \ 0 \le x < 3 \\ 1/2 & , \ 3 \le x < 4 \\ 1 & , \ x \ge 4 \end{cases}$$

   Represent the distribution function graphically. Obtain the mass function of the associated random variable.

   (c) A mathematician always carries two match boxes, each containing n matches. Whenever he needs, he choses a box at random and draws a match from it. Find the probability that when the first box is found to be empty for the first time, the second box will contain exactly r matches. (4+6+6)=16

2. (a) Show that $f(x)$, given by

$$f(x) = \begin{cases} x & , \ 0 < x < 1 \\ k-x, & , \ 1 < x < 2 \\ 0 & , \ \text{elsewhere,} \end{cases}$$

   is a (probability) density function for a suitably chosen value of k. Calculate the probability that the random variable lies between $\frac{1}{2}$ and $\frac{3}{2}$.

   (b) Calculate Var (X) for the above random variable.

   (c) If X has Normal distribution with mean −2 and variance 4, write down the density function of $2X + 5$. If X_1, X_2 are independent standard normal deviates, what is the distribution of $\frac{X_1 + X_2}{2}$? (6+4+6)=16

3. (a) The joint density function of X and Y is given by

$$f(x,y) = (6-x-y)/8, \quad 0 < x < 2, \quad 2 < y < 4$$
$$= 0 \quad , \quad \text{elsewhere.}$$

   i) Find the two marginal densities and the conditional distribution of Y for given X. Are the two variables independent. ?

   ii) Find the mean of X+Y.

   (b) If X_1 and X_2 are two independent Poisson variables with means m_1 and m_2 respectively, find the distribution of $X_1 + X_2$. (6+4+6)=16

Please turn over

4. (a) What is a probability-generating function ? Show that the probability-generating function of the sum of two independent random variables is the product of their probability-generating functions.

(b) It is known th has the moment-generating function $M(t) = (0.3e^{t} + 0.7)^{10}$. What is the moment-generating function of $Y = 4 + 7X$?

(c) Let X be a normal variable with mean 0 and $Y = X^2$. Find the moment-generating function of Y. (6+4+6)=16

5. (a) State and prove Chebycheff's inequality.

(b) Let $X_1, X_2, \ldots$ be a sequence of random variables having expectations $\mu_1, \mu_2, \ldots$, then prove that

$$\bar{X}_n - \bar{\mu}_n \xrightarrow{P} 0, \text{ if } \frac{\operatorname{Var}(X_1+X_2+\ldots+X_n)}{n^2} \to 0$$

as $n \to \infty$.

(c) State clearly the conditions which a binomial distribution tends to (i) a Poisson dist No. and to (ii) a normal distribution. (6+6+4)=16

GROUP B : Statistical Methods (50 marks)

(Answer any three questions from this group)

6. (a) Let $x_1, x_2, \ldots, x_n$ be n independent observations from a normal distribution with mean μ_1 and s.d. σ_1 and $y_1, y_2, \ldots, y_m$ be another set of m independent observations from a normal distribution with mean μ_2 and s.d. σ_2. Find unbiased estimators of σ_1^2 and σ_2^2 and 95 percent confidence limits to σ_1^2/σ_2^2.

(b) For the problem given in (a) above, find maximum likelihood estimators of σ_1^2 and σ_2^2. Examine if $s = \sqrt{(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)/(n-1)}$ is an unbiased estimator of σ_1 ? (4+4+4+4)=16

7. (a) Derive a most powerful similar region test for testing $H_0: \sigma = \sigma_0$ against the alternative $H_1: \sigma = \sigma_1 (\sigma_1 > \sigma_0)$, when n independent observations are drawn from the normal distribution with mean 0 and s.d. σ. Is this test consistent ?

(b) Describe a large-sample test for testing goodness of fit when the hypothetical cell probabilities are functions of unknown parameters. (6+2+8)=16

Please turn over

8. (a) Let $x_1, x_2, \ldots x_n$ be n independent observations from a distribution with density function

$$f(x) = \alpha \exp\{-\alpha(x-\mu) - \exp[-\alpha(x-\mu)]\} .$$

Show that the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\mu}$ are given by

$$\frac{1}{\hat{\alpha}} = \bar{x} - \frac{\sum_{i=1}^{n} x_i e^{-\hat{\alpha}x_i}}{\sum_{i=1}^{n} e^{-\hat{\alpha}x_i}}$$

and

$$e^{-\hat{\alpha}\hat{\mu}} = \frac{1}{n}\sum_{i=1}^{n} e^{-\hat{\alpha}x_i} .$$

Find the asymptotic variance of $\hat{\alpha}$.

(b) If $x_1, x_2, \ldots, x_n$ are n independent normal variates with $E(x_r) = r\theta$ and $V(x_r) = r^3 \sigma^2$, show that the maximum likelihood estimator of θ is given by

$$\hat{\theta} = \left(\sum_{r=1}^{n} x_r/r^2\right)/\left(\sum_{r=1}^{n} 1/r\right).$$

(6+4+6)=16

9. (a) Let $x_1, x_2, \ldots,$ be independent random variables with common probability density function $f(x)$, where

$$f(x) = \begin{cases} \lambda \exp[-\lambda x] & \text{if } x \geq 0 \\ 0 & \text{Otherwise,} \end{cases}$$

λ being an unknown positive real number. Find a sequential test for testing the hypothesis $H_0: \lambda = \lambda_0$ against the alternative $H_1: \lambda = \lambda_1$.

(b) Find the O-C function of a SPR test for testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1 (\theta_1 > \theta_0)$, where θ is the parameter of a Poisson distribution.

(8+8)=16

10. Write notes on any three of the following :

(a) Invariant tests;

(b) Sufficient statistic;

(c) Rao-Cramer inequality;

(d) Mann-Whitney U-test.

(4X4)=16

NEATNESS (for Groups A and B) (4)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1977

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. (a) Distinguish clearly between sampling and non-sampling errors in sample surveys.

   (b) Describe the prevalent methods for controlling these errors.

2. (a) What do you mean by proportional and optimum allocations in stratified sampling ?

   (b) Consider the sampling strategies

   A ≡ ((Simple random sampling without replacement); usual unbiased estimator)

   and

   B ≡ (Stratified sampling with proportional allocation (without replacement); usual unbiased estimator)

   for the population mean of a character. Obtain the relative efficiency of B over A and mention under what situations one is to be preferred over the other. (4+12)=16

3. (a) What do you mean by double sampling and when do you need it?

   (b) Give a suitable sampling strategy for estimating the population mean of a character y, when the population mean of an auxiliary character x is not known but can be estimated. Compare this sampling strategy with that in which no auxiliary information is used. (4+12)=16

4. (a) Describe, in brief, various ways in which an auxiliary information can be utilized to reduce the sampling error.

   (b) A sample of n units is drawn with probabilities $p_i > 0 \left(\sum_{1}^{N} p_i = 1\right)$ and with replacement out of N units in the population and characters y and x are observed. Consider an estimator

$$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i/p_i - \lambda\left(\frac{1}{n}\sum_{i=1}^{n} x_i/p_i - X\right)$$

   where λ is a constant and X is the population total of x.

   i) Show that $\hat{Y}$ is unbiased for estimating the population total of y.

   ii) Obtain an expression for variance of $\hat{Y}$. (4+4+8)=16

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

   i) Circular systematic sampling.

   ii) Cluster sampling.

   iii) Sampling on successive occasions.

   iv) Indian National Sample Surveys. (8+8)=16

Please turn over

GROUP B: Design & Analysis of Experiments (5 marks)

(Answer any three questions from this group)

6. (a) What is 'Statistical Design of Experiments'? What is 'Statistical Inference'? Are they related ?

(b) Explain the principles of randomisation, replication and local control.

(c) Name a design where all the principles in (b) are used.

(d) What is a 'Uniformity trial' ? (3+9+2+2)=16

7. (a) Define a Balanced Incomplete Block Design with parameters, v, b, r, k and λ . Establish the relations among the parameters. Why is such a design balanced ?

(b) State the linear model assumed for analysing the design in (a) and write down the analysis of variance table. (2+6+2+2+4)=16

8. (a) Explain the concepts of complete and partial confounding in factorial experiments.

(b) Consider a 2^4 factorial experiment with four factors, A, B, C and D each at two levels.

i) Define the factorial effects.

ii) Give the layout of a 2^4 factorial experiment in a single replicate, where the factorial effect ABCD is to be confounded. (4+6+6)=16

9. (a) Discuss the role of concomitant variates in design of experiments.

(b) Develop the technique of analysis of covariance in detail. (4+12)=16

10. Write short notes on the following :

(a) Shape and size of plots.

(b) Missing-plot technique. (8+8)=16

NEATNESS (for Groups A and B) (1)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1977

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. What is an index number ? Discuss the various problems that arise in the construction of a price index number. (2+14)=16

2. What is meant by the secular trend of a time series? Describe the different methods of determining the trend from a time series. Comment on their relative advantages and disadvantages. (2+10+4)=16

3. (a) Define the lognormal distribution. Discuss its appropriateness for describing the distribution of expenditure in a country by expenditure classes.

   (b) Give a short description of an input-output table and indicate the possible uses of such a table. (2+6+5+3)=16

4. What is a production function ? How does it enable one to verify statistically the "laws of return" ? Illustrate by reference to the Cobb-Douglus production function. (2+10+4)=16

5. Write short notes on the following :

   i) National income at factor costs and market prices.

   ii) Commonly used Engel curve forms and their properties. (8+8)=16

NEATNESS (2)

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

(Answer any three questions from this group)

1. (a) What is a control chart ? What evidences of a lack of control in a manufacturing process does a control chart provide ?

   (b) Give the expression for C_2 as used on an s-chart. Obtain the formulae for D_3 and D_4 for an R-chart. (6+10)=16

Please turn over

2. (a) Explain the terms LTPD and AOQL in connection with an acceptance sampling plan. When do you prefer an AOQL plan to an LTPD plan ?

(b) Describe a graphical procedure for carrying out a sequential sampling plan by attributes. What is the minimum number of items that must be inspected before one can decide to accept a lot on the basis of such a plan ? (6+10)=16

3. (a) Distinguish between <u>specification limits</u>, <u>tolerance limits</u> and <u>confidence limits.</u>

(b) How do you set up tolerance limits for a distribution of the item quality measure (i) that is known to be normal with a known S.D. and (ii) that is unknown. (6+10)=16

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

(a) Process capability.

(b) Continuous sampling plans.

(c) Military Standard 105D.

(d) Rotatable designs. (16)

NEATNESS (2)

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

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

1. (a) Explain the terms

i) genotypic frequency and

ii) gene frequency

(b) State and prove the Hardy-Weinberg Law of equilibrium for a random breeding population.

(c) A stock of plants initially consisted of a proportion p_0 of the type AA, a proportion q_0 of the type Aa and a proportion r_0 of the type aa, $(p_0+q_0+r_0 = 1)$. With continued self fertilisation, what will these proportions be after n generations ? (25)

2. Give an account of the different methods of estimation of Linkage. Describe the nature of data you will require for each method. (25)

3. What are fixable and non-fixable components of heritable variation for quantitative character ? Describe a procedure for estimating them, illustrating the nature of data required. (25)

<u>Please turn over</u>

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

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

1. (a) Distinguish between a stationary population and a stable population.

   (b) Give expressions for a stable population birth rate and a stable age distribution.

   (c) Describe a method of finding out the true rate of natural increase (r) in a stable population. (5+4+7)=16

2. (a) Explain the different columns in a complete life table.

   (b) Derive the interrelationships between force of mortality(μ_x), central rate of mortality (m_x) and probability of dying (q_x). (8+8)=16

3. (a) Derive the law of population growth in the form :

$$P_t = \frac{L}{1 + e^{(\delta - t)/\alpha}},$$ the symbols

   having their usual significance.

   (b) Would this law be suitable for population data of India ? Give reasons in support of your answer. (10+6)=16

4. Write notes on :

   (a) Net reproduction rate.

   (b) Indirect method of standardisation of death rate.

   (c) Measures of morbidity. (16)

NEATNESS (2)

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

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

1. (a) Describe any two methods of scaling for combining and comparing examination scores in a number of scholastic tests.

   (b) Describe Likert's method of scaling, stating the asumptions involved. (10+6)=16

2. (a) Describe any two methods of experimentally establishing the validity of a test of arithmetic reasoning.

   (b) Describe the split-half method of estimating the test reliability and state its drawbacks.

   (c) Let $X = y_1 + y_2 + \dots + y_n$ be a composite measurement. Let $\rho_{XX'}$ be the reliability of X and $\rho_{jj'}$, j = 1, 2, ..., n, be the reliabilities of the components. Let σ_j^2 be the variance of the jth component and let σ_X^2 be the variance of X. Show that

$$\rho_{XX'} = 1 - \frac{\Sigma\sigma_j^2 - \Sigma\rho_{jj'}\sigma_j^2}{\sigma_X^2}$$ (6+5+5)=16

<u>Please turn over</u>

3. (a) Discuss Thurstone's multiple-factor theory in factor analysis.

   (b) Describe any two methods for the estimation of factor loadings. (6+10)=16

4. (a) Assuming the classical test theory model, define true score, error of measurement and standard error of measurement.

   (b) Obtain the formula for finding the reliability of a test if it is lengthened K times. State the assumptions involved, if any. (6+10)=16

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

   (a) The use of fractile analysis in psychology.

   (b) Classificatory techniques in personnel selection.

   (c) Weighting procedure to maximize the reliability of a composite test.

   (d) Stopping rule in factor analysis.

   (e) Flanagan technique of item analysis.

   (f) Speed versus Power tests (16)

NEATNESS (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1977

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 (25 marks)

(Answer any two questions from this group)

1. Compute the inverse of the following matrix :

$$\begin{bmatrix} 1 & 7 & 10 & 3 \\ 2 & 19 & 27 & 8 \\ 0 & 12 & 17 & 4 \\ 5 & 2 & 6 & 3 \end{bmatrix}$$

($12\frac{1}{2}$)

2. Using Cramer's rule, solve the following system of equations :

$$\begin{aligned} x + 2y - 12z + 8v &= 27 \\ 5x + 4y + 7z - 2v &= 4 \\ -3x + 7y + 9z + 5v &= 11 \\ 6x - 12y - 8z + 3v &= 49 \end{aligned}$$

($12\frac{1}{2}$)

3. Compute the value of π, correct to 7 places of decimals, from the formula

$$\frac{\pi}{4} = \int_0^1 \frac{dx}{1+x^2}$$ by means of a

suitable quadrature formula, using intervals of 0.1 for x. ($12\frac{1}{2}$)

GROUP B : Descriptive Statistics (50 marks)

(Answer all questions from this group)

4. For a certain frequency distribution concerning a continuous variable, the mean and the moments about the mean are as follows :

$\bar{x} = 108.48$ $m_3 = 99.82$

$m_2 = 297.75$ $m_4 = 375,193.0$

(a) Decide upon a type of Pearsonian curve that would be appropriate for this frequency distribution.

(b) Write down the equation to the particular curve of the chosen type that may be supposed to give the best fit to the data. Find the probability that x lies between 100 and 110. (5+8+4)=17

Please turn over

5. (a) A certain telephone switch-board handles, on an average, 750 calls during a rush hour. The board is capable of making 16 connections per minute. Find the probability that the board will be over taxed during a one-minute period in the rush hour. (Clearly state the assumption you have to make.)

