

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1963

Paper I : Official Statistics and Descriptive Statistics (Theoretical)

Time : 4 hours

Full marks : 100

Figures in the margin indicate full marks for each question

GROUP A

(Answer any three questions from this group)

1. Discuss briefly (a) the importance of statistics in a planned economy, and (b) the data on various types of determinants of economic growth needed in the process of the formulation of plan targets. (8)

Discuss critically the requirements and availability of statistics for the setting up of such overall plan targets and targets of different types of goods and services. (8)

2. What are the types of data currently included in the official publications of Government of India dealing with public health and vital statistics? Also, name the publications. (10)

Distinguish between health statistics and vital statistics. What suggestions have you to offer to improve the reliability of these statistics? (6)

3. Describe the scope and coverage of available Inland Trade Statistics of India (by rail, river, road and air). Also, give the names of the publications containing them. (12)

What are your suggestions for improving the reliability of these statistics? (4)

4. Give a detailed account of the nature and scope of industrial statistics available at present in India. (8)

Describe the International Standards in Basic Industrial Statistics. To what extent does India conform to them? (8)

Neatness. (2)

GROUP B

(Answer all questions in this group)

5. Show that if $P > 0$, $Q > 1$ and $Q - P = 1$, then

$$f(x) = Q^{-k} \binom{k+x-1}{x} \left(\frac{P}{Q}\right)^x, \quad x = 0, 1, 2, \dots$$

represents a probability density when $k > 0$. Interpret the case when k is an integer. (8)

Find the moment generating function of the above distribution and hence find the first three moments of the distribution. (9)

6. Discuss the various measures of central tendency and variability and discuss their relative merits. (5)

Show that the mean deviation of a random variable is minimum when measured from the median. (5)

Find the mean deviation about the mean of the number of successes in n independent trials each with probability p of success. (6)

7. Discuss in detail the method of decomposing a Time Series into its various components. (17)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1963

Paper II: Probability Theory and Statistical Methods (Theoretical)

Time: 4 hours.

Full marks: 100

Figures in the margin indicate full marks for each question.

Group A(Attempt any three questions from this group)

1. a) N letters are randomly placed in N correctly addressed envelopes. Show that the probability that exactly K letters are placed in the correct envelopes is

$$\frac{1}{K!} \sum_{j=0}^{N-K} \frac{(-1)^j}{j!} \quad K = 0, 1, 2, \dots, N. \quad [9]$$

- b) If X and Y are Poisson variates with means μ_1 and μ_2 , show that the probability that $X - Y$ has the value r is the coefficient of t^r in

$$\exp(\mu_1 t + \mu_2 t^{-1} - \mu_1 - \mu_2) \quad [8]$$

2. X is a random variable with mean μ and standard deviation σ and measures β_1 and β_2 of skewness and kurtosis given by

$$\beta_1 = \frac{\mu_3}{\mu_2^{3/2}}, \quad \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

Show that for $K \geq 1$

$$\text{and } P \left\{ |X - \mu| \geq k\sigma \right\} \leq \frac{1}{k^2}$$

$$P \left\{ |X - \mu| \geq K\sigma \right\} \leq \frac{\beta_2 - 1}{\beta_2 - 1 + (K^2 - 1)^2}$$

and show that the inequalities can not be improved in general. [12]

Use the above inequalities to justify the 'frequency' definition of probability. [5]

3. Write down the distribution of central chi-square with n degrees of freedom. Hence find the distribution of the ratio of two independent chi-squares with n_1 and n_2 degrees of freedom. [4+8=12]

Find the first two moments of the distribution of this ratio. [5]

4. Define the moment generating function of a random variable and find it for the case of a random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

and utilise it to find the moments about the mean of the random variable. [11]

Also prove that if two random variables X_1 and X_2 are independently distributed as above with parameters μ_1, σ_1 and μ_2, σ_2 respectively, then their sum also has a similar distribution. [6]

5. Define the terms marginal distribution and conditional distribution in the case of a bivariate distribution. [4]
- Show that if $a > 0$, $c > 0$, $ac - b^2 > 0$, then for a suitable value of k

$$f(x, y) = k e^{-\frac{1}{2}(ax^2 + 2bxy + cy^2)},$$

represents a bivariate frequency density function where $-\infty < x < \infty$, $-\infty < y < \infty$. Find this value of k . [8]

Find the marginal distribution of X and the conditional distribution of X given $Y = y$. [5]

Group B

(Attempt any two questions from this group)

6. Write a note on the method of least squares for the estimation of parameters occurring in a statistical model. Show how this leads to a test of significance of the hypothesis $H_0: b = 0$ in the linear regression model of y on x , where b is the regression coefficient. [25]
7. Samples of size n_1 and n_2 are drawn from two normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 . It is known that the variances of the populations are proportional to the means. How can one test for the hypothesis $H: \mu_2 = 3\mu_1$ when the variances are unspecified? [5]
8. Describe the sequential probability ratio test for a simple hypothesis against another simple alternative. How is the test adjusted to get given values for first and second kinds of errors? What is the practical advantage of such a test over the usual fixed sample test? Apply the sequential probability ratio method to test if a binomial proportion has a specified value p_0 against an alternative p_1 . [25]
9. K samples of size n are taken from K normal populations with means $\mu_1, \mu_2, \dots, \mu_K$ and common variance σ^2 . How will you test the hypothesis $H: \mu_1 = \mu_2 = \dots = \mu_K$? Find the expected value of the mean square between samples when the above hypothesis is not true. [25]

000000

Please Turn Over

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - May 1963

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

Time : 4 hours

Full marks : 100

Figures in the margin indicate full marks for each question.

GROUP A

(Answer any three questions from this group)

1. "The sample survey is co-extensive with statistics itself". Discuss, bringing out clearly the role of sampling in statistical practice. (9)

Examine critically the situations in which sampling yields more accurate results than complete enumeration. (9)

2. Describe the circumstances under which sampling with varying probabilities is to be preferred to sampling with equal probability. (8)

A sample of size n is selected with varying probabilities with replacement. Give an unbiased estimate of the population total and its variance. How will you estimate the gain due to sampling with varying probabilities as compared to simple random sampling? (8)

3. Describe critically a recent round of the National Sample Survey conducted in India, giving details in regard to the design, information collected, and method of analysis of the data. (10)

What suggestions have you to offer to improve the design of the survey and the estimation procedures adopted? (6)

4. What are non-sampling errors? Write a note on the control and measurement of these errors? (8)

Show how the technique of interpenetrating sub-samples can be used in this connexion. Indicate its limitations. (8)

Neatness. (2)

GROUP B

(Attempt any three questions from this group)

5. Write an essay on 'THE CONTRIBUTION OF STATISTICAL REASONING TO EXPERIMENTATION' bringing out clearly the part played by randomisation, replication and local control in the planning of experiments. (16)

6. Describe how uniformity trial data are used to determine the size, shape and arrangement of experimental plots. (16)

7. Twenty five mango trees in an orchard were used for a manurial experiment in 1961, laid out in five randomised blocks of five plots (trees) each.

Owing to paucity of experimental material the same set of 25 trees are now required for a spraying experiment, to compare five different sprays.

Though it may be safe to assume that the manures do not interact with sprays, the proposed experiment should take into account, the residual effects which the manures might have possibly left on the plants.

Please turn over

(a) Suggest a suitable design for the spraying experiment and indicate how the random assignment of sprays to plants will have to be decided upon consistent with the suggested design. (3+7)

(b) Describe the analysis of the suggested experiment. (6)

8. (a) Define 'Factorial experiments'. (2)

(b) What are the advantages of a factorial experiment as compared with a series of separate experiments studying one factor at a time? (1)

(c) When can a factorial experiment be most conveniently laid out in a Split-plot design? (1)

(d) When a Split-plot experiment was already in progress the subplots had to be further split to accommodate a third factor (normally applied late) in the same experiment.

Write down the analysis for this experiment. (2)

9. Write short notes on any two of the following: (16)

(a) Fractional Replication.

(b) The analysis of mixed-up yields.

(c) The effect of errors in assumptions underlying Analysis of Variance.

Neatness. (2)

ooooOoooo

Please Turn Over

INDIAN STATISTICAL INSTITUTE
 Statistician's Diploma Examination - May 1963
 Paper IV: Applied Statistics (Theoretical)

Time: 4 hours

Full marks: 100

- i) Answer questions only from the two groups relating to the subjects you have opted for.
- ii) Use a separate answer book for each group.
- iii) Figures in the margin indicate full marks for each question.

GROUP A - ECONOMIC STATISTICS

(Answer any three questions from this group. For each question 2 marks.)

1. What type of price level is measured by a consumer price index number? Give a sketch of the method of calculation of such an index number for important groups of items as well as for all items taken together, indicating the nature of data required. [16]
2. What is a moving average? Describe how the technique is useful in different stages of the analysis of economic time series. [16]
3. If you are required to obtain the demand function for food on the basis of time series data, what determining variables would you include? Given the data, how will you derive the demand function and elasticities of demand? [16]
4. What is a production function? Taking a suitable production function, calculate the marginal productivities of the various factors of production, and indicate how you will estimate the parameters of the production function from statistical data. [16]
5. Give a brief but critical review of the growth of national income, changes in the structure of the economy, employment position and urbanization over the plan periods in India. [16]
6. Write notes on any TWO of the following:
 - i) Lorenz curve
 - ii) Input-output table
 - iii) National income: factor cost. [16]

GROUP B - STATISTICAL QUALITY CONTROL

(Answer any three questions from this group. For each question 2 marks.)

- 7.a. What is a control chart? Explain the terms 'control limits' and 'warning limits' in relation to a control chart. [6]
- b. When is a process said to be in 'statistical control'? Comment upon the statement: 'If a process is in statistical control all the items it produces are satisfactory'. [6]
- c. Describe some of the purposes for which an $\bar{X} - R$ chart can be used. [4]
- 8.a. Explain in detail the procedure of setting control charts for defects per unit. [8]
- b. Write a short note on the use of GO, NO-GO gauges in industrial quality control. [6]
9. Write short notes on
 - a) OC-curve [5]
 - b) Tolerance limits [5]
 - c) Sequential sampling. [6]

Please Turn Over

- 10.a. Describe the use of 'ranges' in comparing the variability of two or more machines in respect of a measurable characteristic. [10]
- b. Examine the usefulness and the limitations of the latin square design in industrial experimentation. [6]

GROUP C - STATISTICAL METHODS IN GENETICS

(Answer any two questions from this group)

11. Compare the relative efficiencies of backcross data ($\frac{AB}{ab} \times \frac{ab}{ab}$) and intercross data ($\frac{AB}{ab} \times \frac{AB}{ab}$) for the estimation of a linkage parameter. [25]
12. Determine the expected frequencies of the M₁, M₂ and M₃ blood group types in a population under random mating. How do you test the hypothesis that M, N gene frequencies are the same for two populations on the basis of independent samples from the two populations? What does a difference in the frequency of M₁ blood group in a community imply? [25]
13. Explain the importance of inbreeding in plant selection. Study the consequences of selfing in an inbreeding program. [25]

GROUP D - VITAL STATISTICS AND DEMOGRAPHY

(Answer any three questions from this group.)

For neatness 2 marks.

- 14.a. What do you understand by 'de facto' and 'de jure' population? Enumerate the relative merits of these two concepts in taking a census, with particular reference to India.
- b. Which of the methods - 'householder' or 'cavasser' do you recommend for India and why?
- c. How would you proceed to study the nature and extent of internal migration in India? [16]
- 15.a. Define the terms 'child-woman ratio', 'crude birth rate', 'total fertility rate', 'gross reproduction rate' and 'net reproduction rate'.
- b. For each of the above, indicate under what conditions it is the most convenient measure of the fertility of a population.
- c. Some of the quantities in (a) are defined in terms of the 'number of person-years lived during a year'. How is this number computed? [16]
- 16.a. Explain the difference between a 'stationary population' and a 'stable population'.
- b. Describe the steps in the computation of a stable population, starting from a stationary population.
- c. Describe briefly the utility of the concept of a stable population. [16]
- 17.a. How would you define an 'infectious disease'?
- b. How are records of such diseases obtained, with respect to incidence, intensity and duration?
- c. Describe a simple mathematical model which would help in predicting the course of an epidemic over time. [16]

GROUP E - EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Answer any two questions from this group)

19. Explain clearly the significance of the terms: (i) discriminating power (ii) reliability and (iii) validity as these are used in connection with the construction of a battery of objective tests. If the number of items in a test with reliability q_1 and validity q_1 is increased k-fold, show that the reliability q_k and validity q_k of the lengthened test are, under certain assumptions (which you have to state), given by

$$q_k = \frac{k q_1}{1 + (k-1)q_1} \quad \text{and} \quad q_k = \frac{\sqrt{k} q_1}{\sqrt{1 + (k-1)q_1}}$$

respectively.

[25]

20. Describe Hotelling's method of Principal Components as applied to the problem of estimation of factor loadings in multiple factor analysis of scores in psychological tests.

[25]

21. Three psychological tests A, B and C were given to a large number of Army recruits who could be distinctly classified into two groups in respect of their neurotic condition (i) Normal and (ii) Abnormal. The means, variances and covariances of the scores in these three tests are available separately for the two groups. If a new recruit whose neurotic condition is unknown, is given these tests, and the scores are made available to you, how will you use the above statistics to examine if he is likely to have any abnormal neurotic traits?

[25]

ooOooo

Please Turn Over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1963

Paper V: Methods of Numerical Computation, Descriptive
Statistics and Official Statistics (Practical)

Time: 5 hours

Full marks: 100

- i) Figures in the margin indicate full marks for each question.
ii) Use of Calculating machines is permitted.

