INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May, 1975

Paper I : Official Statistics and Descriptive Statistics (Theoretical)

Time : 4 hours Full marks:100

(i) Figures in the margin indicate full marks.

(ii) Use of calculating machine is not permitted

GROUP A : Official Statistics (50 marks)

(Attempt any three questions from this group)

1. Discuss critically the usefulness of the available Indian Official statistics in economic planning. (16)

2. Give a brief critical review of the available industrial statistics in India, with particular reference to censuses and surveys conducted in this respect. (16)

3(a) Describe the main features of the National Sample Survey Organisation.

(b) What do you understand by unemployment statistics ? What are the main problems met in measuring unemployment in India ? Write down the sources of such data.
(8+8)=16

4. Make a critical review of State Income estimates in India. Describe the special problems faced in such estimation. (16)

5.(a) Write short notes on any two of the following in so far as they relate to India :

i) Price statistics

ii) Trade statistics

iii) Educational statistics

iv) Vital statistics

(b) Describe the work pertaining to statistics of any one of the United Nations specialised agencies.
(8+8)=16

Please turn over

Group B : Descriptive Statistics (50 marks)

(Attempt any three questions from this group)

6(a) Discuss the relative merits of different methods of collecting primary data in a sample survey.

(b) Why are the results of sample surveys often more reliable than those based on complete enumeration? Mention in this connection the different ways in which non-sampling errors affect statistical data. (8+8)= 16

7(a) What are Sheppard's corrections for moments? State the conditions under which these corrections are valid. When are they important?

(b) Define the Cauchy distribution and examine its properties.

(c) Derive the negative binomial distribution as a compound Poisson distribution. Mention a situation where this approach to the negative binomial is meaningful. (5+5+6)= 16

8(a) Give an account of the technique of orthogonal polynomials for fitting polynomial regressions of y on x. Is the technique available for all types of bivariate data ?

(b) Define any coefficient of rank correlation between x and y and examine its properties. Mention clearly the case of 'tied' ranks. (8+8)=16

9(a) Define the coefficient of multiple correlation $R_{y.x_1 x_2 \ldots x_k}$ and express it in terms of total correlation coefficients. Can the coefficient decrease, if a regressor X_{k+1} is added to the multiple regression equation ?

(b) Given the correlation coefficients r_{12}, r_{13} and r_{23} obtained in an analysis of trivariate data, how would you examine the internal consistency of the three coefficients ? Explain fully. (11+5)=16

10. Write short notes on any two of the following

(i) The difficulties of directly comparing the price levels in two distant periods by means of a price index number.

(ii) The moving average method of determining the trend of a time series.

(iii) Uses of the bar diagram and its variants, in representing Statistical data.

(iv) The Pearsonian system of frequency curves. (8+8)=16

NEATNESS : (Groups A & B) (4)

(M2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1975

Paper II : Probability Theory and Statistical Methods (Theoretical)

Time : 4 hours Full Marks : 100

Figures in the margin indicate full marks

Group A : Probability Theory (50 marks)

(Attempt any three questions from this group)

1. (a) When are two events independent? When is a sequence of events a sequence of independent events?

   (b) Define the conditional probability $P(A|B)$ of an event A, given the event B of positive probability. If A_1 and A_2 are any two events, show that

$$P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B) - P(A_1 \cap A_2 | B)$$

   (c) Two identical decks of N distinct cards each, are shuffled to secure randomness when matched. Obtain the distribution of the number of matchings. Show that the asymptotic mean and asymptotic variance of this distribution, are equal.

(5+4+7) = 16

2. (a) With reference to random variables (r.v.), define the terms distribution function, probability mass function and probability density function. The distribution function of a r.v. X is given below:

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

   Obtain the distribution function of $(X+1)^2$.

   (b) When are two r.v.s. independent? Give three r.v.s. which are pairwise, but not mutually, independent.

   (c) Let $a_1, a_2, \ldots, a_N$ be N positive numbers such that $\sum_{j=1}^{N} a_j = 1$. If ϕ is the probability generating function (p.g.f.) of an integer-valued r.v., show that $\sum_{j=1}^{N} a_j \phi^j$ is a p.g.f.

   Identify the r.v. of which it is the p.g.f.

(5+5+6) = 16

3. (a) Define the standard normal distribution. If X and Y are independent standard normal variates, obtain the distribution of $\frac{1}{2}(X+Y)$, $|X-Y|$ and identify it.

   (b) Let X and Y be independent, identically distributed r.v.s. with a common geometric distribution; obtain $P[X=Y]$ and $P[X \ge 2Y]$.

   (c) If X and Y are as in (b), and $U = \min.(X,Y)$ and $V = (X-Y)$ show that U and V are independent.

(5+5+6) = 16

4. (a) Define the terms (i) conditional distribution and (ii) marginal distribution for discrete r.v.s. Let X be a r.v. having the Poisson distribution with mean 2. If X=0, then Y=0; and if X=j > 0, then Y is distributed like the sum of j independent Poisson variates each with mean 2. Obtain the marginal distribution of Y.

(b) Define the bivariate normal distribution and state its properties.

(c) State and prove Chebyshev's inequality and hence derive the weak law of large numbers for independent and identically distributed r.v.s. with finite, positive variance.

(6+4+6) = 16

5. (a) State any one form of the Central limit theorem. Hence, or otherwise, obtain

$$\lim_{n \to \infty} P\left[\sum_{j=1}^{n} x_j \leq n\mu\right]$$

where $\{X_n, n \geq 1\}$ is a sequence of independent and identically distributed r.v.s. such that $E(X_1) = \mu$ and $\sigma^2(X_1) = \sigma^2$ $(0 < \sigma^2 < \infty)$

(b) Let X(i) be the smallest observation in a random sample of size n, from the uniform distribution on (0,1). Obtain the asymptotic distribution of $X_{(1)}$.

(c) If $X_1, X_2, .., X_n$ is a random sample from a standard normal distribution, obtain the distribution of $\sum_{i=1}^{n} X_i^2$.

(5+6+5) = 16

Neatness = + (2)

Group B : Statistical Method (50 marks)

(Attempt any three questions from this group)

6. (a) Explain the terms (i) error I and error II [of first and second kinds] (ii) the best critical region, (iii) composite hypothesis and (iv) similar tests.

(b) Prove that a UMP test (when one exists) is necessarily unbiased.

(c) Let X_i (i=1, 2, .., n) be independent normal random variables, such that $E(X_i) = \theta a_i$, $Var(X_i) = \sigma^2 a_i > 0$ where σ^2 is known and a_i are known and positive constants. Obtain a UMP size α test of the hypothesis $\theta = 0$ against the alternative $\theta < 0$. Does there exist a UMP size α test of the hypothesis $\theta = 0$ against the alternative $\theta \neq 0$? ?

(8+3+5) = 16

(Contd...3)

7. (a) Let X_i (i = 1, 2, .., n) be independent random variables each being distributed uniformly on the integers 1, 2, ... N, where N is unknown. Show that $T = \max(X_1, X_2, \ldots, X_n)$, is sufficient for N.

(b) Let X_i (i = 1, 2, ..., n) be n independent and identically distributed Bernoulli random variables such that $P[X_i = 1] = p$, $0 < p < 1$. What kind of functions of p is unbiasedly estimable?

(c) The independent observations x_i (i=1, 2, ..., n) come, respectively, from the n distributions with probability density functions:

$$f_i(x) = \begin{cases} \dfrac{\exp(-x/\theta)\, x^{\mu_i - 1}}{\theta^{\mu_i}\,\Gamma(\mu_i)} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

where the positive constants μ_i are known and the positive parameter θ is unknown. Obtain the maximum likelihood estimator of θ, show that it is unbiased and calculate its variance. Comment on the consistency property of this estimator.

(4+3+9) = 16

8. (a) Discuss, in detail, the important uses of the t-distribution in tests of significance.

(b) k independent samples of size $n_i \geq 2$ (for i = 1, 2, ... k) are taken from k normal populations with means m_i and variances σ_i^2. Discuss the likelihood ratio test of the hypothesis that the variances are all equal.

(c) If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two efficient estimators for a parameter θ, show that the correlation coefficient between $\hat{\theta}_1$ and $\hat{\theta}_2$ is unity.

(6+6+4) = 16

9. (a) The regression of Y on X is known to be linear in each of two different situations. On the basis of sets of independent observations, how would you test whether the two lines are parallel?

(b) Let X_i (i = 1, 2, ..., n) be n independent random variables each being distributed uniformly on the interval $(0, \theta)$ where $\theta > 0$ is unknown. Find the constants a and b, which make the confidence interval of the form $aZ < \theta < bZ$ shortest, where $Z = \max(X_1, X_2 \ldots\ldots X_n)$ and the confidence coefficient is β.

(c) Obtain E(Z) under the set up (b) and hence suggest an unbiased estimator for θ.

(6+6+4) = 16

10. Write short notes on any three of the following.

(i) sequential probability ratio tests
(ii) admissible estimators
(iii) one-way analysis of variance and
(iv) minimum chi-square method of estimation 16

MATHS (Groups A and B) (4)

(75)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1975.

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

Time : 4 hours | Full Marks : 50

i) Figures in the margin indicate full marks.

ii) Use of calculating machines is not permitted.

Group A - Sample Surveys (50 marks)

(Attempt any three questions from this group)

1. (a) For a sample of size n selected from a population of size N using simple random sampling without replacement, estimate the proportion of units in the population which possess a particular attribute. Also obtain an unbiased estimator of the variance of this estimator.

(b) A survey was conducted in a village consisting of 625 households, by covering a sample of 50 households selected by using simple random sampling without replacement, to estimate the average monthly expenditure (in rupees) on toilet goods. The estimate was found out to be 4.20 with a standard error of 0.47. Using this information, determine the sample size needed to estimate the same characteristic in a neighbouring village, on the basis of a sample to be selected by simple random sampling with replacement such that the length of the 95% confidence interval is 20% of the true value.

State clearly the assumptions involved in finding out the sample size.

(7+9) = 16.

2. Explain how you would select a sample of n units with probability proportional to its size and without replacement.

Let $y_1, y_2, \ldots, y_n$ be the values of the study variate y for the above n-selected units and $x_1, x_2, \ldots, x_n$ be the corresponding sizes. Build up an unbiased estimator for the population total of the variate y. For the special case of n=2, indicate how you would estimate the sampling error of the suggested estimator unbiasedly?

(5+6+5) = 16.

3. (a) For stratified simple random sampling (without replacement) to estimate the population mean, write down only the optimum allocation of a fixed total sample size n to the strata. (Derivation of formula not required). How does one use this allocation in practice?

(b) Let W_i be the proportion of population units for the i-th stratum, $(i = 1, 2, .., k)$, and let $S_i^2 = N_i \sigma_i^2 / (N_i - 1)$, where σ^2 is the stratum variance.

Obtain an expression for the variance of the estimate of the [illegible] population mean based on an allocation of the sample size proportional to $W_i S_i^{\alpha}$ where α is a real number in (0,2). Hence show that the allocation in (a) is not inferior to this allocation, with respect to the variance criterion.

Common on the case $\alpha = 2$. (6+10) = 16.

(7) (P.T.O)

4. (a) How would use the ratio-method for estimation of the population total of a characteristic Y when auxiliary information on a closely related characteristic X is available?

(b) Derive the conditions under which such an estimator is preferred to the conventional estimator, which does not use the information on X.

What modification would you suggest for the sampling scheme, to make the estimator given in (a) an unbiased one.

(8+8+4) = 16

5. Write notes on any two of the following:

(a) Circular systematic sampling technique.
(b) Regression method of estimation.
(c) Sources of non-sampling errors in surveys.
(d) Use of cost-function in a sample survey. (8 + 8) = 16

Group B : Design and Analysis of Experiments - (50 marks)

(Attempt any three from this group)

6. Explain how the optimum size and shape of plots and blocks for field experiments are determined with the help of uniformity trial data. Indicate also how such data can be utilised for improving the precision of treatment comparisons in a field experiment laid out at the same site, after the uniformity trial has been concluded?

(8+8) = 16

7. Explain clearly the method of analysis of data of a Latin Square experiment involving a single missing value. Give the expression for the standard error of the difference in effects between the treatment involving the missing value and any other treatment. How will you modify the analysis in case more than one observation are missing?

(8+4+4) = 16

8. Prepare a randomized layout for an experiment involving 49 varieties, using a Simple Lattice design. Using intra-block information only, give the analysis of variance, the efficiency of the design compared to randomized block designs and the formula for the variances of the differences between two varietal means.

(4+12) = 16

9. What are the principles governing the construction of a confounded factorial design? A fertilizer experiment involving three treatments N,P and K, at three levels each, is to be conducted to estimate the rate of response of wheat.

Give a design and indicate the method of analysis to be followed such that information on all the interactions become also available.

16

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

(a) Use of the principles of local control in the construction of Randomized block and Latin square designs.
(b) Fractional replication in factorial experiments.
(c) Use of transformation in analysis of data.
(d) Graeco-Latin Squares.

(8+8) = 16

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1975

Paper IV : Applied Statistics Group Papers (Theoretical).

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)

(Attempt any three questions from this group)

1. (a) Discuss the problem of aggregation in connection with the estimation of a production function.

(b) Show that under competitive market conditions, the exponents of the Cobb-Douglas production function represent the proportionate shares of the factors of production in the value of the total output.

