INDIAN STATISTICAL INSTITUTE

QUESTION PAPERS

for

Statistician's Diploma Examinations

March & September 1958

Price Rupee one only

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1958.

PAPER I—THEORETICAL STATISTICS (GENERAL)

Time : 2 Hours — Full marks : 60

GROUP B

(a) Answer *any two* questions.
(b) Figures in the margin indicate full marks.
(c) Use of calculating machines is not permitted.

1. (a) Discuss in some detail the applications and limitations of χ^2 test.

(b) Show that for testing independence in a 2×2 contingency table with cell frequencies as below

a	b
c	d

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}, \text{ where } N = a+b+c+d$$

(c) How do you alter your test for independence given above when the cell frequencies are small. (25)

2. (a) In the multiple regression equation

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}, (i = 1, 2, \ldots n)$$

show how you will estimate $\alpha, \beta_1, \beta_2, \ldots \beta_p$.

(b) How will you test for the homogeneity of a set of correlation coefficients derived from samples of sizes $n_1, n_2, \ldots n_k$ from k bivariate normal populations. (25)

3. (a) Describe a few hypothetical models which give rise to a Poisson distribution. Give suitable examples.

(b) Given two independent observations from two Poisson distributions, how do you test that the parameter has the same value for both the distributions. (25)

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

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) Attempt *any three* questions from group A and *any two* from group B.
(c) All questions carry equal marks.
(d) Use of calculating machines is not permitted.

Group A

1. What is meant by Cost and Variance functions with reference to sample surveys? Indicate how you will use these functions. How will you set about building up these functions with the ultimate object of their use in planning a sample survey for obtaining area cultivated under the major crops in your State? How will you actually use them in planning the survey?

2. For the formulation of an integrated development plan, an economic survey of a district is required to be conducted. Describe what you will recommend in regard to (i) information to be collected and (ii) methods and sources to be used.

3. What are the types of data currently included in the official publications of the Government of India dealing with *either* foreign trade *or* industrial production. Also name the publications.

Describe briefly, in the context of planning, the limitations, if any, of the publications you name and the statistics you describe.

4. Describe briefly how Engel's curves and elasticities of demand of different commodities can be obtained from family budget data. How are they useful for purposes of economic planning?

Can family budget data be used to gauge the disparity in the economic development of two states or regions, and, if so, how?

5. What is meant by seasonal variation in time series? Describe a suitable method of measuring the seasonal variation.

The following are the index numbers of seasonal variation of the production and prices received by farmers for a particular commodity in the post-war period:

month	production	price
January	130	96
February	94	98
March	97	100
April	91	96
May	94	96
June	97	99
July	85	106
August	79	108
September	78	111
October	90	104
November	120	95
December	136	91

Draw line charts of the two series on a graph paper. Examine the charts and say what economic conclusions you reach. Describe some problems where knowledge of the index of seasonal variations of production of the commodity could prove invaluable.

Group B

6. In which situations are split plot designs useful? Explain how split-plot designs may be regarded as a factorial experiment with main effects confounded. Compare the advantages and disadvantages of the design with that of a randomised block design. Explain the mode of analysis of the data obtained by employing the above design with the help of an assumed model.

7. Starting with some fundamental assumptions (to be stated) obtain the commonly accepted law of growth of population. Discuss the different properties of the curve and also the two different methods of fitting the above curve to the growth data, indicating why they have been advocated.

8. State clearly the assumptions behind multiple factor theory. Show that a given correlation matrix can be explained by an infinite number of factor matrices.

Discuss the 'Centroid Method' of factoring. Do you consider that the factors isolated by any standard method of analysis have psychological reality? Give reasons for your opinion and state in what ways you think factor analysis may be useful to psychologists.

9. Describe how Mendel's experiments laid the foundation of the Chromosome theory in genetics.

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) Attempt *any two* questions from group A and *any three* fro group B.
(c) All questions carry equal marks.
(d) Use of calculating machines is not permitted.

Group A

1. Discuss very briefly the suitability of unbiasedness and minimum variance as criteria for estimators.

Show that if u and v are unbiased minimum variance estimators of parametric functions $g(\theta)$ and $h(\theta)$, then $u+v$ is the unbiased minimum variance estimator of $g+h$. Deduce, hence, that the unbiased minimum variance estimator of a given parametric function, if it exists, is unique.

2. Describe the relation between confidence intervals and tests of significance. Show that if a confidence interval is obtained by 'inverting' a family of uniformly most powerful unbiased tests of a parameter, each test being of level α, then the interval is the 'shortest' unbiased interval of confidence coefficient $1-\alpha$.

3. Suppose, that in an analysis of variance you obtain three independent sums of squares, S_1, S_2 and S_3 such that S_i/σ_i^2 has the chi-square distribution with n_i degrees of freedom, $(i = 1, 2, 3)$, where n_1, n_2, n_3 are known constants of the experimental design, and σ_1^2, σ_2^2 and σ_3^2 are known parameters. Suppose also that $\sigma_3^2 = \sigma_1^2 + k\sigma_2^2$, where k is a known constant. How would you estimate σ_3^2?

Group B

4. In the presence of two alternative hypotheses H_1 and H_2, how would you divide the sample-space into two regions of acceptance such that

(i) error of acceptance of one hypothesis be fixed and that of accepting the other rendered minimum,

(ii) the error of acceptance of H_1 and H_2 be equal and this common error be rendered minimum?

Show that in both these cases the best regions of acceptance are bounded by surfaces of a constant value of a function of the minimal set of sufficient statistics.

5. What are similar regions? Under what conditions can one construct similar regions to test a composite hypothesis with one degree of freedom?

Show that for the hypothesis H_0: $\mu = \mu_0$ for a normal population $N(\mu, \sigma)$, a one-sided uniformly most powerful test exists.

6. Describe in short, Fisher's method of discrimination between two groups on the basis of multiple measurements by choosing an optimum linear function of the variables. Show how the statistics D^2 arises out of this procedure.

How can this procedure be represented as a problem in regression by introducing a suitable dummy variable into the system?

7. Write notes on:

(a) Likelihood criterion; (b) Canonical correlation.

Paper VI—Practical

Time: 6 Hours — Full marks: 100

(a) Answer *two* questions from each group.

(b) Figures in the margin indicate full marks.

(c) Use of calculating machines is permitted.

Group A

1. Determine the appropriate type of Pearsonian curve to be fitted to the following data relating to the age of bride at marriage. Calculate the constants involved in the equation of the curve. Draw the histogram and the fitted curve by calculating a suitable number of ordinates on the same graph paper. (30)

age (years)	frequency	age (years)	frequency
15—16	367	30—31	256
16—17	717	31—32	164
17—18	1294	32—33	134
18—19	2121	33—34	94
19—20	3156	34—35	77
20—21	4009	35—36	68
21—22	3593	36—37	59
22—23	3604	37—38	33
23—24	3060	38—39	40
24—25	1774	39—40	27
25—26	1353	40—41	18
26—27	936	41—42	21
27—28	663	42—43	11
28—29	468	43—44	14
29—30	319	44—45	4

Mean age of bride = 22.1877

$\mu_2 = 13.3348, \quad \mu_3 = 67.8145, \quad \mu_4 = 1224.6342.$

$\beta_1 = 1.0390, \quad \beta_2 = 6.8873$

2. The following table gives the mean chest expansion in inches of a number of school boys of ages between 8 and 18, and classified into two groups according as the nutrition is 'good' or 'bad'.

Obtain the lines of regression of chest expansion on age for the two groups separately. Examine the significance of (i) the two regression coefficients, (ii) the difference of the two means of chest expansion, (iii) the difference of the two regression coefficients. (30)

age in years	nutrition 'good'		nutrition 'bad'	
	number of boys	mean chest expansion	number of boys	mean chest expansion
8	4	2.06	16	1.61
9	11	2.02	27	1.64
10	15	2.97	47	1.84
11	10	1.30	35	2.06
12	20	2.19	60	1.84
13	13	2.20	42	2.16
14	17	2.15	74	2.49
15	14	2.66	62	2.57
16	23	2.54	74	2.52
17	9	2.92	52	2.24
18	11	2.27	37	2.46
total	147		526	

3. The following data relate to the land values and crops in 25 States of USA. It is believed that the average valuation for the farmland per acre (x_1) was based on the average yield of corn per acre (x_2) during the previous ten years and on the percentage of farmland under 'small grain' (x_3) and under corn (x_4). Judge how the data confirm this view. Rank the factors in order of their influence on the farmland valuation. (30)

Corrected sums of squares and products.

	x_1	x_2	x_3	x_4
x_1	1205417	195743	109864	164017
x_2		34881	18359	28001
x_3			10418	15242
x_4				23403
Σx_i	5155	927	488	745

Group B

4. (a) 12 animals were fed on diet A and another set of 15 animals on diet B. The gains in weights (in lbs) due to these diets are given below:

Gain in weight, Diet A—25, 30, 28, 34, 24, 25, 13, 32, 24, 30, 31, 35.

Gain in weight, Diet B—44, 34, 22, 8, 47, 31, 40, 35, 32, 35, 18, 31, 35, 29, 22.

Test whether there is any significant difference in the average gain in weights. (12)

(b) Two samples, the first of 20 pairs and the second of 25 pairs, give correlation of .65 and .85 respectively. Are these values significantly different? (8)

5. (a) 100 candidates sat for a certain examination and were numbered from 1 to 100 in alphabetical order, numbers 1 to 50 being in room A and 51 to 100 in room B. Road mending operations were in progress outside the windows of room B and it is feared that they may have affected the candidates in that room. Twenty candidates reached the pass mark in room A and sixteen in room B. What conclusions would you draw and why? (8)

(b) The table below shows the incidence of accidents over different sections of a factory. Do the data indicate any significant difference in the incidence of accidents as far as the various sections are concerned? (12)

sections	number of persons employed	number of accidents	rate per hundred
1	21	11	52.38
2	22	10	45.45
3	18	6	33.33
4	103	20	28.16
5	106	26	24.53
6	21	5	23.81
7	145	33	22.76
8	66	14	21.21
9	153	29	18.95
10	80	8	8.99
11	45	3	6.67
12	35	1	2.68

6. Compute the value of f(x) when $x = 0.475$, given the following tables of values of f(x) for $x = 0.45$ to 0.50. Find for what value of x is f(x) equal to $\frac{1}{2}$. (20)

x	f(x)
0.45	0.475482
0.46	0.484656
0.47	0.493745
0.48	0.502750
0.49	0.511668
0.50	0.520500

Paper VII—Practical

Time : 6 Hours | Full marks : 100

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

1. The table below gives the monthly production of salt in the Indian Union for five years (1952 to 1956) :

Monthly production of salt (1952–1956)
(nearest lakh maunds)

month	1952	1953	1954	1955	1956
January	21	31	19	26	34
February	44	40	37	45	54
March	76	85	70	89	101
April	142	180	126	143	156
May	174	229	179	182	210
June	149	156	161	153	175
July	50	43	63	54	43
August	21	32	17	53	26
September	23	25	17	20	48
October	19	14	23	15	18
November	21	9	12	13	11
December	31	18	15	20	17

(i) Compute the annual average rate of change.

(ii) Find, by the method of 12-month centered moving average, the seasonal indices.

2. The follwoing two sets of 4×4 Latin squares were tried in an experiment on wheat. The object was to study the effect of different manurial treatments. The necessary details and data are given below :

Treatments

A = control, B = Ammonium sulphate
C = Magnesium D = Ammonium sulphate and magnesium.

Yield in lb. per plot

Latin square I				Latin square II			
425	442	540	340	450	359	532	469
B	C	D	A	B	A	D	C
384	512	490	408	501	450	495	432
D	A	B	C	C	D	B	A
506	508	536	600	469	560	470	399
C	D	A	B	A	B	C	D
451	568	499	347	380	412	399	538
A	B	C	D	D	C	A	B

(i) Carry out the analysis of variance of each of the two Latin squares.

(ii) Obtain the combined analysis of variance.

(iii) State your conclusions with the support of appropriate summary table for the various treatment means.

3. A part of the data on plot yields observed in the crop estimation survey on wheat carried out in Rabi 1956–57 in the district of Sehore of former Bhopal State is given below. In each tehsil, a certain number of villages were selected at random with replacement and in each selected village two fields were selected. A plot of $33' \times 16\frac{1}{2}'$ was marked randomly in each field. Estimate the mean yield in lb. per acre of wheat and its sampling error.

Plot yields in chhataks ($\frac{1}{8}$ seers)

tehsil	area under wheat in acres	field number	villages						
			1	2	3	4	5	6	7
1	159	1	74	158	67	90	83	40	47
		2	45	80	14	156	129	8	59
2	76	1	86	99	47	58	79	52	
		2	91	95	51	108	73	48	
3	67	1	44	44	22	65			
		2	42	41	58	72			

4. The following table gives the number of breakdowns in rubber-covered wire in successive lengths of 10,000 feet at a specified test voltage.

length number	number of breakdowns (c)	length number	number of breakdowns (c)
1	1	11	10
2	1	12	5
3	3	13	0
4	7	14	19
5	8	15	16
6	1	16	20
7	2	17	1
8	6	18	6
9	1	19	12
10	1	20	4

(a) What would be the distribution of number of breakdowns per unit length (which in this case is 10,000 feet) if the manufacturing process is in statistical control? State reasons for your answer.

(b) Using the data given to you, obtain suitable control limits for the number of breakdowns per unit length and examine if the process is in control.

(c) Discuss any modification which might be needed in the construction of control limits if the size of the unit varies from sample to sample.

5. You are given below the matrix of intercorrelations of four tests.

tests	1	2	3	4
1	–	0.4	0.4	0.2
2	0.4	–	0.7	0.3
3	0.4	0.7	–	0.3
4	0.2	0.3	0.3	–

Given that there are two common factors, determine the factor loadings of the principal factor.

6. In a two-stage sampling with replacement scheme, the variance of the sample mean is given by

$$V = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{mn}$$

where m is the number of first-stage units selected and n is the number of second-stage units chosen from each of the selected first stage units, and σ_1^2, σ_2^2 are the first stage and second stage variances respectively in the population.

For a certain survey which is proposed to be carried out according to a two stage sampling scheme, the following preliminary information is available :

(a) The cost C of the survey is expected to depend upon m and n in the following way

$$C \text{ (in Rs.)} = 500 + 6m + 1.5mn$$

(b) $\sigma_1^2 = 36.5$ and $\sigma_2^2 = 85.5$ (These are the values as estimated from a pilot survey).

For the same survey, however, the budget sanctioned is only Rs. 2500, and you are not supposed to spend more than this.

Find out the optimum values of m and n which you would recommend in such a case.

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. The following paragraph is taken from Shri C. D. Deshmukh's recent book, Economic Developments in India, 1946–56 :

"The contribution of the statistician to national development is 'through the patient collection and analysis of statistical and technical information and their utilisation for policy and administrative decisions in a scientific manner'. He has to provide the vital quantitative or dimensional element of each question or proposition, both the orders of magnitude of the facts or phenomena under investigation and their potential development through the dimension of time. Thus, statistical data and analysis both

regarding the existing conditions and future trends are an essential ingredient, occasionally the crucial ingredient of policy making. It is, therefore, vital that there should be few occasions when statistical experts themselves differ or disagree among themselves, particularly in their basic approaches: this duly gives the layman excuse to do nothing about a problem or to follow a solution according to his own personal bent. Much of this stultification can only be prevented by establishing statistical centres in the country, organised in an integrated and dovetailing system. Such a system should provide for the planner and the policy-maker both macro-economic statistics at the national level concerning long-term and large-scale movements and fluctuations in overall production, income, employment, population, resources etc. as well as micro-economic statistics which would assist in a better formulation of local problems and plans to deal with them."

Elaborate the ideas considered in the paragraph noting differences, if any.