Group A

(Answer any two questions from this group)

1. Find by numerical differentiation the first and second derivatives of the function $y = f(x)$ tabulated below, at the point $x = 0.6$.

x	y
0.4	1.5836494
0.5	1.7974426
0.6	2.0442376
0.7	2.3275054
0.8	2.6510818

[12]

2. Find the value of the determinant.

1.569	0.329	- 0.038	- 0.319
0.555	- 0.112	0.667	0.418
0.372	- 0.013	1.232	- 0.819
0.842	- 0.314	- 0.939	1.009

[12]

3. The values of an elliptic integral for certain values of the amplitude ϕ are given in the table below. Compute by use of an appropriate interpolation formula the value of the integral when $\phi = 24^\circ 36' 42''$.

ϕ	$F(\phi)$
21°	0.3706344
22	0.3887052
23	0.4068349
24	0.4250264
25	0.4432823
26	0.4616054
27	0.4799982

[12]

Please Turn Over

Group B

(Answer question 7 and any two of the rest)

4. The table below gives the value of notes issued yearly in India during 1933-1943.

Year	Value of notes issued (crores Rs.)
1933-34	177
1934-35	186
1935-36	196
1936-37	208
1937-38	214
1938-39	207
1939-40	252
1940-41	269
1941-42	421
1942-43	650

- a) Show the above data graphically. [6]
- b) If Y denotes the value of notes issued and t the time, fit a curve of the type $Y = ab^t$ to the data. [8]
- c) Draw the graph of the above equation and write your comments on the fit. [4]
5. 100 groups of 10 persons each were observed and the number of persons in each group who wore glasses recorded. The observations are given in the following table:

No. of persons wearing glasses in groups of 10	0	1	2	3	4	5	6	7	8	9	10
No. of groups	8	40	57	25	15	10	4	1	0	0	0

- a) Estimate the proportion of persons who wear glasses in the population from which the groups were formed. [4]
- b) Assuming that the number of persons wearing glasses in a group follows the binomial distribution, obtain the theoretical frequencies. [8]
- c) Assuming that the normal approximation to the binomial distribution holds in this case, calculate the theoretical frequencies of groups of 10 with 0, 2 and 5 persons respectively wearing glasses [a number x of persons wearing glasses may be considered to represent the interval $x - \frac{1}{2}$ to $x + \frac{1}{2}$]. [6]

6. The following table shows sepal length (Y), petal length (X_1) and petal width (X_2) in cms. of 20 flowers of a certain species.

serial no.	Y	X_1	X_2	serial no.	Y	X_1	X_2
1	4.8	1.6	0.1	11	5.0	1.3	0.3
2	5.4	1.5	0.4	12	4.5	1.3	0.3
3	5.2	1.5	0.1	13	4.4	1.3	0.2
4	5.5	1.4	0.2	14	5.0	1.6	0.6
5	4.9	1.5	0.2	15	5.1	1.7	0.1
6	5.0	1.2	0.2	16	4.8	1.4	0.3
7	5.5	1.3	0.2	17	5.1	1.6	0.2
8	4.9	1.4	0.1	18	4.6	1.4	0.2
9	4.4	1.3	0.2	19	5.3	1.5	0.2
10	5.1	1.5	0.2	20	5.0	1.4	0.2

- a) Obtain the equation of linear regression of Y on X_1 and X_2 .
- b) Calculate the multiple correlation coefficient of Y with X_1 and X_2 .
7. The following table shows prices and quantities consumed by a typical household in a month of certain commodities in 1955 and 1956.

Commodity	Price (Rs.)		Quantity	
	1955	1956	1955	1956
Rice	20.00 (md.)	15.00 (md.)	1 md.	1.25 mds.
Salt	4.00 (md.)	4.75 (md.)	10 srs.	8 srs.
Cloth	1.25 (yds.)	1.50 (yds.)	10 yds.	10 yds.
House rent	20.00 (rooms)	30.00 (rooms)	1 room	1 room

Calculate the three following types of index number for 1956 cost of living with 1955 as base:

- i) Weighted average of price relatives;
- ii) Aggregative expenditure index with
 - a) base year quantities as weights;
 - b) current year quantities as weights;
- iii) Fisher's 'Ideal' index number.

Group C

(Answer any one question from this group)

8. a) Name the publications in which any three of the following are available in India, the authority or organization issuing them, the periodicity with which they are issued and the breakdowns by which the informations are available:-
- i) Statistics of migration of population.
 - ii) Index numbers of agricultural population.

- iii) Statistics of births and deaths. [3]
- iv) Statistics of production, according to varieties of iron and steel. [3]
- v) Statistics of insurance. [3]
- b) From the publications supplied to you, collect the data in respect of any three of the following, presenting them in neat tabular forms, with proper headings, and foot-notes, if necessary. Mention the source from which you have compiled the data.
- i) Statistics of production of iron ore products in India for any six consecutive months. [5½]
- ii) Index number of industrial production in India for any five consecutive months. [5½]
- iii) Statistics of production of yarn and cloth in India variety-wise and statewise for any four consecutive months. [5½]
- iv) Statistics of shipping in India for any six consecutive months. [5½]
- v) Index numbers of wholesale prices in India for any twelve consecutive weeks. [5½]
- a) From the publications supplied, collect figures of average yield per acre of principal crops in India for the latest five consecutive years for which these are available and write a critical note on the trends in these yields. [13]
- b) Comment on the method of collection and presentation of data, as also the time lag in publication, of any three of the following publications issued by the Government of India:-
- i) Basic Road Statistics. [4]
- ii) Monthly Abstract of Statistics relating to posts and telegraphs. [4]
- iii) Indian Livestock Census. [4]
- iv) Indian Land Revenue Statistics. [4]
- v) Indian Cotton Pressing Factories and Returns. [4]

INDIAN STATISTICIAN INSTITUTE

Statistician's Diploma Examination - May 1963

Paper VI: Statistical Methods, Design and Analysis of Experiments and Sample Surveys (Practical).

Time: 5 hours

Full marks: 100

- i) Figures in the margin indicate full marks for each question.
 ii) Use of calculating machines is permitted.

Group A(Answer all questions from this group)

1. The following yields of grain of wheat were observed in fertiliser demonstration trial scheme carried on cultivator's fields in Ahmedabad district in Bombay State during Rabi 1956-57. Besides control, a treatment of NP fertiliser mixture (14:7:0) at the rate of 270 lb. per acre was applied. The control and the treatment were assigned at random to two adjacent plots in each village, the yields were converted to maunds per acre for the convenience of analysis. Analyse the data and state your conclusions:

[20]

Yield of grain of wheat in md./acre.

Sl.no.of village	C (control)	NP (Mixed fertiliser)
1	4.4	8.3
2	24.0	31.9
3	9.7	16.7
4	9.7	11.7
5	9.2	8.8
6	47.6	59.3
7	13.4	14.9
8	14.1	12.6
9	9.9	12.9
10	8.8	10.7
11	9.7	14.6
12	9.2	10.7
13	9.3	14.6
14	12.6	15.5
15	12.9	13.1
16	12.6	15.6
17	11.6	14.9
18	9.7	12.6

- 2.a) In an investigation into the causes of juvenile delinquency records for a number of boys convicted on various charges were collected. All these boys belonged to (1) criminal tribes, (2) others and the family background which was indicated by (a) whether both parents living or (b) at least one parent dead was considered to be a possible cause of delinquency. Test if this cause works in the same way for criminal tribes as for others.

Number of convicts

	criminal tribes	others
both parents living	15	

- b) A group of persons were subjected to a certain nutritional regime and it was noted whether they had shown improvement or deterioration after the course of treatment, indicated by I and D. Out of 20 students 13 were I and 7 were D. Test if the regime had any significant effect on the persons.
- c) Correlation between yield of grain (x) and yield of straw (y) was obtained from two varieties of wheat from 15 and 25 plots of size $1/80$ of an acre, and are .72 and .65 respectively. Test if the difference between the correlations is significant. [20]

Group B

(Answer any two questions from this group)

3. The following table gives the result of an experiment relating to the testing of rubber covered fabric in the Martindale Wear Testing Machine, using the layout given below. The factors used are: four different types of fabric A, B, C, D; four different positions on the machine 1, 2, 3, 4 and four runs 1, 2, 3, 4. The loss in weight of the fabrics are given below (in $1/10$ milligram).
Analyse the data and give your comments.

run	position on machine			
	1	2	3	4
1	A 251	B 241	D 227	C 229
2	D 234	C 273	A 274	B 226
3	C 235	D 236	B 218	A 268
4	B 195	A 270	C 230	D 225

4. The table below gives the results of an experiment on three varieties of cotton plants carried out on 5 randomised blocks. Here y denotes the yield of cotton lint in ounces per plot and x the number of plants per plot.

variety		blocks					total
		1	2	3	4	5	
A	y	11	8	9	6	5	39
	x	16	12	13	10	8	59
B	y	5	6	10	8	4	33
	x	12	14	18	15	10	69
C	y	4	7	6	3	4	24
	x	7	13	10	7	6	43
total	y	20	21	25	17	13	96
	x	35	39	41	32	24	171

Certain statistics computed from the above are given below:

$$\begin{aligned} \sum y^2 &= 694 & \sum x^2 &= 2125 & \sum xy &= 1197 \\ \sum B_y^2 &= 1924 & \sum B_x^2 &= 6027 & \sum B_y B_x &= 3400 \\ \sum V_y^2 &= 3186 & \sum V_x^2 &= 10091 & \sum V_y V_x &= 5610 \end{aligned}$$

where B_y denotes a block total for y , V_x a variety total for x etc.

Analyse the yield data after correcting for the variation in the number of plants and give your comments. [15]

5. Analyse the following results of a factorial experiment involving 3 factors N, P, K each at 2 levels, conducted in 2 replications, each of 2 blocks of 4 plots.

	block 1		block 2	
	treatment	yield	treatment	yield
Replication 1:	NK	159	P	153
	O	179	NPK	202
	PK	135	N	153
	NP	130	K	182
	block 3		block 4	
	treatment	yield	treatment	yield
Replication 2:	NPK	155	N	191
	NP	129	NK	138
	K	151	P	188
	O	159	PK	210

[15]

GROUP C

6. EITHER

The table below shows the total cultivated area in 1959 and also the area under wheat in 1960 of 20 villages in a certain district.

sl. no. of village	area in acres	
	total cultivated area	under wheat
	in 1959	in 1960
1	401	75
2	634	163
3	1194	326
4	1770	442
5	1060	254
6	360	101
7	946	359
8	470	109
9	1625	481
10	827	125
11	96	5
12	1304	427
13	377	78
14	259	78
15	186	45
16	1767	564
17	604	238
18	824	255
19	1016	236
20	524	247

Draw a sample of 5 villages from the above list one by one with replacement according to the following schemes:

- i) with equal probabilities
- ii) with probabilities proportional to the total cultivated areas in the villages.

Please Turn Over

Describe the procedure adopted by you.

On the basis of the two sets of samples drawn by you obtain, in each case, an unbiased estimate of the area under wheat in 1960 in the 20 villages. Also estimate the variance of each of the two estimates. [30]

OR

Raw wool contains varying amounts of grease, dirt and foreign material and its quality is measured by the 'clean content' - the percentage which the weight of the clean wool bears to the original weight of the raw wool.

To estimate the clean content an electrical core boring machine is used, which takes cores of about 1/4 lb. from a bale, which is then subjected to laboratory analysis.

In an experiment, 6 bales were drawn from a lot and in each 4 cores were taken and 'clean content' measured. Estimate the average clean content from the data given below and the standard error of the estimate. Would it not have been better to take 12 bales and 2 cores from each?

Bale						
Core	1	2	3	4	5	6
1	52.3	57.0	54.6	54.9	59.9	57.8
2	56.2	58.7	57.5	60.1	57.8	59.7
3	62.9	58.2	59.3	58.7	60.3	59.6
4	505	57.4	57.5	55.6	57.5	58.1

[30]

000000

Please Turn Over

INDIAN STATISTICAL INSTITUTE
 Statistician's Diploma Examination - May 1963
 Paper VII: Applied Statistics (Practical)

Time : 5 hours

Full marks : 100

- (a) Answer questions only from the two groups relating to the subjects you have opted for.
 (b) Use a separate answer book for each Group.
 (c) Figures in the margin indicate full marks for each question.
 (d) Use of calculating machines is permitted.

GROUP A - ECONOMIC STATISTICS

(Answer all questions from this group)

1. Gold holdings (in fine oz) per thousand persons in India and in the world are given below for the 5-year periods since 1895. On the basis of linear trends compare the rate of growth of gold holding in India with that in the world. (12)

	<u>India</u>	<u>World</u>
1895	162	200
1900	180	309
1905	201	343
1910	232	380
1915	300	432
1920	340	471
1925	407	490
1930	451	519
1935	353	532
1940	299	582
1945	276	629
1950	331	644
1955	308	643

2. The following table gives the average consumer expenditure (in rupees per month) on food grains and clothing in India in a recent period, classified by expenditure classes (in rupees per person per month).

