INDIAN STATISTICAL INSTITUTE

QUESTION PAPERS

for

The Statistician's Diploma Examinations

April & September 1957

Price Re

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, APRIL 1957

PAPER I—THEORETICAL STATISTICS (GENERAL)

Time: 4 *Hours* *Full Marks*: 100

(a) Answers to the different groups are to be given in separate books.
(b) All questions carry equal marks.

GROUP A (60 Marks)

Attempt ANY THREE questions

1. An urn contains w white and b black balls. All the balls are drawn out at random one by one without replacement.

(i) What is the probability that the first and the last ball drawn are of different colour?

(ii) What is the probability that among the first three balls drawn there is at least one black ball?

2. Write a short note on the concept of correlation carefully distinguishing between the concepts of correlation ratio and correlation coefficient. Explain the meaning of total, partial and multiple correlation coefficients. Derive an expression for multiple correlation coefficient and show that the multiple correlation coefficient increases as the number of independent variables increases.

3. A farmer sells bean seeds in packets of 100 seeds and *agrees to refund* the price if the number of seeds (in a packet) that germinate is less than 75. The farmer has a large number of packets of seeds to sell and his cost of production has been Rs. 2/- per packet. From past experience, the farmer knows that the average number of seeds that germinate per packet is 80. How should the farmer fix the price of a packet so that in the long run he may expect to earn a profit of eight annas per packet sold? You need not actually compute the price. Only explain how you are going to find it. You may assume that the number of seeds that germinate in a packet of 100 is a binomial variable).

4. (a) *Prove* Newton's Backward Difference Interpolation formula.

(b) *State* Stirling's Central Difference Interpolation formula and the Three-eighth rule of numerical integration.

State the assumptions involved in both these cases.

GROUP B (40 Marks)

Attempt ANY TWO questions

5. (a) Write a short note on the measurement of strength of association between two attributes A and B.

In a sample

$$(AB) = 97 \qquad (A\beta) = 66$$
$$(\alpha B) = 44 \qquad (\alpha\beta) = 40$$

Are the two attributes independent ? If each of the above frequencies were increased five fold, would your conclusions be altered ?

(b) Show that if x_1^2 and x_2^2 are two independent χ^2 variables with n_1 and n_2 degrees of freedom respectively, then $u = x_1^2 + x_2^2$ obey the same distribution with (n_1+n_2) degrees of freedom.

6. Given k univariate samples of size $n_1, n_2, \ldots n_k$ with mean values $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_k$ and standard deviations $s_1, s_2, \ldots s_k$ supposed to have been drawn from k univariate normal populations with mean $m_1, m_2, \ldots m_k$ and standard deviations $\sigma_1, \sigma_2, \ldots \sigma_k$.

Discuss the procedure you would adopt to test the hypothesis that

(i) $m_1 = m_2 = \ldots = m_k$

(ii) $\sigma_1 = \sigma_2 = \ldots = \sigma_k$

Mention the assumptions involved in your test procedure.

Discuss with reasons whether you can test (i) and (ii) irrespective of each other.

7. (a) If $x_i (i = 1, 2, \ldots, n)$ is a random sample of size n from a normal population with mean μ and variance σ^2, starting from the joint distribution of $\bar{x}$ and s, obtain the distribution of :

$$t = \frac{\sqrt{n}(\bar{x}-\mu)}{s} \text{ where}$$

$$n\bar{x} = \sum_1^n x_i \text{ and } (n-1)s^2 = \sum_1^n (x_i - \bar{x})^2.$$

(b) What are the different situations in which the t-distribution is used for drawing inference from statistical data ?

Mention the assumptions involved in each situation.

(c) In what situation can we use the F-distribution as an alternative to t distribution for drawing inference ? Give reasons.

Paper II—Applied Statistics (General)

Time : 4 *Hours* *Full Marks :* 100

(a) Answers to the different groups are to be given in separate books.
(b) All questions carry equal marks.

Group A

Attempt ANY THREE questions.

1. Describe the administrative organisations engaged in India in the collection and publication of agricultural statistics, giving briefly their functions and content of their publications.

How are surveys for estimation of crop yields organised in your State ?

What are the crops covered by these surveys and over what areas of the State? Can you suggest any improvements in the procedure ?

2. What does an index number of industrial production attempt to measure ? Describe the scope and method of construction of the official index of industrial production in India. What are its limitations, if any, and what steps would you suggest for its improvement ?

3. You are required to conduct a sample survey in a medium sized city to ascertain the economics and difficulties of a few selected small scale industries. Describe briefly how you would organise the survey, giving the sample design, the items of information you will put in the schedule (preferably in the form of a schedule), the organisation of field work and check, scrutiny and tabulation of filled up schecules, and the types of results you will like your final tables to show.

4. What are the basic components of a time series ? Describe one method each of (a) eliminating the effects of trend from a time series, and (b) measuring seasonal variations.

In measuring seasonal variation, can cyclical and erratic influences be eliminated and if so, how ?

GROUP B

Attempt ANY THREE questions.

5. (a) What is a Latin Square ? Explain its uses in agricultural experiments. In what situations is Latin Square preferred to a randomised block ?

(b) What are the advantages of factorial design ? Explain the utility of confounding and partial confounding of some interactions.

6. (a) Write notes on (i) stationary population,
(ii) stable population.

(b) What is a life table ? Explain how you would proceed to construct a life table from census returns only.

7. Discuss the different methods of estimating the reliability of a psychological test.

To determine the reliability of a mathematical aptitude test containing 50 items, it was divided into two parallel subtests each containing 25 items. The subtests give validity coefficients 0.65 and 0.66 respectively with some criterion whereas the coefficient of correlation between the two subtests is 0.73. Find the reliability coefficient of the total test as well as its validity coefficient with the criterion. State precisely the assumptions on which your computations are based.

8. (i) If you are told that a man who was badly burned had, afterwards, children with bluish spots of skin, do you accept this as proof of inheritance of acquired characters? What investigations would you make ?

(ii) What is meant by *linkage* in *heredity* ? How is it detected ? Give details of a method for its detection with reference to a back-cross involving two factors.

PAPER III—STATISTICAL INFERENCE (GENERAL)

Time : 4 *Hours* *Full Marks :* 100

(a) Answers to the different groups are to be given in separate books.

(b) All questions carry equal marks.

GROUP A (40 Marks)

Attempt ANY TWO questions.

1. Describe the method of maximum likelihood. State the advantages (if any) and disadvantages (if any) of this method of estimation. Is there any relation between this method and that of Bayes' ?

2. State and prove the information inequality (Cramer-Rao lower bound) concerning the variance of an unbiased estimate of a real parameter.

Give examples to show how this result can *sometimes* be used to verify that a particular unbiased estimate is of minimum variance. No detailed verifications or proofs are required in answering this part of the question.

3. What is a confidence interval ? Are such intervals useful in practice ?

Suppose that $X_1, X_2, ..., X_n$ is a sample of independent observations from a continuous but otherwise entirely unknown population. Let $Y_1 \leq Y_2 \leq ... \leq Y_n$ be the sample values arranged in ascending order of magnitude.

Show that for $i < j$, (Y_i, Y_j) is a confidence interval for the population median. How would you choose i and j ?

GROUP B (60 Marks)

Attempt ANY THREE questions.

4. What are the errors of first and second kinds ? How are they controlled in testing a simple hypothesis with respect to a particular alternative ? What modification is necessary in case of a composite hypothesis ?

5. What is a discriminant function in relation to problems of classification ? Give a mathematical solution to the problem for the case of two groups of multiple measurements ?

6. Explain clearly the importance of Hotelling's T^2 Statistic in multivariate analysis. Discuss the various situations where it can be used.

7. Write notes on :—

(i) Locally most powerful tests,

(ii) Fiducial distributions.

PAPER VI—(PRACTICAL)

Time : 6. *Hours* *Full Marks :* 100

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

1. Samples of agricultural households were drawn from two localities and their total earnings (x) in a month were ascertained. The following summary data is supplied :

locality	n	Σx	Σx^2
A	15	441.7	19795.05
B	10	260.7	19178.57

(i) Test the hypothesis that σ_A = Rs. 20.00 in locality A.
(ii) Can we consider the two samples as coming from populations with the same standard deviations ?
(iii) Are the average earnings in the two localities significantly different from one another ? (10)

2. The following table gives the results of an examination of a sample of 20 ears of corn :

	Mean	Standard deviation
Ear weight—e(gms)	106.55	27.68
Grain weight—g (gms)	90.96	24.68
Cob-weight($e-g$) = c	15.59	
r_{eg} =	0.994	

Test whether cob-weight obtained from the sample is significantly different from 12 gms. (10)

3. The following summary results are supplied to you from an investigation of data relating to investment and production of cement and steel in a country.

y = per capita net investment in units of (p) Rs. at 1948-49 prices.
x_1 = per capita production of cement in units of (q) ton
x_2 = per capita production of steel in units of (r) ton.
n = 14, Σy = 2.0432, Σx_1 = 11.5530, Σx_2 = 3.9648

Σy^2 = 0.3006, Σx_1^2 = 10.2936, Σx_2^2 = 1.1320

correlation matrix

	y	x_1	x_2
y	1.0000	0.8093	0.9082
x_1	0.8093	1.0000	0.8726
x_2	0.9082	0.8726	1.0000

Work out the values of $R_{y.x_1x_2}$ and $b_{yx_1.x_2}$ and test for their significance. (20)

4. Each of the 42 cultivators in five villages tried a special manure recommended by an agricultural experimental station for growing paddy in one plot of size 7′×7′ and also local manure in another adjacent plot of the same size. The gain in yield (in tolas per plot of 7′×7′) due to use of special manure was analysed and the following are the summary results :—

Village	n	Sum of gains Σx	Sum of squares of gains Σx^2
(1)	(2)	(3)	(4)
A	17	1353	277480
B	13	1229	132578
C	3	360	61720
D	7	150	9250
E	2	30	890

(i) Is the average gain significantly different from village to village ?

(ii) Estimate the overall average gain and its standard error. (10)

5. (a) Compute the first and second derivatives of $f(x)$ at $x = 304$ from the following table of values of $f(x)$:—

x	$f(x)$
300	5.70378247
301	5.70711026
302	5.71042702
303	5.71373281
304	5.71702770
305	5.72031178
306	5.72358510
307	5.72684775

(15)

(b) Using suitable interpolation formulae, obtain the 5 percent points of χ^2 for the degrees of freedom $\nu = 27.35$

Five percent points of χ^2

ν	χ^2	ν	χ^2
25	10.5197	29	13.1211
26	11.1603	30	13.7867
27	11.8076	40	20.7065
28	12.4613	50	27.9907

(10)

6. In a study published in the American Journal of Sociology, children of different social groups were classified according to their participation in organised group activities

Group A consists of children from upper-middle class and upper-class families, group B from lower-middle class families, group C from upper-lower class families and group D from lower-class families.

		Number of children studied	Number of non-participants
Social group	A	61	17
	B	55	22
	C	67	24
	D	43	18

Test whether the proportion of children not participating in organised group activities is different for children of different social groups.

Test also the difference in the proportions between groups A and D. (15)

7. The following data relate to the average monthly production of bicycle in India during the years 1948-1955 (in hundreds) :—

1948	46	1952	164
1949	54	1953	220
1950	86	1954	310
1951	93	1955	410

Fit a cubic to the data by the method of orthogonal polynomials. (10)

Paper VII (Practical)

Time : 6 *Hours* *Full Marks :* 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is permitted.

1. Given $l_{91} = 871$ and

d_{91}	d_{92}	d_{93}	d_{94}	d_{95}	d_{96}	d_{97}	d_{98}	d_{99}	d_{100}	d_{101}
296	209	144	93	58	34	18	10	5	3	1

where l_x and d_x have the usual meaning as in a life table.

Find the probability that

(a) a person aged 93 will die in 3 years ;

(b) a person aged 92 will survive age 96 ;

(c) 3 persons aged 92, 93 and 94 will survive 4 years ;

(d) of 3 persons aged 93, 94 and 95, at least two will survive 4 years. (10)

2. The following table gives yields of three varieties of Alfalfa (in tons per acre) in 1944 following four successive dates of final cutting in 1943. Some of the necessary computations have also been done and shown.