2. For working out future investment allocation in India, Professor P. C. Mahalanobis used the following figures for the total capital (fixed plus working) per employed person for four broad industrial categories:

1. basic investment goods	Rs. 20,000
2. factory consumer goods	Rs. 8,750
3. household industries and agriculture	Rs. 2,500
4. services	Rs. 3,750

Discuss in detail the different statistical sources which may be used for estimating item (2) above. Also sketch briefly relevant sources for the other items indicating major gaps, if any.

Could data given in the Census of Manufactures be used for the purpose in a straightforward way?

3. Derive expressions for Laspeyres, Paasche and Fisher's 'ideal' index numbers and establish a relation connecting the last index number with the first two.

Describe factor reversal, time reversal and circular tests and show that the 'ideal' index satisfies the first two tests but fails to satisfy the last one.

4. Let Y, S and I stand respectively for national income, saving and investment, the variables being conceived as deviations from means. Also let subscript t denote time, i.e. Y_t gives the value of national income in the t—th year. Consider the equations

$$S_t = Y_t - a\, Y_{t-1}$$

$$I_t = b\, Y_{t-1} + C\, Y_{t-2}$$

$$S_t = I_t$$

and then answer the following questions:

(i) is the system identifiable? What are the conditions of identifiability?

(ii) prove that $Y_t - (a + b)\, Y_{t-1} - C\, Y_{t-2} = 0$. How would you interpret the equation?

(iii) consider the equation $Y_t - 1.6\, Y_{t-1} + 0.6\, Y_{t-2} = 0$.

What does it tell about the movement of national income over time?

5. Briefly explain the following terms:

(i) secular trend, (ii) seasonal and cyclical variations, (iii) auto-correlation, lag correlation and correlogram, (iv) harmonic analysis, and (v) variate difference method.

6. Show how in a closed economy national income can be conceived either as an aggregate of factor payments or as an aggregate of net outputs or as an aggregate of final expenditures. Construct a simple numerical example to illustrate the process. If there are international transactions, how is the above scheme modified?

Briefly answer the following questions:

(i) in national income estimation, should a factor payment be conceived as gross or net of direct taxes?

(ii) explain the differences between (a) factor cost and market price concepts and (b) national and domestic concepts.

7. Write short notes on *any three* of the following:

(i) elasticity of demand, (ii) concentration ratio, (iii) input-output table, (iv) material balances, (v) linear programming.

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. (a) Explain the different situations on a Control Chart pointing to need for action.

(b) State clearly the difference, in theoretical basis, between Control Charts when standards (parameters) are known and Control Charts when standards are not known.

2. (a) Explain the 'Sequential Probability Ratio Test' as applied to acceptance sampling.

(b) Show how the sequential test on the basis of data from two different operations each of which results in either a success or failure (and thereby constituting a double dichotomy), reduces to that for the Binomial distribution.

3. (a) Describe the AQL sampling plans, specially from the point of view of their construction.

(b) Examine the following statements with respect to AQL plans:

(i) 'For reduced inspection change to a lower Sample Size Letter and for tightened inspection to a higher Sample Size Letter'.

(ii) 'With acceptance sampling plans having the same sampling fraction, the larger the lot size the better for the purchaser'.

4. (a) Describe the single sampling plans that are available for industrial use for inspection by measurement.

(b) Show how under certain assumptions the sample size for the known-sigma plan can be shown to be a simple multiple of that for the unknown-sigma plan.

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

(i) Control Charts for individuals, (ii) Continuous sampling inspection plans, (iii) Operating characteristic, (iv) Narrow-limit gauging, (v) Response surfaces.

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) Treatment should be mathematical wherever possible.

1. Write a short critical note on each of the following as understood in the context of sampling studies :

(i) Cluster sampling compared with simple random sampling,

(ii) Inter-penetrating sub-samples.

(iii) Sampling with probability proportional to size.

(iv) Double sampling.

2. Work out, in sampling for proportions, the formula for the required sample size, n, against a given error margin. How would you adjust your formula if the data are continuous? Write a short critical note indicating as to what should be your suggestions for the sample size when information is to be collected on more than one item. Also indicate how you would approach the general problem in this context.

3. Deduce the expression for variance of the estimated mean for a stratified random sample, and show what the form would reduce to when

(i) the sampling fractions are negligible in all the strata,

(ii) proportional allocation is adopted,

(iii) the variances in all the strata have the same value and sampling is proportional.

4. Into a 'lot container' is placed a sample of N balls, which are drawn at random from a supply, of Mp red and Mq white balls. In turn, into a 'sample container' is placed a sample of n balls which are drawn at random from the 'lot container'. Of these n balls, r are found to be red. Work out the variances of r when the N balls are put into the 'lot container' with replacement and the sample is drawn from the 'lot container', (i) with replacement and (ii) without replacement.

5. (a) Show that, with simple random sampling from a finite population, the ratio-estimate is a 'best linear unbiased estimate', if (i) the relation between y and x is a straight line through the origin, and (ii) the variance of y about the line is proportional to x.

(b) Work out the variance of a ratio estimate of the population total in any suitable form.

6. (a) Work out the variance of the mean of a systematic sample and find out, under what conditions (if at all) the mean of a systematic sample will be more precise than the mean of a simple random sample.

(b) Write out a short critical note on the comparison of systematic sample with stratified random sampling.

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. (a) What are pilot surveys and what are their various uses?

(b) What are cost and variance functions and how are they obtained from pilot surveys?

2. What is meant by optimum number of sample units? Illustrate with suitable example, the general principal of determining them. With reference to your example, describe the procedure you would follow in answering the question —' what is the proportionate marginal decrease in variance of the estimate for a marginal (proportionate) increase in cost?'

3. What are sampling and non-sampling errors in a sample survey? What steps do you suggest to reduce them?

4. Data on household indebtedness and trend of self management of agricultural holdings have been collected from a random sample of rural households in India. Describe the steps and procedure you would recommend for the analysis of data with the help of Hollerith, IBM or Power Samas machine equipment. How do you ensure accuracy of work at each stage of processing?

5. Explain the meaning and use of *any four* of the following :

(i) Random sampling with probability proportional to size,
(ii) Systematic sampling with probability proportional to size,
(iii) Ratio and Regression estimates in sample surveys,
(iv) Cluster sampling,
(v) Two-stage sampling,
(vi) Interpenetrating sampling.

6. It is proposed to carry out a series of sample surveys for the estimation of area and yield of principal cereals in India during the next five years. What, in your opinion, is the appropriate ultimate sample unit for this purpose?

Discuss, from the point of view of sampling and non-sampling errors and administrative problems, the merits and demerits of the following :

(i) a fixed sample for all the seasons of all the years with investigators permanently stationed in suitable villages,
(ii) a fixed sample for all the years with change in assignment of samples to investigators in different years,
(iii) independent samples for different crop seasons with investigators permanently stationed in suitable villages,
(iv) independent samples for different crop seasons with change in assignment of samples to investigators,
(v) partial replacement of previous season's samples with change in assignment to investigators.

7. What considerations should weigh with the officer-in-charge of a sample survey in deciding the type of investigators needed for each of the following type of surveys to be carried out in India :

(i) a survey for studies on family planning problems (attitude survey),
(ii) a survey for studies on conditions of household level of living and household indebtedness,
(iii) a survey for estimating the area and yield of principal crops.

What procedure would you adopt for recruiting them?

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

Time : 4 Hours Full marks : 100

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

1. Discuss the advantages and disadvantages of Latin square Designs. Explain how in such a design one can test whether the row classification or column classification or both have led to increased precision in the experiment.

In a given $s \times s$ Latin Square Design with the yield of one plot missing, explain how the experiment is to be analysed and study the efficiency of different treatment comparisons.

2. Explain how split-plot designs may be regarded as factorial experiments with main effects confounded.

Explain the method of analysis appropriate to the case when whole-plot treatments are arranged according to a Latin Square, each plot being divided into the requisite number of split-plots for the split-plot treatments.

3. Explain the principle of confounding. Five factors each at two different levels require to be tested and in order to control heterogeneity, it is proposed that each block should consist of only 8 plots. Show that a balanced arrangement in 5 replications is possible such that the main effects and 2-factor interactions are preserved, and one-fifth of the information on 3-factor and 4-factor interactions is lost. Discuss the appropriate scheme of analysis.

4. Give the theory behind analysis of designs with recovery of inter block information. Illustrate your answer with the analysis of resolvable balanced incomplete block designs having parameters v, b, r, k and λ. Indicate how you can compare such a design with a simple arrangement in r randomised blocks of size v.

5. Write short notes on *any two* of the following :
(i) Rotation experiments,
(ii) Exploration of the optimum in a response surface,
(iii) Uniformity trials,
(iv) Covariance analysis.

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. Show that the joint characteristic function of

$$\theta_1 = \frac{\Sigma x^2}{2(1-\rho^2)\sigma_1^2}, \theta_2 = \frac{\rho \Sigma xy}{2(1-\rho^2)\sigma_1\sigma_2}, \theta_3 = \frac{\Sigma (y^2)}{2(1-\rho^2)\sigma_2^2},$$

in a sample of size n from a bivariate normal population with mean of $x = 0$, mean

of $y = 0$, variance of $x = \sigma_1^2$, variance of $y = \sigma_2^2$ and covariance $(x, y) = \rho\sigma_1\sigma_2$, is given by

$$\frac{(1-\rho^2)^{\frac{n}{2}}}{\{(1-it_1)(1-it_2)-\rho^2(1+it_1t_2)\}^n}$$

where t_1 refers to θ_1 and so on. Hence show that the sampling distribution of variances and covariances has the same characteristic function, except for constants, but values of n reduced by unity. Obtain the joint sampling distribution of variances and covariances.

2. If s_1, s_2 and r denote the sample standard deviations and the product moment correlation coefficient in a sample of size n from a bivariate normal population, show that the joint distribution of

$$v = \frac{s_1}{\sigma_1} \Big/ \frac{s_2}{\sigma_2} \text{ and } r \text{ is given by}$$

$$dF \propto \frac{v^{n-2}(1-r^2)^{\frac{n-4}{2}}}{(1-2\rho rv+v^2)^{n-1}}\,dr\,dv.$$

Putting $r = \dfrac{u(\lambda+\mu)-(1-u)(\lambda-\mu)}{u(\lambda+\mu)+(1-u)(\lambda-\mu)}$, where $\lambda = 1+v^2$, $\mu = 2\rho v$, and integrating over u show that the distribution of v is given by

$$dF = \frac{2(1-\rho^2)^{\frac{n-1}{2}}}{B\left(\frac{n-1}{2}, \frac{n-1}{2}\right)} \frac{v^{n-2}}{(1+v^2)^{n-1}} \left\{1 - \frac{4\rho^2v^2}{(1+v^2)^2}\right\}^{-\frac{n}{2}} dv.$$

3. Using the geometric methods or otherwise obtain the distribution of non-central χ^2. Indicate how you would use this distribution for deriving the distribution of the ratio of a non-central χ^2 and an independent central χ^2.

4. (a) Derive the distribution of the median of a sample of n independent observations from a normal population and show that the accuracy of the median is less than that of the mean.

(b) Obtain the distribution of the difference between $1-$th and $k-$th values of an ordered sample of size n from a rectangular population.

5. Obtain the non-null distribution of the sample multiple correlation coefficient in a sample of size n from a multivariate normal distribution.

6. Obtain a limiting distribution of the goodness of fit χ^2 under the null hypothesis specifying all the class probabilities.

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

Time : 4 Hours Full marks : 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

1. Discuss briefly the essential features of Vital Statistics data collected through the registration agencies in India with special reference to the findings in recent enquiries relating to their efficiency.

2. Describe a method of statistically evaluating the crude birth and death rates from the census returns indicating briefly the validity of the method.

3. Write short note on the international statistical classification of causes of deaths and suggest a feasible method of collecting statistics on causes of deaths in India.

4. If you are to assess the effectiveness of the execution of a family planning programme, describe briefly the methods of collecting necessary data and statistical evaluation.

5. Discuss the economic classifications adopted in the 1931 and 1951 census of India, indicating the difficulties which arise in making inter-censal comparisons.

PAPER IV AND V—PSYCHOLOGY AND EDUCATION (THEORETICAL)

Time : 4 Hours Full marks : 100

All questions carry equal marks

Attention : In this examination the questions are of various types. Read carefully the directions given for each type of question. The directions will tell you how to answer the questions of each type. Write all your answers in this booklet.

Directions for items 1 *and* 2 : For the questions that follow, you will find that with each question some answers are also given. Your task is to select for each question the ONE out of the ANSWERS GIVEN which is the correct or best answer for that question. The questions are serially numbered and in the answer-space in the right hand column these numbers are reproduced. The answers for each question are marked A, B, C etc. So in answering a question write *only* A, B, or C etc. by the side of the number of that question in the answer-space.

Answer here

1. Achievement test means the same thing as
   - A. Intelligence test
   - B. Personality test
   - C. Knowledge test
   - D. Ability test

   1.

2. I.Q. is the same as
   - A. M.A. − C.A. + 100
   - B. (M.A. − C.A.) × 100
   - C. $\frac{\text{C.A.}}{\text{M.A.}} \times 100$
   - D. $\frac{\text{M.A.}}{\text{C.A.}} \times 100$

   2.

Directions for items 3 *to* 11 : We want to compare the reliability of two tests. Under what conditions such a comparison is justified? A list of conditions is given below. For each condition, write

A, if it is necessary or,
B, if it is not necessary.

	Answer here
3. The same group has taken both tests.	3.
4. The units of measurement of the two tests are the same.	4.
5. The distributions of the scores on the two tests have the same skewness and kurtosis	5.
6. The length of the tests is the same.	6.
7. The trait is the same in both the tests.	7.
8. The task is the same in both the tests.	8.
9. The mean item difficulty is the same on both the tests.	9.
10. The item-intercorrelations are all equal.	10.
11. Each test is factorially pure.	11.

Directions for the rest of the items : The questions that follow require short answers. But the questions are such that you do not have to write much. Your answers should be *brief, clear-cut and to the point.* Below each question you will find a blank space where you should write your answer. Whatever you write in the blank space will be considered as your answer. If you write your answers elsewhere, those will not be considered.

[*N.B.*—Blank space for answer has not been provided in this print]

12. Give the meaning of reliability in terms of true, error and observed variance of scores on a test.
13. What is the difference between a 'speed' and a 'power' test?
14. What is meant by 'item analysis' of a test?
15. What are the sources of error variance in test-retest reliability?
16. What is the advantage of using a correlational measure of item discriminating power over that of critical ratio or chi-squared?
17. Describe a function of reliability which is invariant with regard to test-length.
18. Define 'parallel' tests.
19. Why must the validity of a test be less than its 'index of reliability'?
20. What is the use of 'standard error of measurement'?
21. What is the difference between 'standard error of measurement' and 'standard error of estimate'?
22. What is the effect of doubling a test on the variance of the observed scores?
23. Why should the 'split-half' method of reliability be used for 'power' test only?
24. If a test has 50 items and a reliability of .7, how many more items should be added to attain a reliability of .8?
25. A test of 100 items has a reliability of .6 and a validity of .4. If another 100 items are added, what will be the validity?

26. What are the factors determining the ceiling of the validity of a test?
27. In adding the marks of different examinations for a given group of examinees, is there any advantage in transforming to standard scores before adding?
28. What does Thurstone mean by 'Simple structure'?
29. '... for a given validity r_{xy}, the smaller the reliability r_{xx} of the test, the higher will be the upper limit of validity when a test is lengthened.' How?
30. Given the composite score X_i of the ith individual on a battery of tests, write equations according to the factor theories of Spearman and Thurstone, showing how X_i will be explained in terms of factors. Define the terms in the equations.
31. What is 'communality'?
32. What is the difference between orthogonal and oblique rotation of axes in factor analysis?

