QUESTION PAPERS

for

Statistician's Diploma Examinations

March & September 1961

Price Rupee one only

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1961

PAPER I—THEORETICAL STATISTICS, GENERAL

Time: 4 Hours | Full marks: 100

(*a*) Attempt *any five* questions.

(*b*) Figures in the margin indicate full marks.

1. The following values of u_x are given:

$$u_{10}=544, \qquad u_{15}=1227, \qquad u_{20}=1775$$

(*a*) Find, correct to one decimal place the value of x for which $u_x=1000$ by direct interpolation using a formula with unequal intervals and by inverse interpolation using a formula with equal intervals. (14)

(*b*) Estimate the drviative of the function u_x at $x=15$. (6)

2. (*a*) Define mutually exclusive events and state whether they can be regarded as a special case of dependent or independent events. (5)

(*b*) A particular army corps consists of 24 battalions, each consisting of 1000 men. It is to be divided into 4 divisions each to consist of 6 battalions, it being equally likely that any particular battalion will be allotted to any one of the divisions. What is the probability that 100 men, known to be serving in the army corps, who were only eligible to enlist in 2 particular battalions out of the 24, and each of whom is equally likely to have enlisted in either of the 2, will all be in the same division?

Which, if any, of the information given above have no bearing on the probability in question? (15)

3. (*a*) Derive the probability distribution of the number of successes in 'n' independent trials each with a constant probability 'p' of success. (6)

(*b*) Give two examples from practice in which the Binomial probability law can be expected to hold good. (4)

(*c*) Given a Binomial distribution with $p=0.5$, show how to determine the number of trials such that the probability shall be at least 0.95 that the sample relative frequency p' will not deviate from 0.5 by more than 0.03? (You may use the Normal approximation). (10)

4. (*a*) How is the median of a series of observations defined? In which situations would you prefer the median as a centre of location? (5)

(*b*) The median diameter for a large number of spherical balls was found to be 8 cm.

What will be the median volume? Justify your answer.

If it is given that there are equal number of balls in the diameter-intervals: (6 cm. 8 cm) and (8 cm. 10 cm.) show that the number of balls in the two volume. intervals:

$$\left(\frac{\pi}{6} 6^3 \text{ cc}, \frac{\pi}{6} 8^3 \text{ cc}\right) \text{ and } \left[\frac{\pi}{6} 8^3 \text{cc}, \frac{\pi}{6} 8^3 + \frac{\pi}{6} (8^3 - 6^3) \text{ cc}\right]$$

will not in general be the same.
Which of the two intervals will have the larger number?

If the diameters are symmetrically distributed, show that the volume distribution will be positively skew. (10)

(c) If, for the diameter distribution, the following figures are available: mean $= 8$ cm, standard deviation $= 2$ cm, $\beta_1 = 0$, $\beta_2 = 4.00$, what would be the mean and standard deviation of the surface area? (5)

5. (a) Explain clearly the meaning of 'marginal and conditional distributions' with reference to a bivariate density function $f(x, y)$. (5)

(b) For the density function,

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

obtain the marginal density for x and the conditional density $f(y/x)$. (8)

(c) Calculate:

(i) $P[(x + y) < 3]$ and (ii) $P(x < 1.5, y < 2.5)$ (7)

6. (a) Explain what is meant by 'association of attributes,' Describe the relation satisfied by class-frequencies when two attributes are independent. (5)

(b) Explain how the chi-square test is applied to ascertain the significance of association between two attributes, from a sample contingency table. (5)

(c) Describe an exact procedure suitable for small sample sizes, to test the significance of the difference between two independent binomial proportions. Derive the necessary sampling distribution. (10)

7. (a) Explain fully 'standard error of a statistic' in the context of large sample tests of significance. (5)

(b) Obtain large- sample standard errors of the standard deviation and the coefficient of variation in samples from a normal population. (5)

(c) In a study of height of Indian adult males, a sample of 200 individuals gave rise to a mean height of 5 ft. 5.1 ins. with a standard deviation of 2.7 ins. In a similar study made 40 years later, on a sample of 400 individuals, the mean was found to be 5 ft. 6.0. ins. with a standard deviation of 2.6 ins. To what extent do these data substantiate the hypothesis that the height of the adult male in India had increased? (10)

PAPER II—APPLIED STATISTICS, GENERAL

Time: 4 Hours | Full marks: 100

(*a*) Attempt *any five* questions.
(*b*) All questions carry equal marks.

1. Suppose you are required to conduct a sample survey in your state to study the changes in the economic conditions of the people during the Second Five Year Plan period. What sampling unit will you choose and what types of information will you collect? What type of frame will you use?

   Prepare a sampling design with stratification and with selection on the principle of probability proportionate to size of sampling unit, separately for urban and rural areas, indicating the method of determining the number of units to be chosen in different strata.

2. What do you mean by cost and variance function? How will you obtain them empirically? In a two-stage sampling, the first stage units are of equal sizes and the same number of second stage units are to be selected in each of them. Given the variance function and cost function how will you determine the number of units to be selected in the two stages so as to minimize the variance of the estimate of the mean of the characteristic under study, for a given cost?

3. Suppose it is desired to fix the prices of agricultural commodities in such a way that the change in the prices of these articles tally with the change in the prices of the items entering into the cost of cultivation, and you are asked to prepare two suitable index numbers to be used for this purpose. Give an outline of the methods of construction of these index numbers indicting the items to be included and the formulae to be used.

4. What do you mean by the net output of an industry in a year? If you are asked to calculate the national income of India for a recent year, by adding up the net output of different sectors, how will you divide the whole economy into sectors? Describe in detail how you will calculate net output of agriculture, textile industry, and transport and communication.

5. Give a critical review of the statistics of manufacturing industries in India and comment on its inadequacies in relation to the requirements for planning.

6. What are the main factors that determine the growth of a population? Discuss the growth of population in India in the last decade with reference to these factors and suggest ways for checking this growth.

7. What do you mean by a factorial experiment? Why are these types of experimental designs preferred to singgle-factor type of experimental designs? Give the layout of a 2^3 factorial design with the highest order interaction being confounded and indicate the method of analysis.

8. Explain why it is necessary to determine the reliability and validity of test scores. A battery of tests is given to 100 boys of the final year class of a higher secondary school and their scores noted. What procedure will you adopt to determine the reliability and validity of these test scores?

9. When are two genetic factors said to be linked? Derive the expected frequencies of the four types of off springs from an intercross involving heterozygots for two (possibly linked) factors in coupling.

Derive a test of significance for examining the independence of two factors on the basis of data on the offsprings of an intercross.

PAPER III—STATISTICAL INFERENCE

Time: 4 Hours | Full marks: 100

(*a*) Attempt *any four* questions.
(*b*) Figures in the margin indicate full marks.

1. (*a*) Define a sufficient statistic. (3)

(*b*) Show that a necessary and sufficient condition for the existence of sufficient statistic based on independent samples from the same population involving one unknown parameter θ is

$$\phi' = A + B\phi$$

where, $\phi = \frac{\partial \log p}{\partial \theta}$, p is the joint density of the ssample and A, B do not involve the sample. (12)

(*c*) Explain briefly the role of sufficient statistics in estimation and tests of significance. (10)

2. (*a*) Explain what is meant by 'errors of first and second kind.' (6)

(*b*) Explain the uses of the power function. (10)

(*c*) Discuss critically the optimum properties of the F-test in analysis of variance problems. (9)

3. Explain the tests of significance you will apply to the following problems, stating clearly the nature of the sampling distribution involved and the availability of the tabulated values of the levels of significance:

(*a*) to test the hypothesis that the population intra-class correlation coefficient has a given value. (7)

(*b*) K treatments are tried in r complete blocks (with k plots in each). The observations from the same block are assumed to be equi-correlated, but those from different blocks are independent. The problem is to test the equality of the k treatment means. (11)

(*c*) The random variable y has a linear regression on $x_1, x_2, \ldots, x_p$ of the form $Y = \alpha + \beta_1 x_1 + \ldots + \beta_p x_p$. The problem is to test the hypothesis:

$$\beta_1 = \beta_1^0, \ldots \beta_p = \beta_p^0 \quad (7)$$

4. (*a*) Explain clearly what is meant by confidence intervals, in the case of one unknown parameter. (7)

(*b*) When there are more than one set of confidence intervals, what is the criterion on which the choice of the best confidence interval is based? Illustrate by means of an example. (9)

(*c*) A 90 percent confidence interval for the mean of a normal population with known variance, of the form:

$$(\bar{x} + l_1, \bar{x} + l_2)$$

is to be provided where $\bar{x}$ is the sample mean based on n independent observations. Derive the confidence interval of the shortest length. (9)

5. Measurements are available on the lengths of each of five fingers on the right as well as the left palms of n individuals.

(*a*) Suggest a suitable procedure for testing lateral symmetry, i.e. whether the corresponding left and right measurements are equal on an average, indicating fully the algebraic formula for the proposed test statistic and necessary tables to be referred to for ascertaining its significance. (12)

(*b*) Derive the sampling distribution of the test statistic under the null hypothesis, stating clearly the conditions under which this would be valid. (13)

6. (*a*) State and prove Neyman-Pearson's fundamental lemma. (10)

(*b*) Let $x_1, x_2, \ldots, x_n$ be n independent observations on a normal variate X with mean μ and standard deviation unity. Show how you will construct the following tests for the hypothesis $\mu = \mu_0$ against $\mu \neq \mu_0$ at a preassigned level of significance, α.

(*i*) the most powerful test.
(*ii*) the least powerful test
(*iii*) the test whose power function is always equal to α

Show that test (*i*) is uniformly most powerful with respect to one sided alternatives. (15)

PAPER VI—PRACTICAL

Time: 6 Hours Full marks: 100

(*a*) Attempt *any five* questions.
(*b*) Figures in the margin indicate full marks.
(*c*) Use of calculating machines is permitted.

1. (*a*) The following values were computed from 500 observations on a characteristic.

$\bar{x} = 49.04$; $\mu_2 = 243.06$; $\mu_3 = -199.68$; $\mu_4 = 162220.63$

Determine the type of Pearsonian curve that would fit the distribution of the characteristic (Evaluation of constants not necessary). (5)

(*b*) The marks in English obtained by 500 candidates of a University examination was found to follow the Pearsons' Type II distribution

$$y = y_0 \left(1 - \frac{x^2}{59.4^2}\right)^{0.3}, \quad -59.4 \leq x \leq +59.4, \text{ with}$$

origin at 38.9. Find by consulting Biometrika tables the first quartile of the distribution. (5)

(*c*) The following data give the monthly average of electricity generated (in million kwt) in a country during 1948 to 1956.

year	1948	1949	1950	1951	1952	1953	1954	1955	1956
electricity (million kwt)	381	409	426	488	516	559	627	716	805

Fit a polynomial growth curve of the third degree and estimate the monthly average production in 1957. (10)

2. Fifteen rose plants in each of *two* gardens were grouped into *five* exclusive sets of *three* plants each, chosen at random. Each set of three plants in a garden was given different treatments (*A*, *B*, *C*, *D* and *E*). The number of salable flowers obtained from these plants in a specified period are shown below:

Number of salable flowers from plants.

treatment	garden I	garden II
A	27, 18, 19	102, 71, 79
B	31, 52, 57	84, 76, 81
C	34, 52 33	67, 74, 83
D	38, 60, 45	71, 51, 63
E	31, 50, 45	53, 63, 61

Test whether the treatments differ significantly from one another in producing salable flowers. (20)

3. In an experimental field, the number of larvae of bet webworm (Loxostige sticticalis L) was to be determined. The field was divided into 750 plots of which 65 were chosen at random and the number of larvae counted. The data obtained are summarised below:

Distribution of the number of larvae in 65 plots

number of larvae	0	1	2	3	4	5	6	7	8	9	total
frequency	24	17	10	8	4	1	0	1	0	0	65

(*a*) What theoretical distribution do you think would fit the distribution? Calculate the expected frequencies based on your model and test the goodness of fit. (15)

(*b*) What is your estimate of the number of larvae in the whole field? Also determine the standard error of this estimate. (5)

4. In a study to determine the effects of factors which influence academic success, results of 450 students were recorded in respect of X_1 = marks in examination, X_2 = general intelligence, and X_3 = average daily hours of study. The means, stdanrd deviations and correlations are

X_1	X_2	X_3
$\bar{X}_1 = 18.5$	$\bar{X}_2 = 100.6$	$\bar{X}_3 = 4.5$
$\sigma_1 = 11.2$	$\sigma_2 = 15.8$	$\sigma_3 = 1.12$
$r_{12} = .60$	$r_{13} = .32$	$r_{23} = -.35$

(*a*) Find the extent to which examination marks are related to general intelligence, when the average hours of study remains the same; ascertain if this is significant. (10)

(*b*) Also obtain a formula for predicting X_1 when X_2 and X_3 are given; estimate the standard error of prediction. (10)

5. The distribution in four blood-groups *O*, *A*, *B* and *AB* of 150 convicts and 450 non-nonconvict factory-workers in Bombay are given below.

	O	A	B	AB	Total
convicts	58	63	20	9	150
nonconvict workers	180	183	62	25	450

(*a*) Test if the two sets of individuals come from populations having the same blood-group composition. (10)

(*b*) Examine whether the observed frequencies for the convicts in the four classes *O*, *A*, *B* and *AB* are consistent with a model which specifies the relative frequencies as

r^2, p^2+2pr, q^2+2qr and $2\,pq$ respectively in terms of unknown parameters p, q, r such that $p+q+r=1$. [The maximum likelihood estimates for these parameters obtained from the same data are given by

$$\hat{p} = .28, \quad \hat{q} = .10, \quad \hat{r} = .62] \qquad (10)$$

6. The difference (correct to five decimal places) in siderial longitudes of Sun and Moon at 5.30 A.M. each day from 17th to 20th December 1960 are given below:

time and date		*diffcence in longitude* (x)
5.30 A.M.	17-12-1960	20.22778
-do-	18-12-1960	6.32333
-do-	19-12-1960	− 7.79611
-do-	20-12-1960	−21.95689

(*a*) Find by a suitable method of interpolation the date and time at which the difference (x) was zero. (10)

(*b*) What are the assumptions involved in your procedure? (5)

(*c*) What is the maximum likely error in your estimate due to rounding off at each stage of computation? Give your reasons. (5)

PAPER VII—PRACTICAL.

Time: 6 Hours — Full marks: 100

(*a*) Attempt *all* questions.

(*b*) Figures in the margin indicate full marks for each part of questions.

(*c*) Use of calculating machine is permitted.

Note: In evaluating your answer, credit will be given for neatness, tabular presentation of computations and precise statement of basic assumptions and final conclusions.

Computational accuracy is essential. But some credit will be given for using the correct procedure even when answers are numerically inaccurate. You are therefore advised to describe very briefly the procedure you adopt and to write down the computational formulae you use.

1. The table below gives the farmer's prices and volumes produced of a number of grains in the United States of America for the years 1950 and 1955.

Compute index numbers of farmer's prices of grains in 1955 with 1950 as base, using:

(a) weighted arithmetic average of price-relatives, weights being the quantites produced in lbs. in 1950. (4)

(*b*) Laspeyres' formula (4)

(*c*) Paasche's formula (4)

(*d*) Fisher's 'ideal-index' formula (3)

Volumes produced and farmer's prices of grains, USA, 1950 and 1955

grains	volume produced (million bushels)		farmer's prices (dollars per bushel)		average weight in lbs of a bushel of the grain
	1950	1955	1950	1955	
(1)	(2.1)	(2.2)	(3.1)	(3.2)	(4)
wheat	1019	935	2.00	1.98	60
rye	21	29	1.21	1.06	56
buckwheat	4	2	1.11	1.17	56
corn	2764	2884	1.53	1.34	56
oats	1369	1503	0.79	0.60	32
barley	304	401	1.18	0.92	48
sorghums	234	243	1.05	0.98	56

2. *Either,*

An experiment was conducted to examine whether subjecting the seed to temperature treatment before planting has any effect on yield. Five different levels of temperature, 60°F, 75°F, 90°F, 105°F and 120°F were tried. The experiment was conducted with a Randomised Blocks layout with ten replications of each treatment. The average yields per plot for the five treatments are given below. The *error mean square* as computed was 26.57.

Average yield per plot based on ten replications

temperature	60°F	75°F	90°F	105°F	120°F
average yield per plot	28.2	36.5	43.2	42.5	37.7

(*a*) Draw up the blank Analysis of Variance table and write down the degrees of freedom for the components. (2)

(*b*) Compute the mean square due to the treatments and test for its significance (4)

(*c*) Denoting the expected yield at temperature t by Y_t, fit a quadratic response curve of the form

$$Y_t = \alpha + \beta t + \gamma t^2 \qquad (8)$$

(*d*) Compute the mean square due to deviation from quadratic response and test for its significance. (6)

(*e*) Using the estimated quadratic response curve obtained in (*c*) find the optimum temperature for seed treatment, that is the value of t for which Y_t is a maximum. (5)

Or,

The following table gives the yield of cane in quarters per 1/40 acre plot from two sugarcane variety trials in years 1933 and 1934 respectively.