(8+8) = 16

2. Discuss different methods of obtaining measures of seasonal variation and examine the rationale of each method. (16)

3. For an open Leontief input-output system consisting of three industries, clearly mention the conditions for equilibrium and obtain the solutions for output and price. (16)

4. What is a Lorenz curve? How do you define the measure of income concentration from a Lorenz curve? Obtain the form of this curve when income (ξ) has the following distribution:

$$P\left[\xi > x\right] = \left(\frac{x_0}{x}\right)^{\nu} \quad \text{for } x > x_0 \quad \text{where } \nu > 0$$

(16)

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

(a) equivalence of the three methods of measuring national income;

(b) problem of multi-collinearity in demand analysis;

(c) business forecasting on the basis of time series data;

(d) construction of a consumer index number.

(8+8) : 16

Neatness (2)

S.B.IV – GROUP (b) : STATISTICAL QUALITY CONTROL (Half-Paper 50 marks)

(Attempt any five questions from this group)

1. An experiment has to be conducted involving five factors, A,B,C,D,E each at two levels. Of these, the factors A,B and C are likely to interact. Construct a quarter replicate of the 2^5 factorial design, making use of the 2^3 full factorial design involving the factors A,B and C. Indicate the different effects estimated by the quarter replicate and their respective aliases. (10)

2. A particular welding operation is done by k different welders. The number of weld-joints produced by each welder during a month and the corresponding number of defective weld-joints are as follows:

Welder	1	2	....	k
No. of weld-joints produced	n_1	n_2	....	n_k
No. of defective weld-joints	x_1	x_2	....	x_k

Give a procedure for testing the hypothesis that the k welders are equally efficient in welding. State your assumptions if any. (10)

3. (a) Write a note on the scope of an $\bar{X}$-s chart for its use in quality control work.

(b) Obtain the expressions for the control limits for $\bar{X}$-s control charts (where $\bar{X}$ and s are the subgroup mean and standard deviation respectively), when the process standard deviation σ (sigma)

(i) happens to be known (from past data) (ii) is not known. (2+8) = 10

4. Describe a method of estimating the process capability (inherent variation) of a given large scale manufacturing process, indicating clearly the steps involved. Mension some uses of the estimation of process capability. (6+4) = 10

5. (a) Compare exhaustive (100%) inspection, percentage inspection and sampling plan inspection, in respect of their discriminating power between bad and good lots. Comment also upon the average quality of accepted items under the three inspection procedures.

Define:

(b) (i) AQL (ii) AOQL (iii) ATI (AOI)

in case of single sampling plan by attributes, using acceptance rectification criteria. (4+6) = 10

6. From a lot of size N, a random sample of n items is chosen. The units in the sample are inspected, to be classified into defectives and non-defectives. If the lot contains 100p% defective items and it is accepted if the number of defectives in n does not exceed a given number c, write an exact expression for the probability of accepting the lot and mention various approximations to it.

Indicate the procedure for a double sampling acceptance/screening plan by attributes. (6+4) = 10

7. Describe the salient features of the Dodge-Roming sampling inspection tables. (10)

(110)

S.B. IV - GROUP (c) : STATISTICAL METHODS IN GENETICS (Half-Paper 50 marks)

(Attempt any three questions from this group)

1. In planning a breeding programme with F_2- data, obtained from a progeny test, all the genotypes are assumed to be completely identified (including the double heterozygotes at the coupling or repulsion phase). Estimate the amount of information on recombination fraction, by crossing two heterozygotes at repulsion phases.

   Show that the information obtained from the F_2- data, is twice that for the back-cross. Why, do the breeders usually adopt the back-cross as standard, in comparing the efficiency?

   (10+3+3) = 16

2. Out of a large number of sires under progeny testing, a proportion "p" is selected on the basis of the first lactation records of "m" daughters each. What would be the expected genetic advance as a result of such selection?

   Estimate the heritability from intra-sire regression of daughter on dam, stating clearly all the assumptions involved.

   (8+8) = 16

3. By using the maximum likelihood method, obtain the estimates of the blood-group gene-frequencies of mothers and children, from mother-child combination of A-B-O blood-groups.

   What is the other method commonly followed in estimating these gene-frequencies? Describe its limitations.

   (10+6) = 16

4. Under the assumptions of absence of differential fertilisation, viability and linkage, obtain the following estimates for (n+1)th generation when F_1- plants are selfed, for n successive generations:

   (a) variance of $F_{(n+1)}$ progenies,

   (b) variance of means of $F_{(n+1)}$ progenies,

   (c) mean variance of $F_{(n+1)}$ progenies,

   (d) covariance of $F_{(n+1)}$ progeny means and F_n parents.

   (4+4+4+4) = 16

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

   (a) Role of selection and mutation in population drift.

   (b) Relative merits and demerits of Sib method and Proband method of ascertainment.

   (c) Linkage disequilibrium and viability.

   (d) Path coefficient vis-a-vis Inbreeding coefficient.

   (8+8) = 16

Contnoss (2)

S.P.IV - (3307)(A) : VITAL STATISTICS AND DEMOGRAPHY (Half-Paper 50 marks)

(Attempt any three questions from this group)

1. (a) Describe three indices used for comparing levels of mortality in subdivisions of a much larger experience, mentioning their advantages and disadvantages when applied to: (i) different occupational groups; (ii) different areas of residence.

(b) Four students were independently asked to state which statistical records in his opinion, would prove as the best indicator of the sickness experience of a population.

Student A considers, that record of sickness absence in industry offer the greatest potentialities.

Student B prefers, data collected by means of social surveys as being more representative.

Student C says, that National Insurance records are under better control and are therefore the best choice.

Student D feels that sickness cannot be accurately defined. Hence no useful purpose can be served by measuring it.

Comment on the different opinions and add your own views with reasons.

(8+8) = 16.

2. Explain briefly how a population projection could be made by the component method:

(a) The population estimates of Great Britain for the year 2002 were:

i) 52.2 millions as per the Royal Commission on Population Projection in 1954,

ii) 70.4 millions according to 1962 official Projection.

(iii) 61.5 millions as per revised 1971 official Projection.

What justification is there for making population projections, if there are such wide variations as above?

(b) How would you construct the age-structure of a stable population, supported by l_0 annual births, given its exponential growth rate and a life-table representing its mortality level.

(6+3+7) = 16.

3. (a) Discuss the different ways in which sampling can be employed in a population census.

(b) What official publications would you consult too review the marriage rates for India and how would you summarise the statistics. What are the factors influencing marriage trends to which you would like to draw attention.

(8+6+2) = 16.

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

(a) De facto and De jure methods of census enumeration and their relative advantages and disadvantages.

(b) Approximate relations between the measures of mortality, viz. q_x, m_x, and A_x of a life-table as also between curtate and complete expectations of life. (Will those relations be valid for say age 90?).

(c) Net reproduction rate as an indicator of growth of population.

(d) Different phases in growth rate through which a population passes owing to voluntary restriction of births and gradual improvements in mortality. (Which phase is suitable for use of a logistic curve?).

(8+8) = 16.

Neatness (2)

B.IV - GROUP (e) EDUCATIONAL AND PSYCHOLOGICAL STATISTICS (Half-Paper 50 marks)

(Attempt any three questions from this group)

1. Assuming the classical test theory model, derive the expression for the linear regression function of true score on observed score.

   One examinee receives a score of 60 on test 1 and a second examinee receives a score of 62 on test 2. Assume that tests 1 and 2 measure the same ability with reliability $\rho_1=0.90$ and $\rho_2=0.00$, and that $E(X_1)=E(X_2)=50$ where X_1 and X_2 are the scores, corresponding to tests 1 and 2.

   Which examinee do you think has the higher true score? (8+8) = 16.

2. Describe an experimental method for obtaining reliability of an essay-type examination.

   An examination has two parts, an objective part (o) and an essay part (e). The split-half correlation of scores for part (o) is 0.80 and the corresponding correlation for part (e) is 0.60. The correlation between the total score given by examiner A and examiner B is 0.78, for the essay part (e).
   Estimate the content reliability for (o) and (e) separately. (8+8) = 16.

3. (a) Discuss the advantages of transforming raw scores into percentile ranks.

   (b) Given that a reading test for unselected 10-year olds, yields, a mean of 50 and a standard deviation of 10, whereas an arithmetic test for the same group gives a mean of 48 and a standard deviation of 8. If a student scores 52 on reading and 50 on arithmetic, can it be concluded that he is better in reading ability than in arithmetic? Give reasons for your answer. (8+8) = 16.

4. (a) Distinguish between principal component analysis and factor analysis.

   (b) Assume a matrix of factor loadings for three tests and a criterion on four factors as shown in the table below. The last column gives the reliabilities.

Tests	Factors I	II	III	IV	Reliability
1	.4	.4	.3	.5	.74
2	.6	.2	0	.4	.68
3	.3	.6	0	.2	.90
C	.1	.7	.3	0	.87

   i) compute the communality of test 1
   ii) which test has the largest specific variance?
   iii) compute the uniqueness of test 2.
   iv) compute the table of correlation for the four variables. (8+8) = 16.

5. Write short notes on any three of the following:
   (a) Attenuation formula for giving the correlation between true scores (in terms of correlation between observed scores and the reliability of each measurement).
   (b) Suppressor variables and moderator variables, in regression analysis.
   (c) Correction for guessing in a multiple choice test.
   (d) Bi-serial and tetrachoric correlation coefficients. (16)

Neatness (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1975

Paper V: METHODS OF NUMERICAL COMPUTATION; DESCRIPTIVE STATISTICS and OFFICIAL STATISTICS (PRACTICAL)

(a) Figures in the margin indicate full marks.
(b) Use of Calculating Machines is permitted.

GROUP (A): METHODS OF NUMERICAL COMPUTATION (26 marks)

(Attempt any two questions from this group).

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

$$\begin{aligned} 3x + 2y - z + t &= 1 \\ x - y - 2z + 4t &= 3 \\ 2x + 3y + z - 2t &= -2 \\ 5x - 2y + 3z + 2t &= 0 \end{aligned}$$

(b) Find the inverse of the following variance-covariance matrix:

$$\begin{pmatrix} 63713 & 15449 & 4620 \\ 15449 & 6353 & 1056 \\ 4620 & 1056 & 648 \end{pmatrix}$$

(6+7) = 13.

2. Taking the help of the tabled values of x and e^x given below:

x	e^x	x	e^x
0.12	1.127497	0.16	1.173511
0.13	1.138829	0.17	1.185305
0.14	1.150274	0.18	1.197217
0.15	1.161834		

evaluate by making use of appropriate formulae the values of $e^{0.1255}$, $e^{0.1867}$, and $\int_{0.12}^{0.18} e^x\,dx$.

(3+3+7) = 13.

3. (a) The following table gives the values of the function f(x,y) for different values of x and y.

y ↓ \ x →	6	8	12	24
30	2.42	2.27	2.09	1.89
40	2.34	2.13	2.00	1.79
60	2.25	2.10	1.92	1.70
120	2.17	2.02	1.83	1.61

Obtain f(x,y) when x=7 and y=32, using the reciprocals of x and y as arguments for interpolation.

(b) Use any method of inverse interpolation for determining the real root of the equation $x^3-6x-11=0$ lying between 3 and 4, correct to 3 decimal places.

(6+7) = 13.

GROUP B : DESCRIPTIVE STATISTICS (50 marks)

(Attempt all questions from this group)

4. The frequency distribution of I.Q. for 309 six year old children in a certain locality is given below.

I.Q.	Frequency
40-59	3
60-79	3
80-89	17
90-99	65
100-109	69
110-119	79
120-129	37
130-139	19
140-169	12
Total	309

(a) Draw the histogram of the frequency distribution.

(b) Comment on the skewness of the distribution by computing a suitable measure.

(6+8) = 14.

5. (a) The following data relate to the number of smokers in 100 randomly chosen groups of 8 male college students each, as obtained from a recent survey conducted in a certain educational institution:

Number of smokers per group	0	1	2	3	4	5	6	7	8	Total
Number of such randomly chosen groups	15	21	32	14	9	6	3	1	0	100

Fit a binomial distribution to the data and obtain the expected frequencies.

(b) An employee in an office is supposed to report for duty at 10.30 a.m. He is marked late in attendance if he turns up between 10.50 and 11.00 a.m. and is not allowed to attend if he arrives after 11.00 a.m. During the last 180 days he reached office, he was marked late on 57 days and was not allowed to attend on 11 days. Assuming that his arrival time follows the normal distribution, estimate.

i) the number of days he arrived before time.

ii) the number of days he arrived within five minutes of the scheduled reporting time, and

iii) the number of days he arrived either before 10.45 a.m. or after 10.55 a.m., during the stated period.

(8+8) = 16.

(P.T.O)

6. From the following data, obtain:

i) the coefficient of correlation between age of husband and age of wife;

ii) the expected age of husband for a wife aged 19;

iii) the expected age of the wife for a husband aged 30:

Serial No. of couples	I	II	III	IV	V	VI	VII	VIII	IX	X
Age of husband	25	22	28	26	35	20	22	40	20	18
Age of wife.	18	15	20	17	22	14	16	21	15	14

7. EITHER

The following data relate to percentage of expenditure and percent increase in price over 1939, for different groups of expenditure for the middle-class people of Calcutta in 1957.

Obtain the general cost of living index number for 1957 with 1939, as base. A Calcuttan belonging to this income group was drawing as salary Rs.180.00 in 1939 and Rs.396.00 in 1957. Determine how much he should have received as extra allowance to maintain his pre-war (1939) standard of living.