PAPER VIII AND IX—ECONOMIC STATISTICS (PRACTICAL)

Time: 4 Hours | Full marks: 100

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

1. The following table gives some recent national income and price data for India:

	Years						
	48–49	49–50	50–51	51–52	52–53	53–54	54–55
	at current prices (Rs. 100 crores)						
net national income at factor cost	86.5	90.1	95.3	99.7	98.2	104.8	96.2
net national income at market price	90.6	94.6	100.3	105.6	103.4	110.4	102.4
net capital formation	4.4	5.2	5.2	6.0	5.9	6.3	7.5
gross capital formation	8.0	9.2	9.4	10.5	10.6	11.3	12.6
	at constant (1948–49) prices (Rs. 100 crores)						
net national income at factor cost	86.5	88.2	88.5	91.0	94.6	100.3	102.8
	index number with 48–50 as base						
index number of investment cost	100.0	104.9	105.4	107.1	110.0	115.6	115.2

The population figures in 1941 and 1951 can respectively be taken as 318.8 million and 361.2 million.

(a) work out the following ratios for every year:
  (i) net capital formation at current prices/net national income at factor cost at current prices,
  (ii) gross capital formation at current prices/gross national income at market prices at current prices,
  (iii) net capital formation at constant prices/net national income at factor cost at constant prices, and
  (iv) gross capital formation at current prices/net national income at factor cost at current prices.

(b) Compare the overall national income price deflator and the investment cost index.

(c) Calculate for each year, per capita national income and per capita consumer expenditure, both at current and at constant prices. Consumer expenditure may be supposed to include government consumption and international transactions may be neglected. (30)

2. The following data are taken from Douglas, The Theory of Wages: Manufacturing Industries: USA, 1900–20.

x = Day's index of physical volume of manufactures (1899 = 100),

a = average number of wage earners in manufacturers (1899 = 100)

b = volume of physical capital in manufactures (1899 = 100).

year	x	a	b
1900	101	105	107
1901	112	110	114
1902	122	118	122
1903	124	123	131
1904	122	116	138
1905	143	125	149
1906	152	133	163
1907	151	138	176
1908	126	121	185
1909	155	140	198
1910	159	144	208
1911	153	145	216
1912	177	152	226
1913	184	154	236
1914	169	149	224
1915	189	154	266
1916	225	182	298
1917	227	196	335
1918	223	200	366
1919	218	193	387
1920	231	193	407

Fit a function $x = Ka^{\lambda} b^{1-\lambda}$ to the above data.

Work out the marginal productivity of labour in Indian manufactures for the year 1940 on the assumption that $x = 0.0390\ a^{\cdot 74}\ b^{\cdot 26}$ is a suitable production function and that average productivity of labour is Rs. 1400. How does the marginal productivity compare with an annual wage rate of Rs. 672? (50)

3. The following table gives population, output of electricity and output of coal in five countries:

	population (millions)	electricity (billion kwt)	coal (million metric tons)
USA	165	629.0	447.6
USSR	200	170.1	276.1
UK	50	94.1	225.2
CHINA	600	12.1	93.2
INDIA	380	8.5	38.8

Calculate per capita production of energy in these countries on the assumption that 1000 kwt = 0.6 metric ton of coal. Present all the figures in a table and indicate major limitations of the results. (20)

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 variable quality characteristic (x) is known to be normally distributed with a certain value σ for standard deviation. An item is considered defective if x exceeds the value U. Lots containing 1.5 per cent defectives are acceptable with risk 5 per cent and those containing 3.0 per cent defectives are to be rejected with risk 10 per cent.

(a) obtain the values of the constants in the plan for each of the following single sampling inspection procedures:

(i) attribute inspection by setting the gauge at U,
(ii) inspection by measuring x on each item inspected,
(iii) gauging inspection by setting the gauge at (U−1.5) and counting items not passed by this gauge.
(suitable approximations may be used in the claculations).

(b) Which of the above plans would you recommend if you know further that the net cost of inspection per item, is 1 anna for attribute, 5 annas for measurement and 1.5 annas for modified gauge types of inspection?

2. A certain type of glass tubing will function properly only if its diameter, which is gauged, measures between 4.7 and 5.1 mm. Any lot of tubes is considered good if there are only 2 per cent or fewer defectives and bad if there are 3 per cent or more defectives. Prepare *either* a chart for graphical procedure *or* a form for tabular procedure for item by item sequential inspection, with consumer risk 5 per cent and producer risk 2 per cent.

Assuming that the incoming lots have quality 2 per cent, what would be the ASN for the plan?

3. The number of surface defects observed during inspection of galvanised iron sheets of given areas, are shown below:

sheet number	area in unspecified units	number of defects	sheet number	area in unspecified units	number of defects
1	1	15	9	3	24
2	1	9	10	3	8
3	1	12	11	3	16
4	1	22	12	3	10
5	2	13	13	4	15
6	2	14	14	4	12
7	2	8	15	4	9
8	2	11	16	4	10

Analyse the data with the help of a suitable control chart.

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

Time: 4 Hours Full marks: 100

(a) Attempt *all* questions.
(b) Figures in the margin indicate full marks.
(c) Use of calculating machines is permitted.
(d) No text book on sampling is allowed. Practical books may, however, be allowed.

1. 40 pages were selected at random from the volume, 'American men of science' containing 2800 pages. The number of biographies, M_i, per page varies in general from about 14 to 21, the total number of biographies in the volume being 50,000. On each selected page, 2 biographies were chosen at random and the ages (y_{i1}, y_{i2}) of the scientists were recorded. The data are furnished below:

Ages of 80 scientists in 'American men of science'

($n = 40$, $m = 2$)

unit no.	M_i	y_{i1}	y_{i2}	unit	M_i	y_{i1}	y_{i2}
1	15	47	30	21	14	39	46
2	19	38	51	22	16	37	40
3	19	43	45	23	17	36	54
4	16	55	41	24	17	60	55
5	16	59	45	25	17	51	51
6	19	30	38	26	18	39	43
7	18	43	43	27	20	49	51
8	18	49	51	38	20	71	58
9	18	45	35	28	21	69	72
10	18	46	59	30	16	56	55
11	20	71	64	31	14	62	62
12	18	35	46	32	19	61	54
13	19	61	54	33	19	45	87
14	19	45	87	34	19	31	38
15	18	31	38	35	18	64	56
16	16	64	39	36	18	70	56
17	16	63	47	37	20	45	59
18	19	36	33	38	18	62	56
19	19	61	39	39	19	62	60
20	19	54	34	40	15	48	35

(a) Obtain the following estimates for the average age of the scientists in the complete volume,

(i) ordinary sample mean (unweighted),

(ii) ratio-estimate, in the form,

$$\frac{\sum_{i=1}^{n} M_i \bar{y}_i}{\sum_{i=1}^{n} M_i}$$

(iii) an unbiased estimate.

(b) Also calculate the variances and furnish a short critical note comparing the estimates. Would you recommend sampling with probability proportional to size in such a case? What will then be its advantages, if any? (60)

2. Data for a small artificial population which exhibits a fairly steady rising trend are provided below. We have $N = 40$, $K = 10$, $n = 4$. Each column represents a systematic sample and the rows are the strata.

Data for 10 systematic samples with $n = 4$, $N = Kn = 40$

	systematic sample numbers									
strata	1	2	3	4	5	6	7	8	9	10
I	0	1	2	1	4	5	6	7	7	9
II	7	8	9	10	12	13	15	16	16	17
III	18	18	19	20	21	20	24	23	28	29
IV	29	30	31	31	33	32	35	37	38	36

Work out from the data given above the relative efficiencies of (i) systematic sample, (ii) stratified sample and (iii) random sample. What happens to the relative efficiencies when the order of the observations is reversed in the second and fourth strata, and how would you account for the change, if any, obtained in the relative efficiencies? (40)

Or,

The table given below gives the measured volumes of timber on 25 systematically located plots of 1/10 acre each and corresponding eye-estimates. The total area of the plantation was 5124 acres and the total volume of timber, from eye-estimates of all the trees was 6,110,000 cu.ft.

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

y	x	y	x	y	x	y	x
160	103	194	209	152	80	182	152
46	15	254	209	215	178	74	148
63	58	135	110	125	65	24	207
91	70	146	110	100	196	255	167
126	95	154	109	287	167		
146	92	110	109	261	268		
87	110	112	128	169	152		

(a) Obtain estimates of the total volume of timber from the above data,
(i) based on the measurements of the sample plots only, and
(ii) by regression method.

(b) Calculate also the standard error for the total volume as estimated by both the methods. (40)

Time : 4 Hours Full marks : 100

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

1. (a) Attached is a list of tehsils along with their population. Point out the defects, if any, in this frame.

(b) After correcting for defects that may exist, select 5 tehsils with probability proportional to population with replacement. Give adequate reference to random number book, page, column and row used. Explain your procedure in detail by presenting the computation sheet in a suitable tabular form with proper column headings.

(c) 10 villages have been chosen with probability proportional to geographical area with replacement from a frame of 108 villages. Using the data given below, estimate the average area under paddy per village and the standard error of the estimate. For a given village, x represents area under paddy and y the geographical area in acres. The total area of the 108 villages is 50058 acres.

sample village	y	x	sample village	y	x
1	404	273	6	271	23
2	397	147	7	541	155
3	515	272	8	333	105
4	541	155	9	240	108
5	658	230	10	229	72

2. In a sample survey for study of economic condition of population, it is proposed to take a two stage sample with m villages as first stage units and n households per village as second stage units, both the stage units to be chosen with equal probability and with replacement.

The variance of an estimate is $V = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{mn}$ where $\sigma_1^2 = 5.2$, $\sigma_2^2 = 11.2$.

Cost of survey (in Rs.) = $T = 600 + 10m + 4mn$.

Determine the optimum values for n and m and work out the minimum variance for values of $T = 4000$, 6000, 8000, and 10000.

Fit a function $V = \phi(T)$ by any suitable method and find out $\frac{\delta V}{\delta T} \Big/ \frac{V}{T}$ at $T = 6000$.

3. Draw a suitable schedule with adequate spacings and headings for collecting data on *any one* of the following subjects :

(i) Agricultural household indebtedness,
(ii) Cost of cultivation of cereal crops.

Define important terms that you may use in the schedule in a separate sheet of paper.

4. In a survey for estimating the area and yield of major crops in India, it is proposed to take 8 clusters of 10 fields each for area survey, 5 sample cuts for yield estimation and 3 households for enquiries on income and expenditure from each sample village.

Making use of the data below, determine the number of investigators required and the total budget for the survey. Mention any assumptions that you may make in your calculations.

(i) the field establishment for this enquiry is sanctioned for one year,

(ii) three visits will be necessary for each village (one for each crop season),

(iii) enumeration of 8 clusters and 5 cuts can be done in about five days,

(iv) enumeration of one sample household will take about one day,

(v) number of sample villages to be contacted in each crop season is 3000,

(vi) the budget should provide for equipment, printing of schedule and other contingent expenses of the establishment,

(vii) the cost of an aparatus for crop cutting will be Rs. 80.00,

(viii) account should be taken of the time required for training the field staff, camp setting, journey and leave reserve, holidays etc.

(ix) for any other data such as pay, allowance etc. of investigators, inspectors and other staff that you may require for your calculations, you may use your guess estimate (mentioning these clearly).

Data for question 1(a) and 1(b).

serial number	district	tehsil	population (0000)
1	Burdwan	Sadar	80
2		Kalna	30
3		Katwa	31
4		Katwa	31
5		Asansol	76
6	Bankura	Sadar	97
7		Vishnupur	
8	Midnapore	Sadar	106
9		Contai	74
10		Tamluk	79
11		Ghatal	31
12		Jhargram	46
13	Hooghly	Sadar	45
14		Serampur	73
15		Arambag	37
16	Howrah	Sadar	93
17		Ulberia	
18	Birbhum	Sadar	64
19		Arampurhat	43
20	24 Parganas	Sadar	151
21		Basirhat	71
22		Basirhat	71
23		Barasat	31
24		Bangaon	21
25		Barackpore	88
26		Diamond Harbour	90

Paper VIII and IX—Design of Experiments Applied (Practical)

Time : 4 Hours | Full marks 100

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

1. The results of cotton varietal trials at the Agricultural College Farm, Nagpur, for the season 1937–38 are described below. The experiment consisted of five varieties viz., (i) Verum 434, (ii) Rosoum, (iii) E.B. 31, (iv) Verum 262 and, (v) Late Verum, replicated 4 times in randomised blocks on plots of 89′×18′, there being 12 lines of cotton per plot, 89′ long. The yields are given in lbs. per plot.

					East					
	block III						Block IV			
V434	Rosoum	V262	Late V	E.B.31		Late V	Rosoum	V262	V434	E.B.31
27.5	21.0	33.75	22.5	15.0		23.75	14.0	34.5	32.75	14.5
←North		Block I						Block II		South→
V434	Rosoum	E.B.31	V262	Late V		V262	V434	E.B.31	Rosoum	Late V
40.5	25.25	21.0	41.5	30.25		40.75	31.0	17.75	31.0	27.5

West

Make a statistical analysis of the data. Prepare a summary table of the mean yields and record your principal findings. It was suspected before the experiment was started that since Blocks III and IV are situated on comparatively lowlying land, the yields in these blocks would be lower. Does the experiment corroborate this? Is it true that both Verum 262 and Verum 434 are superior to the rest? What is the efficiency of this design compared to a completely randomised design?

2. One satisfactory way known of assessing the life of a motor tyre is to run the tyre in a test car under normal road conditions. In order to eliminate the large variations which usually exist between one car and another, the tyres must be run on the same car and they must be interchanged frequently in order to eliminate variations due to position. Four treatments were under study and each tyre was divided into three equal parts where treatments could be applied, and the wear on each part was separately assessed. Below is given the full description :

Tyre I	Treatments	A, B, C
Tyre II	-do-	A, B, D
Tyre III	-do-	A, C, D
Tyre IV	-do-	B, C, D

Road tests on tyres : Relative Wears
(arbitary units)

treatments	tyre (blocks)			
	1	2	3	4
A	238	196	254	–
B	238	213	–	312
C	279	–	334	421
D	–	308	367	412

Analyse the experiment with a view to find out differences between treatments and differences between tyres.

3. Below are given the plan and yields of a 1/2 replicate of a 2^6 design in 4 blocks of 8 plots. Identify the defining contrast and the confounded contrasts. Carry out a statistical analysis of the data assuming three and higher factor interactions to be negligible.

Block I		Block II		Block III		Block IV	
(a)	740	ae	1156	ad	1184	ae	1176
ab	1008	be	692	bd	864	be	1008
acde	1468	de	1104	ce	896	cd	1000
bcde	1100	abde	1324	abce	1356	abcd	1292
acdf	1108	df	860	cf	1324	cdef	1001
bcdf	888	abdf	1248	abcf	1008	abcdef	800
ef	784	aef	844	adef	1508	af	996
abef	780	bef	996	bdef	860	bf	984

4. A Latin Square experiment was carried out to test for possible differences between the breaking strengths (in lbs. per sq. cm.) of glass made from six furnaces using six different moulds. Six runs of each furnace were needed to complete the 36 possible combinations of furnace and mould. One result was spoilt through the cooling apparatus becoming out of order and is shown as blank in the table of breaking strength below, the letters standing for the moulds.

		furnace					
		1	2	3	4	5	6
R U N S	1	B 98.8	E 111.0	D 106.0	C 118.2	A 58.0	F 130.5
	2	D 113.0	B 89.5	E 115.0	A 79.5	F 96.8	C 129.0
	3	A 79.0	D 107.8	B 101.5	F 104.5	C 103.0	E 118.5
	4	F 109.2	A 78.5	C 106.2	D –	D 103.0	B 93.5
	5	E 105.8	C 107.5	F 103.8	B 90.2	D 113.0	A 72.5
	6	C 102.0	F 97.0	A 70.8	E 104.8	B 93.2	D 131.2

Analyse these data for mould and furnace effects, giving means and standard errors for each mould and furnace.