Variety	Date	Blocks 1	2	3	4	5	6	total ++	Square of total
Ladak	A	2.17	1.88	1.62	2.34	1.58	1.66	11.25	126.5625
	B	1.58	1.26	1.22	1.59	1.25	0.94	7.84	61.4656
	C	2.29	1.60	1.67	1.91	1.39	1.12	9.98	99.6004
	D	2.23	2-01	1.82	2.10	1.66	1.10	10.92	119.2464
Total++		8.27	6.75	6.33	7.94	5.88	4.82	39.99	406.8749
Sq. of total		68.3929	45.5625	40.0689	63.0436	34.5744	23.2324	274.8747	
Cobnek	A	2.23	2.01	1.70	1.78	1.42	1.35	10.59	112.1481
	B	1.38	1.30	1.85	1.09	1.13	1.06	7.81	60.9961
	C	1.86	1.70	1.81	1.54	1.67	1.88	10.46	109.4116
	D	2.27	1.81	2.01	1.40	1.31	1.06	9.86	97.2196
total++		7.84	6.82	7.37	5.81	5.53	5.35	38.72	379.7754
Sq. of total		62.4656	46.5124	54.3169	33.7561	30.5809	28.6225	255.2544	
Ranger	A	1.75	1.95	2.13	1.78	1.31	1.30	10.22	104.4484
	B	1.52	1.47	1.80	1.37	1.01	1.31	8.48	71.9104
	C	1.55	1.61	1.82	1.56	1.23	1.13	8.90	79.2100
	D	1.56	1.72	1.99	1.55	1.51	1.33	9.66	93.3156
total++		6.38	6.75	7.74	6 26	5.06	5.07	37.26	348.8844
Sq. of total		40.7044	45.5625	59.9076	39.1876	25.6036	25.7049	236.6706	
Grand total++		22.49	20.32	21.44	20.01	16.47	15.24	115.97	

N.B. Rows and columns marked with ++ are totals.

Sum of squares of 72 primary observations = 193.4665

Sum of squares of 24 cell totals of Block × date table = 578.0015

Analyse the data and write a report based on your analysis. (25)

3. The following table presents the data from a complete census of 340 villages in a tehsil. The villages are stratified into four strata by the size of the agricultural area. Here

N_i = the number of villages in the i-th stratum,

$\bar{Y}_{N_i}$ = stratum mean for the area under wheat,

S_{W_i} = stratum standard deviation for the area under wheat,

S_{a_i} = stratum standard deviation for the agricultural area.

Stratum	Size of agricultural area in bighas	N_i	$\bar{Y}_N$	Sw_i	Sa_i
1	0— 500	63	112.1	56.3	129.6
2	501—1500	109	276.7	116.4	267.0
3	1501—2500	53	558.1	186.0	276.1
4	> 2500	25	960.1	361.3	982.2

A sample of 34 villages is taken. Find the sampling variance of the estimated area under wheat (1) when the villages are selected by simple random sampling, (2) (2) when the villages are selected by stratified sampling and allocated (a) in proportion to the size of the strata N_i, (b) in proportion to $N_i Sw_i$ and (c) in proportion to $N_i Sa_i$. Calculate the relative efficiencies compared to sampling with proportional allocation.

Comment on your results. (30)

4. The following are the quarterly index numbers of wholesale prices in eight successive years

Year	Quarters I	II	III	IV
1	122	125	118	117
2	110	114	114	109
3	105	99	93	89
4	86	90	83	84
5	85	80	80	78
6	77	80	81	80
7	82	81	83	82
8	83	84	85	86

Find the trend and compute the indices of seasonal variation. (15)

5. The following gives the weights in grains of consecutively produced pellets rom two machines in a factory, during the same period of time.

Machine 1		Machine 2	
Sample No.	Weight	Sample No.	Weight
1	122.0	1	121.0
2	122.0	2	120.0
3	117.5	3	123.5
4	110.5	4	128.0
5	112.5	5	120.5
6	121.5	6	122.0
7	125.5	7	118.0
8	121.5	8	118.0
9	121.0	9	128.5
10	122.0	10	121.0

Machine 1		Machine 2	
Sample No.	Weight	Sample No.	Weight
11	123.5	11	126.5
12	118.5	12	118.5
13	122.0	13	128.0
14	123.0	14	119.5
15	121.5	15	125.0
16	122.5	16	122.0
17	124.0	17	120.0
18	127.0	18	128.0
19	125.0	19	124.0
20	122.5	20	126.0
21	124.5	21	124.5
22	122.5	22	122.5
23	123.5	23	123.9
24	125.0	24	125.0
25	125.5	25	125.5
26	126.5	26	123.0
27	122.5	27	128.0
28	122.0	28	123.5
29	123.0	29	121.5
30	124.0	30	125.5
31	124.0	31	125.0
32	127.0	32	126.5
33	124.0	33	123.0
34	125.0	34	125.0
35	125.5	35	120.5
36	124.5	36	114.5
37	128.0	37	118.5
38	128.0	38	116.0
39	127.5	39	115.5
40	117.5	40	121.5
41	119.0	41	123.0
42	118.0	42	123.5
43	121.0	43	120.0
44	118.0	44	115.5
45	118.0	45	117.0
46	123.0	46	122.0
47	120.0	47	121.5
48	116.0	48	116.5
49	124.5	49	121.0
50	119.0	50	121.5

By examining the mean and range charts

(i) determine if stable operating conditions were in existence for both the machines during the time data were taken.

(ii) can the same limits be used for both the machines ?

(iii) suppose the purchaser of the material imposes a minimum weight tolerance of 115 grains. The producer is willing to take only 1 in 1000 risk of violating this. Recommend a control procedure. (20)

PAPER IV AND V—STATISTICAL QUALITY CONTROL (THEORETICAL)

Time : 4 *Hours* *Full Marks :* 100

(a) Attempt ANY FOUR questions.
(b) All questions carry equal marks.

1. What are '3-sigma limits' and 'probability limits' ? Examine their uses in the analysis of data by the c-chart.

A control chart for defects per unit u uses probability limits corresponding to probabilities 0.975 and 0.025. The central line on the control chart is at $u' = 2.0$. Determine the position of the U.C.L. and L.C.L. when the number of units (N) in each area examined = 5.

2. Explain what are meant by :—
'Tolerance limits' and 'Confidence limits'
Describe a practical method of obtaining the 'Tolerance limits' of a controlled production of a measurable quality. State the assumptions and approximations involved and derive the necessary equations.

3. Write a critical mathematical review of the methods, on which '(Dodge—Romig) Sampling Inspection Tables' for single and double sampling are based.
Indicate a few situations in which these tables can be used.

4. Discuss the meaning and uses of OC-curve and ASN-curve in Acceptance Sampling. Explain their significance in non-technical terms to a business executive.

For a sequential sampling plan by attributes (for proportion defective), indicate how the five-point OC-curve and the five-point ASN-curve can be obtained.

5. Write short notes :—
(i) The p-chart
(ii) Multiple sampling
(iii) A.O.Q.L.
(iv) Rational sub-groups
(v) Variables Inspection

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. What are the difficulties in comparing the costs of living in two countries ? What method will you recommend for comparing the cost of living in Delhi with that in London ?

2. What do you mean by correlogram analysis ? How will you use this type of analysis to find out the nature of a stationary time series ?

3. How will you forecast the demand for food-grains in India in 1957 ? Describe the procedure that you will adopt and the nature of the data you will require for the purpose.

4. Explain what is meant by adult equivalent scale ? Describe how and in which situation such scales are used. Give an account of some methods for obtaining such scales.

5. What is social accounting ? Give an outline of a condensed form of social accounts which you would like to use for India. How will you obtain the estimate of national income from such accounts ?

6. Write a critical note on the statistics of industrial production in India indicating the sources and the nature of the data with respect to exhaustiveness and accuracy.

7. Suppose that you are asked to collect data which can be used for the purpose of rehabilitation of a group of families who are going to be evacuated from an area to be used for constructing an iron and steel factory. Give details about the data that you will collect indicating the purpose to be served by such data.

Paper IV and V—Sample Survey Theory (Theoretical)

Time : 4 *Hours* *Full Marks :* 16

(a) Attempt FOUR questions.
(b) All questions carry equal marks.
(c) Treatment should be mathematical wherever possible.

1. (a) For a fixed total sample size, compare the precisions of
(i) stratified sample with proportional allocation,
(ii) stratified sample with optimum allocation and
(iii) ordinary random sample, ignoring the finite population corrections.

(b) Write a note on the construction of strata for stratified sampling, including a description of the rule of Dalenius for proportional allocation.

2. (a) Work out the approximate expressions for variance and bias of the ratio estimate for the population mean of the variate under enquiry in the case of a random sample.

Compare the precision of the ratio estimate with that of the linear regression estimate and the simple sample-mean of the variate under enquiry.

(b) Discuss the different ratio estimates of the population mean in the case of stratified random sample.

3. (a) Work out the variance of the mean of a systematic sample drawn from a finite discrete sequence, in terms of the correlation coefficient between pairs of units that are in the same systematic sample. Also compare the precision of the systematic sample-mean with the mean of a stratified random sample with one unit per stratum. (All the strata are of equal size, and together over the whole sequence, the total sample size of the two cases, systematic sample and stratified sample, must be the same).

(b) Discuss the problem of estimating from the sample-data the standard error of a systematic sample drawn from a finite discrete sequence.

4. In the case of sampling with replacement with probabilities proportional to 'sizes' of units, derive the expressions for (i) a suitable estimate, T, of the population total, (ii) the variance of T, $V(T)$, and (iii) an estimte of this variance, $V(T)$. Further compare the precision of this estimate, T, with that of the ratio estimate (for the population total) using the size of the units as the auxiliary variate (Ignore finite-population-corrections in the latter case).

5. Write a note on the planning of surveys on successive occasions, discussing particularly the problem of matching the successive samples.

6. (a) The variate y has a probability density function $f(y) = e^{-y}$, $(0 < y)$. This population is divided into two strata at the point y_0 and a stratified random sample of size n is taken with proportional allocation. Find the variance, V, of the mean of the stratified sample as a function of y_0. Further find out the value of y_0 which will minimise V.

(b) Write a brief note on surveys for the purpose of mapping.

Paper IV and V—Sample Surveys Applied (Theoretical)

Time : 4 *Hours* *Full Marks :* 100

(a) Attempt ANY FIVE questions.
(b) All questions carry equal marks.

1. How do you propose to determine the size of sample required to attain a given accuracy when the sample is fully random ? State at least three general rules which may be of value in increasing the accuracy of estimation from sampling procedures.

2. Explain what is meant by (i) Multistage sampling, (ii) Multiphase sampling, (iii) Systematic samples, (iv) Line sampling and (v) Interpenetrating samples.

3. Give an account of the various phases of work which you may have to think of, right from the setting up of the general administrative organisation upto the preparation of the report, if you are placed in charge of execution and analysis of a sample survey.

4. What are the common sources of error to which a complete census may be subject. Indicate what steps you would like to take to control them.

5. If you are required to evaluate, through a sample survey, the success of the Grow More Food campaign in so far as its distribution of manures and fertilisers is concerned, state what should constitute your sample and indicate how far would the collected data be able to throw light on the problem.

6. If estimates of acreages of an important crop of a state are to be obtained through sample surveys, state

(i) what should be sampling frame,

(ii) how should the samples be taken and

(iii) what essential particulars of information should be provided for in the schedule.

7. Describe, with reference to any sample survey in which you might have taken part, the steps that were taken in training the field staff.

8. Give an account of the advantages, if any, of combining a complete census with a sample survey.

Paper IV and V—Design of Experiments Applied (Theoretical)

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. Write an essay on the principles underlying the theory and applications of modern experimental designs.

2. In a randomised block experiment having b blocks and t treatments the yields of two of the plots got mixed up and only their total is known. Estimate the individual yield for each of the two mixed-up plots for each of the three possibilities namely, the two plots belong to :—

(1) different blocks and different treatments,

(2) same block, and

(3) same treatment.

Derive the expressions for the mean variance of all treatment comparisons in the three cases.

3. Explain the principle of *Generalised Interaction* with reference to a 2^m symmetrical factorial experiment.

A 2^4 experiment is to be carried out in blocks of 4 plots each without confounding main effects. Give a suitable design in 4 replications partially confounding every interaction in some replication or other. Write down what effects are confounded in each replication.

4. Explain the parameters of a partially balanced incomplete block design having two associate classes and prove the relations between them.

Derive the values of these parameters for the following examples of these p.b.i.b. designs (i) simple lattice, (ii) triple lattice.

5. In the following design of a $3\times2\times2$ experiment in blocks of six plots each, find out what effects have been partially confounded. Give details of the method of analysis of this design.

Replication 1		Replication 2		Replication 3	
Block 1	Block 2	Block 1	Block 2	Block 1	Block 2
000	001	001	000	001	000
011	010	010	011	010	011
101	100	100	101	101	100
110	111	111	110	110	111
201	200	201	200	200	201
210	211	210	211	211	210

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

(1) Hidden replication

(2) Youden square

(3) Double confounding

(4) Inter-block information

7. A strip-plot design was employed for comparing the effects of 4 levels of irrigation (l_1, l_2, l_3, l_4) in combination with 3 frequencies of irrigation (f_1, f_2, f_3) on the yield of an agricultural crop. The levels and frequencies of irrigation were applied in strips along rows and columns respectively within each of 4 randomized blocks.