Group		Percentage of expenditure in groups during 1957.	Percentage of increase in price in 1957 over 1939
(i)	Food	61.22	311.8
(ii)	Clothing	4.51	444.8
(iii)	Fuel & lighting	6.58	298.0
(iv)	House rent	8.97	16.9
(v)	Miscellaneous	18.72	184.5

(6+2) = 8

OR

From the time-series data given below, obtain indices of seasonal variation: in the sale of woollen goods:

Year	Summer	Monsoon	Autumn	Winter
1968	30	81	62	119
1969	33	104	86	171
1970	42	153	99	221
1971	56	172	129	235
1972	67	201	136	302

SD V Group C : Official Statistics (24 marks).

(Attempt both the questions from this group)

8. EITHER

You are required to study over a period of years the prices of important food crops in India in relation to the amount available for consumption and the population. Clearly mention :

(a) the items of statistical information you would specifically need for the purpose

(b) the official sources that will provide you with the necessary and relevant data

(c) the reliability of the data going to be consulted by you

(d) the completeness of the available data for the study or if it would require to be built-up

No actual collection of data is necessary. (4+4+2+2)=12.

OR

You are required to show the volume of goods traffic carried by the Indian Railways over years and relate them to the Index Numbers of industrial production and also Index Numbers of agricultural production for the corresponding years :

(a) from the publications placed at your disposal, collect the relevant data for seven consecutive latest available years

(b) comment on the salient features of the data collected by you.

(3+3+3)=12

9. Collect data on any four of the following topics for three consecutive [illegible] years :

(a) Consumer price index numbers in Calcutta for industrial workers

(b) Average yield rate of wheat in the Punjab

(c) Net area irrigated by Govt. canals in India

(d) Number of scholars in recognised educational institutions in India

(e) Number of newspapers and periodicals published in India.

(f) Number of passengers killed in all Railways in India in train accidents.

(4x3)=12

Note : The questions on official statistics are intended to test the candidate's knowledge and familiarity with the use of official statistics. For questions requiring actual consultation of publications, candidates will be required to make use of official publications placed on tables for the purpose, generally within a specified time period.

i) Full references to publications consulted are to be given along with answers. Any other information considered relevant is also to be given.

ii) The information furnished by the candidates should be arranged methodically and, wherever possible, in neat tabular form.

(Turn)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May, 1975

Paper VI : Statistical Methods: Design & Analysis of Experiments and Sample Surveys (P r a c t i c a l)

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)

(Attempt any two questions from this group)

1.(a) The weights of ten boys before a change of diet was introduced and again six months after the change in diet, are recorded below.

Serial No.	Weight (in lb.) Before	After
1	109	115
2	112	120
3	98	99
4	114	117
5	102	105
6	97	98
7	88	91
8	101	99
9	89	93
10	91	89

Test whether there has been any significant increase in weight of boys, as a result of the change in diet.

(b) The correlation coefficient between the scores in two halves of a psychological test, applied on five groups of 30, 20, 28, 25 and 22 students, were 0.63, 0.48, 0.51, 0.71 and 0.65 respectively.

Test if the groups are different with respect to the correlation coefficient. Assuming homogeneity of the groups also find a 95% confidence interval, for the common value of the correlation coefficient. (10+10) = 20

2.(a) Two different teaching procedures were used on two groups of students; each group included 100 students of about the same ability. At the end of the term, an evaluating team assigned 5 grades A to E to the students. The results were tabulated as follows :

GRADE	A	B	C	D	E	TOTAL
Group I	15	25	32	17	11	100
Group II	9	18	29	28	16	100

Test, at 5 percent significance level, the hypothesis that the two teaching procedures have been equally effective.

Please turn over

S.D. VI Group A : Statistical Methods (Contd.)

Q.2.(Contd.)

2.(b) Fit the curve $y = ae^{bx}$ to the following data and test if 'b' differs significantly from 1.

x :	1	2	3	4	5	6	7	8
y :	15.3	20.5	27.4	36.6	49.1	65.6	87.8	117.6

(3+8+4) = 20

3.(a) Scores on a clerical aptitude test administered to a batch of 10 clerks of the Secretariate and 9 Directorate, are given below. Examine by Wald-Wolfowitz run test, whether the two groups of clerks have the same score distribution in the population.

Scores of clerks

Secretariate clerks : 48, 40, 51, 35, 59, 52, 43, 60, 46, 55

Directorate clerks : 45, 47, 56, 42, 57, 50, 57, 62, 57

(b) An experiment was conducted in three villages selected at random to compare three varieties of rice A, B and C. In each village two plots were sown with the same variety. The following table gives the yield rate in maunds per acre:

Yield of paddy
(in maunds per acre)

Variety	Village (1)	Village (2)	Village (3)
A	16.7	38.2	23.8
	15.1	41.2	14.8
B	38.1	43.2	15.6
	35.4	45.7	18.7
C	22.8	52.6	29.4
	34.5	51.7	35.4

Test whether the yield rates of the different varieties differ from village to village. Prescribe the best variety, if any.

(7+10+3)= 20

[illegible] B : [illegible] (30 marks)

(Attempt all questions from this group)

4. EITHER

Prepare a Latin square layout with proper randomization, for a varietal trial involving five varieties. (10)

OR

Obtain the layout of BIB design with v=7, k=4, r=4, b=7, λ=2. (10)

5. EITHER

The following table gives the plan and yield of a randomised block experiment. The yield for treatment number (4) in block number II is missing (x).

Test for significant differences between the treatments and obtain the estimate of the standard error of the difference between the mean of treatment (4) and that of any other treatment.

Plan and yields

Block I	(1) 21	(3) 11	(2) 14	(4) 28	(5) 18
Block II	(3) 14	(2) 24	(1) 18	(4) x	(5) 31
Block III	(4) 17	(1) 13	(3) 24	(2) 24	(5) 23
Block IV	(5) 12	(2) 22	(1) 8	(4) 17	(3) 18

The open figures represent yield values and the fitures in parentheses, indicate the treatments. (20)

OR

The following data relate to power consumption in electric furnace heats. There were 4 factors, at 2 levels each, and the experiment was replicated 2 times. The factors considered were nature of roof A(low, high), power setting B(low,high), scrap used C(tube, plate) and charge D(700 lb, 1000 lb). The response was measured in kilowatts of power consumed per ton of melted product. Each replicate contains 4 blocks of 4 units.

Replicate I

(1)	866	abcd	752	acd	1056	ac	1029
abd	784	ab	709	bd	702	ad	966
cd	922	d	988	bc	817	b	744
abc	829	c	1017	a	946	bcd	798

Replicate II

a	840	(1)	800	b	746	ab	646
bc	771	cd	868	ac	977	abcd	714
acd	908	abd	596	bcd	700	c	954
bd	650	abc	691	ad	876	d	650

Analyse the data given above and give your considered comments on the results. (20)

Please turn over

SDVI GROUP C : Sample Surveys. (30 marks)

(Attempt all questions from this group)

6.(a) The age and sex distribution of the students of a college in the year 1974 is given below:

Age group :	below 16	16-18	19-20	above 20.
Males	43	311	323	37
Females	23	81	94	33

Draw a simple random sample of 10 students from the above population

(b) The following table gives the 1971 population (in lakhs) of ten principal cities in India.

City	Population (in lakhs)	City	Population (in lakhs)
1)Calcutta metropolitan area	70.4	6) Ahmedabad	17.4
2)Greater Bombay	59.7	7) Bangaloro	16.5
3)Delhi	36.3	8) Kanpur	12.7
4)Madras	24.7	9) Poona	11.2
5)Hyderabad	18.0	10) Nagpur	8.6

Select two different cities such that the probability that a city enters the sample is in proportion to the 1971 population.(5+5)=10

7. EITHER

In a census of non-agricultural establishments, 500 small units were listed. A simple random sample (without replacement) of 30 units (establishments) gave the following data on employment (number of persons employed).

2	2	2	2	3	5	2	3	4	5
2	4	2	6	6	7	5	4	5	3
2	2	2	2	5	2	2	2	2	2

(a) Estimate the total employment in all the 500 units and give and estimate of the standard error of the estimator in (a)
(b) In the course of investigation, 7 of the 30 selected units were noticed having loans from banks. If the first 7 figures in the first row refer to these, estimate the number :
(i) of establishments taking loans from banks in the population
(ii) the number of persons employed in such establsihments (12+8)=20

OR

A district contains 2,072 farms divided up into 5-strata on the basis of area in acres. From the N_i farms in the i-th stratum a random sample of n_i farms is taken. The number of cattle on each sample farm is determined. From this the mean $\bar{y}_i$ and the variance s_i^2 are calculated from the sample for each stratum, as shown below:

Sl.No.	Stratum based on area (acres)	N_i	n_i	$\bar{y}_i$	s_i^2
1.	under 16 acres	635	84	4.84	27.54
2.	16 -30 ..	570	125	11.63	55.84
3.	31 -50 ..	475	138	15.95	71.70
4.	51 -75 ..	303	112	23.59	192.32
5.	over-75 ..	89	41	29.61	334.93

Estimate from the data collected above :
(a) the average number of cattle per farm in the district
(b) the standard error of the estimator in (a) (12+8)=20

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1975.

Paper VII : Applied Statistics Group Papers (Practical)

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.

S.D.VII - GROUP (a) : ECONOMIC STATISTICS : (Half-paper 50 marks)

(Attempt any two questions from this group).

1. The following table gives the distribution of income among 37,369 personal-income recipients in the U.S.A. for the year 1918. Plot the data using a suitable scale and find out if Pareto's law will give an adequate fit over the entire range of the observed distribution. If not, try a suitable range and fit the Pareto law to that range.

Income class (in dollars) (1)	No. of persons (000) (2)	Income class (in dollars) (1)	No. of persons (000) (2)
0-500	1828	5001-10000	599
501-1000	12551	10001-25000	192
1001-1500	12498	25001-50000	41
1501-2000	5222	50001-100000	14
2001-3000	3065	100001-200000	5
3001-5000	1383	200001 and above	2

(25)

2. The distribution of assessed income in India is given below for the years 1951 and 1960. Examine if the inequality of income distribution has increased from 1951 to 1960.

Annual Income (in rupees)	1951		1960	
	No. of assessees (000)	Assessed income in crores	No. of assessees (000)	Assessed income in crores
Below 10000	371.2	174.2	643.0	359.7
10000-20000	61.8	84.1	155.2	212.7
20000-70000	32.7	108.2	79.1	270.9
70000-100000	2.5	20.5	6.5	53.9
100000-200000	2.4	33.3	4.6	31.3
Over 200000	1.8	154.4	2.6	234.5

(25)

3. On the basis of family-budget enquiry, the following figures for average per capita monthly total expenditure (in rupees) and average per capita consumption of cereals (in kg) for different income-groups are obtained. Estimate the Engel elasticity of demand for cereals, assuming the constant-elasticity form of the Engel curve.

Per capita monthly income	Number of persons	Average per capita	
		Monthly consumption of cereals	monthly total expenditure
0-20	123	10.5	14.9
20-50	252	13.2	32.4
50-100	173	15.8	70.7
100-150	110	16.1	121.8
150-200	62	17.2	170.2
above 200	22	18.1	240.9

(25)

S.D. VII - GROUP (b) : STATISTICAL QUALITY CONTROL (Half-Paper 50 marks)

(Attempt question one and any other two from this group).

1. EITHER

An experiment was conducted to investigate whether its tensile strength of certain component differs significantly, when two different raw materials M_1 and M_2 are used. Three batches of the component were manufactured from each raw material. Five samples were tested from each batch and the results are given below:

Tensile strength (kg/mm^2) of sample pieces

	Raw material M_1			Raw material M_2		
Batch nos.	B_{11}	B_{12}	B_{13}	B_{21}	B_{22}	B_{23}
	19	8	25	17	33	28
	14	4	21	8	27	35
	20	6	25	13	19	11
	25	17	28	21	29	15
	7	10	23	6	11	22

(a) Write down an appropriate model for analysis of variance.
(b) Analyse the data and give your findings. (4+22)=26

OR The chromium content of test samples can be determined by two methods - spectral and chemical. Following are the results for 15 test samples:

Chromium Content (%) Of Test Samples

Sample number	Spectral method	Chemical method	Sample number	Spectral method	Chemical method
1	0.62	0.64	9	0.62	0.65
2	0.50	0.51	10	0.43	0.49
3	0.60	0.58	11	0.70	0.72
4	0.52	0.51	12	0.58	0.59
5	0.74	0.72	13	0.60	0.59
6	0.49	0.50	14	0.62	0.62
7	0.57	0.56	15	0.62	0.66
8	0.46	0.45			

(a) Estimate the precision for each of the two methods adopted.
(b) Test for equality of precision, for the two methods in (a) (20+6)=26

2. For the single sampling attributes inspection plan with lot size N=200, sample size n=20 and Acceptance number c=1,
(i) draw the average outgoing quality (AOQ) curve and from the curve so obtained, read the approximate value of AOQL.
(ii) Indicate how you could obtain AOQL, without using any AOQ curve. (10+2)12

3. The 'impulse sensitivity' of fuseheads coming from a controlled process, has mean = 1.7 units and standard deviation = 0.15 units. The specifications are 1.2 units to 2.2 units. For more efficient control it is decided to take a sample of 25 fuseheads from each batch and inspect them to a narrower specification limit of 1.4 to 2.0 units. Control chart is to be maintained on "the number of fuseheads not conforming to these narrower specifications." Obtain the control limits for this chart, using the above information. (12)

4. The standard deviation (s.d.) of the measured values of a quality characteristic is 20 units. However, it is known that the (s.d.) of the error of measurement of this characteristic is 3 units.
(a) Estimate the true s.d. of the quality characteristic.
(b) How much improvement in the measuring technique, is necessary to reduce the overall s.d. to within 1% of the true s.d. value. (6+6)=12.