5. The actual lay-out and yields of a design for certain varietal trial on 10 varieties of soyabeans are given below:

	V	Y	V	Y	V	Y	V	Y	V	Y
	1	20	6	80	6	75	2	60	10	145
	2	45	1	15	8	65	5	100	8	60
	3	35	7	50	4	45	9	15	3	30
	4	40	5	105	9	10	10	155	7	40
Block	1		2		3		4		5	

V = variety, Y = yield in lbs., size of plot = $\frac{1}{60}$ acre.

Analyse the experiment with a view to order the varieties according to their yielding capacities.

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

Time: 4 Hours Full marks: 100

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

1. The distribution of women (married only once) in two social classes, classified by the duration of married life and age at effective marriage is shown in the following table. Figures in parenthesis entered in each cell give the total numbers of children born to the group of women to which the cell corresponds.

duration of married life in years	social class I (higher professions, higher services)			social class II (the remaining occupations)		
	age at effective marriage (in years)			age at effective marriage (in years)		
	<15	16–19	20 and over	<15	16–19	20 and over
0— 4	4 (4)	8 (7)	26 (17)	8 (8)	44 (40)	37 (32)
5— 9	4 (12)	24 (52)	28 (72)	62 (134)	100 (250)	30 (62)
10—14	0 (17)	23 (71)	24 (71)	117 (448)	118 (425)	27 (86)
15—19	0 (18)	29 (97)	18 (53)	80 (396)	55 (234)	11 (53)
20 and over	10 (44)	18 (74)	4 (19)	52 (331)	26 (121)	6 (22)

Compute the average fertility rates of the two social classes standardised for (a) duration of married life, (b) duration of married life and age at effective marriage. How has the age at effective marriage affected the difference between class fertility rates

2. In a fertility enquiry, an analysis of the period of risk to pregnancy of married women by type of contraception practised gave the following results :

item	contracepted exposure		non-contracepted exposure			
	social class I (high profession etc.)	social class II (remaining occupations)	couples who overpractised contraception		couples who never practised contraception	
			social class I	social class II	social class I	social class II
month of exposure to the risk of pregnancy	14484	20595	3610	10656	5380	41181
number of pregnancies observed	166	371	221	679	115	1421

Calculate the expected pregnancy rates of the two social classes that would have resulted if none of the couples in either social class had ever practised contraceptions assuming that couples generally resort to contraceptive practices only after attaining desired fertility levels.

3. In an epidemiological investigation relating to the incidence of a certain infectious disease 9469 individuals were selected at random from a certain community. It is known that an attack of this disease confers lasting immunity to the affected individuals. A study of the case histories of the 9469 individuals gave the following results :

present age of individuals	number of individuals in the sample	number reported as having had the disease at any time till the date of survey
(1)	(2)	(3)
0— 4	1203	22
5— 9	1220	60
10—14	1128	90
15—19	1038	111
20—24	960	127
25—29	802	123
30—34	722	125
35—39	640	120
40—44	562	112
45—49	484	104
50—54	309	85
55—59	311	68
all ages < 60	9469	1147

Assuming that non-immune individuals in all ages are equally susceptible to this disease, estimate the incidence rate and obtain the expected frequencies of column 3 in the above table.

4. The distribution of women in two social classes by number of children born and number of children surviving the first year of life is shown in table below:

TABLE

number of children surviving	social class I (higher professions, higher services)						social class II (other occupations)					
	number of children born						number of children born					
	1	2	3	4	5	6	1	2	3	4	5	6
0	8	5	2	0	0	–	48	28	6	2	0	0
1	144	33	10	3	0	–	275	102	49	18	4	0
2	–	178	32	25	4	–	–	362	142	85	24	4
3	–	–	139	18	18	–	–	–	309	134	83	12
4	–	–	–	115	20	–	–	–	–	217	71	30
5	–	–	–	–	63	–	–	–	–	–	129	28
6	–	–	–	–	–	–	–	–	–	–	–	23
Total	152	216	183	161	105	–	323	492	506	456	311	97

Examine from the above data whether infant mortality is inversely related to the social status of the mothers.

PAPER VIII AND IX—PSYCHOLOGY AND EDUCATION (PRACTICAL)

Time: 4 Hours Full marks: 100

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

1. A group of examinees (N = 20) had taken 3 different tests. The number of items in each test was as follows:

Test A—40 items
Test B—40 items
Test C—50 items

On each test the information available on each examinee was on the number of items done right (R) and the number of items done wrong (W). On the tests A and B, information was also available on the last item reached (L) but not on test C. On test C, however, an additional information was available in the form of the number of items on which double answers were given (DA). The data are given in the table below:

number of examinee	tests	A			B			C		
		R	W	L	R	W	L	R	W	DA
1		34	2	37	35	2	38	52	6	0
2		20	8	37	21	6	34	18	40	0
3		26	5	40	26	4	34	55	3	0
4		26	2	40	26	1	34	9	4	29
5		18	6	37	20	5	36	11	7	32
6		25	6	40	14	12	26	23	30	0
7		22	2	27	22	4	32	31	8	0
8		32	5	40	29	7	36	10	3	26
9		33	7	40	39	1	40	51	5	0
10		29	9	39	33	4	38	54	4	0
11		25	11	40	26	7	36	52	6	0
12		23	1	24	20	2	22	19	4	16
13		27	5	40	23	5	34	31	24	0
14		19	9	38	21	10	37	37	18	0
15		28	4	32	28	5	33	15	11	24
16		17	12	37	18	12	34	12	10	11
17		28	7	40	26	11	37	10	22	0
18		35	4	40	35	4	39	16	26	0
19		36	4	40	35	5	40	17	35	0
20		23	4	40	21	6	37	31	16	0

(a) Find the corrected score for each examinee, when the items in Tests A and B are 3-choice and in Test C 5-choice.

(b) Find the composite score for each examinee on a scale of $\bar{x} = 50$ and $\sigma = 10$, equating the contribution of each test to the total.

2. Find the reliability of the test and the standard error of measurement from the following data on score per item:

person	item														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	0	0	1	0	1	0	1	1	1	1	1	0
2	1	0	1	1	0	1	0	1	0	1	0	0	1	0	0
3	0	1	1	1	0	1	1	1	0	0	1	0	0	0	1
4	1	1	1	0	1	0	1	1	1	0	1	0	1	1	0
5	1	0	1	1	0	1	0	0	0	0	0	1	0	1	1
6	0	1	0	1	0	1	1	0	0	0	1	0	1	1	0
7	1	1	0	0	1	0	1	1	0	0	1	1	1	0	0
8	1	1	1	0	1	1	0	1	0	0	0	1	1	1	1
9	1	1	1	0	1	0	1	1	1	1	1	1	0	0	0
10	1	0	1	1	0	1	0	0	0	0	0	0	1	1	1
11	0	1	1	0	1	1	1	1	1	1	1	0	1	0	0
12	0	1	1	1	0	1	1	1	1	1	0	0	1	0	1
13	1	0	1	1	0	1	0	1	0	0	0	1	1	0	1
14	0	0	0	1	0	0	1	1	0	0	1	0	0	1	0
15	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1
16	1	1	1	0	0	0	1	0	0	0	0	0	0	1	0
17	0	1	0	0	1	1	0	0	0	0	1	1	1	0	1
18	1	1	0	1	1	1	0	1	1	1	0	0	1	0	1
19	0	1	1	0	0	1	1	0	1	1	1	0	1	1	0
20	0	1	1	1	1	1	0	1	1	1	0	1	1	0	0

3. The intercorrelations of 7 tests are given below. Their reliabilities are also given. For each test, find the first centroid factor loading, the communality, the specificity, the uniqueness and the error.

Intercorrelations of tests

	1	2	3	4	5	6	7	reliability
1		.56	.68	.41	.44	.41	.31	.804
2			.62	.49	.34	.30	.43	.871
3				.31	.36	.33	.32	.912
4					.33	.32	.29	.917
5						.36	.22	.391
6							.08	.478
7								.608

4. From the inter-correlations of 8 tests on a battery, two contriod factors were obtained. It was found on rotation that 4 tests had high loadings on Factor I but practically zero laodings on the other, whereas the rest 4 tests had high loadings on Factor II but zero loadings on Factor I. The data on factor loadings and the relevant intercorrelations for the first 4 tests are given below. Now, using the first 4 tests find the multiple regression equation for predicting the first factor score given first 4 test scores.

	Tests				Factor loadings	
	1	2	3	4	I	II
1		.608	.413	−.132	.80	.03
2			.481	.137	.78	.00
3				.120	.66	.03
4				—	.42	−.10

PAPER VIII AND IX—THEORIES OF INFERENCE (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. $x_1, x_2, \ldots, x_8$ is a random sample of 8 members from a normal population with zero mean and variance σ^2. It is proposed to test the hypothesis $\sigma = 1$ against the alternatives $\sigma \neq 1$. Obtain Type A_1 Test (Uniformly most powerful unbiased test), and evaluate the power of the test for the following values of σ:

0.4, 0.6, 0.8, 1.2, 1.4, 1.8, 2.2

2. In a certain acceptance inspection programme, it has been decided to allow two per cent defective units. The acceptance of a lot with 8 or more per cent of defectives is a serious error. The management is prepared to take a 5 per cent chance of rejecting good lots and a 10 per cent chance of accepting bad ones.

(a) Construct a suitable sequential sampling plan to meet the above situation. Give both tabular and graphical representation of the procedure you adopt.

(b) An inspector tests 40 units without finding a single defective. Under your plan, would he have decided to accept the lot before reaching the 40th unit or would he have to continue further before reaching a decision? Would he have come to a decision to reject the lot if he had found the 10th, the 18th and the 23rd units defective?

(c) Sketch the OC curve and the average outgoing quality (AOQ) curve and estimate the AOQL.

(d) Sketch the ASN curve for this plan.

———

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1958

PAPER I—THEORETICAL STATISTICS (GENERAL)

GROUP B

Time: 2 Hours Full marks: 50

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

1. (a) Write a note on Fisher's 't' test for paired comparison. In what way does this differ from 't' test for testing significance of difference between means from two samples? What are the other uses of 't' test?

(b) If you are given two frequency distributions of a character x obtained from two samples, how will you test whether the two samples are from the same population? What are the assumptions involved in the test procedure? How does the test differ, if you are given the frequency distribution from a sample and asked to test whether it is from a population whose porportion of cell frequencies are given?

2. (a) Define the product moment correlation coefficient and show that it is invariant under location and scale transformations of the variates. Find out the upper and lower limits of the product moment correlation.

(b) Given that P is the price index and S is the corresponding relative stock of gold money. The means, standard deviations and the correlation coefficient are shown below:

	P	S
Mean	100	103
Standard deviation	8.0	4.0

$$r_{PS} = .4$$

Using these constants, construct a linear equation to read off the values of P for values of S and hence find the most likely value of P when S = 111.0. Test whether this estimate is significantly different from 150. What are the assumptions involved in your test procedure?

(c) Given all the total correlation coefficients between the variates x_1, x_2, x_3, x_4, describe how you will calculate the second order partial correlation coefficient between x_1 and x_2.

3. If T be an estimate of the population parameter θ and F(T) is a suitable function of T, show that

$$\text{Var } F(T) = \text{Var } (T). \quad \{F'(\theta)\}^2.$$

Clearly state the assumptions involved.

Note.—Question paper for Group A of paper I is not available for circulation.

Utilise the above result to obtain suitable transformations of the estimates of the following so as to make the variance of the transformed statistics free from the population parameter.

(i) Proportion of success p in the binomial distribution.

(ii) The mean in the Poisson distribution.

Paper II—Applied Statistics (General)

Time: 4 Hours . Full marks: 100

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

(b) Attempt *any three* questions from Group A and *any two* from Group B.

(c) All questions carry equal marks.

Group A

1. Describe the steps involved in organising a sample survey for collection of data relating to the labour force, employment and unemployment in India. How would you define the terms labour force, employment and unemployment?

Do you think that the definitions which are appropriate for economically advanced countries are also so for the under developed countries? Give reasons.

2. One of the objects of planned development in India is to reduce disparities among different regions. What indicators would you choose to study the level of development of a region? Indicate briefly, with reference to your State, availability or otherwise of the data you would need. (Mention the name of your State).

3. What is meant by consumer price index and what is its usefulness? How is it constructed? Discuss whether we should have one such index for India as a whole or we should have more than one. If you recommend more than one, what are your proposals?

Do you consider an all-India consumer price index to be particularly useful? State your reasons. For which sectors of the population is the construction of such a price index specially urgent? Give your arguments in support of your views.

4. State where the statistics of *any two* of the following are available. Also discuss their limitations and the manner in which they can be improved.

(a) wages, (b) employment, (c) consumer expenditure,

(d) capital formation.

5. If you are placed in charge of the processing of data of a large scale sample survey, describe in detail how you would organise the work, if it is to be done (a) on manual basis or (b) by punched card mechanised method. What are the relative advantages of the two methods?

Group B

6. Discuss briefly the advantages and disadvantages of the following experimental designs: (i) completely randomised design, (ii) randomised block and (iii) latin square.

Explain clearly how the efficiency of the randomised complete block design relative to completely randomised design can be estimated.

7. Describe the general procedure of population census-taking in India. What are the principal sources of error in census statistics? How do you propose to control them? Discuss the various alternative methods of estimating inter-censal and post-censal population.

8. State the criteria for testing the reliability and validity of a standardised test.

Explain how the reliability coefficient of an intelligence test when item scores are available can be determined. Can you utilise this coefficient for testing the adequacy of a behaviour model?

9. What are (i) mitosis and (ii) meiosis. Describe how 'crossing over' takes place in meiotic division. Explain clearly the significance of 'crossing over'.

Paper III—Statistical Inference (General)

Time : 4 Hours | Full marks : 100

(a) Attempt *any two* questions from Group A and *any three* from group B.
(b) Answers to the different groups are to be given in separate books.
(c) All questions carry equal marks.
(d) Use of calculating machines is not permitted.

Group A

1. Define the terms consistency, asymptotic normality, and asymptotic efficiency of an estimate.

State (without proof) the formula for the asymptotic efficiency of the sample median in samples of independent observations from a continuous population. Show that if the population is Normal, the efficiency is $2/\pi$.

2. Show that if t_1 is an unbiased estimate of a real parameter θ, if y is a sufficient statistic, and if t_2 is the conditional expectation of t_1 given y, then t_2 is also an unbiased estimate of θ, and Var $(t_2) \leqslant$ Var (t_1) no matter what the value of θ.

Use the above result to show that if $X_1, X_2, \ldots, X_n$ is a sample of independent observations from a Poisson distribution with mean λ then $t = (X_1 + .. + X_n)/n$ is the unbiased uniformly minimum variance estimate of λ.

3. A sample of nine independent observations is drawn from a continuous population with density $f(x, \theta)$, where θ is unknown. The sample values are :

$$-0.2, -2.4, 4.0, 0.5, 1.5, -1.8, -2.5, 3.7, \text{ and } 0.9.$$

How would you estimate θ if it is known that :

Case 1. $f(x, \theta) = \begin{cases} \frac{1}{2} & \theta \text{ for } -\theta < x < \theta \\ 0 & \text{otherwise} \end{cases} \quad (0 < \theta < \infty)$

Case 2. $f(x, \theta) = 1/\pi\,[1+(x-\theta)^2] \quad (-\infty < \theta < \infty)$.

Discuss briefly the method you suggest in each case.

Group B

4. (a) What is meant by confidence interval?

(b) Given a sample of size n from a normal population with unknown mean and unit variance, what is the confidence coefficient of the interval

$$\left(\bar{x} - \frac{2}{\sqrt{n}}, \infty\right) \text{ for the estimation of } \mu ?$$

(c) Bring out in brief the relation between Neyman's theory of 'shortest' confidence intervals and that of testing statistical hypotheses.