Exp. class	per cent of persons	total expenditure	expenditure on	
			food grains	clothing
Below 8	14	6.26	3.49	0.21
8 - 11	17	9.41	4.69	0.53
11 - 13	12	11.98	5.89	0.79
13 - 15	9	13.96	6.35	1.03
15 - 18	12	16.49	6.93	1.43
18 - 21	8	19.54	7.58	1.92
21 - 24	7	22.51	8.26	2.19
24 - 28	5	25.79	8.52	2.60
28 - 34	5	30.69	8.71	3.85
34 - 43	5	37.72	9.33	4.93
43 - 55	3	47.24	10.08	4.33
55 & above	3	83.29	13.98	10.24
All classes	100	18.74	6.61	1.77

Obtain the income-elasticities of demand for food grains and clothing and estimate the total consumer expenditures on these items for the years 1961 to 1964, assuming that the population in 1961 is 45.8 crores, the rate of growth of population per year is 2.2 per cent and that the rate of increase of income (total expenditure) per person per year is 2 per cent. (25)

Please turn over

3. On the basis of the information given below, obtain national income at market price, national income at factor cost, net domestic product at factor cost, personal income, disposable income, consumption expenditure and saving (public and private combined). (12)

	(₹. crores)
Gross national product at market price	11,215
Indirect taxes	673
Direct taxes	257
National debt interest	44
Subsidies	18
Transfer payments (relief, pension etc.)	129
Capital consumption allowance	569
Net donations from abroad	78
Net income earned abroad	1
Net lending abroad	12
Gross investment in fixed capital	1461
Increase in stocks	27

000000

GROUP B - STATISTICAL QUALITY CONTROL

(Answer any two questions from this group)

4. In a factory there were 3 copstan lathes turning and facing a particular product. For convenience the products were batched and inspected in quantities of 200. The data on the number of defectives for 9 successive batches from each machine are given below.

Batch	No. of defectives		
	W/C 1	W/C 2	W/C 3
1	26	25	22
2	10	30	27
3	14	31	15
4	22	28	7
5	15	20	15
6	11	26	10
7	14	30	8
8	22	29	12
9	15	31	10

(a) Are the machines under control ?

(b) What is the capability of each machine ?

(c) Assuming that the differences between machines can be removed what would be the capability of the process ? (25)

5. A weaving experiment was done to determine the warp breakage rates of four warps numbered (1,2,3,4) each of which had been treated differently. The warps were woven simultaneously in 4 looms A, B, C, D according to the following layout by dividing the total weaving time into 4 periods. The warp breakage rates are also given in the table below.

Weaving period	Warp number			
	1	2	3	4
1	A	C	D	B
	5.5	2.9	9.8	6.7
2	C	D	B	A
	6.0	6.3	5.1	9.2
3	D	B	A	C
	8.9	2.0	5.8	6.6
4	B	A	C	D
	6.1	5.1	2.9	9.9

An increased warp breakage rate affects the weaving efficiency adversely.

Analyse the above data and make your recommendations on the choice of a suitable warp. (25)

✓ 6. (a) Prepare a defects per unit sampling scheme where the inspection unit is 1000 square yds of fabric, given

Allowable defects per unit = 40

Probability of accepting a lot as above = 0.10

Good quality (defects per unit) = 10

Probability of rejecting good quality = 0.05 (15)

(b) Draw the AOQ curve for the following sampling inspection plan

$N = 1000$, $n = 100$, $C = 3$

and obtain the AOQL (10)

oooOooo

GROUP C - STATISTICAL METHODS IN GENETICS

(Answer all questions from this group)

7. The following tables gives the distribution of 214 individuals by O - A - B and M - N blood group systems

	O + A	B + AB
MM	36	40
MN	52	50
NN	16	20

(a) Examine whether the two blood group systems are independent.

(b) Estimate the M - N gene frequencies from the marginal totals.

(c) Estimate the B gene frequency. (25)

8. The genetic and environmental variance - covariance ($v - c$) matrices for three characteristics, A, B and C are as follows.

	Genetic v-c matrix			Environmental v-c matrix		
	10	5	1	2	1	0
	20	4		4	1	
	25			6		

Find the discriminant function in terms of observed values a, b, c of the characteristics A, B and C for maximising the genetic component of the characteristic A in the selected population. (25)

oooOooo

Please turn over

GROUP D - VITAL STATISTICS AND DEMOGRAPHY

(Answer any two questions from this group)

9. The following table shows, for different parts of India in the 1931 census, the distribution of persons according to the digit occupying the units place in the recorded age. [E.g., 253 persons in Bengal per 1000 reported their age in years in figures ending with the digit 0.]

State	Digit in the Unit's place of age recorded in census									
	0	1	2	3	4	5	6	7	8	9
	Number (per 1000) recorded in respect of each digit of age									
Bengal	253	43	121	56	64	187	76	57	106	37
Bombay	292	43	110	56	60	215	66	47	78	33
Madras	264	48	113	64	73	171	89	48	90	40
Punjab	279	44	110	55	67	198	78	49	84	36
U.P.	294	47	113	45	65	186	83	43	91	33
Total	1382	225	567	276	329	957	392	244	449	179

- For each state, rank the digits in order of preference.
- Calculate the average frequency of each digit for all the states together and rank the digits in order of these average frequencies.
- Write your comments on the nature and extent of digit preference of Indians in reporting age. (25)

10. The following table gives age-wise figures of total and married female population in two regions A and B during a certain year.

Age (years)	Region A		Region B	
	Number of females (000)	Number married during the year (000)	Number of females (000)	Number married during the year (000)
15 - 19	1611	81.5	2159	390.7
20 - 24	1613	626.0	1964	1192.4
25 - 29	1571	1059.2	1834	1429.3
30 - 34	1474	1136.0	1619	1324.0
35 - 39	1381	1101.4	1473	1202.1
40 - 44	1318	1050.5	1231	980.1
45 - 49	1225	951.6	1091	836.4
50 - 54	1061	766.5	917	633.3
55 - 59	848	555.4	734	479.6
60 - 64	712	397.7	595	334.8
65 - 69	571	256.9	513	225.4
70 - 74	400	131.3	310	98.9
75 - 79	239	52.4	174	36.8
80 - 84	124	16.0	97	12.3
85 & above	60	3.9	56	3.7

- Calculate the total marriage rates for women in the regions A & B.
- Calculate the marriage rate of region B
 - by "direct standardization" with respect to region A
 - by "indirect standardization" with respect to region A. (25)

11. The following table gives the population (${}_n P_x$) on July 1, 1930 and number of deaths (${}_n D_x$) during three whole years, 1929, 1930 and 1931 in different age-groups of a population.

Age interval x to $x+n$	Population on July 1, 1930 within age interval x to $x+n$ (${}_n P_x$)	Deaths in 1929, 1930, 1931 within age interval x to $x+n$ (${}_n D_x$)
0 -	12343	2585
1 - 4	53766	727
5 - 9	75745	402
10 - 19	152849	805
20 - 29	119913	1121
30 - 39	121612	1503
40 - 49	107867	2875
50 - 59	72931	4320
60 - 69	47412	5481
70 - 79	19555	4814
80 - 89	4151	2176
90 - 99	290	279
100 - 109	3	6

Calculate for each age-group

- (a) the average annual death rate (${}_n m_x$)
 (b) the probability (${}_n q_x$) of a person at age x dying within an interval x to $x+n$, using the formula

$$-m_0 (.9539 - .5509 m_0)$$

$${}_n q_x = 1 - e$$

for the first age-group (0 -), and a standard formula for ${}_n q_x$ in terms of ${}_n m_x$ for the other age-groups. (25)

000000

GROUP B - EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Answer all questions from this group)

12. The following table of correlations of five emotional traits was obtained from a record of 172 normal children, aged nine to twelve, by Burt. It is expected that there is a general emotionality factor common to all these five traits. Find out the loadings of this factor by using Thurstone's Centroid method.

trait	1	2	3	4	5
1. sociability					
2. sorrow	0.83				
3. tenderness	0.81	0.87			
4. joy	0.80	0.62	0.63		
5. wonder	0.71	0.59	0.37	0.49	

(20)

13. A large number of students were examined in a particular subject by three examiners E_1 , E_2 , and E_3 independently, whose marking standards can be compared from the following percentage frequency distribution of scores :

marks	percentage frequency distribution		
	E_1	E_2	E_3
below 10	5	10	5
10-30	15	20	25
30-50	50	60	50
50-70	24	8	10
70-90	5	2	8
90-100	1	-	2

What will be the relative ranks of three students A, B, C who obtained the following marks ?

students	marks given by		
	E_1	E_2	E_3
A	25	62	73
B	48	51	35
C	78	25	50

(15)

14. The following table gives the number of passes and failures in Mathematics and Physics in a certain University examination. On the assumption that the scores in Mathematics and Physics are jointly normally distributed, estimate the coefficient of correlation between the scores in the two subjects.

		Mathematics	
		Pass	Fail
Physics	Pass	2279	142
	Fail	119	89

(15)

oooooOooooo

Please Turn Over

1027-22

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - May 1963
Paper VIII: Subjects of Specialisation - I

Time: 4 hours.

Full marks: 100

i) Answer questions only from the section below on the subject you have opted for.

ii) Figures in the margin indicate full marks for each question.

(a) ECONOMIC STATISTICS - Econometrics

(Answer any five questions)

1. Derive the Slutsky equation

$$\frac{\partial q_i}{\partial p_j} = \left(\frac{\partial q_i}{\partial p_j} \right)_{U=\text{const.}} - q_j \left(\frac{\partial q_i}{\partial y} \right)_{\text{Price=constant}}$$

where q_i stands for the quantity of the i -th commodity consumed, p_j for the price of the j -th commodity ($i, j = 1, \dots, n$), y for income and U for the utility of consumption of q_1, \dots, q_n .

Interpret the equation.

[20]

2. Explain and comment on the following statement. 'The input-output system as developed by Wassil. W. Leontief is an empirically oriented multimarket analysis.'

[20]

3. Describe the social accounting procedure for the economy as a whole and explain how social accounting is useful for economic policy.

[20]

4. Explain what is meant by a decision-model and how the decision-models are to be used for economic policy. Explain also how the decision-models are related to analytical economic models and how the target variables, the instrument-variables and the data variables are related to exogenous and endogenous variables. Give illustrations to cover all the above aspects of the decision-models.

[20]

5. Describe how stochastic models of economic situations are constructed. Distinguish between errors in variables and errors in equations and state how the distinction is relevant to estimation problems, be there one or more relationships in the model. Also discuss what precautions and steps have to be taken in estimating the parameters when the errors in variables are ignored.

[20]

6. What is a recursive system of structural relationships? Give examples of a recursive system. Discuss the method of estimation of parameters appropriate and available for a recursive system.

[20]

7. When is a system of structural equations said to be just identified? State a set of necessary and sufficient conditions. Describe in detail a method of estimation that can be used in the case of a just identified system of equations. State the methods of estimation that are usually followed when there is over-identification.

[20]

8. If the relationship between y_t and x_t is given by

$$y_t = b' \tilde{x}_t + \epsilon_t$$

where b' is the row-vector (b_1, \dots, b_k) ,

\tilde{x}_t is the column-vector (x_{t1}, \dots, x_{tk}) ,

$$E(y_t) = E(x_t) = E(\epsilon_t) = 0 \text{ for all } t \text{ and } x_t$$

x_t are non-stochastic real numbers,

and $E(\epsilon_t, \epsilon_t')$, $t \neq t'$ is not necessarily equal to zero,

derive the generalised least squares estimate vector of b and show that it is an unbiased minimum-variance linear estimator.

Explain why this result is relevant to economic analysis. [20]

9. Examine in detail the steps involved in the estimation of price and income elasticities of demand, assuming the data you need are available and are reliable. How are the price and income elasticities of demand relevant to economic analysis and economic policy? [20]

10. What type of data would you require for measuring inequality or concentration of incomes. How will you use these data? In the light of your analysis discuss the adequacy or otherwise of the following measures of inequality or of concentration:

- (i) Coefficient of variation
- (ii) Concentration-ratio
- (iii) $\left| \frac{\text{mean} - \text{median}}{\text{mean}} \right|$
- (iv) $\frac{\text{income above median or any other suitable fractile}}{\text{total income}}$

Work out the expressions for these measures for Pareto and lognormal distributions of income. [20]

100000

Please turn over

(b) TECHNO-COMMERCIAL STATISTICS - Statistical Quality Control

(Answer any four questions)

1. Explain the terms : specification limits, control limits, confidence limits and tolerance limits.
 From a controlled process a random sample of n units gives measures for a quality as x_1, x_2, \dots, x_n . For large n obtain a simple approximate formula for constructing tolerance limits T_1 and T_2 such that at least a given proportion P of the entire production would be within T_1 and T_2 with a confidence coefficient α . [25]
2. ✓ It is desired to instal an acceptance sampling plan for a measurable characteristic x where the individual product is considered defective if $x > u$.
 Evolve a suitable variables sampling plan given
 $AQL = P_1$, producer's risk = α
 $LTPD = P_2$, consumer's risk = β
 when the process s.d. is not known. State your assumptions.
 Derive a formula to obtain the points on the OC curve of this plan. [25]
3. Describe clearly the salient features of Mil Std 105A or Mil Std 105B tables.
 How will you select a suitable single sampling plan from those tables for normal inspection?
 When using this plan what are the criteria to be adopted to decide whether to reduce or tighten inspection or to continue normal inspection? [25]
4. The finished weight of a particular type of enamelware has been showing considerable variation, all of which has been attributed by the production department to the variation of the raw material, viz., mild steel sheets. The main stages of manufacture are (a) fabrication, (b) pickling (c) base coating, (d) first coating and (e) final coating. The production of these items are more than 5000 in a shift.
 It is desired to estimate the variation introduced at each stage. Explain how you will plan the collection and analysis of the data. Indicate the outlay of the forms for data collection and computations. (The weight at fabrication may all be attributed to the raw materials). [25]
5. Write short notes on
 (a) Evolutionary operations
 (b) Narrow limit gauging
 (c) Hamilton's standard lot plot method of acceptance sampling by variables. [25]

oooOooo

Please turn over

(A) DESIGN AND ANALYSIS OF EXPERIMENTS - Statistical Aspects

(Answer any five questions)

1. Explain the three principles, replication, randomisation and local control with reference to field experimentation. [20]
2. What are the considerations in determining the size and shape of plots for experimentation? What preliminary data do you need and what analysis do you undertake for finding the optimum size and shape? [20]
3. What are the advantages in using a Latin square design? Five varieties are tested in 3 Latin squares using 75 plots in all. How would you analyse the data? Draw the analysis of variance table specifying the sources of variation, degrees of freedom and the expressions for sum of squares etc. [20]
4. Define a balanced incomplete block design. Obtain intra and inter block estimates of varietal differences. [20]
5. How are factorial experiments more informative than experiments using one factor at a time?
Obtain a design for a three factor experiment each at two levels using blocks of four plots achieving complete balance over the estimation of first and second order interactions. [20]
6. It is stated that addition of antibiotics to tanks where fish are grown reduces the mortality of fish. As a statistician attached to Fisheries research station you are asked to conduct experiments to verify the statement made. Give a full description of the procedure you would follow emphasizing the need to repeat the experiment over several seasons. [20]

000000

(B) SAMPLE SURVEYS - Theoretical Aspects.