(M)

S.D. VII - GROUP (c) : STATISTICAL METHODS IN GENETICS (Half-Paper 50 marks)

(Attempt any two questions from this group)

1. The amount of Dutch clover in forage stand was estimated by a mechanical counter and also by eye, the estimates being denoted by X_1 and X_2 respectively. The two treatments to be discriminated were randomized in 15 blocks of two plots each, giving an analysis of variance set up as follows:

Sources	d.f.	S.S. (x_1)	S.S.(x_2)	S.P.(x_1x_2)
Between treatments	1	13.47	8.43	10.65
Between blocks	14	93.11	54.69	69.95
Error	14	20.44	6.41	4.99
Total	29	127.02	69.53	76.19

Obtain the best discriminant function (y) and the corresponding ANOVA.

If instead, a discriminant based on the sum of the sum of the variables (i.e. $z=x_1+x_2$) is set up, find out the relative efficiency of selection. Also work out the correlation coefficient between these two discriminants (y and z). (15+6+4)=25.

2. The leaves of the jute plants are, according to two segregating characteristics, classified for each of F_2 and F_3 generations as under:

generation	bitter curly	non-bitter curly	bitter non-curly	non-bitter & non-curly
F_2	696	85	74	285
F_3	833	74	56	320

If r = recombination fraction and $P=(1-r)^2$, estimate the "amount of information" and the "deviation from maximum likelihood" for the two types of families. Test whether the data from the two types of families, agree in showing the same value of 'r'. (12+8+5)=25

3.a) 47 families of 5, 6 and 7 children, were found to be segregated for albinos in the manner shown below:

Number of albinos	Size of family (number of children) 5	6	7
(1)	(2)	(3)	(4)
1	7	7	4
2	6	6	4
3	4	3	5
4	1	1	2
Total	18	17	15

Test whether the data indicate that albinism is a simple autosomal recessive.

b) A woman had left her baby in its carriage in a big city outside a supermarket. When she returned from shopping, the baby was found missing. She later found a baby which had been left at a 'foundling' home shortly thereafter, and claimed that it was her baby. Blood typing test was done, with results given below:

Mother	A	MN	Rh-negative
Father	O	N	Rh-positive
Foundling	A	M	Rh-positive

Evaluate her claim on the basis of this blood type evidence. (15+10)=25.

S.D. VII – GROUP (d) : VITAL STATISTICS AND DEMOGRAPHY (Half-Paper 50 marks)
(Answer any two questions from this group)

1. (a) From a life table prepared from census returns of U.S.S.R. (1926), the following values of L_x are quoted.

Age	L_x	Age	L_x
20	6813	23	6709
21	6780	24	6672
22	6745	25	6634

The given number surviving to age 20, (l_{20}) is 6829, and the expectations of life at age 25 $(\mathring{e}_{25})$ is 40.2 years. Complete the other columns of the life-table for ages 20-25.

(b) Calculate on the basis of the data in (a) above, the probability of one among three persons aged 20, 21, 22 dying within two years. (20+5)=25.

2. (a) The following table gives the population of Europe at successive periods. The census enumeration was made more or less on the same day of the year every time. Fit a suitable curve to the data below and indicate the goodness of it.

Year	Population (millions)	Year	Population (millions)
1650	100	1850	266
1750	140	1900	401
1800	187	1910	540

(b) State with reasons, whether the common logistic curve will be suitable for the purpose. (20+5)=25.

3. (a) The mortality experiences (in a given calendar year) of two populations, A and B, consisting of men aged 70-74 only, are as follows:

	Population 'A'		Population 'B'	
Age x	Population at age x	Death between ages x and (x+1)	Population at age x	Death between ages x and (x+1)
70	2135	68	3042	82
71	1223	43	3538	112
72	1402	56	3011	117
73	1897	87	4103	190
74	1620	82	2527	143

Express the mortality of population B at ages 70-74, as a single percentage ratio of that of Population A, by means of the following:

(i) Indirect Standardisation
(ii) Direct Standardisation
(iii) Comparative mortality index
(iv) Equivalent average death rate

(b) Comment on the differences in results. (18+7)=25.

S.D. VII – GROUP (c) : EDUCATIONAL AND PSYCHOLOGICAL STATISTICS (Half-Paper 50 marks)

(Answer any three questions from this group)

1. Score distribution of the Head Examiner in History and his 5 co-examiners, marking equivalent sets of scripts each, are available. The table below gives the mean and the standard deviation for each of the six distributions, which may be supposed to be normal. For each co-examiner work out the equivalence curve, relating any score of a co-examiner, to the equivalent score of the head-examiner.

The raw-scores of the 3 students at the top were 74, 63 and 65 respectively. Convert these scores to the head-examiner equivalents and work out the revised rankings based on the derived equivalent scores. given by coexaminers 2, 1 and 4

Examiner	Mean	Standard Deviation	Examiner	Mean	Standard Deviation
Head-Examiner	46.02	9.097	Co-Examiner (3)	38.95	11.011
Co-Examiner (1)	33.90	10.232	Co-Examiner (4)	39.12	15.958
Co-Examiner (2)	40.74	10.797	Co-Examiner (5)	40.00	12.410

(8+8)=16.

2. Three comparable forms of a test are given to 200 persons, with the following results:

	Test form(I)	Test form(II)	Test form(III)
Means	$\bar{x}_1 = 54.0$	$\bar{x}_2 = 55.5$	$\bar{x}_3 = 56.6$
Standard deviations	$S_1 = 13.9$	$S_2 = 13.5$	$S_3 = 11.4$
Correlations	$r_{12} = 0.90$	$r_{13} = 0.88$	$r_{23} = 0.86$

Do these data indicate that the tests are parallel? (16)

3. The following correlation coefficients between intelligence test scores were found, in an investigation to study the relative influences of environmental and heredity factors:

	Two brothers		Twins	
	Reared apart	Living together	Reared apart	Living together
Correlation coefficient:	0.235	0.342	0.751	0.513
Sample size	50	40	45	55

Comment on these figures, using the necessary tests of significance. (16)

4. The table below gives statistics on four tests proposed as predictors of grade-point average. These statistics are based on very large samples. So also is the reliability of the criterion estimated to be 0.72, on the basis of a very large sample. Presume further, the test length to be equal to the number of items.

Test	Mean observed score	Standard deviation of observed score	Number of items	Reliability test	Validity test
A	16.5	4.4	30	0.72	0.68
B	12.6	3.5	20	0.77	0.50
C	66.3	17.2	120	0.95	0.75
D	52.2	13.7	100	0.95	0.77

(a) If test A is lengthened to a 100 item test, what will be the revised mean score value, score standard deviation, reliability and validity of the new set?

(b) If test B is to be lengthened, to increase its reliability to 0.90, how many new items are needed? What will the new validity value be?

(c) Which of the four tests, in its present form, is best for use in predicting the criterion?

(d) Which test has the highest validity, disattenuated for unreliability in both the predictor and the criterion. (4+4+4+4)=16.

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May, 1975

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

Time: 4 hours Full marks: 100

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

b) Figures in the margin indicate full marks.

GROUP A : ECONOMIC STATISTICS

Econometrics - Special Paper I

(Attempt any five questions from this group)

1. Suppose that the random disturbances in a single equation linear regression model $Y=X\beta+u$ are auto correlated in the following manner:

$$u_t = \rho u_{t-1} + \epsilon_t \quad ; \quad |\rho|<1$$
$$E(\epsilon_t) = 0$$
$$E(\epsilon_t^2) = \sigma_\epsilon^2 \text{ for all } t$$
$$E(\epsilon_t \epsilon_s) = 0 \text{ for all } t \neq s$$

but that all the other assumptions of the classical least squares model are appropriate. Assuming further that ρ is known, explain how you would estimate the regression coefficients of the model.

What would be your procedure, if ρ is unknown? (20)

2. Explain the phenomenon of multicollinearity and state the difficulties created by it, in the estimation of single equation regression models. Explain how "extraneous" information may help in overcoming the problem. (20)

3. Examine the identifiability of the relations in the following structural equations model:

$$y_{1t} = \beta_{12} y_{2t} + \beta_{13} y_{3t} + \gamma_{11} x_{1t} + u_{1t}$$
$$y_{2t} = \beta_{23} y_{3t} + \gamma_{21} x_{1t} + \gamma_{22} x_{2t} + u_{2t}$$
$$y_{3t} = \beta_{31} y_{1t} + \gamma_{31} x_{1t} + \gamma_{32} x_{2t} + u_{3t}$$
$$y_{4t} = \beta_{41} y_{1t} + \beta_{42} y_{2t} + \gamma_{43} x_{3t} + u_{4t}$$

where y's denote endogenous variables, x's exogenous variables, u's represent disturbances and β's and γ's the unknown coefficients to be estimated. (20)

4. In the two-stage least square method, the endogenous variables on the right hand side of the equations, are replaced by their estimated values from the reduced form, after that the estimates remain the same, if, in addition, the variables on the left-hand side are also replaced by their estimates from the reduced form. (20)

(1 20) (P.T.O)

5. Describe the procedure of construction of inter-industry transaction tables and show the equivalence of the different concepts of National Income, on the basis of such a table. (20)

6. Suppose that the utility function of an individual consuming two goods X_1 and X_2 is characterised as follows:

$$U(X_1, X_2) = b X_1 + \log X_2 \qquad (b>0)$$

Given the prices of the two goods (P_1 and P_2) and the income (Y):

i) derive the individual's demand functions for the two goods
ii) calculate the income and price elasticities of demands for the two goods. (20)

7. Describe the common indices of concentration in Business and Industry. Discuss in this connection the measures based on the Lorenz Curve. (20)

8. For a Cobb-Douglas production function,

$$Q = A L^{\alpha} K^{\beta} \quad , \quad \alpha\beta > 0 \quad , \quad A > 0$$

where Q is output, L labour input, K capital input and A is a constant

(i) Show that the production function is homogeneous of degree $\alpha+\beta$. What are the "returns to scale" when $\alpha + \beta$ is (a) equal to 1, (b) greater than 1 and (c) less than 1?

(ii) Find out the elasticity of substitution between capital and labour.

(iii) Show that if $\alpha + \beta = 1$, then greater the value of α, the more labour intensive is the production process at any given marginal rate of substitution, between capital and labour.

(iv) Obtain the shares of labour and capital, in the value of output under perfect competition. (20)

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

(i) Lognormal distribution and the geographical test of lognormality

(ii) Friedman's permanent income hypothesis

(iii) Constant elasticity Engel curve, and the method of fitting it to family budget data.

(iv) The instrumental variable method of estimation. (20)

S.D. Paper VIII Group B : TECHNO-COMMERCIAL STATISTICS

Statistical Quality Control - Special Paper I

Time : 4 hours Full marks : 100

Attempt any five questions from this group

Figures in the margin indicate full marks.

1. (a) What do you understand by Process Control? Explain the same with the help of an industrial example.

   (b) What is meant by process capability? How is its evaluation used in economic control of a process?

   (c) How would you proceed to evaluate the standard rejection rate in a factory, from an adequate set of data on the number of items inspected and the number of items rejected per day?

   (6+6+8)=20

2. (a) Explain the c u s u m chart for controlling the mean of a variable, when the shift in the mean is to be detected in one direction.

   (b) Using the usual notations, derive expressions for the parameters of V-mask used for controlling the mean. (8+12)= 20

3. (a) State the acceptance and rejection criteria under variable inspection, when a single-sided specification limit for a measurable characteristic is given.

   (b) What are the published plans available for sampling inspection by variable? Describe briefly the salient features of any one of them.

   (c) What are 'tightened' and 'reduced' levels of inspection? When do you revert to tightened or reduced inspection, while operating a sampling plant under the system described by you in (b)?

   (5+10+5)= 20

4. (a) Explain Dodge's CSP-1 plan for continuous production.

   (b) For the above plan, derive the following results with standard notations:

      i) the average number of pieces inspected in a 100% screening sequence following the finding of a defect

      $$\text{is } u = \frac{1 - q^i}{p \cdot q^i}$$

      ii) the average number of pieces passed under the sampling procedure before a defect is found

      $$\text{is } v = \frac{1}{fp}$$

      iii) the average fraction of total production inspected

      $$\text{is } F = \frac{u + fv}{u + v}$$

(Question 4.(c) continued on the next page).

4. (a) For an inspector the errors of misclassifications have been found to be as follows:

i) the probability of classifying non-defective item as defective = η_1

ii) the probability of classifying defective item as non-defective = η_2

Show that if π is the proportion of defectives in a submitted lot, the effective proportion defective $\pi' = \eta_1 + \pi(1-\eta_1-\eta_2)$.

(5+10+5)=20

5. (a) What is work sampling? Explain the situations in which work sampling technique can be used.

(b) How would you plan the data collection for carrying out work sampling study, for estimating the overall utilisation of a group of similar machines in a factory?

(c) How would you analyse the data collected in the above work sampling study and evaluate the overall utilisation percentage?

(4+3+3) =20

6. (a) Explain briefly the fundamental principles of experimental designs. Discuss their advantages.

(b) Where are 'nested designs' in industrial experimentation? Give a suitable example from an industry illustrating the use of 'nested designs'.