	Experiment (1933) Blocks					Experiment (1934) Block				
	1	2	3	4	5	1	2	3	4	5
varieties										
I	41	44	45	54	44	38	35	41	39	54
II	52	62	60	66	49	51	52	50	63	62
III	44	45	50	50	60	46	38	47	49	49
IV	56	50	60	56	60	47	57	55	65	46
V	51	56	61	64	63	43	51	56	56	72

(*a*) Analyse the experimental data for the two years separately to test the significance of differences between varieties. (10)

(b) Examine whether the residual variation is of the same order in both the years. (5)

(c) Compute the interaction between years and treatments and tests its significance. (5)

(d) Can you suggest the best variety on the basis of the statistical analysis carried out? (5)

3. For 20 subgroups of 5 samples each, the measurements in thousandths of an inch on the thickness of mica discs manufactured by a machine for assembly in radio parts, are given below.

subgroup	sample 1	2	3	4	5
1	11	8	12	15	14
2	11	10	14	13	15
3	17	12	10	8	11
4	15	12	11	12	14
5	13	11	12	12	5
6	13	8	10	12	10
7	11	12	10	9	13
8	14	10	12	7	5
9	12	10	13	10	8
10	10	11	9	11	13
11	10	12	13	14	7
12	8	10	12	9	12
13	13	12	14	16	10
14	7	9	11	13	9
15	8	8	15	7	6
16	7	10	12	10	15
17	7	8	13	12	10
18	10	10	16	12	12
19	12	12	17	11	9
20	19	8	14	10	11

(a) By drawing control charts, examine if the manufacturing process is under control. (15)

(b) Excluding sub-groups, if any, not under control, estimate the mean and the standard deviation of the process. (6)

(c) The specifications for the thickness are $0.008'' - 0.015''$. Using the mean and the standard deviation obtained in (b) and assuming normality, estimate the long run percentage of discs, manufactured under controlled conditions, that will not meet the specifications. (4)

4. A plan for a forest survey to estimate the volume of timber is required. The technique of double-sampling will be used. In a first sample of n randomly selected plots of 1/10th acre, eye-estimates (x) and measurements (y) of the volume of

timber will both be taken and the linear regression equation $a+bx$ of y on x will be worked out. In a second sample of N plots, only the eye-estimates will be taken. The final estimates will be obtained as $T=a+b\bar{x}$ where $\bar{x}$ is the average of the eye-estimates in the second sample. The variance of T is given approximately by

$$V(T) = V_0\left[\frac{\rho^2}{N} + \frac{1-\rho^2}{n}\right]$$

where V_0 is a certain constant and ρ is the coefficient of correlation between x and y.

The cost of conducting the survey, may be taken to be of the form

$$C = a + bn + cN$$

where a, b and c are certain constants.

If the cost of the survey is fixed at C=Rs. 10,000, determine the values of n and N which will make $V(T)$ a minimum. Use the following estimates of the constants:

V_0 = 400 in suitable units
ρ = 0.7
a = Rs. 1000, b=Rs. 40, c= Rs.5.

Find also the value of $V(T)$ for these values of n and N. (15)

5. The following table gives the age-specific mortality rates q for $x=0, 1, 2, 3,$ and 4 in a certain stationary population.

Mortality rates q_x

x	0	1	2	3	4
q_x	0.037915	0.004219	0.002314	0.001712	0.001258

(a) Given that the complete expectation of life at birth

e_0^0 = 68.59 years, compute e_4^0, the expectation of life at age 4. (10)

(b) Compute the chance that of two new born babies, none will die before completing four years of life. (10)

PAPER IV AND V—ECONOMIC STATISTICS (THEORETICAL)

Time: 4 Hours Full marks: 100

(a) Attempt *any five* questions.
(b) Figures in the margin indicate full marks.

1. (a) The following extract is taken from 'The Role of Measurement in Economics' by Richard Stone: "The condition for the exact identifiability of a structural equation in a linear model is that the number of variables which appear in the model but do not appear in the equation be equal to the number of equations in the model less one."

Explain the principle considered and illustrate by a simple example. (12)

(*b*) Write out the basic equations of *any two* of the following economic growth models:

(*i*) Harod-Domar
(*ii*) Mahalanobis two sector.
(*iii*) Samculson Multiplier Accelerator. (8)

2. (*a*) Purchasing power of a rupee to consumers may be different in different states of India. Describe the type of data needed and method to be adopted for measuring such differences. (16)

(*b*) What method will you recommend for comparing the cost of living in Calcutta with that in London? (4)

3. (*a*) 'The main work dealing with the statistical derivation of demand function is . . . due to the late Professor Henry Schultr of the University of Chicago' [Tintner: Econometrics].

Describe any one of the procedures used by Schultr for obtaining a demand function. (14)

(*b*) A demand function is given by $x_a^2 p_a^3 p_b = 10$ where x_a is the quantity of a commodity '*a*' demanded, p_a the price of this commodity and p_b the price of some other commodity '*b*'. Find the partial elasticities of demand. (6)

4. (*a*) Describe the method followed in India and obtain the contribution to national income by *any one* of the following sectors: (*i*) agriculture, (*iii*) other commerce and transport. (10)

(*b*) Distinguish between the concepts of national income and disposable personal income. (10)

5. Write brief notes on *any four* of the following: (20)

(*i*) distribution of national income by size.
(*ii*) distribution of national income by factor shares
(*iii*) correlogram
(*iv*) cyclical variation in a time series
(*v*) National Sample Survey
(*vi*) Rural credit survey
(*vii*) The estimate of money supply in India.

6. (*a*) Describe very briefly the methods used for projecting the population of a country. (10)

(*b*) Write a note on the distinguishing feature of the 1961 population census of India. (10)

7. The following extracts are taken from Chapter II of the Third Five Year Plan: A Draft Outline:

(*i*) 'The increase in national income over the second five year plan is expected to be about 20 percent.'

(*ii*) '. . . agricultural production will be going up by about 40 percent between 1950-51 and 1960-61.'

(*iii*) 'The total irrigated area will increase from 51.5 million acres in 1950-51 to 70 million acres in 1960-61.'

(*iv*) 'Overall industrial production is expected to record an increase of about 120 per cent between 1950-51 and 1960-61.'

(*v*) 'The installed generating capacity will increase from 2.3 million KW in 1950-51 to 5.8 million KW by the end of 1960-61.'

(*vi*) 'From 91 million tons in 1950-51, the total freight traffic originating on the railways is expected to increase to 162 million tons at the end of the second plan.'

(*vii*) 'The percentage of children attending schools to all children in the age groups 6-11 will increase from 43 in 1950-51 to 60 by 1960-61.'

Write brief critical notes on *any four* of the above quotations indicating the usual sources of the information relating to the base period. (20)

PAPER IV AND V—STATISTICAL QUALITY CONTROL (THEORETICAL)

Time: 4 Hours Full marks:100

(*a*) Attempt *any four* questions.
(*b*) All questions carry equal marks.
(*c*) Figures in the margin indicate full marks for each part of questions.

1. (*a*) What is acceptance sampling? Explain the use of OC-curve in sampling. What is the effect of inspection errors on the OC-curve? (10)

(*b*) Describe clearly any scheme of inspection of a continuous production process to insure a prescribed limit on the outgoing quality. (15)

2. (*a*) Bring out the difference between multiple sampling and item by item sequential sampling.

Describe the sequential procedure for testing whether the proportion of defectives produced by a given process is equal to that produced by a different process. (15)

(*b*) Distinguish between

(*i*) control limits and specification limits.
(*ii*) tolerance limits and confidence limits. (10)

3. (*a*) Describe the various uses of Range for routine control and tests of significance in industry. (12)

(*b*) Describe briefly the various steps to be taken in setting up control for the number of defects per unit of a product where the defects have to be classified according to their seriousness. (13)

4. Define a run in a sequence which consists of n_a letters a and n_b letters b. Denote the number of runs of a in the above sequence as r_a and of b as r_b. If all permutations of the n_a letters a and n_b letters b are equally likely show that the joint probability distribution function of r_a and r_b is given by

$$P(r_a, r_b) = \frac{k\,(n_a-1)!\,(n_b-1)!\,n_a!\,n_b!}{(r_a-1)!\,(r_b-1)!\,(n_a-r_a)!\,(n_b-r_b)!\,(n_a+n_b)!}$$

where $k = 2$ if $r_a = r_b$ and $k = 1$ if $r_a \neq r_b$.

Deduce the probability distribution function $P(x)$ of the total number of runs $x = r_a + r_b$ distinguishing between odd and even values of x.

Explain how the above distribution can be used for process control. (25)

5. Write an essay on the investigation of multivariate processes with the aid of

(*i*) multiple regression analysis
(*ii*) factorial experiments, and
(*iii*) evolutionary operations

bringing out the advantages and limitations of each method. (25)

PAPER IV AND V—SAMPLE SURVEYS, APPLIED (THEORETICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *any four* questions.
(*b*) All questions carry equal marks.

1. Distinguish between sampling and non-sampling errors. Discuss about possible sources of non-sampling errors in sample surveys. Which (if any) of these sources are not of relevance in a census (complete enumeration)?

2. What are the points that require special consideration at the planning stage of sample surveys?

With reference to any suitable example (preferably survey actually conducted in India) explain how and to what extent these points have been covered.

3. What are the advantages and dis-advantages of the mail-enquiry method as compared to the interview method of surveys?

Discuss whether there should be any difference in 'forms' for collection of the same information by these two methods.

Enumerate systematically the different phases of work in the actual implementation of a well-planned mail enquiry. Bring out the inter-relationships between the various stages and explain in what way each affects the final outcome.

4. It is proposed to conduct a sample survey in a city for estimating total number of vagrants and beggars of different categories such as those having different types of physical deformities, etc. A sub-sample of size 300 from these persons will be surveyed, for detailed study about their past occupation, past residence, present occupation and conditions of living.

   Prepare a suitable sampling design for the survey. (There are fairly detailed maps of different wards of the city). Also discuss about the method of estimation, field-staff to be employed and the period of survey.

   Discuss about possible sources of non-sampling error and explain how you will try to eliminate them in the survey.

5. Write notes on the following:
   (*a*) Interpenetrating sub-samples.
   (*b*) Pilot survey.
   (*c*) Cost-function for a State-wide two-stage sampling design.

PAPER IV AND V—SAMPLE SURVEYS, THEORY (THEORETICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *any four* questions.
(*b*) All questions carry equal marks.

1. Define bias and accuracy of an estimate. Under what conditions are biassed estimates preferred to unbiassed estimates?
   From a random sample of n elements, a random sub-sample of m elements are drawn without replacement and added to the original sample. Show that the mean based on $(n+m)$ elements is an unbiassed estimate of the population mean, and that the ratio of its variance to that of the mean of the original 'n' elements is approximately

$$\frac{1+3\frac{m}{n}}{\left(1+\frac{m}{n}\right)^2},$$

   if the size of the population is large.

2. State the principle of 'collapsing strata' and explain how and when it is used.
   For the purpose of estimating the mean, a population having a continuous density function $f(y)$ is to be divided into k strata, by including in the i-th stratum all individuals with characteristics between y_{i-1} and y_i, $i=1, 2, \ldots, k$, $y_0 = -\infty < y_1 < \ldots < y_{k-1} < y_k = \infty$. How would you determine the y_i's $i=1, 2, \ldots, k-1$ such that under stratified sampling with proportional allocation, the variance of the usual unbiased estimate of the mean is least?

3. Discuss the advantages of cluster sampling. Derive the variance of an unbiassed estimate of the mean based on cluster sampling and show that the efficiency of cluster sampling compared to simple random sampling will be high if the intra-class correlation between elements in the cluster is negative.

4. Discuss the problem of non-response in surveys. Show how to obtain an unbiased estimate of the mean of the whole population by re-surveying a pre-assigned fraction f of the non-respondents. Find the variance of this estimate. Taking a simple but realistic cost-function, determine the optimum value of f.

5. Describe the circumstances under which sampling with varying probabilities is to be preferred to sampling with equal probability.

   How would you estimate the variance of this estimate?

   From a Tehsil, one village is to be selected for a socio-economic survey with probability of selecting a particular village proportional to the population of the village and another village is to be selected for an agricultural enquiry, with probability of selection proportional to the cultivated area in the village. Describe a scheme for selection which will ensure a high probability for a common village to be selected for both the enquiries.

6. Explain a method of obtaining an unbiassed estimate of the population mean when two units are selected without replacement with varying probability from each stratum.

PAPER IV AND V—DESIGN OF EXPERIMENTS, APPLIED (THEORETICAL)

Time: 4 Hours | Full marks: 100

(*a*) Answer *any three* questions.
(*b*) All questions carry equal marks.

1. Ten varieties of wheat are to be tested in a layout requiring seven blocks of size six and arranged in the following manner: in each block the varieties 8, 9 and 10 occur along with three other varieties from 1, 2, 3, 4, 5, 6, 7 such that each pair of the varieties 1, 2, 3, 4, 5, 6, 7 occurs just once and anyone of them not more than once in the same block and altogether in three blocks. Prepare a layout of the design.

   Describe in detail how you would analyse the data and calculate the estimates of variances of the different varietal differences. Find out the efficiency factor of the design.

2. To compare the effectiveness of three fertilizers: Dung (D), Nitrochalk (N) and Superphosphate (P), an experiment was conducted with the combinations of these fertilizers each at two levels and arranged in two Latin squares, each of side 4 in such a way that the three factor interaction is completely confounded with the squares. Prepare a layout of this design. Describe how you would proceed to analyse these data and test the main effects as well as interactions of the various fertilizers.

3. An experiment was conducted to test the effects of 3 levels of nitrogen, 3 of phosphorous, and 3 of potash on the germination of lettuce seedlings. The seed was thoroughly mixed and divided into 108 samples. Each sample was planted in a copper box in a mixture of soil and sand. The boxes were placed in a germinator and at the end of a week the number of seeds germinating in each box was counted. There were 4 replications, each placed on a different shelf in the germinator. On a shelf the boxes were placed in 3 columns of 9 boxes each so that each column represented an incomplete block. The layout of the design is a balanced group of sets for a $3 \times 3 \times 3$ factorial experiment arranged in blocks of size 9 such that the three-factor interaction is partially confounded. Prepare a layout of the design. Explain in detail the method of analysis to be employed and the structure of the analysis of variance of the design.

4. An experiment was conducted to study the responses to different applications of superphosphate (P_2O_5) on soils of various degrees of fertility. The treatments constitute 5 different levels of applications of P_2O_5 and are arranged in a 5×5 Latin square. Four regions with different types of soil were chosen so as to represent a wide range of fertility and a separate experiment was conducted in each region. Discuss in detail how you would proceed to analyse these data. Make the appropriate analysis of variance table for testing the linear as well as the quadratic effects of P_2O_5, taking into consideration the variation due to soil fertility.

PAPER IV AND V—VITAL STATISTICS AND POPULATION STUDIES (THEORETICAL)

(*a*) Attept *any four* questions.
(*b*) All questions carry equal marks.

1. Explain briefly how you will assess inaccuracies in the reporting of age data in the Indian Census. What methods can be used to make suitable adjustments?

2. What is the 'component method' of working out population projections? State clearly the data required for this purpose.

3. What is meant by aging of the population? Explain the relative roles of the decline of fertility and mortality in the aging of the European populations.

4. Explain the relationship between lifetable death rates (q_x) and age-specific death rates (m_x). How is the relationship used in the preparation of life tables?

5. White short notes on *any three* of the following:

   (*i*) Couple fertility
   (*ii*) Cohort fertility
   (*iii*) Stable population
   (*iv*) Reverse-survival method.

PAPER IV AND V—PSYCHOLOGY AND EDUCATION (THEORETICAL)

Time: 4 Hours | Full marks: 100

(*a*) Answer *any four* questions.
(*b*) All questions carry equal marks.

1. Explain clearly the characteristics of individual and group tests in test theory. Discuss the methods of standardisation of these two types of tests. What are the advantages of using standardised tests.?

2. Why is 'split half method' not considered a suitable technique for estimating test relibility. Discuss how it is possible to modify this method.

   Give a brief summary of the different methods which have been evolved for finding the reliability coefficient of any given test.

3. Why is the concept of parallel tests considered to be important in test theory? Write a critical note on Wilks' L_{max} test which is used for testing simultaneously in a parallel test battery that all means are equal, all variances are equal, all convariances are equal.

4. Starting with the fundamental equation of the multiple factor model, show that

   (*i*) the product of the reduced factor matrix by its transpose is the reduced correlation matrix,

   (*ii*) The communality of a test is always less than the reliability except in the limiting case where the specific factor is absent in which case the communality and relibility are equal.