(b) A city education authority carried out a survey of the IQs of the children in its schools, testing 150,000 children. The distribution of IQs was found to be normal with a mean of 100, but 3.5% of the children were found to be educationally sub-normal with IQs less than 75. What is the standard deviation of this distribution? (6+6)=12

6. The following data relate to the percentage of nitrogen (x_1), percentage of chlorine (x_2) and log of leaf burn in seconds (y) for 30 samples of tobacco taken from farmers' fields :

$$\Sigma x_1 = 98.36 \qquad \Sigma x_1^2 = 332.34$$
$$\Sigma x_2 = 24.23 \qquad \Sigma x_2^2 = 30.10$$
$$\Sigma y = 22.55 \qquad \Sigma y^2 = 22.81$$
$$\Sigma x_1 x_2 = 81.38 \qquad \Sigma x_2 y = 12.41$$
$$\Sigma x_1 y = 31.65$$

Find the linear regression equation of y on x_1 and x_2, and give an idea of the utility of the equation as a predicting formula for y. (10+3)=13

7. The prices of the principal food articles in a town of Eastern India during the year 1961 and during September 1971 are shown in the table below, together with the average monthly expenditure per household on each of these articles in 1961:

Items	Average expenditure per household in 1961 (Rs.)	Prices per Kg. in Rs.	
		1961	September 1971
Cereals	53.27	0.55	1.25
Pulses	7.11	0.62	1.00
Edible oils	4.[illegible]7	3.10	4.20
Vegetables	2.19	0.52	0.68
Milk	4.85	1.25	2.00
Meat and fish	1.41	4.00	5.92
Fruits	0.34	3.20	3.70
Salt	0.56	0.10	[illegible].15
Spices	2.25	3.10	3.00
Sugar	1.18	1.20	2.30

Compute a suitable index to show the change in food-price-level between the year 1961 and September 1971. (8)

Please turn over

GROUP C : Official Statistics (25 marks)

(Answer both the questions from this group)

8. Name the publications in respect of any three of the following series of data. Give also the names of the organisations issuing them and the periodicity of each publication.

i) Index number of agricultural production in India.

ii) Statistics relating to area and outturn of forest produce in India.

iii) Imports and exports of principal articles to and from India.

iv) Production, despatches, stocks and distribution of coal in India.

v) Statistics of production of principal crops for different countries of the World.

vi) Gross and net national product in India. (3X3)=9

9. From the official publications supplied to you, collect annual data pertaining to the recent five years for the Index number of consumer prices, separately for

(a) Industrial workers and

(b) Urban non-manual employees,

in respect of the following four centres :

(i) Bombay, (ii) Calcutta, (iii) Delhi and (iv) Madras.

Present the data in a neat table as well as graphically. Write a critical note on the salient features of the data. (7+5+4)=16

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1977

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) To test the hypothesis H_0: $p = \frac{1}{2}$ against the alternative hypothesis H_1: $p = \frac{3}{4}$, where p is the probability of getting a head in a throw of a coin, the coin is tossed 10 times. The decision rule is to accept H_0 if the observed number of heads is less than 8, otherwise to accept H_1. Find the level and the power of the decision rule.

(b) The following table gives the observed phenotypic frequencies for hapto-globin distribution for two different populations :

	phenotypes			total
	1 – 1	2 – 1	2 – 2	
population I	18	56	38	112.
population II	8	37	64	109
total	26	93	102	221
probability	p^2	$2p(1-p)$	$(1-p)^2$	1

Find the maximum likelihood estimates of the gene-frequencies (p) for the two populations and find their asymptotic standard error. Also test the hypothesis of equal gene-frequencies ($p_1 = p_2$).

(3+3+5+5+4)=20

2. (a) A group of 12 persons of age 26 years were treated with a tonic for one year to find whether the tonic has any effect on the increase of weight. Another group of 10 persons of same age were also observed (without tonic) and their increase in weight are also given below.

persons	increase in weight in Kg.
with tonic	5.5, 3.5, 2.5, 7.0, 6.0, [illegible], 3.0, 3.5, –0.5, 5.0, 4.0, 2.0
without tonic	2.0, 3.0, –1.0, 0.0, [illegible], 3.0, 1.0, 0.5, –1.5, 1.0

Use any non-parametric test to see whether the tonic leads to any increase in weight or not.

(b) The correlation coefficient between the height and weight of 27 students was computed to be 0.55. Find 95% confidence limits to the population correlation coefficient. Test also the hypothesis of no correlation.

(c) Describe a sequential probability ratio test (SPRT) for testing H_0: $\mu = 10$ against H_1: $\mu = 20$, when observations are drawn from a normal distribution with mean μ and s.d. 2.0, $\alpha = 0.01$ and $\beta = 0.05$.

(5+5+4+6)=20

Please turn over

3. The following table gives the total yields of paddy crop last year in quintals (Y), total agricultural area in hectares (i) irrigated (X_1) and (ii) non-irrigated (X_2), separately for 1[illegible] different villages.

Village no.	Yield of paddy in 100 Kg.(Y)	Total agricultural area in hectares	
		Irrigated (X_1)	Non-irrigated (X_2)
1	88.2	1[illegible].4	1[illegible].3
2	111.5	5.2	3[illegible].8
3	13[illegible].3	[illegible]5.7	17.9
4	255.8	3[illegible].6	4[illegible].1
5	95.7	9.5	17.[illegible]
6	125.1	11.2	23.1
7	168.6	13.8	35.6
8	259.9	25.7	[illegible]1.6
9	1[illegible].8	1[illegible].2	21.5
10	213.6	10.8	35.3

(a) Fit a linear regression $Y = \alpha + \beta_1 X_1 + \beta_2 X_2$.

(b) Test the hypothesis H_0 : α = [illegible].

(c) Set 95% confidence limits to β_1 and β_2. (8+4+8)=20

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

(Answer all question from this group)

4. Consider a randomised block experiment with 4 treatments (1, 2, 3, 4) and 5 blocks. The responses are shown below.

B_1	1	4	2	3
	5[illegible]	45	35	[illegible]
B_2	3	2	1	4
	5[illegible]	38	45	46
B_3	1	1	2	3
	48	46	36	52
B_4	1	2	[illegible]	3
	48	[illegible]	44	54
B_5	2	4	1	3
	39	40	49	52

(a) Analyse the randomised block experiments and test whether all the treatment effects are the same. Also make comments on your findings.

(b) Suppose in the block B_1, the observation corresponding to the treatment 1 is missing. Find an estimate of the missing observation.

(c) Write down the three orthogonal treatment effect contrasts. Calculate the sum of squares due to the above treatment effect contrasts. Check that the total sum of squares of treatment contrasts is the same as the treatment SS calculated in (a). (1[illegible]+1[illegible]+8)=3[illegible]

Please turn over

GROUP C : Sample surveys (30 marks)

(Answer all questions from this group)

5. Two dentists, A and B, make a survey of the state of the teeth of 200 children in a village. Dr. A selects a simple random sample of 20 children and counts the number of decayed teeth for each child, with the following results :

Number of decayed teeth/child	0	1	2	3	4	5	6	7	8	9	10
Number of children	8	4	2	2	1	1	0	0	0	1	1

Dr. B, using the same dental techniques, examines all 200 children, recording merely those who have no decayed teeth. He finds 60 children with no decayed teeth.

Estimate the total number of decayed teeth of 200 children in the village

(a) using A's results only;
(b) using both A's and B's results
(c) which estimate do you expect to be more precise? (10)

6. The following data show the stratification of all farms in a county by farm size and the average acres of corn (maize) per farm in each stratum:

Farm size (acres)	Number of Farms N_h	Average Corn Acres $\bar{Y}_h$	Standard Deviation S_h
0 – 40	394	5.4	8.3
41 – 80	461	16.3	13.3
81 – 120	391	24.3	15.1
121 – 160	334	34.5	19.8
161 – 200	169	42.1	24.5
201 – 240	113	50.1	26.0
241 –	148	63.8	35.2
	2010	26.3 = Grand mean	

For a sample of 100 farms, compute the sample sizes in each stratum under

(a) proportional allocation,
(b) optimum allocation.

Compare the precisions of these methods with that of simple random sampling. (20)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1977

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. (a) The table below gives the index numbers of wholesale prices of groups/sub-groups of items in India for the last week of 1972 (with base 1961–62 = 100) along with their respective weights.

Group/Sub-group	Weight in all commodity index	Weight in the group	Index number for the last week of 1972
1. Food articles	413	–	244.0
2. Liquor and tobacco	25	–	239.6
3. Fuel, power etc.	61	–	181.4
4. Industrial raw materials	121	–	?
4.1 Fibres	–	340	178.5
4.2 Oil seeds	–	430	253.9
4.3 Minerals	–	30	142.5
4.4 Others	–	200	218.6
5. Chemicals	7	–	202.0
6. Machinery and Transport equipment	79	–	169.6
7. Miscellaneous	294	–	178.1

For the last week of 1072, calculate the index numbers of wholesale prices of (i) industrial raw materials and (ii) all commodities taken together.

(b) The following table relates to the quarterly consumption of newsprint by US publishers during 1959–64:

Year	Quarter 1	2	3	4
1959	385	484	423	488
1960	416	510	424	497
1961	392	486	437	496
1962	415	499	442	508
1963	350	516	443	524
1964	452	550	445	530

Assuming the seasonal pattern to be constant, calculate the seasonal indices for the above time series data by the method of ratio-to-moving average. Comment on your results. (3+5)+(15+2)

Please turn over

2. The following table shows per-capita consumption of butter and margarine in Sweden during 1921-39 along with figures for current retail prices, per capita national income at current prices and consumer price-index number.

year	per capita annual consumption of butter & margarine (Kg.)	average retail price of butter and margarine	per capita national income	consumer price index
1921	12.10	4.62	909	241
1923	13.40	3.12	719	177
1925	14.04	2.94	750	176
1927	15.65	2.52	780	171
1929	17.62	2.31	842	169
1931	18.11	1.99	816	159
1933	18.77	1.91	726	153
1935	19.91	2.30	858	150
1937	20.14	2.33	994	162
1939	20.44	2.53	1114	171

i) Assuming the constant elasticity form of the demand function, estimate the income and price elasticities of demand along with the respective standard errors.

ii) Re-estimate the price elasticity assuming that the income elasticity is 0.4. (15+10)=25

3. (a) Consider an economy with three sectors of production (primary, secondary and tertiary) and three primary factors of production (labour, capital and land). The requirements of the primary factors per unit of gross output in the three sectors and the input-output coefficient matrix A are given below.

	Primary	Secondary	Tertiary
Labour	3.0	1.0	2.0
Capital	0.7	0.5	0.3
Land	1.2	0.1	0.0

$$A = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0.5 & 0 & 0.6 \end{bmatrix}$$

i) If the final bill of goods in the three sectors is given as 50, 30 and 200 units in the primary, secondary and tertiary sectors respectively, what will be their gross output levels ?

ii) Work out the total demand for the primary factors corresponding to the solution obtained above.

iii) If the total capacity in the three sectors be 200, 200 and 600 units respectively, what should be the final demand to ensure 90% capacity utilisation in each sector.

Please turn over

3. (b) (Contd.) Examine graphically whether the distribution of persons by size-classes of income before tax in India for the year 1963-64 given below follows the Pareto form :

Income range (Rs.'000)	Number of persons (in millions)
20 – 25	29.0
25 – 30	16.7
30 – 40	17.0
40 – 50	8.4
50 – 70	6.6
70 – 100	3.2
100 – 200	2.0
200 – 300	0.3
300 – 500	0.2
above 500	0.1

(6+3+6+10)=25

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

(Answer any two questions from this group)

1. Two single sampling inspection plans for acceptance-rectification, viz. $n = 36$, $c = 0$ and $n = 80$, $c = 1$, ensure the same AOQL of 1% for lots of size $N = 1500$. Compare the amounts of inspection necessary in the two sampling plans for different qualities p of incoming lots. Determine the indifference quality levels for the two plans. How can you compare the two OC curves in respect of steepness ? (25)

2. Below are given the results of final testing and inspection, during a period of 5 months, of certain permanent magnets used in electrical relays :

Week ending	No. of magnets inspected	No. of defective magnets
3/12	724	48
10/12	763	83
17/12	[illegible]	70
31/12	748	85
7/ 1	724	45
14/ 1	727	56
21/ 1	720	48
28/ 1	719	67
4/ 2	759	37
11/ 2	745	52

Analyse the data and comment on the state of control or otherwise of the process. (25)

Please turn over

3. An experiment was conducted to determine the effects of different types of music on the output of a group of workers. Results are given in the following table :

Week	Music types			
	A	B	C	D
I	6.06	3.64	4.20	1.86
II	10.20	4.76	4.53	2.60
III	8.34	3.56	4.62	1.84

Analyse the results and find the music type yielding the maximum output.

How is the analysis modified if we drop out music type D?

(15+10)=25

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

(Answer any two questions from this group)

1. Segregation with respect to own colour in F_3 segregating families of a cross between two rice varities, T_1 and T_{131}, are given below :

Family No.	Segregation	
	Purple	No colour
1	84	28
2	89	18
3	58	22
4	112	32
5	57	15
6	62	33

The segregation of purple to no colour is expected to be in the ratio 3:1.

(a) Are the data in agreement with this hypothesis ?

(b) Is the segregation uniform for all the families? ([illegible])

2. (a) The frequencies of the blood groups MM, MN and NN in a sample from a population are as follows :

Genotype –	MM	MN	NN
Frequencies	30	52	18

Estimate the gene frequencies for M and N genes.

(b) In a random sample of population, the following phenotypic distribution was observed from ABO blood groups :

Blood group	O	A	B	AB
Frequencies	200	170	35	6

Obtain the estimates of gene frequencies. (25)

Please turn over

3. The segregation of sweet pea plants for flower colour and pollen shape in a F_2 population is as follows :

Pollen shape	Flower Colour	
	Purple	Red
Long	296	27
Round	19	85

(a) Test whether there is linkage between the two characters.

(b) If so, estimate the recombination fraction along with its standard error. (25)

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

(Answer any two questions from this group)

1. The following table shows the age-specific fertility rates (births per 1000 females per annum) and the proportion of females surviving from birth to mid point of the age group:

Age of mothers	Fertility rate	Proportion surviving (from birth to mid point of age group)
15 - 20	120.4	0.5790
20 - 25	338.3	0.5555
25 - 30	[illegible]1.9	0.5250
30 - 35	185.5	0.4[illegible]
35 - 40	[illegible]27.8	0.4365
40 - 45	10.[illegible]	0.3[illegible]7
45 - 50	16.7	0.3368

Calculate the gross and net reproduction rates for the population.

[Assume that for this population, no. of male births : no. of female births = 106 : 100.] (25)

2. Following is the female population in 1971 for a certain country by quinquennial age groups :

Age group	Female population (in 000's)
15 - 20	48[illegible]
20 - 25	432
25 - 30	391
30 - 35	37[illegible]
35 - 40	279
40 - 45	211
45 - 50	187

Project the population to 1976, using the figures for female life table survivors (1971-76) as given below :

Please turn over

2.
(Contd.)

Age x	Female life table survivors at age x ℓ_x
15	68,227
20	65,812
25	62,044
30	59,329
35	56,483
40	52,993
45	49,224
50	45,733

(Radix of the life table ℓ_0 = 100,000). (25)

3. Use the age specific fertility schedule in Question No.1 (with male births : female births = 100 : 100) and the life table survivorship column (with radix = 100,000) in Question No.2 to estimate the true rate of natural increase (r) in a stable population. (25)

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

(Answer any three questions from this group)

1. (a) Mr. X and Mrs. Y obtained a score of 117 and 95, respectively, on a psychological test consisting of 150 items. The reliability of this test is .78. The test scores have a mean of 60 and a standard deviation of 16.2

i) What is the standard error of Mr. X's test score ?

ii) What is the standard error of Mrs. Y's test score ?

iii) What upper and lower limits should be assigned to Mr. X's true score at the 1% per cent level ?

iv) What upper and lower limits should be assigned to Mrs. Y's true score at the 1% per cent level ?