(c) How would you analyse the results of an experiment using nested design. (7+6+3)=20

7. Write brief notes on any three of the following:

i) Group control charts;

ii) Use of 'confounding' in industrial experimentation;

iii) Variation flow analysis;

iv) Discovery sampling. (20)

S.D. Paper VIII : BIOMETRY

Group C : Biometric Methods

Special Paper I

(No candidate available)

Paper VIII Group D : DIPLOMA EXAMINATION IN STATISTICS

Statistical Aspects - Special Paper I

(Answer any six questions from V to group)

Figures in the right indicate full marks.

Time : 4 hours Full marks : 100

1. (a) Explain the concepts behind (i) fixed effect, (ii) mixed effect and (iii) random effect models, with appropriate illustrations.

   (b) Describe the method of testing non-additivity of effects in a fixed-effect model, appropriate for the observations from randomised block designs. (7+1)=16

2. In a (rxr) latin square design, two observations from the same row are missing. Obtain a method of analysis of variance of the remaining data.

   Find expressions for the variance of the difference between the estimates of two treatment effects (i) both of which are affected and (ii) only one of which is affected. (10+6)=16

3. Define square lattice designs.

   A varietal trial with 36 varieties of a crop, is to be laid out. Obtain 12 blocks for a square lattice design, with 3-plot blocks in 3 replications.

   Outline the method of intra-block analysis of the design. (4+6+6)=16

4. Define a Youden square design. Outline the method of its analysis. Discuss the advantages of such a design. (4+12)=16

5. Define factorial experiments. What are the advantages of such experiments? Explain the concept of two-factor interaction.

   Describe a method of analysis of data from a confounded factorial experiment, when each factor has two levels. (5+5+5+7)=16

6. (a) Explain (i) Confounding,
   (ii) Partial confounding and
   (iii) Balanced confounding.

   (b) Describe asymmetrical factorial experiments, indicating their particular advantages. (5+5+3+3)=16

7. Define second order rotatable designs and central composite rotatable designs.

   Describe the purposes of analysis of rotatable designs. Outline the method of analysis of data, collected from second order rotatable designs. (3+2+4+6)=16

8. Write notes on any two of the following:

   (i) Uniformity trials.
   (ii) Weighing designs.
   (iii) Analysis of covariance. (8+8)=16

Numbers (2)

(5-XII)

SD Paper VIII Group E : SAMPLE SURVEYS

Special Paper I : Theoretical Aspects

Time : 4 hours. Figures in the margin indicates full marks. Full marks : 100

(Attempt any five questions from this group)

1. (a) Describe 'Midzuno-Sen' procedure for selecting a sample of 'n' distinct units from a population consisting of N units.

(b) Show that under this procedure, the classical ratio estimator is unbiased and its approximate variance is the same as the approximate mean square error of the estimate in question, under simple random sampling without replacement. (5+7+8)=20

2. (a) What do you understand by a self-weighting design. Explain with suitable examples.

(b) Treating tehsils as strata, a stratified random sample of n villages is drawn from a district. In the i-th selected village m_i plots growing wheat are selected by simple random sampling without replacement ($i = 1, \ldots, n$).

Build up an unbiased estimate of the total area under wheat, in the district under reference, and also derive its variance. When will this design become self-weighting? (3+8+2)=20

3. As part of cattle survey, for estimating the total number of sheep in a tehsil, two procedures are suggested :

i) Draw a simple random sample of n villages, from the list of all villages constituting the tehsil.

ii) Draw a simple random sample of n villages from the list of those villages in the tehsil, which actually own and possess sheep.

Taking an unbiased estimator of the total number of sheep in the tehsil, examine the relative merits of these two procedures. (20)

4. (a) Let a population consist of N units and n be a positive integer, such that N is a multiple of n. Describe how you would proceed to select a sample of n units from the population by linear systematic sampling and build up an unbiased estimator of the population mean. Also work out its variance.

(b) If the population units are arranged first at random and then a linear systematic sample of n units is drawn, show that the resulting sample is a simple random sample. (5+7+8)=20

Please turn over

5. (a) Explain what you understand by inter-penetrating net work of samples with a suitable example.

(b) How do you use the technique of inter-penetrating samples to control investigators' bias. (5+15)=20

6. Consider the following selection procedure. Draw a unit from the population containing N units with probability proportional to size, then draw another unit from the remaining units of the population, with probability proportional to size:

(a) Work out the selection probability of drawing a sample of two units (i, j) from the population and hence derive the inclusion probabilkty of the i-th unit. Show that the sum of the inclusion probabilities of all the population units, equals to 2.

(b) Write down Hervitz-Thompson estimator of the population total and derive the Yates-Grundy form of its variance. (10+10)=20

7. Describe the technique of double sampling for stratification to build up an unbiased estimator of the population mean and derive its variance. (15+15)=20

8. Discuss the role of pilot surveys in designing large-scale sample surveys with the help of a suitable example. (20)

SD VIII Group F: Techniques of Computation

(Numerical Analysis - Special Paper I)

(No candidate available)

S.D. VIII - GROUP (G) : STATISTICAL METHODS

General Theory - Special Paper I

(Attempt any five questions from this group)

Figures in the margin indicate full marks.

Time : 4 hours Full marks : 100

1. (a) Let $x_1, \ldots, x_n$ be n independent observations from a population with density function $f(x, \theta)$, where θ is an unknown real-valued parameter:

i) State clearly the assumptions under which the likelihood equation for estimating θ will have a consistent solution with probability going to 1 as $n \longrightarrow \infty$.

ii) Show that the consistent solution of the likelihood equation in the above case is essentially unique and is asymptotically Normal.

(b) Given a random sample from the population,

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

derive the maximum likelihood estimates of μ and σ^2 and show that these maximise the likelihood absolutely. (5+3+7)=20

2. Let $x_1, \ldots, x_n$ be n independent observations from a population with density function $f(x, \theta)$, where θ is an unknown real parameter.

(a) Under suitable regularity conditions, to be stated precisely, derive the Cramer-Rao lower bound for the variance of any unbiased estimate of a parametric function $g(\theta)$.

(b) Obtain the form of the density function for which to have an unbiased estimate of $g(\theta)$, whose variance attains the Cramer-Rao lower bound. Illustrate your answer with an example.

(c) Give an example where the Cramer-Rao lower bound is not attainable. (3+13+4)=20

3. (a) Define (i) Bayes estimate and (ii) Minimax estimate of a parametric function.

(b) Let $f(x, \theta)$ be a family of density functions with unknown parameter vector $\theta \in (\Omega)$. Let $\lambda(\theta)$ be the density functions of a prior distribution of θ on (Ω). Let $g(\theta)$ be a real valued function to be estimated with squared error loss function. Derive the form of the Bayes estimate of $g(\theta)$, in this set up.

(c) Let the family of distributions in (b), be binomial with parameters n and θ, where $0 < \theta < 1$. Let the density function $\lambda(\theta) = \frac{1}{\beta(l,m)} \cdot \theta^{l-1}(1-\theta)^{m-1}$ for some fixed l and m, both positive.

Obtain the Bayes estimate of θ.

(4+10+6)=20

(P.T.O)

4. (a) Define tests with Neyman structure.

(b) Derive a necessary and sufficient condition for all similar tests to have Neyman structure.

(c) A random variable x follows a Normal distribution with unknown mean μ and unknown variance σ^2. Based on a sample of size n, obtain the uniformly most powerful unbiased test for $H_0: \mu \leq \mu_0$ against the alternative $H_1: \mu > \mu_0$. (7+7+6)=20

5. (a) Explain the concept of an optimum confidence interval for a real-valued parameter of a probability distribution.

(b) Let $x_1, \ldots, x_n$ be n independent observations from an exponential distribution with density function $\theta e^{-\theta x}$ where $\theta > 0$ is unknown. On the basis of the sample obtain an optimum confidence interval of confidence coefficient γ $(0 < \gamma < 1)$ for θ.

(c) Discuss a procedure for obtaining a confidence interval (with pre-assigned confidence coefficient) for the median of a continuous distribution. How does the procedure simplify in large samples? (4+10+6)=20

6. (a) Explain the likelihood ratio technique for constructing tests of hypotheses.

(b) Let $\underset{\sim}{x}$ be a random vector following a k-variate Normal distribution with unknown mean vector $\underset{\sim}{\mu} = (\mu_1, \ldots, \mu_k)'$ and unknown but non-singular dispersion matrix Σ. Based on a random sample of size n derive the likelihood ratio test for testing the hypothesis $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$. Derive the distribution of the test statistic under H_0. (4+10+6)=20

7. Write short notes on all the three items below:

(a) Level of significance and power in testing hypothesis.

(b) Admissibility of decision procedures.

(c) Invariance principle in decision making. (7+6+7)= 20

S.D. VIII GROUP H : PROBABILITY THEORY

Basic Probability - Special Paper I

(No candidate available)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1975

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

Time : 3 hours Full marks : 100

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

(b) Figures in the margin indicate full marks

Group A : ECONOMIC STATISTICS

(Indian Economics & Economics of Planning - Special Paper II)

Section I : Indian Economics (50 marks)

(Attempt any three questions from this section)

1. If you are asked to prepare an estimate of India's National Wealth, what are the principal items you will care to include? Indicate briefly the problems of valuation involved in preparing such an estimate. (16)

2. The Green Revolution has in fact so far touched only a minor section of Indian Agriculturists. Discuss this with reasons. (16)

3. What are the major heavy industries established in India under the Plans? Has their establishment been justified? Give reasons for your answer. (16)

4. Discuss the sailent feature of the anti-inflationary measures adopted by the Government in recent months. (16)

5. State with your arguments whether you support or reject the following:

(a) Nationalisation of India's coal industry was essential on technological grounds.
(b) The performance of Indian agriculture cannot be improved, unless large-scale co-operative farming is introduced. (9+8)=16

Section II : Economics of Planning (50 marks)

(Attempt any three questions from this section)

6. In Harrod's growth model, the growth rate depends upon two ratios. Mention the nature of these ratios. Give some idea regarding the value of these ratios for our country in recent years. (16)

7. Show that any given population growth rate calls for a certain rate of investment, given the productivity of investment. In the Indian situation, what is the likely value of this critical rate of investment? (16)

8. Discuss the nature of an inter-industry transactions table. Point out the usefulness of such a table for planning purposes, giving illustrations from Indian Plans. (16)

9. Give an account of the financial resources which are usually taken into consideration by Indian Planners. What is the significance of foreign aid in adding to plan resources? (16)

10. Examine the 4-sector growth model developed by Professor Mahalanobis as a framework for Planning in India. (16)

Neatness (Section I & II) (4)

(M30)

S.D. Paper IX - Group B : Techno-Commercial Statistics
(Special Paper - II)

Section I : Operations Research (70 marks)

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

Section II : Statistical Methods in Business (30 marks)

Section I : Operations Research (70 marks)

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

(b) Attempt any four questions from this section.

1. A plant manufactures washing machines and dryers. The major manufacturing departments are the Stamping Department, the Motor and Transmission Department, and the Final Assembly Departments for the washer and the dryer. The Stamping Department fabricates a large number of metal parts for both the washer and the dryer. Monthly department capacities are:

Stamping Dept.	:	10,000 washers or 10,000 dryers
Motor and Transmission Dept.	:	16,000 washers or 7,000 dryers
Dryer Assembly Dept.	:	Only 5,000 dryers
Washer Assembly Dept.	:	9,000 washers only.

The Stamping Department can produce parts for 10,000 washers or 10,000 dryers per month, as well as for some suitable combination amounts of washers and dryers. It is assumed that the change over of production from washers to dryers and vice-versa, does not involve any additional costs.

A similar situation exists for the Motor and Transmission Department, and the final assembly lines are separate for the two products.

The contribution to monthly profit is Rs.900/- per washer and Rs.1,200/- per dryer.

Determine the number of washers and dryers to be produced in such a way as to maximize total contribution. (17)

2. The elements of the following effectiveness matrix, correspond to the time (in hours) taken by different machine operators to finish different jobs. It is assumed that each operator can perform only one job and each job is to be done by only one operator.

		Operator					
		A	B	C	D	E	F
Job	I	7	2	2	9	4	5
	II	4	5	11	3	8	11
	III	4	10	2	10	7	11
	IV	12	9	12	10	7	13
	V	4	1	3	8	10	9
	VI	4	6	9	12	13	4

Assign the jobs to the operators in such a way, that the total time required for the jobs is minimised. (17)

(Contd...2)

3. (a) A manufacturer is required to supply a total quantity of R units at a constant rate, to his customer during time T. The costs associated with the manufacturing process are:

C_1 = the cost of holding one unit in inventory for a unit of time,

C_2 = the shortage cost for a unit per unit time,

C_3 = the set-up cost per production run.

Find the optimal number of units to be manufactured per production run.

(b) A manufacturer has to supply his customer with 24,000 units of his product per year. This demand is fixed and known. Since the unit is used by the customer in an assembly-line operation, and the customer has no storage space for the units, the manufacturer must ship a day's supply each day. If the manufacturer fails to supply the required units, he will lose the account and probably his business. Hence, the cost of a shortage is assumed to be infinite, and, consequently, none will be tolerated. The inventory holding cost amounts to 10 paise per unit per month, and the set up cost per production run is ₹.350/-.

Determine the optimal time-interval between two consecutive production runs. (10+7)=17

4. A manager has to decide whether to assign the machine repair job of a particular shop producing high-quality products to a specialised team of engineers supplied by an outside firm or assign the job to an internal team of mechanics. The manager knows that the frequency of machine breakdowns in his shop (there have been 240 breakdowns in the last 300 working hours) is distributed according to Poisson's pattern, with an average rate of 0.8 per hour. The non-productive cost of each machine, when it is waiting for a repair job, is ₹.22/- per hour.