5. (a) What are the conditions which a region must satisfy in order to give a most powerful test of a simple hypothesis against one alternative hypothesis? Examine if these conditions are necessary and sufficient.

(b) For a normal population N (0, σ), show that a uniformly most powerful test exists for the hypothesis $\sigma = \sigma_0$ when alternatives are only of the form $\sigma_1 < \sigma_0$.

6. What is Wishart's distribution? Show that the pooled covariance matrix from two independent covariance matrices arising out of normal populations follows Wishart's distribution, and that there is a simple relationship among the degrees of freedom of the three Wishart's distributions involved.

7. Write notes on:
   (a) Null hypothesis,
   (b) Relation between sufficient estimators and tests of hypotheses,
   (c) Reproductive property of the χ^2 distribution.

Paper VI—(Practical)

Time: 6 Hours — Full marks: 100

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

Group A

1. The length of each conversation made through a telephone exchange was recorded during a certain period. With the data thus collected the following frequency distribution was formed:

length of conversation (in seconds)	frequency
1— 100	6
101— 200	28
201— 300	88
301— 400	180
401— 500	247
501— 600	260
601— 700	133
701— 800	42
801— 900	11
901—1000	5

Determine the appropriate type of Pearsonian curve to be fitted to the above data. Calculate the constants involved in the equation of the curve and expected frequencies for all the classes. (30)

2. *Either,*

(a) In a distribution which is exactly normal, 9.1 per cent of the frequencies are under 35 inches and 5.5 per cent are over 57 inches. What percentages of the frequencies of the distribution are (i) greater than 60 inches, (ii) less than 32 inches, (iii) between 40 inches and 50 inches. Calculate the lower quartile, and the sixtieth, sixtyfifth and ninetyfifth percentiles.

If the total frequency is 1000, draw the curve by calculating a suitable number of ordinates.

(b) At the beginning of a year, a life insurance company had 1000 male policy-holders of age 25, and the average sum assured per policy-holders is Rs. 2000.00.

A mortality table shows that of 8582 men alive at age twenty five, 84816 are alive at age twenty six. Find the upper and lower limits for the amount which the company would reasonably be expected to pay out during the year on these policies. (20)

Or,

(a) In order to test the effectiveness of a serum in preventing colds, 1000 individuals were selected at random and the first 500 randomly selected individuals of this group were treated with the serum and the rest were untreated. Information was collected for all the individuals in respect of (i) no attack of cold, (ii) one attack of cold and (iii) more than one attack of cold during one year. It was found that the total number of persons who were attacked with cold more than once during one year was 243, out of these cases 140 were untreated. The total number of persons who were attacked with cold only once during the year was 281, and 252 individuals among the treated group did not have a single attack of cold. From a study of these records do you consider the serum to be effective in perventing cold?

(b) A car-hire firm has 2 cars. A car is hired out for a full day. The number of demands for a car on each day is distributed as a Poisson distibution with mean 1.5.

Calculate the proportion of days on which neither car is on demand, and the proportion of days on which some demands have to be refused.

On what proportion of days is one of the cars not in use? What proportion of total demands has to be refused? (20)

Group B

Answer *any two* questions

3. (a) The following are breaking strength in lb. of twelve strips of a cotton fabric

160	175	161	158	166	174
166	172	165	181	170	158

Assuming that the figures represent random and independent observations of a normally distributed random variable, determine the 95 and 99 per cent central confidence intervals for the population standard deviation? (12)

(b) Two independent samples of 19 and 23 pairs gave estimates of the correlation coefficients 0.55 and 0.70 respectively. Are these correlation coefficients significantly different from each other?

Assuming that they are not significantly different from one another, obtain a pooled estimate of the correlation coefficient. (13)

4. Eighteen sample units are measured for three characters (x_1, x_2, y). Numerical calculations gave the following results:

$$\Sigma y = 1463; \qquad \Sigma y^2 = 131,299$$
$$\Sigma x_1 y = 20706.20; \qquad \Sigma x_1 = 215$$
$$\Sigma x_1^2 = 4321.02; \qquad \Sigma x_2 y = 63825$$
$$\Sigma x_2 = 758; \qquad \Sigma x_2^2 = 35,076$$
$$\Sigma x_1 x_2 = 11010$$

Obtain the Multiple regression equation of y on x_1 and x_2.

Test the significance of the two partial regression coefficients. Calculate the Multiple correlation coefficient of y on x_1 and x_2 and test its significance. (25)

5. The following table gives certain statistics relating to the milk yields (x) of 168 dairy cows classified according to manner of milking and to whether they have calved before or not. In each cell the first figure given is the number (n) of cows, the second the sum (Σx) of the yields of these cows during the lactation period and the third the uncorrected sum of the squares (Σx^2).

Is there evidence that yield is lowered by machine milking for (i) first calvers and (ii) others? In case there is evidence of decrease in yield for both the first calvers and others, examine if the amount of decrease is the same in both the cases.

Do first calvers yield less milk than the other cows?

Statistics of milk yield

	first calvers	others
hand milked	$n = 34$	$n = 50$
	$\Sigma x = 4771$	$\Sigma x = 8357$
	$\Sigma x^2 = 747243$	$\Sigma x^2 = 1475865$
machine milked	$n = 34$	$n = 50$
	$\Sigma x = 4197$	$\Sigma x = 6783$
	$\Sigma x^2 = 591705$	$\Sigma x^2 = 100,0048$

(25)

Paper VII—(Practical)

Time : 6 Hours | Full marks : 100

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

Group A

1. An experiment to study the effect of four treatments on the yield of jowar was conducted for a period of three years with four replication each year. A randomised block design with fresh randomisation each year was adopted. The yield per plot of 1/40th of an acre in oz. is given below. Interpret the results on the basis of combined analysis for the three years. You are advised to prepare first the analysis of variance tables for each year separately.

Yield per plot in oz.

year	treatment	replications			
		I	II	III	IV
1943–44	1	288	416	320	352
	2	432	544	352	528
	3	320	256	400	256
	4	400	512	368	512
1944–45	1	301	214	143	180
	2	361	380	266	105
	3	272	155	123	162
	4	378	132	73	52
1945–46	1	488	258	394	324
	2	390	520	356	438
	3	520	412	420	468
	4	468	445	392	518

2. The table below gives the electricity sold (million kwh) for different months during 1952–56. Assuming linear trend, find out the monthly seasonal indices :

months	years				
	1952	1953	1954	1955	1956
January	410	442	478	545	622
February	391	417	451	522	622
March	393	436	483	568	651
April	393	438	504	564	694
May	413	456	515	569	669
June	409	450	510	576	658
July	430	471	516	583	656
August	429	469	510	574	650
September	431	464	523	587	666
October	442	469	512	592	670
November	436	456	535	597	668
December	458	487	558	648	715

3. Given below are the entries from a Life table wherein the missing values have been indicated by crosses:

x	l_x	d_x	$1000q_x$	L_x	T_x	e_x^0
25	91,241	×	×	×	×	43.28
26	91,019	223	×	×	×	×
27	×	×	×	×	×	×
28	90,568	×	2.50	×	×	×
29	×	×	×	×	×	×
30	×	×	×	89,007	3,495,526	×

Complete the above table and calculate the following probabilities:

(i) that a person aged 25 will attain age 30,

(ii) that a person aged 28 will die within one year,

(iii) that a person aged 26 will die between ages 27 and 30.

GROUP B

4. It is propsed to carry out a sample survey by mail enquiries for estimating the average number of miles run by new medium size cars prior to changing the first set of tyres.

Anticipating a large non-response rate, it is propsed to mail the schedules to a large sample and to follow up, by actual visit, a sample of non-respondents.

$$\bar{x} = p\bar{x}_1 + (1-p)\bar{x}_2$$

is an unbiased estimate and its variance is approximately equal to

$$V(\bar{x}) = \frac{\sigma^2}{n} + \left(\frac{1}{\theta} - 1\right)(1-\pi)\frac{\sigma_0^2}{n}$$

p = response rate for mail enquiries (proportion of respondents) in the sample,

π = response rate for mail enquiries in the entire population found from previous studies = 0.25 approximately,

σ^2 = variance of x in the entire population = 10,

σ_0^2 = variance of x among non-respondents in the population = 10,

θ = proportion of non-respondents in the sample, proposed to be followed by up interview,

n = number of sample units to which schedules were mailed.

Find the optimum values of n and θ (i.e. values which minimise the expected costs of the survey) such that the variance of the estimate will be same as that of a random sample of size 100 from the population. (assuming the population to be very large).

Given that the expected total cost, c, of the survey is given by

$$c = c_1 n + c_2 np + c_3 n(1-p)\theta.$$

where

c_1 = cost of mailing a questionnaire = Rs. 0.15

c_2 = cost of processing a questionnaire returned by mail = Rs. 0.50

c_3 = cost of interviewing a person and processing the same = Rs. 4.50

Estimate the corresponding cost of the survey. If π were equal to 0.4, what would the optimum value of n and θ be and the corresponding cost?

5. In the manufacture of a particular brand of baby food, the label states that the packet contains 1 lb. net. A packet is considered to be defective if it contains less than 15.5 oz's. The packets are filled by machine and the setting of machine is considered to be satisfactory if not more than 5 per cent of the packets are defectively packed.

Past experience suggests that packet to packet variation in weight follows a normal distribution.

(a) If the mean of this distribution μ (oz) = 16 ozs. and s.d. σ(oz) = 1 oz. what proportion (p) of the manufactured packets are defectives?

(b) The setting is suspected to be defective and fresh adjustments are advised whenever

$$\frac{15.5 - \bar{x}}{s} > k$$

where $\bar{x}$ and s are estimated from a random sample of n units.

Assuming n to be large, obtain the value of k such that the probability of the machine being unnecessarily held up for adjustments is less than .05; that is,

$$\text{Prob}\left\{\frac{15.5 - \bar{x}}{s} > k\right\} = .05 \text{ when } p = .05.$$

Hint: Note that for any k and large n,

Prob. $(\bar{x} + ks < 15.5)$

can be computed using the asymptotic normal distribution of $\bar{x}+ks$, which has mean $\mu+k\sigma$ and variance

$$\frac{\sigma^2}{n}\left(1+\frac{k^2}{2}\right).$$

(c) What is the probability of the machine being allowed to work when $p = .10$, under the procedure you suggest.

6. Four tests were administered on a group of 250 final year boys of a higher secondary school and the intercorrelations in scores of these tests were computed. They are as follows:

Table of intercorrelations

tests	1	2	3	4
1	1.00	–	–	–
2	.64	1.00	–	–
3	.65	.74	1.00	–
4	.52	.78	.53	1.00

It is anticipated that these four tests have two common facotrs. Obtain the loadings of these factors.

Paper IV and V—Economic Statistics (Theoretical)

Time : 4 Hours | Full marks : 100

(a) Attempt *any two* questions.

(b) All questions carry equal marks.

1. The following is an extract from a speech by the Prime Minister :

"I wish to lay stress here on our approach to palanning, I mean the democratic approach. This approach has also to be statistical approach, that is an approach based on statistical information, not on vague ideas and ideals. Statistical data have to be collected, and sample surveys and calculations made at every stage, to calculate our actual resources, the employment potential of the plan, the production rate, and the commodity balances. For, it must be remembered, planning does not merely mean putting up a factory here or a factory there. Planning implies interlocking of production, consumption, employment and a large number of other things such as transport, social services, education and health. The whole thing has somehow to be brought together. Of course, human relationships in a vast country of 370 million people cannot be dealt with in a mathematical way. There are numerous uncertain factors, the coming of monsoon, for instance. Yet the element of uncertainity and error can be greatly reduced by planning." (National Development Council, 20 January 1956).

Amplify the idea and bring out the role of statistics in planning.

2. Describe the methods used for deriving statistical demand curves, briefly indicating the nature of data utilised for the purpose.

What is the price elasticity of demand for the demand curve

$$p = \frac{a}{b+cy}$$

where p is price, y is the quantity demanded and a, b, c are constants?

3. Indicate the different methods for obtaining the trend of a time series. Prove that the logistic curve

$$y = \frac{k}{1+e^{a+rt}}$$

in which y depicts an output, t time and k, a and r are constants, has a point of inflexion at

$$t = -\frac{a}{r},\ y = \frac{k}{2}.$$

At what point of the curve the time rate of change of y is greatest and what is this rate?

4. Obtain Pareto's expression for income distribution and describe its important properties.

Work out an equation for the concentration curve and prove that the coefficient of concentration (ρ) is given by

$$\rho = \frac{1}{2\nu-1}\ ,$$ where ν is the Pareto constant.

Write down the equations of two other functions used for describing income distributions.

5. Sketch briefly, the method of estimation used for obtaining the official estimates of national income in India.

What is the difference between net domestic product at factor cost and gross national product at market prices?

6. Indicate what is a cost of living index number and describe how it is constructed. What are the important uses of cost of living index numbers? Describe briefly available data on working class cost of living index numbers in India.

What kind of values of income elasticity of demand you would expect in India for items of consumption such as foodgrains, milk and milk products, salt and clothing?

7. Mention three nationwide large scale sample surveys in India.

Supposing you want to estimate the personal income of the households residing in an Indian State, what kind of a survey would you plan? Give your answer under the following major heads:

(a) frame and design, (b) questionnaire, (c) field staff and field organiation, (d) scrutiny

---

Paper IV and V—Statistical Quality Control (Theoretical)

Time: 4 Hours Full marks: 100

(a) Attempt *any two* questions.

(b) All questions carry equal marks.

1. Discuss the relative merits of attributes and variables inspection.

Give a single sampling acceptance plan for variables inspection, in which the standard deviation of the quality characteristic being not known, the mean range based on sub-samples, has to be used.

Derive the elements in the plan.

2. Describe the Military Standard 105A sampling plans for attributes. Examine in what respect these plans are different from plans in which the sample sizes are pre-assigned.

3. Compare item by item sequential sampling with other methods of sampling, when the purpose is (i) process control, (ii) estimating lot quality and (iii) lot acceptance.

4. Comment in detail on the following:

(i) "action is recommended on the basis of control charts when there is evidence that the improbable has happened."

(ii) "the means to assess errors due to sampling involved in control chart analysis is the operating characteristic of the control chart."

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

(i) Use of sample range and mean range instead of standard deviation.

(ii) Determination of tolerance limits from sample data.

(iii) Formation of sub-lots and proportional sampling to achieve representativeness.

(iv) Consequences of compositing ultimate sample-units to form gross-sample of bulk material.

(v) The problem of allocation of specification tolerance of assembly, between those of components.

(vi) The use of factorial experiments in place of control chart analysis.

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) Treatment should be mathematical wherever possible.

1. Write a short critical note on each of the following:
   (i) Improvement over the performance of systematic sampling in the presence of linear trend.
   (ii) Systematic sampling in two dimensions.
   (iii) Finite population correction.
   (iv) Relative precision of stratified random and simple random sampling.

2. Deduce the variance for the total in two-stage sampling without replacement. Write down the expression for the variance when the sampling is extended to three stages.

3. (a) Show that in stratified random sampling, the variance of the estimated mean $\bar{y}_{st}$ is smallest for a fixed total size of sample, if the sample is allocated with n_h proportional to $N_h S_h$, where

n_h is the number of sample units of the hth stratum,
N_h is the total number of units in the hth stratum,
and S_h is the true variance of the hth stratum.

(ii) Show that, with a cost function of the form

$$\text{cost} = C = a + \Sigma c_h n_h,$$

(that is, where cost mounts directly with the size of sample), the variance of the estimated mean $\bar{y}_{st}$ is a minimum when n_h is proportional to $N_h S_h / \sqrt{c_h}$, where the symbols have their usual significance.

