INDIAN STATISTICAL INSTITUTE

QUESTION PAPERS

for

Statistician's Diploma Examinations

March-April & September 1959

Price Rupee one only

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH-APRIL 1959

PAPER I : THEORETICAL STATISTICS, GENERAL

Time : 4 Hours Full marks : 100

(a) Attempt *any five* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is not permitted.

1. Give an idea of the uses and the limitations of (i) averages, (ii) standard errors, in summarising experimental data.

An experimenter wishes to compare two methods of feeding sheep. He finds that one group of fifty randomly selected animals fed by the first method, has a mean increase in weight of 15 lbs., with standard deviation 3 lbs. while a second group of fifty fed by the second method has a mean increase in weight of 29 lbs. with a standard deviation of 4 lbs.

Explain how one can make valid comparisons between the two methods on the basis of the above data. What is your inference ?

2. In a game of Badminton between two players, the server scores a point and serves again if he wins a rally; but if he loses a rally, the score remains the same and the other player serves.

Assuming that two players A and B are equally matched and that no one is placed in an advantageous position compared to the other in regard to service, find the probability that r rallies will be played before the scores change and hence find the expectation of r.

If A is serving, what are the probabilities that

(i) A will be the first to win a point and
(ii) B will be the first to win a point.

3. (a) $u = ax+by$ and $v = ax-by$ where x and y represent deviations from the means of two measurements on the same individual. The correlation coefficient between x and y is ρ. If u and v are uncorrelated, show that

$$\sigma_u \sigma_v = (a^2+b^2)\, \sigma_x \sigma_y \sqrt{1-\rho^2}$$

(b) What statistical tools would you use to test the significance of an observed value of the correlation coefficient ? State clearly the underlying assumptions.

4. State what methods you would apply to test the significance of the difference in the following cases. Clearly state and explain the validity of assumptions which you find necessary to make about the parent populations. Discuss both the cases—(a) large and (b) small samples.

(i) The difference of proportions obtained from two independent samples,

(ii) The difference of means of two independent samples taken from populations of known equal variance,

(iii) The difference of means of two independent samples taken from populations of the same, but unknown, variance,

(iv) The difference of two sets of measurements made at different times but on the same individuals.

5. Explain how you would obtain the significance levels for t and χ^2 from a table of significance levels for the variance ratio F.

Two groups of four young male rats are chosen from two different litters. In each group of four rats, all receive the same basic diet, and two chosen at random from each group receive a daily dose of a vitamin concentrate. Show how you would carry out an analysis of variance on their weight increases over a given period for testing the difference between the two diets and explain the sources of variation used to estimate the error for test of significance.

6. (a) What is meant by an interpolation formula? Mention the assumptions involved in deriving an interpolation formula.

(b) Derive Newton's divided difference formula and hence obtain Newton's Forward difference interpolation formula.

(c) Under what situation is this formula used?

7. (a) How and under what conditions can the binomial distribution be approximated by Poisson's distribution?

(b) Show that for a Poisson variate the mean equals the variance.

(c) Show that the sum of two Poisson variates is a Poisson variate.

PART II : APPLIED STATISTICS, GENERAL

Time : 4 Hours Full marks : 100

(a) Answer *all* questions.

(b) All questions carry equal marks.

1. *Either,*

Describe suitable sampling 'frames' for the following sample surveys :—

(i) for a working class and middle class family budget enquiry in a city like Calcutta or Bombay, and

(ii) for estimating yield per acre of winter paddy in a major rice-growing state.

Give reasons for the choice of your frame in preference to the other alternative.

Or,

It is desired to conduct a sample survey to estimate the extent and nature of unemployment in Calcutta city. Describe what sample design you would recommend to obtain valid estimates at minimum cost. What items of information would you like to include and what ultimate results would you like to present?

2. *Either,*

Mention the name of any *one* of the Indian publications where statistics relating to (i) Industrial production and (ii) Foreign trade of India are available. State

also the method of collection and compilation, periodicity of publication, the authorities publishing them and broad item coverage included in those publications.

Or,

Describe the functions of the Central Statistical Organisation and the National Sample Survey Directorate of the Government of India and their functional relationship with other departments of Central and State Governments and other institutions.

3. Describe the importance of price statistics in the economy of a developing country from the point of view of plan balances, giving the several uses to which such statistics could be or ought to be put to. Discuss the availability, limitations and improvements, if any, needed in such price statistics with reference to India.

4. *Either,*

Describe King's method of constructing an abridged life table. Deduce approximate relation between m_x and q_x stating clearly the assumptions made.

Or,

Describe any suitable method of predicting future population of India and discuss its limitations. Would the method you describe be equally applicable to the States or to selected cities ? Give reasons.

5. *Either,*

What are the advantages of a 'Factorial experiment' over the 'Single factor experiment'? Explain the purpose, of 'Confounding in factorial experiments'. Illustrate your point by considering a 2^3 experiment in which the highest order interaction has been confounded.

Or,

What is meant by an 'Incomplete Block Design' in design of experiments ? What are their advantages over other designs ? Describe with a simple suitable example, the procedure you will adopt for analysing the data of a balanced incomplete block design and give the structure of the analysis of variance table.

6. *Either,*

What is genetical linkage and how does it help us in preparing autosomal maps ? Indicate broadly the statistical methods by which we can (i) detect and (ii) estimate linkage in well known types of genetical data.

Or,

Describe briefly the method which is adopted for standardising a given test. How do you propose to scale different individuals whose scores are available in a large number of tests ?

PAPER III : STATISTICAL INFERENCE

Time : 4 Hours Full marks : 100

(a) Answer Question 7 and *any four* of the rest.

(b) All questions carry equal marks.

1. What is the justification, if any, of the principle of unbiased minimum variance estimation ?

Define a sufficient statistic. Explain briefly how the principle of sufficiency can sometimes be used to obtain unbiased estimate of uniformly minimum variance, and give an example of the method.

2. What is a confidence interval? Suppose that a sample of n independent observations is drawn from a continuous but otherwise entirely unknown population. Let $X_1 < X_2 < \ldots < X_n$ denote the sample values arranged in ascending order of magnitude. Show that, for any i and j with $i < j$, (X_i, X_j) is a confidence interval for the population median. How would you choose i and j?

3. Define the term 'uniformly most powerful unbiased test of size α' and state conditions under which such a test exists.

Suppose that n independent Bernoulli trials are carried out, the probabilities of success and failure in each trial being θ and $1-\theta$ respectively, where θ is unknown. It is desired to test the hypothesis $\theta = 1/3$ against all alternatives. How would you construct a uniformly most powerful unbiased test?

4. Show that if a sufficient statistics exists, it is a function of the maximum likelihood estimate.

Verify this property for the estimation of the parameter λ of the Poisson distribution

$$e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0,1,\ldots$$

5. Give two examples of composite hypotheses. State a set of sufficient conditions on the parent distribution so that it should be possible to construct similar regions for testing a composite hypothesis with one degree of freedom. Examine if these conditions hold for testing the hypothesis $\sigma = \sigma_0$ about the standard deviation of a normal population with unknown mean μ. In this case how would you proceed to choose the best among the similar regions?

6. What is a locally most powerful and unbiased test? Find a set of criteria for the construction of such tests. What would be the locally most powerful and unbiased test for the hypothesis $\mu = \mu_0$ about a normal population $N(\mu, 1)$?

7. A doctor at a certain hospital asked 100 lung cancer patients and 100 cases of bone fracture whether they were heavy smokers, with the following result:—

	Lung Cancer	Broken Bones
Heavy smoker	85	23
Not a heavy smoker	15	65
Total	100	100

Statistician A says that these data prove that smoking is a cause of lung cancer. Statistician B disagrees, and maintains that the above scheme should be 'reversed'. One should obtain instead a group of people who are heavy smokers, and also a group of people who are not, and then compare the proportions of Lung Cancer cases in the two groups, these numbers becoming available, of course, only after each member of both groups is dead. Statistician C says that both A and B are wrong, and that neither experiment can possibly determine whether smoking is a cause of Lung Cancer.

Discuss briefly the relative merits, if any, of A, B and C.

PAPER VI : (PRACTICAL)

Time : 6 Hours | Full marks : 100

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

1. (a) Find a polynomial of the minimum degree which coincides with $\sin x$ at $x = 21°, 22°, 24°, 25°$. Using this, calculate $\sin 23°$ and compare it with the tabular value from any standard tables of trignometric functions. (10)

(b) Using Simpson's rule with $h = 0.1$, evaluate

$$\int_0^1 \frac{dx}{1+x^4} \qquad (10)$$

2. *Either,*

(i) Find out the approximate probability that a correlation coefficient of 0.70 or less can arise in a sample of size 25 from a normal population in which the true correlation coefficient is 0.80. (4)

(ii) Test the significance of the following at 5 per cent level :—

(a) A multiple correlation coefficient (= 0.48) of y on 5 variates x_1, x_2, x_3, x_4, and x_5 based on a sample of size 20. (4)

(b) A partial correlation coefficient of -0.38 between x_1 and x_2 eliminating effects of x_3, x_4, and x_5 in a sample of size 30. (4)

(iii) The following correlation coefficients between weight of ears of wheat as harvested and weight of grains of wheat after harvesting and cleaning were obtained from the results of a crop cutting experiment in a certain district which was divided into six sub-blocks.

Sub-block number	1	2	3	4	5	6
No. of samples	274	179	233	54	255	50
Correlation coefficient	0.9055	0.8710	0.9210	0.9365	0.9508	0.9780

Test whether the correlation coefficients differ among themselves. (8)

Or,

(a) Experience has shown that 40 per cent of the young of a certain breed of animals die of a certain disease during the first year of their life. The discoverer of a new vaccine claims that its use at least halves mortality and in support of his claim cites an experiment in which out of 40 young treated with vaccine, only 8 died. Examine how far his claim is justified. (10)

(b) Five thousand candidates appeared for a certain examination in which there were three subjects (a), (b), and (c). Each subject carried 100 marks. The means, standard deviations and correlations of the marks obtained in the various subjects are given below :—

	(a)	(b)	(c)
Mean	39.46	52.31	45.26
S.D.	6.20	9.40	8.70

$\rho_{ab} = 0.43$, $\rho_{bc} = 0.38$, $\rho_{ac} = 0.29$

On the hypothesis of a normal distribution of marks, estimate the number of candidates who obtain 50 per cent or more of the aggregate marks. Calculate also the variance of the estimate. (10)

3. You are given below the mean chest expansion (y) in inches of a number of school boys of ages (x) between 8 and 18 and classified into two groups according as the nutrition is 'good' or 'not good'. Obtain the regression coefficient of chest expansion on age for the two groups separately. Examine the significance of (*i*) the two regression coefficients (*ii*) the difference of the two regression coefficients. (20)

	Nutrition good		Nutrition not good	
Age in years (x)	No. of boys (f)	Mean chest expansion (y)	No. of boys (f)	Mean chest expansion (y)
8	4	2.06	16	1.61
9	11	2.02	27	1.64
10	15	2.97	47	1.84
11	10	1.30	35	2.06
12	20	2.10	60	1.84
13	13	2.20	42	2.16
14	17	2.15	74	2.49
15	14	2.66	62	2.57
16	23	2.54	74	2.52
17	9	2.02	52	2.24
18	11	2.27	37	2.46
total	147	....	526	....

4. *Either,*

(a) Ten determinations have been made, each under standard conditions and by the same method, of the electrical resistance of a certain piece of silver wire and the results, measured in ohms, were 3.09, 3.13, 3.12, 3.08, 3.13, 3.10, 3.18, 3.10, 3.05, 3.15.

If the wire were made of pure silver, it is known that the resistance would be 3.09 ohms, but the presence of any impurities in the silver would increase the resistance of the wire.

Test whether the above results are consistent with the hypothesis that the wire is made of pure silver. (10)

(b) A manufacturer sells an article at a fixed price of Re. 1/-. He guarantees to refund the purchase money to any purchaser who finds that the weight of his article is less than 8 chattacks. Actually the weight of the articles made by the manufacturer are normally distributed with standard deviation 1 chattack. His cost of production, Rs. C per article, is related to the mean weight w chattack of the articles by the relation,

$$C = 0.05w + .30.$$

Graphically, or otherwise, determine the mean weight which the manufacturer should aim at, if he wishes to maximise his expected profit per article. (10)

Or,

Eight electric lamps have been chosen at random from each of four batches of lamps, each batch having been produced with a different type of filament of wire.

The 32 lamps have been run until they have burnt out and the lengths of life are recorded to the nearest 10 hours. The results in 10-hour units were :—

Wire A	175,	180,	164,	185,	170,	168,	162,	166
Wire B	168,	172,	163,	166,	173,	160,	167,	181
Wire C	170,	172,	190,	178,	165,	168,	165,	172
Wire D	162,	178,	170,	173,	169,	175,	180,	178

Test whether there is any significant difference between the average life of the lamps made with the different types of wire. (20)

5. The following table gives the distribution of the duration of sickness among (a) employed and (b) unemployed workers. Examine whether unemployment influences the duration of sickness.

Duration of sickness (in weeks)	Frequency	
	Employed	Unemployed
(1)	(2)	(3)
0—1	530	442
1—2	218	208
2—3	47	115
3—4	34	73
4—5	25	16
5—6	22	5

(20)

PAPER VII (PRACTICAL)

Time : 6 Hours Full marks : 100

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

1. *Either,*

The following data give the number of books and journals issued by a library on different working days in November 1959, starting from Monday, November 3, 1958.

Assuming linear trend, obtain seasonal indices for the different working days (Monday, Tuesday etc.) in a week. Suggest a possible use of these indices.

Date	Total number of books issued	Date	Total number of books issued
3-11-58	188	17-11-58	197
4-11-58	159	18-11-58	228
5-11-58	246	19-11-58	275
6-11-58	160	20-11-58	282
7-11-58	213	21-11-58	236
8-11-58	178	22-11-58	258
10-11-58	207	24-11-58	220
11-11-58	258	25-11-58	366
12-11-58	249	26-11-58	327
13-11-58	199	27-11-58	271
14-11-58	192	28-11-58	230
15-11-58	171	29-11-58	272

Or,

Given below are the data relating to wholesale prices of six items. The prices for different items are quoted in Rupees and relate to week ending the dates specified for five weeks.

(i) Compute index numbers of wholesale prices for different weeks with week ending 28th December 1957, as the base.

(ii) Also compute the chain base index numbers for different weeks.

(iii) With the chain base index numbers, compute again the index numbers with fixed base for different weeks (with week ending 28th December, 1957 as the fixed base). Compare the two series of fixed base index numbers by the two different methods and comment on them.

Mention whether you have used geometric mean or arithmetic mean in the construction of the above index number and give reasons for your choice.

Items	Units	Prices in Rupees for the week ending				
		28-12-57	4-1-58	11-1-58	18-1-58	25-1-58
Rice (coarse)	md.	24.00	23.50	23.25	23.50	22.75
Wheat (C.591)	md.	15.25	15.12	15.37	15.31	14.87
Bajra	md.	12.25	12.37	12.12	12.50	11.50
Coal	ton	20.62	20.37	20.50	20.25	21.00
Electricity	kwt	5.22	5.22	5.22	5.25	5.25
Rubber	100 lbs.	152.50	152.50	152.00	150.50	151.00

2. The following data give the population of U.S.A. at each decennial census, om 1800 to 1950. Fit a logistic curve, $\frac{1}{P_t} = A + B.C^t$ where P_t denotes the popula- on at time t, to the data. Using this, estimate the population of U.S.A. in 1960.

year	population (in millions)	year	population (in millions)
1800	5.31	1880	50.16
1810	7.24	1890	62.95
1820	9.64	1900	76.00
1830	12.87	1910	91.97
1840	17.07	1920	105.71
1850	23.19	1930	122.78
1860	31.44	1940	131.67
1870	38.56	1950	150.70

Either,

The following table contains the plan and the yields of a 2^4 field experiment i beans. The yields are given in lbs. and the factors employed are as follows :—

Dung (d) : 0, 10 tons per acre
Nitrochalk (n) : 0, 0.4 cwt. N per acre
Superphosphate (p) : 0, 0.6 cwt. P_2O_5 per acre
Muriate of potash (k) : 0, 1.0 cwt. K_2O per acre

	Block I				Block II			
Replication I	p 46	k 56	d 54	npk 37	dp 50	nk 44	dk 43	pk 51
	dnk 41	dnp 48	dpk 55	n 42	dnpk 43	(1) 57	dn 40	np 49
	Block III				Block IV			
Replication II	npk 42	d 41	p 38	dnk 34	nk 44	dp 53	(1) 58	np 40
	n 48	dnp 53	k 51	dpk 45	pk 56	dk 52	dnkp 54	dn 42

(a) Identify the confounded interaction.
(b) Calculate the main effects and unconfounded interactions and test them for significance.
(c) Give a concise presentation of the main effects and two factor interactions in a tabular form and record their standard errors.