   Supposing the rellability of a test is known, is it possible to predict the number of terms in a model?

5. Describe the essential features of (*i*) principal component model and (*ii*) multiple factor model. Indicate how factor loadings can be calculated in each case.

   Show that Spurman's model is only a special case of multiple factor model in all its aspects.

6. Write notes on

   (*i*) Objective tests
   (*ii*) Coefficients of validity
   (*iii*) Item analysis
   (*iv*) Rotation of Axis.

———

PAPER IV AND V—PROBIT ANALYSIS (THEORETICAL)

Time: 4 Hours | Full marks: 100

(*a*) Answer *any four* questions.
(*b*) All questions carry equal marks.

1. (*a*) Mention some alternative forms of dosage response curves and discuss their relative merits and demerits.

   (*b*) What are the considerations involved in the choice of a dose metameter?

2. (*a*) Explain the concepts of 'tolerance distribution' and 'median effective dose'.

   (*b*) How do you combine several independent parallel estimates of 'median effective dose' to obtain a pooled estimate, knowing the standard errors of the individual estimates?

3. Fuses are tested by submitting them to a pulse of current of x amps. In such a test k batches of fuses were tested, the ith batch containing n_i fuses of which r_i were found to have been damaged by the current x_i (x_i being the same for each fuse in batch i). If it is assumed that each fuse independently has a chance $Q(\lambda, x)$ of damage by a curent x, where $Q(\lambda, x)$ is a known function of x and and unknwon parameter λ, show that the maximum likelihood estimate of λ is

$$\sum_{i=1}^{k} \left(\frac{r_i}{Q_i} - n_i\right) \frac{1}{P_i} \frac{\delta P_i}{\delta \lambda} = 0$$

   where $Q_i = Q(\lambda, x_i)$ and $P_i = 1 - Q_i$. Find the large sample variance of the maximum likelihood estimate of λ.

   Illustrate your answer by considering the exponential response law, $P(\lambda, x) = e^{-\lambda x}$.

4. (*a*) Find the expected kill when the logarithms of the doses received by the subjects are normally distributed about a mean ξ and variance γ^2, given that the tolerance distribution of the logarithm of dose is normal with mean μ and variance, σ^2.

   (*b*) Suppose the mortality probit can be expressed in terms of two dosage factors x_1 and x_2 as

$$y = a + b_1 x_1 + b_2 x_2$$

   If the dosages to which a population can be exposed are such that x_1 can be controlled but x_2 is normally distributed about ξ_2 with variance γ_2^2, find the mortality probit as a function of x_1.

5. Write critical notes on:

   (*a*) Synergistic action, (*b*) Adjustment for natural mortality and (*c*) Parker-Rhodes equation.

PAPER IV AND V.—THEORIES OF INFERENCE (THEORETICAL)

Time: 4 Hours | Full marks: 100

(*a*) Attempt *any four* questions.
(*b*) All questions carry equal marks.

1. Give an account of the theory of inverse probability as modified by Jeffreys, and Fisher's criticism of the same. To what extent does the invariance theory proposed by Jeffreys meet the objections levelled against it?

2. Define unbiased critical regions of Type C, and show how to obtain such regions.

   A random sample of size n_1 is taken from a normal population $N(\mu_1, 1)$, and a second random sample of size n_2 is taken from another normal population $N(\mu_2, 1)$. Determine the regular unbiased Type C region to test the hypothesis $\mu_1 = \mu_2 = \mu_0$ (specified), assuming that errors of the same size in μ_1, μ_2 are equally important.

3. Explain the concept of a decision function. Show how Fisher's theory of point estimation can be viewed as a special case of the General Decision Problem.

   $x_1, x_2, \ldots, x_n$ are independent indentical random variables with range (0, 1) and mean θ. Obtain the minimax point estimate of θ when the loss function is given by the mean square difference.

4. State Wald's fundamental identity in sequential analysis. Obtain a sufficient condition for the differentiability of the fundamental identity under the expectation sign.

   Hence or otherwise derive expressions for the ASN function when

   $$(i)\quad E\left(z = \log \frac{f(x, \theta_1)}{f(x, \theta_0)}\right) = 0,$$

   and

   $$(ii)\quad E(z) \neq 0.$$

5. Obtain the decision rules for testing sequentially the hypothesis that the standard deviation of a normal distribution does not exceed a given value, assuming that the mean of the normal distribution is known. What modifications are necessary in the decision rules when the mean is unknown? Justify the modifications.

   Obtain the five point OC and ASN curves of the above sequential test procedures when the mean is known.

6. Write a critical note on the advantages and disadvantages of non-parametric tests in comparison with parametric tests.

   Show how to obtain distribution-free confidence intervals for the median and ther percentage points of a distribution.

PAPER IV AND V—ACTUARIAL STATISTICS(THEORETICAL)

Time: 4 Hours | Full marks: 100

(*a*) Attempt *two* questions from each group.
Altogether *six* questions are to be answered.
(*b*) All questions carry equal marks.
(*c*) Use of calculating machine is not permitted.
(*d*) All symbols used in this paper have their usual significance.

Group A

1. A lease-holder of property, which has 30 years to run, of annual value of Rs. 170.75 nP wishes to sell the remainder of his lease after reserving to himself the rental for the next 10 years. A purchaser wishes to realise 5 percent on his outlay for the entire term, it being assumed that he will not be able to re-invest at more than $3\frac{1}{4}$ percent. What must he give for the property?

2. A loan of Rs. 4, 800 is repayable by 40 equal half yearly instalments of principal, interest at 4 percent per annum, being payable half yearly on the balance of the loan unpaid at the beginning of the half-year. Find the sum for which the remaining payments may be commuted at the time when the 20th is due, taking interest at 3 percent per annum.

3. Find an expression for the value of an annuity certain for n terms, the terms being in geometrical progression and represented by c, c^2, c^3 .. c^n. Under what conditions could the value of a perpetual annuity of this nature be ascertained and what would be its value?

Group B

4.((*a*) Prove *any three* of the following:

(*i*) $$\overline{A}_x = \mu_x \bar{a}_x - \frac{d\bar{a}_x}{dx} = 1 - \delta \bar{a}_x$$

(*ii*) $$P_x^{(m)} = P_x + (P_x + d)\left(\frac{m-1}{2m} P_x^{(m)}\right).$$

(*iii*) $$P_x = \frac{\lfloor n A_{x+m} + v^n {}_n p_{x+m} \cdot {}_{n+m}l'_x - {}_m l'_x}{1 + \lfloor n-1 a_{x+m}}$$

(*iv*) $a_{\overline{xy}} > a_{\overline{ww}}$ where $a_{xy} = a_{ww}$ under Makeham's law.

5. Define in words the term 'Force of mortality'. Obtain an expression for the force of mortality when the law of mortality may be expressed by the equation

$$l_x = l_0 a^x . b^{x^2} . \beta^{c^x}$$

Show that if a mortality table follows Makeham's law and complete tables of expectation of life be formed for combinations of lives of equal ages, it would be possible to dispense with tables of the values of annuities.

6. Interpret and obtain integral expressions for *any four* of the following:

(*i*) $\overline{A}^2_{x:y:z}$ (*ii*) $\overline{A}^2_{x:yz:}$ (*iii*) $\overline{A}^2_{x:yz}$

(*iv*) $\overline{A}^1_{x:\overline{yz}:}$ (*v*) $\overline{A}\,\frac{1}{xy}:z$.

Give a formula of approximate summation that may be made use of to obtain

$$\overline{A}^2_{x:y:z}$$

and indicate the steps of the process.

Group C

7. If a community consists of l_x persons at age x, l_{x+1} at age $x+1$ and so on down to the extremity of life, show that the total funds to be raised to provide for an assurance of 1 on the death of each member will be $l_x\,(a_x+A_x)$; and for an annuity-due to each person

$$l_x\left(1+e_x+\frac{e_x-a_x}{i}\right).$$

What would be the uniform contribution of each individual to these several funds.

8. Obtain the values of l, m and n in the 17 term summation formula for graduation

$$\frac{[4]\,[5]\,[6]}{120}\,\{1[3]-m[n]\}$$

and calculate second and fourth difference errors. Criticise the statement: 'This formula is more acceptable to Actuaries than to Statisticians'.

9. What data would you require in order to construct a Table representing the mortality experience of the policy-holders in a Life Insurance Company?

Construct an 'Exposed to risk' formula for the 'Policy-year Method' giving the assumptions you make in details.

What policies, if any, would you exclude from the investigation?

10. How does 'Reserve' arise under a policy in level premium system? Can a Life office pay the whole of the reserve held under a policy as surrender value? Give your answer with reasoning.

How is surrender value related to paid-up value?

Should there be any provision for payment of surrender value under a policy issued on 'Natural Premium' system?

PAPER IV AND V—GENETICS (THEORETICAL)

Time: 4 Hours | Full marks: 100

(*a*) Attempt *any five* questions.
(*b*) All questions carry equal marks.

1. What are auto tetraploids? With two alleles (A, a) at an auto tetraploid locus and starting with an initial gametic array

$$x_0AA + 2y_0 \quad Aa + z_0aa$$

show that under random mating, as number of generations gets indefinitely large, the population of genotypes tends to

$$(p^2AA + 2\ pq\ Aa + q^2aa):$$

where

$$p = x_0 + y_0, \quad q = y_0 + z_0, \quad p + q = 1.$$

2. How does marriage between near relatives increase the frequency of homozygotes and decrease that of heterozygotes?

Haldane says 'It is illogical to tolerate marriage between double first cousins and between uncle-niece while prohibiting marriage between half-sibs'.

Demonstrate the validity of this statement by computing the probability of appearance of rare recessive abnormality in the offspring from each one of the three types of marriages.

3. Suppose there are two linked loci with genes (A, a) and (B, b) which behave according to the two-factor model. AB/ab is selfed and the offspring tested by selfing. Genotypes with AA or BB are resistant and produce resistant progeny only. Genotypes with a single A or a single B are resistant also and produce some resistant individuals and some susceptible individuals, while genotypes with no A or no B are susceptible. The testing is presumed to be without error and leads to classifiction of the individuals of the F_2 as resistant, segregating or susceptible.

Obtain the maximum likelihood estimate of the recombination fraction and evaluate the information on it.

4. For a simple Mendelian character in which the genotype array in generation n is

$$p_n^2AA + 2p_n\ q_n \quad Aa + q_n^2aa$$

suppose that only $(1-k)$ of the recessive survive to reproduce, but other-wise there is no selection. Show that if k is small, kn is approximately equal to

$$U_n - u_0 + \log_e\left(\frac{U_n}{U_0}\right)$$

where U_n is given by

$$U_n = \frac{p_n}{1-p_n}.$$

5. Suppose that there are N (unknown) families in a population, with normal parents capable of producing a portion p of albino children. Assume that the families are all of size s and let the ascertainment be by affected individuals only, with the probability of any such affected individual entering the records being p'.

Describe a procedure for estimating from such data the parameters p, p' and N with their respective standard errors.

6. (*a*) With reference to one variable locus with two alleles (A, a) explain the concepts of effect of gene substitution, additive genetic variance, dominance variance.

(*b*) Under suitable assumptions, find an expression for the regression of offspring on parent for an attribute which is controlled by any number of alleles occurring at a single locus.

7. For a population segregating for two alleles A, a evaluate the effect of continued parent-offspring mating in the frequencies of different mating types, starting with an initial population with frequencies p^2, $2pq$ and q^2 for the genotypes AA, Aa and aa respectively.

PAPER VIII AND IX—ECONOMIC STATISTICS (PRACTICAL)

Time: 4 Hours — Full marks: 100

(*a*) Figures in the margin indicate full marks.
(*b*) Attempt all questions.
(*c*) Use of calculating machine is permitted.

1. The following table is taken from Shah Khambata's Wealth and Taxable Capacity in India.

items	quantity in million tons			value in Rs. crores		
	pre-war average	war and post-war average	1921-22	pre-war average	war and post-war average	1921-22
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. rice	32.5	33.3	35.1	354	541	697
2. wheat	8.4	9.3	10.0	80	142	207
3. barley	3.2	3.4	3.5	21	31	48
4. jowar	6.9	7.7	7.3	49	82	117
5. bajra	3.6	3.4	3.5	24	44	68
6. maize	2.4	2.8	3.0	23	32	41
7. ragi	3.3	3.4	3.5	16	30	47
8. gram	4.7	4.8	5.8	37	61	102
9. other foodgrains and pulses	10.3	10.6	10.5	65	105	147
total	75.3	78.7	82.2	669	1068	1474

Obtain all the value figures at constant pre-war prices. Compute the percentage contribution by different crops to the total value of all crops listed in the above table, both at current and at constant prices, separately for each of the three periods. Give your comments. (20)

[War here naturally refers to the First World War].

2. The following table gives the household consumer expenditure in rupees per month on food for various levels of household consumer expenditures collected in a survey of Faridabad Township in 1954:

level of household expenditure (in Rs.)	number of households	total expenditure in Rs. per household	expenditure on food in Rs. per household
(1)	(2)	(3)	(4)
8— 25	17	20.62	16.32
26— 50	61	39.57	29.31
51— 75	117	63.42	47.53
76—100	101	87.77	62.00
101—125	73	110.38	70.89
126—150	47	137.56	84.41
151—200	50	173.13	102.77
200 and above	34	291.24	147.15

Compute Engel elasticity of food expenditure (30)

3. (a) The following table gives the prices of 11 food items for the years 1939 and 1946 and the weights to be given to each item:

commodity	unit	price in Rs. per unit		weights
		1939	1946	
(1)	(2)	(3)	(4)	(5)
1. rice	seer	0.14	0.40	16.67
2. atta	-do-	0.13	0.34	3.24
3. fish	-do-	0.57	2.10	6.51
4. meat	-do-	0.60	2.06	2.29
5. pulses	-do-	0.18	0.78	2.28
6. eggs	score	0.53	2.20	0.53
7. potatoes	seer	0.12	0.64	2.98
8. other vegetables	-do-	0.10	0.30	4.28
9. mustard oil	-do-	0.48	1.43	3.42
10. ghee	-do-	1.59	5.85	4.24
11. sugar	-do-	0.29	0.64	1.88

Calcutta an index number of consumer's prices of the food items for 1946 with 1939 as base using figures in col. (5) as weights. Compute also the index number for 1939 with 1946 as base using the same weights and compare the reciprocal of this with the earlier index. (18)

(*b*) Average hourly wage rates in dollars for certain categories of workers in an industry are given below for 1950-53 with the typical distribution of weekly manhours.

categories	years				typical number of weekly manhours
	1950	1951	1952	1953	
foremen	1.60	1.73	1.80	1.84	800
machanics	1.45	1.59	1.67	1.70	2000
helpers	1.13	1.23	1.29	1.32	8000
tiremen	1.13	1.23	1.29	1.32	400
greasers	1.02	1.10	1.16	1.18	1200

Calcutta the index of hourly wage rates for each of the years 1951 to 1953 with 1950 as base. (12)

4. Make freehand sketches of the trends shown by the distribution given below and write out the forms of the equation of the curves which might give reasonably good fits (the parameters need not be estimated). For case (a) estimate also the parameters of the curves. (20) = 5 + 5 + 10

(*a*) Value of life insurance policies in force in the USA in billions of dollars

1880	·2
1890	4
1900	9
1910	16
1915	23
1920	42
1925	72
1930	108
1935	101
1940	118

(*b*) Number of horses on farms in the continental United States in millions.

1916	21.3
1920	20.1
1924	17.4
1928	14.8
1932	12.7
1936	11.6
1940	10.4
1944	9.3

PAPER VIII AND IX—STATISTICAL QUALITY CONTROL (PRACTICAL)

Time: 4 Hours | Full marks: 100

(*a*) Answer *any two* questions.
(*b*) All questions carry equal marks.
(*c*) Use of calculating machine is permitted.

1. The existing specifications for bursting strength on a certain type of paper are 105—110 pounds per square inch (psi).

The present method of testing requires that four samples be selected at random and tested. If any of the samples falls below the minimum specifications the lot is rejected.

There is considerable disagreement as to whether these specifications are reasonable and whether the present method of testing for acceptance is adequate. To investigate this problem, results of tests made on 25 consecutive lots of paper according to the present method have been recorded and given below.

lot no.	bursting strength (psi) on sample			
	1	2	3	4
1	102	104	114	98
2	104	107	112	96
3	107	108	113	112
4	107	109	107	101
5	113	106	112	95
6	110	112	102	120
7	110	102	120	112
8	109	93	98	100
9	89	94	110	92
10	105	124	95	105
11	109	110	104	104
12	102	96	105	104
13	95	94	110	105
14	117	118	107	110
15	115	97	104	110
16	107	97	106	97
17	97	106	97	114
18	110	111	109	117
19	105	111	95	120
20	115	110	112	110
21	106	102	103	98
22	116	92	118	104
23	93	108	100	108
24	105	91	93	97
25	112	112	108	96

Is the control over the characteristic during production satisfactory?
Assuming that no fundamental changes are possible in the process at present, are the present specifications reasonable?