Indicate the structure of the table of analysis of variance of the yield per ultimate plot. Derive expressions for the standard error of the difference between mean yields of :—

(i) two frequencies of irrigation under the same level of irrigation

(ii) two levels of irrigation under the same frequency of irrigation

(iii) two frequencies of irrigation under two different levels of irrigation.

PAPER IV AND V—GENETICS (THEORETICAL)

Time : 4 *Hours* *Full Marks :* 100

(a) Attempt ANY FIVE questions

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. If each of two factors shows complete dominance, what are the types of crosses, which yield information on linkage, if any, between them ?

Factor $A-a$ with A showing complete dominance over a, and factor $B-b$ showing no dominance are known to be linked. The progeny from a cross

$$\frac{AB}{ab} \times \frac{aB}{ab}$$

turned out to be as follows :—

	Phenotypic class					
	$A(BB)$	$A(Bb)$	$A(bb)$	$a(BB)$	$a(Bb)$	$a(bb)$
Frequency	24	30	5	8	36	28

Obtain the maximum likelihood estimate of the recombination fraction (assumed the same for both the sexes) and its standard error.

2. In two crosses of a crop plant, the following results were obtained in respect of a trait in which improvement was sought :—

	Cross A	Cross B
Variance between plants in F_2	5.5	6.7
Regression coefficient of selfed F_3 progeny mean on F_2 parental value	0.51	0.16

Asses the relative potentiality of the two crosses for improvement by selection. Estimate approximately the improvement that may be expected to be achieved by selecting the best five percent of plants in cross A.

3. What are the methods of studying single factor segregations in human traits that you know ? 18 families of five children each were found to be segregating for various numbers of albinos as follows :—

No. of albinos	1	2	3	4
No. of families	7	6	4	1

Assuming complete ascertainment, test whether albinism can be regarded as a simple autosomal recessive.

4. Explain Hardy-Weinberg law of equilibrium under random mating. Discuss the disequilibrium due to linkage between two autosomal genes. If the recombination fraction is 0.2, how many generations would be required to halve a given amount of disequilibrium ?

5. Discuss the equilibrium in the frequency of a gene under the forces of selection and mutation, assuming random mating.

6. Explain briefly Bernstein's explanation of the mechanism of inheritance of ABO blood groups in man. Outline the procedure suggested by Stevens for obtaining maximal likelihood estimates of the gene frequencies.

7. Explain the utility of the linear discriminant function technique in plant selection. Indicate in outline the procedure to be followed in arriving at a suitable discriminant function and in judging the effectiveness of the technique relatively to straight selection.

8. Discuss random chromatid segregation for autotetraploids and find the recurrence relations between the gametic proportions of two consecutive generations.

Paper IV and V—Mathematical Theory of Sampling Distribution (Theoretical)

Time : *4 Hours* *Full Marks :* 100

(a) Answer ANY FOUR questions.
(b) All questions carry equal marks.

1. Define a hypergeometric distribution and indicate the relationship, if any, between hypergeometric and binomial distributions.

Obtained the first four cumulants of the hypergeometric distribution.

2. Derive the distributions of the means of samples of size n from Cauchy's and rectangular populations by characteristic functions.

3. Derive the distribution of the multiple and partial correlation coefficients. Indicate the relationship, if any, between these distributions and non-central chi-square.

4. Explain the relationship between Hotelling's T^2 and Mahalanobis's D^2 statistic. Derive the distribution of the D^2—statistics.

5. Discuss in detail the distribution of the roots of the determinantal equation. Also derive the distribution of the largest root of the determinant equation for three variables.

6. On a line of length a, n points are taken at random. Obtain the distribution of the distance between the r-th and $(r+1)$th points from the origin. Get the first four cumulants for this distribution and discuss the limiting form of the distribution as $\frac{r}{n}$ tends to finite limit when r and n are very large.

Paper IV and V—Psychology and Education

Time : *4 Hours* *Full Marks :* 100

(a) Answer ANY SIX questions.
(b) All questions carry equal marks.

1. Present four methods for estimating test reliability where the test consists of a number of questions and has been administered only once. Give formulas, assumptions, advantages and disadvantages of each.

2. Propose a method for estimating the reliability of an essay examination. Describe both the method of collecting the data and the method of data analysis.

3. Compare speed and power tests with respect to :—
   (a) interpretation of the scores
   (b) estimation of test reliability

4. Differentiate and compare aptitude and achievement tests with respect to :—
   (a) definition
   (g) effect of learning
   (c) norms
   (d) interpretation of scores

5. Compare projective and questionnaire methods of assessing personality from the point of view of :—

(a) reliability (b) validity

Giving an example, discuss the methods and problems in each case.

6. Describe and compare two methods of factr analysis.

7. Present an experimental design for the following problem, indicating how the principles of design have been incorporated, the nature of the data to be collected, and the tests of significance to be used.

Problem : an educationist wants to determine which of the following methods of teaching mathematics is best :(1) lecturing to large classes, (2) lecturing plus discussion in small classes, (3) seminar or group discussion in small classes.

8. Discuss the vocational guidance technique in terms of the following two problems :—

(a) use of psychological tests

(b) its application when there is a restricted job market, i.e. the number of jobs is limited.

Paper IV and V—Theories of Inference (Theoretical)

Time : 4 *Hours* *Full Marks :* 100

(a) Attempt ANY FOUR questions.

(b) All questions carry equal marks.

1. Random samples of n_1 and n_2 observations have been taken from two normal populations whose variances may be markedly unequal. Discuss, in detail and critically, the logic of using the Fisher-Behren's distribution in testing the evidence against the hypothesis that the means are equal,and in assigning fiducial limits to the difference between the means (development of mathematical details of the distribution is not required).

2. Discuss the Neyman-Pearson theory of statistical tests, indicating in what respects you consider it inadequate.

Explain the importance of similar regions when there is a nuisance parameter.

3. Explain the term 'bias' as used in the Neyman-Pearson theory of testing of hypothesis, and discuss briefly its relevance to practical situations.

Describe Pitman's method of obtaining an unbiased test for a location or a scale parameter, and deduce that Bartlett's test for the homogeneity of a set of variances is unbiased in the Neyman-Pearson sense.

4. Show how the method of inverse probability leads to the posterior probability distribution of the unknown parameters of a distribution, and point out its relation to the method of maximum likelihood.

θ is the unknown parameter of the rectangular distribution

$$f(x,\theta) = \frac{1}{\theta} \qquad (0 \leqslant x < \theta)$$

Assume that the probability density function of the prior probability distribution of θ is proportional to $\frac{1}{\theta}$. Given a random sample $x_1, x_2, \ldots, x_n$, obtain the posterior probability distribution of θ. What is the maximum-posterior-probability estimator of θ?

5. A random variable x has the probability density function $f(x, \theta)$ and a sufficient statistic exists for the unknown parameter θ. Develop a sequential test for the hypothesis $\theta = \theta_0$ against the alternative $\theta = \theta_1$.

Obtain approximate expressions for the OC function and the ASN function for the test you have developed.

Apply the general results obtained above when the random variable x has the density function

$$f(x, \sigma) = \frac{1}{\sigma} e^{-\frac{x}{\sigma}} \qquad (0 \leqslant x < \infty)$$

6. Given two populations π_1 and π_2 characterized by the probability density functions

$$f(x, \theta_1) \text{ and } f(x, \theta_2)$$

respectively, which are known except for the values θ_1 and θ_2 of the unknown parameter θ. Derive a sequential procedure to test the hypothesis $\theta_1 \leqslant \theta_2$ against the alternative $\theta_1 > \theta_2$.

Obtain the power of the proposed test and the average number of pairs of observations required to reach a decision.

Use the above results to obtain the sequential test for the problem of discriminating between variances of two normal populations when the means are unknown.

7. Prove that the zero-sum two persons game is strictly determined when the players have finite number of pure strategies and are using mixed strategies.

Explain the bearing of the theory of zero-sum two persons game on Wald's Decision Theory.

Sketch the method of proof of the completeness of the class of Bayes's Decision Rules.

8. If ν is the population median, obtain a distribution—free test for testing $\nu = \nu_0$ against alternatives $\nu > \nu_0$

What is meant by the property of consistency of a non-parametric test? Describe the Wald-Wolfowitz Run Test for testing the hypothesis that two given samples have come from the same population and show that the test is consistent.

PAPER VIII AND IX—STATISTICAL QUALITY CONTROL (PRACTICAL)

Time : 4 *Hours* *Full Marks :* 100

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

1. Below is given $\bar{X}$, R for 20 rational sub-groups each of size $n = 5$, from a production process in the order of manufacture.

[$\bar{X}$ is expressed in units of 0.0001 in. in excess of 0.4000 in. R is expressed in units of 0.0001 in.]

$\bar{X}$	34.0	31.6	30.8	33.0	35.0	32.2	33.0	32.6
R	4	4	2	3	5	2	5	13

$\bar{X}$	33.8	37.8	35.8	38.4	34.0	35.0	33.8	31.6
R	19	6	4	4	14	4	7	5

$\bar{X}$	33.0	28.2	31.8	35.6
R	5	3	9	6

(i) Check these for control on $\bar{X}-R$ charts.

(ii) Can you estimate the inherent variation of the process ? If so, give that estimate.

(iii) Comment on the prospects of this process meeting the specification limits of $0.4037 \pm .0013$ ins.
What further remarks would you offer in this case ?

2. A Double Sampling Plan is given by :—

$N = 1000$
$n_1 = 50$
$c_1 = 0$
$n_2 = 100$
$c_2 = 2$

(i) Draw the *OC*-curve, calculating at least 7-suitable points (Justify any approximations that may be used).

(ii) Under an acceptance/rectification scheme where rejected lots are detailed and all defectives found are replaced by good ones, obtain the A.O.Q.L.

(iii) Give the A.Q.L. and L.T.P.D. of this plan, if the producer's risk (α) and consumer's risk (β) are respectively = .05 and .10.

3. The measurements of the diameters of a part produced on each of six spindles of an automatic screw machine are given below in units of 0.0001 in. in excess of 0.4000 in.

Subgroup number	Spindle Number	Diameter a	Diameter b	Subgroup number	Spindle Number	Diameter a	Diameter b
1	1	55	67	4	1	50	61
1	2	44	59	4	2	57	50
1	3	57	43	4	3	48	47
1	4	48	40	4	4	45	36
1	5	57	45	4	5	55	44
1	6	37	51	4	6	48	65
2	1	55	56	5	1	45	60
2	2	59	59	5	2	46	35
2	3	52	58	5	3	57	52
2	4	46	45	5	4	69	42
2	5	48	62	5	5	72	64
2	6	64	52	5	6	48	69
3	1	52	61	6	1	52	45
3	2	63	55	6	2	59	47
3	3	50	53	6	3	39	65
3	4	54	53	6	4	65	60
3	5	51	53	6	5	50	54
3	6	56	60	6	6	60	51

Analyse the above data on a group control chart and write a report on your findings.

PAPER VIII AND IX—ECONOMIC STATISTICS (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 following table gives the disposable income in billions of dollars and the consumer price index in USA from 1923 to 1952.

Year	disposable income (billions of dollars)	consumer price index (1947–49 = 100)	Year	disposable income (billions of dollars)	consumer price index (1947–49 = 100)
1923	68.0	72.0	1938	65.5	60.3
1924	69.9	73.1	1939	70.2	59.4
1925	73.7	75.0	1940	75.7	59.9
1926	76.1	75.6	1941	92.0	62.9
1927	76.3	74.2	1942	116.7	69.7
1928	78.1	73.3	1943	132.4	74.0
1929	82.5	73.3	1944	147.0	75.2
1930	73.7	71.4	1945	151.1	76.9
1931	63.0	65.0	1946	158.9	83.4
1932	47.8	58.4	1947	169.5	95.5
1933	45.2	55.3	1948	188.4	102.8
1934	51.6	57.2	1949	187.2	101.8
1935	58.0	58.7	1950	205.8	102.8
1936	66.1	59.3	1951	225.0	111.0
1937	71.1	61.4	1952	235.0	113.5

Analyse the important movments in the real income of USA and write a note on these movements.

2. The table below gives monthly per capita expenditure in rupees on some groups of items, of urban families in India in 1951, calssified by monthly per capita total expenditure.