The outside team charges ₹.[illegible]/- per hour and guarantees that due to the experience and facilities used, it can replace and fix an average of 1.5 machines per hour. The internal team can however manage at Rs.19/- and fix an average rate of 1 machine per hour (both exponentially assumed). The manager would like to know whether it is a good decision to assign the job to the external firm. Give your opinion with reasons. (17)

5. (a) State and explain Bellman's 'principle of optimality' in dynamic programming.

(b) A man is engaged in buying and selling identical items, each of which requires considerable storage space. The buying and selling prices are indicated in the table below. He operates from a warehouse which has a capacity of 500 items. He can order on the 15th of each month, for delivery on the first day of the following month. During a month, he can also sell any amount upto his total stock on hand.

	January	February	March
Cost prices (₹)	150	155	165
Sales prices (₹)	165	165	185

If he starts the year with 200 items in stock, how much should he plan to purchase and sell each month, in order to maximise his profit for the first quarter of the year? (6+11)=17

6. Write notes on any three of the following:

i) Replacement Problems ii) Erlangian Distribution of Service Time
iii) Monte Carlo Methods iv) Transportation Technique as Linear Programming. (17)

S.D. Paper IX Group D : Section I (Alternative)

ELEMENTS OF BOOK-KEEPING AND ACCOUNTANCY (70 marks)

(Attempt Question 1 and any other three from this section)

i) Use separate answer book for this section.

ii) Figures in the margin indicate full marks.

1. Sri. Samal submitted the following Trial Balance, which he has not been able to agree:

	Dr. Rs.	Cr. Rs.
Capital		25,000
Drawings	3,250	
Stock on 1-1-1974	17,425	
Returns Inward		554
Carriage Inward	1,340	
Deposit with A.Biswas		1,375
Returns Outward	340	
Loan given to S.Paul @ 5% on 1.1.74		1,000
Interest thereon		25
Rent	920	
Rent due	130	
Stock (31.12.74)		17,792
Purchases	12,970	
Building	10,000	
Debtors	4,000	
Goodwill	1,000	
Insurance	730	
Creditors		3,000
Advertisement Expenses	900	
Reserve for Doubtful Debts		1,000
Bad Debts	400	
Patents & Patterns	500	
Cash	56	
Sales		17,914
Pre-paid Insurance		350
Wages	754	
Carriage Outward		725
	55,095	79,915

a) Rewrite the Trial Balance correcting the mistakes committed by him.

b) Prepare a Trading and Profit and Loss Account after giving effect to the following adjustments noted below:

i) Increase Bad debts by Rs.600. Reserve for doubtful debts is to be 10% and Reserve for Discount on Debtors at 5% and on creditors at 10%. 1

ii) Advertising expenses are to be spread over a period of 4 years. (12+5+8)=25.

2. What is Trial Balance?

Describe the errors which are not detected in Trial Balance. (15)

5. Discuss the distinction between any three of the following items:

i) Consignment and Sales.
ii) Fixed and Floating Assets.
iii) Intangible and Fictitious Assets.
iv) Reserve and Provision. (15)

4. The following irregularities were detected in the book of a merchant:

i) Rs.500 paid for rent, was debited to Landlord's account.
ii) A purchase of goods for Rs.550 from Z, has been posted to the debit of the creditors account, as Rs.55.
iii) Discount allowed Rs.75, was credited to Discount A/c.
iv) A cheque returned by Bank as dishonoured, debited to Discount Account for Rs.200.
v) A sale of Rs.2500 was posted to his personal A/C as Rs.250.

Give the entries, necessary to correct the irregularities. (15)

5. Write notes on any three of the following:

i) Proforma Invoice.
ii) Del Credere Agent
iii) Contingent Liability
iv) Purchase Day Book
v) Capital Reserve
vi) Capital Redumtion Reserve (15)

6. Define Secret Reserve. Is it exhibited in the Balance Sheet? If so, on which side of the Balance Sheet? (15)

SD IX : Group B : Techno-Commercial Statistics

Sec. II - Statistical Methods in Business (30 marks)

(Attempt any two questions from this section)

USE A SEPARATE ANSWER BOOK FOR THIS SECTION.

1. A footwear manufacturer produces different varieties of footwear in different sizes wishes to schedule his monthly production for the next six months. How will you plan a survey to forecast the demand of his footwear in a metropolitan city? (15)

2. (a) Explain the factors to be considered in the evaluation of a job.

(b) While conducting a work sampling study in a spinning mill, it was found that on an average 20 per cent of the time, the units were idle and unattended. This was on the basis of 400 observations. Are these observations sufficient to predict that the idle time proportion estimated is within 10 per cent of the actual idle time with 95 per cent confidence? (10+5)=15

3. Write notes on any two of the following:

(a) Payment by Results.
(b) Sampling Techniques in Auditing.
(c) Market Surveys. (15)

(M43)

IX Group C : BIOMETRIC METHODS - Special Paper II

Statistical methods in Genetics and Bio-assays)

(No candidates available)

B.D. Paper IX - Group D : DESIGN AND ANALYSIS OF EXPERIMENTS - May 1977

Combinatorial Aspects - Special Paper II.

(Attempt any five questions from this group)

1. (a) Show that the polynomial $f(x)=x^2+x+1$, is irreducible over GF(2).
   (b) Use the above polynomial to construct $GF(2^2)$ and verify that x is a primitive element. Write down the sum and product tables.
   (c) Show that the number of mutually orthogonal latin squares of order s is at most s-1.
   (d) Write down a complete set of mutually orthgonal latin squares of order 4 (four). (5+5+5+5)=20

2. (a) Define a Balanced Incomplete Block Design (BIBD), with parameters v,b,r,k and λ.
   (b) Show that $b \geq v$.
   (c) Show that any two blocks of a Symmetric BIBD (SBIBD), have the same number of treatments in common.
   (d) Define the dual of a BIBD. Show that the dual of a SBIBD is a SBIBD with the same parameters. (5+5+5+5)=20

3. (a) Define the finite Projective and Euclidean Geometries PG(n,s) and EG (n,s) where $s=r^p$ (r is a prime). Show that the number of n- flats ($0 \leq n \leq n-1$) in PG(n,s) is $\phi(n,n,s)$ where:

$$\phi(n,n,s) = \frac{(s^{n+1}-1)(s^n-1)\ldots\ldots\ldots\ldots(s^{n-n+1}-1)}{(s^{n+1}-1)(s^n-1)\ldots\ldots\ldots\ldots(s-1)}$$

   and the number of n-flats in EG(n,s) is $\phi(n,n,s) - \phi(n-1,n,s)$.

   (b) Considering the 'hyperplanes' in PG(n,s) as blocks and 'points' as treatments, show that we can construct a symmetrical balanced incomplete block design with parameters.

$$v = \frac{s^{n+1}-1}{s-1}, \quad k = \frac{s^n-1}{s-1} \text{ and } \lambda = \frac{s^{n-1}-1}{s-1} \qquad (10+10) = 20$$

4. (a) Discuss the concept of fractional replication in factorial designs. Give a plan for a $\frac{1}{2}$ replicate of a 2^6 experiment, where all interactions involving three or more factors, are assumed to be zero, but two-factor interactions must be estimable.

   (b) In a 2^6 factorial experiment, identify all the confounded interactions if the key-block is :

(1)	bce	bcf	ab
ade	bde	bdf	aef
ace	cdef	cd	adf
abcf	abcdef	abcd	af

   Develop the plan completely.

(10+10) = 20

(P.T.O)

(a) Define a resolvable balanced incomplete block design and show that for such a design $b \geq v+r-1$
When is such a design called 'affine' resolvable?

(b) Show that for an affine resolvable balanced incomplete block design,

(i) the parameters can be expressed as

$$v = nk = n^2\{(n-1)t+1\},\ b = nr = n(n^2t+n+1),\ r = nt+1,\ \lambda = t+1.$$

Where n and t are integers and $n \geq 2$, $t \geq 0$

(ii) The number of treatments common to any two blocks of different sets is:

$$\frac{k^2}{v} = (n-1)t+1 \qquad (8+8+4)=20$$

3. (a) Define an orthogonal array (N,K,s,t) of index λ.

(b) Give a method of constructing orthogonal array $(s^2, s+1, s, 2)$ of index 1, where s is a prime power.

(c) Show that the orthogonal array $(s^2, s+1, s, 2)$ of index 1, is equivalent to s-1 mutually orthogonal latin squares of order s.
(4+3+3)=20

7. (a) Define a Hadamard matrix of order n.

(b) Show that if a Hadamard matrix of order n exists, then n is either 1 or 2 or a multiple of 4.

(c) Show that the existence of a Hadamard matrix of order 4t, is equivalent to the existence of a symmetric balanced incomplete block design with parameters $(4t-1, 2t-1, t-1)$. (5+5+10)=20

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

(a) Falsity of conjectures of Euler and Mansish.

(b) Difference sets and their applications to the construction of BIB Designs.

(c) Rotatable Designs. (10+10)=20

S.D. IX Group B : SAMPLE SURVEYS : May 1975

Organisation Aspects - Special Paper II

Time : 2 hours Full marks : 100

(Attempt any five questions from this group)

(i) Figures in the margin indicate full marks.

(ii) Use of Calculating Machines is not permitted.

1. It is proposed to conduct a sample survey in India to collect comprehensive data on livestock number owned and possessed by the households, livestock enterprises of the households and consumption of livestock products by the households.

Enumerate the various points to be considered for planning the survey and your recommendations in respect of such points, with proper justification. (3+12)=20

2. In a multi-subject sample survey to be conducted in India, it is proposed to collect data on the following subjects:

Subject schedules:

i) Household consumer expenditure.
ii) Employment & unemployment of household members.
iii) Weekly price of selected commodities (on each Saturday) from the market,
in the place of visit of the Investigators (on Saturdays).

The informations have to be collected in different samples of households for each subject, within each sample village and sample urban block.

The sample villages and sample urban blocks will be selected from the whole of India and will be divided into four sub-sets to be canvassed in four sub-rounds of three continuous calendar months each. The total survey period is one full year.

It is stated that the field work will be done by an organisation which has 300 primary field Investigators with necessary supervisory staff. Explain the method of fixing the work load to the Investigators. Illustrate your method by estimating the work load per investigator in this survey. Assume any reasonable figure for any factor that you may require for your calculation. (12+8)=20

3. Explain how you would proceed step by step in preparing a schedule for a sample survey of households, to study the level of living of working class households in rural areas in India.

Illustrate your steps by mentioning the broad outlines of the schedule for this survey. (10+10)=20

4. What should be the contents of a well prepared report on a sample survey? Illustrate your case by mentioning the good points and drawbacks of a report on a sample survey, which you have read. (Give reference to the same like title, organisation issuing the report, year of issue etc. of the report taken for illustration).

5. (a) Explain what is meant by 'cost function', 'variance function' and optimum sampling design for maximum precision in two-stage sampling. Give illustrations.

(b) Give your arguments for and against the use of these principles partly or wholly, in conducting sample surveys. To what extent are these actually useful in practice? (4+4+4+4)=20

(M46) (P.T.O.)

6. What are pilot surveys? What are their various uses in sample surveys? Give examples of any actual pilot surveys (with which you may be familiar) which have been put to such uses. (3+4+5)=20

7. a) The National Sample Survey Organisation is conducting large scale sample surveys in India in the form of annual rounds. In connection with this work a suggestion to decentralise the pre-tabulation scrutiny of schedules and get the same done in the various field offices in the country and centralise only the subsequent mechanical tabulation work.

Give your considered views for and against this proposal from the point of view of cost, accuracy and speed of publication of the reports etc.

b) In your opinion, can any further decentralisation of tabulation be conveniently thought of? Give reasons for your views.

c) Note: It is to be remembered that the final tables are to be built up for N.S.S. report separately for each State of India and also on All-India basis under the rural sector, urban sector and rural plus urban sector. (20)

8. Write notes on any four of the following:

i) Sampling & non-sampling errors and bias in estimates.
ii) Self weighting sampling design
iii) Systematic sampling
iv) Interpenetrating sampling
v) Mail enquiry vs interview method of data collection
vi) Card design (5+5+5+5)=20

S.D. IX - GROUP F Techniques of Computation (Practical)
Numerical Analysis - Special Paper II
(Candidates not available)

S.D. IX Group G : STATISTICAL INFERENCE

Special Topics - Special Paper II

Time : 4 hours Full marks : 100

(Attempt any five questions from this group)

Figures in the margin indicate full marks.

1. a) Describe Wald's Sequential probability ratio test for testing a simple hypothesis against a simple alternative and prove that the test terminates with probability one.

b) Derive the OC function of the sequential probability ratio test.

c) Consider the sequential probability ratio test for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, where θ is the mean of a normal distribution with variance 1. If $\alpha = \beta$ (in usual notations), prove that the ASNs under H_0 and H_1 are equal.

(14+6)=20

(Contd...)

2. (a) Describe the Wilcoxon (Mann-Whitney) test for the two-sample problem, mentioning clearly the underlying assumptions, the hypothesis to be tested and the alternative hypothesis.

(b) Show that the Wilcoxon and the Mann-Whitney test statistics are linearly related.

(c) Obtain the null distribution of the Wilcoxon test statistic when both samples have size 2.