Or,

The following data came from an experiment to compare 5 treatments (gasolines A,B,C,D and E where A was the control or standard gasoline, B and C were standard plus two kinds of ingredients manufactured by company P, and D and E were standard plus two kinds of ingredients manufactured by company Q). The response observed was milenge per gallon expressed as deviations from 27. The tests were conducted on each of 5 days. Also 5 cars were used in the experiment.

cars	Days				
	1	2	3	4	5
1	6(D)	4(B)	−2(A)	0(E)	7(C)
2	2(A)	3(C)	2(E)	−1(B)	0(D)
3	7(E)	1(D)	5(C)	−3(A)	5(B)
4	8(B)	5(E)	1(D)	5(C)	−1(A)
5	1(C)	0(A)	3(B)	2(D)	4(E)

(a) Obtain the analysis of variance table to test for the hypothesis of no differential treatment effect.

(b) Test if the standard, A, differs significantly from the rest to determine whether the ingredients are useful.

(c) Does the two gasolines (B and C) manufactured by P differ among themselves ?

(d) Make an overall comparison between the gasolines (B and C) manufactured by P with those manufactured by Q, namely (D and E).

4. *Either,*

In connection with planning a sample survey for estimating the proportion of land under jute in a particular region, it was decided to study first the relation between x, the size (i.e., area) of the sample unit and the corresponding variance V_x. For this purpose a pilot survey was undertaken in which separate (and independent) random samples were taken for different sizes of the sample unit and the corresponding variances were calculated.

It is suggested that the variance function giving the relation between V_x and x may be of the form

$$V_x = \frac{a}{g_x}$$

where a and g are constants.

Estimate the constants, a and g from the data given below and examine the goodness of fit of the suggested variance function.

size of sample unit x (in acres)	sample size n	estimate of variance V_x
1.00	1756	0.1119
2.25	1377	0.0813
4.00	1161	0.0650
6.25	951	0.0577
9.00	832	0.0505
12.25	732	0.0455
16.00	618	0.0419
25.00	476	0.0398
36.00	365	0.0342

What should be the optimum value of x and n in order to build up an estimate of average proportion for a given cost of Rs. 10,000/-. It is known that total cost T is related to x and n, in the form

$$T = 1000 + 2.1n + .7x.$$

Or,

The table below presents the summary of data from a complete census of all the 340 villages in the Ghaziabad Tehsil of the Meerut District. The villages were stratified by size of their agricultural area into four strata as shown in column 2 of the table. The numbers (Ni) of the villages in the different strata ($i = 1,2,3,4$) are given in column 3.

The population values of the strata mean ($\bar{Y}Ni$) and standard deviations (σwi) for the area under wheat and standard deviations for the agricultural area (σai) are shown in the subsequent columns.

strata	size class of agricultural area (in bighas)	No. of villages (Ni)	$\bar{Y}_{Ni}$ Mean area under wheat per village	Standard deviation for area under wheat (σwi)	Standard deviation for agricultural area (σai)
(1)	(2)	(3)	(4)	(5)	(6)
1	0— 500	61	110	56	130
2	501—1500	201	279	115	265
3	1501—2500	55	560	188	278
4	2501 onwards	23	960	361	982

If a sample of 34 villages are taken and the average area under wheat per village is estimated, calculate the sampling variance of the estimated area under wheat per village, when the samples are drawn according to the following schemes :—

(a) If the villages were selected by the method of simple random sampling without stratification,

(b) If the villages were selected by the method of simple random sampling within each stratum and allocated in proportion to the number of villages in the strata and

(c) If the villages were selected by the method of simple random sampling within each stratum allocated in proportion to (i) the products $Ni\sigma wi$ (ii) the products $Ni\sigma ai$.

5. The following table gives the results of a daily inspection of galvanised sheets products in a steel mill. Examine by drawing suitable control charts whether the manufacturing process is in control and give your comments.

Date	number inspected	number of defectives
May 1	4,892	86
2	10,555	77
3	1,508	3
4	6,857	105
5	8,247	127
6	2,337	14
7	4,078	20
8	5,772	65
9	8,672	137
10	9,632	136
11	9,516	158
12	9,750	123
13	6,013	84
14	10,407	220
15	10,138	102
16	3,832	30
17	4,811	107
18	8,490	109
19	8,004	161
20	12,036	125

PAPER IV AND V: ECONOMIC STATISTICS (THEORETICAL)

Time: 4 Hours — Full Marks: 10

(a) Attempt *any five* questions.

(b) All questions carry equal marks.

1. State three methods of estimation of national income and show clearly how the lead to the same measure of national income.

Discuss briefly the method used for the estimation of income from manufacturing enterprises in the Indian Union.

2. 'There are three types of errors involved in the construction of index number —formula error, sampling error and homogeneity error.' Discuss this statement and suggest how the errors can be minimized

3. Let the distribution of income (x) be denoted by the density function $f(x; \mu, \sigma$ where log x is normally distributed with mean μ and standard deviation σ. Show that

$$E(x) = e^{\mu+\frac{1}{2}\sigma^2}$$

$$\text{Var}(x) = e^{2\mu+\sigma^2}\ (e^{\sigma^2}-1)$$

Obtain equations to represent the concentration curve of the distribution and derive an expression for the concentration ratio.

4. Construct a simple statistical model from which you can obtain the supply and demand functions for a commodity.

State the procedure you will adopt for the estimation of structural parameters and obtain expressions for your estimates. Assess in this connection the advisability of using (i) regression analysis and (ii) the method of maximum likelihood for the estimation of structural parameters.

5. What is meant by the cyclical component of a time series? Briefly discuss the various methods used for obtaining the cyclical component of a time series.

6. In fitting income-consumption curves from family-budget data, explain in detail how you propose to deal with the following problems:-

(i) Household expenditure on a particular item depends, other than on income, on the number of persons in the family.

(ii) Consumption in families of the same size and income differs because of varying age and sex composition.

7. You are required to prepare a paper on the relationship between capital invested and volume of output in selected organised industries of Indian Union for a number of years on a comparable basis.

State in detail your line of work, mentioning the sources of data.

8. Write a critical note on foreign trade statistics of India, describing their sources and main features.

State the procedure of computation of the index of trade for India.

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.

1. Is rational subgrouping necessary for all control charts?

With the help of an illustration from control charts for variables, explain how rational subgrouping is an important part of the design of control charts, on which depends very much the fineness of interpretation.

2. Explain how Simon's Charts are useful in estimating lot quality.

In view of statistical tables available in published form, indicate the methods of determining confidence intervals for (a) the process mean and (b) the process standard deviation, using (i) the sample standard deviation, (ii) the sample range and (ii) the sample mean range.

3. 'Sentence him and give the verdict later'. Examine in detail the appropriateness of this quotation to convey the purpose of lot-acceptance sampling, as also to distinguish the problem of lot-acceptance from that of lot quality estimation.

4. Explain the concept of AOQL. Comment on the statement: "The AOQL may be exceeded at times."

Show how a double sampling acceptance plan, for attributes, based on the AOQL concept, is constructed.

5. Comment on *any four* of the following statements:-

(i) The use of random number tables is unavoidable in practice in drawing a valid sample of items from a lot.

(ii) Item by item sequential inspection, as developed by Wald, is not gaining popularity in industry due to complexity of theory.

(iii) In the ideal, specification limits and control limits should be identical.

(iv) The theory of runs is not relevant to all control charts.

(v) Designing specifications, experimentation, process control, acceptance inspection, quality estimation etc. should not be regarded as independent statistical problems, but as parts of a unified quality control programme.

(vi) In using acceptance sampling plans, it is always recommended that smaller lots should be pooled to form larger lots whenever possible. The reason is to achieve better conformity with statistical theory.

PAPER IV AND V: SAMPLE SURVEY, THEORY. (THEORETICAL).

Time: 4 Hours. Full marks: 100

(a) Attempt *any four* questions

(b) All questions carry equal marks.

(c) Treatment should be made mathematical wherever possible.

1. Write a short critical note on each of the following:—

(i) Large sample distribution of the ratio $\bar{y}/\bar{x}$, where $\bar{y}$ and $\bar{x}$ are sample means in a sample of independent observation from a bivariate population.

(ii) Use of *sampling with replacement* in practice.

(iii) Observation of a fixed sample versus independent samples at different times, to study population changes over a time period.

(iv) Sampling in auto-correlated populations.

2. (i) Derive the expression for variance of a simple random sample drawn without replacement from a finite population.

(ii) Into a 'lot-container' is placed a sample of N balls which are drawn at random from a supply of Mp red and Mq white balls. Then a sample of n balls is drawn at random from the lot-container, and is placed in a 'sample container'. Of these n balls, r are found to be red.

Find the expressions for Var(r), when the N balls are put into the lot-container *with replacement*, and the sample of n is drawn from the lot-container (a) without replacement, and (b) with replacement.

3. Let $\bar{x}_n$ and $\bar{y}_n$ have the same meaning as in Question 1 (i) where n is the sample size. Show that the product moment correlation in the sampling distribution of $\bar{x}_n$ and $\bar{y}_n$ equals the population correlation.

(ii) Derive an expression for the variance of the mean of a systematic sample in any form.

(iii) Discuss situations where you expect the mean of a systematic sample to be a more precise estimate than that of a simple random sample.

4. Let a random sample of size n be drawn *with replacement* from a population of N identifiable units. Let $\bar{x}$ be the mean of the sample and $\tilde{x}$ the mean of the r distinct units in the sample (ignoring repetitions of units that happen to occur more than once in the sample).

Let $x_{(1)}, x_{(2)}, \ldots x_{(\nu)}$ be the observations corresponding to the ν distinct units, so that $\bar{x} = \frac{1}{\nu} \sum_{i=1}^{\nu} x_{(i)}$

(a) Show that both $\bar{x}$ and $\bar{\bar{x}}$ are unbiased estimates of the population mean.

(b) Show that, given $x_{(1)}, x_{(2)}, \ldots, x_{(\nu)}$, the conditional expected value of $\bar{x}$ is equal to $\bar{\bar{x}}$

(c) Hence or otherwise establish that $\bar{\bar{x}}$ has a smaller variance than $\bar{x}$.

(d) Obtain an expression for the variance of $\bar{\bar{x}}$ and state how you would estimate this variance.

5. (i) What is understood by optimum allocation in the context of stratified sampling? Explain why it is called optimum and how such allocation is arrived at.

(ii) Write a short note on critical comparison between simple random sampling and stratified random sampling with proportional and optimum allocation.

6. Find out the estimate of the variance of a mean of a sample of size two when the selection of the sample units is made with varying probabilities and the sample units are drawn without replacement from a finite population. All important steps have to be indicated.

PAPERS IV AND V: SAMPLE SURVEY, APPLIED (THEORETICAL)

Time: 4 Hours. Full marks: 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

1. Write short notes on *any four* of the following terms:—
   (i) Interpenetrating sampling
   (ii) Multistage sampling
   (iii) Cluster sampling
   (iv) Sampling with probability proportional to size
   (v) Size and shape of sample cuts in sample surveys for estimating yield of crops.
   (vi) Non-sampling error.

2. Write a note on the scope and various uses of *pilot surveys*.

3. Discuss the merits and demerits of the three main methods of data collection in sample surveys (namely, methods of interrogation, mail enquiry and direct observation). Your discussion should be with reference to such aspects of a sample survey as domain of study, item coverage, cost of survey, schedules and instructions, non-sampling errors, etc.

4. What are cost and variance functions with reference to a sample survey? What data do you require for building up these functions and how do you obtain them? Describe, with a suitable example, the method of determining the sample size which is likely to give the estimate with maximum efficiency for a given cost.

5. Explain in detail the arrangements that you would propose for supervision and control of field work, and for the scrutiny of data collected, in a socio-economic sample survey by the method of interrogation. Also, explain the merits (and demerits, if any,) of your proposals.

6. Explain the utility or otherwise of organising training for investigators before the start of a large scale sample survey requiring about 500 investigators who will be working all over India. If training is necessary, what procedure do you recommend for imparting the necessary training to the ultimate investigator and by whom? Give reasons.

7. It is proposed to organise a sample survey, to be conducted annually, for estimating the number of livestock in India.

(i) What should be the sampling unit and sampling frame?

(ii) What stratification would you suggest?

Note : (a) The data of the last livestock census for Tehsils are published and available.

(b) Separate estimates for the individual states may be required.

(iii) How should the sample be drawn?

(iv) What items of information would you collect?

(v) What reference period would you adopt?

(vi) What are the likely practical difficulties to be faced while collecting data?

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

Time : 4 Hours. Full marks : 100

(a) Attempt *any four* questions

(b) All questions carry equal marks.

(c) Use of calculating machines is not permitted.

1. (a) Explain what you understand by complete and partial confounding in factorial designs. Illustrate your answer by giving one example of each.

(b) Prepare a layout plan for a completely confounded design in two replications for a 2^6 factorial experiment (with factors A,B,C,D,E and F, each at two levels) in blocks of eight plots, confounding the interactions ABC, ADE, and BDF. Verify if any other interactions are completely or partially confounded, and if so, which? Give the structure of analysis of variance for this design and a brief outline of its analysis.

2. (a) What do you understand by a missing plot in a design? What are the practical situations in which such a contingency arises in field experiments?

(b) Discuss methods of estimating missing values in a designed experiment. Obtain the formula for the missing plot value in a randomised block design when a single plot is missing. Show that by using the missing plot value thus obtained, the treatment sum of squares is overestimated, and derive the expression for this overestimation. Also, give the expression for the standard error of the difference in means between two treatments, one of which involves the missing value.

3. Thirteen litters of four rats each are available for an experimental study of the comparative effects on growth of 13 different dietary treatments 1, 2, 3, ..., 13. Give the randomized layout of the experiment and explain in detail the method of analysis you would adopt.

4. Construct a balanced confounded $3\times3\times2$ design (with factors A, B and C respectively) in block of 6 plots, arranged in two replications, in which only the interactions AB and ABC are partially confounded. By fitting constants by the method of least squares, derive the estimates and sums of squares for these partially confounded interactions. Also, calculate the loss of information on each of the partially confounded degrees of freedom and the total loss of information.

5. Prepare a randomized layout for an experiment involving 36 varieties, using the triple lattice design. Using intra-block information only, develop the analysis of variance, the efficiency of the design compared to randomized blocks and the formulae for the variances of the differences between two varietal means.

6. Write short notes on *any three* of the following :—
   (a) Intra-and inter-group balanced designs.
   (b) Hyper-Graeco-Latin squares.
   (c) Inter-block information.
   (d) Response curves.
   (e) Fractional replication.

PAPER IV AND V : VITAL STATISTICS AND POPULATION STUDIES (THEORETICAL).

Time : 4 Hours — Full marks : 100

(a) attempt *any four* questions.
(b) all questions carry equal marks.

1. Discuss the factors which had contributed to the ageing of the populations of certain industrially advanced Western countries.

2. Discuss the demographic trends observed in India during the last 50 years in relation to the variation of social, public health and other factors.

3. Describe briefly the component method for population projection mentioning the demographic data which will be required for this purpose.

4. Write short notes on the following :—
   (a) use of hospital data for morbidity studies.
   (b) use of household morbidity data for studying communicability of diseases.