Monthly per capita total expenditure (Rs.)	No. of households	monthly per capita expenditure (Rs.) on food grains	other food items	non-food items	total
0– 7	15	3.22	1.26	1.83	6.31
8–10	22	4.51	2.48	2.84	9.83
11–12	24	4.94	3.58	3.64	12.16
13–14	15	5.70	4.16	4.02	13.94
15–17	36	5.58	5.48	5.05	16.11
18–20	25	6.46	6.90	6.33	19.75
21–23	25	5.96	7.93	8.79	22.68
24–27	35	7.30	8.73	9.64	25.67
28–33	29	7.53	10.26	13.32	31.11
34–42	37	7.90	13.85	17.24	38.99
43–54	30	10.36	16.00	23.26	49.71
55 and above	63	0.39	26.01	69.69	105.09

Obtain the income elasticities of these groups of items, interprete the significance of these elasticities, and indicate some of the ways of using them.

Paper VIII and IX—Sample Surveys Theory (Practical)

Time : 4 *Hours* . *Full Marks :* 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is permitted.

1. The data given below are the number of seedlings for each foot of bed in a bed 120 feet long.

serial no. (in ft.)	no. of seedlings	serial no. (in ft.)	no. of seedlings	serial no. (in ft.)	no. of seedlings
1	8	41	26	81	31
2	6	42	26	82	23
3	6	43	10	83	41
4	23	44	41	84	18
5	25	45	30	85	15
6	16	46	55	86	21
7	28	47	34	87	8
8	21	48	56	88	22
9	22	49	39	89	11
10	18	50	41	90	3
11	26	51	27	91	4
12	28	52	20	92	5
13	11	53	25	93	11
14	16	54	39	94	9
15	7	55	24	95	25
16	22	56	25	96	16
17	44	57	18	97	13
18	26	58	44	98	22
19	31	59	55	99	18
20	26	60	39	100	9
21	20	61	34	101	24
22	19	62	21	102	19
23	25	63	27	103	28
24	11	64	25	104	18
25	31	65	32	105	29
26	26	66	43	106	24
27	29	67	33	107	33
28	19	68	45	108	37
29	17	69	23	109	32
30	28	70	27	110	26
31	16	71	37	111	36
32	9	72	14	112	20
33	22	73	14	113	43
34	26	74	24	114	27
35	17	75	18	115	20
36	39	76	17	116	21
37	21	77	14	117	18
38	14	78	38	118	19
39	40	79	36	119	24
40	30	80	29	120	30

Find the variance of the mean of a systematic sample consisting of every 20th foot. Compare this with the variance for (i) a simple random sample, (ii) a stratified random sample with 2 units per stratum, each stratum being 40 foot long, and (iii) a stratified random sample with 1 unit per stratum, each stratum being 20 feet long. (The total sample size in all the cases will be 6). (70)

2. Try either (a) or (b)

(a) For the sample of 30 households, the following table shows the number of persons in a household and the number of persons who visited a dentist.

family no. (serial)	1	2	3	4	5	6	7	8	9	10
no. of persons	5	6	3	3	2	3	3	4	4	4
no. visiting dentist	1	0	1	2	0	0	1	1	1	0
family no. (serial)	11	12	13	14	15	16	17	18	19	20
no. of persons	3	2	7	4	3	5	4	4	3	3
no. visitin dentist	1	0	2	1	0	1	4	1	1	0
family no. serial)	21	22	23	24	25	26	27	28	29	30
no. of persons	4	3	3	1	2	4	3	4	2	4
no. visiting dentist	1	0	1	0	0	0	1	1	0	0

Estimate the variance of the proportion of persons who saw a dentist and compare this with the binomial estimate of the variance.

(b) A population consists of 6 units, with the following values of y and x.

unit no.	1	2	3	4	5	6
y	2	4	5	8	10	12
x	0	1	3	4	5	6

By working out all possible cases, compare the precisions of the ratio and linear regression estimates for simple random sample of size 2. (30)

Paper VIII and IX—Sample Surveys, 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. In the absence of village (mauza) maps (Cadastral survey maps) you are requested to furnish estimates of the acreages under different crops of one district of your State through a sample survey of families. For the purpose of such a survey, draw up a simple schedule and write out instructions for the field staff. The instructions shall have to be divided into two parts, (i) General instructions, and (ii) Detailed instructions for the schedule. The general instructions should contain a note on the limitations, if any, of such a survey with reference to the purpose in view.

2. In connection with the construction of parity Index of the receipts and expenditure of the cultivating families of a district you are required to estimate the proportions of cost under different items of Agricultural expenses, such as, on Labour, Seeds, etc. from a survey of the families in the rural areas of the district. Draw up a schedule, as complete as possible, for the purpose.

3. Prepare a budget estimate of the cost of field work for the survey as indicated in question no. 1 above. The following particulars may be utilised for the purpose of calculations :—

(i) The district may be supposed to have 20 police stations.

(ii) The average number of villages (mauzas) of a police station may be taken as 150.

(iii) The average number of families of a village (mauza) may be taken as 125.

(iv) The content of the schdule may be taken as such that a primary worker (primary investigator) will be able to enumerate 8 families per working day.

(v) Consolidated pay may be provided for the primary workers as well as the supervisory staff.

4. The following table shows the measured volumes of timber of conifer on 25 systematically located plots of 1/10 acre and eye estimates of the volumes per 1/10 acre in the stands in which they occured. The total area of the conifer stands was 5124 acres and the total volume of timber, from eye estimates of all these stands, was 6,110,000 cu. ft. Obtain an unbiased estimate of the total volume from the data by regression method. What will be the difference of this estimate from the estimate obtained from the mean of the measured volumes on the sample plots ?

Measured volumes, y, on 25 sample plots, and eye-estimates, x, of correspoding stands (cu. ft. per 1/10 acre).

y	x	y	x
170	102	112	128
47	14	153	79
64	57	216	177
91	70	125	65
126	95	100	196
146	92	287	167
87	110	261	208
105	208	169	152
255	208	182	152
135	110	74	148
146	110	24	207
154	110	255	167
110	110		

25

Time : 4 *Hours* *Full Marks* : 16

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

1. The following table gives the layout plan and mean height in inches of one year old plants in a 2^3 factorial experiments conducted using a split-plot design in two replications, each main plot being assigned one of the 4 combinations of two levels ((l_0 and l_1) of lime and of two levels (h_0 and h_1) of Hamus and the 8 sub-plots within a main plot being assigned all combinations of two levels each of N, P and K.

Analyse the data confining attention to only the main effects and two-factor interactions so far as the treatment effects are concerned and find out which of them are significant. Write a report on the results.

	l_0h_0				l_0h_1				
	npk 27.3	*n* 27.1	*p* 27.6	*k* 26.5	*p* 24.6	*k* 28.7	*n* 30.4	*npk* 33.6	l_0h_1
l_0h_0	(1) 21.1	*nk* 25.0	*np* 29.7	*pk* 29.8	*np* 29.3	*nk* 28.3	*pk* 32.6	(1) 26.7	
	n 25.3	*k* 20.4	*p* 33.9	*npk* 33.1	*pk* 30.5	*np* 29.9	(1) 20.2	*nk* 28.8	
l_1h_1	*np* 32.1	(1) 27.2	*pk* 34.4	*nk* 26.4	*npk* 31.0	*n* 29.1	*p* 28.8	*k* 30.1	l_1h_1
	k 28.3	*p* 28.6	*npk* 39.5	*n* 26.5	(1) 27.3	*np* 34.9	*nk* 31.5	*pk* 34.0	
l_1h_1	*np* 32.5	(1) 29.5	*pk* 35.9	*nk* 30.3	*p* 31.7	*k* 33.6	*n* 31.5	*npk* 34.9	l_1h_0
	n 34.7	*npk* 40.4	*p* 37.2	*k* 32.5	*nk* 30.7	*pk* 38.0	*np* 29.3	(1) 27.1	
l_0h_1	*np* 34.8	*nk* 35.6	(1) 32.8	*pk* 37.0	*n* 31.6	*npk* 33.3	*p* 27.5	*k* 28.4	l_0h_0
	l_0h_1				l_0h_0				

Point out any peculiarities or defects you notice in the design.

2. The following table gives the data of an experiment conducted to study the resistance to insect attack of bamboo treated with 13 different insecticides (including control) numbered serially (1) to (13). Each culm of bamboo was cut into 4 pieces and numbered serially 1 to 4 from bottom to top. 13 culms of bamboo labelled A, B . . . M were used. The 52 pieces were assigned to the treatments according to a Youden square design. The degree of attack was measured in terms of x = number of successful holes bored by the insects.

Analyse the values of $y = \log_{10}(1+x)$. Obtain estimates of adjusted treatment means of y with and without recovery of inter-block information and also their standard errors. Calculate the critical differences at 5 per cent level of probability and prepare the bar diagrams of significance between the various treatment means of adjusted y.

Culm	Piece number							
	1		2		3		4	
	Treat no.	x	Treat no.	x	Treat no.	x	Treat no.	x
A	(11)	0	(10)	0	(6)	0	(13)	50
B	(7)	0	(6)	0	(2)	0	(9)	0
C	(1)	0	(13)	125	(9)	0	(3)	0
D	(12)	0	(11)	0	(7)	0	(1)	0
E	(4)	0	(3)	0	(12)	0	(6)	0
F	(9)	0	(8)	0	(4)	0	(11)	0
G	(6)	0	(5)	0	(1)	0	(8)	0
H	(10)	0	(9)	0	(5)	0	(12)	400
I	(13)	150	(12)	400	(8)	450	(2)	0
J	(3)	60	(2)	0	(11)	150	(5)	0
K	(8)	0	(7)	0	(3)	0	(10)	0
L	(2)	0	(1)	0	(10)	0	(4)	0
M	(5)	0	(4)	0	(13)	250	(7)	473

3. The data tabulated below belong to an experiment conducted using a partially balanced incomplete block design. Find out the values of the parameters of the design and verify that it satisfies all the parametric relations of a p.b.i.b. design.

Block	Treatment numbers and yields			
(1)	15 2.4	9 2.5	1 2.6	13 2.0
(2)	5 2.7	7 2.8	8 2.4	1 2.7
(3)	10 2.6	1 2.8	14 2.4	2 2.2
(4)	15 3.4	11 3.1	2 2.1	3 2.3
(5)	6 4.1	15 3.3	4 3.3	7 2.9
(6)	12 3.4	4 3.2	3 2.8	1 3.0
(7)	12 3.2	14 2.5	15 2.4	8 2.6
(8)	6 2.3	3 2.3	14 2.4	5 2.7

Contd.

Block	Treatment numbers and yields			
(9)	5 2.8	4 2.8	2 2.6	13 2.5
(10)	10 2.5	12 2.7	13 2.8	6 2.6
(11)	9 2.6	7 2.6	10 2.3	3 2.4
(12)	8 2.7	6 2.7	2 2.5	9 2.6
(13)	5 3.0	9 3.6	11 3.2	12 3.2
(14)	7 3.0	13 2.8	14 2.4	11 2.5
(15)	10 2.4	4 2.5	8 3.2	11 3.1

The analysis of variance appropriate to this design has been partially done in the table given below. Complete the table and calculate the adjusted mean yield for each treatment without and with recovery of inter-block information. Also calculate the standard errors of differences between pairs of adjusted means.

Source of variation	Degrees of freedom	Sum of squares	Mean square
Blocks (Unadjusted)	14	4.4923	
Treatments (adjusted)	14		
Intra-block error	31		
Total	59	9.3733	
Treatments (unadjusted)	14	3.1383	
Blocks (adjusted)	14		

PAPER VIII AND IX—GENETICS (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. Brunton observed the following segregation in F_2's for alcurone colour and pale green seedling in maize. The factor for seedling colour is known to be linked with one of the two complementary factors for alcurone colour. C and R denote the complementary factors for alcurone colour and P is a factor for pale green seedling :

	CR	Cr+cR+cr
P	1907	1053
P	300	686

Estimate the recombination fraction (assuming equal crossing over in male and female) by the maximum likelihood method and obtain its standard error (The linked factors were in coupling phase).

2. An F_2 family of sweet pea showed the following segregation for the two genes purple-red flower colour and long-round pollen shape, indicating linkage in coupling phase :

Purple long	*Purple round*	*Red long*	*Red round*
296	19	27	85

An F_2 also indicated coupling linkage :—

Purple long	*Purple round*	*Red long*	*Red round*
583	26	24	170

Do the data agree in showing the same recombination fraction for the two genes ?

3. Haldane gives data on 28 families on the segregation of the recessive achromatopsia :

		Family number									
		1	2	3	4	5	6	7	8	9	10
Male	Normal	1	1	2	2	1	2	0	3	2	0
	Affected	3	0	1	2	0	2	1	2	1	6
Female	Normal	4	2	3	3	1	1	1	1	0	2
	Affected	0	2	1	0	2	0	1	1	0	0

		Family number									
		11	12	13	14	15	16	17	18	19	20
Male	Normal	2	1	1	1	0	2	0	2	0	3
	Affected	2	2	0	3	0	3	1	1	0	0
Female	Normal	1	0	3	0	0	1	0	0	0	2
	Affected	4	1	1	0	2	0	1	4	2	1

		Family number							
		21	22	23	24	25	26	27	28
Male	Normal	0	0	0	2	0	1	0	5
	Affected	1	2	0	2	0	2	2	1
Female	Normal	2	1	2	1	0	0	4	3
	Affected	0	0	1	1	2	0	1	0

Test for sex linkage of achromatopsia and estimate the crossover value, (a) assuming complete ascertainment and (b) by employing a method independent of the method of ascertainomnt.