(d) Mention any other non-parametric test for the two-sample problem and discuss the merits of this test relative to the Wilcoxon test. (5+5+5+5)=20

3. (a) If $\underset{\sim}{m}$ is a real n-vector and Σ is a positive semi-definite real symmetric matrix, show that the function

$$\phi(\underset{\sim}{t}) = \exp\left[i\underset{\sim}{t}'\underset{\sim}{m} - \tfrac{1}{2}\underset{\sim}{t}'\Sigma\underset{\sim}{t}\right]$$

defined on R^n, is the characteristic function of some distribution in R^n.

(b) Let $\underset{\sim}{X}' = (X_1, \ldots, X_n)$ follow a multinormal distribution with a non-singular covariance matrix Σ.
Show that (i) $X_1, \ldots, X_n$ can be expressed as linear functions of n independent normal random variables; and (ii) has the
(ii) $(\underset{\sim}{X}-\underset{\sim}{m})'\Sigma^{-1}(\underset{\sim}{X}-\underset{\sim}{m})$ has the chi-square distribution with n degrees of freedom, where $\underset{\sim}{m}$ is the mean vector of $\underset{\sim}{X}$.

(c) Show that uncorrelated multinormal random variables are independent. (7+4+4) = 20

4. What is a linear discriminant function? Bring out the relationship between sample discriminant function and Hotelling's T^2 statistic. Explain how you will test for an assigned discriminant function.
(4+9+7) = 20

5. (a) Obtain the likelihood-ratio test statistic for testing the equality of the covariance matrices of two multinormal populations. Obtain the i-th moment of your test statistic.

(b) Derive the likelihood ratio test statistic, for testing the equality of several multinormal populations with common dispersion. (9+4+9) = 20

6. (a) Define the principal components. Given the covariance matrix Σ, describe and justify the iterative procedure for obtaining the first principal component.

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

(c) Discuss the use of principal components in factor analysis.
(5+6+5+4+4)=20

7. Define U-Statistics for one and two samples. Obtain the limiting distribution of the two-sample U-statistic.
Illustrate the use of U-statistics in some well known non-parametric tests. (7+11+5) =20

8. Write notes on any two of the following:

(a) Stein's two sample procedure.
(b) Tests of randomness.
(c) Mahalanobis distance.
(d) Tolerance limits (7+8) (10+10)=20

I [illegible] STATISTICAL INSTITUTE

Statistician's Diploma Examination - [illegible], 1975.

Paper X - Subjects of Third paper of Special Section - (Practical)

Time : 3 hours Full marks : [illegible]

i) Candidates are required to answer questions from that group only for which they have registered their options.
ii) Figures in the margin indicate full marks.
iii) Use of calculating machine is permitted.

Paper X - Group A : ECONOMIC STATISTICS

SPECIAL PAPER XII - PRACTICAL

1. The following table based on the 7th round of the National Sample Survey (NSS) gives the average monthly per capita expenditure on foodgrains only and on all items, in urban areas by consumer expenditure classes.

Estimates for Urban India (NSS seventh round)

Monthly per capita total consumer expenditure (rupees)	Percentage of population	Monthly per capita average expenditure (rupees)	
		Foodgrains only	All items
0-8	7.52	3.73	6.24
8-11	12.09	4.45	9.36
11-15	13.56	5.15	11.92
15-15	9.29	5.90	13.01
15-18	11.53	6.35	13.77
18-21	10.44	6.94	19.91
21-24	7.79	7.40	22.53
24-30	9.32	8.71	25.32
30-54	5.41	9.37	30.67
34-43	7.15	8.44	33.15
43-54	4.36	7.38	40.70
above 54	3.51	8.00	70.53
all classes	100.00	6.33	23.31

Fit a constant elasticity curve $\log Y_i = \alpha + \beta \log X_i + \epsilon_i$ ($i = 1, 2, \ldots, n$ where n is the number of individuals covered and Y_i and X_i stand for monthly expenditures on foodgrains and on all-items respectively) to the data under the usual assumptions for ϵ_i's using weighted least squares.

Also estimate α and β, ignoring the percentage distribution of income in different classes and compare the results as derived by these two methods.

Work out also the standard errors of α and β and test if β is significantly different from zero. Comment on the results. (15+9+34)=58.

2. Using probits, fit a log-normal distribution to the data on the total consumer expenditure (Q. 1). Plot the Lorenz curve for the fitted lognormal and work out the concentration ratio for this curve. Plot the Lorenz curve for the actual consumer expenditure data and compare it with the Lorenz curve already drawn. (10+12+10)=32

(M49) (Contd...2)

3. Using Durbin's two stage estimation procedure, obtain estimates of α and β in the model

$$X_t = \alpha + \beta t + \epsilon_t \quad \text{(where } u_t \text{'s have zero mean, variance } \sigma^2 \text{ and are}$$

$$\epsilon_t = \rho \epsilon_{t-1} + u_t \quad \text{independent over time)}$$

From the following production data, alternatively, use any other method of estimation which you may know.

Year	X_t	Year	X_t	Year	X_t
1930-31	93	1964-65	240	1969-70	545
1931-32	145	1965-66	235	1970-71	713
1943-55	173	1966-67	303	1970-71	800
1963-64	222	1967-68	367	1971-72	952

[Where X_t = All-India production of Nitrogen fertilizers (000 tons of nutrients)]

Test if ρ is significantly different from zero. Comment on the data and on the results obtained. (3+4+4)=32.

4. Obtain two stage least squares estimates of α and β in the following model from the data given below:

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

$$X_t = \alpha_1 + \beta_1 X_{t-1} + \beta_2 t + \eta_t$$

It is assumed that

$$0 = E(\epsilon_t) = E(\eta_t) = E(\epsilon_t \eta) = E(\epsilon_t \epsilon_{t'}) = E(\eta_t \eta_{t'}) \quad \text{where } t \neq t'$$

and $E(\epsilon_t^2) = \sigma_t^2$, $E(\eta_t^2) = \sigma_\eta^2$ for all t

Year	Total Revenue from Union Excise and Sales Tax (Rs. Abjes) Y_t	National Income at current prices (Rs. Abjes.) X_t
1950-51	1.5	95.5
1951-52	1.4	90.7
1952-53	1.4	93.2
1953-54	1.6	101.0
1954-55	1.8	98.1
1955-56	2.2	99.8
1956-57	2.7	113.1
1957-58	3.9	115.9
1958-59	4.4	126.0
1959-60	5.0	133.5
1960-61	5.7	141.4
1961-62	6.7	141.0

Work out also the asymptotic variance of the estimate of β

Comment on the data and the results obtained (7+4+4)=32

(M50) (Contd...3)

5. Prepare a note useful for economic analysis and policy purposes, making any calculation or drawing any graphs that you think fit, on the basis of the following Indian Economic data:

Population, net availability, procurement and public distribution of food grains in India.

Year	Population in millions	(Figures in million tonnes) Net production of food grains	Import	Net availability of food grains	Procurement	Public distribution
1964	471.5	70.57	6.27	73.07	1.43	9.37
1965	482.0	73.17	7.40	71.58	4.05	10.03
1966	491.7	61.23	10.76	73.41	4.01	11.09
1967	503.6	74.05	8.67	75.33	6.45	12.07
1968	514.9	95.15	5.39	98.70	6.91	10.23
1969	528.4	92.24	5.07	95.75	6.85	9.59
1970	539.5	87.03	5.35	89.49	8.71	9.94
1971	550.2	91.83	2.05	91.50	9.13	7.93
1972	561.9	81.57	0.50	82.07	7.70	10.73

(Net availability = Net production + Net imports + change in Government stock). (52)

B.D. Paper X - Group B : TECHNO-COMMERCIAL STATISTICS

Section I : Statistical Quality Control (50 marks)

(Attempt question no. 1 and any other two questions)

1. The following table gives the average and ranges in samples of size 4 of test records of copper content in commercial brass sheets.

Sample No.	$\bar{X}$	R	Sample No.	$\bar{X}$	R
1.	11.10	0.8	16.	11.65	1.7
2.	11.70	1.2	17.	11.58	1.3
3.	11.35	1.0	18.	9.93	0.4
4.	11.25	1.0	19.	10.70	1.2
5.	11.20	2.0	20.	11.75	0.7
6.	11.00	0.8	21.	10.95	1.7
7.	11.30	1.0	22.	11.50	1.7
8.	11.35	1.2	23.	10.73	0.7
9.	11.50	2.0	24.	10.93	1.1
10.	10.33	1.1	25.	11.23	1.9
11.	10.25	1.0	26.	10.30	0.4
12.	11.53	1.2	27.	11.30	2.0
13.	11.15	0.3	28.	11.33	1.5
14.	11.23	1.0	29.	11.25	0.2
15.	11.00	0.3	30.	11.30	0.4

a) Using control chart technique, test whether the process is under statistical control.

b) A minimum of 9% copper in any sheet is the market specification. Excess of 0.1% on an average, results in a loss of Rs.9000/- per year to the factory. Estimate how much saving can be effected by maintaining statistical control at a proper level, so as to satisfy market specifications.

(12+9)=20

2. a) The tolerance specified for outer diameter of a shell is 4"±0.005". Shells having diameter higher than specified, would be reworked while those with lower diameter would be scrapped. The cost of reworking 5 shells is equal to the loss incurred by scrapping one shell. The process standard deviation was 0.002".

Where should the process be centred so that the total cost of rework and scrap is minimum?

b) The following table gives the data on a measurable characteristic, with specification limits 0.7329 - 0.7571 cm. The sample size is four:

Sample No.	$\bar{X}$	R	Sample No.	$\bar{X}$	R
1.	0.7540	0.0011	9.	0.7547	0.0007
2.	0.7542	0.0014	10.	0.7549	0.0015
3.	0.7542	0.0009	11.	0.7541	0.0017
4.	0.7548	0.0010	12.	0.7542	0.0010
5.	0.7550	0.0008	13.	0.7545	0.0011
6.	0.7539	0.0009	14.	0.7543	0.0009
7.	0.7541	0.0012	15.	0.7551	0.0011
8.	0.7543	0.0011			

Set up a modified $\bar{X}$ and R control chart and comment. (3+9) = 15

Contd....

3. (a) A measure of discrimination power of a sampling plan between good and bad quality, is the ratio $p_{.10}/p_{.05}$ where p_a is the value of proportion of defectives in incoming lots, for which probability of acceptance is 'a'.

(c) Using poisson probabilities, prepare a table of values of the ratio with acceptance number c = 0, 1, 2, 3, 4 and 5.

(b) Given AQL = 2% Producer's Risk = 0.05
LTPD = 8% Consumer's Risk = 0.10

Obtain an appropriate single sampling attributes inspection plan, assuming that lot size is large compared to the sample size. (10+5)=15.

4. In a first order rotatable design for 2 factors, a 2^2 design with 4 central points was run. The data obtained are given below:

Response (y)	Levels of Factor 1 (x_1)	Levels of Factor 2 (x_2)	Response (y)	Levels of Factor 1 (x_1)	Levels of Factor 2 (x_2)
20.0	-1	-1	32.9	0	0
35.1	+1	-1	30.9	0	0
32.9	-1	+1	31.1	0	0
71.9	+1	+1	31.0	0	0

Estimate the first order coefficients, the interaction coefficient and the sum of quadratic coefficients. Carry out an ANOVA test to determine if there is lack of fit of first order model and offer your comments. (15)

Group I : Section II : Operations Research (30 marks)

(Attempt any two questions from this section)

Figures in the margin indicate full marks

1. Solve the following linear programming problem by two-phase method:

Maximize $x_0 = x_1 + 2x_2 + 3x_3$

Subject to $x_1 + x_3 \leq 5$

$x_1 + 2x_2 \geq 8$

$x_1 + x_2 + 2x_3 = 10$

$x_1, x_2, x_3 \geq 0.$ (15)

2. An electric company which generates and distributes electricity conducted a study on the life of poles. The appropriate life data are given in the following table:

Life data of electric poles

Year after installation :	1	2	3	4	5	6	7	8	9	10
Percent of poles failing :	1	2	3	5	7	12	20	30	18	4

(Contd...)

2. a) If the company now installs 5000 valves and follows a policy of replacing valves only when they fail, how many valves are expected to be replaced each year during the next ten years?

To simplify the computations you can assume that failures occur and replacements are made only at the end of a year.

b) If the cost of replacing individually is Rs.130/- per valve and if we have a group replacement policy it costs Rs.70/- per valve, find out the optimum period for group-replacement. (7+8)=15

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

		Cost in Rs./kg of product delivered				Maximum capacity of all products (kgs/week)
		A	B	C	D	
Factory 1	Normal time	1.3	1.5	1.4	1.2	70,000
-do-	Overtime	1.4	1.6	1.5	1.4	14,000
Factory 2	Normal time	1.3	1.4	1.5	1.4	90,000
-do-	Overtime	1.5	1.6	1.7	1.7	11,000
Requirements (kgs/week)		50,000	40,000	50,000	60,000	

Find out the amounts which each factory should produce by 'regular production' and 'overtime production', so as to meet the demand at minimum total cost. (15)

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

(Attempt both the questions from this group)

1. A certain drug is claimed to be effective in curing colds. In an experiment, which was a part of a market study, 24 people with cold were given drug and another 24 people with cold were given sugar pills. The reactions of the patients are recorded as follows:

	Relief	No effect
Drug	52	32
Sugar pills	48	38

Carry out statistical analysis of the data. (10)

2. Six different brands of whisky were ranked by two testers independently.

Brand	1	2	3	4	5	6
Rank by tester A	5	2	3	3	1	4
Rank by tester B	3	4	2	5	1	6

Test whether there is significant agreement between the testers and offer your comments. (10)

For Group B Section II - see next page.