4. (i) If $\bar{y}_z$ is the mean of a sample drawn with probability proportional to z, show that

$$V(\bar{y}_z) = \frac{1}{n} \sum_{i=1}^{N} z_i (y_i - \bar{Y}_z)^2, \text{ where } E(\bar{y}_z) = \bar{Y}_z.$$

(ii) A sample of size n is drawn with probability proportional to measures of size $z_i = M_i / \Sigma M_i$. The item totals for the units in the sample are $y_1, y_2, \ldots, y_n$ where the same unit may appear more than once, since sampling is with replacement. As an estimate of the population total Y we take Y_{pps} (probabilty proportional to size), where

$$Y_{pps} = \frac{1}{n} \left(\frac{y_1}{z_1} + \frac{y_2}{z_2} + \ldots + \frac{y_n}{z_n} \right)$$

Then Y_{pps} is an unbiased estimate of Y, with variance

$$V(\bar{Y}_{pps}) = \frac{1}{n} \sum_{i=1}^{N} z_i \left(\frac{y_i}{z_i} - Y \right)^2.$$

Prove this result.

5. (i) Deduce the variance of the mean of a systematic sample in the following form :

$$V(\bar{Y}_{sy}) = \frac{S^2}{n}\left[\frac{(N-1)}{N} + (n-1)\rho_w\right].$$

where

$$\rho_w = \frac{2}{Kn(n-1)S^2}\sum_{i=1}^{k}\sum_{j<u}(y_{ij}-\bar{Y})(y_{iu}-\bar{Y})$$

and

$$S^2 = \frac{1}{N-1}\sum_{i}\sum_{j}(Y_{ij}-\bar{Y})^2.$$

(ii) If the variates Y_i $(i = 1, 2, .., N)$ are drawn at random from a super-population in which

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

$$\varepsilon(y_i-\mu)^2 = \sigma_i^2 : \ \text{Then}$$

$$\varepsilon V_{sy} = \varepsilon V_{ran}.$$

In deducing the above result, it will have to be remembered that the symbol ε denotes averages over all finite populations which can be drawn from the super-population and that V_{sy} and V_{ran} denote respectively the variances of the systematic sample and the random sample.

6. (i) Show that the bais of $r = u/w$ as an estimate of $R = U/W$, where u and w are random variables and $Eu = U$, $Ew = W$, is given approximately by

$$R(V^2_w - \rho_{uw} V_u V_w)$$

(i) Write a short critical note on the performance of systematic sampling as against that of random sampling in auto correlated populations.

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. Write notes on *any five* of the following :
   (i) Stratified random sampling,
   (ii) Sampling on successive occasions,
   (iii) Multiphase sampling,
   (iv) Systematic sampling after suitable rearrangement of frame,
   (v) Self weighting sample,
   (vi) Non-response in sample surveys.

2. What is meant by 'sampling frame'? What are the defects to which a sampling frame is likely to be subject and the possible reasons for such defects? What are the possible steps that you can think of to rectify the defects?

3. You are asked to carry out a sample survey to study the level of living of the working class population in an industrial area. What are the various problems which you should consider while planning the survey?

4. What, in your opinion, should be the nature of information to be presented in a report on a sample survey? Give an outline of the headings of sections which you would include in the report.

5. Give a brief history of the National Sample Survey conducted in India, mentioning the year since it started, periodicity, subject coverage, the likely users of the results and the nature of use. If the data on subjects are being collected repeatedly, discuss the utility or otherwise of such repetitions.

6. In analysing the data from a large scale sample survey, what are the different methods of checking systems that you can think of at different stages of analysis to ensure accuracy? Discuss their relative advantages and disadvantages.

7. While framing a schedule for a survey on consumer expenditure of house holds, there are two views. One view is that the items on which information is needed should be detailed out exhaustively and printed on the schedule. The other view is that only balnk space should be left so that information may be recorded for such items that are relevent to the household. Discuss the relative merits of the two views from the point of view of data collection and tabulation. What procedure would you recommend, one of the above or a modified procedure?

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

Time: 4 Hours Full marks: 100

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

1. Seven roasts can be cut from each of 8 animals. The experimenter whises to study the effect of 14 non-interacting treatments on the tenderness of roasts. Give the layout plan of the experimont and discuss in detail the method of analysis you would adopt. Also, give the paramotors of the design and write down the subdivision of the degrees of freedom in the analysis of variance. What is the efficiency factor of this design?

2. (a) What is the physical significance of the concept of orthogonality of two effects in a design? When is a design said to be orthogonal? Give examples of situations in which (i) a design laid out to be orthogonal becomes non-orthogonal, (ii) an experiment may be suitably planned to be orthogonal though commonly non-orthogonal, (iii) two or more effects are inevitably likely to be non-orthogonal and consequently, a more complicated procedure of analysis has necessarily to be employed to separate the effects.

(b) In a $t \times t$ Latin square design, if the expectation of the yield y_{ijk} of the plot in the i-th row, j-th column and receiving the k-th treatment can be put in the form

$$\mu+\rho_i+\gamma_j+\tau_k,$$

where

$$\sum_{i=1}^{t} \rho_i = 0, \quad \sum_{j=1}^{t} \gamma_j = 0, \quad \sum_{k=1}^{t} \tau_k = 0,$$

and Var. (y_{ijk}) σ^2, estimate μ, ρ_i, γ_j and τ_k, stating the principles on which your method of estimation rests. Show that in this design, the degrees of freedom for rows, columns and treatments are orthogonal to one another.

3. (a) In a symmetrical factorial experiment in which each factor is at two levels, and in which each block is of size 2^r, what is the maximum number of factors which can be accommodated so that no degrees of freedom belonging to a main effect or first order interaction are confounded?

(b) In a confounded factorial design with 7 factors A, B, C, D, E, F and G at two levels each, tested in blocks of 16 plots, the contents of one of the blocks are g, bdeg, ad, abe, be, cde, abcdg, aceg, ef, bdf, adefg, abfg, beefg, edfg, abcdef, acf. Give the contents of the remaining 7 blocks and determine which of the interactions have been confounded.

If the experiment is carried out in only a single replication, discuss the method of analysis and give the partition of the degrees of freedom in the analysis of variance.

4. (a) What are the practical situations under which it becomes necessary to resort to fractional replication in a factorial design? Why is the device limited in its usefulness and what precautions have to be taken to ensure that an experiment with fractional replication does not become valueless from the practical point of view?

(b) The following is the layout plan for a 1/4 replicate of a 2^7 factorial experiment with factors, A, B, C, D, E, F and G, each at two levels:

Block I	Block II	Block III	Block IV
defg	acfg	ceg	abcdg
aeg	bdg	ae	abcef
abceg	cef	bdf	fg
(I)	abde	acdefg	acf
abcdf	abfg	beg	befg
bcdefg	acde	abdefg	adg
adf	bef	ab	de
be	cdg	cdf	bcde

Determine the defining contrasts and the interactions which have been confounded. Assuming that second and higher order interactions are non-existent, give the partition of the degrees of freedom in the analysis of variance.

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

(a) Balanced lattice squares,

(b) Qualitative-cum-quantitative experiments,

(c) Strip-plot designs,

(d) Linear response.

Paper IV and V—Probit Analysis (Theoretical)

Time: 4 Hours | Full marks: 100

(a) Attempt *any four* questions.
(b) Figures in the margin indicate ful marks.
(c) Credit will be given for neat and brief answers.

1. (a) What are the important considerations involved in the choice of dose and response-metameters in biological assays? (15)

(b) Give your critical comments on the comparative usefulness of the probit and logit transformations. (10)

2. (a) Explain what is meant by 'tolerance distribution' and 'median effective dose.' (10)

(b) The tolerance distribution for a certain insecticide is normal and the values of LD 30 and LD 60 are respectively

$$\text{LD } 30 = 0.62$$
$$\text{LD } 60 = 1.58$$

Using the following table of Probits, find the value of LD 80. (15)

Per cent kill	60	70	80
Probit	5.253	5.524	5.842

3. k different dosages $x_1, x_2, .., x_k$ of an insecticide were tried independently of k different batches of $n_1, n_2, .., n_k$ insects and the numbers killed were found to be $r_1, r_2, .., r_k$ respectively. If the expected rate of mortality at a dosage x is given by

$$P_x = 1 - exp\ (-x/\theta)$$

where θ is an unknown parameter, devise a routine scheme for computing the maximum likelihood estimate of θ and its standard error. (25)

4. (a) Define 'relative potency' of one insecticide with respect to another and examine the assumptions under which it is meaningful. (15)

(b) The relative potency of an insecticide A with respect to another insecticide B is ρ and their action in a mixture is 'similar'. Under suitable assumptions to be carefully stated, derive a formula for the percentage kill P in terms of the dosage x of a $1 : r$ mixture of the two insecticides A and B. (10)

5. (a) Explain what is meant by the 'synergistic action' of a mixture of two poisons. (10)

(b) Describe several statistical models for synergistic action and discuss the problem of estimation of the parameters involved in *one* such model. (15)

6. Describe the Robbins-Monro sequential procedure for determining the impulse which would produce a pre-determined response and state (without proof) the advantages and disadvantages of such a procedure.

Give your critical comments on the usefulness of this procedure in estimating the median lethal dose of an insecticide. (25)

7. Write an essay on the use of Probit analysis in fields other than that of biological assays. (25)

PAPER IV AND V—MATHEMATICAL THEORY OF SAMPLING DISTRIBUTIONS

(THEORETICAL)

Time: 4 Hours | Full marks: 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

1. (a) Obtain the distribution of the mean of a sample of n observations for a rectangular population with the aid of characteristic functions.

(b) If x and y have continuous frequency function $f(x, y)$, their characteristic function is

$$\phi(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} exp\,(iux + ivy)\, f(x, y)\, dx\, dy.$$

Show that the distribution of x when y is given has a characteristic function

$$\phi(u/y) = \frac{\int_{-\infty}^{+\infty} e^{-ivy}\, \phi(u, v)\, dv}{\int_{-\infty}^{+\infty} e^{-ivy}\, \phi(0, v)\, dv}.$$

2. Show that in samples from a normal bivariate population

$$dF \propto exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right\}\right] dx\, dy$$

the functions $$u_j = \frac{x_j}{\sigma_1} + \frac{y_j}{\sigma_2},\ v_j = \frac{x_j}{\sigma_1} - \frac{y_j}{\sigma_2}$$

are distributed independently and that their correlation coefficient R may be written

$$R = \frac{a-a}{\sqrt{\{(a+a)^2 - 4\,aar^2\}}}$$

where $$a = \frac{\sigma_1^2}{\sigma_2^2},\ a = \frac{\Sigma(x-\bar{x})^2}{\Sigma(y-\bar{y})^2}$$ and r is the correlation between the observed x's and y's.

Hence show that

$$t = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} = \frac{(a-a)\sqrt{n-2}}{\sqrt{4(1-r^2)\,aa}}$$

is distributed as student's t with $n-2$ degrees of freedom.

3. Obtain the joint distribution of the mean and the standard deviation for samples from a normal population. Hence derive the distribution of f and its first two cumulants.

4. Explain the relationship between Hotellings T^2 and Mahalanobis' D^2-statistic. Derive the distribution of the D^2-statistic.

5. For a trivariate normal distribution, discuss in detail the distribution of the maximum root of the determinantal equation.

Paper IV and V—Genetics (Theoretical)

Time: 4 Hours | Full marks: 100

(a) Answer *any five* questions.

(b) All questions carry equal marks.

1. Let the series of alleles at a locus be S_1, S_2, S_3. These are self-sterile, i.e. pollen grains containing allele S_j cannot function on plants which contain this allele. A plant with genotype $S_1 S_2$ can be fertilized only by S_3 and produces $S_1 S_3$ and $S_2 S_3$ individuals in equal numbers. Starting with an initial population

$$x_0 S_1 S_2 + y_0 S_1 S_3 + z_0 S_2 S_3$$

show that

$$x_n = \tfrac{1}{2}\{1-(-\tfrac{1}{2})^n\} + (-\tfrac{1}{2})^n x_0$$

What will be the composition of the equalibrium population and the nature of equilibrium?

2. (a) Describe the genetics of O—A—B—AB system. Under conditions of random-mating, how are the frequencies of the different blood groups in the offspring determined by the gene-frequencies in the parental generation? How are these gene-frequencies estimated from observed data?

(b) In a case of disputed parentage, two babies were of type MN and N respectively. Their mothers also were of types MN and N but it was uncertain to which mother either baby belonged. The husband of women N was of types M. To which mother did the type N belong?

3. (a) Consider a six-linked character in which selection is practiced only in the homogametic sex and suppose a proportion k of the recessives are discarded. Evaluate the progress of the population.

(b) Suppose a recessive triat occurs in 1 in 1000 of a random mating population. How many generations of complete selection against the recessive individuals would be necessary to reduce the proportion to 1 in 1000,000?

4. The following data relate to three pairs of allelomorphs A, a; B, b; C, c; occurring at three different loci Scute, Beaded and Rough respectively on the sex-chromosome of Drosophila. Numbers of cross-overs for double back-cross are given below:

	number of observations	number of cross-overs
Scute to Beaded	260	5
Beaded to Rough	450	12
Scute to Rough	6300	450

How will you analyse the data to find out if this is consistent with the hypothesis of linear arrangement of genes on the chromosome? (You need not actually compute, but you have to formulate the statistical hypothesis to be tested and describe the procedure of testing this hypothesis).

5. Define coefficient of inbreeding. Consider a population segregating for two alleles A and a with frequencies p and $q = 1 - p$ respectively and let f be the coefficient of inbreeding in the population. Show that the frequency distribution of genotypes is given by

$$(p^2+fpq)\,AA + 2(1-f)\,pq\,Aa + (q^2+fpq)\,aa$$

6. For a population segretating for two alloles A and *a*, evaluate the effect of continued sib-mating in the frequencies of different mating types.

How can one use these methods of mating between near relatives to improve yield of plants?

7. (a) Suppose we wish to distinguish between *AA* and *Aa* sires and the genotype *aa* is lethal. We have *Aa* dams for test purposes. How many progeny should be obtained from each sire and what would be the rule of classification to ensure approximate probabilities of 10 per cent of making either of the two possible errors?

(b) The following results are for the genes Br P in 3 back-cross tests by three investigators

investigators	Br P	Br P	br P	br P
1	57	77	112	64
2	51	81	74	46
3	70	71	76	44

Do these three tests agree with each other? (You need not actually compute, indicate the procedure).

Paper IV and V—Psychology and Education (Theoretical)

Time : 4 Hours Full marks : 100

Directions for Examinees : There are two parts in this paper. Each part carries 50 marks. Part I contains 20 questions each carrying equal marks. For each question in Part I write your answer in the space provided in the question paper itself. Note that only short answers are required for Part I. First answer Part I, then answer Part II. In Part II only five questions are given each carrying equal marks. You are to answer only *two* out of the five questions. Write your answer to questions in Part II on the separte answer book. In evaluating your answers to Part II, logical presentation and neatness will also be considered.

N.B.—In this print, space for answer below each question in Part I is not provided.

Part I

1. Define '*true*' score.
2. What is 'index of reliability' and what is its use?
3. What is meant by the 'communality' of a test?
4. What are the sources of unreliability of essay-type examination?
5. What is '*error*' in a test score?
6. Show how the method of analysis of variance can be used to estimate the *reliability* of a test.
7. For item-criterion correlation, when items are scored 0 or 1 and the criterion is a continuous variable, which of the following coefficients would you find and why?

(a) Bi-serial, (b) Point-biserial, (c) Phi, (d) Tetra-choric, (e) any other.

8. When the validity of a test is defined as its correlation against an external criterion, how can the *validity* be increased if the reliability of the test is held constant?

9. What are the *assumptions* of Kuder-Richardson Formula 20:

$$\frac{n}{n-1}\left(1-\frac{\Sigma pq}{\sigma_t^2}\right)$$

used in finding the reliability of a test?