4. The following data on blood group frequencies were obtained from an examination of 6000 Chinese :—

	Blood Group			
	AB	A	B	O
No.	607	1920	1627	1846

Assuming Bernstein's explanation of three alleles of a single factor, obtain the maximal likelihood estimates of the gene frequencies and their standard error.

5. In a varietal trial on wheat carried out in randomised blocks, in addition to grain yield the number of ears per plant, the grain number per ear, the wieght per grain and the wieght of straw were observed. The following tables give the estimates of phenotypic (t_{ij}) and genotypic (g_{ij}) variances and covariances of the characters. Fit a discriminant function for selection for improvement of the varieties with respect to yield of grain.

x_1 = logarithm of yield of grain per plant

x_2 = logarithm of ear number per plant

x_3 = logarithm of average number of grains per ear

x_4 = logarithm of average weight per grain

x_5 = logarithm of weight of straw per plant

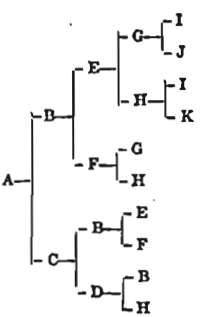
TABLE 1 : $10^6 \times t_{ij}$ for means of 6 plots.

i	2	3	4	5
2	3311	—1810	—1391	385
3		1535	1299	216
4			2087	488
5				859

TABLE 2 : $10^6 \times g_{ij}$

i	2	3	4	5
1	—48	946	1093	013
2	3086	—1743	—1392	252
3		1363	1326	171
4			2059	490
5				654

6. (a) Given the following pedigree, compute how inbred are E, B, D, C and A. Also find how closely related are I and B to A.



(b) How intense is the inbreeding of animals inbred full brother and sister for 3 generations. ?

How closely related are brother ans sister then ?

Paper VIII and IX—Mathematical Theory of Sampling Distribution (Practical)

Time : 4 *Hours* *Full Marks :* 1000

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

1. Take 20 random samples of 6 observations each from a normal population and examine whether their vaiances are in accordance with those from a normal population by the following methods :—

(a) Bartlett's method

(b) By examining the proportions in three equal groups

2. Taking any distribution other than normal, show by Monte Carlo's method that the usual t-distribution can be applied without any adjustments. Thirty samples may be used to prove your statement.

3. Draw a random sample of 10 observations from a trivariate normal population whose parameters are as follows :—

$$\mu_x = 10,\ \mu_y = 15,\ \mu_z = 30$$
$$\sigma_x = \sigma_y = \sigma_z = 1$$
$$\rho_{xy} = .9,\ \rho_{xz} = .8 \text{ and } \rho_{yz} = .7$$

Test whether the sample means are in statistical agreement with the values of the means in the population assuming the variances and covariances in the population to be (i) known and (ii) not known.

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION—SEPTEMBER 1957

PAPER I—THEORETICAL STATISTICS (GENERAL)

Group B only

Time : 2 *Hours* *Full Marks* : 40

(a) Attempt any TWO questions from this group.
(b) Figures in the margin indicate full marks.
(c) Use of calculating machines is not permitted.

1. (a) Give some situations where the χ^2 test is useful.

(b) If χ_1^2 and χ_2^2 are two independent χ^2 vairables with n_1 and n_2 degrees of freedom respectively, show that

$$u = \frac{\chi_1^2}{\chi_2^2} \text{ and } v = \chi_1^2 + \chi_2^2 \text{ are also independently distributed.}$$

Hence obtain the distribution of

$$u = \frac{\chi_1^2}{\chi_2^2}. \qquad (20)$$

2. (a) What are the assumption underlying the analysis of variance test used in the analysis of a randomised block design?

Give some methods of taking care of any violation of these assumptions.

(b) If y_i $(i = 1,2, \ldots, n)$ is normally distributed with mean $\alpha + \beta x_i$ and variance σ^2, obtain the best estimates of α and β and devise the test for $\beta = 0$ when x_i's are fixed. (20)

3. (a) If x_{ij} $(i = 1,2\,;\, j = 1, 2, \ldots n_i)$ are random samples of sizes n_i $(i = 1, 2)$ from normal populations with means μ_i $(i = 1, 2,)$ and variances σ^2, show how you will test the hypothesis $\mu_1 = \mu_2$.

(b) In a feeding experiment on rats the folowing results were obtained :—

Diet	*Gain in grams*						
High protein	13, 8,	14, 11,	10, 12,	11, 9,	12, 12	16,	10
Low protein	7,	11,	10,	8,	10,	13,	9

Investigate if there is any evidence of the superiority of one diet over the other.

The 5 per cent values of t for different values of degrees of freedom are :—

d.f.	16	17	18	19
value of t	1.75	1.74	1.73	1.73

(20)

Note : Question paper for Group A of this paper is not available for circulation.

PAPER II—APPLIED STATISTICS (GENERAL)

Time : 4 *Hours* *Full Marks* : 10

(a) Answers to the different groups are to be given in separate book
(b) All questions carry equal marks.
(c) Attempt any THREE question from *each group*.

Group A

1. It is desired to conduct a sample survey to ascertain the extent of rur indebtedness in your State. Describe what kind of sampling design and questionnair you would suggest. Also give blank specimens of three tables which you would consid most essential for presentation of the statistical results of the survey.

2. Give a brief account of the activites of the Directorate of National Samp Survey of the Government of India.

How are these activities essential for policy making ?

3. What publications of the Government of India would you look up to fin information on the following ? Indicate briefly the main features of the informatio availbale (on the items) and give your comments on its quality :

(1) Agricultural production
(2) Wholesale prices
(3) Employment
(4) Annual balance of payments
(5) Plan expenditure

4. Describe the various methods of measuring seasonal variation from tim series data.

Explain the relative merits and demerits of each of these methods.

5. Describe a method that you recommend for computation of national incor of India with the available data. Show the details of items that you would inclu for computation of national income.

Group B

6. What is a 'balanced incomplete block design ?' What is the use of such desig in agricultural experiments ? Indicate the structure of analysis of variance for su a design and mention the important steps in the calculations involved.

7. What is meant by 'confounding ?' Explain its advantages in factorial expe ments. Derive a confounded design for five factors at two levels each in blocks size 4.

8. (a) Explain what is meant by (i) linkage, (ii) segregation. (iii) independe assortment.

(b) In a back-cross involving two factors, suppose one parent is doub heterozygous Aa Bb and the other is doubly recessive aa bb. How does each of t two factors segregate according to Mendelian expectation in this case and wh statistical test do you propose to apply to detect linkage ?

9. What is a 'logistic curve' and what are its uses ? Illustrate the uses by describing the procedure and steps you adopt with reference to population data for India.

10. (a) Write short notes on :—

(i) Two factor theory, and

(ii) Multiple factor theory in a connected account.

(b) Explain how you will combine the ranks of a number of subjects by several judges.

PAPER III—STATISTICAL INFERENCE

Time : 4 *Hours* *Full Marks :* 100

(a) Answers to the different groups are to be given in separate books.

(b) All questions carry equal marks.

Group A

(Answer any TWO questions)

1. What is a maximum likelihood estimate ?

Suppose that length of life (in days) of a certain mechanical part is known to be distributed according to the probability density

$$(1/\theta)e^{-(x/\theta)}, \qquad 0 < x < \infty,$$

where θ is an unknown parameter, $0 < \theta < \infty$. Suppose that in a given case ten parts were installed at the same time. It was learnt one year later that seven of these parts were still in service, while three had failed, after 62, 70 and 238 days of service respectively. What is the maximum likelihood estimate of θ ?

2. Define a sufficient statistic.

Show that corresponding to any unbiased estimate of a parameter there exists an unbiased estimate based only on the sufficient statistic such that the variance of the latter estimate never exceeds the variance of the given estimate.

Explain briefly the utility of the above result.

3. Suppose that in a certain city there are N passenger cars, with serial license numbers 1, 2, . . N. A visitor to the city notices that the license numbers of the first n cars he has seen are $x_1, x_2, \ldots, x_n$ say. How should he estimate N from this data ?

Explain the basis of the method you suggest.

Group B

(Answer any THREE questions)

4. How are the classical and studentised D^2 statistics defined ? State clearly the assumptions and hypotheses underlying the test.

Describe different practical situations where the above statistics may be effectively used.

5. What is a 'Likelihood ratio test'? A sample of size n is drawn from each of k normal populations with the same variance. Derive the likelihood ratio criterion for testing the hypothesis that the means are all zero. Show that the criterion is a function of a ratio which has the F distribution.

6. What is a power curve? How are power curves useful in comparing tests? How will you construct power curve of F statistic for testing the equality of variance of two normal populations?

7. You are asked to build up a reliable linear formula of prediction of performance (y) of workers in a factory using their scores in as small a number of tests as possible. Observations are available on y and scores in a large number of tests. Describe the test procedure you will adopt to select the set of tests required to build up the formula.

Paper VI—Practical

Time : 6 *Hours* *Full Marks :* 100

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

1. (a) The following table gives the values of a function $f(x)$ for $x = 5.0, 5.1, \ldots, 5.6$.

(i) Find the value of $f(x)$ for $x = 5.34$.
(ii) Also determine the value of x for which $f(x) = 3.38$.

x	$f(x)$
5.0	3.3551
5.1	3.3648
5.2	3.3742
5.3	3.3836
5.4	3.3928
5.5	3.4018
5.6	3.4107

(9)

(b) Using a suitable quadrature formula, evaluate

$$\int_1^3 x^2 \log_e x \, dx$$

given the following values :—

x	$\log_e x$
0.1	—2.30259
0.2	—1.60944
0.3	—1.20397
0.4	—0.91629
0.5	—0.69315

(6)

and $\log_e 10 = 2.30259$

2. Fit a curve of the form $y = ar^x$ to the following data :—

x	0	2	4	6	8
y	114.5	130.1	150.1	171.0	196.7
x	10	12	14		
y	225.2	257.9	293.2		

(8)

3.(a) The frequency distribution given below relates to 500 experimental observations made on the specific gravity of a liquid :

Below .800	7	
.800—.801	35	
.801—.802	82	mean = .803
.802—.803	122	
.803—.804	126	standard
.804—.805	83	deviation = .0015
.805—.806	31	
.806—.807	10	
Above .807	4	
	500	

Test the hypothesis that the errors of observation follow a normal law. (12)

(b) In an epidemic of a certain disease, 927 children contracted the disease. Of these, 408 received no treatment and of them 105 suffered from after-effects. Of the remainder who did receive treatment, 166 suffered from after-effects. Test the hypothesis that the treatment was not effective and comment about the conclusion. (5)

4. (a) The coefficient of correlation between the harvest prices of rice and jowar was found to be 0.67 on the basis of 60 pairs of price quotations in a certain year. In the following year the coefficient of correlation was found to be 0.58 based on 70 pairs of quotations. Test for the significance of the difference between the correlation coefficients. (5)

(b) A survey of 250 sample tribal households containing 1406 persons in a certain Agency tracts in an Indian State showed that there are 1036 females to 1000 males. In the whole State according to the 1951 census, there are 986 females to 1000 males. Apply an appropriate test to find out whether the tribal people in that Agency tract differ significantly in their sex composition from the people of the State. (5)

5. The mean and standard deviation of earning per month per head in rupees estimated from samples of workers in two factories are given below :—

$n_1 = 13$ $\quad \bar{x}_1 = 27.92$ $\quad s_1 = 0.073$

$n_2 = 9$ $\quad \bar{x}_2 = 32.13$ $\quad s_2 = 1.143$

Test the suggestion that the average rate of earning differs between the two factories. State clearly the assumptions underlying your procedure. (10)

6. The mean, standard deviation and correlation coefficients between three variables estimated from a sample of 28 are given below :

$n = 28$

$\bar{x}_1 = 113.6 \qquad s_1 = 15.2 \qquad r_{12} = 0.240$

$\bar{x}_2 = 90.2 \qquad s_2 = 14.9 \qquad r_{13} = 0.860$

$\bar{x}_3 = 7.5 \qquad s_3 = 2.6 \qquad r_{23} = 0.007$

Note : Values of s are calculated by dividing the corrected sums of squares by $(n-1)$.

(i) Find $r_{13.2}$ and test for its significance.

(ii) Set out the multiple regression equation of x_1 on x_2 and x_3.