5. What are the essential characteristics of the growth of urban populations in India, particularly large towns and cities, and describe a method of studying these from census data. What additional information may be collected from sample surveys to assess the factors affecting the growth of towns and cities.

PAPER VIII AND IX : ECONOMIC STATISTICS (PRACTICAL).

Time : 4 Hours. Full marks : 100

(a) Answer *all* questions.

(b) Figures in the margin indicate full marks.

(c) Using of calculating machines is permitted.

1. The distribution of household expenditure per person per month in rural India is given below :—

monthly expenditure per person (Rs. x)*	average monthly expenditure per person (Rs.)	percentage population
0– 7	6.20	15.47
8–10	9.63	17.80
11–12	11.92	12.04
13–14	13.96	10.31
15–17	16.21	10.83
18–20	19.06	8.75
21–23	22.26	6.94
24–27	26.06	5.77
28–33	30.53	4.82
34–42	36.89	3.73
43–54	48.98	1.07
55–	89.05	1.57

*Note : In calculating the monthly per capita expenditure, only the integral part of the average is retained and the decimal number excluded.

(i) Draw the concentration curve of the distribution and comment on its main features. (20)

(ii) Let $Q(x)$ denote the proportion of persons in households with a monthly expenditure of Rs. x per person or more. Plotting $\log Q(x)$ against $\log x$ and η defined by

$$Q(x) = \int_{\eta}^{\infty} \frac{1}{\sqrt{2\pi}} e^{(-u^2/2)}\, du$$

against $\log x$, examine graphically if the distribution of household expenditure per person is of the Pareto-type or log-normal. (20)

(iii) Assuming $\log x$ to be normal, obtain the estimates of mean and variance of x for the above distribution. (20)

2. The following table supplies the parameters of constant elasticity Engel curve $\log_{10} y = \alpha + \beta \log_{10} x$, where y = expenditure on item and x = total expenditure in rupees per month per capita, for a number of items for rural India :—

Items	values of parameters α	β
Food grains	0.23	0.52
Milk and milk products	−2.47	1.00
Sugar	−2.28	1.55
Clothing	−1.90	1.66

Using the results obtained from the solutions of question 1 above, estimate the expenditures per person per month for the above items on the assumption that $\log x$ is normal. (15)

3. The estimated personal expenditure on fuel and light in the U.K. is given in the following table.

Quarter	Expenditure on fuel and light (in million pounds) 1945	1946	1947
1	78	84	92
2	62	64	70
3	56	61	63
4	71	82	85

Assuming a linear trend, calculate the seasonal index. (25)

PAPER VIII AND IX : STATISTICAL QUALITY CONTROL (PRACTICAL).

Time : 4 hours Full marks : 100

(a) Attempt *question 5* and *any three other* questions.
(b) Figures in the margin indicate full marks.
(c) Use of calculating machines is permitted.

1. A control scheme performed 100 per cent inspection on lots varying in size from about 2000 units to 5000 units, a unit being tested with respect to two *independent* characteristics A and B. A unit found defective at any time with respect to any one of the characteristics was immediately rejected. Hence, units found defective for A were not inspected for B and those found defective for B not for A.

Examine the data, given below, for control with respect to each of A and B as well as for both the characteristics and write down your inferences. (27)

Lot number	lot size	number rejected A	B
1	4823	307	256
2	2058	125	115
3	3189	184	150
4	3257	213	145
5	2038	146	102
6	2611	177	156
7	2510	142	121
8	4012	309	248

2. Values of a quality characteristic as measured recently five times on each of six machines are given below. Analyse the data to find out if any assignable causes due to age of machines were operating in the production. (27)

observation number	1950		1951		1952	
	M_1	M_2	M_3	M_4	M_5	M_6
1	4.2	5.3	2.5	3.5	4.1	5.2
2	4.1	3.8	4.2	3.7	3.0	3.6
3	3.9	5.6	3.9	3.9	3.7	4.2
4	3.8	4.2	3.8	4.2	4.1	4.2
5	3.7	4.3	3.7	3.1	4.0	4.3

Note : M_1 to M_6 are the six machine, two from each of the years 1950, 1951 and 1952.

3. A single sampling acceptance plan for attributes inspection has the following elements : sample size $n = 40$, acceptance number $c = 1$. Making the necessary calculations (without using Probability Tables) draw the approximate OC curve of this plan by plotting at least five points on it.

Use the curve to find out the producer's and consumer's risks corresponding to an AQL value = 2.0 percent and LTPD = 9.5 percent. (27)

4. Find out the elements of the sequential item by item defects per unit plan for which the stipulations are :—

$$m_1 = 1.5, \qquad m_2 = 3.5, \qquad \alpha = 5 \text{ per cent}$$
$$\beta = 10 \text{ per cent.}$$

Plot the five-point ASN curve of the plan.

Assuming that an alternate plan with smaller ASN is desired (to suit lots of smaller size), reconstruct the plan by relaxing the value of β only to the necessary extent such that the resulting ASN is approximately two thirds of the original. (27)

5. An assembly consists of a circular disc with a cylindrical shaft passing through a hole at the centre of the disc. A large number of such are to be made. The tolerance specified for the clearance between the hole and the shaft is 0.0010 cms. Assuming that the shafts and holes are made on machines with equal precision, find out what should be the tolerance specifications for the diameters of the hole and the shaft. (19)

PAPERS VIII AND IX : SAMPLE SURVEY, THEORY (PRACTICAL)

Time : 4 Hours. Full marks : 100

(a) Attempt *all* questions.
(b) Figures in the margin indicate full marks.
(c) Use of calculating machines is permitted.
(d) No text book on Sampling is allowed. Practical books may, however, be allowed.
(e) Neatness in presentation will be an important consideration in awarding marks.

1. Table (A) given below shows the number of villages and the area under wheat in each of 89 administrative circles in a subdivision in India and Table (B) shows the analysis of variance on a village basis. It is required to estimate area under wheat in the subdivision using an administrative circle as the unit of sampling. We shall assume that a sample of 20 circles is to be selected, *with replacement*. Calculate the sampling variance of the estimate of the *total area* under wheat in the sub-division for each of the following procedures of sampling and estimation :—

(a) equal probability; ratio estimate

(b) equal probability; 89 × (mean per circle in the sample),

(c) probability proportional to the number of villages in the circle; (89) × (the usual weighted average of circles in the sample).

Calculate the expected number of villages in the sample according to schemes (a) and (c).

Calculate the variance of an equivalent sample with the village as the unit of sampling and compare the relative efficiency of the various methods. [The number of villages in this sampling scheme = Expected number of villages in the sampling scheme of (a).] (60)

TABLE (A)

Number of villages and the area under wheat in the different administrative circles.

circle no. (i)	no. of villages (Mi)	area under wheat (acres) (Miyi)	circle no. (i)	no. of villages (Mi)	area under wheat (acres (Miyi)
1	6	1562	46	2	687
2	5	1003	47	3	941
3	4	1691	48	1	710
4	5	271	49	1	387
5	4	458	50	10	3516
6	2	736	51	5	2002
7	4	1224	52	9	3622
8	2	996	53	2	1400
9	5	475	54	2	1584
10	1	34	55	3	830
11	3	1027	56	8	167
12	4	1303	57	3	622
13	3	692	58	2	591
14	1	524	59	5	273
15	1	602	60	2	781
16	3	1522	61	2	1101
17	4	2087	62	2	789
18	8	2474	63	2	601
19	2	461	64	3	928
20	4	846	65	4	1141
21	3	1036	66	1	1208
22	4	948	67	5	1633
23	4	1412	68	4	902
24	3	438	69	3	1286
25	5	2111	70	5	1299
26	2	977	71	7	1947
27	3	814	72	3	741
28	1	319	73	2	574
29	2	583	74	7	2554
30	4	1150	75	4	669
31	3	670	76	1	1187
32	2	499	77	2	852
33	4	714	78	1	51
34	4	1081	79	1	1265
35	1	389	80	8	1423
36	7	2675	81	2	794
37	3	868	82	1	1604
38	2	1412	83	3	1621
39	2	445	84	2	1764
40	5	706	85	6	2668
41	2	642	86	1	1076
42	4	2050	87	1	348
43	6	2530	88	4	1224
44	1	247	89	4	1490
45	2	421			

TABLE (B)

Analysis of variance of areas under wheat in villages in the sub-division.

source of variation	degrees of freedom	sum of squares	mean square
between circles	88	10924581	124143
within circles between villages	210	9588011	45657
total population	298	20512592	68834

2. (i) 500 fish were taken by netting from a lake, marked and released. On a subsequent occasion 12,329 fish were taken up and it was found that 63 of these were marked. Estimate the total number of fish in the lake. Also calculate the (estimated) standard error of your estimate. State the assumptions under which your calculations are valid, at least approximately. (20)

(ii) The following data refer to a simple random sample of 30 households from a locality and give the number of persons in the household as also the number of persons who visited the dentist at least once during 1958.

size of the household	number visiting dentist	size of the household	number visiting dentist
5	1	5	1
6	0	4	4
3	1	4	1
3	2	3	1
2	0	3	0
3	0	4	1
3	1	3	0
3	1	3	1
4	1	1	0
4	0	2	0
3	1	4	0
2	0	3	1
7	2	4	1
4	1	2	0
3	0	4	0

Estimate the proportion of persons in the population who visited the dentist at least once in 1958, and also the standard error of this estimate. Indicate your method of estimation clearly. (20)

PAPER VIII AND IX : SAMPLE SURVEY APPLIED (PRACTICAL).

Time : 4 Hours. Full marks : 100

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

1. You are given in Table 1 list of 70 villages in a tehsil of India along with their population in 1951 and cultivated area in the same year. Making use of the population figures as preliminary information, rearrange the villages in serial order, so that in effect the villages are stratified for estimating the total cultivated area from a systematic sample.

Draw five circular systematic samples, of size 7 each, from the rearranged frame.

From each of the five samples, estimate the total cultivated area in the tehsil, using the figures for cultivated area for the selected villages as given in Table 1.

Obtain a *single combined* estimate from the five sample estimates. Also, calculate a valid estimate of *sampling error* of this combined estimate.

Suitable references to other tables used should be given.

2. Draw up a suitable schedule, with appropriate headings and spacings, for a sample survey enquiry to collect data annually for estimating livestock numbers in India, with state breakdowns. Indicate the important points, concepts and definitions which you would like to stress in the instructions to investigators.

3. A schedule proposed to be used in a survey for a study of housing conditions in India, is attached. If you are required to write a short statistical report based on this survey, what tables would you prepare for this purpose? Draw the blank specimen tables, which you suggest, with suitable tabular and columnar headings.

4. It is proposed to carry out a sample survey for estimating the yield per plant of cinchona in a plantation, by adopting a double sampling scheme, using x = the girth of the plant 6" above the ground as the auxiliary character.

Let N be the total number of plants for which x is measured and let n be the number of plants (chosen out of the N) for which measurements on yield of bark (y) are also taken. The variance of the ultimate estimate will then be

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

where σ = standard deviation of yield of bark and ρ = correlation coefficient between x and y.

It has been found from empirical studies during pilot surveys that the cost of the survey depends on N and n in the form

$$T = a + bn + cN$$

where a = Rs. 498/-; b = Rs. 20/-; c = Rs. 2.60.

It is estimated that $\sigma = 17$ units and that $\rho = 0.6553$.

Calculate the optimum values of N and n and the corresponding variance of the estimated average for a given total cost of

$$T = \text{Rs. } 10{,}000/\text{-}.$$

Show the steps of your calculations clearly and neatly.

TABLE 1

Cultivated area and population in seventy villages of a tehsil of India.

serial no. of village	population	cultivated area (acres)	serial no. of village	population	cultivated area (acres)
1	776	678	36	904	760
2	670	663	37	773	602
3	4505	1290	38	1040	532
4	1732	1170	39	760	438
5	2874	1390	40	2084	638
6	2282	1110	41	828	277
7	793	760	42	4877	1640
8	895	730	43	911	424
9	1157	950	44	1205	822
10	3201	1700	45	1139	555
11	1117	900	46	4064	347
12	1236	1169	47	1114	744
13	5210	1840	48	547	372
14	843	660	49	1178	644
15	1238	1140	50	1159	732
16	1917	1360	51	441	622
17	1800	1509	52	555	342
18	2335	1810	53	827	387
19	4396	2240	54	2869	322
20	1607	1225	55	726	636
21	2071	1250	56	633	410
22	2155	1690	57	680	427
23	7780	3200	58	587	496
24	2746	1744	59	1901	936
25	2549	2400	60	2419	1226
26	1007	680	61	1258	836
27	1567	970	62	1225	634
28	5271	1850	63	1477	978
29	650	340	64	1314	724
30	3200	2450	65	1208	422
31	2902	1760	66	728	493
32	2955	2120	67	851	396
33	1746	1220	68	786	732
34	1045	860	69	663	422
35	666	620	70	740	370

Household Schedule : Housing and Vital Statistics

Rural*
Urban

[1] identification : sample village			[1] identification : sample block
1. serial no.		8. State.....	1. State.....
2. zone-State-natural division		9. district.....	2. district.....
3. stratum		10 tehsil/taluk/thana	3. town.....
4. sub-sample			4. serial no.....
5. village (S.U.I)		11. village.....	5. zone-State-natural division.....
6. household class		12. hamlet.....	6. stratum.....
7. sample household (S.U.II)		13. house number.....	7. sub-sample.....
			8. sample block.....
			9. household class.....
			10. sample household.....

[3] classificatory characters		remarks :
1. household size		
2. ,, land possessed (acres 0.00)		
3. ,, industry occupation code		
4. ,, expenditure code		
5. per capita expenditure code		

[4] household members and their particulars

serial no.	relation to head	sex : m-1, f-2	age last birthday (yrs)	marital status	education (general)	activity status	industry occupation (code)	serial no.	relation to head	sex : m-1, f-2	age last birthday (yrs)	marital status	education (general)	activity status	industry occupation (code)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)

*delete whichever is inapplicable.

[5] housing condition

sl. no.	item	code or actual	sl. no.	item	code or actual
(1)	(2)	(3)	(1)	(2)	(3)
1.	plinth		10.	courtyard area (sq.ft.)	
2.	wall		11.	source of drinking water	
3.	roof		12.	distance of the source	
4.	period built since		13.	bath	
5.	period repaired since		14.	kitchen	
6.	number of rooms		15.	latrine	
7.	room area (sq.ft.)		16.	tenure status	
8.	covered verandah area (sq.ft.)		17.	owner	
9.	total floor area (sq.ft.)		18.	monthly rent (Rs.)	

reasons for dissatisfaction, if any, with regard to:

19. condition of the house.....

.....

20. local facilities (medical, educational, market).....

.....

[6] living condition

sl. no.	item	code or actual	sl. no.	item	code or actual
1.	lighting arrangement (electric light-1, kerosene light-2, others-3)		6.	number of persons sleeping (i)adult male (ii) adult female (iii) children-male (iv) children-female	
2.	sleeping arrangement (a) (i) winter (ii) summer				
3.	number of rooms used for sleeping		7.	ventilation (b) (i) in bed room (ii) in living room (iv) in bath room	
4.	total floor area of rooms				
5.	floor area available for sleeping				

(a) sleeping arrangements : bed room-1, living room-2, verandah-3, terrace-4, courtyard-5, lawn-6, others-7.

(b) ventilation : good-1, satisfactory-2, not satisfactory-3.