(Answer any four questions)

1. (a) Discuss how stratified random sampling with properly defined strata leads to improved estimate of the population mean.
(b) Deduce Neyman's formula for the optimum allocation of sample size to different strata.
(c) Indicate the practical difficulties in adopting the Neyman allocation.
(d) Consider the situation where the various strata are well defined and an unstratified simple random sample which subsequent to selection, has been classified strata-wise.
Show that, if the strata-sizes were known, the usual weighted average of strata means (so derived) would be nearly as efficient as with stratified sampling under proportional allocation provided of course the total sample size is large.
(e) From the following details concerning a stratified random sample, estimate the variance of the estimate in an unstratified simple random sample of the same size.

	Stratum 1	Stratum 2
Size	500	1000
Sample size	50	100
Sample mean	102.50	112.00
Sample variance	2.50	3.50

[25]

Please Turn Over

2. (a) Define 'Regression estimate'.
- (b) Suggest the regression estimator and derive an expression for its variance in each of the following cases,
- two phase sampling (double sampling)
 - rotation sampling (sampling on successive occasions) observations on any particular occasion being used as ancillary information for estimating the mean on the immediately following occasion.
- (c) In b (ii) what fraction of the sample should be replaced on each occasion in order that the estimate on the current occasion may have the maximum precision? [25]
3. Consider probability sampling with replacement and let x_1, x_2, \dots, x_j be the observations on the j distinct units that appear in a sample of size n so drawn. Denote by n_i the frequency of the i -th distinct unit ($i = 1, 2, \dots, j$; $\sum n_i = n$).
- Show that given x_1, x_2, \dots, x_j and their respective probabilities of selection, the conditional probability distribution of n_1, n_2, \dots, n_j is independent of the x 's.
 - Hence or otherwise, show that it is possible to construct an unbiased estimator based only on the distinct units that is more efficient than the usual unbiased estimator based on all the n observations.
- [Hint: Use the fact that the variance of the conditional expectation never exceeds the unconditional variance]. [25]
4. (a) For the following stratified two-stage sampling design, suggest a suitable estimate of the population mean and derive an estimate of its variance:
- The first stage units are selected from each stratum with probability proportional to size and the second stage sampling fractions are so chosen that the overall sampling fraction is uniform. Sampling at second stage is supposed to be random with equal probability. Sampling at both stages are with replacement.
- (b) Indicate situations where such a design might be advocated. [25]
5. (a) Describe Lahiri's method of sampling with probability proportional to size.
- With Lahiri's method derive the expected number of draws that would be necessary to obtain a sample of size n (with replacement).
 - Compare this method with the method of forming cumulative totals.
 - How does splitting of the larger units reduce the expected waiting time as derived in (b)?
 - Let any one of the above methods be repeated in the following case until two distinct units are selected.

Unit	Assigned Probabilities of selection
1	1/8
2	1/2
3	1/8
4	1/4
Total	1

For each of the four units, compute the probability of its inclusion in the sample so selected. [25]

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

- Tests of randomness of random sampling numbers
- Interpenetrating network of subsamples
- Selfweighting designs
- Systematic sampling
- Optimal allocation for more than one variate in stratified sampling.

[25]

ooo09oo

(g) STATISTICAL INFERENCE - General Theory.

(Answer any five questions)

- A population has the probability density function $f(x, \theta)$.
 - Define the 'consistency' and 'efficiency' of an estimator.
 - Is a consistent estimator necessarily unbiased? Illustrate by two examples.
 - Show that if T_n is a statistic calculated from a sample of size n , such that

$$E(T_n | \theta) \rightarrow \theta \text{ and } V(T_n | \theta) \rightarrow 0$$
 as $n \rightarrow \infty$, then T_n is a consistent estimator of θ .
 - If T_n and T'_n are two minimum variance unbiased estimators of θ , show that the coefficient of correlation between T_n and T'_n is unity.
 - In the case (d), show that T_n and T'_n are equivalent in the sense that the set of sample points for which $T_n \neq T'_n$ is zero. [20]

Please Turn Over

2. A population has the probability density function $f(x; \theta)$, where θ is known to lie in a set Ω .
- State the criterion of sufficiency in terms of conditional probabilities. Verify this criterion for the statistic \bar{X} from a random sample of size n taken from a normal population with mean μ and standard deviation σ .
 - Show that a necessary and sufficient condition for \bar{X} to have a sufficient estimator is that the joint probability density function of the sample has Koopman's exponential form.
 - If $f(x; \theta)$ is a normal probability density where θ is a vector-valued parameter $\theta = (\mu, \sigma)$, and x_1, \dots, x_n are independent observations of a random variable with probability density $f(x; \theta)$, what is a sufficient statistic for the estimation of θ ? Verify Koopman's form for the joint probability density of the observations.
3. The parameter θ occurring in a probability density $f(x; \theta)$ belongs to a set Ω .
- Suppose a statistic T is complete for Ω . If Ω' is a proper subset of Ω , is T necessarily complete for Ω' ? Give reasons for your answer.
 - Let Ω' be a proper subset of Ω and the following condition be true for all sets E in the sample space:

$$P(E | \theta) = C \text{ for every } \theta \in \Omega'$$
 implies

$$P(E | \theta) = 0 \text{ for every } \theta \in \Omega \setminus \Omega'$$
 Show that under this condition, if T is complete for Ω' it is also complete for Ω .
 - Verify that the sample mean is complete for normal probability densities $N(\mu, 1)$, where $-\infty < \mu < +\infty$.
 - Suppose θ has unbiased estimators for every $\theta \in \Omega$, and let T be a sufficient statistic over Ω . Show that from each unbiased estimator an improved unbiased estimator with less sampling variance can be obtained as a function of T .
 - In the case (d) show that, if T is also complete for Ω , all the improved unbiased estimators are equivalent or, in other words, that there exists a unique minimum variance unbiased estimator. [20]
4. a) State and prove Neyman-Pearson's lemma and show how it can be used to find a most powerful test of a simple hypothesis against a simple alternative. Show that the definition of the test you obtain is both necessary and sufficient.
- Show that in the above case the power of the test is greater than the size (or level) of the test.
 - From a normal population $N(\mu, \sigma)$, a random sample of size n is taken. Obtain a most powerful test for the hypothesis $H_0: \sigma = \sigma_0$ against an alternative $H_1: \sigma = \sigma_1 (\sigma_1 > \sigma_0)$. Is this test uniformly most powerful for all alternatives? What is the largest class of alternatives for which it is uniformly most powerful?

- 4.d) If a sufficient statistic exists for the parameter θ of the single parameter density function $f(x; \theta)$, examine if a uniformly most powerful test can be constructed (under certain conditions, to be stated, on the exponential form of the sample probability density) for testing
- $$H_0 : \theta \leq \theta_0 \text{ against } H_1 : \theta > \theta_0. \quad [20]$$
- 5.a) Define a 'locally most powerful' test.
- b) Show how to use Neyman-Pearson's lemma to construct locally most powerful unbiased tests.
- c) What is the 'Neyman structure' of a population which ensures the existence of similar regions in the case of composite hypotheses with one free parameter?
- d) Verify the above structure for $N(\mu, \sigma)$ for the hypotheses $\mu = \mu_0$, and show how you would obtain similar regions. [20]
- 6.a) Define a 'confidence region' for a parameter.
- b) Suppose the criterion for the choice of the 'best' confidence region from a class of such regions of the same confidence level α is that this region should have the smallest probability of including a wrong parameter value. Show how the theory of uniformly most powerful tests can be used to obtain the best confidence region.
- c) Obtain the best confidence region of level α for the parameter μ of the normal distribution $N(\mu, \sigma)$ from a sample of size n . [20]
- 7.a) Describe how a two-person zero-sum game can represent statistical decision situations.
- b) Define the following terms:
- i) randomized and non-randomized decision functions;
 - ii) Bayes and minimax decision functions
 - iii) Admissible, complete and minimal complete classes of decision functions. Establish some inclusion and equivalence relations among these classes.
- c) State a theorem on the existence minimax decision rules which are Bayes with respect to some least favourable a priori distribution.
- d) Describe how the problem of minimum variance estimation can be posed as a decision problem with a convex loss function. [20]

oooOooo

Please Turn Over

INDIAN STATISTICAL INSTITUTE
 Statistician's Diploma Examination - May 1963
 Paper IX: Subjects of Specialisation - II

Time : 4 hours

Full marks : 100

- (1) Answer questions only from the section below on the subject of specialisation you have opted for.
 (ii) Figures in the margin indicate full marks for each question.

(a) ECONOMIC STATISTICS - Indian Economics(Answer any three questions from each group)GROUP A

1. Discuss the measures taken in recent period for the consolidation of agricultural holdings in India. What other measures are necessary? (16)
2. Review the changes in the structure of Indian industries as a result of the growth of the public sector. Do you consider the achievement of the public sector satisfactory? (16)
3. Should an industry have any right to exist if it cannot pay its workers wages required for subsistence? Discuss the factors that should be considered while fixing minimum wages for various industries in India. (16)
4. "The monetary and credit policy of the Reserve Bank of India during the plan period has been characterized by the twin considerations of restraining inflationary forces and assisting a developing economy". Discuss with reference to the measures adopted by the bank in the plan period. (16)
5. What are the main changes in taxation policy in India that have been adopted since the beginning of economic planning? How far have these changes actually helped in the mobilisation of resources for development? (16)

Neatness. (2)

GROUP B

6. What is meant by material balances in planning? How can input-output analysis be used for this purpose? (16)
7. Give a brief account of the growth models of Harrod and Domar, indicating their limitations. How far can these models be used in planning for economic growth in India? (16)
8. "All crucial decisions in planning are essentially political in character and cannot be subjected to economic calculus". Examine India's Five-year plans with reference to the above statement. (16)
9. "The problem of balancing competing objectives is implicit in all planning and the quality of a plan depends upon the soundness of its judgments regarding the relative emphasis on the various objectives". Discuss critically with reference to the experience of the Second Five year plan of India. (15)
10. How should the aggregate investment and its allocations be determined in the planned development of a country? How will you use linear programming for allocation of investment? (16)

Neatness (2)

ooooo

Please turn over

(b) TECHNO-COMMERCIAL STATISTICS - Operations Research
 (Or Elementary of Bookkeeping & Accountancy) and
Statistical Methods in Business.

(i) Operations Research

(Attempt any four questions. Use separate answer book for these questions)

1. "Operations Research is an aid for the executive in making his decision by providing him with the needed quantitative information based on the scientific method of analysis". Discuss the above statement in detail, illustrating it with operations research methods that you know. (17)

2. "In the management of inventories, the questions which always have to be answered are "when" and "how much" for the parcel or goods under consideration, whether it or they are procured or produced or both". Illustrate with the inventory models that you know how the above questions are tackled. (17)

3. For the case of one servicing station, Poisson arrivals with mean arrival rate λ and exponential service-times with mean service-rate μ , show that P_n the (limiting) probability that there are n units in the system (at time t or any other time) satisfies the following equation.

$$P_1 = \frac{\lambda}{\mu} P_0$$

$$\text{and } P_n = \frac{\lambda + \mu}{\mu} P_{n-1} - \frac{\lambda}{\mu} P_{n-2} \quad n > 1$$

Obtain the solution of the equations. Obtain also the distribution of the waiting time and the distribution of the total time an arrival spends in the system.

Thus or otherwise obtain the average queue-length, the average waiting time of an arrival, and the average time an arrival spends in the system. Work out the variance of the queue-length. (17)

4. If C_1 is the unit cost of replacement in a group and C_2 , the unit cost of individual replacements after failure and $f(x)$ = number of failures in the x^{th} period, write down the expression for $K(t)$ = total cost from the time of group installation until the end of t periods where t is the number of periods between group replacements. If t is chosen so that $\frac{K(t)}{t}$ is minimised, verify the following replacement rules.

(i) One should group-replace at the end of the t^{th} period if the cost of individual replacements for the t^{th} period is greater than the average cost per period through the end of t periods and

(ii) One should not group replace at the end of the t^{th} period if the cost of individual replacements at the end of the t^{th} period is less than the average cost per period through the end of t periods. Discuss also the possibility of never group-replacing as an alternative to group-replacing at some time t . (17)

5. Describe the simplex method of solving a linear programming problem. Explain how the simplex-solution can also be made to give the solution of the dual programming problem. (17)

6. Explain with illustrations how Monte Carlo methods are useful in Operational Research. (17)

Neatness. (2)

(For questions on Statistical Methods in Business please see page 5)

Please turn over.