(b) Compute each of the different types of test errors for the above data, e.g. error of substitution, error of measurement and error of prediction. (8+8)=16

2. Using the data presented in the table below, answer the questions appearing under (a) – (h). Each question carries two points.

Test	Mean	Standard deviation	Number of items	Reliability	No. of Respondents
A	73.2	12.7	120	0.92	300
B	17.3	3.8	25	0.86	250
C	21.3	7.1	50	0.80	430
D	20.3	7.9	75	0.84	150
E	56.5	13.7	100	0.89	200

Please turn over

2. (a) (Contd.) Estimate the variance of Test A if it is increased to 240 items.

(b) Estimate the true variance of Test B if 75 more items are added.

(c) What will be the error of measurement if Test C is lengthened to 150 items ?

(d) How many items would need to be added to Test D to double the true variance ?

(e) What would have been the standard error of scores in Test E if instead of [illegible] respondents, it had been based on data from 2,[illegible] respondents ?

(f) What will be the reliability of Test A if it is lengthened to 24[illegible] items ?

(g) How long would Test B need to be to have a reliability of 0.95 ?

(h) Estimate the index of reliability of Test C for double length. (2X8)=16

3. (a) If the reliability of a test is [illegible], what should be the correlation :

i) between two parallel thirds ?

ii) between two parallel fourths ?

iii) between two parallel halves ?

(b) If x and y are parallel tests, the variance of (x-y) is 73.28 and the variance of (x+y) is [illegible]1.5[illegible].

i) What is the reliability of x+y score ?

ii) What is the reliability of x score ?

iii) What is the reliability of y score ?

(c) If the product-moment correlation between the first and second halves of an educational test is 0.7[illegible], what is the corrected reliability estimate of the whole test ?

(d) If the standard deviation of the total scores of a test is 15.3 and the standard deviation of the "odds-minus-evens" score is 12.1, what is the test reliability on the assumption that odds and evens are parallel sub-tests ? (4X4)=16

4. (a) Extract the Spearman "g" factor from the following matrix of intercorrelation :

R =

-	.64	.48	.40
.64	-	.36	.40
.48	.36	-	.30
.40	.40	.30	-

Please turn over

4. (b) (contd.) Given below is a factor-loading matrix extracted by the Principal axes method. The test reliabilities are also reported at the extreme right column :

Test	Factor I	Factor II	Test Reliability
A	.5	.1	.6
B	.2	−.5	.7
C	.3	.7	.8
D	.8	−.2	.9

On the basis of these data, answer the following :

i) What is the observed Communality (h_D^2) for Test D?

ii) Which test has the largest uniqueness variance ?

iii) Which test has the smallest specific variance ?

iv) Compute the intercorrelations among the four tests.

v) What percentage of total variance is accounted for by the first factor ? (6+10)=16

5. The frequency distribution of test scores in two tests applied to 250 respondents are given below :

Scores	Frequency Test A	Frequency Test B
0 − 9	3	1
10 − 19	15	2
20 − 29	37	7
30 − 39	56	10
40 − 49	85	37
50 − 59	36	135
60 − 69	12	35
70 − 79	4	12
80 − 89	2	8
90 − 99	−	3
	250	250

Find by using

(i) percentile scores

and (ii) T-scores

the relative performances of three respondents whose raw scores are given below :

Respondent	Raw score Test A	Raw score Test B
1	40	60
2	50	50
3	60	50

(8+8)=16

NEATNESS (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1977

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

Time: [illegible] hours Full marks: [illegible]

(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. (a) State the basic assumptions of a general linear model and obtain the least square estimators.

   (b) Prove that the least square estimators are best linear unbiased.

   (c) Explain the procedure of significance tests for the regression coefficients clearly specifying the assumptions you make.
(5+1[illegible]) = [illegible]

2. (a) Show with the help of a two-variable linear model, that if there are errors of measurement in the variables least square estimators will be biased and inconsistent.

   (b) Clearly explain the difference between the classical approach and the use of instrumental variables in the estimation of parameters where the variables are subject to error. ([illegible]) = [illegible]

3. (a) Derive the cost function of a firm operating in a competitive market and state its properties, making explicit your assumptions.

   (b) Derive the cost function of a firm faced with Cobb-Douglas type of production function and explain the procedure you would use for the estimation of parameters and tests of significance. ([illegible]) = 20

4. (a) Why are single cross-section samples not suitable for estimation of price elasticities of demand ? Would a time series of cross section samples be more suitable for this purpose? Give clearly the reasons for your answer.

   (b) Do logarithmic demand functions of the form

$$\log E_{ij} = \alpha + \beta \log y_j + x_{ij},$$

where E_{ij} = expenditure on the ith commodity by the jth family unit

y_j = income of the jth family unit

x_{ij} = random disturbance

provide good graduations to all components of family expenditure and savings in the estimation of Engel curves from family budget data?

   (c) How would you account for family size in the estimation of engel elasticities from family budget data ? ([illegible]) = [illegible]

Please turn over

5. (a) Explain the relation between inter-industry on input-output accounts and national income.

(b) Discuss some of the uses of input-output analysis with the help of appropriate illustrations. (10+10)=20

6. (a) Explain clearly the salient properties of log-normal distributions.

(b) Discuss the comparative merits of log-normal and Pareto distributions for an analysis of income distribution. (10+10)=20

7. Consider an income-determination model consisting of the following relations :

$$C_t = \alpha + \beta Y_t + u_t$$

$$Y_t = C_t + Z_t$$

where C = consumption expenditure

Y = income

Z = non-consumption expenditure

x = stochastic disturbance

t = time period

(a) Explain the problems that rise if the method of least suares is applied to the equations individually.

(b) Show how these are handled by the methods of indirect least squares and two-stage least squares. (6+14)=20

8. (a) Define the problem of multi-collinearity in a single equation linear model.

(b) Explain the limitations of the least squares method in case of multicollinearity.

(c) Discuss the methods that can be used for the estimation of parameters in such cases. (2+6+12)=20

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

i) Problem of identification in simultaneous equation problems.

ii) Use of dummy variables in the estimation of parameters.

iii) Maximum likelihood estimators in a two-variable linear model.

iv) Problem of auto-correlation in time-series data. (10+10)=20

Please turn over

GROUP B : TECHNO-COMMERCIAL STATISTICS

Statistical Quality Control - Special Paper I

(Answer any five questions from this group)

1. (a) Using the well-known empirical relationship between mean, median and mode, show that the median of the χ^2 distribution with ν degrees of freedom is approximately given by $\nu - \frac{2}{3}$.

(b) Consider a single sampling plan by attribute for the case where the lot size is large. Let the sample size be n and acceptance number c. Let $p_{0.5}$ denote the fraction defective for which the probability of acceptance is .5. Using the result of (a) and the Poisson approximation to the binomial distribution, show that

$$p_{0.5} = \frac{c + 2/3}{n}$$

(10+10)=2

2. Describe in detail the layout and other relevant features of MIL-STD 105D sampling plans. ( )

3. (a) In a control chart for averages, only the upper control limit is used, so that we conclude that the process is out of control when we observe for the first time an out of control point. If for some process conditions, the probability of getting an out of control point is p, show that the average run length for the chart to indicate an out of control situation is 1/p.

(b) For the case considered in (a), if we modify the rule and say that the chart indicates that the process is out of control if r successive out of control points are observed for the first time, derive an expression for the average run length. (5+15)=

4. (a) Find a known-sigma variables acceptance sampling plan for the case of a one-sided specification limit, say U, given that under the plan lot with 100 p_1 and 100 p_2 percent defectives would be accepted with probabilities $1 - \alpha$ and β, respectively. Give an expression for the OC curve of the sampling plan.

(b) What is a chain sampling plan ? When would you recommend such a plan ? Derive an expression for the OC curve of this plan. (10+10)=20

5. (a) Discuss the use of fractional factorial experiments in industry.

(b) There are four factors A, B, C and D, each at three levels. It is known that the factor D has no interactions with the other three factors. Further, all interactions involving three or more factors are known to be negligible. Design an experiment of 27 experimental units such that all the main effects and the two-factor interactions AB, AC and BC can be estimated. Indicate the ANOVA layout for this experiment.

(c) Indicate how you can construct fractional replicates of 4^m experiments by considering 2^m experiments. (5+10+5)=20

Please turn over

6. (a) Explain, with examples, the following terms encountered in Response Surface Methodology :

i) First and second order designs

ii) Rotatability

iii) Central composite designs.

(b) Describe in detail the strategy of steepest ascent in the determination of optimum operating conditions. (10+10)=20

7. Write short notes on :

(a) Group control charts.

(b) Sloping control charts for the case of tool wear.

(c) Work sampling (20)

GROUP C : BIOMETRIC METHODS

Special Paper I

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

1. (a) Based on n independent Bernoulli trials each with probability of success p, describe a test of the hypothesis H_o: $p = p_o$ against $H_1 : p > p_o$ where $0 < p_o < 1$ and p_o a given number.

(b) In an investigation into the efficacy of a new cow vaccine, two groups A and B were carefully selected so that as far as possible one group was a duplicate of the other. This was done by matching pairs of animals as nearly as possible and spinning a coin to decide which animal was put in A and which in B (the control group). All 200 animals were then equally exposed to the ailment concerned.

At the end of the investigation, 10 animals had contacted the ailment, 8 of which were from the control group, B. Do these results indicate that the vaccine has an effect in preventing the ailment ? What relevance does the information on matching have to your selection of tests of significance ? (5+15)=20

2. The potato cyst eelworm forms encysted cells or cysts on the roots of potatoes. Some of these cysts contain live eggs which will later hatch to form a new generation. To estimate the number of live cysts per unit weight of soil, a random sample of weight x is taken and t cysts are counted. A subsample of n cysts are opened and r are found to contain live eggs. Derive an approximate expression for the sampling variance of $\frac{rt}{nx}$, the estimated number of live cysts per unit weight of soil.

What instructions would you give to sampling teams to ensure that the coefficient of variation is unlikely to exceed 20 per cent ? (20)

Please turn over

5. In a grazing experiment with 24 dairy cows two replicates in a pasture were available; it was desired to study the effect of three grazing treatments on equal areas of land – (i) continuous grazing, (ii) 15 days of grazing and then 15 days of no grazing, and (iii) two days of grazing and then two days of no grazing – with four cows per grazing treatment. In addition, two of the four cows on each grazing treatment were to be fed concentrates and the other two were not. The yield character studied is pounds of butter-fat per cow per year. Set up a design for such an experiment conducted over a three year period and lay out the degrees of freedom in the analysis of variance. What are the correct mean squares for testing the variation among the various means under the hypothesis to be stated by you ? (20)

6. (a) It is not uncommon in biological work for the correlation of all pairs of p variables to be more or less the same. In such a case, the correlation matrix is of the form

$$R = ((r_{ij})), \quad r_{ij} = r, \quad i \neq j.$$

Let p = 3; assuming multivariate normality, find a suitable discriminant function (actually find the function explicitly) to separate two populations with possibly different mean vectors and with the same dispersion matrix yielding a correlation matrix of the above form.

(b) Iris versicolour and Iris Virginica are two varieties of the Iris flower. Fifty specimens from each variety are observed and classified in a contingency table as follows, according to petal width and petal length.
(First number denotes frequency of Versicolour, second of Virginica in each cell).

Frequency of

Iris Versicolour : Iris Virginica

Petal Length	Petal Width: Small < 1.5 cm	Petal Width: Large ≥ 1.5 cm	Total
Small (< 4.5 cm)	11:0	0:0	11: 0
Medium (≥ 4.5; < 5.5 cm)	24:0	13: 6	37: 6
Large ≥ 5.5 cm	0:1	2:43	2:44
Total	35:1	15:49	50:50

Work out a suitable rule to classify a new specimen into one of the two groups. Estimate the proportion of misclassifications by your rule. (10+10)=20

Please turn over

5. (a) Let X, Y, Z be each a $1 \times p$ random variable. Let (X,Y,Z) have a 3p-dimensional normal distribution with mean vector (λ, μ, ν) and nonsingular dispersion matrix Σ where $E(X) = \lambda$, $E(Y) = \mu$, $E(Z) = \nu$. We have a random sample of size n from (X,Y,Z). Derive a suitable test of the hypothesis

$$H_0 : 3\lambda = \mu + 2\nu \text{ against } H_1 : 3\lambda \neq \mu + 2\nu.$$

(b) In an anthropological study of Egyptian skulls, the researcher has four series of skulls, 91 predynastic, 162 from the sixth to the twelfth dynastics, 70 from the twelfth and the thirteenth dynastics and 75 from Ptolemaic dynastics, 398 in all. On each skull, four measurements (in mm)

X_1 : maximum breadth

X_2 : basio-alveolar height

X_3 : nasal height

X_4 : basic-bregmatic height

are made.

Making suitable assumptions, formulate the observational set-up as a multivariate linear model. Discuss with the help of clearly stated formulae, etc., how you would analyse data from this study to find out if there are differences in the mean vectors of the skull measurements in the four series. (10+10)=20

6. State what techniques you would use to analyse data in each of the situations below. Give reasons for your choice and discuss briefly the steps involved in your analysis.

(a) Scores on five tests in mathematics were obtained from a sample of 200 students at two different stages of their course-mid-year and final. The mid-year test consisted of two papers and the final test of three papers. It is required to find whether it is possible to predict the scores of a student on the final test from his scores on the mid-year test.

(b) Twenty-four rats were divided into two groups of twelve, each group consisting of 3 young females, 3 adult females, 3 young males, and 3 adult males. One group was subjected to 500 units whole body radiation and the other group was subjected to 300 units. The cumulative weight losses in grammes at 1, 3, 6 and 7 days radiation were recorded for each rat. It is required to test whether the stronger radiation produces weight losses different from that of the weaker radiation.

(c) 48 patients suffering from muscle wastage of one of two types - myopathy and neuropathy - were subjected to a muscle biopsy, the result of which were observations on eleven variables. A diagnosis was made, being the consensus of opinion of a group of doctors based on the result of muscle biopsy as well as detailed case histories. It is required to lay down an objective procedure of diagnosis for future patients using only the results of muscle biopsy. (6+7+7)=20

Please turn over

GROUP D : DESIGN AND ANALYSIS OF EXPERIMENTS

Statistical Aspects – Special Paper I

(Answer any five questions from this group)

1. Consider a design with v treatments and b blocks, wherein the i-th treatment appears n_{ij} times in the j-th block, $i = 1, \ldots, v$, $j = 1, \ldots, b$. Assuming the usual model (without interaction) show that

   i) the reduced normal equations for estimating treatment effects are $Ct = Q$ where C is a matrix whose elements depend only on n_{ij} values, t is the vector of treatment effects and Q is the vector of adjusted treatment totals;

   ii) $E(Q) = Ct$, $D(Q) = \sigma^2 C$, where E stands for expectation, D( ) denotes the dispersion matrix and σ^2, the per observation variance;

   iii) the adjusted treatment sum of squares for testing the hypothesis of equal treatment effects is $\hat{t}'Q$ where $\hat{t}$ is a solution of the normal equations. (10+5+5)=20

2. What is the rationale behind analysis of covariance ? For a randomised block design derive the analysis of covariance when only one concomitant variable is used. How do you apply the analysis of covariance in analysing data with missing values? (2+12+6)=20

3. Show that a balanced confounded design for a 3×2^2 experiment in blocks of size 6 can be obtained in three replications. Explain the procedure of obtaining the sums of squares of the affected interactions. (12+8)=20

4. (a) Define the following terms in the context of fractional factorials :

   i) Defining Contrasts
   ii) Aliasing
   iii) Resolution R plans.