What is the chance that a lot would fail to pass the sampling test method currently used?

Every test on a sample gives a value of psi that could be treated as a variable or that could be treated as an attribute (inside or outside specifications). Could you suggest a better method of treating these data in devising an acceptance sampling plan?

2. The following data relates to the number of defective insulators in samples of 200 items (a half shifts' production) giving 12 samples in a week of six working days.

first week		second week		third week	
sample number	defective insulators in 200 items	sample number	defective insulators in 200 items	sample number	defective insulators in 200 items
1	29	13	20	25	12
2	25	14	15	26	10
3	22	15	14	27	13
4	28	16	10	28	17
5	30	17	11	29	4
6	27	18	13	30	11
7	30	19	15	31	14
8	29	20	15	32	6
9	22	21	8	33	7
10	26	22	15	34	11
11	31	23	10	35	9
12	24	24	8	36	13

Certain changes were made in the production process at the beginning of the second week. Analyse the data, by using control charts, to examine if the changes in the process results in improvement in quality. Suggest a suitable control procedure Draw the OC curve for this procedure and read off from this the indifference quality.

3. Devise an attribute sequential sampling plan for

$$AQL = 0.03 \text{ and producer's risk} = 0.05$$
$$LTPD = 0.08 \text{ and a consumer's risk} = 0.10.$$

Draw the OC and ASN curves for the above plan.

Obtain the corresponding single sampling attribute plan and compare the amount of inspection under the two schemes when the lots submitted contain 5 percent defectives.

PAPER VIII AND IX—SAMPLE SURVEYS, APPLIED (PRACTICAL)

Time: 4 Hours Full marks: 100

(*a*) All questions carry equal marks.
(*b*) Use of calculating machine is permitted.

1. *Either,*

A family-budget survey will be conducted in some selected cities and towns of India for preparing weights for middle-class cost of living index. Prepare suitable schedules (for sample selection and enquiry proper) for the survey and write down necessary instructions for the guidance of the investigators.

Also, prepare suitable working sheets and proforma for final tables for compilation and presentation of such indexes for usual major groups and all items combined (items of different groups are to be clearly shown in the tables).

Or,

(*a*) Prepare suitable punching and sorting designs for mechanical tabulation of the data on family-budget mentioned in question 1.

(*b*) It is proposed to investigate a sample of 2000 budgets in Calcutta for the survey mentioned in question 1 and to prepare necessary weights for the city.

Prepare a budget for the survey and write a short note justifying staff and expenditure. (Take into consideration the following).

Monthly rental charges (per machine) of punching, verifying, sorting, tabulating ann auto-reproducing machines are Rs. 20, Rs. 30, Rs. 320, Rs. 2050 ann Rs. 350 respectively and cost per thousand of punching cards is Rs. 11. (A punching/verifying operator can perform 23 thousand strokes per day).

Basic pay-scales of different categories of staff are to be chosen from the list of standard scales shown in Enclosure A.

An employee is entitled to get montuly allowances in addition to basic pay as follows:

(*i*) Dearness allowances at the rate shown in Enclosure A.

(*ii*) Compensatory allowances at 10 percent of basic pay for all scales with a minimum upto Rs. 55.

(*iii*) Other allowances of Rs. 10 for all persons with basic pay upto Rs. 250.

2. Enclosure B shows the average monthly income of families residing in ten different holdings of a small urban area.

Draw a sample of six holdings and take samples of five families from each selected holding. The sampling at each stage is at random (with equal probability) and without replacement.

From the data of these 30 sample families estimate the total monthly income of all families of the area and also estimate its variance. Obtain also the 'true' value of this variance.

Enclosure A

I. Standard scales of pay (monthly in Rs.)

serial number	scale of pay
1.	350-30-680-40-1200
2.	250-20-650-25-750
3.	130-5-150-10-350
4.	125-5-240-10-250
5.	150-6-120
6.	100-4-180-5-200
7.	80-4-160-5-180
8.	70-3-118-4-150
9.	55-3-118-4-130
10.	20-½-25

II. Rates of Dearness Allowance.

basic pay	dearness allowance
upto Rs. 100	35
101–150	40
151–200	45
201–250	50
251–300	60
301–400	70
above 400	17½% of basic pay.

Enclosure B

Average monthly income of families residing in different holdings.

holding number	serial number of family	average monthly income (Rs.)
(1)	(2)	(3)
1	1	120
	2	30
	3	75
	4	150
	5	35
	6	75
	7	70
	8	75
2	1	60
	2	60
	3	60
	4	100
	5	70
	6	100
	7	100
	8	125
	9	80
	10	150
	11	45
	12	45
	12	90
3	1	40
	2	40
	3	150
	4	60
	5	20
	6	20
	7	50
	8	50
	9	90
4	1	60
	2	40
	3	80
	4	60
	5	110
	6	80
	7	90
	8	150
	9	80
	10	115
5	1	50
	2	90
	3	160
	4	200
	5	125
	6	100
	7	100
	8	140
	9	80
	10	80
	11	230
6	1	90
	2	90
	3	90
	4	105
	5	150
	6	60
	7	120
	8	90
	9	130
	10	60
	11	150
	12	195
7	1	115
	2	60
	3	60
	4	90
	5	60
	6	70
	7	80
	8	50
8	1	90
	2	80
	3	60
	4	60
	5	50
	6	60
	7	60
	8	30
	9	80
	10	35
9	1	50
	2	60
	3	85
	4	125
	5	15
	6	40
	7	80
	8	500
	9	100
	10	280
	11	80
	12	150
	13	110
10	1	60
	2	80
	3	85
	4	165
	5	50
	6	80
	7	150
	8	125

PAPER VIII AND IX—SAMPLE SURVEYS, THEORY (PRACTICAL)

Time: 4 Hours Full marks: 100

(*a*) Answer *all* questions.

(*b*) All questions carry equal marks.

(*c*) Use of calculating machine is permitted.

1. The following data relating to holding size, cultivated area and consumption of chemical fertilizers were collected in a survey of Fertilizer practices carried out in one taluk. A simple sample of 36 holdings gave the following information:

serial number	holding size (acres)	cultivated area (acres)	consumption of chemical fertilizers (lbs.)
(1)	(2)	(3)	(4)
1	21.04	2.70	0
2	12.59	1.76	0
3	20.30	1.47	48
4	16.16	1.64	54
5	23.82	1.56	0
6	1.70	1.79	0
7	26.91	5.44	544
8	7.41	1.97	117
9	7.68	2.45	0
10	66.55	3.26	0
11	141.80	3.20	197
12	28.12	3.00	102
13	8.29	1.95	0
14	7.27	2.20	160
15	1.47	0.48	0
16	1.12	1.12	0
17	10.67	1.98	0
18	5.94	2.45	0
19	3.15	2.60	0
20	4.84	1.67	36
21	9.07	5.13	406
22	3.69	3.69	0
23	14.61	1.34	0
24	1.10	0.25	0
25	22.13	2.70	248
26	1.68	1.21	56
27	49.58	4.48	216
28	1.68	1.38	42
29	4.80	1.39	66
30	12.72	2.15	180
31	6.31	2.12	74
32	14.18	1.49	182
33	22.19	3.80	222
34	5.50	1.75	102
35	25.29	5.17	224
36	20.09	1.50	40

Estimate the proportion of holdings in the four holding-size classes

0—4.99 ; 5.00—9.99 ; 10.00—24.99 ; 25 and above.

Also estimate for each holding size class the average consumption of fertilizers per acre of cultivated area. Obtain the standard errors of all the above estimates. State whether different holding-size classes significantly differ with respect to their chemical fertilizer consumption per acre of cultivated area.

2. With the object of estimating the average catch of fish landed per operating fishing unit on certain representative parts of the Indian coast, the following scheme of sampling was adopted:

All the fish landing centres were grouped under two strata (A and B), of equal number of centres on the basis of geographical contiguity of centres. From among the large number of centres in each stratum two independent samples of 5 centres each were selected for observations during each of two time periods (I and II) which together comprise the full fishing season. At each of the centres so selected under each time period, an independent sample of 3 operating units was selected from among the large number of such units there. The selection at all stages was according to a simple random sampling procedure. The weights (in seers) of catches landed by the selected units are given below:

serial number of selected centre	Time Period I			Time Period II		
	total catch (seers) of selected units			total catch (seers) of selected units		
			stratum A			
1	610	454	618	112	124	200
2	297	400	515	56	75	58
3	1180	112	480	187	160	363
4	1297	533	1130	226	175	520
5	860	357	657	67	137	616
			stratum B			
1	1085	754	980	434	204	185
2	817	736	926	160	101	290
3	900	616	109	391	565	150
4	510	320	412	420	186	617
5	906	900	735	144	36	21

The total number of operating units is the same for all centres in a given time period, but the number of operating units at any centre in Time Period I is twice that in time period II. Obtain an estimate of the average catch per unit for the entire coast and full fishing season.

Give also the standard error of the estimate.

For a given number of centres to be selected what is the break up of these centres between the four substrata (2 geographical strata × 2 time periods) when the number of units selected at a centre is (*a*) 3 and (*b*) 4. Calculate the number of centres and the optimum break-up necessary to estimate the average catch with 3 percent standard error.

PAPER VIII AND IX—DESIGN OF EXPERIMENTS, APPLIED (PRACTICAL)

Time: 4 Hours | Full marks: 100

(a) Attempt *both* the questions.
(b) All questions carry equal marks.
(c) Use of calculating machine is permitted.

1. An experiment was conducted to study the responses to different levels of phosphate (P) on three varieties (V_1, V_2, V_3) of potatoes.
Phosphate was applied at six different levels: 0, 20, 40, 60, 80 and 100 lbs. per acre. The design employed was a split-plot arrangement where the varieties are the whole plot treatments and the levels of phosphate comprise the subplot treatments, and the whole experiment being replicated 5 times. The yield of potatoes (in thousand lbs. per acre) are given in the table below.

Carry out the appropriate analysis of these data and test for the linear as well as the quadratic effects of the phosphate.

Can you conclude from this experiment that there is an optimum level of application of the phosphate for which the corresponding yield would be a maximum? If, so, determine this optimum levels.

Yield of potatoes in thousand lbs. per acre

replication		lbs. of phosphate applied per acre					
		0	20	40	60	80	100
V_1	1	13.3	12.7	13.0	14.5	12.8	13.9
	2	12.8	13.3	13.5	14.7	13.9	13.1
	3	12.0	13.1	13.7	14.3	14.0	12.8
	4	12.4	12.9	14.0	13.8	12.9	13.0
	5	12.6	13.2	13.7	13.3	14.3	12.7
V_2	1	12.0	13.5	14.2	15.1	13.4	12.7
	2	12.3	13.1	14.0	14.8	12.7	12.8
	3	13.2	13.0	13.7	14.0	11.9	13.2
	4	12.5	12.8	13.5	13.5	12.6	12.1
	5	12.1	12.7	13.0	13.7	12.9	12.9
V_3	1	14.6	14.5	14.8	15.8	14.6	14.2
	2	14.3	14.6	14.7	14.8	13.7	14.3
	3	13.8	13.9	14.1	15.2	13.0	13.8
	4	13.3	13.1	13.5	15.7	13.5	12.9
	5	13.4	14.3	13.9	14.3	12.7	13.5

2. The following table gives the plan and yields (in bushels per acre) for a 5×5 simple lattice experiment on soyabeans, the treatments being the 25 varieities of soyabeans. The experiment is carried out in four replications. Analyse the data to find if there is any significant difference between the varieites. Also, calculate the estimates of the error variances for the difference between two treatment means when (1) they occur in the same block and when (2) they do not occur in the same block.

Yield of a 5×5 simple lattice experiment on soyabeans.
(figures in brackets represent the varieties)

Replication I

(1)	1.6	(2)	1.7	(3)	1.5	(4)	1.8	(5)	1.6
(6)	2.6	(7)	2.2	(8)	2.2	(9)	2.3	(10)	1.8
(11)	2.7	(12)	1.7	(13)	1.7	(14)	1.9	(15)	2.4
(16)	2.8	(17)	2.6	(18)	2.3	(19)	2.3	(20)	2.4
(21)	2.4	(22)	2.5	(23)	2.1	(24)	2.4	(25)	2.4

Replication II

(1)	3.4	(6)	2.3	(11)	3.4	(16)	2.1	(21)	1.8
(2)	3.1	(7)	2.1	(12)	2.4	(17)	2.2	(22)	3.3
(3)	2.6	(8)	1.4	(13)	2.2	(18)	2.2	(23)	2.2
(4)	2.7	(9)	2.0	(14)	4.0	(19)	1.9	(24)	3.3
(5)	2.5	(10)	2.5	(15)	3.2	(20)	2.6	(25)	2.9

Replication III

(1)	2.3	(2)	3.6	(3)	1.9	(4)	2.3	(5)	2.1
(6)	2.5	(7)	2.8	(8)	3.2	(9)	2.1	(10)	2.5
(11)	2.9	(12)	2.0	(13)	2.0	(14)	2.0	(15)	2.6
(16)	3.1	(17)	2.6	(18)	2.7	(19)	1.4	(20)	2.7
(21)	2.5	(22)	2.2	(23)	2.3	(24)	3.0	(25)	1.8

Republication IV

(1)	2.6	(6)	1.7	(11)	3.0	(16)	2.3	(21)	3.1
(2)	2.5	(7)	2.0	(12)	2.1	(17)	1.7	(22)	2.4
(3)	1.7	(8)	2.1	(13)	2.5	(18)	2.5	(23)	2.6
(4)	2.9	(9)	2.4	(14)	3.0	(19)	1.6	(24)	2.6
(5)	2.7	(10)	2.8	(15)	3.0	(20)	2.5	(25)	2.5

PAPER VIII AND IX—VITAL STATISTICS AND POPULATION STUDIES (PRACTICAL)

Time: 4 Hours — Full marks: 100

(*a*) Attempt *any two* questions.
(*b*) All questions carry equal marks.
(c) Use of calculating machine is permitted.

1. Prepare an abridged Life Table from the following data on population and death rates. You may assume (with the usual Life Table notations) that

$$L_0 = 0.25\, l_0 + 0.75\, l_1$$

$${}_4L_1 = 1.9\, l_1 + 2.1\, l_4$$

$$L_{85} = l_{75} \times \log (l_{85}); \text{ and}$$

for other age groups $${}_nL_x = \frac{n}{2}(l_x + l_{x+n}).$$

age last birthday (years)	population	death rate/1000 population
0	40,000	68.3
1 — 4	160,100	6.4
5 — 9	197,400	2.1
10—14	208,500	1.3
15—24	424,800	2.7
25—34	416,000	3.3
35—44	369,500	4.6
45—54	309,800	8.5
55—64	253,200	19.7
65—74	167,200	48.6
75—84	65,700	122.3
85 and above	8,300	317.3

You may use the values given in appendix on page 37.

2. (*a*) Fit a logistic curve to the Census data of the USA from 1800 to 1910 utilizing the full material.

year	population (million)
1800	5.308
1810	7.240
1820	9.638
1830	12.866
1840	17.069
1850	23.192
1860	31.443
1870	38.558
1880	50.156
1890	62.948
1900	75.995
1910	91.972

From the fitted curve, estimate the population in 1914.

(*b*) The following table shows separately for the rural and urban sectors of West Bengal (1956) (*i*) the average number of children born and number of children surviving by age of mother as estimated from two interpenetrating sub-samples 1 and 2 of a sample survey. The estimates provided by the combined sample (i.e. sample obtained by pooling together sub-samples 1 and 2) are also shown. (*ii*) The age distribution of mothers in West Bengal as a whole (rual and urban sectors combined) is shown in the second table.

Explain how the latter information can be utilised for purposes of 'standardization' with a view to obtaining comparable figures for (*i*) average number of children born. (*ii*) surviving, per mother of all (13 and above) age groups. Illustrate by actually computing the standardised figures.

Do the sub-sample data throw any light on the margin of uncertainity of these standardised figures?