(iii) Test for the significance of $R_{1.23}$. (25)

7. In a drought resistance experiment on Aman paddy, the yield in grms. from equally sized plots with different intervals of watering were as shown below. Carry out the analysis of variance and test whether there is any significant difference in yields due to intervals of watering.

Yield of paddy in grams.

Interval of watering		
8	10	12
1.180	0.910	0.910
1.420	1.200	1.900
1.350	1.690	1.480
1.210	2.830	3.810
1.440	1.740	0.750
2.650	1.790	1.110
		2.780

(15)

Paper VII—Practical

Time : 6 *Hours* *Full Marks :* 100

(a) Figures in the margin indicate full marks

(b) Use of calculating machines is permitted.

1. In an experiment on the growth of teak plants after one season, two planting methods A, B and three root lengths *P*. *Q*. *R* were used, these being

A : Pits *P* : 4 in. root

B : Crowbar holes *Q* : 6 in. root

R : 8 in. root

The layout of the experiment in four randomised blocks is shown below, together with mean height in inches of the 50 plants grown on each plot.

Analyse these data in full, and write a report on your conclusions containing tables which summarise the essential features of the data.

Block I	*BR* 16.8	*AR* 24.8	*AQ* 23.6
	BP 12.7	*BQ* 15.4	*AP* 18.2
Block II	*AQ* 22.7	*BR* 17.6	*AR* 23.4
	BQ 16.8	*AP* 16.2	*BP* 11.9
Block III	*BP* 13.4	*AQ* 24.7	*BR* 19.5
	AP 19.4	*BQ* 18.3	*AR* 25.5
Block IV	*BQ* 16.9	*BR* 17.4	*BP* 10.2
	AP 12.3	*AR* 21.8	*AQ* 21.4

(20)

2. An extract from the Life Table of males (all India) from the census of 1931 is given below :—

Age x	20	21	22	23	24	25
Number living at age x	51203	50554	49886	49202	48502	47787
Age x	26	27	28	29	30	
Number living at age x	47057	46310	45543	44751	43931	

Starting with 10,000 persons of age 20, construct a mortality table up to the age 30, using the extract given above. Obtain from it (1) the probability that a person aged 20 will live for 5 years more and (2) the probability that he will die (a) in his 30th year, (b) before his age is 30. (10)

3. (a) An enquiry into the budgets of middle class families in a city gave the following information :—

Expense on	Food	Rent	Clothing	Fuel	Miscellaneous
	35 per cent	15 per cent	20 per cent	10 per cent	20 per cent
Prices (1928)	150/-	30/-	75/-	25/-	40/-
Prices (1929)	145/-	30/-	65/-	23/-	45/-

Determine the change in the cost of living of middle class families from 1928 to 1929. (5)

(b) The following figures show the imports of cotton piece goods into a country in 1913–14, 1929–30, 1930–31 and 1931–32 from another country. Ignoring the changes in exchange rates find the index numbers of (a) quantity (b) value and (c) price of imports using the figures of 1913–14 as base. (15)

year	Quantity in million yards			Value in million pounds sterling		
	grey (bleached)	white (unbleached)	coloured	grey	white	coloured
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1913–14	1534	793	832	17.0	3.5	11.9
1929–30	926	414	483	15.7	10.0	11.4
1930–31	365	272	246	5.2	4.7	5.1
1931–32	249	280	223	2.0	4.0	3.8

4. The following gives the measurements of capacity of containers taken from a production process. Samples of 4 containers each were taken at regular intervals of time during two periods. Specifications were set at 47.0 ± 0.5 c.c.; measurements are recorded as deviations in units of 0.1 c.c. from 46.5 c.c.

Period I

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

Period II

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

(i) Draw control charts for mean and range separately for the two periods. Examine if the process was in control during both the periods.

(ii) Can the same limits be used for both the periods? (25)

5. The following table of intercorrelations of four emotional traits is obtained from a record of 172 normal children aged nine to twelve.

Intercorrelations of four emotional variables

Variable	1	2	3	4
1. Sociability	1.00			
2. Sorrow	0.83	1.00		
3. Tenderness	0.81	0.87	1.00	
4. Joy	0.80	0.62	0.03	1.00

It is expected that there is a general emotionality factor common to all these emotional traits. Find out the loadings of this factor by centroid method or Hotelling's method of principal components. (25)

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. Discuss the basis for $\bar{X}$—R charts. The dimensional tolerances for an assembled product C is specified as $m \pm d$. The components are two separate manufactured

parts X and Y. Explain, step by step, how you would proceed to study 'the natural tolerances' of the assembled product C and compare the production possibilities with the specifications.

2. On what considerations you would base your choice of size of the sample (s) and the frequency of sampling for current control by p—chart.

A production process is in control with process average at 0.03. Obtain the probability of detecting a sudden shift in the process average to 0.06 (on the current control chart for p), if,

(i) 800 units are sampled daily

(ii) 400 units are sampled twice every day—once in the forenoon and once in the afternoon.

Comment on the results.

3. Explain the salient features of the single sampling plan and the double sampling plan by attributes.

Indicate how you would obtain the *OC*—curve and the *A.O.Q.L.* of these sampling plans, and the *ASN* curve of the double sampling plan.

Discuss how these help in the comparison of the two plans.

4. Write a mathematical essay on :—

'The use of variables sampling plans in acceptance inspection for lot per cent defective.'

5. Write short notes on :—

(i) Sequential sampling (ii) Theory of runs
(iii) The c—chart (iv) Modified control limits.
(v) The objectives of statistical quality control

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

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. What is meant by the *design* of an experiment ? What are the purposes served by (a) replication, (b) randomisation, and (c) local control in the design or lay-out of an agricultural field experiment ?

Give some examples of local control generally used in field experimentation.

2. A balanced incomplete block design (v, k, r, b, λ) is modified to include a new set of a treatments by increasing the block size to $k+a$ plots so that each of the new treatments occurs once in each of the b blocks. Show that the resulting design is special case of intra-and inter-group balanced disigns and prove that the variance of difference between effects of any two of the original set of v treatments is

$$\frac{2(k+a)\sigma^2}{v\lambda+ra}$$

where σ^2 is the intra-block variance for blocks of size, $k+a$ plots.

3. In a $n \times n$ Latin square experimemet with wheat, the threshed yields of two adjacent plots inside one of the rows got mixed up and only their total is known. Estimate the individual yield for each of the two mixed-up plots. Derive the expression for the mean variance of all pairs of treatment comparisons and hence the average loss of information. What difference does it make in the loss of information if the two mixed-up plots were treated as missing plots.

4. How many factors each at three levels can be tried in blocks of 9 plots without confounding any main effect or two-factor interaction? Obtain such an arrangement using the maximum number of factors in two replications partially confounding some of the three-factor interactions. Indicate which sets of degrees of freedom have been partially confounded.

5. (a) For a two-associate *p. b. i. b.* design prove that

$$n_1 p^1_{12} = n_2 p^2_{11}$$
$$n_1 p^1_{22} = n_2 p^2_{12}$$

(b) Show that the dual of the unreduced balanced incomplete block design for v treatments in blocks of 2 plots each is a two-replicate *p. b. i. b.* design with two associate classes. Prove that the efficiency factor of the latter is

$$\frac{v(v+1)}{v^2+3v-2}$$

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

(1) Quasi-Latin squares
(2) Uniformity trials
(3) Efficiency of a design
(4) Interaction as error
(5) Recovery of inter-block information

7. The actual design (before randomisation) of a $\frac{1}{4}$ replicate of a factorial experiment involving sex factors :

a, b, c, d, e, f each at two levels, arranged in 4 blocks of 8 plots each, is given below :

Block I	*Block II*	*Block III*	*Block IV*
(1)	*ac*	*ad*	*ae*
ab	*bc*	*bd*	*be*
acde	*de*	*ce*	*cd*
bcde	*abde*	*abce*	*abcd*
acdf	*df*	*cf*	*cdef*
bcdf	*abdf*	*abcf*	*abcdef*
ef	*acef*	*adef*	*af*
abef	*bcef*	*bdef*	*bf*

Determine which of the interactions have been confounded and which main effects and interactions form *aliases* while being estimated from this design.

Assuming three-factor and higher interactions to be non-existent write down the partition of the degrees of freedom in the table of analysis of variance against each ascribable source of variation.

Paper IV and V—Sample Surveys, Theory (Theoretical)

Time : 4 *Hours* *Full Marks :* 100

(a) Attempt any FOUR questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is not permitted.
(d) Treatment should be mathematical wherever possible.

1. (a) Describe the uses of (linear) regression estimates with suitable illustrations. Work out the approximate expression for the bias and the variance of the above noted estimate for the population mean derived from a random sample.

(b) Discuss the different regression estimates of the population mean derived from a stratified random sample.

2. Describe the uses of double sampling for stratification with suitable illustrations. Derive for this case the expression for (i) the variance of the estimate of the population mean of the variate under enquiry, and (ii) an estimate of the above noted variance ; also consider the problem of an approximate optimum allocation of the various sample sizes using a simple cost function. (You may ignore the finite-population-corrections).

3. Write a note on the method of inter-penetrating samples discussing its various uses (e.g. for comparing different methods of enumeration or different teams of enumerators, for estimating the margin of error of the estimates, etc.). Illustrate your answer with suitable examples.

4. (a) Discuss the studies which have been made to describe the variance of the sample units in area sampling as a function of the size of the sample-unit. Discuss in a general manner the problem of determining the optimum shape (including orientation) of the sample units in area sampling.

(b) When the sample unit is a cluster of a number of elements, derive the expression for the variance of the sample units in terms of the intra-cluster correlation.

5. (a) Describe the structure of various types of systematic samples in two dimensions with suitable examples.

(b) Describe any suitable sampling method of estimating the total number of fish in a lake. (Only approximate formulae need be used).

6. (a) In sampling for an attribute that is rare, one method is to continue drawing a random sample until m units which possess the rare attribute have been found, where m is decided in advance. If the finite-population-corrections can be ignored, show that the probability that the total sample required is of size n is

$$\frac{(n-1)!}{(m-1)!\ (n-m)!}\ P^m Q^{n-m}, \qquad (n>m),$$

where P is the proportion of the rare attribute, and $Q = 1-P$. Further, show that the average size of the total sample is $\frac{m}{P}$, and that $p = \frac{m-1}{n-1}$ is an unbiased estimate of P.

(b) Discuss the following statement :—

'It is expected that in a series of random sampling numbers there may occur patches which are not suitable for use by themselves. But such patches should be present in a sufficiently long series, and may be used in conjunction with preceding and/or succeeding portions of the series for drawing a fairly large sample'.

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

Time : 4 *Hours* *Full Marks :* 100

(a) Attempt any THREE questions.

(b) All questions carry equal marks.

1. Explain what is meant by 'bias' in selection as understood in the context of sample surveys. Enumerate a few instances to show how faulty selection may give rise to bias, stating in each case as to how it could be avoided. Mention at least one case when a bias, if it exists at all, will be of little consequence.

2. Write short notes on (i) Systematic sample, (ii) Stratified sample, (iii) Non-response, (iv) Pilot survey, and (v) Sampling fraction.

3. What is a 'sampling frame' ? Give an account of the broad classification of defects to which a sampling frame may be subject, and explain how these defects generally creep in.

4. (a) What are non-sampling errors ? What precautions would you like to take to control them ?

(b) Draw up the plan of an experiment to detect investigator-bias.

5. Give an outline of the main features which you think should be incorporated in a report on a sample survey.

6. You are required to conduct a family budget enquiry in a big city of your State. Discuss in this context briefly the problem of

(a) finding out sources for the frame,

(b) selection of samples,

(c) checking the field work and scrutinising the field records.

7. Describe with reference to any sample survey the important steps which you would like to take to train up the field staff.

8. In connection with the preparation of a scheme for the collection of statistics in your State regarding handloom products and their stock, discuss briefly the following problems :—

(i) the level, that is, whether distributors or producers at which it will be reasonable to collect the information

(ii) whether you would like to recommend complete enumeration or a sample survey

(iii) whether it would be reasonable to propose to furnish fortnightly figures

N.B.—The State may be taken to have about 2 lakhs of handlooms.

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. What are the various types of error to which a cost of living index number is subject ? What methods do you suggest to reduce these various types of error?

2. What is autocorrelation in a time series ? Why is autocorrelation suspected in most of the economic time series ?

Given a time series, how will you test for the existence of autocorrelation ? How will you test the significance of correlation between two autocorrelated time series ?

3. What explaining variables will you consider to obtain the demand function for food in India on the basis of time series data ? How will you determine the form of the function to be fitted and what method will you adopt to obtain the estimates of the parameters ? Give illustrations of some uses that you would like to make of the above demand function.