(Alternative)

S.D. Paper X Group B : Section II (Alternative)

ELEMENTS OF BOOK-KEEPING AND ACCOUNTANCY - Practical (Marks 30)

(Attempt any two questions from this section)

Figures in the margin indicate full marks.

USE A SEPARATE ANSWER-BOOK FOR THIS SECTION.

1. On April 1, 1975, A draws a bill on B for Rs.500 for three months for valuable consideration. The bill is accepted by B on the same date and A has it forthwith discounted at the Bank @ 6%.

On the date of maturity, B is unable to meet the bill for want of funds. Instead, he pays Rs.200 down to A and gives him a bill for two months for the balance due (including bank charges Rs1.50) plus interest @ 4%. This bill is duly met at maturity.

Show the necessary i) Journal
ii) Cash Book
and iii) Ledger
entries in A's book. (15)

2. Define depreciation.

State the various methods of providing for the same in accounts.

What is the difference between depreciation and obsolescence. (15)

3. Mention the various methods by which Stores issues may be priced ? Which method would you consider best and why? (15)

S.D.X GROUP C : BIOMETRIC METHODS

Practical - Special Paper III

(No candidate available)

S.D. X Group D : DESIGN AND ANALYSIS OF EXPERIMENTS : May 1975

(Practical) Special Paper III

(Attempt any three questions from this group)

Figures in the margin indicate full marks

Time : 5 hours Full marks : 100

1. The coded data in table below are flow rates of a fuel through three types of nozzle, as measured by five different operators, each of whom made three observations on each nozzle:

Operators ⟶

Nozzle	1	2	3	4	5
A	6, 6, -15	26, 12, 5	11, 4, 4	21, 14, 7	25, 13, 25
B	13, 6, 15	4, 4, 11	17, 10, 7	-5, 2, -5	15, 9, 1
C	10,10, -11	-35, 0,-14	11, -10,-17	12, -2, -16	-4, 10, 24

(a) Analyse the data, assuming that the operators are regarded as a random sample from a large population.

(b) Would the differences between the nozzles be significant if such assumptions are not made?

(c) If the answer to (b) dis agrees with the corresponding one found in (a), can intuition alone provide an explanation, why differences that are significant under one model are not under the other. (14+14+4)=32

2. An agricultural experiment was conducted in a two-associate class P.B.I.B. design. The lay-out of the design is given below:

	Treatments		Treatments
Block 1.	(8, 4, 10, 12, 14, -1)	Block 6.	(12, 6, 13, 4, 3, 9)
2.	(4, 15, 10, 6, 7, 5)	7.	(7, 3, 1, 15, 12, 11)
3.	(15, 8, 9, 2, 3, 10)	8.	(13, 14, 11, 8, 6, 5)
4.	(6, 2, 13, 15, 14, 1)	9.	(14, 7, 11, 4, 2, 9)
5.	(2, 12, 5, 8, 7, 13)	10.	(1, 5, 9, 10, 13, 11)

(a) Find out the association parameters and identify the association scheme.

(b) Utilise the above association scheme to construct the B.I.B. design:

$v = b = 15$, $r = k = 7$, and $\lambda = 3$

(16+4+12)=32

3. The following table gives the lay out and yields of a 3 x 4 balanced factorial experiment, conducted in 9 four-plot blocks. In every block, each of the levels b_j (j=1,2,3,4) of the second factor B, is represented just once and the figures are arranged according to their order. Entries in the table show the levels a_i, (i=1,2,3) of the first factor A and the yields. Analyse the data to test for the main effects and interaction.

Block	Levels of B			
	b_1	b_2	b_3	b_4
1.	a_1	a_1	a_1	a_1
	3.87	4.42	5.54	6.23
2.	a_1	a_2	a_2	a_3
	4.91	5.28	6.06	6.54
3.	a_1	a_3	a_3	a_2
	4.42	4.94	5.21	5.87
4.	a_2	a_1	a_2	a_2
	4.02	4.27	5.39	6.38
5.	a_2	a_2	a_3	a_1
	4.84	5.38	6.16	6.44
6.	a_2	a_3	a_1	a_3
	4.37	4.99	5.26	5.82
7.	a_3	a_1	a_3	a_3
	4.14	5.89	4.70	5.32
8.	a_3	a_2	a_1	a_2
	5.11	5.50	6.31	5.91
9.	a_3	a_3	a_2	a_1
	4.68	4.82	5.55	5.41

(32)

4. In a chemical experiment fourteen combinations of three factors were tried. Table below gives the yields and the levels of x_1, x_2, and x_3 (on a suitable scale) for each combination.

x_1	x_2	x_3	Yield	x_1	x_2	x_3	Yield
-1	-1	1	45	2	0	0	42
-1	1	-1	39	2	2	0	42
1	-1	-1	49	2	2	0	43
1	1	1	43	0	2	0	48
0	0	0	43	0	0	2	43
0	0	0	45	0	0	- 2	46
0	0	0	41	0	0	0	43

(a) Fit a second degree polynomial regression equation to predict the yields in terms of the levels of the factors.

(b) Write down the analysis of variance table showing separately the expressions due to first and second degree terms.

(c) Obtain the variance function of the predicted yield at the point x_1, x_2, x_3.

(11+10+8)=32

SD(M75)X(10)

SD X GROUP E : SAMPLE SURVEYS - MAY 1975

(Practical) - Special Paper III

Time : 5 hours

Full marks : 100

(Attempt three questions in all, Q 1 and Q 3 being compulsory

Figures in the margin indicate full marks

1. A survey of fertilizer practices carried out in a certain region, yielded the following data relating to holding size and consumption of chemical fertilizers.

Serial Number	Holding size (acres)	Consumption of chemical fertilizer (lbs).	Serial Number	Holding size (acres)	Consumption of chemical fertilizer (lbs.)
1.	21.04	0	19.	3.15	0
2.	12.59	0	20.	4.84	36
3.	20.30	48	21.	9.07	108
4.	16.16	54	22.	3.69	0
5.	23.82	0	23.	14.61	0
6.	1.79	0	24.	1.10	0
7.	26.91	544	25.	22.13	248
8.	7.41	117	26.	1.68	56
9.	7.68	0	27.	149.58	216
10.	66.55	0	28.	1.68	42
11.	141.80	197	29.	4.80	66
12.	21.12	192	30.	12.72	180
13.	8.29	0	31.	6.31	74
14.	7.27	160	32.	14.38	182
15.	1.47	0	33.	22.19	222
16.	1.12	0	34.	5.50	192
17.	10.67	0	35.	25.29	224
18.	5.94	0	36.	20.29	40

(a) Treating the above 36 holdings as the population and the size of holdings as 'size', select a sample of two holdings without replacement drawing them one by one, the probability of drawing any unit in a particular draw being proportional to its size.

Calculate an estimate of the total consumption of fertilizer in 36 holdings on the basis of the sample of 2 holdings, using (i) Des Raj's estimator and (ii) Des Raj's symmetrized estimator.

(b) From these 36 holdings draw a sample of 5 holdings, with suitable selection probability so as to get an unbiased ratio estimate for the total consumption of fertilizer, for the above 36 holdings and furnish an unbiased estimate of the variance of the estimate for the total consumption of fertilizer. (6+10+7+5+12) = 40

2. In a simple random sample of 374 households from an American town, 292 were found occupied by White families and the remaining 82 by Non-whites. A subsample of about 1/4th of these households gave the following additional information regarding ownership of the houses.

	Owned	Rented
White	31	43
Non-Whites	4	14

Using the double sampling technique, estimate the proportion of rented households in the town under study. Also find an estimate of the standard error of the estimate. (20)

(M58)

3. Data are required to be collected on the cost of elementary education in a certain district of West Bengal, for the latest five years.

(a) Draw up a brief schedule for the same and a scrutiny programme for checking the above filled-in schedules, prior to processing.

(b) Suggest a suitable tabulation programme to show the costs under:-

(i) teachers' salary

(ii) recurring cost of maintenance of building, furniture, teaching accessories and miscellaneous other items.

(iii) non-recurring cost of land, buildings, furniture and miscellaneous other items.

(10+10+20)=40

4. A population consists of 100 units, serially numbered from 1 to 100. Consider the following scheme of sampling:

(i) Draw one unit out of the first ten units, with probability of drawing i-th unit proportional to i (i+1) where (i=1,2,...,10).

(ii) Form a sample of size ten, by taking the unit drawn in stage (i) and the remaining nine units, as those bearing the serial numbers at intervals of ten from the selected one.

(a) Adopt this method to select a sample of size ten.

(b) Suppose the ten possible sample totals are respectively given by 595, 692, 538, 439, 582, 607, 659, 482, 553, 573.

(i) Obtain an unbiased estimate of the population mean based on the sample in (a) above.

(ii) Also work out the standard error of the estimate. (20)

S.D. X Group F : TECHNIQUES OF COMPUTATION

Numerical Analysis (Practical) (M50)

Special Paper III based on use of Unit Record Machines

(Proposed to be given separately)

SD X Group G : Statistical Inference - May 1975

Practical - Special Paper III

(Attempt all questions from this group)

Figures in the margin indicate full marks.

1.(a) Let X be normally distributed with mean θ and variance unity. You are to test $H_0[\theta=0]$ against $H_1[\theta = \pm 1]$. The loss is simple, that is, for a correct decision it is zero and for incorrect decision it is one. Assuming that least favourable distribution assigns equal probabilities $\frac{1}{2}(1-g_0)$ to $\theta = \pm 1$ and probability g_0 to $\theta = 0$, where g_0 is to be properly chosen, find the minimax decision function for this problem.

(b) In the above problem find the Bayes solution when $g_1 = \frac{1}{3}$. (10+6)=16

2. For the problem of testing the hypothesis $H_0[\mu=0]$ against $H_1[\mu>0]$ about the population $N(\mu,1)$ on the basis of a random sample of size n, consider the following two tests :

(a) Reject H_0, if and only if $\sqrt{n}\,.\,\bar{x} \geq C$ where $\bar{x}$ is the sample mean.

(b) Reject H_0 if and only if, the number of positive sample observations, $S \geq K$.

i) Draw the power curves of the two tests on the same graph paper, taking n = 10 and level of significance 5.5%, for both the tests.

ii) Draw on the same graph paper the power curve of the test

(a) with n = 64 and power curve of test

(b) with n = 100, taking 5% level of significance for both.

Give your comments in each case. (10+7)=17

3. The results of a survey to assess public opinion regarding some problem of current interest are included in the following table.

Opinion	age in years			
	19-25	26-35	36-55	56 & above
i) support unconditionally	76	125	96	10
ii) support conditionally	69	117	120	17
iii) condone	14	27	35	4
iv) oppose unconditionally	60	168	210	46

Examine whether the nature of public opinion on the subject depends on the age; if so, comment on how it is affected. Test whether the distribution in the age groups 19-25 and 26-35 are identical or not. (9+6)=15

Please turn over

4. From the following table find the maximum likelihood estimate of the unknown parameter 'p' by the method of scoring.

Class	1	2	3	4
Observed frequency	125	18	20	35
Probability	$\frac{1}{4}(3-2p-p^2)$	$\frac{1}{4}(2p-p^2)$	$\frac{1}{4}(2p-p^2)$	$\frac{1}{4}(1-2p+p^2)$

(10

5. Let $X_1, \ldots, X_n$ be a random sample of size n from the normal distribution, N $(\theta, 1)$

i) for testing $\theta = 0$ against $\theta = 1$, let n (α, β) be the least sample size required, so that the most powerful test based on a random sample of size n (α, β) will have the probabilities of first and second kind of error at most equal to α and β respectively. Obtain n (α, β) for $\alpha = 0.01$, $\beta = 0.10$ and $\alpha = 0.05, \beta = 0.01$.

ii) for each of the two combinations of α and β, obtain the expected sample size in Wald's SPRT for testing $\theta = 0$ against $\theta = 1$, at the two points $\theta = 0$ and $\theta = 1$.

[You can neglect the excess over the boundaries at the termination of the test.] (8+10)=18

6. In an investigation of the growth of fruit trees the following correlation matrix was obtained relating to:-

i) the logarithms of the weight of the mature tree at grubbing,

ii) basal trunk girth of the mature tree,

iii) total shoot growth during the first four years,

iv) basal trunk girth at 4 years of age.

$$R = \begin{bmatrix} 1.000 & 0.939 & 0.266 & 0.178 \\ & 1.000 & 0.424 & 0.338 \\ & & 1.000 & 0.835 \\ & & & 1.000 \end{bmatrix}$$

The trees were set out in a randomized block design that led to 28 d.f. for each sum of squares and products of measurements.

(a) Under the usual assumptions, test the hypothesis of independence of the 4 year and maturity sites of responses.

(b) Extract the two canonical correlations and their associated coefficient vectors. (10+8)=18

7. The following is a random sample of size 20 from a population :

37.0, 31.4, 34.1, 33.3, 34.9, 31.6, 31.3, 34.0, 32.0, 31.6
30.2, 31.0, 33.5, 33.7, 33.4, 33.4, 32.1, 33.3, 32.3, 31.5

Assuming that the parent distribution is symmetric about the median, test the hypothesis that the population median is 32. Find the 90% confidence interval for the median. (6)

SD X Gr.H : Probability Theory
SPECIAL TOPICS - Special Paper III
(No candidate available)