   (b) Explain a method of getting a resolution IV plan from a resolution III plan for a 2^n factorial. (3+3+3+11)=20

5. (a) What do you understand by 'recovery of inter-block information' in a Balanced Incomplete Block (BIB) design ?

   (b) Discuss the analysis of a BIB design with recovery of inter-block information. (5+15)=20

6. (a) What are response surface designs ?

   (b) Define a second order rotatable design and obtain necessary and sufficient conditions for a response surface design to be second order rotatable.

   (c) How do you analyse the data collected from a second order rotatable design ? (2+1[illegible])

Please turn over

7. What are the basic assumptions underlying the analysis of variance ? Discuss in detail the steps that you will take when one or more of the assumptions do not hold. (… …)=25

8. Write notes on any two of the following :

i) Cross over designs

ii) Split plot designs

iii) Weighing designs (10+…)=25

GROUP E : SAMPLE SURVEYS

Theoretical Aspects : Special Paper I

(Answer any four questions from this group)

1. (a) From a bivariate population containing N units (with values x_i, y_i, $i = 1, 2, \ldots, N$) a random sample of n units is selected without replacement. Derive estimators of V(x) and covariance (x,y). Define a suitable estimator of $m_x - K m_y$ (when $m_x = \sum_{1}^{N} x_i/N$, $m_y = \sum_{1}^{N} y_i/N$ and K is a constant) and derive an expression for the estimate of the variance of the estimator of $m_x - K m_y$.

(b) Determine optimum allocation of n sample units over L-strata, if it is desired to estimate population total Y. Show that

Variance (Optimum) $\leq$ Variance (Proportional)

(c) There are two strata - the first stratum consists of low income group households and the second stratum consists of high income group households. If the object is to estimate the difference between per household assets in the two strata, how should a sample of n be distributed over the two strata ?

(You are given approximate standard deviations σ_1, σ_2 of the assets per household and number of households H_1, H_2 in the two strata.) (…+2+2+2+4+6+5)=25

2. (a) Using Taylor's expansion (or otherwise) derive the expression for the sampling variance of the combined ratio estimate.

(b) Derive an expression for the bias of the combined ratio estimate in terms of the sampling variance of the combined ratio estimate.

(c) Derive a similar expression for the bias of the separate ratio estimate.

(d) A population contains units belonging to three mutually exclusive categories - A, B and C. n units are selected at random with replacement and n_1 units turn out to be A and n_2 units B. Let π_A and π_B be the proportion of A and B units respectively in the population. If $T = \pi_A - \pi_B$, estimate T and find estimate of the variance of the estimate of T. (6+4+4+2+7)=25

Please turn over

3. (a) A population consists of ng units. The units are arranged at random into n groups of g units each. From each group one unit is selected with probability proportional to size. Derive an unbiassed estimator T of population total Y. Derive an estimator of variance of T.

(b) From a population $(y_i, p_i, i=1,2,\ldots,N;\ p_i > 0,\ \Sigma\ p_i = 1)$ a random sample of two units are selected (one unit after another) with probability proportional to p_i without replacement. Considering the order of selection of the two units, set up an estimate T of population total Y, such that it is possible to have a non-negative estimate of the variance of T. (6+8+1[illegible]) = [illegible]

4. (a) A population is divided into k strata in each of which there are M units (firststage units (f.s.u.)). From each stratum m f.s.u. are selected at random with replacement. From each selected f.s.u. two second stage units (s.s.u.) are selected at random with replacement. Using suitable notations derive an unbiassed estimator T of population total Y (and show that T is unbiassed). Derive an expression for the estimate of the variance of T.

(b) From a population containing N f.s.u.'s, m f.s.u.'s are selected at random with replacement. From each of the m selected f.s.u.'s two s.s.u.'s are selected with probability proportional to size with replacement and variate values y are observed on these $2m$ units. Derive an estimate T of population total Y. Write down the variance of T. Derive an estimate of the variance of T. (4+12+2+2+[illegible]) = 2[illegible]

5. (a) For a sample survey with a two stage sample design the cost function is of the form $C = c_1 m + c_2 mn$, where

c_1 = cost per f.s.u.

c_2 = cost per s.s.u.

m = number of f.s.u. surveyed

n = number of s.s.u. surveyed in each selected f.s.u.

If the variance of the estimator under consideration is of the form

$$\frac{A}{m} + \frac{B}{mn},$$

derive optimum values of m, n in terms of A, B, c_1, c_2 and C, for a fixed cost.

(b) In a repetitive survey n units are selected out of [illegible] afresh every year, retaining the other n units of the previous year. Using suitable notation set up estimator for the ratio R

$$R = \frac{\text{population total in the } (t+1)\text{th year}}{\text{population total in the } t\text{-th year}},$$

and derive the variance of the estimate of R. (12+1+3) [illegible]

Please turn over

3. Write notes (highlighting the important points) on any three of the following :

(a) Circular systematic sampling with probability proportional to size.

(b) Regression method of estimation.

(c) Horvitz-Thompson estimator.

(d) Interpenetrating net work of samples.

(e) Self-weighting designs. (25)

GROUP F : TECHNIQUES OF COMPUTATION

Numberical Analysis : Special Paper I

(Answer any five questions from this group)

1. (a) A familiar result of mathematical analysis is the following theorem :

"If a continuous function $f(x)$ assumes values of opposite signs at the end-points of an interval $[\alpha, \beta]$ i.e. if $f(\alpha)f(\beta) < 0$, then the interval will contain at least one root of the equation $f(x) = 0$".

Use this result to formulate an algorithm for finding an approximation with a given accuracy, to a root of an equation $f(x) = 0$ in a given interval, $f(x)$ being asumed continuous in the given interval.

(b) Describe an iterative process to find the positive cube root of a positive number c. (12+8)=20

2. Discuss the different ways in which round-off errors may crop up in numerical computation. Define absolute and relative round-off errors and show how relative errors are propagated in addition and multiplication.

Suppose that you use a certain machine to add numbers and this machine can store only the four most significant digits of any number. If the machine encounters any number with more than four digits, it chops off all digits to the right of the four most significant digits. You want to add 02654 and 71628 on this machine. Find the answer the machine will produce and calculate the relative error in the answer. Find also the amount of the propagated part of the relative error. (8+4+[illegible]+4)=20

3. Explain the terms "Direct Method" and Iterative Method" as they are used in numerical analysis. Write down two algorithms – one as an example of a direct method and the other of an iterative method – for solving a system of linear equations numerically. Explain the advantages/disadvantages of both the algorithms. (4+12+[illegible])=20

4. Explain the mathematical basis of Jacobi's method for finding the eigen values and eigen vectors of a symmetric matrix. Give briefly a computational scheme for the method. ([illegible])

Please turn over

5. (a) Describe an efficient computational scheme for interpolation with non-equidistant points. The procedure should be such that if necessary one could compute an improved value by using the previous values.

(b) The table below contains an error in a value of $f(x)$ which should be located and corrected :

x	$f(x)$	x	$f(x)$
3.60	0.112046	3.65	0.152702
3.61	0.120204	3.66	0.160788
3.62	0.128350	3.67	0.168857
3.63	0.136462	3.68	0.176908
3.64	0.144600		

(10+10)=20

6. One wants to construct a qudrature formula of the following type :

$$\int_{-1}^{+1} f(x)dx = a[f(-1)+f(1)] + b[f(-\alpha)+f(\alpha)] + cf(0) + R$$

Find a, b, c and α so that $R = 0$ when $f(x)$ is any polynomial with degree $\leq N$ and N should be as large as possible. (20)

7. Describe mathematically the class of methods known as predictor-corrector methods for the numerical solution of ordinary differential equations with given initial conditions. Discuss, in particular, Milne's method showing how an estimate of the error can be made at each step of the computation. What are the advantages or disadvantages of these methods compared to the Runge-Kutta methods. (20)

GROUP G : STATISTICAL INFERENCE

General Theory : Special Paper I

(Answer any five questions from this group)

1. (a) State and prove Neyman's factorisation criterion for sufficiency in the case of a continuous density.

(b) If $X_1, \ldots X_n$ are independent random variables with a common density $p(x; \theta)$ and if this density admits of a sufficient statistic, show that under suitable assumptions the joint density of $X_1, \ldots X_n$ will be of a particular "exponential" form.

(c) Let $x_1, \ldots, x_n$ be i.i.d. random variables having a common rectangular distribution.

$$f(x \mid \theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1}, \quad \theta_1 \leq x \leq \theta_2$$
$$= 0, \quad \text{Otherwise}$$

Find a non-trivial sufficient statistic. (7+7+6)=20

Please turn over

2\. (a) Let θ be a parameter having a maximum likelihood estimator (m.l.e.) 't'. Show that $h(t)$ is a m.l.e. of $h(\theta)$ where $h(\theta)$ is a 1-1 funct n of θ.

(b) Discuss how normal distribution can be utilized to yield confidence interval for a parameter of a general distribution, when samples are of large size.

(c) If x_b is the largest observation in a random sample of size n from a rectangular $R(0, \theta)$ distribution, then show that the confidence limits for θ, with confidence coefficient α, $0 < \alpha < 1$, are x_b and $\dfrac{x_b}{(1-\alpha)^{\frac{1}{n}}}$. (7+7+6)=20

3\. (a) If a minimal co[illegible]te class exists, then prove that it consists of exactly the admissible decision rules.

(b) Let X have a negative binomial distribution with parameters (Ψ, ν), $0 < \Psi < 1$, ν known. Consider the loss function

$$L(\Psi, \hat{\Psi}) = \frac{(\Psi - \hat{\Psi})^2}{\Psi(1-\Psi)}$$

i) Derive the Bayes estimators against the prior beta (a, b) distribution. Are these estimators admissible ?

ii) Is the M.L.E. a generalized Bayes estimator? Is it a minimax estimator ? (10+5+5)=20

4\. (a) What is meant by the 'critical function' of a test ? What is the form of the critical function of a non-randomized test ? Express the power function of a test in terms of its critical function.

(b) When is a test said to be (i) unbiased, (ii) uniformly most powerful (UMP) against a composite alternative ? Prove that UMP test against a composite alternative is also unbiased against it. (10+5+5)=20

5\. (a) State Wald's fundamental equation for the ASN of a sequential test involving a sequence of independently and identically distributed variables, clearly specifying the underlying assumptions.

(b) For the problem of testing a simple hypothesis $\theta = \theta_0$ against a simple alternative $\theta = \theta_1$ subject to given error probabilities, deduce lower bounds of the ASN at θ_0, θ_1 of a sequential test. Show that the sequential probability ratio test attains these bounds approximately. (11+9)=20

6\. (a) Distinguish between 'a complete class' and 'an essentially complete class' of decision rules.

For a non-sequential decision problem show that the existence of a sufficient statistic implies the essential completeness of all randomized rules based on sufficient statistics.

(b) Formulate the hypothesis-testing problem as a two-decision problem. Show that if the loss function satisfies some reasonable conditions, a test is uniformly better than another in terms of risk function if and only if it is so in terms of power. (5+7+8)=20

Please turn over

7. (a) Define the concepts of 'unbiasedness' and 'shortness' of confidence intervals in the sense of Neyman.

(b) Establish the 1-1 correspondence

i) between level $(1-\alpha)$ confidence interval for a parameter θ and critical region of size α for testing $\theta = \theta_0$,

ii) between U.M.P.U. tests for $\theta = \theta_0$ and "shortest unbiased" confidence intervals for θ.

(c) Use the above results to construct an unbiased and shortest confidence interval for the parameter μ of $N(\mu, \sigma^2)$. (6+8+6)=20

GROUP H : PROBABILITY THEORY

Basic Probability : Special Paper I

(Answer any two questions from Section 1 and all questions from Section 2)

Section 1

(a) Let μ be a countably additive set function on a σ-field F with $\mu(\phi) = 0$.

i) if $A_1, A_2, \ldots \in F$ and $A_n \uparrow A$ show that $\mu(A_n) \to \mu(A)$ as $n \to \infty$.

ii) if $A_1, A_2, \ldots \in F$ and $A_n \downarrow A$, and $\mu(A_1)$ is finite, show that $\mu(A_n) \to \mu(A)$ as $n \to \infty$.

(b) Let μ be a measure on a field F_0 of subsets of Ω, and assume that μ is σ-finite on F_0, so that Ω can be decomposed as $\bigcup_{n=1}^{\infty} A_n$, where $A_n \in F_0$ and $\mu(A_n) < \infty$ for all n. Then show that μ has a unique extension to a measure on the minimal σ-field F over F_0.

(a) If $h_1, h_2, \ldots$ are Borel measurable functions from Ω to $\overline{R}$, and $h_n(\omega) \to h(\omega)$ for all $\omega \in \Omega$, then prove that h is Borel measurable

(b) Let f be a function from R^k to R^p, not necessarily Borel measurable. Show that $\{x: f$ is discontinuous at $x\}$ is an F_σ (a countable union of closed subsets of R^k) and hence is a Borel set.

(a) If μ is a σ-finite measure on F, g and h are Borel-measurable, $\int_\Omega g\, d\mu$ and $\int_\Omega h\, d\mu$ exist, and $\int_A g\, d\mu \leq \int_A h\, d\mu$ for all $A \in F$ then show that $g \leq h$ a.e. $[\mu]$.

(b) Let μ be a σ-finite measure on a σ-field F, λ a σ-finite signed measure (i.e. $|\lambda|$ is σ-finite) then show that λ has a unique decomposition as $\lambda_1 + \lambda_2$ where λ_1 and λ_2 are signed measures, such that $\lambda_1 \ll \mu$, $\lambda_2 \perp \mu$.

Please turn over

4\. (a) If $\{ f_n \}$ is convergent in measure w.r.t. a finite measure then show that there is a subsequence converging almost uniformly to the same limit function.

(b) If $|f_n| \leq g$ for all $n = 1,2, \ldots$, where g is μ-integrable, and $f_n \xrightarrow{\mu} f$, show that f is μ-integrable, and $\int_\Omega f_n d\mu \to \int_\Omega f d\mu$

Section 2

5\. (a) If $X_n \xrightarrow{dist} X$ and $Y_n \xrightarrow{dist} 0$, then show that

$$\text{(i) } X_n + Y_n \xrightarrow{dist} X; \quad \text{(ii) } X_n Y_n \xrightarrow{dist} 0.$$

(b) Prove that the probability of convergence of a sequence of independent random variables is equal to zero or one.

6\. (a) If a sequence of r. v's $\{ X_n \}$ converges to a r.v.X in probability show that there exists a subsequence $\{X_{n_k}\}$ which converges to X with probability one.

(b) Give an example with proof of a sequence of random variables which converges in probability but does not converge with probability one.

7\. (a) If $\{X_i, 1 \leq i \leq n\}$ are independent random variables, and $\{f_i, 1 \leq i \leq n\}$ are Borel measurable functions, then prove that $\{f_i(X_i), 1 \leq i \leq n\}$ are independent random variables.

(b) If X_1 and X_2 are independent random variables each assuming the values +1 and −1 with the same probability 1/2, then show that the three random variables X_1, X_2, X_1X_2 are pairwise independent but not totally independent. Give an example (with proof) of n random variables such that every n-1 of them are independent, but not all of them are independent (for some n > 3).

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination : November 1977

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

Time: 4 hours | Full marks: 10

(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

Indian Economics and Economics of Planning : Special Paper I.