(11) Elements of Bookkeeping & Accountancy

(Use separate answer book for these questions)

7. From the following Trial Balance of Mr. Senyal as on 31st March 1963, prepare a Trading & Profit & Loss account for the year ended on 31.3.1963 and a Balance Sheet as at that date after making necessary adjustments :- (20)

	Dr.	Cr.
Mr. Senyal's Capital Account		60,000
" " Drawing Account	12,000	
Furniture & Fixtures	4,000	
Plant & Machinery	30,000	
Stock on 1. 4. 1962	20,000	
Purchases	80,000	
Return outwards		2,000
Salaries & wages	22,400	
Sundry Debtors	20,400	
Sales		1,30,000
Sundry Creditors		12,000
Return Inwards	5,000	
Loan @ 6% taken from Mr. Sethi on 1.10.1962		10,000
Postage & Telegram	1,500	
Rent, Rates & Taxes	3,600	
Bad Debts written off	400	
Discounts		600
Trade Expenses	200	
Interest on Loan from Sethi.	150	
Insurance charges	800	
Travelling expenses	500	
Sundry Expenses	300	
Cash in Hand	3,050	
Cash at Bank	10,300	
	2,14,600	2,14,600

Adjustments

1. Stock on 31.3.1963 was valued at Rs. 21,000.
2. Of the Sundry Debtors Rs. 400 are bad and should be written off. Create Reserve for Bad & Doubtful debts @ 5% on Sundry Debtors and a Reserve for Discount on Debtors @ 2%.
3. Salaries Rs. 800 for March 1963 were not paid.
4. Interest on Capital to be calculated @ 6% per annum and on Drawings Rs. 350.
5. Insurance prepaid was Rs. 100.
6. Depreciate Furniture & Fixture @ 5% and plant & machinery by 10%.

Please turn over

8. For their mutual accommodation, Krishnan accepted on 1.3.1963, a bill drawn on him by Ramn for Rs. 4,000 for 3 months. The bill was discounted for Rs. 60 discount and the proceeds were shared equally.

On the same day and for the same purpose, Krishnan received an acceptance from Ramn for Rs. 6000 for 3 months.

The bill was discounted for Rs. 90 discount and the proceeds were shared as two-third to Krishnan and one-third to Ramn.

Before the due date, Krishnan sent a cheque to Ramn in full settlement of his account but at maturity Ramn became insolvent and his estate paid a dividend of 50 naye paise in the rupee.

Give journal entries in the books of Ramn & Krishnan. (15)

9. EITHER,

On 31.12.1960, the trial balance of I. L. Gupta contained the following ledger balances :-

Reserve for Bad debts account (1.1.1960) = Rs. 1000
Reserve for Discount on Debtors (1.1.1960) = Rs. 475
Sundry Debtors account = Rs. 40,400

(of the above Debtors, Rs. 400 are bad and should be written off.)

On 31.12.1961, the Sundry Debtors were Rs. 20,500 of which Rs. 500 are bad and should be written off.

It is desired to maintain the Reserve for Bad debts account @ 5% on Sundry Debtors and Reserve for Discount on Debtors @ 2½%.

Show by means of Ledger Accounts how the Reserve accounts will appear in the books of Mr. Gupta for the years 1960 & 1961 and bring down the balance on 1.1.1962. (15)

OR

A company whose accounting year is the calendar year purchased on 1.4.1959, Machinery costing Rs. 30,000.

It purchased further machinery on 1.10.1959 costing Rs. 20,000 and on 1.7.1960 costing Rs. 10,000.

On 1.1.1961, one-third of the machinery which was installed on 1.4.1959 became obsolete and it was sold for Rs. 3000.

Show how the Machinery Account would appear in the books of the company, it being given that the Machinery was depreciated by the Fixed Instalment method @ 10% per annum. What will be the balance on 1.1.1962 on Machinery Account. (15)

10. EITHER,

Give journal entries to rectify the following errors :- (10)

- i) Rs. 540 received from Kunta was posted to the debit of his account.
- ii) Rs. 100 being purchase returns were posted to the debit of Purchases Account.
- iii) Discounts Rs. 200 received were posted to the debit of Discount Accounts.
- iv) Rs. 374 paid for Repairs to Motor Car was debited to Motor car account as Rs. 174.
- v) Rs. 400 paid to C. Das was debited to the account of G. Das

OR

Write explanatory notes on the following :- (10)

- i) Liquid assets
- ii) Documentary bill
- iii) Withdrawn Capital

11. If the present price of an article be reduced by 40%, a purchaser will be able to obtain 48 more articles for 18 shillings.

What is the present price ?

(10)

(11) Statistical Methods in Business

(Attempt any two questions. Use separate answer book for these questions)

12. "The joint use of previous sales and general business indicators combined with some very simple techniques of extrapolation prove of considerable help in providing reasonably good forecasts of sales at strategic points during the year".

Enlarge upon the above statement and discuss its adequacy.

(15)

13. Describe in detail the purpose and the different steps in the planning and execution of a market survey.

(15)

14. Explain in brief the following four systems of job-evaluation and discuss their applicability.

(1) Nonquantitative evaluation measures :-

- (a) The ranking system.
- (b) The job classification-system.

(2) Quantitative evaluation measures.

- (a) The point system.
- (b) The factor comparison system.

(15)

ooooo

(d) DESIGN & ANALYSIS OF EXPERIMENTS - Combinatorial Aspects

(Answer any four questions)

1. Characterise axiomatically a group, a subgroup of a group and finite field.

(9)

Show that the elements of $GF(2^n)$ of the form $x+x^2$ where x is in $GF(2^n)$ form a subgroup of order 2^{n-1} of the additive group of $GF(2^n)$.

(4)

Construct $GF(2^3)$ and verify that the non zero elements of the form $x + x^2$ where x is in $GF(2^3)$ form a difference set in the multiplicative group of $GF(2^3)$.

(12)

2. Define a balanced incomplete block design (B I B D) and a resolvable balanced in complete block design, and obtain the relations and inequalities amongst their parameters, using the incidence matrix of the designs.

(11)

Show that the number of common treatments between any two blocks of an unsymmetrical design with $\lambda \geq 2$ is never greater than α where

$$\alpha = \frac{(b-3)(r-1)}{(r-2)^2 - (\lambda-2)(b-2)}$$

(11)

Show how you can construct two B I B designs given a symmetrical B I B D.

(6)

Please turn over

3. Define a partially Balanced Incomplete Block design (P B I B D) and obtain the relations among the parameters for two associate classes P B I B D. (7)

Explain Group Divisible, Triangular and L_1 association schemes and obtain the parameters of the second kind in these cases. (8)

Give a necessary and sufficient condition that a two associate classes P B B I D is a Group Divisible design. (5)

Show how to construct a Group Divisible design from a B I B D with $\lambda = 1$. (5)

4. Define an orthogonal array $[\lambda n^2, c, n, 2]$ in n symbols with c constraints, strength 2 and index λ and show that

$$c \leq \left[\frac{\lambda n^2 - 1}{n - 1} \right]$$

where $[x]$ denotes the greatest integer contained in x . (10)

Prove that the existence of an Affine Resolvable B I B D with parameters $v = nk = n^2(n-1)t+1$, $b = nr = n(n^2t+1)$, $\lambda = nt + 1$ implies the existence of an orthogonal array of strength 2 given by

$$[(n-1)t+1, n^2t+1, n, 2]$$

in n symbols and the maximum number of constraints n^2t+1 . Indicate the connection with mutually orthogonal Latin Squares when $t = 0$. (15)

5. (a) In a symmetric factorial experiment with n factors each at two levels each block size is 2^r , $r < n$. Find the maximum number of factors which can be accommodated so that no degree of freedom belonging to a main effect or first order interaction is confounded? (12)

(b) Construct the principal block for a suitable confounded design for five factors each at two levels in four blocks of size eight so that no interaction of less than three factors is confounded. Indicate the interactions actually confounded. (13)

6. Explain clearly what is a rotatable design and indicate some methods of constructing them. (25)

ooooOoooo

(e) SAMPLE SURVEYS - Organizational Aspects

(Answer any four questions)

1. Write a critical note on the problems that arise in the organization of large-scale sample surveys in India. Discuss briefly the appropriate method of inquiry to be adopted, the choice between sampling and complete enumeration, and the considerations to be kept in view in drawing up the schedules and questionnaires. (25)

2. Write a critical note on preparation of reports of sample surveys, detailing the important features of a report and the steps to be taken to improve the quality of the reports. Illustrate your remarks by referring to a National Sample Survey report. (25)

3. Draw up a schedule of inquiry into the volume of rural indebtedness in your state. Give your own definitions and concepts, with justification, and draw up a set of instructions to field workers and supervisors. (25)

4. You are required to collect data relating to consumer expenditure in the households in India. Discuss the concepts and definitions involved in the items of information to be included in the schedule. Comment on the organization of the survey and processing of data, illustrating your observations by referring to the consumer expenditure survey under the National Sample Survey. (25)

5. What are the advantages and disadvantages of the mail-enquiry method as compared to the interview method of surveys? (6)

Discuss whether there should be any difference in forms for collection of the same information by these two methods. (6)

Enumerate systematically the different phases of work in the actual implementation of a well planned mail enquiry. Bring out the inter-relationships between the various stages and explain in what way each affects the final outcome. (13)

6. Write short notes on any three of the following :-

- (a) Ad hoc field organization (0 $\frac{1}{2}$)
- (b) Post-tabulation scrutiny (0 $\frac{1}{2}$)
- (c) Testing of schedules and instructions (0 $\frac{1}{2}$)
- (d) Storing of records (0 $\frac{1}{2}$)
- (e) Follow-up of non-response cases (0 $\frac{1}{2}$)

ooooo:ooo

(c) STATISTICAL INFERENCE - Special topics

(Answer any four questions)

1. Show that the sequential probability ratio test for testing a simple hypothesis against simple alternative terminates with probability one. (13)

Obtain an optimum sequential test (in the sense to be defined by you) for testing the mean of a normal population with unknown variance. (12)

2. Show that there does not exist a fixed sample procedure which can provide fixed length confidence interval with a given confidence-coefficient for the mean of a normal population with unknown variance. (13)

State and prove the two sample procedure developed by Stein for this purpose. (12)

Please turn over

3. Discuss the importance of non-parametric methods in statistical inference. Mention a few practical situations where such methods are usually used. (7)

Obtain the sign test for the hypothesis that the populations from which two samples of n observations are drawn, are identical. (9)

If a sequence of observations taken at successive intervals of three is available, how will you test, using the method of runs, that the sequence is random? (9)

4. Two multivariate normal populations are known to have identical dispersion matrices. Write down the form of the D^2 statistic to test the hypothesis that the means in the populations are the same. (9)

Show that D^2 statistic remains invariant under the same linear transformation of variables and that its sampling distribution involves only the population value of D^2 as a parameter. (16)

5. What are canonical correlation coefficients between two sets of variates? (4)

Find the determinantal equation which they satisfy. (7)

Find the distribution of the sample canonical correlation coefficients when the population correlation coefficients are all zero. (7)

Show also that the sample canonical correlation coefficients are invariant under linear transformations of each set of variables. (7)

6. If $(x_{1t}, x_{2t}, \dots, x_{kt})$ $t = 1, 2, \dots, n$ constitute n independent observations of a vector having a k -variate normal distribution with zero mean and

$L_{r,s}$ ($r = 1, 2, \dots, k$; $s = r, r+1, \dots, n$) denote the determinant of the matrix

$$\left| \left| \sum_{t=1}^n x_{it} x_{jt} \right| \right| \quad i, j = 1, 2, \dots, r,$$

show that

$$\frac{L_{1,s}}{L_{2,s}}, \frac{L_{2,s}}{L_{3,s}}, \dots, \frac{L_{k-1,s}}{L_{k,s}}$$

are mutually independent, as also the variates

$$\frac{L_{r,r}}{L_{r,r+1}}, \dots, \frac{L_{r,n-1}}{L_{r,n}} \quad (25)$$

7. Write critical notes on any three of the following :-

- Principal Factors
- Likelihood ratio test
- U-statistics
- Problem of optimum classification
- Tolerance units. (25)

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - May 1965
Paper X: Subjects of Specialisation - III

Time: 5 hours.

Full marks: 100

- (i) Answer questions only from the section below on the subject you have opted for.
(ii) Figures in the margin indicate full marks for each question.
(iii) Use of Calculating machine is permitted.

(a) ECONOMIC STATISTICS - Practical
(Answer any four questions)

1. Fit a logistic curve

$$x_t = \frac{k}{1 + be^{-at}}$$

to the following Agricultural income-series for India (1900-1920).

Year t	Agr. income (Rs. Abja)	Year	Agr. income (Rs. Abja)
1901	39.76	1911	44.33
1902	36.87	1912	42.26
1903	42.69	1913	42.20
1904	41.03	1914	38.07
1905	36.19	1915	42.33
1906	39.27	1916	46.71
1907	42.49	1917	48.93
1908	32.17	1918	46.74
1909	39.01	1919	35.47
1910	45.28	1920	45.18

(Note: $\frac{1}{x_t} = \frac{(1 - e^{-at})}{k} + e^{-at} \frac{1}{x_{t-1}}$)

and b may be estimated by the formula of Rhoden

$$\log_e b = \frac{n(n+1)}{2} + \frac{1}{n} \sum_{t=1}^n \log_e \left[\frac{k}{x_t} - 1 \right]$$

where N is the total number of observations.)