(for other codes, please turn over).

block	column	item	description	sl. no.	code nos.
(1)	(2)	(3)	(4)	(5)	(6)
3	—	3	industry occupation	1	as in industry-occupation classification
3		4	household expenditure (in Rs.) code	2	0-25 (1), 26-50 (2), 51-100 (3), 101-200 (4), 201-300 (5), 301-400 (6), 401-500 (7), 501-700 (8), 701 and above (9).
		5	household per capita expenditure (in Rs.) code		0-7 (01), 8-10 (02), 11-12 (03), 13-14 (04), 15-17 (05), 18-20 (06) 21-23 (07), 24-27 (08), 28-33 (09), 34-42 (10), 43-54 (11), 55 and above (12).
4	5		marital status	4	never married-1, married-2, widowed-3, divorced-4, separated-5.
	6		general education	5	illiterate-1, below primary-2, primary-3, middle-4, matric-5, intermediate-6, graduate and above-7.
	7		activity status (code)	6	(a) *gainfully employed*- (1) *at work*: employee-11, employer-12, own account worker-13, unpaid family enterprise worker-14. (2) *not at work but with job or enterprise*: employee-15, employer-16, own account worker-17. (b) *unemployed* (1) *with job or enterprise at a future date*: (i) seeking work for the first time as employee-21, not as employee-22, (ii) seeking work not for the first time as employee-23, not as employee-24, (iii) not seeking but available as employee-25, not as employee-26. (2) *not with job or enterprise at a future date*: (i) seeking work for the first time as employee-27, not as employee-28, (ii) seeking work not for the first time as employee-29, not as employee-30 (iii) not seeking but available as employee-31, not as employee-32. (c) *not in labour force*- with job or enterprise at a future date, not seeking and not available-41, student-42, houseworker-43, rentier, pensioner-44, beggar-45, too young-46, too old-47, permanently disabled-48, others-49.

block	column	item	description	sl. no.	code nos.
(1)	(2)	(3)	(4)	(5)	(6)
4	8		industry occupation	7	as in industry occupation classification.
5	—	1	plinth	8	mud-1, timber, wood, bamboo, reed-2, brick, cement, stone-3, others-4.
5	—	2	wall	9	mud, bamboo, reed-1, timber, wood, corrugated iron sheets-2, brick, cement, concrete, stone-3, others-4.
5	—	3	roof	10	straw, grass thatched, bamboo or reed and mud, country tiles (khola)-1, corrugated sheets, asbestos, tiles-2, cement, concrete-3, other pucca (brick and mortar) -4, others-5
5	—	4	period built since	11	upto 6 months-1, more than 6 months and upto 1 year-2, more than 1 year and upto 2 years-3, more than 2 years and upgo 6 yrs.-4, more than 6 yrs, and upto 10 yrs.-5, more than 10 years and upto 15 yrs.-6, more than 15 years and upto 26 yrs-7, over 26 years-8.
5	—	5	period repaired since	12	as under serial number 11.
5	—	11	source of drinking water	13	tanks and ponds-1, wells-2, tube-wells-3, rivers, lakes and springs-4 municipal tap water-5, other sources-6.
5	—	12	distance of the source	14	within the premises-1, upto 50 yds.—2, 51-100 yds.-3, 101-400 yds.—4, over 400 yds.-5.
5	—	13	bath	15	(a) no bathroom-1, (b) bathroom: (i) in individual use-2, (ii) in common use with other households-3.
5	—	14	kitchen	16	kitchen in individual use-1., kitchen in common use with other households-2, part of living room used as kitchen-3, covered verandah used as kitchen-4, no specific part of house used as kitchen-5, individuals forming a mess-6.
5	—	15	latrine	17	(a) no built up latrine-1, (b) built up latrine-1 (i) in individual use-2, (ii) in common use with other households-3.

block	column	item	description	sl. no.	code nos.
(1)	(2)	(3)	(4)	(5)	(6)
5	—	16	tenure status	18	owner of both land and building-1 owner on payment by instal. ment-2, owner of building in leased in land-3, tenant-4, occu. pier of free quarters in lieu of services-5, others-6.
5	—	17	owner	19	public sector : government and other public bodies-1 private sector: industrial and commer. cial organisation having- (i) less than 20 workers in non-power operated concerns or less than 10 workers in power-operated concerns-2, (ii) 20 or more workers in non. power operated concerns or 10 or more workers in power-operated concerns-3, private landlords-4, others-5.

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

Time : 4 Hours. Full marks : 100

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

1. An experiment was conducted on 5 bromo grass strains (*a, b, c, d, e,*) both alone and with alfalfa. The two main plots (with and without alfalfa) were laid out in a randomized complete block design in 4 replicates. The five bromo grass strains were randomly assigned to sub-plots within each main plot. The following yields (dry weight in grams) were obtained :—

Strain of Bromo	Rep. I		Rep. II		Rep. III		Rep. IV	
	with alfalfa	without alfalfa	with alfalfa	without alfalfa	with alfalfa	without alfalfa	with alfalfa	without alfalfa
a	730	786	1004	838	871	1033	844	867
b	601	1038	978	1111	1059	1380	1053	1229
c	840	1047	1099	1393	938	1208	1170	1433
d	844	993	990	970	965	1308	1111	1311
e	768	883	1029	1130	909	1247	1124	1289

Complete the analysis and write an interpretation of the results.

2. The annexed table shows the plan and yields in lbs. per plot (of size 57′ × 18′ of a 2^3 factorial experiment on paddy carried out in a single replication consisting of

4 blocks of 8 plots each, the treatments being all combinations of two seed rates (1, p), two types of seed beds (1, q), two kinds of manuring of seed bed (1, r) two kinds of transplanting (1, s), and two kinds of field manuring (1, t) :

Block I		Block II	
qt 89.2	qr 68.0	pqrs 51.8	p 77.8
rst 78.8	pqrst 74.0	pqst 57.2	rs 50.0
pqs 68.0	pt 87.0	st 72.3	q 64.3
pr 77.4	s 69.0	prt 74.3	qrt 68.0
qrst 84.8	prst 82.5	qrs 68.0	pst 65.8
ps 78.2	r 79.8	rt 68.2	prs 56.4
qs 82.0	t 87.1	pq 77.3	(1) 76.0
pqt 95.2	pqr 84.3	pqrt 79.2	qst 73.0
Block III		Block IV	

Identify the effects confounded, analyse the data and state your conclusions.

3. The yields of grain in grams per plot from 4 replicates of a varietal trial on barley with 16 varieties, using a simple lattice design, are given below :-

Rep. I

(1) 2323	(2) 2470	(3) 2223	(4) 2394
(5) 2959	(6) 3560	(7) 3248	(8) 2687
(9) 3184	(10) 3021	(11) 2694	(12) 2791
(13) 2583	(14) 2819	(15) 2452	(16) 2737

Rep. II

(1) 2579	(2) 2767	(3) 2329	(4) 1966
(5) 2462	(6) 2507	(7) 2522	(8) 2421
(9) 2967	(10) 2634	(11) 3178	(12) 2950
(13) 2168	(14) 2709	(15) 2610	(16) 2491

Rep. III

(1) 1203	(5) 1530	(9) 2080	(13) 2445
(2) 1545	(6) 2884	(10) 2395	(14) 3223
(3) 1540	(7) 2310	(11) 3527	(15) 3783
(4) 1668	(8) 2324	(12) 1471	(16) 2771

Rep. IV

(1) 2000	(5) 2360	(9) 2410	(13) 2474
(2) 1951	(6) 2138	(10) 2644	(14) 1694
(3) 1407	(7) 2238	(11) 1903	(15) 2138
(4) 1781	(8) 1808	(12) 2658	(16) 2109

Carry out the analysis of the data with recovery of inter-block information. Find the adjusted varietal means and the standard errors of the difference of two varietal means. Calculate the efficiency of the design in comparison to randomized blocks.

4. Six varieties of wheat were simultaneously tested for quantity of yield at three representative research stations in a wheat tract, the design adopted at each station being five randomized blocks of six plots each. The yields in lbs. per plot (of 1/20th of an acre) obtained from these experiments are as under :—

Stations	Blocks	Varieties 1	2	3	4	5	6
A	I	31	33	20	38	31	31
	II	33	31	35	45	26	2
	III	31	31	30	44	25	31
	IV	30	41	10	28	31	23
	V	34	18	28	23	15	20
B	I	33	23	24	24	13	8
	II	29	29	27	24	23	23
	III	28	21	38	24	16	18
	IV	37	34	38	33	28	30
	V	48	46	45	45	50	25
C	I	30	23	34	25	20	13
	II	39	22	28	25	28	32
	III	56	43	43	31	49	17
	IV	38	45	36	35	32	20
	V	44	51	23	58	40	30

Analyse the data and interpret the results of analysis.

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

Time : 4 Hours. Full marks : 100

(a) Attempt *any three* questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. The following table gives the corrected sums of squares and products for two variables x_1 and x_2 for five experiments each having 10 subjects.

Experiment	Σx_1^2	Σx_2^2	$\Sigma x_1 x_2$
1	168.9	2020.5	458.5
2	192.1	715.6	102.6
3	132.1	2869.6	160.4
4	158.1	1964.9	333.7
5	191.6	5722.1	680.6

Examine the following :—

(a) whether the variances of x_1 in the different experiments are significantly different.

(b) whether the regression coefficients of x_3 on x_1 are significantly different in the five experiments using the method of analysis of variance.

2. Given that the sample variance-covariance matrix for four variables x_1, x_2, x_3 and x_4 is as follows:—

$$S = \begin{pmatrix} 95.3 & 52.9 & 69.7 & 46.1 \\ 52.9 & 54.4 & 51.3 & 35.1 \\ 69.7 & 51.3 & 100.3 & 56.5 \\ 46.1 & 35.1 & 56.5 & 45.0 \end{pmatrix}$$

and $\bar{x}_1 = 185.7$, $\bar{x}_2 = 151.1$, $\bar{x}_3 = 183.8$ $\bar{x}_4 = 149.2$,

(a) Estimate the parameters of the *conditional distribution of* (x_3, x_4) given (x_1, x_2).

(b) Calculate the partial correlation coefficient $r_{34 \cdot 12}$

(c) Calculate the multiple correlation coefficient between x_3 and (x_1, x_2).

3. Draw two random samples of six observations each, from each of two bivariate Normal populations with the following parameters:—

Population 1	Population 2
$\mu_1 = 15$ cms.	$\mu_1 = 14$ cms.
$\mu_2 = 30$ cms.	$\mu_2 = 28$ cms.
$\sigma_1 = 3$ cms.	$\sigma_1 = 3$ cms..
$\sigma_2 = 4$ cms.	$\sigma_2 = 5.5$ cms.
$\rho = .5$	$\rho = .55$

Test whether the difference between the two samples you obtain is significant or not. Indicate in detail the method of drawing the samples.

4. In an experiment with a fertiliser, the amount of fertiliser used, and the weights of grain and straw produced, from each of eight experimental plots of equal size, are noted below (units unspecified):—

weight of grain	40	17	9	15	6	12	5	9
weight of straw	53	19	10	29	13	27	19	30
amount of fertilizer	24	11	5	12	7	14	11	18

Test the effects of fertiliser on grain yield and straw yield for significance.

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

Time : 4 Hours. Full marks : 100

(a) Attempt *all* questions.

(b) All questions carry equal marks.

(c) Use of Calculating machines is permitted.

1. The all-India crude death rate estimated by the 1951 census actuary was 27/1000. In columns 2-5 of the table below are given the population by age-sex groups and the corresponding specific death rates of England and Wales for 1956. In columns 6 and 7 are given the population of India by age-sex groups as returned in the 1951 census. Compute the standardised death rate for India (1951) using the age-sex distribution of England and Wales as the standard.

age-group in years	population of England and Wales (1956) in thousands		age specific death rates per thousand (1956) in England and Wales		population of India in thousands (1951)	
	males	females	males	females	males	females
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0— 4	1687	1604	6.40	4.98	230409	237066
5— 9	1883	1790	0.46	0.33	231626	223504
10—14	1633	1559	0.41	0.27	208993	195610
15—19	1366	1389	0.79	0.36	184531	174229
20—24	1366	1389	1.07	0.53	162693	157670
25—34	3089	3100	1.21	0.87	281798	265824
35—44	3119	3216	2.49	1.90	219211	195129
45—54	3129	3263	7.45	4.52	152474	136048
55—64	2200	2703	22.09	11.22	89109	85107
65 and over	2045	3128	82.88	60.51	62933	64933

2. In a certain community marriage is universal for girls and widow remarriage is strictly prohibited by social customs. Assuming that all girls and boys marry precisely at ages 15 years and 25 years respectively, compute the net reproduction rate from the data presented below :—

age in years	survival number l_x		no. of female live births per annum per 1000 married females	age in years	survival number l_x		no. of female live births per annum per 1000 married females
	males	females			males	females	
0	10000	10000	—	38	4723	4748	70
15	..	6395	20	39	4641	4633	63
16	..	6353	27	40	4560	4519	56
17	..	6316	48	41	4480	4406	48
18	..	6281	68	42	4399	4296	39
19	..	6247	87	43	4319	4186	30
20	..	6212	100	44	4238	4077	24
21	..	6174	111	45	4153	3966	18
22	..	6133	122	46	4064	3854	14
23	..	6088	124	47	3970	3740	10
24	..	6039	123	48	3872	3626	7
25	5741	5984	121	49	3769	3511	4
26	5671	5922	119	50	3659	3396	—
27	5600	5855	116	51	3543	3280	—
28	5528	5780	112	52	3421	—	
29	5454	5700	108	53	3293	—	
30	5378	5612	105	54	3162	—	
31	5299	5518	102	55	3028	—	
32	5219	5418	98	56	2893	—	
33	5137	5313	93	57	2757	—	
34	5054	5204	89	58	2620	—	
35	4971	5092	84	59	2483	—	
36	4888	4979	80	60	2345	—	
37	4805	4864	75	61	2208	—	

Comment on the result.

3. To estimate the death rate of a certain state a sample of 20 villages, together comprising a population of 18051 persons, was selected. Information on the number of deaths which occurred within the selected villages during the year January 1st 1956 to December 31st 1956 was elicited by two investigators *A* and *B*. The investigation was designed so that every household could be visited twice, once by each investigator, with a time lag of about 2 weeks. A cross-check of the data collected by *A* and *B* revealed the following:—

	Reported by		
	A only	B only	A as well as B
deaths at ages below 5 years	41	48	28
deaths at ages 5 years or above	43	59	70
deaths at all ages	84	107	98

Estimate the number of deaths which could have been missed by both A and B stating the assumptions involved in your procedure. If you could know some of the important reasons for omissions on the part of A and B, state how you would proceed to obtain a more valid estimate of the number of deaths missed by both.

4. The following table gives the population of the Indian Union returned in the successive decennial censuses during the half century 1891-1941. If the growth of the Indian population during this period and in the few decades after 1941 can be assumed to be determined by the logistic function, estimate the population of the Indian Union in 1951. If the actual population returned in the 1951 census was 3569 lakhs, interpret the error in your estimate in the light of the demographic changes which occurred during this period.

TABLE

year	population (in lakhs)
1891	2359
1901	2355
1911	2490
1921	2481
1931	2755
1941	3128

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1959

PAPER I : THEORETICAL STATISTICS, GENERAL

Time : 4 Hours. Full marks : 100

(a) Attempt *any five* questions.

(b) Use of calculating machines is not permitted.

1. (a) In n mutually independent trials where n is even, there is a probability of success of θ_1 in each of the first $\frac{n}{2}$ trials and there is a probability of success of θ_2 in each of the remaining trials.

Prove that the mean and the variance of the total number of successes are respectively $\frac{1}{2}n(\theta_1+\theta_2)$ and $\frac{1}{2}n(\theta_1+\theta_2)-\frac{1}{4}n(\theta_1^2+\theta_2^2)$. Hence show that unless $\theta_1=\theta_2$ the variance of the number of successes is less than it would be in a Binomial distribution with the same number of trials and the same number of successes.

(b) A population of insects is subjected to successive sprayings of an insecticide. The probability that an insect gets an attack in one spraying is 'p'. An insect will die if it gets r attacks. Find the probability that an insect will die in the n-th spraying.

2. (a) Discuss the importance of the Normal distribution in Statistics.

(b) The calorific value of a stream of gas varies in a normal frequency distribution having mean of 150 units and a standard deviation of 30 units. If several streams are mixed, the instantaneous calorific value is the average of the calorific values of the separate streams which may be assumed to fluctuate independently, all with the same mean and the same standard deviation.