[N.B. The term 'mother' includes also married women without children]

Average number of children per mother

age of mothers (years)	sub-sample 1		sub-sample 2		combined	
	born	living	born	living	born	living
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			rural West Bengal			
13—17	0.32	0.26	0.12	0.12	0.20	0.17
18—22	1.24	1.13	1.47	1.08	1.35	1.10
23—27	3.07	2.61	2.70	2.07	2.91	2.35
28—32	4.40	3.25	4.30	3.24	4.35	3.25
33—37	5.16	3.21	4.38	3.48	4.91	3.32
38—42	5.83	3.71	5.32	4.00	5.65	3.85
43+	6.30	4.30	5.45	3.55	0.16	3.97
			urban West Bengal			
13—17	0.09	0.00	0.25	0.13	0.16	0.05
18—22	1.09	1.02	1.19	1.13	1.14	1.08
23—27	2.23	1.80	2.09	2.22	2.47	2.20
28—32	3.71	2.71	2.98	2.28	3.39	2.52
33—37	4.62	3.62	3.75	2.96	4.17	3.28
38—42	5.10	4.35	5.68	4.95	5.40	4.67
43+	3.86	3.24	5.55	4.60	5.15	4.26

age of mothers (years)	percentage distribution of mothers in West Bengal
13—17	10.35
18—22	22.73
23—27	16.82
28—32	15.16
33—37	10.17
38—42	10.91
43+	13.86
	100.00

3. The smoothed age distribution of the Census 1951 population for India is given sex-wise for the individual ages from 0 to 19 years. The survival probabilities are also shown. Obtain the projected sex-wise population for each of the above individual ages 0-19 years in 1952 and 1953, given the number of females aged 15–44 years in 1951 = 79,285 thousand and number of females aged 15–44 years in 1952 = 80,184 thousand; general fertility rate (live birth per female aged 15–44 years) = 0.185; sex ratio at birth (number of male live births per female live birth) = 1.06 ; survival probability for the new born = 0.8626 for males and 0.8866 for females.

age last birthday (years)	number in 1951 (thousand)		survival probability	
	male	female	male	female
0	5863	5697	.9086	.8899
1	3967	3934	.9526	.9367
2	4707	4665	.9700	.9600
3	4746	4832	.9780	.9718
4	4658	4579	.9804	.9772
5	4803	4720	.9833	.9803
6	4720	4591	.9842	.9819
7	4635	4466	.9851	.9835
8	4547	3345	.9859	.9849
9	4458	4229	.9866	.9862
10	4366	4117	.9872	.9875
11	4274	4010	.9878	.9886
12	4180	3908	.9884	.9896
13	4086	3810	.9888	.9905
14	3992	3716	.9892	.9913
15	3895	3630	.9895	.9919
16	3794	3553	.9898	.9924
17	3689	3484	.9899	.9926
18	3587	3414	.9898	.9926
19	3488	3343	.9896	.9923

Appendix

values from Reed—Merrel tables for abridged life Tables.

m_0	q_0
0.068	0.060416
0.069	0.061241

$4m_1$	$4q_1$
0.006	0.022967
0.007	0.026678

$5m_x$	$5q_x$
0.001	0.004989
0.002	0.009954
0.003	0.014897

$10m_x$	$10q_x$
0.002	0.019833
0.003	0.099025
0.004	0.039334
0.005	0.048961
0.008	0.077356
0.009	0.086661
0.019	0.175426
0.020	0.183885
0.048	0.392518
0.049	0.399029
0.122	0.737911
0.123	0.741027

PAPER VIII AND IX—MATHEMATICAL THEORY OF SAMPLING DISTRIBUTIONS (PRACTICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *any two* questions.

(*b*) All questions carry equal marks.

(*c*) Use of calculating machine is permitted.

1. (*a*) Draw 12 samples each of size 9 from a bivariate Normal population with $\mu_1 = \mu_2 = 5$, $\sigma_1 = \sigma_2 = 1$ and $\rho = 0.6$ and compute the correlation coefficient for each of the samples.

(*b*) Using the results of the model sampling experiments in (*a*) above estimate the power of the test for independence in a bivariate Normal population when the level of significance is 5 percent and the value of the correlation coefficient in the population is $\rho = 0.6$.

2. A particle performs a random walk on the points $x = 0, 1, 2, 3$. The particle is absorbed when it reaches the point 0 or 3. The particle starts from $x = 2$ and the probability of a step to the right is .7 and that to the left is .3. Find the probability generating function of the number of steps before the particle is absorbed at $x = 0$ or $x = 3$.
Estimate $E(n)$ and $V(n)$ by carrying out a model experiment to obtain 10 observations on n.

3. The observed gestation period (in days) of several calfiings of 6 cows are presented below.

cow	gestion period (in days)				
1	282,	277,	279,	281,	286
2	282,	283,	283,	282,	276
3	283,	281,	284,	281,	285
4	284,	280,	289,	292,	290
5	275,	284,	293,	303,	284
6	303,	304,	284,	284,	292

Calculate the coefficient of intra class correlation coefficient of gestation period of cows. Determine also the 90 percent confidence interval of the population intra class correlation coefficient.

PAPER VIII AND IX—PROBIT ANALYSIS (PRACTICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *any two* questions.
(*b*) All questions carry equal marks.
(*c*) Use of calculating machine is permitted.

1. For estimating the relative potency of a sample of 'A calamus' sent by a manufacturer against the specified standard preparation, 4 does of each of the preparations were taken together with a control dose for each preparation. House flies were taken in batches of 5 and sprayed on dishes. Each concentration was used on five dishes giving a total of 25 flies subjected to each concentration. For each control dose 20 dishes were taken giving a total of 100 flies subjected to each of the control doses. Mortality among the experimental flies was observed after 24 hours. The observations are given below:

concentrations (gm. per 100 cc)	number dead	
	test preparation	standard preparation
.3	16	23
.15	8	20
.075	7	16
.0375	4	12
control	4	5

Assuming that log tolerance is normally distributed, estimate the relative potency of the test preparation of 'A. Calamus'.

2. Data from an assay of insulin by mouse convulsion method are given below:

dose (i.u)	number of mice	number convulsed
.004	12	1
.006	24	8
.008	24	15
.009	24	15
.0135	10	8

Estimate the linear relationship between log dose and (1) probit and (2) logit of fraction convulsed separately. Examine graphically or otherwise as to which is a better transformation.

3. The data presented below were obtained from mortality tests of groups of 6 subjects at each of 9 doses of a stimulus:

log dose (x)	number of subjects	number dead
1	6	0
2	6	0
3	6	1
4	6	0
5	6	2
6	6	4
7	6	4
8	6	6
9	6	5

Estimate the mean of the log tolerance distribution from the above data through the Spearman—Karber method.

PAPER VIII AND IX—PSYCHOLOGY AND EDUCATION (PRACTICAL)

Time: 4 Hours — Full marks: 100

(*a*) Answer *all* questions.
(*b*) All questions carry equal marks.
(*c*) Use of calculating machine is permitted.

1. The following table gives the frequency distribution of scores in a particular test of 200 boys and 200 girls.

score	frequency	
	boys	girls
55—59	2	1
50—54	25	9
45—49	48	27
40—44	47	44
35—39	26	43
30—34	19	40
25—29	15	20
20—24	9	10
15—19	7	4
10—14	2	2

(*i*) Draw the percentile curves.

(*ii*) From the graph, compare the performances of boys and girls.

2. Fifteen subjects are divided at random into 3 groups of 5 subjects each. One group is taught by method A, another by method B, and the third by method C. Retention for the material presented is tested at 3 different times: immediately after learning period, 24 hours after learning period and 1 week after learning period. The series of retention scores for the subjects in the various groups are given below:

	Subjects	Time of Testing		
		immediate	24 hours	1 week
method A	1	28	25	22
	2	32	29	24
	3	36	35	27
	4	45	42	40
	5	46	43	40
method B	6	27	24	20
	7	29	26	22
	8	36	36	30
	9	42	43	39
	10	48	44	40
method C	11	40	38	33
	12	36	26	20
	13	50	48	44
	14	45	43	35
	15	42	39	30

(*i*) Analyse the total sum of squares into its component parts and test the various mean squares for significance.

(*ii*) Write a note on the results obtained.

3. The following table gives the inter test correlations of seven tests.

	1	2	3	4	5	6	7
1	–	.685	.545	.532	.551	.471	.501
2	–	–	.475	.450	.461	.404	.472
3	–	–	–	.576	.527	.316	.375
4	–	–	–	–	.280	.341	.431
5	–	–	–	–	–	.338	.444
6	–	–	–	–	–	–	.519
7	–	–	–	–	–	–	–

(*i*) Estimate the communality of each test.

(*ii*) Calculate the factor loadings and write down the equation of the factor pattern.

(*iii*) If possible, obtain a 'simple structure'.

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER, 1961.

PAPER I—THEORETICAL STATISTICS, GENERAL

Time: 4 Hours Full marks:100

(*a*) Attempt *any five* questions.

(*b*) Figures in the margin indicate full marks for each question.

1. (*a*) State and prove the Bienayme-Tchebycheff's inequality. (10)

(*b*) Explain the statistical implications of the inequality. (5)

(*c*) Is it possible to improve the inequality? Give reasons for your answer. (5)

2. Explain the terms (*i*) probability density function, (*ii*) probability differential, (*iii*) mathematical expectation. (6)

A batch of small calibre ammunition is accepted as satisfactory if none of a sample of five shots falls more than 2 feet from the centre of the target at a given range. Assume that the probability density function of r, the distance from the target centre of a given impact point, is

$$f(r) = re^{-\frac{r^2}{2}} \quad (0 \leq r < \infty).$$

What is the probability that the batch will be accepted? (10)

Find also the expected value of r. (4)

3. Derive the Poisson distribution as the limiting form of the binomial distribution. (8)

Assume that the probability of a motor failure on a routine flight between two cities is 0.005. Use the Poisson approximation to the binomial to find approximately the probability of at least one failure in 1000 flights. (6)

If x is a Poisson variate with $P(x=1) = P(x=2)$, find $P(x=3)$. (6)

4. (*a*) Find the form of the function u_x given that $u_0 = 8$, $u_1 = 11$, $u_4 = 68$, $u_5 = 123$. (10)

(*b*) Show that

$$\int_{1\cdot5}^{2\cdot5} v_x dx = \frac{1}{24}(v_1 + 22v_2 + v_3) \text{ approximately} \qquad (10).$$

5. (*a*) How do you test the hypothesis (*i*) that a partial regression coefficient of k-th order has a given value ρ and (*ii*) that the partial correlation of k-th order is zero? (6)

(*b*) Show that if $\rho = 0$, the two tests in (*a*) are identical. (5)

(c) The hypotheses that (*i*) the population multiple correlation coefficient is zero and (*iii*) all the population partial regression coefficients are zero, are the same. Prove this. (9)

6. (*a*) What are the advantages of a comparatively large sample over a small sample? (10)

(*b*) Explain the uses of transformation of statistics. (5)

(*c*) Derive the appropriate transformation for the large sample test of the correlation coefficient. (5)

7. Give the algebraic set up for the anlysis of variance for one-way classification, stating clearly the assumptions involved. Show how you will apply it to test the linearity of regression. (8+2+4)

A random sample of 150 pairs of values from a bivariate normal population when grouped in 15 arrays of y's gave values $r=0.4$ (correlation coefficient) and $\eta_{y_x}=0.5$ (correlation ratio). Test whether these results are consistent with the assumption of linearity of regression of y on x. (6)

PAPER II— APPLIED STATISTICS, GENERAL

Time: 4 Hours Full marks: 100

(*a*) Attempt *any five* questions.
(*b*) All questions carry equal marks.
(*c*) Figures in the margin indicate full marks for each question.

1. How will you prepare a plan for a 8×8 Latin square? When the results of such an experiment arrive, you discover that one of the plot yields is missing. How will you complete the analysis of such an experiment? Indicate the method for working out the standard error of the difference of two treatment means, one of which relate to the missing value. (20)

2. Discuss critically the details of the present method of estimation of national income by industrial origin in India. Draw particular attention to the sources of data used bringing out clearly gaps in data requirements for various sectors. What would you propose to fill up these gaps? (20)

3. What are the main uses of a retail price index? Do we have a retail price index in this country? If so, give the necessary details of this index. (20)

You are required to design a fresh enquiry for collecting data for the construction of such an index. Discuss the necessary technical details for the completion of such an index. How will you use the data from such an enquiry for the study of demand analysis?

4. Suppose you are required to estimate the rate of profit earned and the proportion of profit invested by the joint stock companies in India in 1960, on the basis of a sample survey.

Describe clearly the different steps of the sample design that you will suggest for the purpose, mentioning the frame to be used. (8)

What types of information will you collect? (4)

How will you obtain the estimates of the above characteristics and their sampling errors? (8)

5. What do you mean by selection with probability proportionate to the size of the sampling unit? How will you select 50 villages out of 1000 villages in a region, with probability proportionate to population in the village? (4+6)

From each of these 50 villages, 10 households are selected at random and number of members in the household and expenditures on various items in a given month are recorded. How will you estimate per epita expenditure on an item (say cloth) in this region? (10)

6. Describe the construction of a life table giving the meanings of the figures in various columns. How will you obtain q_x from census data? (10+10)

7. Under what situation is it necessary to use incomplete block designs in experimentation? Give a description of a balanced incomplete block design and state the relations that must hold good between number of blocks, number of varieties etc. Give a sketch of the method of analysis of such a design. (4+8+8)

8. Explain clearly the basic hypothesis underlying factor analysis and the concepts of communality of factor loading. Briefly describe the method of extraction of factors by the centroid method. (10+10)

9. Explain the principle of random mating. The frequencies in per cent of the blood group alleles in a population were computed to be IA, 20.62; IB, 7.56; i, 71.82. What are the expected phenotype frequencies in this population, on the assumption of random mating?

The phenotype frequencies actually found were O : 1337, A: 894, B: 309, AB : 70. Test the agreement of these frequencies with those expected. (4+8+8)

PAPER III—STATISTICAL INFERENCE

Time: 4 Hours Full marks: 100

(*a*) Attempt *any four* questions
(*b*) All questions carry equal marks.

1. (*a*) Show that the coefficient of correlation between two unbiased minimum variance estimators T_1 and T_2 (calculated from a sample of size n) of a parameter θ is unity.

(*b*) For the above T_1 and T_2, show that the probability is zero that $|T_1 - T_2| > \frac{1}{m}$ for at least one value of m, m being an integer.

(c) If $f(x, \theta)$ is the p.d.f. of x, find an expression for the minimum variance attainable by any unbiased estimator of θ.

2. (a) Examine whether the following distributions admit of sufficient statistics for the parameters mentioned:

(i) $f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp. \left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$, $-\infty < x < +\infty$, for the parameter μ.

(ii) $f(x, p, \alpha) = \frac{\alpha^p}{\Gamma(p)} e^{-\alpha x} x^{p-1}$, $x > 0$, for the parameter α

(iii) $f(x; \beta) = \frac{1}{\beta}$, $0 < x < \beta$ for the parameter β.

(b) Show that if t is any unbiased estimator and T a sufficient statistic for θ, then $E(t|T)$ is an unbiased estimator of θ which has a variance smaller than that of t.

(c) $x_1, \ldots, x_n$ is a random sample from a normal population $N(\mu, 1)$. It is desired to estimate $\mu^2 + 1$.

(i) Show that $t = \frac{1}{n} \Sigma x_i^2$ is an unbiased estimator.

(ii) Given that $\bar{x}$ is a sufficient statistic for μ, find an unbiased estimator of $\mu^2 + 1$ which has a variance smaller than that of t.

3. Write a critical note on the method of maximum likelihood in the theory of estimation. Describe an interative method for solving the maximum likelihood equations and for obtaining the variance and covariances of the estimates.

A person is required to shoot at a target till he scores two successes. Suppose he needed n trials: how do you estimate p, the probability of success in a single trial, assuming that the trials are all independent and p remains the same from trial to trial.

4. (a) Define a most powerful test and an unbiased test for a simple hypothesis.

(b) Examine whether a most powerful test is necessary unbiased.

(c) State Neyman-Pearson's lemma and show how it can be used to construct a locally most powerful test of a simple hypothesis regarding a single parameter.

(d) Obtain a locally most powerful test for the hypothesis $\mu = \mu_0$ where μ is the parameter of a normal population $N(\mu, 1)$.

5. Give a brief description of the theory of confidence intervals as developed by Neyman.

Apply the theory to determine the 95 percent confidence interval, based on a sample of size n, for the mean of a normal population when the variance is unknwon.

6 Describe the tests you would use for the following hypotheses for p-variate normal populations.

(*i*) the partial correlation coefficient $\rho_{12.....p} = 0$.

(*ii*) $\mu_1 = \mu_1^0, \ldots, \mu_p = \mu_p^0$, where $\mu_1, \ldots, \mu_p$ are the means of the p-variates and $\mu_1^0, \ldots, \mu_p^0$ are preassigned values.

(*iii*) $\mu_1 = \mu'_1, \ldots, \mu_p = \mu'_p$, where $\mu_1, \ldots$ and $\mu_p \mu'_1, \ldots, \mu'_p$ are the p means of two p-variate normal distributions respectively.

(*iv*) $\beta_2 = \ldots = \beta_p = 0$ where $\beta_2, \ldots \beta_p$ are the coefficients of regression of x_1 on $x_2, \ldots, x_p$.