10. How can the *reliability* of a test be estimated by the use of 'parallel' forms of the same test?

11. Why the split-half method of estimating the reliability of a test is not recommended for '*speeded*' tests?

12. It is possible to compare the *standard error of measurement* of different tests used on different groups but not their reliabilities. Why?

13. How should the items of a multiple choice *personality* test be analysed?

14. When the validity coefficients of several tests are known, how should the test scores be combined so as to *maximise the prediction of the criterion?*

15. When the criterion is complex (involving more than one dimension), how would you judge the *validity* of a battery of tests?

16. In any instance of psychological and educational measurement, what minimum conditions must be met or assumptions made in order to use Fisher's discriminant function?

17. If one group has taken a series of tests in one sequence and another group has taken the same series in a different sequence, is there any way of judging the *effect of sequence on test performance?*

18. When tests are used for selection and a small number of persons is selected, the estimate of validity of the tests based on the selected group's performance on some criterion is frequently an underestimate. Why?

19. In vocational guidance, a psychologist gives three tests, A, B and C to a person. Tests A, B and C are already known to measure aptitudes a, b and c which again are known to predict success in the vocations M, N and P. On the three tests, the person's scores are at the 75th, 70th and 65th percentiles on national norms. What statistical considerations should the psychologist have in giving vocational advice to the person?

20. How does sampling of persons and tests affect the factor-matrix?

Part II

Answer *any two* questions.

1. Propose a *design* of experiment to decide whether instruction through the medium of mother-tongue will be better in schools rather than in colleges. Indicate methods of *sampling, collection of data and statistical analysis.*

2. Discuss the problems in the *theory* and *methods* of centriod factor analysis.

3. Discuss the differences between the methods of *scaling attitudes* and those of objective *test-construction.*

4. Discuss the problems in the *standardization* of a projective technique of personality assessment.

5. Write an essay on the *uses* of non-parametric statistics in psychology and education.

PAPER VIII AND IX—ECONOMIC STATISTICS (PRACTICAL)

Time : 4 Hours Full marks : 100

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

1. The following table gives the seri distribution of income among personal income receipients in the United States in 1918 : (45)

income class ($)	number of persons (000)
0 — 500	1828
500 — 1000	12531
1000 — 1500	12498
1500 — 2000	5222
2000 — 3000	3065
3000 — 5000	1383
5000 — 10000	588
10000 — 25000	192
25000 — 50000	41
50000 —100000	14
100000 —200000	5
200000 and over	2

Consider the range $500 and over and obtain estimates for a and ν in the Pareto equation $y = a X^{-\nu}$ where y is the number of persons having income X or more.

Work out the expected frequencies.

From the value of the parameter ν or otherwise, obtain an estimate of the concentration ratio.

2. The following table gives the general index of industrial production and an index of freight turnover (net ton kilometers) for the period 1937–1956 in India :

year	index of industrial production	index of freight turnover
1937	74	77
38	78	75
39	76	80
1940	81	86
41	87	95
42	82	95
43	87	90
44	87	96
45	89	99
45	81	91
47	78	73
48	87	77
49	86	86
1950	85	92
51	95	98
52	98	98
54	100	100
54	107	109
55	115	122
1956	126	135

Obtain the coefficients of correlation between the two index numbers for the periods 1937–56 and 1937–50. (25)

3. The following table gives some facts about area and population of India :

Some facts about area and population of India, 1941

regions	area in square miles	population in million	density of population per square mile	percentage of population in	
				villages	towns
NORTHWEST INDIA	284342	39.3	123	79.1	20.9
Rajasthan	130207	15.3	117	82.7	17.3
Punjab	37378	12.6	338	81.0	19.0
Pepsu	10078	3.5	347	81.0	19.0
Jammu & Kashmir	92780	4.4	522	84.1	15.9
Part C States	13899	3.5	—	—	—
NORTH INDIA	113409	63.2	557	86.4	13.6
Uttar Pradesh	113409	36.2	557	86.4	13.6
EAST INDIA	261057	90.0	395	88.9	11.1
Assam	85012	9.0	176	93.4	6.4
Bihar	70330	40.2	572	93.3	6.7
Orissa	60136	14.6	244	95.9	4.1
West Bengal	30775	24.8	806	75.2	24.8
Part C States	15404	1.4	—	—	—
CENTRAL INDIA	289399	52.2	181	84.3	15.7
Madhya Pradesh	130272	21.2	163	86.5	13.5
Madhya Bharat	46478	7.9	171	81.9	18.1
Hyderabad	82168	18.7	227	81.4	18.6
Part C States	30481	4.4	—	—	—
WEST INDIA	149609	40.7	272	68.8	31.2
Bombay	111434	36.0	323	67.9	31.1
Saurashtra	21451	.1	193	66.3	33.7
Part C States	16724	0.6	—	—	—
SOUTH INDIA	168009	75.6	450	70.0	30.0
Madras	127790	57.0	446	80.4	19.6
Mysore	29489	9.1	308	76.0	24.0
Travancore & Cochin	9144	9.3	1015	84.0	16.0
Part C States	1568	0.2	—	—	—
ALL INDIA	1269640*	361.0	312	82.7	17.3

*including Andaman and Nicobar islands, area 3125 square miles.

Sources : Census of India 1951, Part IIA. Census of India, Paper No. 9, 1952 and Census of India Paper No. 7, 1954.

(a) Check the area and population figures and correct obvious mistakes, if any.

(b) Recompute the figures for densities and replace the incorrect figures by correct figures.

(c) Correct all other obvious mistakes, if any, and indicate doubtful figures, if any, where correction is not possible.

(d) Copy the corrected table neatly.

(e) Obtain the estimates of densities for the aggregates of Part A, Part B and Part C States. (30)

PAPER VIII AND IX—STATISTICAL QUALITY CONTROL (PRACTICAL)

Time : 4 Hours | Full marks : 100

(a) Attempt *all* the questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. The table below gives the number of missing rivets noted in the final assembly of twelve motor-boats.

1.	9	7.	11
2.	15	8.	8
3.	13	9.	20
4.	18	10.	12
5.	11	11.	21
6.	15	12.	16

Examine the data for statistical control. Assuming that for the assembling process an average of 10 for the number of missing rivets was specified, estimate the percentage of abnormal cases that would result if the quality is such as is revealed by the data.

2. A producer wishes to control the resistance of a certain product at the average level of 120 ohms. From each of 10 consecutive batches, a random sample of 5 units were selected and tested. The mean and standard deviations calculated for each sample, are given below. (In some cases all the units in the sample did not go through the test.)

Analyse and interpret the data by plotting suitable control charts.

sample	number of items	mean ($\bar{x}$)	standard deviation (s)
1	5	124.6	7.75
2	5	123.4	12.20
3	4	145.8	9.36
4	5	132.7	8.25
5	3	114.6	11.52
6	5	131.6	13.61
7	5	107.8	12.32
8	5	123.9	10.08
9	4	150.3	9.25
10	5	151.2	8.36

3. With the help of necessary tables,

(i) obtain a point-estimate of the lot standard deviation, given that in 4 samples each of 5 items, the observed ranges for a measurement were 8.6, 9.2, 7.5 and 9.3 cms.

(ii) obtain 95 per cent confidence limits for the lot standard deviation, given that in a sample of 10 packets of margarine, the standard deviation of percentage moisture, was 4.3.

(iii) construct 95 per cent confidence limits for average number of defects per item in lot, given that in a sample of 20 items inspected a total number of 22 defect were noted.

4. Set up a chart (for graphical procedure) *or* table (for tabular procedure) suitable for item-by-item sequential sampling inspection by measurement, such that lots with a mean value 5.6 for a normally distributed characteristic will be accepted 95 per cent of the times and lots with mean value 7.5 for the characteristic will be accepted only 10 per cent of the times, it being known that the standard deviation of the characteristic is stable with value 2.2.

Paper VIII and IX—Sample Surveys, Theory (Practical)

Time : 4 Hours — Full marks : 100

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

1. The table given below shows the total cultivated area during 1931 as also the area under wheat in two consecutive years 1936, 1937 for a sample of 34 villages in a sub-division. The villages were selected with replacement with probability proportionate to the cultivated area as recorded in 1931. The total cultivated area in 1931 and the total area under wheat in 1936 for all the 170 villages of the sub-division were known to be 78000 and 21288 acres respectively. Estimate the area under wheat for the sub-division for the year 1937 using the ratio method of estimation and calculate the standard error of the estimate so made.

What would be the standard error of the estimate if the information for the previous year were not used? Find out also the increase in efficiency, if any, in using the previous year's information. (60)

Values of total cultivated area and of area under wheat in two consecutive years for a sample of 34 villages in a sub-division

serial number of village	total cultivated area in 1931 (acres)	area under wheat (in acres)	
		1936	1937
1	401	70	50
2	630	163	149
3	1194	320	284
4	1170	440	381
5	1065	250	278
6	827	125	111
7	1737	558	634
8	1060	254	278
9	360	101	112
10	946	359	355
11	470	109	99
12	1625	481	498
13	827	125	111
14	96	5	6
15	1304	427	339
16	377	78	80
17	259	75	105
18	186	45	27
19	1767	564	515
20	604	238	249
21	700	92	85
22	524	247	221
23	571	134	133
24	962	131	144
25	407	129	103
26	715	190	175
27	845	663	335
28	1016	235	219
29	184	73	62
30	282	62	79
31	194	71	60
32	439	137	100
33	854	196	141
34	820	255	263

2. The table given below presents the summary of data for complete census of all the 340 villages in a sub-division. The villages were stratified by size of their agricultural area into 4 strata as shown in column 2 of the table. The numbers of villages in the different strata are given in column 3. The population values of the strata means for the area under wheat ($\bar{Y}_{N_i}$) and those of the standard deviations for the area under wheat (S_{w_i}) and for the agricultural area (S_{a_i}) are given in the subsequent columns.

Calculate the sampling variance of the estimated area under wheat for a sample of 34 villages.

(i) if the villages are selected by the method of simple random sampling without stratification,

(ii) if the villages are selected by the method of simple random sampling within each stratum and allocated in proportion to (a) the sizes of the strata (N_i), (b) the products $N_i S_{w_i}$ and (c) the products $N_i S_{a_i}$. (40)

Strata Means and standard deviations of Areas for villages in a sub-division.

stratum number	size of village in bighas	N_i	Y_{N_i}	S_{w_i}	S_{a_i}
(1)	(2)	(3)	(4)	(5)	(6)
1	0— 500	63	115.5	55.8	129.8
2	501—1500	198	275.6	115.8	265.7
3	1501—2500	54	555.0	195.5	274.5
4	> 2500	25	965.5	360.4	975.7

Or,

A simple random sample of 30 households was drawn from a census taken in 1951 in two wards of a town. The population contains about 15000 households. In the table the persons in each household are classified (i) as to whether they had consulted a doctor in the past 12 months, (ii) as to sex.

Make a critical comparison of the ratio formula with the binomial formula in the context of estimating the proportion of people who had consulted a doctor as also the proportion of males in the population. (40)

Data for a simple random sample of 30 households.

household number	number of persons	number of persons		doctor seen in last year	
		males	females	yes	no
(1)	(2)	(3)	(4)	(5)	(6)
1	5	1	4	5	0
2	6	3	3	0	6
3	3	1	2	2	1
4	3	1	2	3	0
5	2	1	1	0	2
6	3	1	2	0	3
7	3	1	2	0	3
8	3	1	2	0	3
9	4	2	2	0	4
10	4	3	1	0	4
11	3	2	1	0	3
12	2	1	1	0	2
13	7	3	4	0	7
14	4	3	1	4	0
15	3	2	1	1	2
16	5	3	2	2	3
17	4	3	1	0	4
18	4	3	1	0	4
19	3	1	2	3	0
20	3	1	2	3	0
21	4	1	3	2	2
22	3	2	1	0	3
23	3	2	1	0	3
24	1	0	1	0	1
25	2	1	1	2	0
26	4	3	1	2	2
27	3	1	2	0	3
28	4	2	2	2	2
29	2	1	1	0	2
30	4	2	2	1	3

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. You are given a list of 80 tehsils in a State along with their population in 1951 and net area sown in the same year. Draw a sample of 10 tehsils with probability proportional to population with replacement and estimate the 'total net area sown' in the State taking the figures of net area given against the selected tehsils as having been obtained by an enquiry. What is the variance of this estimate ? What should be the sample size if the estimate is required with a confidence limit of 98 per cent ? Suitable reference to tables used should be given.

2. A random sample of 40 cuts, each of size 100 sq. ft. are taken from a locality for estimating the average yield rate of dry paddy. For a sub-sample of 20 cuts out of this, both the green weight and dry weight have been recorded. For the remaining 20 cuts, only the green weights were recorded. Estimate the average yeild rate of dry paddy and its variance by the following two methods.

(i) taking the sub-sample of 20 cuts for which the dry weights are available as a simple random sample; (ii) ratio estimate. Compare the relative efficiency of the two methods. Assuming that the cost of taking green weight for a cut as (c) and that for taking dry weight from a cut as (.25c), what will be the percentage increase in cost over that of the present sample, to get as good an estimate as that of the ratio estimate by taking a simple random sample of size 'n' for dry weight ?

serial number of cut	green weight (tolas)	dry weight (tolas)	serial number of cut	green weight (tolas)
1	167	153	21	86
2	127	119	22	115
3	189	174	23	114
4	138	126	24	145
5	114	105	25	179
6	110	102	26	83
7	124	113	27	86
8	173	159	28	145
9	140	131	29	120
10	118	109	30	78
11	133	124	31	88
12	134	124	32	110
13	82	75	33	129
14	138	125	34	104
15	145	134	35	141
16	144	134	36	126
17	170	163	37	118
18	148	136	38	154
19	114	104	39	170
20	139	128	40	107

3. Work out a suitable schedule with appropriate headings for a study of housing condition of industrial labourers in a locality. Indicate the scope of the instructions which you would give to investigators for data collection.

4. Enclosed is a schedule on household indebtedness used in survey for studying the indebtedness of agricultural population. Draw up blank specimen tables with suitable tabular and columnar headings that you would think of presenting in a short report of the survey.

Net area sown and total population in 80 tehsils of a locality in 1951
[data for question 1]

tehsil number	total population (000)	net area sown (000 acres)	tehsil number	total population (000)	net area sown (000 acres)
1	160	46	41	73	79
2	141	65	42	75	69
3	112	63	43	98	118
4	90	57	44	172	108
5	110	85	45	170	128
6	107	87	46	135	80
7	136	93	47	125	81
8	131	72	48	65	48
9	93	65	49	101	93
10	169	138	50	108	86
11	98	62	51	132	90
12	132	78	52	117	83
13	94	78	53	70	51
14	85	61	54	138	176
15	105	69	55	125	153
16	75	53	56	49	64
17	70	70	57		143
18	23	18	58	146	167
19	127	90	59	63	90
20	79	32	60	92	120
21	90	62	61	87	128
22		62	62	97	180
23	186	133	63	132	87
24	131	100	64	34	48
25	79	59	65	141	134
26	126	162	66	70	78
27	96	128	67	53	73
28	92	95	68	86	66
29	126	127	69	79	50
30	94	90	70	120	84
31	86	79	71	96	88
32	135	98	72	110	142
33	95	99	73	82	98
34	108	83	74	41	34
35	74	71	75	20	18
36	49	50	76	64	63
37	97	42	77	65	61
38	80	43	78	72	38
39	102	140	79	38	31
40	85	117	80	60	38
total				7878	6798

Household indebtedness

(Reference Period : July 1953–June 1954)

(Rural)

(1) Identification particulars of sample village and household

0. serial number	1. zone-state-natural division
2. stratum	3. state
4. district	5. tehsil/taluk/thana
6. village	7. head of household
8. informant	9. household size
10. livelihood class*	11. area owned (0.00 acres)
12. area leased out (0.00 acres)	13. area leased in (0.00 acres)
14. area in possession of household (0.00 acres)	

(2) Particulars of loan taken during the year and outstanding loans on 1/7/53. (each transaction in separate columns)

details of loan	loand during year				outstanding loands			
	loan no.	loan no.	loan no.	loan no.	loan no.	loan no.	loan no.	loan no.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. amount in Rs. (00) (imputed value if kind)								
2. month and year in which loan taken								
3. source of laon (a)								
4. purpose of loan (b)								
5. loan : secured (1); not secured (2)								
6. if secured, security code (c)								
7. loan : in cash (1); in kind (2)								
8. rate of interest % per year Rs. (00.0)								
9. interest simple (1); compound (2)								
10. time interval (months) for compounding								
11. mode of repayment : cash-1; crop-2; service-3								

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(9)
12. actual use of loan (b)** (i)							
(ii)							
(iii)							
(iv)							
13. amount of interest paid during the year (Rs. 0.00)							
14. amount of interest paid in earlier years (Rs. 00)							
15. amount of loan repaid during the year (Rs. 00)							
16. amount of loan repaid during earlier years							
17. number of instalments, repaid so far.							
18. source of funds from which repayment made (d) (i)							
(ii)							
19. if source is a secured loan, security code (c)							

Footnotes :

* cultivating owners of land–1; non-owning cultivators–2; non-cultivating owners–3; others including agricultural labourers–4.