How many streams should be mixed to ensure that the calorific value is below 110 units only for 0.5 per cent of the times?

3. (a) Prove the Bienayme Tchebycheff inequality that

$$P(|x-\xi|<\sigma t)>1-\frac{1}{t^2}$$

where ξ is the expectation of x, σ^2 is the variance of x and t is any number greater than 1.

(b) In the case where x is the number of successes in 192 independent trials in each of which the probability of success is 1/4, consider the probability that x will be in the range 36 to 60. Compare the value of this probability obtained from the normal approximation with the limit given by a Tchebycheff inequality.

4. (a) Under the usual assumptions of linear regression theory, which should be stated, obtain the least squares estimate of the intercept of the regression line of y on x on the y axis. Obtain the variance of your estimate.

(b) Explain without proof how this variance can be estimated from the data and hence give a procedure for testing whether the regression line passes through the origin.

(c) If tests of significance are applied to the slopes of the regression lines of y on x and x on y, does it necessarily follow that both the tests will show the same degree of significance ?

5. (a) Two samples of sizes n_1 and n_2 respectively are drawn randomly and independently from the same infinite normal population and the mean and the variance of each sample are calculated. It may be assumed that the sample mean and the sample variance are independent.

Show that the sum of the two sample variances is independent of their ratio.

(b) Obtain the constant for the χ^2 probability law with n degrees of freedom given by

$$P(\chi^2) = \text{constant.}\ (\chi^2)^{\frac{n-2}{2}} \exp\left(-\frac{\chi^2}{2}\right)$$

(c) Indicate a few of the important uses of χ^2 distribution.

6. (a) Derive Simpson's rule for numerical quadrature, giving details of derivation.

(b) Describe a method of inverse interpolation suitable to obtain the upper 5 per cent point of the normal distribution with zero mean and unit standard deviation and state the maximum decimal accuracy which can be extracted from this table.

7. The means, standard deviations and sample sizes obtained from a study on a particular characteristic on three groups of people were found to be :—

group	mean	s.d.	sample size
1	4.6	8.9	500
2	5.6	10.3	1000
3	7.2	12.1	2000

On the basis of the above data, can you conclude that the three groups differ significantly from each other with regard to the means of the characteristic under study?

PAPER II : APPLIED STATISTICS, GENERAL.

Time : 4 Hours Full marks : 10

(a) Attempt *any five* questions.

(b) All questions carry equal marks.

1. Suppose that the State Trading Corporation in your State is entrusted with the responsibility of supplying the demand for foodgrains by retail dealers in the towns and cities of the State and you are required to estimate the volume of such demands in different months of the year. Give an outline of a stratified multistage sample design that you may like to use for the purpose. Indicate the basis of stratification and broadly the procedure for allocating a given total number of sampling units to the various strata and to the different stages.

2. What does an index of industrial production measure ? Give an outline of the method of construction of the Index of Industrial production in India. Taking into

considerations the changing composition of industries in India, what modifications in the existing method do you suggest for a more accurate index ?

3. Explain clearly the concepts of gross national product and net national product. Describe the method you will adopt for calculating net national product of agriculture and of cottage industries in India for the year 1957-58.

4. What is net reproduction rate ? How does it measure the growth or decline of a population ? Suppose that you are required to calculate the net reproduction rate of India. What type of data will you need for the purpose in addition to the existing information ? Give an outline of the schedule that you will use for collection of the required data.

5. Explain clearly the implications of the three fundamental principles of experimentation, taking Randomized block and Latin square designs as illustrations.

6. It is desired to give vocational guidance at the end of primary education and also at the end of secondary education in India. What method will you use for the purpose ? Give an outline of the nature of tests you will use and the method of deriving conclusions from the test results.

7. Explain systematic sampling, its advantages and disadvantages. The milk yield of a cow giving milk for a total of N days is recorded on some days, both morning and evening at successive intervals of K days, commencing on a randomly selected day in the first interval. Two estimates of the total lactation yield can be obtained :(i) by multiplying with N the average yield per recorded day and (ii) by multiplying with K the total of the recorded yields. Show that the former estimate is biased excepting when N is a multiple of K and obtain the variance of the latter estimate.

8. What are index numbers ? Describe the criteria which an ideal index number should satisfy. Describe in detail the official index of wholesale prices in India, commenting on its limitations.

9. Define a stationary population and describe its features. In a population, hitherto stationary, the birth rate begins to increase at the rate of two per cent per year. Determine the population at the end of the third year of this increased birth rate.

10. Factor pairs $A-a$ and $B-b$ are suspected to be linked. What are the types of crosses which would provide information on the linkage, if any, between the factors? Discuss, assuming that (i) A and B are dominant to a and b, (ii) A is dominant to a but $B-b$ shows no dominance.

PAPER III : STATISTICAL INFERENCE.

Time : 4 Hours. Full marks : 100

(a) Attempt *any five* questions

(b) All questions carry equal marks.

(c) Use of calculating machines is not permitted.

1. Define the term "consistent and asymptotically normal estimate" in the theory of estimation from large samples.

Suppose that $X_1, X_2, \ldots X_n$ is a sample of independent observations from a population with mean μ and variance σ^2. Construct consistent and asymptotically normal estimates of (i) μ, (ii) μ^2 and (iii) $\mu + \sigma$ and give their asymptotic variances.

2. Prove the following properties of unbiased estimates :—

(a) If the variance of an unbiased estimate of a parameter tends to zero as the sample size tends to infinity, this estimate is consistent.

(b) If T and T' are both unbiased minimum variance estimates of a parameter, then

(i) the correlation between T and T' is unity;

(ii) hence, $P(T \neq T') = 0$, i.e. T and T' are equal at almost all points of the sample space.

(c) If T is an unbiased estimate and T_0 a sufficient statistic for the parameter, then there exists a function $f(T_0)$ of T_0 which is also an unbiased estimate of the parameter, and

$$V[f(T_0)] \leqslant V(T).$$

3. A sample $x_1, \ldots, x_n$ of independent observations has been obtained from a population whose frequency function $f(x; a)$ obeys the regularity conditions necessary for the discussion below. Consider the likelihood function $\phi(x_1, \ldots, x_n; a)$ and the random variable $\delta \log f(x; a)/\delta a$; also define $A = V(\delta \log f/\delta a)$.

(a) Do you think that the function

$$t = \frac{\delta \log \phi/\delta a}{V(nA)}$$

is likely to be distributed normally for large values of n ? Why?

(b) What will be the mean and variance of t ?

(c) How can the above results be used to obtain large-sample confidence regions for a ?

(d) If x is a Poisson variate $f(x; a) = a^x . e^{-a}/x!$

and a sample $x_1, \ldots, x_n$ is available, find a confidence interval for a using the above results.

4. (a) How are shortest confidence intervals for a parameter defined by Neyman?

(b) How are these intervals related to best critical regions for testing hypotheses regarding the parameter ?

(c) Illustrate the above by considering the problem of finding a confidence interval for the mean of a normal population whose variance is not known.

5. The distribution of probability of a sample $x = (x_1, \ldots, x_n)$ is denoted by $P(\sigma)$ and its frequency function by $(x; \sigma)$ where σ is a parameter. In order to test the simple hypothesis $H_0(\sigma = \sigma_0)$ against a simple alternative $H_1(\sigma = \sigma_1)$, a "test function" $\phi(x)$ is defined in the following way :— The sample space X is divided in some manner into three disjoint and exhaustive regions R_1, R_2 and R_3, and $\phi(x)$, which stands for the probability of rejecting H_0 when the sample point is x, is defined as :—

$$\begin{aligned} \phi(x) &= 1 \text{ if } x \in R_1 \\ &= \alpha \text{ if } x \in R_2, \qquad 0 \leqslant \alpha \leqslant 1 \\ &= 0 \text{ if } x \in R_3 \end{aligned}$$

The above is under the restriction that

$$P(\text{rejecting } H_0 \text{ when it is true}) = \int_x \phi(x) dP(\sigma_0) = a,$$

where a is a fixed constant. The "power" of the test function $\phi(x)$ is $\int_x \phi(x) dP(\sigma_1)$.

A particular test function $\phi_0(x)$ is now defined as follows :—

$$\begin{aligned} \phi_0(x) &= 1 \text{ if } f(x; \sigma_0)/f(x; \sigma_1) < c, \\ &= a \text{ if } \quad ,, \quad = c, \\ &= 0 \text{ if } \quad ,, \quad > c, \end{aligned}$$

under the above restriction.

(a) Show that $\phi_0(x)$ is the most powerful test function.

(b) Illustrate the above by constructing a most powerful test for the hypothesis $H_0(m = m_0)$ against an alternative $H_1(m = m_1)$ where m is the parameter of a Poisson distribution.

(c) Define a uniformly most powerful test and verify if the test in (b) is such a test.

6. State and prove the lemma of Neyman and Pearson in the theory of testing hypotheses.

Suppose that it is desired to test the hypothesis that a given coin is unbiased, i.e. $P(\text{"heads"}) = P(\text{"tails"}) = \frac{1}{2}$ in a single throw. Suppose that the results of 7 independent tosses of the coin are to be made available to you. Construct a suitable critical region, such that the size of the region is .05 (up to two decimal places).

7. It is required to make the following tests :—

(i) The means of all the characters of a p-variate normal population are equal to one another. A sample of size n is available.

(ii) The corresponding means of two p-variate normal populations are equal, assuming that the two populations have the same dispersion matrix. A sample of size n_1 from one population and a sample of size n_2 from the other are available.

(a) Describe how best linear functions of the characters are obtained in the two cases for the purpose of the tests.

(b) Describe in short and precise terms how the problem of classification is dealt with in the situation of (ii).

PAPER VI : PRACTICAL.

Time : 6 Hours. Full marks : 100

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

1. (a)

Age	$1+\log_{10} p_x$
50	0.99441
55	0.99203
60	0.98836
65	0.98288
70	0.97407
75	0.95929
80	0.93465

Given the above values of $1+\log p_x$ (base of logarithm = 10), where p_x denotes the probability that a person of exact age x years will survive till he attains the age $x+1$, determine $1+\log p_x$ for all integral values of x between $x = 60$ and $x = 70$.

Hence compute the probability that a person who is exactly 62 years will survive till he attains the age 65 years.

(b) Approximate the function $f(x) = \frac{4}{2^x+x}$ in the interval $2 < x < 6$, by a polynomial of degree 5, using least squares, and compare the values given by this approximating polynomial with the true values of the function.

2. The following table gives the distribution of life (t) in hours, of 337 electronic tubes of a certain manufacturing company. Assuming the distribution of 't' to follow the exponential law

$$\frac{1}{\theta} e^{-t/\theta} (\theta > 0,\ t \geq 0)$$ obtain an estimate of the

parameter θ from the given data. Compare the observed and expected frequencies. Find the number of electronic tubes likely to burn away within 80 hours of their life and also the average life of these tubes.

..... Distribution of life in hours of 337 electronic tubes

life (in hours) t	observed frequency
0– 50	100
50–100	68
100–150	48
150–200	31
200–300	42
300–400	21
400–600	27
total	337

3. (a) An industrial establishment uses 2564 lights, the number of hours a light is used per day varying according to season as follows :

January-February	8 hours
March-April	7 hours
May-June	6 hours
July-August	6 hours
September-October	7 hours
November-December	8 hours

The average life of a bulb is 725 hours with a standard deviation of 67 hours. The manager wants to purchase a larger number of bulbs which will suffice them for two years. How many bulbs should he stock so as not to run a risk of more than 1 in 100 of running short of bulbs. Calculate how far this risk will be diminished by stocking 100 extra bulbs.

(b) To test two promising lines of hybrid corn under normal conditions of farming, a seed company selected 10 farms at random and planted both the lines in experimental plots on each farm. The yields for the different farms are given below :—

farm	1	2	3	4	5	6	7	8	9	10
line A	19	27	24	23	25	20	21	23	23	25
line B	11	18	14	14	12	15	14	3	12	17

Calculate 95 per cent confidence interval for the difference between the mean yields assuming that the yields of the two lines in a farm are (i) independently normally distributed and (ii) jointly normally distributed.

4. Ten candidates for a job were interviewed by a panel of three judges A, B, and C who independently ranked the candidates in order of suitability for the job. The ranks awarded are shown below. Examine to what extent the judges agree amongst themselves.

serial no. of candidate	ranks as awarded by judge		
	A	B	C
1	7	6	8
2	10	9	7
3	1	2	1
4	8	7	6
5	6	8	9
6	5	3	5
7	2	1	4
8	4	4	2
9	3	5	3
10	9	10	10

5. (a) A manufacturing process is intended to produce electrical fuses with no more than 1 per cent defectives. It is checked every hour by trying 10 fuses selected at random from hour's production. If one or more of the ten fail, the process is halted and carefully examined. If in fact its probability of producing a defective fuse is 0.01, what is the probability that the process will needlessly be examined in a given instance? How many fuses (instead of 10) should be tested if the manufacturer desires that the probability be about 0.05 that the process will be examined while it is producing 10 per cent defectives.

(b) Two independent samples have 28 and 19 pairs of observations with correlation coefficients r_1 and r_2 respectively. If $r_2 = 0.75$ within what limits must r_1 lie in order that the two correlation coefficients may not be considered significantly different from each other at the .05 level of significance.

6. Four varieties of wheat were grown at each of thirteen locations and the yields in bushels per acre are given below. To test the hypothesis that all variety means are equal, the location variety means are considered as a randomized block classification given by the mathematical model

$$Z_{ij} = a + L_i + V_j + e_{ij}$$
$$j = 1, 2, 3, 4$$
$$i = 1, 2, \ldots 13$$

where Z_{ij} is the yield of variety j at location i, a is the general mean, L_i is the effect of of the i-th location (assumed random) V_j is the effect of the j-th variety and e_{ij} are the error terms which are assumed to be normally distributed with means zero, variances σ_j^2, and such that the correlation between e_{ij} and e_{km} is zero if $i \neq k$.

Carry out the conventional analysis of variance for testing whether all variety means are equal disregarding the correlated errors. Obtain estimates of σ_1^2, σ_2^2, σ_3^2, σ_4^2 and test the homogeneity of these variances.

TABLE

location	Variety			
	1	2	3	4
1	43.60	24.05	19.47	19.41
2	40.40	21.76	16.61	23.84
3	18.08	14.19	16.69	16.08
4	19.57	18.61	17.78	18.29
5	45.20	29.33	20.19	30.08
6	25.87	25.60	23.31	27.04
7	55.20	38.77	21.15	39.95
8	55.32	34.19	18.56	25.12
9	19.79	21.65	23.31	22.45
10	46.24	31.52	22.48	29.28
11	14.88	15.68	19.79	22.56
12	7.52	4.60	20.53	22.08
13	41.17	32.50	29.25	43.95

7. In a barrage pond meant for supplying water to a power plant the inlet temperature of water (T_1) entering into the plant and the corresponding atmospheric temperature (T_2) were recorded for 30 days during the month of November, 1955 and 1956. In 1956 the hot water discharged from the plant was recirculated in the pond, whereas this practice was not followed in 1955. It is expected that the inlet temperature depends linearly on the atmospheric temperature in both the years. From the data given below do you notice any significant change in the linear relation owing to recirculation of hot water in 1956?