Work out the proportion of variance explained both in the reciprocal series and in the original series. Interpret the results and comment. Work out also the projected values of agricultural output from 1950 to 1960 and comment on their adequacy. [25]

2. The following table gives international data on labour inputs and wage rates in two industries.

Fit the log linear relationship

$$\log \frac{V}{L} = \log a + b \log W$$

to the data for the two industries. Here

V is value added in thousands of U.S. dollars

L labour input in man-years

W wage-rate (total labour cost divided by L).

Obtain the estimates of b and test if they are different from 1.
Interpret the results and comment.

Obtain also the estimates of the regression-coefficients of $\log W$ on $\log \frac{V}{L}$ and compare the two sets of estimates and comment.

Sl. no.	Country (year of data)	Grain mill products industry		Textiles spinning and weaving industry	
		$\frac{L}{V}$	$\$ W$	$\frac{V}{L}$	$\$ W$
1.	U.S.A. (1954)	0.0903	3955	0.2250	2920
2.	Canada (1954)	0.1254	3138	0.2349	2708
3.	New Zealand (1953/56)	0.2311	1951	0.3053	1755
4.	Australia (1955/56)	0.2562	1657	0.5024	1242
5.	Denmark (1954)	0.2562	1657	0.5024	1242
6.	Norway (1954)	0.4932	1503	0.4993	1127
7.	U.K. (1951)	0.3775	1110	0.5560	874
8.	Ireland (1953)	0.4964	965	0.6813	708
9.	Puerto Rico (1952)	-	-	0.6290	932
10.	Columbia (1953)	0.2090	653	0.3052	1089
11.	Brazil (1949)	-	-	0.9334	343
12.	Mexico (1951)	0.6821	340	0.8673	461
13.	Argentina (1950)	0.6507	505	0.8212	481
14.	E-I Salvador (1951)	0.5040	495	0.8702	456
15.	Southern Rhodesia (1952)	0.8475	402	-	-
16.	Iraq (1954)	-	-	2.1531	215
17.	Ceylon (1952)	+ 2.0870	163	1.2190	226
18.	Japan (1953)	0.6606	461	1.3901	287
19.	India (1953)	5.1523	98	2.5202	276

(Note: The data on wage-payments for different countries include varying proportions of non-wage benefits. The data on employment are not corrected for inter-country differences in the number of hours worked per year or the age and sex composition of the labour force.)

[25]

3. The supply equation

$$y_1 = \gamma y_2 + \beta_1 x_1 + \beta_2 x_2$$

is fitted to the data on

y_1 = per capita consumption of meat (pounds per year)

y_2 = retail price of meat (index 1931-39 = 100)

x_1 = cost of processing meat (index, 1935-39 = 100)

x_2 = cost of producing agricultural products (index, 1935-39=100)

and all the variables are measured from their means and the sample consists of annual data, 1919-1941.

The following table gives the following sums of squares and products of the deviations of the variables from the means.

	y_1	y_2	x_1	x_2
y_1	1369.54	- 352.55	- 536.48	983.06
y_2		1581.49	850.33	1235.76
x_1			2534.80	730.78
x_2				2626.99

Derive the estimates of γ , β_1 and β_2 and test their significance. Work out also the multiple correlation coefficient.

Estimate also the coefficients γ and β_1 and the corresponding multiple correlation coefficient when the term $\beta_2 x_2$ is omitted. Hence or otherwise obtain the proportion of variance explained by x_2 alone. Examine if the independent variables, y_2 , x_1 and x_2 are correlated and comment on the results.

(Note - Here the random errors in the supply-equation are assumed to be temporally independent).

Discuss if y_2 can be treated as an exogenous variable as has been done in the above formulation. [25]

4. Fit the log-normal distribution to the following U.S.S. data on income distribution in urban groups I and IV.

Test the goodness of fit of the log-normal distribution to the data.

Work out also the means and coefficients of variation for the two groups and compare them. [25.]

Percentage of persons in urban area

Per capita monthly Expenditure class in Rs.	Urban group I (% age)	Urban group (IV) (% age)
0 - 8	15.62	1.85
8 - 11	14.53	6.40
11 - 13	10.89	3.09
13 - 15	12.91	2.61
15 - 18	14.80	8.10
18 - 21	9.34	6.90
21 - 24	6.96	10.61
24 - 28	4.00	2.58
28 - 34	1.94	4.37
34 - 43	3.79	18.74
43 - 55	3.52	6.80
55 -	1.70	28.95
total	400.00	100.00
total number of persons	400	454

(Note: Urban group I - towns with population below 15,000.

Urban group IV- The big cities; Calcutta, Bombay, Delhi and Madras).

5. Solve the following linear programming problem by any method that is convenient for you. Consider the solution of the dual also. Comment on the results.

A single product (corn) is produced with three fixed factors (labour, land and tractors) according to the technology.

	Inputs	Unit	Process	Process
			I	II
input	Labour	man-month	25	5
	Land	acres	50	100
	Tractors	tractor-months	20	32
output	Corn	100 tons	-1	-1

The price of corn is given as ₹ 10 per ton so that each process has receipts of ₹ 1000 at the unit level. The total resources available are 10 man-months, for labour, 110 acres for land and 10 tractor-months for tractors. The total receipts are to be maximised.

[25]

6. The transaction-matrices table for U.K., for the year 1955 is given below.

Study the table and prepare a small note commenting on the table, working out any percentages and regressions that you think are relevant.

[25]

Input from industries	Output to industries										Final demand Exports Capital formation	Government consumption	Total Gross Output		
	1	2	3	4	5	6	7	8	9	10				total-allocated	
1. Agriculture and food	-	*	2	4	8	*	*	*	*	3	6	29	3	7	796
2. Coal and power	14	-	12	8	9	4	16	8	3	39	8	31	39	15	129
3. Building, building materials and timber	10	3	-	5	1	1	7	10	2	129	12	9	280	56	53
4. Chemicals and rubber	25	2	14	-	15	6	2	19	3	7	6	38	2	6	73
5. Textiles and clothing	4	*	8	5	-	4	*	8	*	16	5	125	2	2	298
6. Paper, printing and miscellaneous	18	1	4	13	6	-	*	5	2	48	14	15	2	6	69
7. Metal making	2	5	28	3	*	1	-	62	30	*	7	38	6	*	5
8. Engineering	5	16	21	*	4	1	2	-	1	39	17	79	173	40	79
9. Metal goods	7	1	16	5	3	1	2	22	-	*	17	45	9	2	32
10. Services	72	20	46	16	41	13	18	28	7	-	35	352	26	304	1907
Not allocated	13	8	7	5	18	5	20	10	6	30	-	15	-3	3	-5
Imports	130	5	41	48	102	23	30	16	37	19	9	52	18	8	266
** Government	176	5	10	7	13	4	2	7	2	256	13	-	7	2	85
*** Primary input	385	268	369	98	258	147	86	282	70	2280	-18	-	-	-	-
Total Gross input	958	534	577	217	476	208	188	477	162	2665	132	600	569	450	5708

(Note - Total gross output for industries = total gross input for the corresponding industry)

* Less than 1/2 million, ** indirect taxes - subsidies, *** payment to primary factors.

(b) TECHNO-COMMERCIAL STATISTICS - PracticalGROUP A - Statistical Quality Control

(. Answer any two questions from this group. Use
separate answer book for this group).

1. An article is submitted for inspection in lots of size 1000 and is proposed to be accepted or rejected on the basis of the following double sampling plan

$$\begin{array}{ll} n_1 = 25 & c_1 = 0 \\ n_2 = 35 & c_2 = 2 \end{array}$$

The production process is known to be under control at an average of 2.5 % defectives.

What is the probability of acceptance of a lot?

What is the AOQ under this scheme and the average number of items inspected if the rejected lots are completely screened?

The cost of inspecting an item under this scheme is Rs.0.04, where as under a scheme of 100 % inspection of all items submitted it will be only Rs.0.01. 100 % inspection would remove only 80 % of the defectives where as the sampling plan can be assumed to remove all defectives from the inspected items. The loss incurred in using a defective article is Rs.1.50. Compare the total costs of accepting the product under 100 % inspection and the above sampling scheme and make suitable recommendations. [25]

2. The following data refers to the outside diameter of a component made on a 5 spindle automatic lathe. 5 items were examined at a time one from each spindle. The observations in coded units, taken every 15 minutes are as follows.

Sample No.	Spindles				
	1	2	3	4	5
1	0	4	7	28	29
2	7	16	21	49	20
3	20	22	12	35	24
4	5	15	9	2	6
5	14	15	16	30	20
6	4	27	26	26	18
7	1	26	12	20	25
8	1	20	20	35	14
9	6	22	27	31	19
10	3	14	17	28	22
11	6	15	22	37	21
12	13	16	15	31	19

Make a suitable analysis of the data and examine if the spindles work with the same variability and the same mean.

What would be the process capability if the differences between spindles could be removed?

Could you make any suggestion regarding process settings on the basis of the above data? [25]

15 electric bulbs were tested for life and the test was terminated after six of these were burnt out. The life of these six bulbs (hrs.) were 652, 764, 791, 805, 807 and 862.

Obtain an estimate of the mean and variance of the life in the population, on the assumption that the length of life has a simple exponential distribution. [25]

Please Turn Over .

GRCUP B - Operations Research Or Elements of Bookkeeping & Accountancy (Practical)

(use separate answer-script for this group)

(1) Operations Research
(Answer any two questions)

1. Solve the following linear programming problem by the simplex technique or any other technique that you know.

$$\text{Maximize } Z = 3x_1 + 5x_2 + 4x_3$$

Subject to the restrictions

$$2x_1 + 3x_2 \leq 8$$

$$2x_2 + 5x_3 \leq 10$$

$$3x_1 + 2x_2 + 4x_3 \leq 15$$

$$x_1, x_2, x_3 \geq 0.$$

(15)

2. The following table gives the distribution of service-time from a random sample of 31. Test if the distribution of the service-time is exponential with constant mean μ . Work out the estimates of mean and variance of the service-time. Comment.

Service-time in Minutes	Number from sample
< 1	5
1 - 2	7
2 - 3	6
3 - 4	4
4 - 5	2
5 - 6	3
6 - 7	1
7 - 8	1
8 - 9	1
9 - 10	0
> 10	1
Total	31

(15)

3. The following table gives the distribution of lead time for a stock item for a retail merchandise (a soft-good) in a department store. Examine whether the normal distribution gives a suitable generating function for the lead-time. Comment.

Lead time	number of weeks
1	0
2	0
3	1
4	3
5	5
6	5
7	4
8	1
9	2
10	1
Total	25

(15)

Please turn over

(11) Elements of Bookkeeping & Accountancy

4. On 31.12.1962, the cash book of B. Basu showed an overdraft of 18,000 with the Bank of India Ltd. This balance did not agree with the pass book balance and you find that Basu had paid into the Bank on 26th December, four cheques for Rs. 10,000; Rs. 12,000; Rs. 6,000 and Rs. 8,000. Of these the cheque for Rs. 6000 was credited by the Bank in January 1963. Basu had issued on 24.12.1962, three cheques for Rs. 15,000; Rs. 12,000 and Rs. 7,000. The first two cheques were presented to the bank for payment in December and the third cheque in January 1963. You also find that on 31.12.1962, the bank had debited Basu's account with Rs. 500 for interest and Rs. 20 for charges, but these were not recorded in Basu's books.

You are required to prepare a Bank Reconciliation Statement as at 31.12.1962. (8)

5. Prove that 25% profit on cost price is equivalent to 20% on sale price. (8)

6. Two candles of equal length are consumed uniformly in 5 & 6 hours respectively. How long after being lighted together will one be four times as long as the other? (8)

7. L 1 can buy 4.85 dollars or 100 francs. Which shall be better to pay a French bill in London or New York if a dollar buys 19 francs? (5)

CRCP C - Statistical Methods in Business

(Use separate answer-script for this group)

(Answer any two questions)

8. The following table gives percentage errors of sales forecasts for sales of last half of the year, based on three alternative forecasting procedures.

Which method is to be preferred most? Comment.

Year	percentage errors of		
	Forecasts based on first quarter investment	Forecasts based on same level as in preceding year	Forecasts based on second quarter investment
1948	- 16.8	-	- 14.0
1949	- 0.1	3.5	0.4
1950	- 11.0	- 21.0	0.6
1951	3.2	20.6	1.3
1952	1.2	- 8.9	0.8

(10)

9. The following two-way table gives the distribution of jobs by the point values and wages per hour of 50 jobs in a company. Work out the straight line wage curve for the company. Interpret the results.

Please turn over

Point-values

Age	80	100	120	140	160	180	200	220	240	260	280	300
per hour	-99	-119	-139	-159	-179	-199	-219	-239	-259	-279	-299	-319
1.80 - 1.89										2	1	
1.70 -								1	1	2		
1.60 -							1	1	1		1	
1.50 -								1	2			
1.40 -							1	1	1			
1.30 -							2	3				
1.20 -					1							
1.10 -			1	2	1	1						
1.00 -			1	1		1						
.90 -		2	3	1								
.80 -	3	3	1									
.70 -	4	3										
Total	7	8	6	4	1	3	4	5	5	1	4	2

(10)

3. The following table gives the results of two studies made twelve months apart. Each study was made in over 75 cities and towns (in USA) properly distributed by city size, geographic area, 50% men and 50% women properly distributed by socio-economic groups. The question asked was "What brand of dentifrice have you seen or heard advertised lately?" Test if there is a change in the distribution and also if the drop in the percentage for Brand '1' is significantly different from zero.