PAPER VI—(PRACTICAL)

Time: 6 Hours — Full marks: 100

(*a*) Answer question 1 and *any three* of the rest.
(*b*) Figures in the margin indicate full marks for each question.
(*c*) Use of calculating machine is permitted.

1. (*a*) The frequency distribution of examination marks in mathematics of 500 students suggested that Pearson's type III curve fits best. The fitted function is

$$y = y_0\left(1 + \frac{x}{109.47}\right)^{33.19} e^{-0.31x}; \quad -109.47 < x < \infty$$

with origin at mean = 42.57 [*i.e.* $x = (X - 42.57)$], y = frequency X = marks.

Find by consulting Biometrika tables the proportion of students who have scored 60 or more marks. (10)

(*b*) The coefficient of variation (y) of 'proportion of area under paddy' as estimated from samples of straight armed line units of different lengths (x) in furlongs are given below:

x	1	2	3	4
y	137	118	108	102

(*i*) Fit a curve of the form $y = ax^b$ and draw the graph of the observed and fitted curve. (10)

(*ii*) Using the graduated values of y, estimate the number of units of size x to be taken for a random sample of units to expect an estimate of 'average proportion of area under paddy' within an error of 5 percent for $x = 1, 2, 3, 4$. (10)

(*c*) The following table gives the value of a function $y = {}_0F_1\ (c, x)$ for $c = 0.5$ and $x = 4$ to 7 at intervals of 1.

x	4	5	6	7
y	27.3082	43.7775	67.0801	99.3234

Using $\log_{10} y$ correct to four decimal places as the variable, find $\log_{10} y$ for $x = 5.5$ by a suitable method of interpolation and hence find $y = {}_0F_1\ (0.5, 5.5)$. (10)

2. The frequency distribution of marks in English of a random sample of 500 students is given below :—

score	frequency	score	frequency
0— 4	3	40—44	78
5— 9	6	45—49	53
10—14	12	50—54	46
15—19	23	55—59	29
20—24	35	60—64	18
25—29	45	65—69	5
30—34	74	70—74	1
35—39	72		
		total	500

(*a*) Test whether the distribution could be assumed normal. (14)

(*b*) If 80 marks are required for obtaining distinction, what percent of students could have secured distinction in the entire population? (6)

3. In a public preference survey, a number of persons were interviewed and their opinions on abolition of English as a compulsory subject in the university curicula were recorded.

Distribution of persons according to age and opinion

opinion	age in years			
	19—25	26—35	36—55	over 55
unconditional support	76	125	96	10
conditional support	69	117	126	17
indifferent	14	27	35	4
unconditionally opposed	60	108	210	46

Examine whether, and in what way, the nature of opinion changes with age. (20)

4. The following table gives the yield of latex for 9 rubber plants under uniform conditions in 1950 and under the three treatments A, B, C, each applied to three plants in 1951. Analyse the data for examining differences between treatments. (20)

Plan No.	1	2	3	4	5	6	7	8	9
Yield in 1950	15	6	15	6	15	10	8	4	12
Yield in 1951 and treatment applied	16 (A)	23 (B)	18 (B)	17 (A)	24 (C)	25 (C)	16 (B)	29 (C)	12 (A)

5. In an experimental study to build up a formula for predicting the yield of dry bark in ounces (x_1) of a plant from its height in inches (x_2) and its girth in inches (x_3) at a level six inches above the ground, measurements were taken on 32 plants and the means and corrected sums of squares and products were computed as given in table below.

character	mean	corrected sums of squares and products		
		x_1	x_2	x_3
x_1	21.65	6498.00	9483.84	321.26
x_2	166.40		102767.64	599.58
x_3	3.14			33.95

(*a*) Estimate the linear equation for predicting x_1 from x_2 and x_3. Find the multiple correlation coefficient between x_1 and x_2, x_3 and test for its significance. (10+5)

(*b*) Test the significance of the coefficient of x_2 in the linear prediction equation. (5)

PAPER VII—(PRACTICAL)

Time: 6 Hours Full marks: 100

(*a*) Attempt *all* the questions.

(*b*) Figures in the margin indicate full marks for each question.

(*c*) Credit will be given for brevity, neatness, tabular arrangement of computations and brief statement of essential steps and underlying assumptions.

1. The table gives the quarterly volumes of civil aviation traffic in billion passenger kilometers in USA, 1954-59.

(*a*) Find the yearly total volumes and their annual rate of growth by fitting straight line trend to the totals. (8)

(*b*) Compute indices of seasonal variation (quarterly) in the volume of traffic. (13)

(*c*) Using results obtained in (*a*) and (*b*) above, estimate the volume of traffic in the first quarter of 1960. (6)

(*d*) Compile from official publications the actual figure for the first quarter of 1960 and compare it with the estimate you obtained in (*c*). (6)

Civil aviation traffic, USA, 1954-59
(billion passenger kilometers)

years / quarters	1954	1955	1956	1957	1958	1959
1. Jan-Mar	7.07	8.60	9.82	11.22	11.64	12.71
2. Apr-June	8.56	10.13	11.56	13.02	13.02	14.87
3. July-Sept	9.14	10.09	12.32	14.21	14.42	16.72
4. Oct-Dec	8.26	9.47	10.76	11.85	11.62	14.24

2. *Either,*

A randomised block experiment was conducted in 8 blocks, each of 4 plots, to examine the effects of two chemical fertilisers N and P on the yield of potatoes. Two doses of N: n_0(no N) and n_1 (0.5 lbs. of N per plot) and two doses of P: p_0 (no P) and p_1 (0.4 lbs. of P per plot) were tried in all the four possible combinations. The average yields in lbs. per plot are given below.

Average yields of potatoes in lbs/Plot

P	N	
	n_0	n_1
p_0	38.25	47.50
p_1	45.75	52.50

Analysis of variance of the experimental result gave 10.532 as the *mean square* due to error.

(*a*) Compute the sum of squares due to treatments and analyse it into the three components (*i*) due to the main effect of N (*ii*) due to the main effect of P and (*iii*) due to the interaction between N and P. (6)

(*b*) Draw up a table showing the structure of the complete analysis of variance: the degrees of freedom for the different components and the mean square due to as many of these components as are possible to obtain from the data supplied to you. Test the significance of the mean squares computed in (*a*). (4)

(*c*) Given that the prices, of equal weights, of potatoes and the chemical fertilizers N and P are in the ratios 1:8:10, what is your recommendation about the best combination of the two fertilizers? (5)

(*d*) Write a short report (for non-statistician) on your findings. (5)

Or,

An experiment was run to compare three similar magnesium salts L, M, N in the production of an antibiotic by fermentation. In the first replication, three fermentations were started, one containing salt L, another salt M, and the third salt N. After five days, samples for analysis were withdrawn from each fermentation and likewise after six days. The whole operation was repeated a total of four times. The replications should be regarded as blocks.

(*a*) Make an appropriate analysis of variance for these data, and report the F-values, with the corresponding degrees of freedom, for the main effects of salt, age (five days versus six days) and their interaction.

(*b*) Give 95 percent confidence limits for the following differences (*i*) salt L—salt M, (*ii*) six days—five days, (*iii*) salt L—salt M, both at six days.

replication	magnesium salt					
	L		M		N	
	5 days	6 days	5 days	6 days	5 days	6 days
1	69	84	91	98	81	86
2	82	78	75	82	72	77
3	67	74	78	92	66	79
4	69	77	85	92	73	81

(20)

3. The variance V of the customary estimate of the mean, in the case of stratified simple sampling is given by

$$V = \frac{1}{N^2} \sum_{i=1}^{k} \frac{N_i^2 \sigma_i}{n_i}$$

where k is the number of strata, N_i is the member of units in the i-th stratum of which n_i are sampled, and σ_i^2 is the variance between the units in the i-th stratum, $i = 1, 2, \ldots, k$. The total cost C for the survey may be approximated by the formula

$$C = c_0 + \sum_{i=1}^{k} c_i n_i$$

where c_0 is the overhead cost and c_i the cost for sampling and processing an unit from the i-th stratum.

In a certain sample survey, the population was divided into $k=4$ strata. A pilot survey was conducted to determine the parameters of the variance-function and the cost-function and this gave the following estimates:

i	N	σ_i^2	c_i (in rupees)
1	500	108.7	12.61
2	350	234.5	8.43
3	650	176.9	7.26
4	400	228.1	11.72

c_0 = Rs. 2500

(a) If the total cost for the survey is fixed apprixmately at C = Rs. 10,000, determine the sizes (n_i) of the samples to be drawn from the different strata so that the variance V of the estimate of the mean is minimum. For these values of n_i's compute the actual cost C (to the nearest rupee) and the variance V (to the four places of decimals). (12)

(*b*) If instead it is decided to allocate the variable part of the cost of the four strata as nearly equally as possible (to the nearest rupee) what would be the actual total cost C and the variance V? (8)

4. (*a*) The following table gives averages (X) and ranges (R) for subgroups of size 4. To facilitate arithmetic, deviations from the specified standard are given. Specified tolerances on individual items were ± 8 units from the standard. Plot suitable control charts and give your comments. Is the process working according to specification? (15)

subgroup no.	1	2	3	4	5	6	7	8	9	10
X	+ 4.6	+ 6.5	+ 0.0	+ 6.1	− 2.5	− 3.1	+ 3.6	+15.0	+ 7.0	− 4.0
R	2	11	4	6	8	5	5	8	11	10
subgroup no.	11	12	13	14	15	16	17	18	19	20
X	+ 3.4	− 0.6	− 2.0	− 2.0	− 7.0	−10.0	+ 7.0	− 6.6	+ 3.0	− 4.4
R	4	15	6	6	4	16	10	21	9	11
subgroup no.	21	22	23	24	25	26	27	28	29	30
X	+ 4.6	+ 4.6	+ 0.4	− 3.0	− 1.6	+ 0.6	+ 6.0	− 8.0	− 2.0	− 2.4
R	5	6	7	15	12	14	25	25	10	11
subgroup no.	31	32	33	34	35	36	37	38	39	40
X	− 7.0	−16.0	−14.0	− 1.0	− 9.6	− 8.6	− 8.4	−10.0	− 8.4	−10.0
R	14	16	15	15	7	11	12	10	12	13

(*b*) Fill up the missing entries (.) in the following life table: (10)

x	l_x	d_x	1000 q_x	L_x	T_x	e^0_x
26	78045	.	.	.	.	39.55
27	77615	440	.	.	.	.
28	.	.	.	.	.	.
29	76724	.	6.05	.	.	.
30	.	.	.	.	.	.
31	.	.	.	75531	2705310	.

PAPER IV AND V—ECONOMIC STATISTICS (THEORETICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *any six* questions.

(*b*) All questions carry equal marks.

1. You are required to assist a committee set up to enquire into and report about the changes in the levels of living of people as also the trend in the inequality of income distribution in India.

Survey briefly the existing material on consumption in the context of your problem. What are the limitations of your analysis?

2. What do you understand by the term 'balance of payments'? How would you examine the balance of payments position for India in the current year?

3. Stuvel defines the following price index

$$I = \frac{L_p - L_q}{2} + \sqrt{\left(\frac{L_p - L_q}{2}\right)^2 + \frac{V_1}{V_0}}$$

where L_p and L_q are the Laspeyre-type price and quantity indices, V_0 and V_1 being the values at the base period and current period respectively.

Examine whether the above index meets the time-reversal, factor-reversal and the proportionality criteria.

4. What do you understand by 'business cycles'? Describe the Harvard methods of constructing economic barometers.

5. What are the salient features of family budget data? How would you proceed to construct empirical demand relationships using such data?

Explain the terms—(*i*) Engel elasticity and (*ii*) constant elasticity hypothesis. Describe Tornqvists' three forms of demand function and show that these do not give constant elasticities.

6. The test for the Pareto distribution of income is the existence of a linear relationship, with negative slope $(-\alpha)$ between the logarithm of income and the logarithm of the number of persons having that income or more.

What is the density function for the Pareto distribution? Find the mean of the distribution. In what sense does the parameter α provide a measure of income inequality?

7. Econometricians have often been interested in investing the existence of increasing, decreasing, or constant returns to scale in technological process. Given the sample of observations on

X = output
N = labour input
K = capital input, and
M = materials input

and the technological equation

$$X = A N^{\alpha_1} K^{\alpha_2} M^{\alpha_3} u$$

u being the multiplicative error term, how would you test the constancy of returns to scale?

8. State the three definitions of national income and explain their equivalence. Distinguish between (*a*) gross national product and gross domestic product, and (*b*) gross national product and net national product.

PAPER IV AND V—STATISTICAL QUALITY CONTROL (THEORETICAL)

Time: 4 Hours | Full marks: 100

(*a*) Attempt *any four* questions.
(*b*) All questions carry equal marks.
(*c*) Figures in the margin indicate full marks for each question.

1. (*a*) Explain the terms (*a*) average run length, (*b*) maximum allowable average range and (*c*) modified control limits in respect of a control chart. Bring out their usefulness in controlling a process. (12)

   (*b*) Explain what is meant by a 'moving average and moving range' chart? When can such charts be used in preference to the conventional $\bar{X}-R$ charts? How would you interpret an out of control point on such a chart? Give a few examples from any industry known to you of the use of such charts. (13)

2. (*a*) From an industrial process k samples each of size n are drawn and the quality characteristic is measured for each of the $n \times k$ items. Write down the analysis of variance for these data. Obtain the expected values of the mean squares due to different sources. (10)

   (*b*) Explain clearly how variance component analysis can be used for (*i*) setting realistic specifications, (*ii*) process improvements, and (*iii*) process control. (15)

3. Describe clearly the salient features of the Mil Std 105 A or Mil Std 105 B tables. (15)

   How will you select a suitable single sampling plan form these tables for normal inspection? When using this plan how would you decide whether to reduce or tighten inspection or to continue normal inspection? (10)

4. Write short notes on:
   (*a*) Simon's I_0 charts (5)
   (*b*) Hamilton's standard lot plot method of acceptance sampling by variables (10)
   (*c*) Evolutionary operation (10)

5. (*a*) Describe the precautions to be taken in designing and conducting a plant scale experiment. (10)

   (*b*) Explain, with the help of one or two examples from any industry known to you, the use of nested design in industrial experimentation. How would you analyse the results of such an experiment? (15)

PAPER IV AND V—SAMPLE SURVEYS, APPLIED (THEORETICAL)

Time: 4 Hours | Full marks: 100

(*a*) Attempt *any four* questions.
(*b*) All questions carry equal marks.

1. (*a*) What are the advantages and disadvantages of sampling as compared with complete enumeration?

(*b*) Which of these two methods will you recommend in the following cases and why?

(*i*) For estimating production of opium in India,

(*ii*) For estimating per capita consumption of cereals in a particular State.

2. It is proposed to conduct continuous sample surveys for preparing quarterly indices of employment and unemployment for Bombay city. There are maps showing the census blocks (1961) of the city. Prepare suitable designs of the surveys necessary for the purpose. Also, discuss about the reference period, methods of estimation and the field-staff to be employed.

Enumerate the possible sources of error and describe the methods of controlling these errors.

3. (*a*) Describe 'Frame' and 'Sampling unit'.

(*b*) Give a short account of the various types of sampling procedure that are generally used in practice.

(*c*) A sample survey will be conducted in Calcutta for studying the income distribution of earners. Discuss the advantages and disadvantages of different sampling units that may be used for the purpose.

4. (*a*) Describe the various stages of computation work of a survey.

(*b*) Describe methods of controlling numerical errors in large scale statistical computations.

(*c*) Give some examples to show that defects in the sample may sometimes be compensated by adjustment of the results.

5. Write notes on *any three* of the following:

(*a*) Master sample

(*b*) Cost function and variance function in a double sampling procedure

(*c*) Sampling on successive occasions

(*d*) Size and density of grids to be used for crop survey.

PAPER IV AND V—SAMPLE SURVEYS, THEORY (THEORETICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *any four* questions.

(*b*) All questions carry equal marks.

(*c*) Figures in the margin indicate full marrks for each question.

1. (*a*) State briefly the relative advantages of stratified sampling as compared to unstratified sampling. (4)

(*b*) Show that stratified simple sampling with replacement which ensures equal probabilities of selection for each unit in a single draw, is more efficient than unstratified sampling with equal probability with replacement. (8)

(*c*) Suppose there are two strata, and let n_1 and n_2 be the respective number of units to be selected with equal probability with replacement from these strata to estimate the population total. If ϕ is the ratio of n_1/n_2 to the optimum value of n_1/n_2 for the same total sample size., viz., n_1+n_2, show that the relative efficiency of the actual allocation as compared to that of optimum allocation is not less than. (13)

$$\frac{4\phi}{(1+\phi)^2}.$$

2. (*a*) Let X_i denote the value of a given measure of size for the ith unit in the population ($i=1, 2, \ldots, N$). Show that the following procedure gives rise to selection of a unit with probability proportional to this measure of size: (10)

(*i*) select 2 numbers at random—one from 1 to N (say, i) and the other from 1 to M (say, R) where M is the maximum of the sizes X_i ($i=1, 2, \ldots, N$)

(*ii*) select the i-th unit if $R \leqslant X_i$

(*iii*) reject the i-th unit if $R > X_i$ and proceed afresh starting from step (*i*).