Section I : Indian Economics (50 marks)

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

1. Examine the growth of the agricultural sector in India during the Plan period. Suggest the main lines of an appropriate agricultural policy for India. (1.+1)

2. Briefly review India's export performance during the last decade, indicating the important changes in the structure and direction of exports.

3. "Given excessive government expenditures, monetary and credit policies have a fairly limited role in correcting inflationary pressures".
   – Examine this statement on the basis of Indian experience.

4. What are the broad principles on the basis of which divisible funds are allocated between the Union and the States in India? Can you suggest a better arrangement ? (12+3)

5. Write a short note on any one of the following :
   (a) Poverty in India.
   (b) Agricultural taxation in India.
   (c) Impact of foreign aid on the Indian economy.

Section II : Economics of Planning (50 marks)

(Answer question No.10 and any two of the remaining questions)

6. The well-known Harrod-Domar model assumes that a constant proportion of output is saved and invested, that the capital-output and labour-output ratios are constant, and that the labour force grows at a proportionate rate.
   (a) Formulate this model mathematically, explicitly stating whatever additional assumptions you make.
   (b) Derive the warranted rate of growth.
   (c) Derive the natural rate of growth.
   (d) What do economists mean when they say that this model is unstable ?
   (e) Can you make an additional assumption so that the model will become stable? (4+1+4+...)

7. The crucial decision in planning for an economy over a given period lies in the choice regarding consumption at a relatively low level in the immediate periods followed by the prospect of a high rate of growth of consumption later. Discuss critically Mahalanobis planning model for India in this light. (20)

8. (a) Define the concept of "viability" in input-output analysis.

(b) State and interpret a "viability condition" for an input-output model, clearly indicating whether the condition is necessary or sufficient or necessary and sufficient for viability.

(c) Sketch an analytical derivation of the "viability condition" that you have stated above. (4+8+8)=20

9. (a) Define a linear programming (LP) problem.

(b) Consider a factory with a given endowment of certain inputs which can be used in a given number of alternative methods of producing some given commodity, where each method of production is defined by a set of fixed input-coefficients. Show that the problem of maximising total output of the factory reduces to an LP problem.

(c) Define and interpret slack variables in this context.

(d) Define a feasible solution and an optimal solution for an LP problem, and show that both a feasible solution and an optimal solution always exist in the case considered. When is a feasible solution for an LP problem called a basic feasible solution ?

(e) Given a basic feasible solution, can you suggest some method of checking whether it is an optimal solution (in the sense of maximising total output) ? (2+4+2+6+6)=20

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

(a) Self-reliance as an explicit or implicit objective of planning in India.

(b) The concept of time-horizon in using a growth model for planning. (10+10)=20

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 separate answer-book for this Section.

(b) Answer any four questions from this Section.

1. Each circle in the figure 1 represents a city. A specific material can be shifted from city 'i' to city 'j' only if there is a directed arc from city 'i' to city 'j'. The three numbers associated with arc (i,j) are l_{ij}, k_{ij} and

Please turn over

1. (contd.) c_{ij} where l_{ij} = least number of tons of the material that can be shifted along this arc, k_{ij} = maximum number of tons of the material that can be shifted along this arc and c_{ij} = the cost (in rupees) of shifting one ton of material along this arc.

There are 27 tons of the material at city '1' and all of it should be shifted to city '6'. The total material brought into the city 'i' should be equal to the total material taken out of the city 'i' for the cities i = 2, 3, 4 and 5. The problem is to decide the quantity to be shifted along the arc (i,j) such that the above restrictions are satisfied and the total cost of shifting the material is minimized.

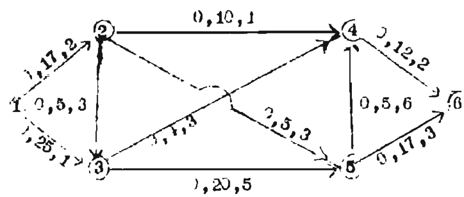

Figure 1

(a) Formulate the above problem as a linear programming problem.

(b) State clearly the assumption made in the formulation. (12+5).

2. A company has three warehouses numbered 1, 2, 3 containing 100, 50 and 160 units of its product. In the next month 30, 60, 75, 100 units must be shipped to 4 retail outlets numbered 1, 2, 3 and 4. The unit cost (in rupees) of shipment from any warehouse to any retail outlet is contained in the following matrix

Ware house	Retail outlets 1	2	3	4
1	10	8	16	3
2	7	25	18	7
3	20	17	20	5

It is desired to determine the quantity to be shipped from various warehouses to various retail outlets such that the total cost of shipping is minimized.

(a) Convert the above problem into a standard form of a transportation problem and write down the initial solution by North West corner rule.

(b) Find the optimal solution using the 'uv' method of solving the transportation problem. (7+ =17)

Please turn over

3. For a certain item in a store the following information is given. Ordering cost per order independent of the order quantity is Rs.A/=. Purchasing cost per unit of the item is Rs.C/=. Cost per unit of the item short per unit time is Rs.b/=. Inventory carrying cost per rupee per annum is Rs.a/=. Annual demand of the item is D units. Assume that the rate of consumption is uniform and the back orders are fulfilled. It is desired to find the fixed order quantity Q such that the total cost per annum is minimized.

(a) Prove that the optimal fixed order quanity $Q^* = \sqrt{\frac{2DA}{Ca}} \cdot \sqrt{\frac{b+aC}{b}}$

(b) Show that the maximum inventory level corresponding to the optimal fixed order quantity $= M^* = \sqrt{\frac{2DA}{Ca}} \cdot \sqrt{\frac{b}{b+aC}}$. (12+5)=17

4. A company uses two large machines in a processing operation. Each machine is running all the time except when down for repairs. Assume that a machine's running time follows a negative exponential distribution with mean running time equal to $\frac{1}{\lambda}$. Assume that there are 2 repair men available and repairing time of a machine by any of the repair men follows a negative exponential distribution with mean repair time equal to $\frac{1}{\mu}$. Further assume that only one repair man will work on a machine at a time. It is known that down time of machine costs Rs.100/= per hour. Suppose that it costs Rs.25 μ/= per hour to operate the repair men at the service rate μ. Assume $\lambda = \frac{1}{2}$ per hour. Determine the optimal value of μ such that the average total cost per hour is minimized. 17

5. Each circle in figure 2 represents a city. One can travel from city 'i' to city 'j' only if there is a directed arc from city i to city j. The numbers on the arc denote the distance between the cities. A person has to travel from city '1' to city '8'. He wants to travel through the shortest route.

(a) Formulate the above problem as a dynamic programming problem.

(b) Find the optimal route and the corresponding distance travelled.

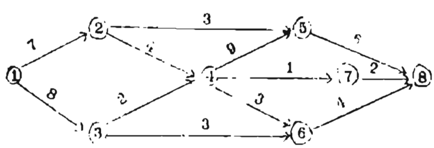

Figure 2

(9+8)=17

6. Write short notes on :

(a) Monte-carlo method.

(b) Replacement of r item that depreciates. (9+8)=17

Please turn over

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

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

(b) Question No.1 is compulsory. Answer any three from the rest.

1. The ABC Company makes up its accounts to 31st March, 1977. The following balances were extracted from the books on 31st March, 1977.

	Rs.
Balance at bank	104418
Sundry creditors	42726
Sundry debtors	58140
Land & Building	95000
General reserve	16000
Motor vehicles	22400
Share capital	200000
Profit and Loss account (balance on 31.3.76)	7452
Stock on hand on 31.3.77.	66570
Trading profit before adjustments	96350
Goodwill	16000

Prepare a Trial Balance of the ABC Company as at 31st March 1977. You are also required to prepare Company's Profit and Loss account for the year ended 31st March 1977 and a Balance Sheet as on that date after making necessary adjustments for

i) a provision for fees of Rs.1250 for the Chairman and Rs.1000 for the other non-executive director.

ii) the doubtful debts provision to be made to 5% on outstanding debtors.

iii) depreciation to be provided for the year on motor vehicles at 20%.

iv) a transfer to general reserve of Rs.14000. (6+6+10)=22

2. Write the difference between revenue and capital expenditure. State which of the following items should be charged to capital and which to revenue.

i) Salary and wages

ii) Royalty

iii) Advertisement

iv) Preliminary expenses

v) Legal expenses incurred in the acquisition of property

vi) Stationery. (..+12)=

3. Write notes on

i) Ledger

ii) Nominal Accounts

iii) Bad Debts

iv) Petty Cash Book. (..x..)=1

Please turn over

4. Rectify the following errors passing journal entries.

a) A machine purchased for Rs.4000 from the XYZ Company, has been entered in the Purchase Day Book.

b) A sale of Rs.40 to X was wrongly debited to the account of Y.

c) Rs.360 paid for repairs to Building was debited to Building Account as Rs.630.

d) A sale of Rs.200 for furniture has been passed through the Sales Book. (4x4)=16

5. Distinguish a Trial Balance from a Balance Sheet. Cite four items which are found in Trial Balance but not in Balance Sheet. (8+8)=16

6. A firm purchased a plant on 1st April 1972 for Rs.50,000. Depreciation is written off at the rate of 10 per cent. Show for 5 years the Plant Account under the Straight Line Method. The firm closes its books on 31st December each year. (16)

Section II: Statistical Methods in Business (30 marks)

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

(b) Answer any two questions from this Section.

1. (a) What are the important points to be taken care in designing a market survey ?

(b) Design a questionnaire for a market survey whose objective is to assess the market share and brand image of the various toilet soaps. (7+7)=14

2. (a) Compare moving average method and simple exponential smoothing method for short term business forecasting.

(b) Show that simple exponential smoothing is not an adequate method for forecasting when trend is present.

3. Write short notes on :

(a) Incentive schemes

(b) Sampling techniques in auditing. (7+7)=14

NEATNESS (2)

Please turn over

GROUP C : BIOMETRIC THODS

Statistical Methods in Genetics and Bio-assays
Special aper II

Section I : Statistical Methods in Genetics (50 marks)

(Answer any three questions from this Section)

1. (a) Differentiate between
   i) Genotype and Phenotype
   ii) Autosomes and Chromosomes
   iii) Dominance and Recessiveness
   iv) Linkage and Crossing over

(b) Explain, giving appropriate derivation, the method of testing single factor segregations, when data on two families are available, one being F_2 and other being a double back-cross.

2. (a) Explain the Hardy-Weinberg law of equilibrium under random mating.

(b) Derive the equilibrium frequency of a mutant under pressures of selection and mutation when the mutant is recessive.

3. (a) Describe environmental and genetic components of variation. How is the regression of offspring on one parent used to estimate additive genetic component of variation ?

(b) Explain the use of a linear discriminant function in plant selection.

4. Explain how you overcome the difficulties involved in analysing human data for testing single factor segregations. Describe Fisher's method of ascertainment of the frequency of a rare recessive gene in human beings.

Section II : Bio-assays (50 marks)

(Answer any two questions from this Section)

5. (a) Differentiate between bio-assays based on quantitative and quantal responses.

(b) Assuming that tolerance distribution is normal, explain how relative potency is estimated through direct assays.

6. Explain the role of probit transformation in the analysis of qualitative bio-assays and describe a method of estimating ED50 (Effective Dose 50).

7. Discuss in detail a 2k-point parallel line assay. Give the analysis of variance table and indicate how to test for validity. What are the draw-backs of such an assay ?

NEATNESS (for Section I and II)

Please turn over

GROUP D : DESIGN & ANALYSIS OF EXPERIMENTS

Combinatorial Aspects : Special Paper II

(Answer any five questions from this group)

1. (a) For a B.I.B design with parameters v, b, r, k and λ, if x denotes the number of treatments common between any two blocks, then show that

$$-(r-\lambda-k) \leq x \leq \frac{2\lambda k + r(r-\lambda-k)}{r}$$

(b) Let $B_i \Delta B_j$ ($\equiv B_i \cup B_j - B_i \cap B_j$) denote the symmetric difference between the i-th and j-th blocks of a SBIBD with parameters v = b, r = k, λ. Show that the set of $\binom{v}{2}$ blocks obtained by taking symmetric differences of all possible blocks is a B.I.B. design. Derive its parameters. (14+6)=.

2. (i) Define a partially balanced association scheme.

(ii) Define two associate triangular, L_1 and cyclic association schemes.

(iii) You are supplied with a two associate partially balanced association scheme with association parameters $v, n_1, n_2, p^1_{11}, p^2_{11}$. Derive the parameters of the designs that can be constructed if you consider (a) all the first associates of any treatment as a block, (b) all the second associates of any treatment as a block, (c) any treatment with all its first associates as a block and (d) any treatment with all its second associates as a block. (2+6+ ...)

3. (i) How do rotatable designs differ from standard factorial experiments ?

(ii) Describe any general method of constructing a S.O.R.D. in k factors.

(iii) Given a second order rotatable design in k-factors in n design points, how can you construct from this design a S.O.R.D. in (k+1) factors ? (4+12+.)=..

4. Consider the following statements :

i) there exists a set of n-1 m.o.l.s. of order n

ii) there exists a set of n-2 m.o.l.s. of order n

iii) there exists a set of n-3 m.o.l.s. of order n

If n > 4 show that (ii) implies (i) and (iii) implies (ii). (8+12)=20

5. Show that the existence of a set of t m.o.l.s. of order n implies and is implied by the existence of an orthogonal array $[n^2, t+2, n, 2]$ of index unity as well as the S.R.G.D. design.

$v = n(t+2)$, $b = n^2$, $r = n$, $k = t-2$, $\lambda_1 = 0$,

$\lambda_2 = 1$, $m = t+2$, n (10+10)=20

Please turn over

6. Show that for an (s^n, s^{n-1}) design one can find $(s-1)^{n-1}$ replications in which the main effects remain unconfounded and a complete balance is achieved over interactions of all orders, from the first to (n-1)th. The loss of information on the (k-1)th order interaction is given by

$$\frac{(s-1)^{k-1} - (-1)^{k-1}}{s(s-1)^{k-1}}$$

(20)

7. Show that if a fractionally replicated design consists of a subset of treatment combinations forming an array of strength (d +1), then one can obtain independent estimates of main effects and first order interactions if interactions of order (d-1) or higher are considered to be negligible.

Show further that if these treatment combinations are arranged in b blocks of equal size, then a sufficient condition that the main effects and first order interactions are preserved is that each block is an array of strength 2. (20)

GROUP E : SAMPLE SURVEYS

Organisational Aspects : Special Paper II

(Answer all questions from this group)

1. It is decided to conduct an inquiry to ascertain the birth rate and death rate of the population. How will you proceed to plan the inquiry starting from the selection of items of information to be collected to its presentation and dissemination of the results of the inquiry.

Discuss the utility of the data and how these can be utilized in important domains of national planning.

2. A survey on the consumption expenditure of households is proposed to be conducted on a nation-wide basis by obtaining information from sample households through interrogation. You are asked to build up a field organization for the collection of data. Discuss the problems that are involved in the collection of data and the steps needed to build up the organization mentioning the aspects of coordination between the field organization on the one hand and the planners of the survey and the statistical office tabulating the data on the other hand.

3. Write short notes on the following :

(a) Types of errors in surveys.

(b) Sampling frame.

(c) Collection of information by physical observation and inquiry.

(d) Ad hoc and repetitive surveys.

(e) Editing and scrutiny of data.

Please turn over

4. A survey on employment and unemployment is to be carried out in urban areas. Mention the concepts and definitions to be used in the survey. Frame a suitable schedule of enquiry and prepare the layout of the final tables to be generated from the data and discuss the use to which the data can be put.

5. A survey on land-holdings is to be carried out in rural areas. In this context answer the questions in the order given below :

(1) Discuss the results which you would like to get from the survey keeping in view the question of land reform and re-distribution of land.