		Year			
		1955		1956	
		T_1	T_2	T_1	T_2
November	1	68	81	75	81
	2	66	83	75	83
	3	66	82	68	82
	4	66	83	78	83
	5	66	80	79	80
	6	67	81	77	81
	7	66	81	77	81
	8	66	80	77	80
	9	66	81	77	81
	10	67	82	77	82
	11	67	83	78	80
	12	68	84	78	80
	13	68	82	77	81
	14	68	84	76	80
	15	62	84	76	81
	16	62	80	74	78
	17	65	81	74	77
	18	65	80	74	79
	19	65	81	74	78
	20	64	80	72	77
	21	64	74	70	78
	22	64	74	71	78
	23	64	74	70	75
	24	63	76	73	77
	25	64	77	72	79
	26	64	78	70	77
	27	64	79	72	78
	28	67	80	71	77
	29	66	78	70	76
	30	65	76	70	77

PAPER VII (PRACTICAL)

Time : 6 hours . Full marks : 100

(a) Attempt *any five* questions

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. Given below are base period production (Monthly average production in 1951) and productions for the month of January 1959. Construct for January 1959 the index

of productions separately for different groups as also the combined index for all the groups after making due adjustments for number of days in the month.

sl. no.	name of items	weight	monthly average production in 1951	production in January 1959
1.	*Mining and Quarrying*	7.16		
	1.1 coal	6.69	2,859 (th. tons)	4086 (th. tons)
	1.2 iron ore	0.47	305 (th. tons)	605 (th. tons)
2.	*Food manufacturing industries except beverage industries*	11.85		
	2.1 sugar	4.27	92.9 (th. tons)	44.6 (th. tons)
	2.2 wheat flour	0.20	40.5 (th. tons)	74.6 (th. tons)
	2.3 tea	5.04	52.4 (mill. lbs)	10.3 (mill. lbs)
	2.4 salt	0.82	6,198 (th. tons)	1,779 (th. tons)
	2.5 vegetable oil products	0.62	14,360 (tons)	27,792 (tons)
3.	*BASIC METAL INDUSTRIES*	8.04		
	3.1 Iron and steel	5.02		
	(i) pig iron and ferro alloys	1.69	152.1 (th. tons)	180.7 (th. tons)
	(ii) finished steel	4.23	89.7 (th. tons)	120.5 (th. tons)
	3.2 Non ferrous metal	1.20		
	(i) aluminium	0.19	320.7 (tons)	840.7 (tons)
	(ii) copper	0.34	966.2 (tons)	837.0 (tons)
	(iii) brass	0.67	936.9 (tons)	900.0 (tons)
	3.3 Gold	0.02	18,864 (ozs)	10,425 (oz)
4.	*Manufacture of textile*	48.01		
	4.1 cotton textile	36.10		
	(i) yarn	12.03	108655 (th. lbs.)	150548 (th. lbs.)
	(ii) cloth	24.07	339682 (th. yds.)	430092 (th yds.)
	4.2 jute textiles	11.91	72.9 (th. tons)	92.1 (th. tons)
5.	*Manufacturing of transport*	2.92		
	5.1 automobiles	2.69	1856 (Nos.)	2,092 (Nos.)
	5.2 bicycles	0.23	9523 (Nos.)	75233 (Nos.)
6.	*Electric light and power*	2.16	4482 (lakh kwh)	11,536(lakh kwh)

2. The table below shows the numbers of inhabitants in each of the 197 United States cities which had populations over 50,000 in 1940. Calculate the standard error of the estimated total number of inhabitants in all 197 cities for the following methods of sampling. (i) a simple random sample of size 50, (ii) a sample which includes the 5 largest cities and is a simple random sample of size 45 from the remaining 192 cities. (iii) a simple which includes the 9 largest cities and is a simple random sample of size 41 from the remaining cities.

Frequency distribution of city sizes

size class (1000's)	f	size class (1000's)	f
50—100	105	800— 850	1
100—150	36	800— 850	2
150—200	13	900— 950	0
200—250	6	950—1000	0
250—300	7	1000—1050	0
300—350	8	..	..
350—400	4	1500—1550	1
400—450	1	..	..
450—500	3	1600—1650	1
500—550	0	..	..
550—600	2	1900—1950	1
600—650	1	..	..
650—700	2	3350—3400	1
700—750	0	..	..
750—800	1	7450—7500	1

Gaps in the intervals are indicated by .....

3. The following table gives for the 28 trainees of a teaching institution, their ranks (in order of merit) at the entrance examination and also their ranks in a statistics test which was held after three months of intensive training in Statistical Methods.

roll no. of the trainee	rank in entrance examination	rank in statistics test
1	13	7
2	27	19
3	20	22
4	23	12
5	10	1
6	14	17
7	17	21
8	26	6
9	5	3
10	3	8
11	12	24
12	24	16
13	9	18
14	28	26
15	1	2
16	18	27
17	6	4
18	8	15
19	21	25
20	19	9
21	22	28
22	25	14
23	15	10
24	7	20
25	16	23
26	2	13
27	11	11
28	4	5

(a) Examine (applying suitable statistical tests if necessary) whether entrance examination is effective at all in selecting potentially good students of statistics.

(d) Comment on the adequacy of such data in deciding about the validity of the entrance examination.

4. Table 2a shows, against each age-group, the percentage of ever-married women in that age group in a certain population.

Table 2b gives the average number of female children born, on completion of the effective reproductive periods, to women of this population married at various ages.

Compute the gross reproductive rate for this population.

TABLE 2a

age-group	ever-married
15—19	4
20—24	43
25—29	73
30—34	80
35—39	81
40—44	81.8
45—49	82.1

TABLE 2b

age at marriage (years)	average number of female children born (completed fertility)
15—19	1.8085
20—24	1.2067
25—29	0.9330
30—34	0.6335
35—39	0.3002
40—44	0.0692
45—49	0.0022

5. It is known that, in a steam-turbine type (electricity) generating plant the relationship between the quantity (H) of heat consumed and the (electricity) output (U) of the generator may be described fairly accurately by a linear law, over a limited range of output, provided the plant is otherwise working in a controlled condition. This relation (the regression of H on U)

$$H = a + bU$$

is often referred to as the William's line. 'a' is called the fixed heat consumption of a plant.

(a) From the following data compute the values of a and b.

Monthly heat consumption and output of a generating station (over 20 months)

unit supplied (U) (millions)	heat consumed (H) (100 millions BTU)	unit supplied (U) (millions)	heat consumed (H) (100 millions BTU)
3.17	596	13.09	1907
3.39	553	11.63	17.40
8.77	1428	13.47	1965
7.70	1248	11.58	1721
5.66	941	10.95	1612
11.10	1671	14.52	2127
11.52	1754	4.06	596
13.60	2016	14.34	2114
12.49	1823	13.32	1950
11.88	1795	15.85	2336

(b) Compute, using the usual formula, the residual variance in H after accounting for U.

(c) Draw, on a graph paper, the Willian's line and the two control lines, at a distance of $\pm 3X$ (residual standard deviation of H).

(d) Examine whether the process is in control and whether it is maintained through the following months for which the figures are given below.

U	H
13.3	1949
14.7	2130

6. The distribution of household expenditure per month in rural and urban India as estimated by the NSS 4th round, April–September 1952 is given below :

household expenditure per month in Rs. (x)	proportion of household (p) rural	urban
Rs.0 and above	1.0000	1.000
Rs. 50 and above	0.7085	0.7878
Rs. 100 and above	0.3702	0.4622
Rs. 150 and above	0.1932	0.2776
Rs. 300 and above	0.0453	0.0896
Rs. 500 and above	0.0115	0.0314
Rs. 1000 and above	—	0.0054

(a) If $\log x$ is normally distributed with mean μ and s.d. σ, what would be the proportion of households spending more than Rs. a per month.(Give an algebraic expression).

(b) Define $t(p)$ by the equation

$$p = \int_{t}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right) du.$$

Compute, referring to tables of the standard normal distribution the values of $t(p)$ for all the proportion entries in the above table.

(c) Plot the values of $Y = t(p)$ so obtained against the corresponding values of $\log x = X$. If $\log x$ is normally distributed the relation between Y and X should be linear. Why?

(d) Examine graphically (no test is required) if the distribution of $\log x$ can be taken to be normal, separately for rural and urban India.

(e) Estimate, from these graphs, the values of μ and σ (the mean and standard deviation of $\log x$).

(f) Using the estimates of μ and σ, derived in (e) estimate the mean and standard deviation of x (both rural and urban estimates are required).

7. A specified dimension of a machined component is 0.05 in. Five components are measured every half hour, and the following table gives the results for 20 consecutive samples, as deviations from the specified values in units of 0.0001 in. Make control charts for the mean and the range, marking on them appropriate limits. Discuss what evidence the chart gives of unstable production and recalculate limits for a part of the series for which the range appears to be in control.

Sample No.

1	2	3	4	5	6	7	8	9	10
+10	−11	−5	+7	−4	−1	+8	+11	+3	+8
+17	−1	−14	+4	+1	+3	+9	+4	+6	0
−1	+10	+3	+5	+6	+12	+2	+8	−4	+2
−1	+10	+14	+4	+2	−5	+3	+13	−3	−10
−15	+9	−5	+1	+7	+2	−4	+11	−8	−3

Sample No.

11	12	13	14	15	16	17	18	19	20
−1	−4	+1	−3	−5	+4	0	−4	−2	−8
−8	+7	+4	−4	+4	+12	−3	−15	−2	−2
−1	−1	+7	+11	+2	−6	−5	−4	−14	−4
0	−1	+5	+3	+13	+5	+8	−6	−5	−2
−4	−7	+1	0	+1	+5	+13	−4	+1	0

PAPER IV AND V: ECONOMIC STATISTICS (THEORETICAL)

Time: 4 Hours | Full marks: 100

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

1. Discuss improvements in agricultural statistics that have taken place since 1947 with particular reference to their importance in improving the estimates of national income.

2. What are social accounts? Indicate a system of social accounting for the Indian economy suitable for assessing priorities in investment for maximising rate of increase of national income.

3. Describe changes carried out in the method of construction of Economic Adviser's Index Number of Wholesale Prices and the uses to which the index can be put. Discuss how far the changes in the method of construction are justified with reference to the uses of the index.

4. What are the data available for projection of demand for foodgrains in India at the end of the Third and Fourth Five Year Plans? Describe a method of such projection using the data available.

5. Describe the main features of Prof. Mahalanobis' four-sector model for economic planning in India.

6. Define what is meant by a stationary time-series. Discuss the various types of models in terms of which economic stationary time-series are usually analysed. Show how to discriminate between the various types of series.

7. What data are available on rural credit in India? Show how the rural credit requirements of a district can be estimated.

8. Discuss what are meant by identification and structural estimation in relation to economic models.

PAPER IV AND V: STATISTICAL QUALITY CONTROL (THEORETICAL)

Time: 4 Hours | Full marks: 100

Attempt *any four* questions.

1. Discuss the importance of the normal distribution in the theory of Statistical Quality Control.

Examine the justification for using in control charts the three-sigma limits without regard to the actual probability distribution of the quality characteristic plotted.

2. Explain the reasons for differences in the formulae and the procedures for setting up control charts for the arithmetic mean ($\bar{x}$) in the two situations (i) standards given (i.e. parameters known) and (ii) standards not given.

How far is it correct to say that in the latter case, the analysis is approximate since estimates are used in place of parameters ?

3. A given product may be either measured or gauged with respect to an item quality characteristic having a single upper specification limit U. What are the considerations for choosing between an attributes plan and a variables plan for accepting lots of the product ? (Assume the two plans will have the same CC).

State a double sampling acceptance procedure for inspection by variables corresponding to a one-sided specification for item quality, and indicate how the elements in the plan are derived to satisfy stipulated values of AQL, LTPD, and producer and consumer risks.

4. Discuss the problem of setting specification limits. Explain with illustrations how statistical methods help in setting appropriate limits.

5. Comment on the following statement : 'The OC of an acceptance sampling plan, which is a function of incoming quality does not provide an adequate picture of quality protection afforded by the plan, unless the probability distribution of incoming quality is also taken into account.'

6. Write notes on any *four* of the following :—
   1. Critical, major and minor defects.
   2. Inspection levels
   3. Homogeneity of lots
   4. Screening
   5. go, no-go gauges and dial gauges
   6. Specification limits, control limits and tolerance limits.

PAPER IV AND V : SAMPLE SURVEY, THEORY

Time : 4 Hours Full marks : 1000

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

1. From a population of N units each of size M elements, a simple random sample of n units is chosen without replacement. From each chosen unit, a simple random sample of m elements are chosen without replacement.

(a) Derive an expression for the variance of the sample mean in terms of the variance between units and variance between elements within units. State the assumptions involved, if any.

(b) If the cost (T) of survey is dependent on n and m in the form

$$T = a + c_1 n + c_2 mn$$

where a, c_1, c_2 are certain constants, obtain an expression for the optimum values of m and n using the variance function derived above.

2. (a) How will you obtain an unbiased estimate of the area under a given crop from a sample of n fields selected with probability proportional to area of the fields and

placement ? Also, how will you obtain an unbiased estimate of the sampling variance of the above estimate ?

(b) Derive an expression for an unbiased estimate of the mean and its actual variance in a sample of 2 units selected with varying probability but without replacement from a finite population.

3. A population contains L strata of equal size. If V_{ran} denotes the variance of the mean of a simple random sample of size n and V_{st} the corresponding variance for stratified random sampling with proportional allocation, show that approximately

$$nV_{ran} = S_b^2 + \frac{L}{L}\sum_{h=1}\frac{(\bar{y}_h - \bar{y})^2}{L}$$

$$nV_{st} = S_b^2$$

where S_b^2 is the average variance within strata, $\bar{y}$ is the population mean $\bar{y}_h$ the mean of the h-th strata.

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

5. (a) Study the relative precision of systematic, stratified and simple random sampling in the case where the population consists solely of a linear trend.

(b) If the variate $y_i (i = 1, 2, \ldots, N)$ are drawn at random from a super-population in which

$$\varepsilon y_i = \mu,$$

$$\varepsilon(y_i - \mu)(y_j - \mu) = 0 (i \neq j),$$

$$\varepsilon(y_i - \mu)^2 = \sigma_i^2$$

then $\varepsilon V_{sy} = \varepsilon V_{rand}$.

[Here ε denotes average over all finite populations which can be drawn from the super-population. V_{sy} and V_{rand} are respectively the variances of systematic and random sample means].

6. Write critical notes on :—

(i) ratio versus regression estimate.

(ii) interpenetrating subsamples.

(iii) partial replacement in sampling on successive occasions.

PAPER IV AND V : SAMPLE SURVEY, APPLIED (THEORETICAL)

Time : 4 Hours Full marks : 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is not permitted.

(c) Attempt *all questions* from group A and *any two* from group B.

GROUP A

1. Describe precisely the most economical operational rule of drawing a simple random sample of n units from a population of N units (*i*) without replacement, and (*ii*) with replacement. (4)

2. (a) What are the reasons which call for stratification in sample surveys ? (3)

(b) What type of information is available and how you will use it for stratification in a sample survey for estimating crop acreages in your State ? What other information do you require ? (2)

3. (a) Describe *precisely* the procedure of drawing a systematic sample of n units from a population of N units, where N need not necessarily be a multiple of n. (3)

(b) How will you estimate the mean of a character in the population from the sample ? Is this estimate unbiassed ? Give reasons. (4)

(c) How will you combine the estimates from three systematic samples of size n into a single final estimate which is unbiased ? (2)

(d) How will you estimate the error of the combined estimate of the mean mentioned in (c) ? (3)

(e) Is it possible to get stratification like effect from a systematic sample ? Given suitable data, explain how you will utilise these to obtain maximum advantage from such stratification like effect. Illustrate with reference to a live problem. (2)

4. (a) What is meant by regression estimates in sample surveys ? In what types of situations can this be used with advantage ? How will you build up the estimates in each case ? (6)

(b) Is the estimate biased or unbiassed (either conditionally or unconditionally) in each case ? Explain (5)

5. (a) What is a sampling frame ?

(b) What are the different types of sampling frames which are likely to be useful in sample surveys?

(c) Is any scrutiny of the sampling frame necessary ? If so, for what type of defects ?

(d) Discuss the possible effects of using such defective frames in sample surveys on the estimates.

(e) What precautions will you take to eliminate or minimise such defects ? (15)

6. What are the points in favour of and against retaining the same sample units (households) in three successive surveys in a region for estimating the birth, death and growth rates and extent of unemployment ? Explain with reference to both technical nd operations aspects of the question. (15)

GROUP B

7. You are asked to carry out a sample survey for estimating the number of unemployed persons and under employed persons in India classified by age and sex.

(i) Suggest a suitable definition for the terms 'unemployed' and 'under-employed' persons.

(ii) What stratification do you suggest ? Give reasons.