[10]

Brand	percentage of persons*	
	October 1937	October 1938
1 A	39.4	25.7
2 B	18.4	18.6
3 C	13.7	17.4
4 D	6.2	16.7
5 E	4.2	4.9
6 F	3.5	4.1
7 Others or none	14.6	12.6
Total	100.0	100.0
Total interviews	5000	5000

*percentage of persons giving the name of the brand to the left in answer to the question put to them.

oooooooo

Please Turn Over

(d) DESIGN AND ANALYSIS OF EXPERIMENTS - Practical

(Answer any three questions)

1. Give a plan for a $\frac{1}{4}$ th. replicate of an experiment involving 7 factors each at 2 levels in blocks of 8 plots, explaining carefully the alias set-up and the confounding scheme. (32)

2. The layout and yields of a varietal trial involving 7 varieties A, B, C, D, E, F, G arranged in a Youden Square with 4 replications are given below. Analyse the data and write report.

A	C	G	D
60	72	78	51
F	B	C	G
35	26	67	83
D	A	B	F
41	53	24	39
E	G	F	A
18	80	36	59
C	F	D	E
65	39	44	18
B	E	A	C
30	20	58	74
G	D	E	B
72	48	16	30

(32)

Please turn over

3. The following experiment was conducted to test whether subjecting the seed to temperature treatment before planting has any effect on yield. Five pots were taken and in each 10 seeds were planted at random, 5 pairs of the seeds being subjected to 5 different levels of temperature. The yields from the individual plants in grams are given below.

Analyse the data to find the optimum temperature for treatment of seeds, if any.

Pots	Temperature									
	60°F		75°F		90°F		105°F		120°F	
	P ₁	P ₂	P ₁	P ₂	P ₁	P ₂	P ₁	P ₂	P ₁	P ₂
1	29	30	30	37	43	41	43	43	38	39
2	34	32	41	44	47	46	47	47	42	41
3	25	26	35	37	40	39	40	41	37	37
4	31	29	39	38	43	42	44	46	42	41
5	34	34	46	44	49	49	48	50	45	43

(32)

4. A chemical experiment was carried out on a laboratory scale to maximise the product of a chemical reaction between two constituents A and B in the presence of a catalytic agent C. Determination of the yield was made by three chemists, using three different types of catalytic agents (C₁, C₂ & C₃) and three different amounts of each of the two constituents A and B, namely: A₁ = 5.00, A₂ = 4.00, A₃ = 2.50 and B₁ = 336.6, B₂ = 250.0, B₃ = 200.0 in suitable units. The yields under the different experimental arrangements are given below.

Analyse the data and give your recommendation about the optimum quantities of the constituents A and B and the type of the catalytic agent. What is the maximum yield to be expected under optimum experimental conditions?

chemist 1				chemist 2				chemist 3			
serial no.	levels of A	levels of B	yield	serial no.	levels of A	levels of B	yield	serial no.	levels of A	levels of B	yield
1	1	1	6.3	10	1	1	7.2	19	1	1	6.3
2	2	1	8.7	11	2	1	7.6	20	2	1	7.7
3	3	1	5.8	12	3	1	5.7	21	3	1	6.0
4	1	2	9.3	13	1	2	9.5	22	1	2	10.7
5	2	2	10.5	14	2	2	11.3	23	2	2	10.1
6	3	2	9.7	15	3	2	8.0	24	3	2	9.0
7	1	3	9.0	16	1	3	8.2	25	1	3	8.7
8	2	3	9.6	17	2	3	10.0	26	2	3	10.6
9	3	3	7.7	18	3	3	8.0	27	3	3	7.4

(32)

Neatness.

ooooo

(4)

(e) SAMPLE QUIZZES - Practical

(Answer all questions)

1. For the purpose of constructing a cost of living index number for the working class of a certain region, a rapid family budget survey is to be undertaken.

(a) Design a proper schedule of not more than two pages, in which all the information would be collected.

(b) Draw up a scrutiny programme for your schedule.

(c) Draw up a 80-column punched card design for the schedule.

(25)

2. The following table shows the number of firms classified in several strata according to the total assets of these firms.

Total assets (000 dollars)	No. of firms	Average net income (000 dollars)	S.d. of net income (000 dollars)
	N_i	\bar{Y}_i	σ_i
unknown	5600	1	5
under 50	28700	1	5
50 - 99	11100	5	8
100 - 249	15000	15	20
250 - 499	7500	50	65
500 - 999	5100	100	130
1000 - 4999	5800	300	390

In each stratum the cost of sampling is $500 + 2n_i$ dollars where n_i is the sample size in the i th stratum. The total amount available for sampling survey is 5000 dollars.

(a) Determine the proportional and optimum allocation for the survey, and calculate the corresponding variances of estimated average net income.

(b) The coefficient of variation (ignoring fpc) of the estimated average of the i th stratum if the sample size is n_i is $C_i / \sqrt{n_i}$,

where $C_i = \sigma_i / \bar{Y}_i$.

Obtain an allocation such that the coefficient of variation in each stratum would be equal.

(25)

3. In a yield survey of paddy in a certain district, 5 villages were selected in each of the 7 strata into which the district was divided; 3 fields were selected in each village and one plot of $\frac{1}{100}$ acre harvested in each field.

Stratum number	No. of villages population	villages sample	sample mean or/plot	m.s. between village mean $\sigma_{b_i}^2$	m.s. within village $\sigma_{w_i}^2$
1	88	5	337.5	1452	2791
2	142	5	297.8	1937	27442
3	119	5	201.1	7107	1864
4	90	5	433.9	9604	11824
5	114	5	292.9	20702	13628
6	102	5	211.9	2511	2008
7	146	5	186.7	3342	7442

(a) Calculate the pooled values σ_b^2 and σ_w^2 for the entire district.

(b) Obtain the population estimates S_w^2 and S_b^2 from

$$S_w^2 = \sigma_w^2; \quad S_b^2 = \sigma_b^2 - \sigma_w^2 / m,$$

m denoting the number of fields sampled in each village.

(c) Let n be the total number of villages sampled, the number in each stratum being proportional to the stratum size N_i .

Please turn over

If m is the number of fields selected in each village, the variance of the estimate \bar{y}_{nm} of the population mean yield is given by

$$V(\bar{y}_{nm}) = S_b^2 \left(\frac{1}{n} - \frac{1}{N} \right) + \frac{S_w^2}{nm}$$

- (i) Express n in terms of S_b^2 , S_w^2 , N and $V(\bar{y}_{nm})$
 (ii) If a desired accuracy - $V(\bar{y}_{nm}) = 144 -$ is given, obtain the values of n corresponding to $m = 1, 2, 3, 4, 5$.
 (iii) If cost of sampling is of the form

$$C = 7n + 2nm,$$

tabulate the cost for each of the pairs (m, n) of (ii) above and obtain the pair which takes the minimum cost for the given accuracy. (25)

4. The following table gives the populations in 1920 and 1930 of a simple random sample of 20 cities of the U.S.A. out of a total number 196 of cities. The total 1920 population of these 196 cities is 22919000.

(Population in thousand)

Population in 1920	Population in 1930	Population in 1920	Population in 1930
243	291	36	46
87	105	161	232
30	111	74	93
71	77	45	53
256	288	36	54
43	61	50	58
25	57	48	75
94	85	76	80
43	30	138	143
298	317	67	67

- (a) Obtain the ratio estimate of the total 1930 population of the 196 cities. Also calculate an approximate sampling variance of this estimate.
 (b) Is this ratio estimate unbiased? If not, obtain an estimate of the bias. (25)

ooooo

(f) TECHNIQUE OF COMPUTATION - Practical

(Answer 11 questions)

1. Using the IBM alphabetic code, punch on card columns 5 - 32 the words INDIAN STATISTICAL INSTITUTE with one blank column between every two consecutive words. (10)

Please turn over

(g) STATISTICAL INFERENCE - Practical

(Answer any two questions)

1. Let X_1, X_2, \dots, X_n be a sample of size n from a Normal population with mean μ and standard deviation σ and let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. For $n \geq 20$, one can approximate the sampling distribution of $\frac{s\bar{X} - \mu}{\sigma/\sqrt{2n}}$ by a Normal distribution with mean 0 and standard deviation $\sigma/\sqrt{2n}$. Using this approximation, obtain an expression for the probability that $\bar{X} > k s$ in terms of the Normal probability Integral.

Use this expression to evaluate the power function for the student's t test for the hypothesis $\mu = 0$ against the alternative $\mu > 0$ at the 5% level of significance for $n = 50$ and

$$\frac{\mu}{\sigma} = 0.2 \quad (0.2) \quad 1.0, \quad 1.5 \quad \text{and} \quad 2.0. \quad [50]$$

2. The following table gives the mean values of measurements on three biometrical characters, and the matrix of pooled variances and covariances for two groups of female desert locusts, one in the phase gregaria and the other in a phase intermediate between gregaria and solitaria.

characters	means		pooled variances and covariances based on 20+72=92 df.		
	gregaria (n=20)	intermediate (n=72)	x_1	x_2	x_3
x_1	25.80	28.35	4.7350	0.5622	1.4605
x_2	7.81	7.41		0.1431	0.2174
x_3	10.77	10.75			0.5702

Evaluate the linear discriminant function for the two phases, and the D^2 -statistic and test for the significance of the latter. [50]

3. In a population under pan-mixis the frequencies of the blood groups O, A, B and AB are given respectively by r^2 , $p^2 + 2pr$, $q^2 + 2qr$ and $2pq$, where p , q , and r , ($p+q+r = 1$) are the frequencies of the genes C^A , C^B , and g respectively. The following table gives the frequencies of the four blood groups in a sample from a certain population.

Blood group	Frequencies
O	176
A	182
B	60
AB	17

Estimate the gene frequencies by the method of maximum likelihood and work out their standard error. [50]

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination, November 1963

Paper I : Official Statistics and Descriptive Statistics (Theoretical)

Time : 4 hours

Full marks : 100

Figures in the margin indicate full marks for each question.

GROUP A

(Answer any three questions from this group)

1. What are the various publications giving data in regard to area and yield of principal crops in India? (8)
- Discuss the method of compilation, reliability and coverage of published data which you have enumerated above. (8)
2. What are the different methods of estimation of national income? Describe them briefly and establish that they lead to the same aggregates. (8)
- What are the special problems arising in the estimation of national income and what are the uses, in the context of planning, of national income estimates? (8)
3. Define consumer price index numbers. Name the publications giving these indices for the whole of India and for selected centres. (9)
- How is the all India working class consumer price index constructed? Discuss its reliability and limitations. (8)
4. Write a critical note on the present position of statistics of agricultural prices in India. Give the names of the publications containing these statistics. (10)
- What suggestions have you to offer to improve their reliability? (6)
- Neatness. (2)

GROUP B

(Answer any three questions from this group)

5. For the binomial distribution with parameters n and p , prove the following recurrence relation for central moments

$$\mu_{r+1} = [(1+E)^r - E^r] (npq \mu_0 - p \mu_1)$$

where $\mu_0 = 1$, $\mu_1 = np$, $q = 1-p$ and E is an operator such that $E \mu_r = \mu_{r+1}$. Hence find μ_2 and μ_3 . (10)

In a precision bombing, the probability of hitting a target is $\frac{1}{2}$. Two direct hits are required to destroy the target. How many bombs must be dropped to ensure a probability of 0.99 or more of completely destroying the target? (6)

6. Define the concept of regression of a random variable on another random variable. (4)
- If X and Y follow a bivariate normal distribution, show that the regression of X on Y is linear. Find the variance of X about its regression value when Y has a fixed value y . (12)

Please turn over

- 7.(a) If X_1, X_2, \dots, X_n are n random variables with the same variance σ^2 and the correlation coefficient between any two of them is ρ , then prove that

$$V(\bar{X}) = \frac{\sigma^2}{n} + (1 - \frac{1}{n}) \rho \sigma^2$$

$$E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = (n-1)(1-\rho)\sigma^2$$

and

$$\rho > -\frac{1}{n-1}$$

where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad [11]$$

- (b) Define the partial correlation coefficient between X_1 and X_2 when X_3 is kept fixed and hence express this partial correlation coefficient in terms of the three total correlation coefficients. (5)

Hence.

(2)

o o o o o

Please turn over.

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1963

Paper II: Probability Theory and Statistical Methods (Theoretical)

Time: 4 hours

Full marks: 100

Figures in the margin indicate full marks for each question.

Group A

(Attempt any three questions from this group)

1. (a) Find the probability of occurrence of at least one of the n events A_1, A_2, \dots, A_n in terms of S_1 's where S_1 is the sum of $n C_1$ probabilities each corresponding to the simultaneous occurrence of some 1 of the n events. [9]
- (b) Two dice are thrown n times. Find the probability that each of the six combinations (1, 1), (2, 2), ..., (6, 6) occur at least once if $n \geq 6$. [7]
2. (a) Explain the concept of the mutual independence in probability of $n \geq 2$ events. Show that in this case there are $2^n - n - 1$ independent relations involving probabilities which must be satisfied in order that the n events may be said to be mutually independent. [8]
- (b) Give an example to show that pairwise independence of n events does not imply the mutual independence of n events in probability. Assuming that the probability of a child being male or female is $\frac{1}{2}$, consider whether the following events A_1 and A_2 are independent. [4]
- (c) In a family of three children A_1 : 'The family has children of both sexes' and A_2 : 'There is at most one girl'. [4]
3. (a) State and prove Tchebycheff's inequality. [8]
- (b) $\{X_j\}$ is a sequence of independent random variables with $E(X_j) = 0$ and $V(X_j) = \sigma^2$. Prove that for any $\epsilon > 0$,
- $$P \left\{ \left| \frac{\sum_{j=1}^n X_j}{n} \right| > \epsilon \right\} \rightarrow 0 \text{ as } n \rightarrow \infty. \quad [4]$$
- (c) Show that the above result holds even when $E(X_j) = 0$ and $V(X_j) = \sigma^2 \sqrt{j}$. [5]
4. State the distribution of chi-square with n degrees of freedom and hence derive the distribution of F ratio with n_1 and n_2 degrees of freedom for the numerator and denominator respectively. Obtain the first two moments of F . [4+8+5]
5. (a) State any form of the central limit theorem and indicate its applications. [7]
- (b) Prove the normal approximation to the binomial distribution. [9]

Please Turn Over

Group B

Answer any two questions from this group.