4. Suppose you are asked to prepare a scheme for payment of children allowance to government employees drawing a monthly pay not exceeding Rs. 500/-. The scheme is to show the details about the allowances to be paid for the first child, for the second child and so on, to different groups of employees with different scales of pay. The object is to relieve the parents of the burden of maintaining children below 14.

What type of data will you require for this purpose ? How will you use the data in preparing the scheme ?

5. Write a critical note on the price statistics of India.

6. (a) Explain the following concepts :—

Gross national product, net national product at market price, net national product at factor cost, personal income, disposable income, domestic income, national income.

(b) How will you compare national income of India for the years 1948–49 to 1954–55, with a view to investigate how far real income has increased ?

7. What is Lorenz curve ? How is concentration of income measured from this curve ? Derive the equation of the Lorenz curve for an income distribution following Pareto's Law and obtain the value of the concentration ratio for such an income distribution.

Paper IV and V—Psychology and Education (Theoretical)

Time : 4 *Hours* *Full Marks :* 100

(a) All questions carry equal marks
(b) Answer any SIX questions

1. Present the equations defining parallel tests which are based on the definition of random error used in test theory. Explain the definition of random error and the meaning of symbols used.

2. Describe the test errors of measurement, substitution, and prediction, giving the equations defining them and their standard deviations. Explain the meaning of symbols used in the equations.

3. Define and compare item difficulty and item discrimination. Give examples of statistical measures used and their interpretation in each case.

4. Describe the method of computation, properties, and interpretation of any two of the following methods of standardizing test scores :—

(a) standard or z scores
(b) percentile ranks
(c) normalised scores
(d) McCall's *T* score

5. Explain how each of the following measures interprets test validity : expectancy tables, validity coefficients and the standard error of estimate. Give formulae where appropriate. A general definition of validity should also be given.

6. Propose an experimental design to measure the effectiveness of the regional language as the medium of instruction at three levels : primary, secondary and university level. Indicate treatments, design, and method of analysis.

7. Describe Spearman's two factor theory of factor analysis and compare it with the multiple factor theory of Thurstone. Since general intelligence tests are based on Spearman's theory, and differential aptitude tests are based on Thurstone's theory, explain how these different theories of factor analysis affect test design and the interpretation of scores on these tests.

8. Give the formula for Spearman's rank correlation method and explain how and why it is used in preference to the product-moment correlation for most psychological and educational data.

PAPER IV AND V—GENETICS (THEORETICAL).

Time : 4 *Hours* *Full Marks :* 100

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

1. Assuming that the maximum likelihood estimator is employed in each case, discuss the information provided by *F* data as compared with the double backcross for estimating the recombination fraction for two linked loci under different situations.

2. 340 families of five children were found to be segregating for what is believed to be a simple autosomal recessive character as follows :—

Number of recessives	Number of recessives ascertained independently				
	1	2	3	4	5
1	140	—	—	—	—
2	80	52	—	—	—
3	35	12	7	—	—
4	4	7	0	2	—
5	0	1	0	0	0

Stating the necessary assumptions and explaining the procedure you adopt, test whether the hypothesis of simple autosomal recessive holds.

3. Discuss the principal methods of estimating the genetic and non-genetic components of variation in the case of characters showing continuous variation in plants and animals.

4. It is desired .to propagate an inbred live while maintaining segregation in respect of four factors each with two alleles and located on distinct chromosomes, the individuals bieng diploid. This is proposed to be done by mating between recessives and heterozysobes for each factor. Find the mean number of individuals which are needed to be raised in order to secure at least one 'eligible' mating (between recessives andheterorzysobes for each factor).

5. Explain Hardy-Weinberg Law of equilibrium. Deduce what happens under random mating in the absence of selection when a factor is sex-linked.

6. Assuming that the average number of offspring per mating is 2 and that the number of offspring is distributed in a Poisson series, obtain the chance of survival of a single mutant gene which is neither beneficial nor harmful to the individuals carrying them.

7. The following table gives the results obtained in a large number of experiments on mice :—

Parents	Number of offsprings	
	yellow	non-yellow
Yellow × non-yellow	2378	2398
Yellow × yellow	2396	1235

Further it was found that the litters born from matings between yellows are smaller by about one-fourth than litters from yellow × non-yellow.

Explain the genetic basis of these results. If a large population of mice is allowed to mate at random, permitting matings only among the members of the same generation, what would be the expected relative frequency of yellows and non-yellows in F_1 to F_3 generations ?

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

(1) Mapping of chromosomes

(2) Blood groups in man

(3) Self-sterility alleles

(4) Segregation in autotetraploids.

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. Explain the term 'true rate of natural increase' and show its relation to the age distribution of the stable population.

2. Discuss briefly the methods adopted by the census actuaries for the construction of Indian life tables.

3. How will you use the availbale migration statistics from census and sample survey data to study the important aspects of migration in relation to population growth.

4. Suggest a scheme for a population survey to study the following demographic aspects of the population, broadly indicating the methods of statistical analysis.

(a) Birth and death rates by socio-economic classes

(b) Trend in fertility over a period of forty years terminating with the date of survey.

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

(a) Application of 'International Statistical classification of causes of death' to Indian mortality data.

(b) Application of morbidity rates to public health problems.

(c) Differential fertility as studied from sample survey data.

Paper VIII and IX—Statistical Quality Control (Practical)

Time : *4 Hours* *Full Marks :* 100

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

1. A sample of 50 pieces is drawn every two hours from the production of a single spindle automatic screw machine and each item is checked by go and no-go gauges for several possible sources of defectives.

The number of defective items found in 20 such successive samples were :—

1, 2, 5, 6, 3, 5, 2, 1, 1, 0, 1, 0, 0, 2, 0, 1, 2, 1, 1, 1.

(i) Analyse the data on a suitable *np*–chart

(ii) Suggest a suitable projection of the *np*–chart for use in the future (Assume assignable causes have been found far out of control points).

(iii) Sketch the *OC*–curve for this projected *np*–chart.

2. The engineering tolerance for a product is 80 units. This is an upper limit. In lot by lot inspection of this product for proportion defective, the *A.Q.L.* and the *L.T.P.D.* are specified respectivley as 0.001 and 0.06 with the producer's and consumer's risks respectively as 0.05 and 0.10.

(i) Design an Acceptance Criterion for single sampling by variables, using the sample mean and the sample standard deviation.

(ii) Devise a single sampling plan by attributes for the same Acceptance Inspection.

(iii) Which of the above two plans you would prefer if the cost of inspection by measurement, per unit, is *twice* that of examination by attributes ?

3. Every hour 5 items are taken as a sample from a production process and each item is measured. The $\bar{X}$ and R for the samples are as below (in certain units) :—

Sample number (In the order of production)	$\bar{X}$	R
1	0.7540	0.0011
2	0.7542	0.0014
3	0.7542	0.0009
4	0.7546	0.0010
5	0.7550	0.0008
	Machine readjusted	
6	0.7539	0.0009
7	0.7541	0.0012
8	0.7543	0.0011
9	0.7547	0.0007
10	0.7549	0.0015
	Machine readjusted	
11	0.7541	0.0017
12	0.7542	0.0010
13	0.7545	0.0011
14	0.7548	0.0009
15	0.7551	0.0012

Specifications are 0.7538 and 0.7553.

(i) Set up an $\bar{X}$-chart with modified limits and an R-chart.

(ii) Make a suitable analysis and write down your findings in the form of a report to the Foreman.

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

Time : 4 *Hours* *Full Marks :* 100

(a) Attempt any TWO questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. An experiment was conducted to test 64 varieties of corn using a 8×8 simple square lattice design. The table below gives the individual plot yoilds, in pounds, of ear corn in the random arragnement of blocks and plots within the blocks.

Analyses the data with recovery of inter-block informaton and write a report on the results.

Replication I

Block								
1	(40)	(33)	(37)	(36)	(34)	(38)	(35)	(39)
	42.9	41.4	48.6	38.0	39.1	39.2	32.9	40.0
2	(56)	(52)	(55)	(49)	(51)	(54)	(50)	(53)
	41.1	42.0	33.1	35.2	38.9	41.9	41.4	34.9
3	(12)	(14)	(9)	(15)	(13)	(10)	(11)	(16)
	30.3	37.7	36.7	39.8	42.5	38.0	43.2	34.4
4	(25	(30)	(20)	(29)	(27)	(32)	(28)	(31)
	35.7	35.3	34.0	35.5	32.0	39.9	34.5	39.9
5	(61)	(59)	(64)	(58)	(63)	(60)	(57)	(62)
	33.4	32.6	38.9	31.7	34.0	37.7	35.8	33.3
6	(42)	(44)	(47)	(43)	(41)	(46)	(48)	(45)
	34.0	29.5	35.2	35.0	37.2	40.2	32.7	40.2
7	(19)	(22)	(17)	(24)	(21)	(18)	(20)	(23)
	31.3	32.0	33.2	29.5	30.5	30.6	30.8	33.1
8	(6)	(8)	(4)	(1)	(3)	(7)	(2)	(5)
	37.7	39.4	32.8	50.5	43.2	41.4	41.7	39.8

Replication 2

1	(25)	(49)	(17)	(9)	(57)	(33)	(1)	(41)
	40.8	37.2	27.0	37.1	37.2	36.4	44.8	38.0
2	(11)	(35)	(51)	(59)	(19)	(43)	(27)	(3)
	44.1	31.4	34.7	39.4	33.2	28.9	33.1	36.0
3	(37)	(53)	(21)	(61)	(45)	(13)	(29)	(5)
	36.8	33.9	29.8	28.5	30.2	33.6	33.0	36.6
4	(14)	(38)	(22)	(62)	(54)	(6)	(46)	(30)
	35.9	37.6	35.5	37.8	34.2	31.9	36.9	30.1
5	(42)	(2)	(50)	(26)	(18)	(58)	(34)	(10)
	38.8	31.8	35.2	33.5	26.9	34.8	32.5	28.4
6	(24)	(40)	(56)	(8)	(48)	(32)	(64)	(16)
	32.7	37.6	32.6	37.0	34.2	33.3	35.7	34.3
7	(47)	(15)	(63)	(31)	(30)	(23)	(55)	(7)
	32.0	33.0	32.0	37.2	36.3	38.0	27.5	39.8
8	(28)	(4)	(52)	(36)	(12)	(20)	(44)	(60)
	34.8	36.3	39.6	33.5	32.5	34.9	31.8	38.4

N.B. : The figures in brackets are treatment numbers.

2. An experiment to study the effect of different levels of different types of nitrogen fertilizers on seed rates of Jowar as given below, was tried in a 3^3 confounded design in one replication with 3 blocks of 9 plots each.

Levels of nitrogen

N_0 —No manure
N_1 —25 lbs. N/acre
N_2 —50 lb. N/acre

Forms of nitrogen

a —Ammonium sulphate (21 per cent N)
n —Ammonium Nitrate (30.6 per cent N)
u —Sodium Nitrate (15.5 per cent N)

Seed rates

S_1 —20 srs/acre
S_2 —30 srs/acre
S_3 —40 srs/acre

Analyse statistically the data of green wieght and dry weight in nearest pounds given in the table below and summarise the results.

Block	*Treatment*	*Green weight*	*Dry weight*
1	N_2uS_2	304	143
	N_1uS_1	320	141
	N_2aS_1	238	110
	N_1aS_2	282	125
	N_0aS_3	340	144
	N_2nS_3	348	143
	N_0uS_2	367	143
	N_0nS_1	291	112
	N_1nS_3	359	152
2	N_0aS_2	301	124
	N_0nS_3	258	120
	N_2aS_3	304	140
	N_1aS_1	276	112
	N_2uS_3	327	131
	N_2nS_1	212	95
	N_1nS_2	318	139
	N_0uS_1	231	107
	N_1uS_3	333	132
3	N_2nS_3	354	136
	N_0nS_2	251	76
	N_1aS_3	261	83
	N_1uS_2	201	84
	N_1nS_1	191	80
	N_0aS_1	145	53
	N_2uS_1	425	150
	N_0uS_3	242	92
	N_2aS_2	302	120

3. Physical properties of samples from 10 bales of particular brand of synthetic rubber were measured in order to determine values of the properties for the lot and to measure the uniformity of the material throughout the lot.

Two compounded batches were prepared from samples of each of the 10 bales. Since there was not sufficient material from any one bale to make more than two tests and since it was not possible to test all the bales in a single day, the following scheme was used.

Allocation of bales

Day	Bale numbers			
1	1	2	3	4
2	1	5	6	7
3	2	5	8	9
4	3	6	8	10
5	4	7	9	10

Show that if 'days' and 'bales' denote 'blocks' and 'treatments' respectively, this is a *p.b.i.b.* design obtained by dualising the balanced incomplete block design

$$v = 5,\ k = 2,\ r = 4,\ b = 10,\ \lambda = 1.$$

The table below gives the values for one of the properties, namely strain at 400 psi. Each value is the median for 3 specimens from a vulcanised sheet.