(2) Give the layout of the tables to be generated from the data to enable you to get the results mentioned in (1).

(3) Enumerate the items for which you are going to collect the information and the population to be covered for the needed result.

(4) Prepare the layout of the schedule of inquiry.

GROUP F : TECHNIQUES OF COMPUTATION

Numerical Computation : Special Paper II

Time : 3 hours (Practical with Desk Calculators) Full marks : 100

(Answer any four questions)

1. (a) Consider the following two linear equations in two unknowns x_1 and x_2, where the coefficients a_{ij}'s are obtained by measurements which may be subject to errors as high as 5%.

$$a_{11} x_1 + a_{12} x_2 = a_{13}$$

$$a_{21} x_1 + a_{22} x_2 = a_{23}$$

where $a_{11} = 1.234 \pm 5\%$ $a_{12} = 2.341 \pm 5\%$ $a_{13} = 5.678 \pm 5\%$

$a_{21} = [illegible].521 \pm 5\%$ $a_{22} = 3.210 \pm 5\%$ $a_{23} = 10.987 \pm 5\%$

Evaluate x_1 and x_2 and their possible margins of error.

(b) What is the possible percentage of error in the computed volume of a sphere whose diameter is measured to be 1 metre, but the measurement is subject to an error of 2.5% ? (18+7)=25

2. The following table gives the values of a function f(x) for x = 0.0 (0.1) 0.9 :

x	f(x)	x	f(x)
0.0	1.63662	0.5	.20306
0.1	1.56425	0.6	.19099
0.2	1.[illegible]8946	0.7	.12248
0.3	1.[illegible]	0.8	1.05852
0.4	1.33962	0.9	1.00000

Prepare a table of differences to locate and correct errors, if any, in the table. After correcting errors, evaluate by interpolation f(0.42) and assess its margin of error. (25)

Please turn over

3. The following table gives the values of $\alpha\ (h, k, \rho)$ for different values of h, k and ρ :

$\rho = 0.80$				
k \ h	0.0	0.1	0.2	0.3
0.0	0.3976	0.3766	0.3538	0.3294
0.1		0.3583	0.3389	0.3162
0.2			0.3204	0.3011
0.3				0.2813

$\rho = 0.85$				
k \ h	0.0	0.1	0.2	0.3
0.0	0.4117	0.3935	0.3679	0.3117
0.1		0.3723	0.3518	0.3202
0.2			0.3342	0.3145
0.3				0.2978

$\rho = 0.90$				
k \ h	0.0	0.1	0.2	0.3
0.0	0.4282	0.4067	0.3822	0.3552
0.1		0.3867	0.3678	0.3441
0.2			0.3554	0.3302
0.3				0.3135

$\rho = 0.95$				
k \ h	0.0	0.1	0.2	0.3
0.0	[illegible]	0.4271	0.4035	0.3795
0.1		0.4099	0.3880	0.3622
0.2			0.3712	0.3500
0.3				[illegible]

Given further that $\alpha\ (h, k, \rho) = \alpha\ (k, h, \rho)$, find by interpolation the value of $\alpha\ (0.10, 0.11, 0.875)$ as accurately as you can. (25)

4. Evaluate by any suitable method

$$F(x) = \int_0^x \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}t^2}\, dt$$

for $x = 0.1\ (0.1)\ 1.0$, each correct atleast to 5 places of decimals. (25)

– Please turn over

5. Find, correct to four places of decimals, a solution of

$$\sin x = y + 1.32$$

$$\cos y = x - [illegible]$$

( )

6. Find the largest latent root and the associated normalised latent vector of the symmetric matrix whose elements on the principal diagonal and above are as follows :

2.296	0.067	– 0.557	0.709	–0.991
	0.498	0.113	0.211	–0.158
		0.417	0.079	0.279
			1.472	–0.593
				1.295

(25)

GROUP G : STATISTICAL INFERENCE

Special Topics : Special Paper II

(Answer any five questions from this group)

1. (a) Obtain the inequalities connecting the constants A and B of a SPRT with error probabilities α and β.

(b) Obtain the O.C. function and A.S.N. function of a sequential probability ratio test of $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 \ (> \theta_0)$, when observations are taken independently from a Poisson distribution with mean θ.

(8+12)=20

2. (a) What is the problem sought to be solved through Stein's two-stage procedure ? Explain how the problem is solved through it, giving proofs of the statements.

(b) Suppose that a hypothesis $H_0 : \theta = \theta_0$ is to be tested against $H_1 : \theta = \theta_1$, on the basis of independent observations arising from a $N(\theta, 1)$ distribution. If the error probabilities are α and β, obtain the A.S.N. under H_0, of the SPRT of H_0 against H_1 and compare it with the sample size of a fixed-sample-size test having the same error probabilities.

(10+10)=20

3. (a) Discuss how you would employ order statistics to find a confidence interval for the population median (Give proofs of statements made in this discussion).

(b) If $X_{(r)}$ and $X_{(s)}$ are rth and sth order statistics $(s > r)$ in a random sample of size n from a population with density function

$$f(x) = \begin{cases} e^{-x} & \text{for } 0 \leq x \leq \infty \\ 0 & \text{otherwise} \end{cases}$$

then show that $X_{(r)}$ and $X_{(s)} - X_{(r)}$ are statistically independent.

(10+10)=20

Please turn over

4. (a) What is an emperical distribution function ? Show that it converges to the population distribution function with probability 1, stating in full the results used, if any.

(b) Evaluate the exact distribution function of the maximum difference

$$D^{+}_{n,n} = \max_{x} \{ S_n(x) - S'_n(x) \}$$

between two emperical distribution functions, under the hypothesis that the two samples of equal size have arisen from populations with identical continuous distribution functions.

(...)

5. (a) Given independent samples of size m and n from populations with continuous distribution functions F( ) and F(x-θ) respectively, obtain a recurrence relation for the probability function of the Mann-Whitney statistic under the null hypothesis H_0 : θ = 0.

(b) Write notes on <u>any two</u> of the following :

i) Use of runs in non-parametric tests.

ii) Treatment of ties.

iii) Tolerance limits.

(...)

6. (a) If $\underline{X}$ has a non-singular $N_p(\underline{\mu}, \Sigma)$ distribution and $\underline{Y} = D\underline{X}$ where D is a k x p matrix of rank k < p, then show that $\underline{Y}$ has a k-variate non-singular normal distribution.

(b) If $\underline{x}$ has a $N_p(\underline{0}, I)$ distribution, then show that a quadratic form $\underline{x}' A \underline{x}$ has a chi-square distribution if and only if A is idempotent.

(...)

7. (a) Let $\underline{x}_1, \ldots, \underline{x}_n$ be n independent observations from a non-singular $N_p(\underline{\mu}, \Sigma)$ distribution. Write down the test statistic and its null sampling distribution (without proof) for the hypothesis of equality of the components of $\underline{\mu}$ in the following situations :

(i) Σ known (ii) Σ unknown.

(b) Write notes on <u>any two</u> of the following :

i) Principal components and their uses.
ii) Discriminant functions and their uses.
iii) Salient features of factor analysis.

(5+5+...)

GROUP H : PROBABILITY THEORY

<u>Limit Distributions : Special Paper II</u>

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

1. (a) State and prove the Kolmogorov's strong law of large numbers.

(b) Let {X_n} be a sequence of independent identically distributed random variables with zero mean. Show that

$$\frac{1}{n} \sum_{k=1}^{n} S_k$$

where $S_k = \sum_{i=1}^{k} X_i$ does not tend to zero in probability as n→∞.

(...)

Please turn over

2. Let $\{A_k\}$ be a sequence of pairwise independent events such that $\sum_{k=1}^{\infty} P(A_k) = \infty$.

(a) Show that $[(\sum_{k=1}^{n} I(A_k))/(\sum_{k=1}^{n} P(A_k))]$ tends to 1 in probability as $n \to \infty$, where $I(A)$ denotes the indicator function of A.

(b) Show that there exists a sequence $n_j \to \infty$ such that with probability 1

$$\sum_{k=1}^{n_j} I(A_k) > \frac{1}{2} \sum_{k=1}^{n_j} P(A_k)$$

for all sufficiently large j.

(c) Show that $P(\limsup_{n \to \infty} A_n) = 1$. (10+10+5)=25

3. (a) State the inversion formula for characteristic functions.

(b) Suppose X and Y are two independent random variables having the same distribution. Show that the distribution of X is continuous everywhere if and only if $P(X-Y=0) = 0$.

(c) Let f be the characteristic function of a distribution function F. Show the equivalence of the following :

i) F is continuous everywhere

ii) $\frac{1}{T} \int_0^T |f(t)|^2 \, dt \to 0$ as $T \to \infty$

iii) $\frac{1}{T} \int_0^T |f(t)| \, dt \to 0$ as $T \to \infty$. (4+9+12)=25

4. (a) Define an infinitely divisible law. Show that if f is a characteristic function of an infinitely divisible law, then $f(t) \neq 0$ for all t.

(b) Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables. Let Z be a poisson random variable with mean $\lambda > 0$ which is independent of the sequence $\{X_n\}$. Let

$$Y = 0 \quad \text{if } Z = 0$$
$$= \sum_{i=1}^{Z} X_i \quad \text{if } Z \geq 1$$

Then show that Y is an infinitely divisible random variable. Also find its characteristic function. (...)

Please turn over

5. (a) Let $\{a_n\}$ be a bounded sequence of non-negative real numbers. Let $\{X_n\}$ be a sequence of independent random variables with $P(X_n = a_n) = P(X_n = -a_n) = \frac{1}{2}$.
Find necessary and sufficient conditions on $\{a_n\}$, for the series $\sum_{n=1}^{\infty} X_n$ to converge with probability 1. Does the series $\sum_{n=3}^{\infty} X_n$ converge with probability one, if $a_n = (\sqrt{n} \log n)^{-1}$?

(b) Let $\{Y_n\}$ be a sequence of independent identically distributed continuous random variables. Let F be the distribution of Y_1. Let $F_n(x) = \frac{1}{n}$ number of i such that $1 \leq i \leq n$ and $Y_i \leq x$. Fix an x with $0 < F(x) < 1$. Find the constants $A_n(x)$ and $B_n(x)$ such that $Z_n(x) = A_n(x) F_n(x) + B_n(x)$ tends to the standard normal variate in distribution.

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1977

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

Time: 5 hours | Full marks: [illegible]

(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 any three questions from this group)

1. Table 1 gives percentages of total expenditures incurred on clothing by rural households in different fractile groups in India during 1974–75. Construct suitable concentration curves and hence obtain the elasticity of expenditure on clothing. Interpret your result, stating your assumptions very clearly.

Table 1

Fractile groups (% of households)	Annual per capita total expenditure (Rs.)	% of total expenditure on clothing
Poorest 5%	216	1.4
5 – 10	295	1.4
10 – 20	367	1.5
20 – 30	413	2.0
30 – 40	51[illegible]	2.4
40 – 50	586	[illegible]
50 – 60	667	5.[illegible]
60 – 70	76[illegible]	5.9
70 – 80	878	7.3
80 – 90	1052	9.3
90 – 95	1260	10.7
Richest 5%	1370	14.1
All	68[illegible]	6.5

2. An inter-industry transactions matrix for an economy for the year 1961 is given below.

Table 2

Producing sector	Using sector: Agriculture	Using sector: Manufacturing	Using sector: Services	Final use	Gross output
Agriculture	[illegible]	[illegible]	100	400	[illegible]
Manufacturing	[illegible]	[illegible]	900	2000	[illegible]
Services	[illegible]	[illegible]	160	950	[illegible]
Number of persons employed	[illegible]	[illegible]	[illegible]		

The figures in the last row are in crores. All other figures are in Rs. crores.

Please turn over

2. (Contd.) i) Obtain the input-output coefficients matrix.

ii) If it is planned to produce gross output of 9000, 3250 and 4000 Rs. crores in the three sectors respectively show that the plan is infeasible. Discuss reasons.

iii) In 1974 government decided that final demand in the three sectors to be 1000, 2500, and 1000 Rs. crores, what should be the gross output in the three sectors and the total employment in the economy ?

iv) What are the implied annual percentage rates of growth of total gross output, total net output (final use) and total employment between 1971 and 1974 ?

3. The following production function was estimated by using data relating to Indian Wollen Textiles.

$$\log V = -0.026 + 0.413 \log K + 0.708 \log L$$
$$(1.255) \quad (0.161) \quad (0.138)$$

$$\bar{R}^2 = 0.94$$

where V = value of output (000 Rs.)
L = wage bill (000 Rs.)
K = book value of capital (000 Rs.)

Figures in parentheses are standard errors.

Interpret the regression coefficients and the summary statistics. Can you test for the existence of constant returns to scale from the above information ? If not, what additional information would you need ?

4. A positive-valued random variable X is said to be lognormally distributed with parameters θ and λ if the natural logarithm of X is normally distributed with mean θ and variance λ^2. Prepare rough sketches of the lognormal distributions whose parameters are.

(i) θ = 2.5, λ = 0.6

(ii) θ = 2.5, λ = 0.8

(iii) θ = 3.0, λ = 0.6

(iv) θ = 3.0, λ = 0.8

GROUP B : TECHNO-COMMERCIAL STATISTICS

(Special Paper III : Practical)

Section I : Statistical Quality Control (50 marks)

(a) Use separate answer book for this Section.

(b) Answer any two questions from this Section.

1. A study was undertaken in a gas filling plant to control the weight of gas filled into cylinders by a filling machine. Every hour four cylinders constituting a subgroup were picked up at random weighed, filled up and the weights of filled up cylinders were determined. The specification limits on the weight of filled up cylinders are 30 ± 5 (in suitable units). Observations collected on 25 subgroups in the same units give the following averages and ranges :

Please turn over

1. (Contd.)

Sub-group no.	Weights of Empty Cylinders: Average ($\bar{X}_1$)	Range R_1	Weights of Filled up Cylinders: Average $\bar{X}_2$	Range R_2
1	23.5	1.6	36.2	3.1
2	22.6	1.8	38.9	2.9
3	21.4	1.7	39.5	2.[illegible]
4	23.6	2.1	39.8	3.[illegible]
5	21.8	1.6	39.6	2.8
6	[illegible]	1.4	35.[illegible]	2.7
7	21.6	1.5	37.2	3.2
8	25.4	2.0	41.6	[illegible].1
9	23.9	1.9	[illegible].4	[illegible]
10	23.4	1.8	41.2	2.3
11	24.9	1.7	39.4	[illegible]
12	22.6	1.3	37.5	2.[illegible]
13	22.8	1.3	36.9	[illegible]
14	21.9	1.1	37.1	[illegible]
15	2[illegible].5	1.[illegible]	38.4	2.8
16	22.4	1.4	37.5	[illegible]
17	23.4	1.3	39.1	[illegible]
18	23.3	1.2	39.3	[illegible]
19	23.8	0.9	36.5	2.[illegible]
20	2[illegible].1	1.4	36.9	[illegible].5
21	25.[illegible]	1.3	35.[illegible]	[illegible].7
22	23.7	1.6	36.1	[illegible]
23	22.9	1.8	37.4	3.[illegible]
24	25.1	1.7	37.5	[illegible]
25	25.6	1.5	38.3	[illegible].1

Assume that the weights are normally distributed random variables.

(a) Make a control chart test for the homogeneity of the data on weight of empty cylinders and examine if there is significant fluctuation in the average weight of empty cylinders and in its variability.

(b) Establish trial control limits for the weight of filled up cylinders.

(c) Estimate the standard deviation of weight of gas in the cylinders (Assume that the weight of gas and the weight of cylinder are independent variables).