(*b*) Suppose one unit is selected with probability proportional to a given measure of size, and then without replacing this unit, a second unit is selected with probability proportional to the same measure of size. Let y_i and y_j be the respective values of these units in respect of the characteristic under consideration, and let p_i and p_j be the respective probabilities of selection of the units in the first draw. Show that the estimator,

$$\hat{Y} = \frac{1}{2-p_i-p_j}\left[\frac{y_i}{p_i}(1-p_j)+\frac{y_j}{p_j}(1-p_i)\right]$$

is unbiased for the population total Y. (7)

(*c*) Derive an unbiased estimator of the variance of the estimator given above. (8)

[Hint : $E(y^2) = V(y)+Y^2$]

3. (*a*) If y and x are unbiased estimators of the population totals Y and X respectively based on the same set of sample units drawn according to any sample design, derive the expressions correct to the second degree of approximation for bias and mean square error of the ratio estimator.

$$R = \frac{y}{x}X$$

for estimating the population total Y.

State clearly the underlying assumptions and the nature of the approximations (18)

(*b*) Under what condition(s), if any, is the above ratio estimator more efficient than the estimator $\hat{Y}$ in large samples? (7)

4. (*a*) Describe briefly the advantages and disadvantages of multistage sampling. (5)

(*b*) Suppose n first stage units are selected with probability proportional to a given measure of size with replacement and from each selected first stage unit m second stage units are selected with equal probability without replacement. Give an unbiased estimator of the population total Y and derive an unbiased estimator of the variance of your estimator. (13)

(*c*) For a given cost C, find the optimum values of n and m which would minimise the variance of the above estimator for the population total Y. The cost is given by the formula: (7)

$$\text{Cost} = C_0 + n\,C_1 + nm\,C_2$$

where C_0 is the fixed overhead cost, C_1 is the cost per first stage unit and C_2 is the cost per second stage unit.

5. (*a*) Briefly enumerate the sources of non-sampling errors in censuses and surveys. (5)

(*b*) In a mail enquiry, suppose out of n units selected with equal probability with replacement, n_1 units respond in the first attempt. For making special efforts to collect the required information, let a sample with a sampling fraction f be selected with equal probability with replacement from the n_2 $(=n-n_1)$ non-responding units in the first attempt. Suppose that at this second attempt they all respond.

(*i*) Give an unbiased estimator of the population mean $\overline{Y}$. (5)

(*ii*) Drive the variance of the estimator, assuming $\frac{n_2-1}{n_2} \simeq 1$ (8)

(*iii*) Find the optimum values of n and f which would minimise the variance for a fixed cost C, the cost function being given by (7)

$$\text{Cost} = C_1 n + C_2\,(1-p)\,n\,f$$

where C_1 is the cost per unit for the first attempt, C_2 is the cost per unit for follow-up and p is the proportion of units responding in the first attempt in the population.

6. Write short notes on the following: (25)

(*i*) Optimum stratification,
(*ii*) Multi-phase sampling
(*iii*) Sampling on successive occasions
(*iv*) Cluster sampling
(*v*) Interpenetrating (net work) of sub-samples

PAPER IV AND V—DESIGN OF EXPERIMENTS, APPLIED (THEORETICAL)

Time: 4 Hours | Full marks: 100

(*a*) Attempt *all* questions.
(*b*) All questions carry equal marks.

1. 28 varieties of soybean are to be tested in 8 blocks each of size 7, and each variety being replicated twice. The design to be employed for this purpose is the 'dual' of a balanced incomplete block design with parameters: $v=8$, $b=28$, $r=7$, $k=2$, $\lambda=1$. Prepare a layout of this design. (The 'dual' of a prescribed design is obtained by considering the varieties as blocks and blocks as varieties in the the given design). Describe how you would proceed to analyse the data and obtain the estimates of variances of the different varietal contrasts.

2. An experiment was conducted to study the responses to five different levels (10, 20, 30, 40, 50 lbs. per acre) of nitrogen (N) on 3 varieties of potatoes (V_1, V_2, V_3). The design used is a strip-arrangement (that is, a variant of the split plot arrangement) where the varieties are the whole plot treatments while the levels of nitrogen are the subplot treatments, the arrangement being in strips across each replication. The whole experiment is replicated four times. Indicate the appropriate analysis of the data. How would you test for (1) the linear as well as the quadratic effects of nitrogen, and (2) the interaction between the varieties and manures?

3. To test the effects of five different fertilizers A, B, C, D, E an experiment is to be conducted with the combinations of these fertilizers each at two levels. These combinations are to be arranged in blocks of size 8 such that one four-factor and two three-factor interactions are confounded in a single replication. The whole experiment is to be carried out in such a set of five replications so as to achieve as much balance as possible over the partially confounded effects. Prepare a layout of this design. Discuss in detail how you would proceed to analyse the data and test the different main effects and interactions.

PAPER IV AND V—VITAL STATISTICS AND POPULATION STUDIES (THEORETICAL)

Time: 4 Hours | Full marks: 100

(*a*) Attempt *any four* questions.
(*b*) Figures in the margin indicte full marks for each question.

1.(*a*) Given for each age x in the reproduction span for a particular year:

F_x = the number of females
B_x = the total number of births
fB_x = the total number of female births
fp_0 = the chance of survival from birth to age x.

Show how to calculate (*i*) female gross reproduction rate, (*ii*) female net reproduction rate, and (*iii*) overall (or crude) female fertility rate. (15)

(*b*) How would you modify your calculations of (*i*) and (*ii*) if only the following information is available by quinary age groups for the ages 15-44 years: number of females, total number of births (male plus female births), the overall sex-ratio at birth, and the chance of survival. (10)

2. (*a*) What are: (*i*) a stable population, (*ii*) a stationary population? (7)
(*b*) Develope the Lotka formula for the intrinsic rate of natural increase in a stable population. (18)

3. (*a*) What are the different types of errors that occur in population censuses? Illustrate with special reference to the Census of India 1931, 1941 and 1951. (12)
(*b*) Outline a scheme for a sample check of population census data with particular reference to India. (13)

4. (*a*) What are the deficiencies of the vital registration system in India? (10)
(*b*) Give, in brief, a scheme for obtaining birth and death rates in India through a sample survey, with particular emphasis on the procedure you would adopt to control both the sampling and the non-sampling errors. (15)

5. (*a*) Explain briefly two important methods of population projection with proforma of working sheets. (15)
(*b*) Name the important population projections made for India during the period 1951-60 and make a critical review of at least one of these projections. (10)

6. Write short notes on *any two* of the following: (25)
   (*i*) UN Model Life Tables
   (*ii*) 'Conventional' and 'True' infant mortality rates
   (*iii*) Standardized mortality indices

PAPER IV AND V—THEORIES OF INFERENCE (THEORETICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *any four* questions.
(*b*) All questions carry equal marks.

1. State and prove the Rao-Balckwell Theorem and point out its importance.

Let $X_1 < X_2 < X_3 < X_4 < X_5$ be the order statistics of a random sample of size 5 from the rectangular distribution over the interval $(0, \theta)$. Show that $2X_3$ is an unbiased estimator of θ. Determine the joint probability density function of X_3 and the sufficient statisite X_5 for θ. Find the conditional expectation $E(2X_3/X_5) = \phi(X_5)$. Compare the variances of $2X_3$ and $\phi(X_5)$.

2. By the method of inverse probability obtain the posterior probability distribution of the standard deviation σ of a normal population with unknown mean μ, assuming the prior probability distribution $P(d\mu d\sigma/\Pi) \propto \frac{d\mu d\sigma}{\sigma}$ and show that it conforms to the t-distribution.

Show that in large samples the results given by the methods of inverse probability and maximum likelihood are indistinguishable in spite of any arbitrariness in the assumption fo prior probability distribution of unknown parameters.

3. What do you mean by 'bias' in statistical tests? Define unbiased critical regions of Type A_1 and show how to obtain such regions when they exist.

Examine whether it is possible to obtain Type A_1 test for testing $H_0 : \sigma = \sigma_0$ against $H_0 : \quad 0 < \sigma < \infty$,

for the normal population with known mean μ and unknown S.D. σ

4. Explain the role of the *OC* and *ASN* functions in sequential test procedure.

Derive exact expressions for the *OC* function and the *ASN* function of a general sequential probability ratio test, proving the main results you employ.

5. Explain the concept of a 'decision function'. How will you choose between alternative decision functions?

Let a single trial be made on an event with probability p. Let x denote the number of successes $x=0$ or 1. If the loss function is

$$\frac{(p - \hat{p})^2}{p\,(1-p)}$$

where $\hat{p}$ represents an estimate of p, obtain the minimax estimator of p.

6. Show how to obtain distribution free consistent estimators of the percentage points of a distribution.

What is Wilcoxon's *T*-test? State its optimum propeties and discuss its asymptotic efficiency.

PAPER IV AND V—GENETICS (THEORETICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *any four* questions.
(*b*) All questions carry equal marks.

1. Describe how a discriminant function for selection with respect to an assigned commercial value of an individual can be computed when multiple measurements on a set of individuals and their progeny are available.

Indicate how exactly such a discriminant function is used and what are the advantages associated with it.

2. For a population segregating for two alleles A, a, examine the consequences of continued brother-sister mating, starting with given initial frequencies of the genotypes AA, Aa and aa.

3. Suppose there are two linked loci with genes (A, a) and (B, b). Write down the

the probabilities for the four phenotypic classes into which the offsprings of the mating.

$$\left(\frac{A\ b}{a\ B}\right) \times \left(\frac{A\ b}{a\ B}\right)$$

may be classified. If data for matings of the type

$$\left(\frac{A\ B}{a\ b}\right) \times \left(\frac{A\ B}{a\ b}\right) \text{ and } \left(\frac{A\ B}{a\ B}\right) \times \left(\frac{A\ b}{a\ B}\right)$$

are available, how do you test the hypothesis that linkage between the factors is same in both the situations.

4. Given the data on M, N blood groups of parents and their children how do you proceed to investigate the inheritance of this system of blood groups?

   Show how the data on M, N blood groups in a population enable us to examine the extent of inbreeding present in the population.

5. Suppose there is a factor with alleles (A, a) A being completely dominant over a. The problem is to determine whether a given individual is AA or Aa, and we 'self' and examine the progeny for recessives. How many offsprings should be examined to keep the error of classifying an Aa individual as AA, below 1 in 1000.

   Given the frequencies of the four phenotypic classes

   $$AB,\quad Ab,\quad aB,\quad ab$$

   obtained by the backcross method to study linkage between the factors A and B, explain how the total χ^2 can be partitioned into 3 individual degrees of freedom to study segregations of individual factors and linkage between factors.

6. Given the genotypic values for AA, Aa, aa as 1, 0, -2 respectively, compute the correlation of parent and offspring starting with a population

   $$p^2AA + 2pq\,Aa + q^2aa$$

   and obtaining the progeny by selfing.

PAPER VIII AND IX—ECONOMIC STATISTICS (PRACTICAL)

Time: 4 Hours — Full marks: 100

(*a*) Attempt *any two* questions.

(*b*) All questions carry equal marks.

1. Suppose the following problem is given. How would you reformulate it as a problem of linear programming?

   'Let the Indian economy be considered as consiting of the four sectors:

   S_1 : basic investment goods
   S_2 : factory consumer goods
   S_3 : household industries (including agriculture), and
   S_4 : services.

It has been determined on the basis of some studies that the income investment ratios in the above sectors are respectively 0.20, 0.35, 1.25 and 0.45. The capital requirements per employed person in the sectors are Rs. 20,000, Rs. 8,750, Rs. 2,500 and Rs. 3,750 respectively. The Planning Comission desires that the additional employment in the household industries sector and services sector should not together exceed 9 million. Further the Commission has the information that the additional employment that would have to be created during the Plan period will not exceed 11 million. Given that the total resources available for investment during the period is less than Rs. 5,000 crores, suggest an optimum investment plan for the Planning Commission.

2. Using the data of Table 2 below estimate the constants α, β of the demand function,

$$Q = \frac{\alpha E}{E+\beta}, E > 0.$$

Obtain the average Engel elasticities for the expenditure groups Rs. 0-15 and Rs. 24-34.

Monthly consumption per capita of cereals in rural India by classes of total expenditure per capita per month, 1955-56.

monthly per capita expenditure classes (Rs.)	population (percentage)	total expenditure per capita per month (E) (Rs.)	monthly per capita consumption of cereals (Q) (seers)
0— 8	15.57	6.27	12.66
8—11	18.07	9.41	14.69
11—13	12.23	11.97	17.35
13—15	9.36	13.96	18.83
15—18	11.35	16.52	24.78
18—21	7.02	19.59	28.78
21—24	7.03	22.36	22.85
24—28	5.12	25.73	23.58
28—34	4.66	30.54	24.06
34—43	4.35	37.54	25.47
43—55	2.08	46.30	31.82
55 and over	2.24	81.05	46.66

The following table gives an estimate of the size distribution of income in India during 1955-56. Draw concentration curves for the distributions of pre-tax income and post-tax income. Study these curves carefully and see whether direct taxation accentuates or reduces the income inequality.

Estimates of income by income groups, India, 1955 56.

monthly pre-tax income per capita (Rs.)	number of persons (millions)	annual per capita income (Rs.)	
		pre-tax income	post-tax income
Under 10	96.7	96.1	96.1
10—15	107.01	149.8	149.3
15—20	66.3	210.2	208.3
20—25	40.2	264.4	261.7
25—30	23.7	323.6	319.8
30—40	24.4	408.2	403.3
40—50	11.2	528.6	522.3
50—60	6.2	640.3	623.2
60—70	3.9	769.2	761.5
70—80	2.4	833.3	862.5
80—100	2.8	1060.7	1050.0
100 and over	5.1	2449.0	2237.3
all classes	390.0		

PAPER VIII AND IX—STATISTICAL QUALITY CONTROL (PRACTICAL)

Time: 4 Hours Full marks: 100

(a) Attempt *any two* qiestipns.
(b) All questions carry equal marks.
(c) Use of calculating machine is permitted.

1. Work out the single sampling plan that would minimise the average amount of inspection given

$$N = 10{,}000, \qquad AOQL = 6.0 \text{ percent}$$

$$\bar{p} \text{ (process average)} = 3.0 \text{ percent}$$

Draw the OC-curve of this plan. (50)

2. In order to study a filling operation the process was visited every hour and during each visit weights of four randomly chosen empty tins (tare weight) and four filled tins chosen at random from those filled during the preceding hour (gross weights) were taken.

The averages and ranges of these samples are given in the following table:

Gross weight			Tare weight		
sample number	$\bar{X}$	R	sample number	$\bar{X}$	R
1	595	11	1	95	4
2	590	27	2	97	7
3	595	41	3	91	5
4	595	16	4	93	14
5	590	16	5	95	9
6	584	8	6	94	10
7	588	9	7	96	3
8	591	23	8	94	6
9	600	8	9	96	3
10	597	13	10	91	9
11	595	40	11	92	11
12	596	21	12	95	7
13	592	33	13	94	9
14	588	17	14	98	5
15	593	45	15	93	3
16	592	14	16	90	14
17	601	11	17	94	7
18	606	10	18	92	4
19	610	12	19	90	7
20	605	14	20	93	4
21	613	13	21	95	9
22	607	17	22	99	7
23	612	17	23	93	10
24	606	21	24	91	11
25	600	14	25	94	13

Estimate the natural variation for the tare and gross weights.

The production in charge contends that the filling operation needs no improvement and all the troubles here would be solved if only he can get more uniform tins. State with reasons, whether you agree with him or not.

What is the existing variation in the net weight of tins? If this is desired to be halved how much reduction should there be in the average range of gross weights assuming that it is not possible to reduce the variation in the tare weights? (50)

3. The outside diameter of a base plug after knurling is known to be dependent on the value before knurling. To study its actual relationship, measurements were taken on 50 randomly chosen items (over a period of 3 days) before and after knurling, which are given below in suitable units.

sample number	before knurling	after knurling	sample number	before knurling	after knurling
1	498	588	11	478	593
2	482	586	12	493	605
3	506	600	13	498	583
4	493	586	14	520	605
5	468	565	15	503	580
6	478	560	16	406	495
7	505	590	17	510	592
8	492	572	18	472	563
9	482	583	19	543	644
10	400	488	20	522	615
21	415	496	36	495	548
22	455	502	37	535	607
23	483	580	38	502	590
24	464	558	39	468	520
25	528	578	40	523	606
26	413	495	41	456	515
27	486	566	42	465	540
28	404	480	43	450	509
29	508	594	44	518	593
30	495	592	45	453	587
31	500	570	46	450	498
32	460	525	47	508	592
33	488	570	48	475	570
34	460	478	49	523	593
35	508	600	50	472	576

The production-in-charge wishes to know: 'If the diameter after knurling is to be contained within 550 ± 50 units, within what limits should the value be held before knurling'? Analyse the data and establish the required limits. State clearly the assumptions involved in arriving at your answer. (50)

PAPER VIII AND IX—SAMPLE SURVEYS, APPLIED (PRACTICAL)

Time: 4 Hours Full marks: 100

(*a*) All questions carry equal marks.
(*b*) Use of calculating machine is permitted.