(a) source of loan : banks : scheculed–1; non-scheduled–2; insurance companies–3; co-operative societies : registered–4; unregistered–5; money lenders–6; owner of land–7; taccavi laon–8; other government laons–9; others–0.

(b) purpose and use of loan : agriculture : capital account–1; current account–2; marketing–3; other enterprises : capital account–4; current account–5; marketing–6; domestic consumption : ceremonial occasions–7; normal maintenance–8; construction or purchase of residential real assets and other durable assets for domestic use (e.g. land, building, furniture, ornaments)–9; repayments of loan–0.

(c) security code : land–1; building–2; ornaments–3; other goods and equipments–4; others–5.

(d) source of fund for repayment : current savings from : agriculture–1; other enterprises–2; past savings from : agriculture–3; other enterprises–4; sale of assets : land–5; building–6; ornaments–7; other household equipments–8; new loans : without security–9; with security–0.

** Write code number in smaller cell and the amount (Rs. 00) in the larger cell.

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

Time : 4 Hours | Full marks : 100

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

1. The table below gives the plan and yields in lbs. per plot of a 3^3 NPK experiment on turmeric (*Curcuma longa*) laid out in three randomised blocks of nine plots each, the treatments being all combinations of :

$$\left\{\begin{array}{l} n_0 = \text{no nitrogen} \\ n_1 = 60 \text{ lbs. } N/\text{acre} \\ n_2 = 120 \text{ lbs. } N/\text{acre} \end{array}\right\} \times \left\{\begin{array}{l} p_0 = \text{no superphosphate} \\ p_1 = 45 \text{ lbs. } P_2O_5/\text{acre} \\ p_2 = 90 \text{ lbs. } P_2O_5/\text{acre} \end{array}\right\}$$

$$\times \left\{\begin{array}{l} k_0 = \text{no potash} \\ k_1 = 100 \text{ lbs. } K_2O/\text{acre} \\ k_2 = 200 \text{ lbs. } K_2O/\text{acre} \end{array}\right\}$$

Plan and yields of raw turmeric in lbs. per plot.

(Plot size 1/96 acre)

Block I	Block II	Block III
$n_1p_1k_2$ 114	$n_1p_0k_2$ 79	$n_0p_2k_1$ 73
$n_0p_2k_2$ 90	$n_2p_0k_0$ 76	$n_1p_0k_0$ 74
$n_2p_1k_0$ 76	$n_0p_0k_1$ 57	$n_2p_0k_1$ 66
$n_0p_1k_1$ 68	$n_2p_2k_2$ 97	$n_2p_1k_2$ 106
$n_1p_2k_0$ 105	$n_1p_2k_1$ 88	$n_0p_0k_2$ 66
$n_2p_2k_1$ 95	$n_0p_1k_2$ 75	$n_2p_2k_0$ 107
$n_1p_0k_1$ 91	$n_0p_2k_0$ 71	$n_1p_2k_2$ 88
$n_2p_0k_2$ 81	$n_1p_1k_0$ 76	$n_0p_1k_0$ 65
$n_0p_0k_0$ 92	$n_2p_1k_1$ 76	$n_1p_1k_1$ 84

Analyse the data as fully as you can and state your conclusions.

2. The following are the plan and yields in lbs. per plot of 13 varieties of corn in a yield trial, the upper figures within brackets being the varieites and the lower figures the yields in lbs. per plot :

Block I	(3) 25.1	(6) 19.8	(9) 28.7	(11) 24.2
Block II	(3) 22.7	(4) 19.4	(8) 33.2	(12) 22.6
Block III	(10) 16.1	(11) 19.0	(12) 31.8	(13) 26.9
Block IV	(2) 27.1	(5) 27.2	(8) 35.4	(11) 17.1
Block V	(7) 23.3	(8) 30.4	(9) 30.7	(10) 32.2
Block VI	(4) 30.3	(5) 32.5	(6) 27.1	(10) 32.6
Block VII	(1) 34.3	(5) 31.0	(9) 25.8	(12) 30.6
Block VIII	(3) 34.2	(5) 32.5	(7) 33.1	(13) 36.5
Block IX	(1) 38.4	(2) 32.7	(3) 37.5	(10) 31.2
Block X	(2) 28.6	(4) 30.4	(9) 26.7	(13) 35.2
Block XI	(1) 36.5	(4) 31.2	(7) 31.3	(11) 28.5
Block XII	(1) 31.9	(6) 33.6	(8) 27.7	(13) 41.2
Block XIII	(2) 30.3	(6) 31.6	(7) 39.5	(12) 26.8

Identify the layout adopted, giving its characteristics. Analyse the data and interpret the results of analysis.

3. In a randomised block experiment for teting the difference among seven varieties A, B, C, D, E, F and G of gunyule, the following table gives the layout plan adopted and the data on resin percentage and shrub weight (gms) for a randomly

selected plant in each plot, the upper figure denoting the resin percentage and the lower figure within brackets the shrub weight in gms.

Block I	B 5.24 (61)	F 4.85 (84)	G 5.99 (86)	E 3.97 (34)	D 5.50 (65)	A 4.49 (88)	C 5.74 (39)
Block II	E 5.71 (58)	G 5.15 (89)	A 5.15 (28)	F 4.86 (67)	D 6.00 (90)	B 4.49 (142)	C 6.15 (88)
Block III	G 5.88 (58)	C 4.88 (61)	F 5.82 (125)	D 4.75 (125)	B 5.60 (95)	A 6.22 (111)	E 6.05 (87)
Block IV	B 6.13 (70)	E 5.64 (101)	A 7.24 (115)	D 6.30 (101)	G 5.36 (99)	C 6.20 (127)	F 5.17 (104)
Block V	G 5.80 (93)	B 5.78 (98)	D 5.85 (8)	A 5.67 (132)	F 5.86 (144)	E 3.85 (105)	C 6.18 (169)

Perform the anlysis of variance and covariance and compare the varieties for resin percentage adjusted for shrub weight. Find the standard error of the difference between the adjusted mean resin percentages for two varieties.

4. In an experiment on the marketing of apples, four treatments A, B, C and D were compared in four stores on four days in each of two consecutive weeks. The plan of the experiment and the data, pounds of apples sold per 100 customers are as under :

day of week	first week				second week			
	store number				store number			
	1	2	3	4	1	2	3	4
Monday	B 8	A 28	D 58	C 34	D 44	C 68	B 40	A 34
Tuesday	C 52	D 112	B 68	A 60	B 16	A 48	C 100	D 94
Wednesday	D 56	C 52	A 40	B 42	C 20	D 40	A 28	B 32
Thursday	A 28	B 20	C 60	D 52	A 45	B 44	D 64	C 25

Analyse the data and state your conclusions, presenting the results in the form of a summary table.

Paper VIII and IX—Probit Analysis (Practical)

Time : 4 Hours | Full marks : 100

(a) Attempt *any two* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.
(d) Credit will be given for neat, brief answers, tabular arrangement of computations and neat graphical presentation of data.

1. The results of an experiment to assess the effect of an insecticide are given below, in which n stands for the number of insects used, r for the number of insects killed and x for the logarithm (base 10) of the dose in milligrams of the insecticide.

x	n	r
control	105	11
0.648	100	43
1.321	148	80
1.680	120	70
2.148	128	85
3.805	102	91

Estimate the dose required to kill 80 per cent of a large poulation of insects subject to a natural mortality rate of 15 per cent. Find also the standard error of your estimate.

2. A biological assay (of the quantal type) is to be carried out to estimate the mean and the standard deviation of a tolerance distribution, assumed to be normal. It has been decided to use only five different dosages x_1, x_2, x_3, x_4 and x_5 and the same number n of subjects at each dosage. From previous experience, approximate values μ and σ of the mean and the standard-deviation are available.

Compare the effectiveness of the following two schemes of choosing the dosages (by computing the dispersion-matrix of the estimates of the mean and the standard deviation).

	scheme I	scheme II
x_1	$\mu + \sigma$	$\mu + 1.5\,\sigma$
x_2	$\mu + 0.5\,\sigma$	$\mu + 0.8\,\sigma$
x_3	μ	μ
x_4	$\mu - 0.5\,\sigma$	$\mu - 0.8\,\sigma$
x_5	$\mu - \sigma$	$\mu - 1.5\,\sigma$

3. Two poisons P_1 and P_2 act independently, one being a stomach poison and the other a contact poison. The regression equations of the probit (Y) of the percentages killed on (X) the logarithm of the dose in milligrams for the poisons are as follows :

P_1 : $Y = 3.25 + 1.25X$
P_2 : $Y = 4.25 + 0.75X$

Write down the expression for the percentage kill when a mixture of w_1 mg of P_1 and w_2 of P_2 is used.

Draw a graph from which one can read-off the different combinations of (w_1, w_2) that would ensure on average kill of (a) 50 per cent, (b) 60 per cent, (c) 75 per cent.

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

Time : 4 Hours | Full marks : 100

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

1. The means of sepal length x_1, sepal width x_2, petal length x_3 and petal width x_4 of fifty specimens of flowers from each of two species are as follows :

Means of specimens in cms.

variate	specimens	
	1	2
x_1	5.936	5.006
x_2	2.770	3.428
x_3	4.260	1.462
x_4	1.326	0.246

The sums of squares and products about the means are noted below :

	x_1	x_2	x_3	x_4
x_1	19.14	9.04	9.76	3.24
x_2	9.04	11.86	4.62	2.47
x_3	9.76	4.62	12.30	3.88
x_4	3.24	2.47	3.88	2.46

Examine as best as you can whether the difference between the means of the two samples is significant or not.

2. (a) Examine whether the correlation coefficients given below are significantly different from each other :

(1) 0.896 (50) (2) 0.756 (60)
(3) 0.689 (85) (4) 0.834 (65)
(5) 0.567 (90) (6) 0.798 (95)

The values shown in the bracket give the number of observations on which the correlations are based.

(b) Obtain the multiple correlation coefficient and the regression equation of the following data of Y on X_1 and X_2 and test their significnce by finding the probabilities :

x_1	x_2	y
0.4	53	64
0.4	23	60
3.1	19	71
0.6	34	61
4.7	24	54
1.7	65	77
9.4	44	81
10.1	31	93
11.6	29	93
12.6	58	51
10.9	37	76
23.1	46	96
23.1	50	77
21.6	44	93
23.1	56	95
1.9	36	54
26.8	58	168
29.9	51	99

3. Draw 12 samples each of size 8 from a bivariate normal population where $\rho = .8$, $\sigma_1 = \sigma_2 = 1$ and $\mu_1 = \mu_2 = 5$ cms. and examine how many of the observed correlations for the samples lie within the 5 per cent level.

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

Time : 4 Hours Full marks : 100

(a) Attempt *all* questions.

(b) All questions carry equal marks.

(c) Use of caluclating machines is permitted.

1. The following table gives the female population of USA by age groups and the age specific death rates for females due to all causes, T.B., and heart diseases for the years 1900 and 1940. Evaluate from the above data the change in the mortality pattern with respect to age and specific causes.

TABLE

age group (years)	female population in thousands		age specific death rate due to all causes per 1000		age specific death rate due to T.B. (all forms) per 100000		age specific death rate due to heart diseases per 100000	
	1900	1940	1900	1940	1900	1040	1900	1940
under 1 year	948	993	145.4	47.7	311.6	24.6	147.8	17.5
1— 4	3589	4193	19.1	2.7	101.8	12.3	15.0	3.7
5—14	8392	11059	3.9	0.9	36.2	5.5	23.3	8.0
15—24	7516	12049	5.8	1.8	205.7	38.2	28.8	14.0
25—34	5861	10818	8.2	2.7	294.3	56.3	43.4	29.7
35—44	4339	9168	9.8	4.5	253.6	59.4	80.8	91.7
45—54	2905	7550	14.2	8.6	215.6	66.3	173.0	279.5
55—64	1940	5163	25.8	18.1	223.0	76.1	414.1	713.5
65—74	1069	3209	53.6	41.9	256.1	80.8	957.3	1723.5
75—84	388	1187	118.8	104.5	279.3	80.4	1751.9	4233.7
85 & over	68	217	255.2	218.1	204.5	62.0	2249.8	8313.0
all ages	37105	65606	15.8	9.5	194.4	45.9	137.4	292.5

2. Most of the infectious diseases which were responsible for the bulk of infant deaths in U.K. were effectively controlled during the period 1930–1950. Select a sensitive mortality index to illustrate the above fact by the data presented in the table below :

social class*	infant mortality rate/1000 live births		proportion of neonatal deaths (deaths within 4 weeks after birth) to total infant deaths (percentage)	
	1930–32	1050	1930–32	1950
I	32.7	17.9	66.4	72.1
II	45.0	22.2	60.4	73.0
III	57.6	28.1	51.0	62.6
IV	66.8	33.7	47.8	58.8
V	77.1	40.7	42.2	53.8

* These social classes are arranged in the above order by the registrar general of England and Wales according to occupational status as follows :

Class I	Professional occupations
Class III	Skilled occupations
Class II	Intermediate between class I and III
Class V	Unskilled occupations
Class IV	Intermediate between class III and V.

3. 4145 families in each of which a primary case of diphtheria was reported were kept under observation for one month. Some of these primary cases were removed to hospitals immediately after the diagnosis was established. The incidence of diphtheria during the month under observation among members of the household of the above primary cases are shown in the following table :

TABLE

age group	class I (primary case not removed to hospitals)			class II (primary case removed to hospitals)		
	total persons	primary cases	secondary cases	total persons	primary cases	secondary cases
under 5 years	2528	780	238	1807	531	112
5— 9 years	2172	701	252	1979	693	120
10—14 years	1679	365	122	1221	252	57
15—19 years	1123	151	47	621	80	8
20 years & above	8011	344	139	4122	158	45

Evaluate the effectiveness of prompt removal of diphtheria cases as a preventive measure in the control of the disease.

4. The following table gives the age specific fertility rates and life table populations of the corresponding age groups of the Uttar Pradesh female population. In the computations of fertility rates only the female births have been counted.

TABLE

age group X to $X + 5$ years	specific fertility rate per year per 1000 females in the age group X to $X + 5$ (only female births counted)	life table population of females in the age group X to $X + 5$ ($l_0 = 1000$).
15—19 years	55.3	3150.1
20—24 „	122.0	3053.1
25—29 „	114.6	2905.5
30—34 „	95.7	2680.5
35—39 „	73.1	2403.0
40—44 „	34.2	2120.7
45—49 „	10.4	1841.3

If the above fertility rates and mortality rates continue to operate for an indefinitely long period, calculate the limiting rate of natural increase of the U.P. female population.