(d) Estimate the process capability of filling process.

(e) Suppose that the distribution of the weight of empty cylinders can not be altered. Do you think that the process capability of the filling process is good enough to meet the specifications? Give the modified control limits for the filling process if a modified control chart is appropriate. ([illegible]

Please turn over

. (a) The lower specification limit L [illegible] the [illegible]nsile strength of a certain wire is 27.[illegible] Kg's. The[illegible] is no upper specification limit. MIL Std 414 is u[illegible]d with normal [illegible]spection, code H and AQL 1%. The variance of the tensile stren[illegible]th is not known. A sample of 25 in five subgroups is [illegible]iven as follows:

Sub-group 1 :	2[illegible].27,	31.4,	31.[illegible],	3[illegible].0,	28.82
Sub-group 2 :	28.[illegible]2,	29.27,	3[illegible].17,	3[illegible].62,	3[illegible].17
Sub-group 3 :	28.[illegible]2,	31.85,	34.[illegible],	29.72,	31.[illegible]5
Sub-group 4 :	[illegible].72,	[illegible].62,	29.27,	29.72,	3[illegible].17
Sub-group 5 :	28.82	28.82	31.10	32.75	32.75

i) Use ran[illegible]e method and form II and v[illegible]ify whether the lot should be accepted.

ii) Treat the observations as 25 random [illegible]servations i[illegible]norin[illegible] sub-grouping and use standard devi[illegible]on method, form 2.

(b) The variance of a certain characteristic is unknown. The most recent 1[illegible] lots of items have been accepted using standard deviation method, lower specification limit, code letter H and [illegible]QL 1.5% MIL.Std 414 plan wi[illegible] the following % defectives. (i.e. % of items fallin[illegible] below th[illegible] lower specification limit on the characteristic under consi[illegible]ration)

[illegible], [illegible]5, [illegible]3, .77, [illegible], [illegible].44, [illegible]9, [illegible].13, [illegible].22, 1.[illegible]5

Is a shift to reduced inspection [illegible]missible ? What are the form 1 and form 2 criteria f[illegible] this [illegible]ift ? ([illegible])

3. [illegible]n experiment was c[illegible]nducted to [illegible]termine [illegible]e effects of the following factors on the [illegible]ain of [illegible]i-[illegible]nductor device.

Factor	Level	[illegible]evel 1
A. Location of [illegible]ssembly	Laboratory	Production line
B. [illegible]artial pressure of controllin[illegible] material	10^{-15}	$10^{-[illegible]}$
C. Relative humidity	1%	3[illegible]%
D. Ageing time	72 hours	1[illegible] hours

Exp. condition	(1)	a	b	ab	c	ac	[illegible]c	abc	[illegible]	[illegible]d	bd	ab[illegible]	cd
Rep. 1	39.[illegible]	31.8	47.[illegible]	4[illegible].[illegible]	43.8	27.6	3[illegible].8	45.[illegible]	[illegible].1	[illegible]2.8	54.9	3[illegible]	[illegible]
Rep. 2	[illegible]3.2	43.7	[illegible]1.[illegible]	[illegible].5	[illegible]1.2	51.3	4[illegible].9	57.5	41.7	39.[illegible]	53.[illegible]	[illegible]	[illegible]

Exp. condition	bcd	abcd
Rep. 1	35.[illegible]	4[illegible].2
Rep. 2	53.5	4[illegible].6

Perform an appropriate analysis, [illegible]timate [illegible]n effects and interpret the results. ([illegible])

[The presence of a letter (any [illegible] of a[illegible],c,d) denote[illegible] that the corresponding factor [illegible] at le[illegible] 1, the absence denotes that the correspondin[illegible] [illegible]ctor [illegible] at level 0 (1) denotes that all factors [illegible] level 0.]

Please turn over

Section II : Operations Research (30 marks)

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

(b) Answer Question No.1 and one question from the remaining.

1. The following failure rates have been observed for a certain type of light bulb :

End of week	1	2	3	4	5	6	7	8
Probability of failure to date	0.05	0.13	0.25	0.43	0.68	0.88	0.96	1.00

The cost of replacing an individual failed bulb is Rs.15/-. A decision is made to replace all bulbs simultaneously at fixed intervals and also to replace individual bulbs as they fail in service. If the cost of group replacement is Rs.3/- per bulb, what is the best interval between group replacements ? At what group replacement price per bulb would a policy of strictly individual replacement become preferable to the adopted policy ? (20)

2. At a repair shop, trucks arrive at random for service at the rate of 4 per hour. The service time is exponentially distributed with a mean of 10 minutes per truck.

(a) What is the average waiting time of a truck ?

(b) What is the average time spent by a truck in the service shop ?

(c) At any time what is the expected number of trucks waiting for repair work in the shop ? (10)

3. The following situation is observed in a rail road network at the end of a month.

Station	1	2	3	4	5	6	7
No. of cars on hand	47	82	31	20	66	-	-
No. of cars required	28	-	-	36	-	79	68

The mileage network is as follows :

Station	1	2	3	4	5
1	0	213	39	91	34
4	176	72	132	0	76
6	49	149	105	63	92
7	76	68	163	82	132

Furthermore there is no train from station 4 to station 7 and from station 3 to station 1.

Cars have to be sent by trains to the stations where they are needed minimising the total mileage.

Formulate the problem as a linear programming problem. (10)

Please turn over

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

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

(b) Answer any two questions from this Section.

1. The following is a summary of a Cash Book for the month of April 1977.

	Rs.		Rs.
Balance brought forward	1613	Payments	1262
Receipts	1469	Balance carried down	1820
	3082		3082

All receipts are banked and payments made by cheque. On investigation you discover :

(1) Bank charges of Rs.136 entered on the bank statement had not been entered in the Cash Book.

(2) Cheques drawn amounting to Rs.267 had not been presented to the bank for payment

(3) Cheques received totalling Rs.762 had been entered in the Cash Book and paid into the bank but had not been credited by the bank until May, 1977.

(4) A cheque for Rs.22 had been entered as receipt in the Cash Book instead of as a payment.

(5) A cheque for Rs.25 had been debited by the bank in error.

(6) A cheque received for Rs.60 had been returned by the bank dishonoured. No adjustment had been made in the Cash Book.

(7) During April 1977 interest on fixed deposits totalling Rs.62 was credited by the bank and no entries made in the Cash Book.

(8) A cheque drawn for Rs.6 had been incorrectly entered in the Cash Book as Rs.66.

(9) The bank statement as on 30th April, 1977 showed a balance of Rs.1162.

You are required to :

(a) show the adjustments in the Cash Book, and

(b) prepare a bank reconciliation statement as on 30th April, 1977 (15)

2. A Ghosh & Company close their financial year on 31st March. Stock-taking continues for two weeks after this date. In 1977 the value of closing stock came to Rs.20000 without making adjustments for the following :

i) Purchases made during the two weeks after 31st March 1977 were Rs.1000.

ii) Sales made during these two weeks were Rs.4000. The firm makes a gross profit of 20% on sales.

Find out the value of the closing stock on 31st March 1977. (15)

3. Write notes on any three :

a) Journal
b) Trading Account
c) Accrual Income
d) Fixed Assets

(3x5) = 15

Please turn over

Section III : Statistical Methods in Business (2/ marks)

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

(b) Answer all questions in this Section.

1. The time required by an average worker at a service counter to serve a customer is known to be normally distributed with s.d. = 4. The following table gives the service time in minutes, taken by a new employee for 12 customers.

time in minutes

customer :	1	2	3	4	5	6	7	8	9	10	11	12
time taken :	.98	1.32	1.40	1.52	1.53	0.76	1.38	1.20	1.48	1.45	.56	1.23

with $\alpha = .10$ would you accept the hypothesis that the new employee is comparable to an average worker with respect to the variability of his service time ? (10)

2. Two methods x and y of measuring surface finish (in micro mm) are being considered. Method y is standard. Method x is an innovation, expected to be cheap and can be rapidly performed. If there is a good relationship between x and y, the less costly method x can be used with standard values of y being estimated from x. Plot the graph of the following observations and determine the relationship. Comment on the relationship you obtain.

x_i	60	62	64	66	68	70	72
y_i	54.[illegible]	55.2	55.5	58.3	58.0	58.8	61.3

(10)

GROUP C : BIOMETRIC METHODS

Special Paper III : Practical

(Answer any four questions from this group)

1. (a) Consider selfing with one locus, and let $f_1^{(n)}$, $f_2^{(n)}$ respectively denote the relative frequencies of homozygotes and heterozygotes at the nth generation, n = 1, 2, 3, ...

i) Write down the generation matrix.

ii) Obtain $f_1^{(20)}$ when $f_1^{(0)} = 0.2$.

(b) Consider the following table for the genes AB in 3 backcross experiments.

Experiment	AB	Ab	aB	ab
1	57	77	112	84
2	51	81	74	46
3	70	71	70	[illegible]

i) Do these three experiments agree with each other in the frequency distribution of genotypes ?

ii) Test for homogeneity of segregations of A and a in the three experiments. (2+6+10 [illegible])

Please turn over

2. Consider a random mating population with respect to the O, A, B locus, and suppose the gene frequencies are r, p and q respectively. Frequencies observed are

O	202
A	179
B	35
AB	6

i) Find the relative frequencies of the four groups and test if the observations substantiate the model.

ii) Estimate r, p and q by the method of maximum likelihood and obtain their standard errors. (7+18)=25

3. The following table gives the results of a symmetrical parallel line assay. Assume that the regressions of response on log dose is linear.

Standard doses (μ_S)			Unknown doses (μ_U)		
0.0025	0.0050	0.0100	0.0125	0.0250	0.0500
11	9	7	13	8	2
15	13	11	13	11	10
6	6	3	12	8	8
11	11	4	14	10	11

i) Perform the usual analysis of variance with 6 treatments.

ii) Test whether the regression lines are parallel.

iii) Estimate the potency ratio.

iv) Obtain a 95% confidence interval of the potency ratio.
([illegible])

4. The following table gives the means and corrected sums of squares and products of the systolic blood-pressure (mm of Hg) Y and the age (in years) X of three groups of subjects :

Groups	Sample sizes	Means		Corrected SS & SP		
		X	Y	XX	XY	YY
1	80	26.62	90.38	137.88	74.12	127.88
2	120	28.60	92.16	30[illegible].0[illegible]	124.0[illegible]	515.67
3	70	28.86	90.6[illegible]	64.86	54.29	390.86

Examine if the linear regressions of Y on X for the 3 groups are – (i) Identical, (ii) Parallel. (16+9)=25

5. The following table gives the means and common standard deviations (s.d.) of four characters in three populations.

Population	Means			Common s.d.
	Brahmin	Arti[illegible]n	Ko[illegible]a	
Stature	164.51	160.[illegible]	15[illegible].17	5.74
Sitting height	8[illegible].43	81.[illegible]	81.16	3.2[illegible]
Nasal depth	25.40	23.[illegible]	21.44	1.75
Nasal height	51.24	48.[illegible]	[illegible].72	3.50

The common correlations matrix is

$$\begin{bmatrix} 1.000 & 0.585 & 0.177 & 0.194 \\ 0.585 & 1.000 & 0.209 & 0.217 \\ 0.177 & 0.209 & 1.000 & 0.291 \\ 0.194 & 0.217 & 0.291 & 1.000 \end{bmatrix}$$

i) Obtain a formula for clarifying a new person into one of these groups by means of observations on these four characters.

ii) Find the probabilities of misclassification by your formula in (i) (It is enough to give expressions.) [illegible]

Please turn over

6. (a) Let $W_i(t)$ denote the weight of individual i at age t (in years). The joint distribution of $W_i(t)$ and $W_i(t+1)$ is bivariate normal with mean vector (μ_t, μ_{t+1}) and covariance matrix

$$\begin{pmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{pmatrix}, \text{ for all } t \text{ and } i.$$

It has been decided to measure (100-n) individuals at age t only, (100-n) individuals at age t+1 only and n individuals at both t and t+1. Find an optimum value of n when the problem is to estimate

(i) the mean weight
(ii) the growth rate.

(b) The correlation coefficient between systolic blood-pressure and diastolic blood pressure of three groups of 30, 25 and 25 persons were 0.63, 0.48 and 0.71, respectively. Examine if the groups are different in respect of the correlation coefficient. (1[illegible]+ [illegible])

GROUP D : DESIGN AND ANALYSIS OF EXPERIMENTS

Special Paper III – Practical

(Answer any three questions from this group)

1. (a) Construct a complete set of 7 mutually orthogonal latin squares of order 8.

(b) Construct two mutually orthogonal latin squares of order 14.

2. The following data are taken from a dish washing experiment. Treatments are detergents and the response observed is the number of plates washed under standard test conditions before the foam disappeared. In the experiment, detergent solutions are made up, and plates soiled with a standard soil are washed one at a time until they are clean. The testing procedure calls for three basins to be used (i.e. for 3 treatments to be tested simultaneously), and the three operators wash at a common speed during a test. A block is thus a set of three operators testing three detergents simultaneously. The experiment is conducted with 9 detergents and a balanced incomplete block design is used. The experimental layout is given below. The letters denote treatments and the responses are given in parentheses. While recording the data it is found that the observation corresponding to treatment F in Block No. 10 is missing.

Block No.

1.	A(19)	B(17)	C(11)
2.	D(6)	E(26)	F(23)
3.	G(21)	H(19)	J(28)
4.	A(20)	D(7)	G(20)
5.	B(17)	E(26)	H(19)
6.	C(15)	F(23)	J(31)
7.	A(20)	E(26)	J(31)
8.	B(16)	F(23)	G(21)
9.	C(13)	D(7)	H(20)
10.	A(20)	F(X)	H(19)
11.	B(17)	D(6)	J(29)
12.	C(14)	E(24)	G(21)

Make an intrablock analysis of the data and test for the treatment effects.

Please turn over

3. A 3^3 factorial experiment was conducted to investigate the effects of three fertilisers N, P and K on the dry matter yield of a grass-legume pasture. The experiment was laid out in 3 blocks of 9 plots each. Treatment combinations of N(0, 40, 80 Kg/ha), P(0, 60, 90 Kg/ha) and K(0, 40, 80 Kg/ha) were randomly allotted to the plots in such a manner that a component of 3 factor interaction was confounded with the blocks. The block compositions along with the dry matter yields of the plots as weights/unit area in Kgs. are provided below. The usual convention of denoting three levels of each factor in ascending order of magnitude by the three numbers 0, 1 and 2 is followed. In any treatment combination recorded, the first number refers to the level of N, the second to the level of P and the third to that of K.

Block 1	Block 2	Block 3
012 (18.10)	010 (8.4[illegible])	221 (24.40)
102 (18.01)	112 (18.4[illegible])	212 (21.62)
222 (23.68)	211 (23.42)	110 (1[illegible].12)
000 (8.57)	220 (25.4[illegible])	122 (15.02)
111 (19.43)	022 (13.96)	011 (16.54)
120 (11.26)	121 (22.98)	020 (11.17)
201 (21.22)	100 (10.11)	101 (10.17)
210 (19.20)	202 (20.25)	002 (9.45)
021 (12.97)	001 (9.14)	200 (19.78)

The figures within parentheses give the dry matter yield values of the plots. Under suitable assumptions, analyse the data and make necessary tests of significance.

4. An experiment was conducted to study the effects of oven temperature and bake time on the life of a component. The experimental design was a split plot design with bake times as the whole plot treatments and temperatures as sub-plot treatments. Four replications were used. The data are presented in the accompanying table. Test for temperature, bake time and interaction effects.