Day	Strain (percentage) at 400 psi			
1	335	320	313	325
2	316	316	321	327
3	310	305	320	315
4	326	324	337	331
5	321	316	320	317

Analyse the data with recovery of inter-block information and write a report on the results.

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

Time : 4 *Hours* *Full Marks :* 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is permitted.

1. Try *Either* (a) *Or* (b).

(a) The following table shows the number of inhabitants (in 1000's) for two years, 1920 (x_i) and 1930 (y_i), in each of a random sample of 49 cities drawn from a population of 196 cities.

Talbo of 1(a). Sizes of 40 cities (in 1000's) in 1920 (x_i) and 1930 (y_i).

x_i	y_i	x_i	y_i	x_i	y_i
76	80	2	50	243	291
138	143	507	634	87	105
67	67	179	260	30	111
20	50	121	113	71	79
381	464	50	64	256	288
23	48	44	58	43	61
37	63	77	89	25	57
120	115	64	63	94	85
61	69	54	77	43	50
387	450	56	142	298	317
93	104	40	60	36	46
172	183	40	64	161	232
78	108	38	52	74	93
66	86	136	139	45	53
60	57	116	130	36	54
46	65	46	53	50	58
				48	75

The problem is to estimate the total number of inhabitants in the 196 cities in 1930, and to estimate the standard error of the above noted estimate for the total. The true 1920 total for the 196 cities is known to be 22,919 thousands. The following four methods are to be used in this connection :— In method (i) y_i-values are to be used only; in methods (ii), (iii) and (iv) the information regarding x's is also to be utilised by using the ratio method of estimaion (method ii), the regression method of estimation (method iii),, and the modified regression method of estimation (method iv) in which the value of the regression coefficient of y_i on x_i is arbitratily taken to be 1. Work out the different solutions to the problem by using all the four methods, (i), (ii), (iii) and (iv). (70)

(b) In a social sample survey, a sample of n individuals was selected at random out of a population of very large size, and questionnaires were mailed to the sample-individuals. Let n_1 be the number of individuals who responded, and n_2 the (number of non-respondents. (Thus $n_1+n_2=n$). Then, a sub-sample of r_2 was randomly selected out of n_2 non-respondents, and the questionnaires were filled up by personal interviews of these r_2 individuals. Let $n_2 = kr_2$, $(k>1)$. Then a suitable estimate T of the population mean (of some variate) derived from the sample-results was used, the variance $V(T)$ of which, under certain simplifying assumptions, comes out as

$$V(T) = \frac{S^2}{n}\left\{1+(k-1)\,Q\right\}$$

where S^2 is the variance in the population, and $Q = 1-P$, P being the expected value of the proportion of respondents. The precision desired is that which wald be given by a random sample of size 1000 if there were no non-response. The cost of mailing

a questionnaire is $ 0.10, the cost of processing the completed questionnaire is $ 0.40, and the cost of carring out a personal interview is $ 4.10, the other costs being negligible.

Work out the optimum values of n and k and the cost of the optimum survey, for the following three values of $P = 0.2$, 0.5 and 0.8. Further, for the same values of of P work out the cost of an alternative survey in which 1000 questionnaires are mailed, and all the non-respondents (out of 1000) are interviewed. (70)

2. The following table shows the data derived from a stratified random sample of tire dealers, in which N_h and n_h stand for the population size and sample size respectively in the h-th stratum, and $\bar{y}_h$ and s^2_h stand for the mean and the variance respectively of the sample in the h-th stratum. (The variate is the number of new tires).

Table

Stratum No.	N_h	n_h	$\bar{y}_h$	s^2_h
1	19,850	3,000	4.1	34.8
2	3,250	600	13.0	92.2
3	1,007	340	25.0	174.2
4	606	230	38.2	320.4

Estimate the gain in precision by using the stratified random sample relative to ordinary (unstratified) random sampling. Further compare this result with the gain that would have been attained from proportional allocation of the stratified sample. (30)

Paper VIII and IX—Sample Surveys, 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. Draw the proforma of a simple schedule and the necessary instructions for field staff to be engaged on a sample survey of handlooms to furnish statistics of monthly production and stock of handloom products.

2. Draw up a schedule, as complete as possible, for an enquiry into family budgets.

3. Supposing that Rs. 1.5 lakhs have been sanctioned for expenditure for the purpose as indicated in question No. 1 above, prepare within this amount the detailed budget estimate for the above survey including field work and analysis. If instead of monthly reports you are required to furnish quarterly reports, indicate to what extent the expenditure could be saved. Show also in this latter case the picture of the revised budget.

4. The table below shows the number of workers (including absentees), x, belonging to 43 factories from a random sample of 325 factories of a district and the

corresponding number of absentees, y. Estimate the percentage of workers absent from the working force, the number of workers belonging to the working force and the number absent from the working force. Estimate also the sampling error of the estimated percentage absentees :—

x	y	x	y	x	y	x	y
95	18	89	7	75	12	159	36
79	14	57	9	69	16	54	26
30	6	132	26	63	9	69	27
45	3	47	7	83	14	61	2
28	5	43	17	124	25	164	69
142	15	116	24	31	3	132	41
125	18	65	16	96	45	82	10
81	9	103	18	42	25	33	8
43	12	52	16	85	35	86	22
53	4	67	27	91	28	51	19
148	31	64	12	73	13		

Paper VIII and IX—Economic Statistics (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 following table gives the disposable income in billions of dollars and the consumer price index in USA from 1923 to 1952.

Year	disposable income (billions of dollars)	consumer price index (1947–49 = 100)	Year	disposable income (billions of dollars)	consumer price index (1947–49 = 100)
1923	68.9	72.9	1938	65.5	60.3
1924	69.9	73.1	1939	70.2	59.4
1925	73.7	75.0	1940	75.7	59.9
1926	76.1	75.6	1941	92.0	62.9
1927	76.3	74.2	1942	116.7	69.7
1928	78.1	73.3	1943	132.4	74.0
1929	82.5	73.3	1944	147.0	75.2
1930	73.7	71.4	1945	151.1	76.9
1931	63.0	65.0	1946	158.9	83.4
1932	47.8	58.4	1947	169.5	95.5
1933	45.2	55.3	1948	188.4	102.8
1934	51.6	57.2	1949	187.2	101.8
1935	58.0	58.7	1950	205.8	102.8
1936	66.1	59.3	1951	225.0	111.0
1937	71.1	61.4	1952	235.0	113.5

Analyse the important movements in the real income of USA and write a note on these movements.

28. The table below gives monthly per capita expenditure in rupees on some groups of items, of urban families in India in 1951, classified by monthly per capita total expenditure.

Monthly per capita total expenditure (Rs.)	number of households	Monthly per capita expenditure (Rs.) on			
		food grains	other food items	non-food items	total
0– 7	15	3.22	1.26	1.83	6.31
8–10	22	4.51	2.48	2.84	9.83
11–12	24	4.94	3.58	3.64	12.16
13.14	15	5.76	4.16	4.02	13.94
15–17	36	5.58	5.48	5.05	16.11
18–20	25	6.46	6.06	6.33	19.75
21–23	25	5.96	7.03	8.79	22.68
24–27	35	7.30	8.73	9.64	25.67
28–33	29	7.53	10.26	13.32	31.11
34–42	37	7.90	13.85	17.24	38.99
43–54	30	10.36	16.09	23.26	49.71
55 and above	63	9.30	26.01	69.69	105.09

Obtain the income elasticities of these groups of items, interprete the significance of these elasticities, and indicate some of the ways of using them.

PAPER VIII AND IX—PSYCOOLOGY AND EDUCATION (PRACTICAL)

Time : 4 *Hours* *Full Marks :* 100

(a) Answer any THREE questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. By the use of the method of tetrad differences, determine whether a single common factor exists for the correlation matrix. Calculate all required sets of tetrad differences and write the conclusion on the basis of the results.

	Test				
	a	*b*	*c*	*d*	*e*
a	–	.50	.27	−.11	.48
b	.50	–	.48	.13	.46
c	.27	.48	–	.28	.20
d	−.11	.13	.28	–	−.08
e	.48	.46	.20	−.08	–

2. An orthogonal factor matrix, consisting of two factors, is given below. Write two paragraphs, one interpreting Factor I and the other interpreting Factor II, on the basis of the factor loadings and the descriptions of the tests given below.

Test	Factor I	Factor II
1	.39	.77
2	.47	.74
3	.77	.29
4	.44	.63
5	.43	.67
6	.71	.31
7	.05	.52
8	.11	.44
9	.64	.00
10	.69	.30
11	.80	.20

Test	Test description
1	English Usage I. A test of English grammar. The task is to locate grammatical errors in sentences. The score is the number of errors correctly located.
2	English Usage II. A test of English grammar designed to be parallel to English Usage I, and scored in the same way.
3	Matching. A test in which pairs of words or numbers are given. The task is to check whether the two members of each pair are identical or different. The score is the number of pairs correctly identified as identical or different.
4	Sentence classification. Six topic headings are given, e.g. educational, political. The task is to classify different sentences under the various headings. The score is the number of sentences correctly classified.
5	Following directions. Simple instructions based on common knowledge are given. The score is the number of instructions followed correctly.
6	Perception. In this test, four alternatives follow each item. Each item consists of a word or series of numbers. The task is to determine which alternative is identical with the original item. The score is the number of alternatives correctly identified.
7	Dexterity. In this test the task is to make tick marks as fast as possible. The score is the number of tick marks made in one minute.
8	General knowledge. A test of general information about geography and political situations. The score is the number of right answers.
9	Number Series. Arithmetic and geometric progressions are given, with some members of the series missing. The task is to fill in the missing members. The score is the number of members correctly filled in.
10	Greater and lesser numbers. Sets of four numbers are given. The task is to check the largest number and the smallest number in each set. The score is the number of largest and smallest numbers correctly identified.

11 Simple arithmetic, A test of simple additions, subtractions, multiplications and divisions. The score is the number of correct answers.

3. A selection board consisting of 4 members interviewed 13 candidates and rated each candidate on 5 characteristics. The resulting data can be interpreted as a three-way factorial design without replication with variations over interviewers, traits, and candidates.

Sums of scores and sums of squares necessary for computing all variance estimates are given below. Calculate the variance estimates for all summary tables and compute the F ratios.

Candidate no.	$\underset{it}{\Sigma X}$	$\underset{it}{\Sigma X^2}$	where	i = interviewers 1, 2, 3, 4 c = candidates 1, 2, ... 13 t = traits 1, 2, ... 5
1	75	291		
2	58.5	181.25	Notation:	
3	79.5	325.25	$\underset{it}{\Sigma}$:	summing over interviewers and traits for each candidate
4	62	198		
5	44	102		
6	60	192		
7	80.5	332.25		
8	39.5	87.25		
9	69	245		
10	68.5	240.25		
11	73	281		
12	50	138		
13	52.4	143.66		

Trait no.	$\underset{ic}{\Sigma X}$	$\underset{ic}{\Sigma X^2}$	$\underset{ic}{\Sigma}$	summing over interviewer and candidates for each trait
1	164.5	550.25		
2	159.5	519.25		
3	130.9	372.41		
4	150	460		
5	207	855		

Interviewer no.	$\underset{ct}{\Sigma X}$	$\underset{ct}{\Sigma X^2}$	$\underset{ct}{\Sigma}$:	summing over candidates and traits for each interviewer
1	188.4	591.66		
2	202	688		
3	209.5	733.25		
4	212	744		

Sums of squares of totals:

$\underset{it\ c}{\Sigma(\Sigma X)^2} = 33{,}876.61$ $\quad \underset{c}{\Sigma}$: Σ total over c candidates for fixed i, t

$\underset{ic\ t}{\Sigma(\Sigma X)^2} = 13{,}330.61$ $\quad \underset{t}{\Sigma}$: total over traits t for fixed i, c

$\underset{ct\ i}{\Sigma(\Sigma X)^2} = 10{,}835.71$ $\quad \underset{i}{\Sigma}$: total over i interviewers for fixed c, t

4. The following set of numbers represents scores on Part A and Part B of a t of coding. Parts A and B represent split halves of the test, and were given at the sa time. Calculate the reliability estimate, correcting for length, of the test.

Part		Part		Part	
A	B	A	B	A	B
17	18	16	12	19	20
19	19	17	18	8	15
9	11	17	18	18	20
15	16	18	17	0	5
17	18	17	19	18	18
11	20	13	17	16	16
15	16	18	18	13	18
15	18	20	20	11	12
20	20	19	19	10	10
16	19	17	19	15	18
18	19	17	19	20	20
				19	20