1. *Either,*

Data on small scale manufactures and handicrafts will have to be collected in a sample of establishments in a given sample of (1961 census) blocks in the urban areas of India. Prepare necessary schedules for different stages of field work and write down necessary instructions for the guidance of the investigators.

Also, prepare an estimate of field staff required for the survey (assuming that the detailed enquiry is to be conducted in 10,000 establishments spread over 1000 tons).

Or,

The object of a sample survey is to study some important aspects of production in the small industry sector. Prepare blank summary tables which in your opinion are of fundamental importance for such a study. Detail out the items of information necessary for this purpose in a suitable schedule. Also write down necessary instructions for the computers for scrutiny and compilation of data.

Also, prepare an estimate of computation-staff required for the purpose (assuming that the sample is self-weighting) if the sample size is 10,000 (establishments).

2. Enclosure I shows the population (in hundreds) of different wards of the towns in a certain area. The twons have been grouped into five strata, indicated by *A*, *B*, *C*, *D* and *E* in the enclosure.

Using wards as the sampling units, draw (with full explanatory notes) two different random samples, each of size 32, by following the two methods mentioned below:

(*a*) Simple random sample from the whole of the urban area

(*b*) Stratified simple random sample with uniform sampling fraction using the strata as indicated in the enclosure.

All the samples are to be drawn without replacement.

Prepare estimates of pupolation total on the basis of these samples.

Compute the variances of these estimates and compare the relative efficiences of these methods of sampling.

Also find out the estimate of these variances from your samples.

Enclosure I

Population of different wards of the towns in the given area.

serial number of town	ward number	population (00)
(1)	(2)	(3)
	Stratum—A	
1	I	102
	II	120
	III	33
	IV	31
	V	36
	I	20
	II	31
	III	28
	IV	26
	V	47
	VI	28
2	VII	34
	VIII	47
	IX	48
	X	46
	XI	62
	XII	10
	I	67
3	II	24
	III	31
	Stratum—B	
	I	188
1	II	238
	III	188
	I	179
	II	149
2	III	151
	IV	106
	V	53
	VI	116
	I	63
	II	122
3	III	51
	IV	121
	V	45
	VI	52
	Stratum—C	
	I	332
	II	197
1	III	402
	IV	170
	V	396
	I	227
	II	242
2	III	219
	IV	224
	V	179

serial number of town	ward number	population (00)
(1)	(2)	(3)
	Stratum—D	
	I	540
	II	187
	III	247
	IV	496
1	V	433
	VI	674
	VII	459
	VIII	637
	IX	351
	X	313
	Stratum—E	
	I	51
1	II	41
	III	24
	IV	24
	I	305
2	II	217
	III	104
	I	97
	II	62
3	III	68
	IV	70
	V	145
	VI	112
	I	61
	II	54
4	III	303
	IV	148
	I	259
	II	49
5	III	23
	IV	16
	I	17
	II	46
6	III	43
	IV	27

PAPER VIII AND IX—SAMPLE SURVEYS, THEORY (PRACTICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *all* questions.
(*b*) All questions carry equal marks.
(*c*) Use of calculating machine is permitted.

1. Suppose a survey is to be conducted to estimate the number of literates in a town. The following rough estimates are available for planning the survey:

community	total number of persons	percentage of literates
C_1	60000	40
C_2	10000	80
C_3	30000	60

(*a*) If a stratified simple random sampling design is to be adopted with communities as strata, suggest the 'best' allocation of the fixed sample size of 200 persons over the strata.

(*b*) Assuming the above figures as the actual figures, calculate the variances of the estimates for stratified and unstratified sampling and hence find the gain due to stratification.

2. In order to get a quick estimate of the yield rate of paddy in an agricultural farm, 10 paddy plots were selected with probability proportional to area with replacement from a total of 25 paddy plots with a total area of 4.00 acres and from each selected plot a circular cut of 4′ radius was located at random. (The number of cuts taken from any sample plot was exactly the same as the number of times it was selected). The crop within the selected circular cuts was harvested and weighed. With results given below, estimate the yield rate of paddy (in mds/acre) and also the standard error of the estimate.

serial number of circular cut	area in acres of plot in which cut is located	yield in tolas
1	0.21	91.7
2	0.14	156.4
3	0.15	77.9
4	0.36	65.5
5	0.30	102.8
6	0.18	84.5
7	0.21	69.2
8	0.12	74.7
9	0.30	108.6
10	0.17	88.1

PAPER VIII AND IX—DESIGN OF EXPERIMENTS, APPLIED (PRACTICAL)

Time: 4 Hours | Full marks: 100

(*a*) Attempt *all* questions.

(*b*) Figures in the margin indicate full marks for each question.

(*c*) Use of calculating machine is permitted.

1. The following table gives the plan and yields of an experiment in turmeric with three 'equidistant' levels of nitrogen (n_0, n_1, n_2) phosphorus (p_0, p_1, p_2) and potash (k_0, k_1, k_2). There are four replications with three blocks in each replicate; the blocks are shown as columns. The different treatment combinations are given in brackets. The symbol (uvw) denotes the treatment combination ($n_u\, p_v\, k_w$). Carry out an appropriate analysis of the data and test for the linear as well as quadratic effects of the fertilizers. (60)

Yield of raw turmeric in lbs. per acre.

Replication I			Replication II		
(012) 21	(201) 53	(111) 57	(121) 51	(100) 43	(010) 37
(122) 20	(121) 41	(102) 59	(220) 45	(012) 48	(120) 27
(220) 22	(210) 35	(221) 71	(102) 42	(210) 49	(200) 17
(202) 23	(011) 49	(001) 64	(201) 44	(221) 44	(021) 24
(101) 20	(112) 51	(010) 63	(000) 52	(202) 38	(211) 25
(021) 41	(100) 72	(212) 48	(022) 41	(020) 32	(222) 26
(000) 50	(002) 51	(022) 52	(212) 37	(111) 23	(002) 27
(110) 30	(222) 57	(120) 36	(011) 46	(001) 41	(112) 30
(211) 31	(020) 55	(200) 43	(110) 49	(122) 33	(101) 25

Replication III			Replication IV		
(101) 37	(020) 29	(220) 30	(112) 48	(020) 58	(222) 24
(210) 28	(211) 35	(122) 31	(121) 42	(002) 44	(201) 52
(221) 32	(121) 28	(100) 53	(010) 54	(221) 25	(012) 50
(000) 50	(110) 38	(201) 22	(022) 36	(122) 40	(210) 33
(112) 33	(222) 30	(010) 47	(100) 50	(001) 59	(000) 62
(011) 30	(001) 51	(111) 48	(211) 39	(212) 37	(111) 39
(120) 29	(012) 39	(002) 40	(202) 46	(200) 49	(120) 51
(202) 32	(102) 38	(021) 38	(001) 45	(101) 42	(102) 61
(022) 25	(200) 36	(212) 24	(220) 37	(110) 41	(021) 40

Corrected sum of squares:

Total = 15346.63

Replications = 2001.67

Blocks (within replications) = 5380.74

2. The following table gives the plan and yield of cotton for a PBIB design with parameters $v = b = 15$, $r = k = 4$, $\lambda_1 = 0$, $\lambda_2 = 1$.

Analyse the data and estimate the variances of the different varietal contrasts. (40)

Plan and yield of cotton in a PBIB Design (variety number is shown in bracket)

blocks	variety and yield			
1	(15) 24	(9) 25	(1) 26	(13) 20
2	(5) 27	(7) 28	(8) 24	(1) 27
3	(10) 26	(1) 28	(14) 24	(2) 22
4	(15) 34	(11) 31	(2) 21	(3) 28
5	(8) 41	(15) 33	(4) 33	(7) 29
6	(12) 34	(4) 32	(3) 29	(1) 30
7	(12) 32	(14) 25	(15) 24	(8) 26
8	(6) 23	(3) 23	(14) 24	(5) 27
9	(5) 28	(4) 28	(2) 26	(13) 25
10	(10) 25	(12) 27	(13) 28	(6) 26
11	(9) 26	(7) 26	(10) 23	(3) 24
12	(8) 27	(6) 27	(2) 25	(9) 26
13	(5) 30	(9) 36	(11) 32	(12) 32
14	(7) 30	(13) 28	(14) 24	(11) 25
15	(10) 24	(4) 25	(8) 32	(11) 31

Total corrected $SS = 2303.583$

Association parameters:

$\lambda_1 = 0, \quad \lambda_2 = 1, \quad n_1 = 2, \quad n_2 = 12$

$((p'_{jk})) = \begin{pmatrix} 1 & 0 \\ 0 & 12 \end{pmatrix} \qquad ((p^2_{jk})) = \begin{pmatrix} 0 & 2 \\ 2 & 9 \end{pmatrix}$

PAPER VIII AND IX—THEORIES OF INFERENCE (PRACTICAL)

Time: 4 Hours | Full marks: 100

(*a*) Attempt *all* questions.

(*b*) All question carry equal marks.

(*c*) Use of calculating machine is permitted.

1. In a chemical process it is very important that a certain solution to be used as a reactant has *p*H less than or equal to 8.30 and the solution is definitely of no use if the *p*H exceeds 8.33. A method of determining *p*H is available which for solutions of this type is known to give measurements which are approximately normal with a standard deviation of 0.02. In order to make a sample inspection of the quality of the product, manufacturer adopts a sequential probability ratio test of strength (α, β) with $H_0 = 8.30$ and $H_1 = 8.33$. Suppose that the rejection of one single consignment of the products entails a loss of 100 rupees,

and that a single determination of the *p*H factor costs 0.1 rupee. Find out approximately the average monetary cost incurred by the manufacturer due to sampling inspection and rejection of consignments when the real *p*H value is 8.20 and (α, β) have the following different values:

(0.05, 0.01) (0.10, 0.03) (0.15, 0.05).

Find out also in each of the above cases the cost involved if the producer adopts the equivalent Neyman Pearson procedure.

2. In a study which tested the equipotentiality theory, Ghiselli compared the learning (in a brightness-discrimination task) of 21 normal rats with the relearning of 8 postoperative rats with cortical lesions. That is, the number of trials to relearning required post-operatively by the 8 E rats was compared with the number of trials to learning required by the 21 C rats. The results are given in the table below. Use a suitable non-parametric test to examine whether the two groups of animals differ significantly in their rate of learning (relearning).

Trials to learning (relearning) required by *E* and *C* rats

E rats	C rats	C rats (cont.)
20	23	24
55	8	15
29	24	21
24	15	15
75	8	18
56	6	14
31	15	22
45	15	15
	21	14
	23	
	16	
	15	

PAPER VIII AND IX—MATHEMATICAL THEORY OF SAMPLING DISTRIBUTIONS (PRACTICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *all* questions.

(*b*) All questions carry equal marks.

1. The following measurements have been taken on two trivariate normal distributions

Population I			Population II		
x_1	x_2	x_3	x_1	x_2	x_3
121	74	254	132	77	249
108	80	300	123	74	315
122	87	223	129	96	319
77	66	209	131	67	310
140	71	261	110	96	268
108	57	245	125	87	217
124	52	242	129	102	300
130	89	242	130	104	270
149	91	277			
129	72	208			
154	87	244			

Calculate the value of the statistic

$$D^2 = \sum_{i,j=1}^{3} \sum^{3} w^{ij} (\bar{x}_{i1} - \bar{x}_{i2}) (\bar{x}_{j1} - \bar{x}_{j2})$$

where $[w^{ij}]$ is the matrix reciprocal to $\{w_{ij}\}$ the dispersion matrix estimated from the two samples together.

Under the hypothesis that corresponding population means of the variates are equal, find the probability of getting a value of D^2 equal to or larger than the above value.

2. Draw 20 samples of size 5 from a normal population with mean 1 and variance unity. Calculate

$$u = \frac{\bar{x} - 1}{S}$$

for each sample, $\bar{x}$ and S^2 being the mean and sample variance. Find the moments and determine what Pearsonian type curve could be fitted to the sampling distribution of u.

PAPER VIII AND IX—PSYCHOLOGY AND EDUCATION (PRACTICAL)

Time: 4 Hours Full marks: 100

(*a*) Attempt *any three* questions.

(*b*) All questions carry equal marks.

(*c*) Use of calculating machine and statistical tables is permitted.

1. The following data on a sample of 40 students taking Mathematics paper of an Engineering Admission Test is available (full marks = 100)

serial number	scores in 1st half	scores in 2nd half	total	serial number	scores in 1st half	scores in 2nd half	total
1	23	19	42	21	20	23	43
2	40	41	81	22	26	25	51
3	20	35	55	23	13	19	32
4	25	28	53	24	23	39	62
5	26	29	55	25	35	28	63
6	41	36	77	26	33	47	80
7	25	23	48	27	34	32	66
8	35	32	67	28	25	27	52
9	35	38	73	29	35	26	61
10	36	43	79	30	19	28	47
11	30	24	54	31	25	15	40
12	35	44	79	32	40	37	77
13	45	34	79	33	25	18	43
14	30	27	57	34	20	28	48
15	35	37	72	35	33	39	72
16	38	41	79	36	35	27	62
17	34	34	68	37	35	34	69
18	27	38	65	38	30	38	68
19	33	37	70	39	38	30	68
20	26	38	64	40	23	34	57

It is required to calculate the reliability coefficient for the total test by the split-half method using the Spearman-Brown Correction. Compute the standard error of measurement for this test. Estimate the reliability this test would have if it were applied to a group whose scores had a standard deviation (*a*) half that of the original group and (*b*) double that of the original group.

2. The intercorrelation matrix of marks of 7439 students in eight major school subjects of the 1957-examination is given below:

	Eng. I	Eng. II	Beng. I	Beng. II	History	Geography	Sanskrit	Mathematics
Eng. I	—	.70	.48	.49	.53	.48	.57	.56
Eng. II	.70	—	.47	.48	.49	.45	.53	.51
Beng. I	.48	.47	—	.49	.47	.36	.50	.41
Beng. II	.49	.48	.49	—	.42	.41	.52	.43
History	.53	.49	.47	.42	—	.47	.51	.45
Geography	.48	.45	.36	.41	.47	—	.49	.56
Sanskrit	.57	.53	.50	.52	.51	.49	—	.58
Mathematics	.56	.51	.41	.43	.45	.56	.58	—

Extract three centroid factors using highest correlation in each column as the estimated communality.

Rotate the factor axes orthogonally to obtain a simple structure and positive manifold and interpret the factors.

3. (*a*) following data represent the rankings of three judges (*A*, *B* and *C*) according to mechanical ability of eight persons P_1, P_2, ..., P_8:

Judges	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
A	7	4	2	6	5	3	1	8
B	4	2	1	7	6	3	5	8
C	7	2	1	6	4	5	3	8

Calculate the coefficient of concordance among the judges and test for its significance. In case of significant concordance, combine the rankings to estimate the true ranks of the persons.

(*b*) At the beginning of the school year the mean score of a group of 65 sixth-grade children upon an educational achievement test in reading was 45.00 with a standard deviation of 6.00. At the end of the school year, the mean score on an equivalent form of the same test was 50.00 with a standard deviation of 5.00. The correlation between scores made on the initial and final testing was 0.60.

Does teaching in the school improve significantly the performance of students?

4. (*a*) The following is a frequency distribution of marks scored by 4968 students of a School Final Examination in English Prose: Text and Grammar, Carrying full marks 100.

marks	frequency
00—09	84
10—19	334
20—29	825
30—39	1569
40—49	1466
50—59	591
60—69	94
70—79	5
total	4968

Compute the mean and the standard deviation of the scores obtained by

(*i*) the top 25 percent of the candidates
(*ii*) the middle 50 percent of the candidates.

(*b*) Four problems, *A*, *B*, *C* and *D* have been solved by 50 percent, 60 percent 70 percent and 80 percent respectively of a large group. Compare the difference in difficulty between *A* and *B* with the difference in difficlty between *C* and *D*.

(*c*) Consider 100 students taking an examination, on the results of which they are to be classified into five sub-groups *P*, *Q*, *R*, *S*, and *T* so that the range of ability in each sub-group is equal. On the assumption that the trait measured by the examination is normally distributed, how many students are to be placed in respective sub-groups?