Component Life in Minutes

Bake time	Temp	Replication I	II	III	IV
5 min.	580	217	186	15[illegible]	234
	600	158	126	1[illegible]	159
	620	229	163	167	191
	640	223	201	182	210
10 min.	580	233	201	170	178
	600	138	130	185	100
	620	186	170	181	154
	640	227	181	201	140
15 min.	580	175	195	213	178
	600	152	147	180	137
	620	155	161	[illegible]2	156
	640	156	172	[illegible]	155

NEATNESS

Please turn over

GROUP E : SAMPLE SURVEYS

Special Paper III – Practical

(Answer **any four** questions from this group)

1. The mean of a population is to be estimated by a two-sample procedure so that the estimated value may differ from the population mean by not more than .001 with a confidence coefficient 95 per cent. The following table shows the values of 20 samples drawn at random from the population. How many more samples are needed ?

Sample number	Values of random variate	Sample number	Values of random variate
1	.0725	11	.0712
2	.0755	12	.0746
3	.0759	13	.0876
4	.0739	14	.0710
5	.0732	15	.0754
6	.0843	16	.0712
7	.0727	17	.0757
8	.0769	18	.0737
9	.0750	19	.0704
10	.0727	20	.0723

In a survey for estimating the mean number of employees in all the hospitals in the United States of America the hospitals were classified into six strata according to the number of beds. The table below gives the data about the results of a sample survey based on samples selected at random without replacement independently from the respective strata.

Number of beds	Number of hospitals	Number of hospitals sampled	Sample average of number of employees	Sample standard deviation of the number of employees
Under 50	1614	25	54	25
50 – 99	1566	24	123	51
100 – 199	1419	22	262	95
200 – 299	683	10	538	152
300 – 499	679	10	912	384
500 and over	609	9	1540	826
Total	6570	100	[illegible]	

Examine if this sampling procedure should give appreciably more efficient result than might be given by a simple random sample of 100 hospitals from all the hospitals ignoring the stratification.

Please turn over

3. In order to estimate the total volume of timber in a region consisting of logs of wood scattered at 9 different places, it was decided to make use of the auxiliary information about the number of trees at the different places as shown below :

Serial no. of places	1	2	3	4	5	6	7	8	9
No. of trees	7	9	8	7	[illegible]	12	9	[illegible]	10

Using the numbers of trees in the localities as size-measures draw (giving all the necessary details) a sample of 3 places by following a scheme so that you may be sure that a ratio estimator based on it for the volume of timber be unbiased.

Suppose a person following the same scheme chooses the places with serial numbers 2, 5 and 8 in this order and on surveying observes the volumes of timber in those places to be 96, 12[illegible] and 102 cubic feet respectively. In this case obtain an unbiased ratio estimate for the total volume of timber in the entire region and obtain an unbiased estimate of the variance of the estimate. (10+5+10)= [illegible]

4. Suppose a finite population is to be surveyed on two occasions so close apart that its size and variance remain unaltered and one wants to estimate the mean on the current occasion by a suitable sampling strategy utilizing the results of both the surveys.

Suppose it is decided to have the same sample-size on both the occasions such that the first sample is an SRSWOR sample from which a certain fraction is sub-sampled on the second occasion and a completely new sample is chosen at random (WOR) from that part of the population which was not sampled earlier. Denoting by ρ the correlation coefficient between the variate values for the two occasions prepare a table to show how your choice of an optimal sub-sampling fraction is dependent on the magnitude of ρ, an estimator for the current population mean-values being suitably chosen.

[You are to briefly indicate the theoretical basis for your table without deriving algebraic results. At least 10 values of ρ must be used in your table.] ([illegible])

5. In a field of barley the grain y_i and the grain plus straw x_i were weighed for each of a large number of sampling units located at random over the field. The total produce (grains plus straw) of the whole field was also weighed. The following data were obtained :

$C_{yy} = 1.13, \quad C_{yx} = 0.78, \quad C_{xx} = 1.11$

[where C_{yy}, C_{xx} are $(C.V)^2$ for y, x respectively and C_{yx} is a similar expression.]

One requires 20 minutes to cut, thresh and weigh the grain on each unit, 2 minutes to weigh the straw on each unit and 2 hours to collect and weigh the total produce of the field. How many units must be taken per field in order that the ratio estimate for the yield of grain for the field may be more economical than the one based on the mean per unit estimate ? (25)

Please turn over

GROUP F : TECHNIQUES OF COMPUTATION

Special Paper III – Practical based on U.R. Machines

(Answer all questions from this group)

The pay roll of a certain Organisation is processed on punched card machines. The relevant input files are maintained in the form of cards. These are

(A) Staff particulars
(B) Work statement for the month
(C) Tax deductions
(D) Repayment of loans

The card designs for these four files are given below :

(A) Staff particulars

			Columns
Card design index		XXX1	1 – 4
Department			79 – 80
Section			77 – 78
Roll Number			7 – 10
Name			15 – 40
Daily rate of pay			45 – 48

(B) Work statement :

CDI		XXX2	1 – 4
Roll No.			7 – 10
Days worked			11 – 12
Gross pay	to be computed & punched		51 – 55

(C) Tax deductions:

CDI		XXX3	1 – 4
Roll No.			7 – 10
Tax deductions	..	.ise ..	11 – 14

(D) Repayment of loans :

CDI		XXX4	1 – 4
Roll No.			7 – 10
Loan repaid	Rs. &	ise ..	11 – 14

Stage 1

You are given CDI XXX1 & XXX2 cards. Arrange the two decks separately in ascending order with respect to roll no. (5)

Stage 2

Merge the two decks of stage 1, so that CDI XXX1 appear first with in each roll no. Select out unmatched cards. Return the merged & unmatched cards to the examiner. (1)

Stage 3

You are given a merged pack from Stage 2. Calculate for each roll number the gross pay (daily rate X days worked) correct to paise and punch the same in columns 51 – 55 of cards bearing CDI XXX2. Return to the examiner. (

Stage 4

Your are given a pack of cards with CDI XXX1, XXX3, XXX4 & XXX2 where CDI XXX2 cards contains the gross pay. Arrange the cards in sequential order (ascending) roll number, in such way that CDI XXX1 appear first within each roll number. Prepare the pay roll sheet according to proforma given below:

Please turn over

Department	Section	Roll no.	Name	daily pay Rs. P	days worked	Gross pay Rs. P	deductions Taxes Rs. P	Loans Rs. P	Total Rs. P	Net pay Rs. P.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)

Total number of workers : Total for each Col. 7, 10 & 11 to be printed .
Grand Total for each Col. 7, 10 & 11 to be printed.

Please turn over

GROUP G : STATISTICAL INFERENCE

Special Paper III - Practical

(Answer all questions)

1. The table below gives the estimates of the means and the common dispersion matrix of 3 characters :

x_1 : length of hind femur,

x_2 : maximum width of head in the genal region,

x_3 : length of pronotum at the peel,

for two groups of female desert locusts one in the phase gregaria and the other in an intermediate phase between gregaria and solotaria.

Means and Dispersion Matrix of Biometrical Characters of Desert Locusts

Character	Means		Dispersion matrix (based on $n_1+n_2-2=90$ d.f.)		
	Gregaria $n_1 = 20$	Intermediate $n_2 = 72$	x_1	x_2	x_3
x_1	25.80	28.35	4.7350	0.5622	1.4685
x_2	7.81	7.41		0.1413	0.2174
x_3	10.77	10.75			0.5702

(a) Test for the significance of the difference between the mean vectors of the two groups of desert locusts.

(b) Find the linear discriminant function based on the 3 characters x_1, x_2, x_3 between the two groups of desert locusts. To which group would you assign the locust with measurements $x_1 = 27.06$, $x_2 = 8.03$, $x_3 = 11.35$.

(c) Are the data consistent with the hypothesis that the linear discriminant function is $y = -3x_1 + 7x_2 + 5x_3$? (1[illegible])=[illegible]

2. $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be independent samples from $N(\mu, \sigma^2)$ and $N(\nu, \tau^2)$, respectively. Determine the sample size necessarily to obtain the power, when $\frac{\tau}{\sigma} = \Delta$, to be greater than or equal to β for the UMP unbiased test of level α for testing $H : \frac{\tau}{\sigma} = 1$ against $K : \frac{\tau}{\sigma} > 1$, where $\alpha = 0.05$, $\beta = 0.9$, $\Delta = 1.5, 2, 3$. (1[illegible]

3. The following table gives the frequencies of eggs laid by [illegible] flies in flower-heads. The count of flower-heads with no eggs is not available.

no. of eggs :	1	2	3	4	5	6	7	8	9
no. of flower heads:	22	18	18	11	9	6	3	[illegible]	1

Assuming that the number of heads follows Poisson law, estimate by the method of Maximum Likelihood the average number of eggs laid and calculate the standard error of this estimate.

Please turn over

4. Given in the table below are the means and standard deviations and correlation coefficient of scores in two halves x, y of a psychological test on 20 boys :

Score	Mean	Standard deviation	Correlation coefficient
x	45.5	9.51	
y	50.2	5.25	.70

(a) Examine whether the scores of the two halves are equally variable.

(b) Examine whether the average score of the two halves differ significantly.

(c) Examine whether the scores on the two halves are uncorrelated. Obtain a 95% confidence interval for the coefficient of correlation. (5+5+5)=[illegible]

5. The following is a random sample of size 10 from a certain population with density $f(x)$, $0 < x < \infty$:

2.7, 3.0, [illegible], [illegible], 1.7, 1.2, [illegible], [illegible], [illegible], [illegible].

On the basis of this sample, test the null hypothesis $H_0 : f = f_0 = 2e^{-2x}$, $0 < x < \infty$, against the alternative $H_1 : f = f_1 = \frac{1}{2}(2e^{-2x} + 3e^{-3x})$, $0 < x < \infty$, using the [illegible] test procedure of level α = [illegible]. Also obtain the power of this test.

6. $\{X_n, n \geq 1\}$ is a sequence of independent Normal variables with unknown mean μ and variance $\sigma^2 = 9$.

(a) Choose an SPRT to test $H_0 : \mu = 0$ against $H_1 : \mu = 2.5$ such that α = [illegible] and β = [illegible]. Calculate the power of this test for $\mu = 3$. Calculate the ASN for μ = [illegible], 2.5, [illegible].

(b) $X_1, X_2, \ldots, X_9$ were observed and the variance in this sample of size 9 was found to be 13.5. How many more observations are needed to get 95% confidence interval for μ of length [illegible].5 (15+[illegible])=[illegible]

7. Given in the following table are the weights in gms. of the anterior muscles of both the hind legs of 16 Rabbits.

Rabbit no.	Wt. in gms. Left leg	Wt. in gms. Right leg	Rabbit no.	Wt. in gms. Left leg	Wt. in gms. Right leg
1	5.[illegible]	4.9	9	5.3	5.2
2	4.8	5.6	10	5.3	5.5
3	4.3	4.3	11	5.3	5.5
4	5.1	5.3	12	5.9	5.9
5	4.1	4.1	13	6.5	6.[illegible]
6	4.[illegible]	4.[illegible]	14	[illegible].3	6.3
7	7.1	6.9	15	6.6	[illegible]
[illegible]	5.9	6.3	16	5.2	6.3

Test whether there is any difference between the left and right legs in respect of the average weight of the anterior muscle,

(a) assuming normality

(b) without assuming normality. ([illegible])=[illegible]

Please turn over

GROUP H : PROBABILITY THEORY

Special Paper III : Practical

(Answer **any four** questions from this group)

1. Answer **any three** parts :

(a) Consider the continuous parameter stochastic process $\{X(t),\ t \geq 0\}$ defined by

$$X(t) = A \cos wt + B \sin wt$$

where the frequency w is a known positive constant and A, B are independent normal variates with means 0 and variances σ^2. Find

$$P\left[\int_0^{2\pi/\omega} X^2(t)dt > C\right].$$

(b) State the axioms which an integral valued process should satisfy in order to become a Poisson process and prove your assertion. Also state when the generalized and non-homogeneous Poisson processes are obtained.

(c) Define the Wiener (i.e. Brownian motion) process and obtain its mean value function and the covariance kernel.

(d) Show that the sum (but not the difference) of two independent Poisson processes is a Poisson process, but the difference has stationary independent increments. (25)

2.(a) Obtain Kolmogorov's system of forward differential equations for a Markov process and prove that the vector of probabilities $\underline{p}(t)$ satisfies the equation $\partial\underline{p}(t)/\partial t = \underline{R}(t)\underline{p}(t)$, where $\underline{R}(t)$ is the infinitesimal transition probability matrix.

(b) Events occur independently in a certain stochastic process. The probability of an event occurring in the time-interval $(t, t+\Delta t)$ is independent of the behaviour of the process prior to t and equals $ae^{-t/b}\Delta t + o(\Delta t)$. Prove that the probability of no event occurring in the interval (u,v) is $\exp\{-ab(e^{-u/b} - e^{-v/b})\}$.

(c) In the simple birth and death process, the intensities of birth and death per individual are λ and μ respectively. Prove that the probability of there being n individuals at the time t, given that initially there are a,

$$P_n(t) = \sum_{j=0}^{\min(a,n)} \binom{a}{j}\binom{a+n-j-1}{a-1}(\mu\alpha)^{a-j}(\lambda\alpha)^{n-j}(1-\mu\alpha-\lambda\alpha)^j$$

where $\alpha = (T-1)/(\lambda T - \mu)$, $T = e^{(\lambda-\mu)t}$ (25)

3.(a) Prove that, in a symmetric random walk in one and two dimensions only, a particle will return to its initial position and will do so infinitely often .

(b) Prove that in a simple random walk on integers, the generating function of the probability $u_{z,n}$ of reaching $-z$ from the origin for the first time at the nth step is

$$[\{1 - \sqrt{(1 - 4pqs^2)}\}/2ps]^z$$

Hence or otherwise show that

$$u_{z,n} = \frac{z}{n}\binom{n}{\frac{1}{2}n + \frac{1}{2}z}(p)^{(n-z)/2}q^{(n+z)/2}$$ (25)

Please turn over

4. (a) Justify the statement that random walks are crude models for the study of the Brownian motion or other processes. In particular, obtain from this equivalence the solution of the Fokker-Planck equation of diffusion.

(b) A particle starting at the origin performs a random walk in continuous time. The successive step-lengths Y_1, Y_2, ... are independent variates with pdf $f(y)$ and they occur at the epochs of a Poisson process of rate α. Prove that the pdf $p(x,t)$ of the variate $X(t)$, indicating the position of the particle at time t, satisfies the differential equation :

$$\frac{\partial}{\partial t} p(x,t) = -\alpha\, p(x,t) + \alpha \int_{-\infty}^{\infty} p(y,t)\, f(x-y)dy$$

Hence, or otherwise obtain $p(x,t)$.

5. (a) Show that the entropy of a finite complete probability scheme defined by

$$H(X) = -\sum_{k=1}^{n} p_k \log p_k$$

possesses all the desirable properties of a measure of average uncertainty and show that this entropy function possesses a maximum.

(b) Define the different entropies for a two-port communication system and calculate them for discrete channel with independent input-output.

(c) Give a measure for mutual information $I(X;Y)$ and show that

$$I(X;Y) = H(X) + H(Y) - H(X,Y).$$

If $P(X=0) = \frac{1}{2}$, $P(X=1) = \frac{1}{2}$ and the channel matrix has rows

$$\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) \text{ and } (0, 0, 1), \text{ compute } I(X;Y).$$

6. (a) Explain the various important types of channels in vogue in information theory, giving an example of each.

(b) Prove that the information processed by a channel is a convex function of the input probabilities.

(c) State and prove Feinstein's lemma and discuss its use in establishing the fundamental theorem of information theory.