(iii) What is the ultimate sampling unit that you suggest ?

(iv) What sampling design will you recommend ? Give reasons. (18)

8. What is meant by non-response in sample surveys? What are the effects of non-response on the estimates ? What are the various ways of dealing with non-response in sample surveys with a view to reducing their bad effects, if any? (18)

9. Illustrate by means of suitable examples the different types of checks which you can make to asses the quality of data from a sample survey. Also, what checks can you employ at the processing stage ? (18)

10. What topics should be included in a report on sample surveys and in what way will these topics be useful to the general readers and to sampling technicians? Give an outline of the headings of sections which you would include in a report on a sample survey of employment and unemployment. (18)

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

Time : 4 Hours Full marks : 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is not permitted.

1. 'The Latin square provides more opportunity than randomized blocks for skilful planning.' Discuss, bringing out carefully the advantages of the Latin square design in comparison with the randomized block design.

Give in detail the procedure of preparing a layout plan for testing the effect of seven treatments A, B, C, D, E, F and G in a 7×7 Latin square. Give the method of analysis and obtain an expression of the efficiency of the design as compared with the randomized block design in seven replications.

2. What is the use of transformations in the analyses of data ?

Describe, with illustrations, the situations in which the square root, the angular and logarithmic transformations are useful in the analysis of experimental data .

3. Seven treatments have been tried in Latin square designs in 10 different areas within a region with a view to determining the best treatment for the whole region. Explain in detail how you would analyse the series of experiments, justifying the various steps involved as far as practicable.

4. Explain the importance of the 'analysis of covariance' technique.

An experiment is arranged in randomized blocks for four matrices to test the effect on milk yield of cows. From the control period yield (x) and the experimental period yield (y), we get the following table of analysis of variance and covariance :—

	D.F.	$S(x^2)$	$S(xy)$	$S(y^2)$	b_{yx}	$b_{yx}S(xy)$
Replication	3	163.50	92.30	52.69		
Treatment	3	54.00	24.00	26.19	0.44444	10.67
Error	9	73.50	49.25	32.56	0.67007	33.00
Treatment + Error	12	127.50	73.25	78.75	0.57451	42.08

(i) Is it worthwhile to adjust the treatment means for regression ?

(ii) Calculate the gain in precision due to adjustment.

(iii) Test for the significance of treatment effects.

[The usual statistical tables will be supplied, if necessary].

5. Show that a 3^4 experiment can be arranged in four replications, each of 9 blocks of 9 plots, such that only the three-factor interactions are confounded and a complete balance is achieved over them. Explain the method of analysis and give the structure of analysis of variance of the design.

6. Write short notes on *any three* of the followings :—

(a) Uniformity trials.

(b) Rotation experiments.

(c) Inter and intra-block information.

(d) Orthogonal sets of Latin squares.

(e) Cross-over designs.

PAPER IV AND V : MATHEMATICAL THEORY OF SAMPLING DISTRIBUTION (THEORETICAL)

Time : 4 Hours Full marks : 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

1. Let $y_1, y_2, \ldots, y_n$ be n independent observations on Normally distributed variables $Y_1, Y_2, \ldots, Y_n$ such that

$$\text{Var}(Y_i) = \sigma^2$$

$$E(Y_i) = a_{i1}t_1 + a_{i2}t_2 + \ldots + a_{im}t_m$$

for $i = 1, 2, \ldots, n$, where $t_1, t_2, \ldots, t_m$ are unknown parameters, and $\| a_{ij} \|$ is a known matrix of rank $r, r \leqslant m < n$.

Let $\sum_1^m l_j t_j$ be a parametric function which admits an unbiased linear estimator.

Let H_0 denote the hypothesis that

$$\sum_{i}^{m} l_j t_j = 0$$

and define

$$R_u^2 = \min_{l_1, l_2, \ldots, l_m} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{m} a_{ij} t_j)^2$$

and

$$R_{H_0}^2 = \min_{l_1, l_2, \ldots, l_m} \left\{ \sum_{i=1}^{n} (y_i - \sum_{j=1}^{m} a_{ij} t_j)^2, \right\}$$

subject to

$$\sum_{j} l_j t_j = 0$$

obtain the distribution of

$$\frac{R_{H_0}^2 - R_u^2}{R_{H_0}^2}$$

2. (a) Derive the distribution of $\sum_{i=1}^{n} x_i^2$

where the x_i are independent Normal variables, the mean and variance of x_i being μ_i and 1.

(b) Show that the percentage points of the central chi-square distribution are increasing functions of the degrees of freedom.

3. Consider a sample of independent observations from a multivariate Normal populations.

(a) Without actually deriving their distributions, show that the sample mean vector and the sample covariance matrix are independently distributed.

4. (a) Show that if a random variable X has a continuous distribution function $F(x)$, then $Y = F(X)$ has the uniform distribution over (0,1).

(b) Let X_1, X_2, X_3 and X_4 be a sample of independent observations from a population having a continuous distributions functions F. Let Z denote the proportion of the population for which the variate values range from the smallest to the largest sample value. Derive the sampling distribution of Z. What is $P(Z < \frac{1}{2})$?

5. The following are 12 independent observations on a variate, in the order in which they were obtained :

20, 23, 18, 17, 24, 16, 17, 21, 22, 26, 15, 16.

By replacing each observation less than the sample median (19) by the letter a, and observations greater than 19 by b, we have the following sequence of letters :

$$b \; b \quad a \; a \quad b \quad a \; a \; b \; b \; b \quad a \; a.$$

This sequence exhibits 6 runs, 3 of a and 3 of b. If the above sequence of numerical

values is permuted at random, obtain the joint distribution of R_a and R_b, where R_a and R_b denote the total number of runs of a and b respectively. (In the given case, $R_a = R_b = 3$).

Discuss an application of this distribution in tests of significance.

6. (a) Discuss in detail how the log transformation

$$z = \log \frac{1+r}{1-r}$$

helps to obtain better tests for the correlation coefficient.

(b) Obtain the sampling distribution of a partial correlation coefficient, and also of a multiple correlation coefficient, in a sample from a multivariate Normal population.

PAPER IV AND V: VITAL STATISTICS AND POPULATION STUDIES (THEORETICAL).

Time: 4 Hours Full Marks: 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

1. Discuss briefly the trends in mortality and fertility in Asian countries since World War II and their implications on the future distribution and growth of population of the world.

2. What is meant by 'migration differential'? Describe the major findings of the more important studies on migration differentials with particular reference to internal migration.

3. Explain the concept of 'stable' population. Narrate some of the uses of this concept in demographic analyses.

4. Describe briefly how life tables may be constructed when mortality statistics are inaccurate. On what data will you then, have to fall back upon?

5. Write short notes on any three of the followings:—

(i) Demographic transition.

(ii) Model life-tables.

(iii) Inaccuracies in age returns.

(iv) Assessment of morbidity.

PAPER IV AND V: PSYCHOLOGY AND EDUCATION (THEORETICAL)

Time: 4 Hours Full marks: 100

(1) All answers should be written on this question paper itself in the space provided.

Directions. There are two parts in this examination. In Part I, there are 25 questions for which answers should be written in the space given below each question.

Necessarily, the answers to these questions will have to be short. The marks for these 25 questions are distributed as follows:

Questions 1—20 : each carries 2 marks.
Questions 21—25 : each carries 4 marks.

A total of 60 marks is, therefore, allotted for Part I. You have to answer all questions, there is no choice.

In Part II, there are two sections A and B. In each section, there are two questions. You have to answer *One* question from each section You will, therefore, answer two questions in all for Part II. Each question carries 20 marks. The total marks allotted for Part II is 40.

[N.B. : space for Answers are not shown in this printed copy]

PART I

1. Write the fundamental equation of test theory, relating observed score to 'true' score.

2. Show that the 'true' variance is almost always smaller than the obtained variance of a set of scores.

3. Show how $\sigma_e = \sigma_t \sqrt{1-r_{tt}} = \sigma_t \sqrt{1-r_{to}^2}$ where σ_e = standard error of measurement, σ_t = standard deviation of observed scores, r_{tt} = reliability, r_{to} = correlation between 'true' and observed scores.

4. Show how the variance of observed scores for a test of doubled length is given by the equation : $\sigma_{2t}^2 = 2\sigma_t^2(1+r_{tt})$, where σ_{2t}^2 = variance of observed scores for a test of doubled length and r = reliability.

5. Write an equation showing the components of an observed score according to multiplefactor theory.

6. Explain correlation between tests in terms of factor-loadings.

7. 'Unless test and criterion have at least one factor in common, the test will have zero validity for predicting that criterion'—Why ?

8. What are the statistical criteria of 'parallel' tests?

9. What is the split-half method of estimating the reliability of a test ?

10. Show how the following formula yields an estimate of reliability :

$$r_{tt} = \left(\frac{n}{n-1}\right)\left(\frac{\sigma_t^2 - \Sigma pq}{\sigma_t^2}\right) \text{ where}$$

n = number of items in a test
σ_t^2 = variance of observed scores
p = proportion of correct answer per item
$q = 1-p$.

11. A test is used for selection of students. The test scores of the selected students are correlated to the extent of .5 with examination marks at the end of a year. Is there any reason to believe that .5 is an underestimate of the true correlation between selection test and marks ?

12. Suggest a correction for chance success in multiple-choice tests and state your assumption.

13. What are the statistics used in attitude-scale construction by Thurstone ?

14. What is the difference between a factor matrix and a correlation matrix ?

15. What information is given by the item analysis of a test ?

16. How does speed affect item difficulty ?

17. How is it that the more difficult a multiple-choice test for a population, the greater are the chances of its having a lower correlation with anything else ?

18. What is the relation between the length of a test and its regression weight in a multiple regression equation?

19. Why should the weights derived from a multiple regression analysis on one sample be used with caution for a battery of tests for use on another sample from the same population?

20. Suppose that ten different tests have been developed for the prediction of a criterion. A multiple regression analysis shows that an equation can be written with only 5 out of the 10 tests, and that adding more tests to the 5 does not appreciably improve prediction. How can this happen, if all the tests are valid ?

21. From a battery of tests we have to find a weighted composite score for use in selection of personnel. How should this be done ?

22. When there are several tests and several criteria, how can the validity of a composite measure against a composite criterion be determined ?

23. What are the essential features of Guttman's method of scale analysis ?

24. What is the problem in estimating 'communality' in factor analysis ?

25. How can rotation help in the interpretation of factors ?

PART II

Section A

26. Plan an investigation to find the effect of age, sex, education and wealth on intelligence. Give your answer with reference to problem, method, design of study and method of analysis of data.

Or,

27. A vocational guidance clinic is being set up. What kinds of statistical information will be necessary for the clinic to function efficiently ?

Section B

28. Suppose you want to find whether teachers' ratings on personality traits of school boys can predict academic success. Discuss the problems involved with reference to your design of study and analysis of data.

Or,

29. A school wants to know whether its examination system is all right or it needs improvement. What kinds of information will be needed for decision ? Discuss with reference to problems that are likely to arise in the collection and analysis of data relevant for the purpose.

PAPER VIII AND IX : ECONOMIC STATISTICS (PRACTICAL)

Time : 4 hours. Full marks : 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is permitted.

1. The following table gives the distribution of incomes of income-tax assessees in the years 1950-51 and 1955-56.

Grade of total income	No. of assessees in 1950–51	Total income assessed in 1950–51	No. of assessees in 1955–56	Total income assessed in 1955–56
0— 25,000	6,062	4,60,98,283	4,87,828	401,57,81,681
25,001— 40,000	17,480	54,25,39,147	21,359	65,88,15,593
40,001— 55,000	5,734	26,54,34,514	7,514	34,89,56,592
55,001— 70,000	2,876	17,88,07,148	3,363	20,70,48,870
70,001— 85,000	1,639	12,59,17,999	1,782	13,69,59,519
85,001—1,00,000	983	9,06,47,872	1,134	10,43,89,904
1,00,001—1,50,000	1,729	21,09,98,934	1,791	21,60,16,204
1,50,001—2,00,000	746	12,84,40,862	731	12,55,85,901
2,00,001—3,00,000	603	14,70,80,059	739	17,90,33,895
3,00,001—4,00,000	316	10,77,48,961	376	13,00,59,220
4,00,001—5,00,000	208	8,33,45,159	243	10,84,56,253
5,00,001 and over	729	119,48,84,993	877	1,61,77,74,257
Total	39,107	313,10,43,940	527,737	7,84,90,77,789

Fit Lorenz curves of concentration for the above data relating to 1950–51 and 1955-56 and interpret the changes in concentration of incomes during the first Five Year Plan period. (40)

2. Using month-end wholesale price data for wheat at Hapur, analyse seasonal variation in wheat prices. (30)

Year	Month											
	January	February	March	April	May	June	July	August	September	October	November	December
1952	16.00	16.75	17.00	17.00	16.50	17.00	19.25	19.00	19.25	19.75	19.50	21.00
1953	22.25	23.00	20.50	16.00	15.87	16.12	17.00	16.50	15.00	16.37	15.50	17.00
1954	16.25	16.25	15.50	15.00	14.50	12.50	12.50	12.75	12.62	12.75	12.25	13.37
1955	13.37	14.00	11.87	9.12	9.62	10.62	11.69	11.50	11.75	12.50	12.62	14.25
1956	14.62	15.00	14.00	13.00	13.56	13.59	14.31	15.00	14.75	15.00	15.75	15.81
1957	16.12	16.37	16.00	15.50	14.31	14.50	16.00	15.75	15.60	15.75	15.75	15.75

3. Suppose that the income elasticities of demand and the percentage rates of change of income in the rural and urban sectors, and also the ratio of urban to rural incomes are known. Obtain a mathematical expression for deriving the income elasticity of demand for the two sectors taken together.

Use the data given below to calculate the income elasticity of demand for all India :—

(i) income elasticity of demand in urban sector is 0.23 and that in rural sector, 0.45,

(ii) percentage increase in rural incomes, urban income and total income are 4.79, 6.06 and 5.61 respectively , and

(iii)urban per capita consumption is 0.8 of rural per capita consumption and urban population is 21.58 per cent of total population. (30)

PAPER VIII AND IX : STATISTICAL QUALITY CONTROL (PRACTICAL)

Time : 4 Hours Full marks : 100

(a) Attempt *all* questions.

(b) Figures in the margin indicate full marks.

(c) Use of calculating machines is permitted.

1. *Either*

The standard deviation of the error in measurement of a characteristic is determined to be 5 units. The standard deviation of measured values of the characteristic under a state of control is 40 units. Estimate the true standard deviation of the characteristic.

It is believed that the standard deviation of the error in measurement could be reduced to half if the average of four measurements repeated on each item is used instead of taking only one measurement on the item. If this is done, what reduction can be expected in the standard deviation of the measured values. (18)

Or

In an acceptance sampling plan a sample of 45 items is inspected out of a lot of 1000 items. If there are no defectives in the sample the lot is accepted, otherwise rejected. If a lot with 1.5 percent defective is submitted for inspection, find out the probability (using Tables of Poisson distribution or otherwise) that it will be accepted. (18)

2. Find out the elements (n and t) for a known-sigma single sampling variables plan for acceptance, for a one-sided specification, such that the OC of the plan will match approximately that of the single sampling plan having $n = 15$ and $c = 2$.

For this plan what will be the probability of accepting a 5 percent defective lot. (25)

3. (i) Obtain 95 percent confidence limits for the lot standard deviation given that in a sample of 12 pieces of coal, the range of percentage ash content was 16.5.

(ii) Obtain a point-estimate of the lot standard deviation, given that the mean range based on 5 samples each of 7 observations of a dimension, was 16.2 mmg.

(iii) Obtain a 99 percent confidence interval for the number of defects per unit in the lot given that in a sample of 15 units, the total number of defects was 12. (24)

4. A component in an electric switch consists of a semi-circular brass piece with 3 teeth on it, numbered 1, 2, 3 in a particular order. The teeth on each piece are machined together. It is desired to control the thickness of teeth and data collected for this purpose are given below :—

thickness of teeth in mms.

piece no.	tooth 1	tooth 2	tooth 3
1	1.021	1.056	1.043
2	1.096	1.083	1.075
3	0.998	1.007	1.026
4	1.008	1.001	1.052
5	1.102	1.112	1.116
6	1.086	1.093	1.085
7	1.101	1.031	1.086
8	1.093	1.062	1.042
9	1.109	1.112	1.106
10	1.052	1.063	1.065

Utilise the data to set up appropriate control charts and examine control for thickness of teeth. (33)

PAPER VIII AND IX : SAMPLE SURVEY, THEORY (PRACTICAL)

Time : 4 Hours Full marks : 100

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

1. (a) A sample survey was carried out to estimate the area under a specified crop in a district. In each of the four subdivisions of the district, villages were selected with equal probability without replacement and in each selected village four clusters of eight khasra numbers each were selected (equal probability and without replacement).

Obtain a suitable estimate of the area under the crop in the district and the estimated variance of this estimate. Is your estimate of crop acreage biased ? Give reasons.

Villages / Strata	Area under crop in cluster (in acres) (1)	(2)	(3)	(4)	(5)	(6)
I	0	110	447	864	—	—
(26924)	625	237	444	439	—	—
	0	134	252	0	—	—
	134	77	130	482	—	—
II	158	0	0	126	61	—
(38255)	32	0	100	490	152	—
	325	0	175	90	260	—
	139	0	47	30	275	—
III	320	177	391	75	113	287
(41990)	222	46	0	39	0	0
	270	305	331	69	540	0
	125	170	20	22	232	90
IV	215	230	561	37	343	669
(104710,	132	0	327	22	732	38
	143	310	195	26	376	99
	27	15	255	0	332	

Total number of survey numbers have been indicated in parenthesis under the numbered strata.

(b, If the villages are selected with probability proportional to survey numbers and with replacement, give a broad outline of the analysis.

2. (a) The following data show the stratification of all the farms in a district by farm size and the average area under Maize (in acres) per farm in each stratum.

Farm size (acres)	number of farms	average area under maize (acres)	standard deviation of area under maize
0— 40	394	5.4	8.3
41— 80	461	16.3	13.3
81—120	391	24.4	15.1
121—160	334	34.5	19.8
161—200	169	42.1	24.5
201—240	113	50.1	26.0
241—	148	63.8	35.2

For a sample of 100 farms, compute the sample sizes in each stratum under (i) proportional allocation, (ii) optimum allocation. Compare the precision of these methods with that of simple random sampling.

(b) If the strata in the above question are to be combined into a lesser number of strata, say, two, what is the best point of division for proportional allocation ?

PAPER VIII AND IX: SAMPLE SURVEY, APPLIED (PRACTICAL)

Time : 4 Hours Full marks : 100

(a) Attempt *all* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.
(d) All workings should be shown in neat tabular form.

1. You are given in Table I (given on next page) a list of villages in a Tehsil of India along with their population in 1931 and cultivated area.

(i) In the table '—' means that the figures are not available. State clearly as to what action you will take in respect of cases where population figures are not available, if you are asked to choose 10 villages from this list with probability proportional to population.

(ii) After taking necessary action as in (i), choose a sample of 10 villages with probability proportional to population and with replacement. State clearly the rule of procedure adopted by you and reference to any tables used.

(iii) Estimate the total cultivated area in the Tehsil from this sample (using the figures of cultivated area given in Table 1, for the selected villages as having been obtained from a sample and assuming that figures for other villages are not available).

(iv) Estimate the standard error of this estimate.

2. It is proposed to list the households in a village and collect the following auxiliary information with a view to using them for sampling of households or for other purposes:—

(a) number of persons in the household

(b) the group to which the household belong to: self-employed in registered manufacturing establishment-1; self-employed in unregistered manufacturing establishment-2; others-3.

(c) principal means of livelihood of household which is self-employed in unregistered manufacturing establishment (group code 2):

Class 1: manufacturing as principal means

class 2: others as principal means.

Note: A household is regarded as self-employed if at least one of its members is self employed usually. This self-employment may be in a manufacturing establishment or elsewhere. The establishment may be a registered factory under the factories act or may be an unregistered factory.

The principal means of livelihood of the household is decided on the basis of the source of major income.

From the list thus constructed, it is proposed to construct for each of the following enquiries suitable sampling frames by assigning suitable sampling serial numbers to every household and draw samples of households according to the following schemes:—

(i) For consumer expenditure schedule: Sampling serial numbers are to be assigned in such a manner as to give effect of a rearrangement of the list of households such that households of size less than 4 get the earlier serial numbers and the rest will get the subsequent serial numbers. From this frame of households, 4 are to be selected systematically.

(ii) For enquiry of household enterprise: Households with group code 2 and class code 1 will be given the earlier serial number and those with group code 2 and class code 2 will be given the subsequent serial numbers. From this frame of households, 3 are to be selected systematically

Draw a suitable schedule with adequate spacing and heading for the purpose. With reference to this schedule, write the instructions to the investigators explaining as to how he should proceed to construct the frame for sampling of households and how he should proceed to draw the samples. The schedule and instructions should be so framed as to make the work of the investigator as simple and light as possible.

3. *Either,*

(a) List the items on which you should collect data on a sample basis to estimate the total cost of cultivation of paddy per acre and per 100 maunds of paddy produced.

(b) Indicate the appropriate reference period for each item.

(c) Define the important terms used, if any.

(d) What is the best possible source of information —the unit of enquiry. Give reasons.

Or,

It is proposed to collect data from a sample survey in your State to know (i) the extent of literacy among male and female population of different household income groups, (ii) reasons for low standard of education, if any, (iii) educational facilities available to them, (iv) educational facilities and assistance needed by them.

Prepare a list of important Tables which will throw light on the above aspects. The structure of such tables should also be presented.

TABLE 1

List of villages in a Tehsil of India along with their population and cultivated area in 1951

serial no. of village	population	cultivated area (acres)	serial no. of village	population	cultivated area (acres)
1	231	40	43	79	96
2	242	123	44	350	215
3	214	184	45	172	43
4	165	155	46	905	210
5	—	..	47	9000	2570
6	—	—	48	240	137
7	226	145	49	—	29
8	35	33	50	91	317
9	75	31	51	15	191
10	370	70	52	440	196
11	671	313	53	333	403
12	84	—	54	291	416
13	1604	1027	55	110	225
14	554	246	56	192	332
15	111	109	57	186	175
16	—	48	58	222	211
17	235	389	59	79	101
18	553	401	60	199	254
19	740	505	61	1199	853
20	123	9	62	876	726
21	199	64	63	164	228
22	5	135	64	—	18
23	127	62	65	—	31
24	111	123	66	200	160
25	146	234	67	108	229
26	1509	1384	68	161	279
27	2446	1016	69	1152	389
28	760	442	70	302	520
29	19	3	71	1121	—
30	143	13	72	261	240
31	24	11	73	—	258
32	100	13	74	182	369
33	82	40	75	3700	3437
34	7	5	76	264	156
35	58	5	77	202	—
36	558	221	78	556	303
37	1090	250	79	117	241
38	731	284	80	131	117
39	123	125	81	87	137
40	79	63	82	113	105
41	164	105	83	70	96
42	80	56	84	—	—
		grand total		38774	

'—' means that figures are not available.

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

Time : 4 Hours. Full marks 100

(a) Attempt *any three* questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. A $(3 \times 2 \times 2)$ factorial experiment was carried out on wheat in blocks of six plots, the treatments being all combinations of three levels of nitrogen (n_0, n_1, n_2), two levels of potash (k_0, k_1) and two levels of super phosphate (p_0, p_1). The layout plan and yields in lbs per plot of $\frac{1}{50}th$ of an acre are given below :-

Replication I				Replication II				Replication III			
Block 1		Block 2		Block 1		Block 2		Block 1		Block 2	
nkp		*nkp*		*nkp*		*nkp*		*nkp*		*nkp*	
001	29.3	000	24.8	000	23.3	001	26.8	000	30.7	001	38.7
010	22.7	011	34.1	011	25.5	010	23.4	011	28.9	010	31.1
100	40.3	101	52.3	101	52.2	100	41.0	100	61.8	101	75.0
111	52.1	110	40.4	110	40.5	111	45.1	111	73.1	110	56.3
200	37.0	201	45.8	200	41.0	201	48.4	201	56.6	200	63.1
211	48.6	210	48.2	211	43.7	210	39.8	210	53.8	211	83.1

Find out, giving reasons, what effects have been partially confounded. Analyse the data and interpret the results of analysis. Also calculate the loss of information on the confounded effects.

2. An entomological trial was carried out in five randomized blocks on four plots each for testing the effect of four insecticides *A*, *B*, *C* and D on control of Grundly bers of paddy. The table below shows the layout plan adopted and the data on total number of adults before the application of insecticides and two days after the application for five randomly selected sample areas of $2' \times 2'$ in each plot, the upper figure denoting the total number of adults before the application of the insecticides and the lower figure the total number of adults two days after the application.

Block 1	Block 2	Block 3	Block 4	Block 5
A	C	B	D	B
9	22	19	24	23
1	2	1	16	3
D	B	C	A	C
10	19	16	21	24
9	2	3	3	1
C	D	A	B	D
10	19	13	25	18
1	16	1	2	15
B	A	D	C	A
14	16	22	31	23
2	1	11	3	1

Obtain the percentage mortality of adults for each plot. Analyse the data after transforming the percentages by using the appropriate transformation and interpret the results of analysis. Give the summary of results for both the transformed and

original variates. State with reasons, whether or not it is possible to obtain from the analysis the critical difference for testing the significance of difference between treatment means for the original variate (percentage mortality).

3. A 5×5 Latin square experiment was carried out on the sugarcane crop to test the effect of five manurial treatments A, B, C, D and E, of which A was control (or no manure). The layout plan and the yields of plant in seers per plot of 1/10th of an acre are given in the table below :—

A 1119	E 1840	D 1438	C 1281	B 1639
D 1762	B 1678	A 1200	E 1761	C 1720
B 1964	A 1322	C 1600	D 1637	E 1876
C 1482	D 1521	E 1875	B 1802	A 1400
E 1883	C 1482	B —	A 1604	D 1476

One of the observations for treatment B in the fifth row (or third column) is missing. Analyse the data and state your conclusions, presenting the results in the form of a summary table.

Indicate the procedure of analysis when a value corresponding to another treatment is missing.

4. A 3^3 factorial experiment was carried out on paddy to study the effect of phosphatic fertilizers at different levels in conjunction with different levels of nitrogen, the treatments being all combinations of :

$$\left\{\begin{array}{l} n_0 = 0\text{lbs./acre} \\ n_1 = 20\text{lbs./acre} \\ n_2 = 40\text{lbs./acre} \end{array}\right\} \times \left\{\begin{array}{l} s_1 = \text{triple super-phosphate} \\ s_2 = \text{bone meal} \\ s_3 = \tfrac{1}{2}\text{triple super-phosphate} + \tfrac{1}{2} \text{rock phosphate} \end{array}\right\} \times \left\{\begin{array}{l} p_0 = 0\text{lbs/acre} \\ p_1 = 20\text{lbs/acre} \\ p_2 = 40\text{lbs acre} \end{array}\right\}$$

The layout plan and the dry weight of grain of paddy in lbs. per plot (of 1/60th of an acre) are given below:—

Block I		Block II		Block III	
$n_1s_2p_2$	37.7	$n_0s_2p_0$	30.2	$n_0s_2p_0$	32.2
$n_0s_2p_2$	31.8	$n_2s_1p_1$	32.7	$n_1s_1p_2$	39.2
$n_1s_1p_1$	32.6	$n_0s_1p_2$	26.6	$n_1s_3p_1$	40.3
$n_0s_2p_1$	23.0	$n_1s_2p_1$	28.1	$n_1s_2p_0$	32.2
$n_2s_1p_2$	32.8	$n_2s_2p_2$	30.4	$n_2s_2p_1$	41.2
$n_2s_3p_1$	30.2	$n_1s_3p_2$	29.2	$n_2s_3p_2$	39.2
$n_0s_1p_0$	24.1	$n_2s_3p_0$	22.7	$n_0s_3p_2$	38.2
$n_2s_2p_0$	28.3	$n_1s_1p_0$	17.5	$n_0s_1p_1$	34.2
$n_1s_3p_0$	35.2	$n_0s_3p_1$	26.9	$n_2s_1p_0$	40.8

Analyse the data and draw some conclusions.

PAPER VIII AND IX: MATHEMATICAL THEORY OF SAMPLING DISTRIBUTION (PRACTICAL)

Time: 4 Hours | Full marks: 100

Answer all questions.

1. (a) The following table gives the sums of squares and product moments of two variables x_1 and x_2 in five experiments with 10 subjects in each experiment. Test for the heterogeneity of the correlations observed in the five experiments.

		Σx_1^2	Σx_2^2	$\Sigma x_1 x_2$
Experiment	1	168.9	2020.5	458.5
	2	192.1	715.6	102.6
	3	132.1	2860.6	169.4
	4	158.1	1904.9	333.7
	5	191.6	5722.1	689.6

(b) Assuming x_1 to be the body weight in kilograms of a cat and x_2 the heart weight in grams, the relevant data for two samples, 47 female and 97 male cats, are summarised as follows:—

sex of the cat	no. in the sample	mean values		uncorrected product matrix		
		x_1	x_2	x_1^2	$x_1 x_2$	x_2^2
female	47	110.0	432.5	265.13	1029.62	4064.71
male	97	281.3	1098.3	836.75	3275.55	13056.17

Test the difference between the above two samples.

2. The corrected sums of squares and product moments of four variables x_1, x_2, x_3 and y are given below:-

	x_1	x_2	x_3	y
x_1	16.68	1.93	0.82	1.51
x_2		0.99	0.34	0.60
x_3			0.22	0.18
y				0.13

Calculate the regression coefficients $b_{y1\cdot23}$, $b_{y2\cdot13}$ and $b_{y3\cdot12}$ and test the significance of these regression coefficients, given that the above data are based on 44 observations.

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 statistical tables and calculating machines is permitted.

1. A test was given to *four* different groups of students. The frequency distribution of scores for each group is given below. Make relevant analysis of the data and suggest what conclusions can we draw about differences between the groups on the ability measured by the test:-

Test score class interval	Frequencies in four groups			
	A	B	C	D
27—29	3	24	4	26
24—26	7	24	6	34
21—23	12	41	5	57
18—20	17	63	20	101
15—17	27	93	18	119
12—14	22	101	16	91
9—11	25	69	9	40
6— 8	8	24	2	8
3— 5	8	10	1	8
0— 2	0	3	2	2

2. Normalise the distribution of scores given below and express the normalised distribution on a scale of $\bar{X} = 50$ and $S.D. = 10$:

Score	Frequency	Score	Frequency
0	3	16	81
3	4	17	59
4	5	18	67
5	15	19	49
6	30	20	53
7	22	21	37
8	30	22	33
9	39	23	16
10	49	24	21
11	47	25	17
12	64	26	6
13	60	27	1
14	68	28	4
15	62	29	1

3. Is there any reason to believe that *more than one* common factor is involved in the set of inter-correlations among ability tests given below?

	Tests							
	1	2	3	4	5	6	7	8
1	—							
2	.49	—						
3	.59	.45	—					
4	.43	.40	.41	—				
5	.44	.45	.47	.57	—			
6	.40	.43	.35	.60	.51	—		
7	.45	.41	.34	.52	.58	.59	—	
8	.46	.40	.40	.57	.66	.66	.65	—

4. From the data given below make an estimate of the validity of Mechanical score with the effect of Intelligence partialed out:

Person No.	Intelligence score	Mechanical score	Criterion score
1	45	15	30
2	60	25	42
3	34	12	25
4	55	18	40
5	72	27	47
6	50	20	25
7	61	22	28
8	43	16	19
9	54	17	27
10	53	14	23
11	63	19	32
12	39	12	15
13	42	15	17
14	24	9	11
15	55	18	35
16	47	10	16
17	59	24	32
18	20	6	10
19	48	15	24
20	15	13	12