6. What is a confidence interval? Find a confidence interval for σ_1^2/σ_2^2 for two normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 , respectively, when two independent random samples of size n_1 and n_2 have been taken from these populations. Give the modification of the procedure when the means μ_1 and μ_2 are known and show that this gives an improved confidence interval. [25]
7. Explain what is meant by unbiasedness and efficiency of an estimate and illustrate with examples. Do these properties hold for maximum likelihood estimates? Samples of size n_1, n_2, n_3 are taken from three normal populations with the same mean μ and variances $\sigma_1^2, 3\sigma_1^2, 5\sigma_1^2$ respectively. Find the maximum likelihood estimate of μ and examine if this is unbiased. [25]
8. Explain the procedure of analysis of variance in a two way classification with equal number of observations in each cell, stating clearly the assumptions involved and the rationale of the test. What happens if the numbers of observations in the cells are unequal? [25]
9. Explain what is meant by a uniformly most powerful test of a hypothesis. If x_1, \dots, x_n is a sample from a normal population with mean μ and variance unity, show that a uniformly most powerful test of the hypothesis $H: \mu = 0$, exists. [25]

ooo bcc

Please turn over

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination, November, 1965

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

Time : 4 hours

Full marks : 100

Figures in the margin indicate full marks for each question.

GROUP A

(Answer any three questions from this group)

1. (a) What are the various uses of pilot surveys ? (3)
- (b) What are meant by cost and variance functions with reference to large-scale sample surveys and in what ways are they useful for such surveys? What are the important components of the cost function ? (10)
2. (a) Discuss the advantages of stratified sampling in comparison with unstratified sampling. (4)
- (b) A population is divided into K strata. If m_i units are selected at random from the N_i units in the i^{th} stratum ($i = 1, 2, \dots, K$), find an unbiased estimate of the population total and obtain its standard error. (4)
- (c) Compare the efficiencies of stratified sampling with Neyman allocation, stratified sampling with proportional allocation and simple random sampling without stratification. (8)
3. (a) Define a ratio estimate for estimating the population total of a character y , and derive an expression of the standard error of the estimate. (6)
- (b) If the coefficient of variation of the auxiliary variate 'x' is more than twice the coefficient of variation of the character 'y', show that in large samples, with simple random sampling, the ratio estimate is less precise than the 'mean per sampling unit' estimate. Is the converse true ? (5)
- (c) State the conditions under which the ratio estimate is an 'unbiased estimate'. (5)
4. Write critical notes on :-
- (i) Partial replacement in sampling on successive occasions. (5/3)
- (ii) Double sampling. (5/3)
- (iii) Use of biased estimators. (5/3)
- Neatness (2)

GROUP B

(Answer any three questions from this group)

5. (a) Describe a randomised block experiment. (2)
- (b) Explain fully the implications of the null hypothesis of equality of treatments, discuss how proper randomisation in a randomised block experiment provides a valid test of this hypothesis. (7)

Please turn over

(c) A manurial experiment on paddy involving 5 organic manures is laid out in six randomised blocks of five plots each. The same experiment is repeated in the following year, without any additional fertiliser being added during the second year of experimentation. The hypothesis of non-interaction between treatments and years would imply that the decline in yield in the second year of experimentation remains the same irrespective of the treatment applied initially.

Discuss how you would use the two years' experimental data to test this particular hypothesis. (7)

6. (a) Describe how, by analysis of covariance, one may use data on supplementary variables to increase the precision of estimates. (5)

(b) What should be the nature of such concomitant variables in order that the analysis of covariance may most profitably be used for the above purpose? (5)

(c) Describe the analysis of covariance for a Latin Square Experiment with one concomitant observation. (6)

7. The following was the layout of an experiment involving 16 varieties in blocks of 4 plots each.

Replication 1				Replication 2			
10	12	11	9	15	10	8	1
1	2	4	3	9	2	7	16
15	14	13	16	3	6	12	13
5	8	7	6	5	11	4	14

(a) Identify the type of design. (4)

(b) Describe the analysis appropriate for this design. (8)

(c) Compute the efficiency factor of this design. (4)

8. (a) Explain the principle of 'confounding - both total and partial' in connection with factorial experiments. Illustrate your answer with 2^3 experiments. (5)

(b) Discuss, by means of an example, how in the presence of block X factor interactions, a confounded design might give rise to misleading interpretation of results. (5)

(c) The following table gives the plan for each one of the two replications of a 2^5 experiment in blocks of size 8.

Block			
1	2	3	4
(1)	ab	a	b
	bc	abc	c
	abd	d	bd
	acd	bd	abcd
	abc	e	ac
	ace	bce	abc
	de	abde	bcde
	bcde	acde	cd

Identify the confounded effects. (6)

Neatness.

Diagrams allowed

(2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination, November 1963.

Paper IV : Applied Statistics (Theoretical)

Time : 4 hours

Full marks : 100

- i) Answer questions only from the two groups relating to the subjects you have opted for.
- ii) Use a separate answer book for each group.
- iii) Figures in the margin indicate full marks for each question.

GROUP A - ECONOMIC STATISTICS

(Answer any three questions from this group)

1. Briefly describe the methods of calculation of index numbers of (i) volumes of export and import and (ii) prices of exported and imported articles. How are these index numbers useful? (6+6+ $\frac{1}{2}$)
2. What do you mean by cyclical movement in a time series? Describe a method of calculation of cyclical component in a time series and indicate its use in forecasting. (4+8+ $\frac{1}{2}$)
3. Explain the concept of elasticity of demand with respect to income and indicate the type of data required and the method of estimation for such an elasticity. How is this elasticity useful in planning. (3+9+ $\frac{1}{2}$)
4. What is an input-output table? Briefly describe the method of construction of such a table. (6+10 $\frac{1}{2}$)
5. How will you calculate national income of India as the sum of net outputs of various industries? How does the breakdown of national income by industrial origin serve as an indicator of economic growth? What changes are found in this respect in India during the plan period? (6 $\frac{1}{2}$ +5)
6. Write notes on any two of the following : (0 $\frac{1}{2}$ +8 $\frac{1}{2}$)
 - i) production function
 - ii) capital formation in India during the plan period
 - iii) Pareto's law of income distribution.

Please turn over

GROUP B - STATISTICAL QUALITY CONTROL

(Answer any three questions from this group. Neatness carries 2 marks)

7. What do you understand by the term 'process capability' when referred to (a) a measurable characteristic and (b) an attribute characteristic. Detail logically the steps to be taken in estimating the capability of a process. Also outline the different types of action that may be needed on the basis of your findings. (16)
8. Describe briefly the considerations involved in starting and maintaining an \bar{X} - R control chart with special reference to the choice of the characteristic, the sub group size, the method and frequency of sampling, the accuracy of measurements, the determination of trial control limits and the revision of the chart on the basis of experience. (16)
- 9.(a) Explain the terms 'producers risk' and 'consumers risk' in respect of an acceptance sampling plan. (6)
- (b) Describe briefly the criteria on the basis of which Dodge-Romig plans have been constructed. (6)
- (c) Consider the following Dodge-Romig plan
- | | |
|-----------------|---------------|
| A O Q L | 2.0% |
| Process average | 1.21% - 1.60% |
| Lot size | 4001 - 5000 |
| n | 155 |
| c | 5 |
- What is the effect of the use of the above plan on the A O Q L and the amount of inspection (a) when the lot size is 50000 and also (b) when the process average is 2.5%. (4)
10. Write short notes on
- acceptance plans for variables
 - interval estimates of lot quality
 - response surface studies. (16)

Please turn over

GROUP D - VITAL STATISTICS AND DEMOGRAPHY

(Answer any three question from this group.
Neatness carries 2 marks)

11. (a) What do you understand by "vital events" and "vital registration" ?
 (b) How is vital registration organized in India ?
 (c) Describe the best method you know of checking the completeness of
 (i) birth and (ii) death statistics. (16)

12. (a) Describe three indices for the measurement of mortality, indicating their suitability with reference to the availability of data.
 (b) What are "direct" and "indirect" standardization of death rates ? How is the "standard population" chosen ?
 (c) Explain the difference between
 m_x = death rate at age x last birthday,
 and
 q_x = probability of dying within one year of attaining age x .
 (d) Show that, under certain assumptions (to be stated),

$$q_x = \frac{2m_x}{2 + m_x} \quad (16)$$

13. (a) What are the assumptions regarding a population under which a life table is constructed ?
 (b) Describe one method of "smoothing" the probabilities q_x of dying, from which a life table is constructed.
 (c) Enumerate the columns of an abridged life table, and describe the steps in their computation. (16)
14. Describe a method of age- and sex-wise projection of population using
 (a) only census data;
 (b) census and registration data.

In each of these cases, state the assumptions which need to be made. (16)

ooo000

Please turn over.

INDIAN STATISTICAL INSTITUTE
 Statistician's Diploma Examination - November 1963

Paper V: Methods of Numerical Computation, Descriptive
 Statistics and Official Statistics (Practical)

Time: 5 hours

Full marks 100

- i) Figures in the margin indicate full marks for each question.
 ii) Use of Calculating machines is permitted.

Group A

(Answer any two questions from this group)

1. By means of a suitable quadrature formula, find the value of the integral

$$I = \int_{-1}^1 \frac{x \sqrt{1-x^2}}{(2-x)^{3/2}} dx. \quad [12\frac{1}{2}]$$

2. The following table gives the values of the probability integral

$\frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2} dx$ corresponding to certain values of x . Find the value (to six places of decimals) of x for which this integral is equal to $\frac{1}{2}$.

x	$\frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2} dx$
0.45	0.4754818
0.46	0.4846555
0.47	0.4937452
0.48	0.5027498
0.49	0.5116683
0.50	0.5204999

[12 $\frac{1}{2}$]

3. Obtain a solution of the following set of simultaneous equations in x, y, z :

$$0.10719x - 0.1295y + 0.26600z = -0.20808$$

$$0.23515x + 0.0527y + 0.03436z = 1.04327$$

$$0.23599x - 0.0317y + 0.52409z = -0.56525$$

[12 $\frac{1}{2}$]

Group B

(Answer question 4 and any two of the rest)

4. WATER

The following tables give the breaking strength of 150 samples of paper, in certain units:

Breaking strength (cm^2/cm^2)	Frequency
1750 - 1800	2
1700 - 1750	1
1650 - 1700	6
1600 - 1650	13
1550 - 1600	23
1500 - 1550	21
1450 - 1500	24
1400 - 1450	19
1350 - 1400	24
1300 - 1350	11
1250 - 1300	6

- a) Calculate β_1 and β_2 and state (without performing any statistical test) whether a normal distribution would fit the data. [10]
- b) Assuming a normal distribution, calculate the expected frequencies in the classes 1750-1800, 1700-1750, 1550-1600, 1500-1550 and 1250-1300. [6]
- c) Calculate the contribution of these class-intervals to the χ^2 test of goodness of fit. [4]

OR

The following table shows the frequencies of the values 0, 1, ..., 10 of a variable in 100 observations

Value of variable	0	1	2	3	4	5	6	7	8	9	10
Frequency	2	5	11	40	8	9	7	7	6	4	1

- a) Represent the data in a suitable diagram. [6]
- b) Assuming the observations to have come from a Poisson distribution, calculate the expected frequencies and plot them on the diagram in (a). [10]
- c) Calculate the variance of the observations and state whether, from the value obtained, anything can be said regarding the assumption of a Poisson distribution. [4]
5. The following table shows age of husband (X) and age of wife (Y) in 100 couples:

Age of wife (years)	Age of husband (years)					Total
	20-30	30-40	40-50	50-60	60-70	
15 - 25		5	9	3		17
25 - 35			10	25	2	37
35 - 45			1	12	2	15
45 - 55				4	16	20
55 - 65					4	4
Total	5	20	44	24	7	100

- a) Calculate the coefficient of correlation between X and Y, and find the equation of linear regression of Y on X. [9]
- b) Using the equation of regression, estimate the values of Y for X = 35, 45, 55. [2]
- c) Calculate the average value of Y in each of the arrays corresponding to the mid-points of X values indicated in (b), and comment on the results of (a) and (b). [4]

Please Turn Over

6. The following table shows the scores X_1 and X_2 obtained by 15 workers in two psychological tests, as well as their efficiency index Y , calculated by very complicated methods.

Serial No.	Score in Test I (X_1)	Score in test II (X_2)	Efficiency index (Y)
1	6	41	43
2	4	46	36
3	97	66	72
4	53	31	54
5	26	31	35
6	49	33	40
7	3	27	34
8	21	55	48
9	51	55	52
10	72	57	49
11	97	58	67
12	18	23	27
13	53	29	39
14	60	60	57
15	9	12	28

- a) Calculate the partial correlations between X_1 and Y , eliminating X_2 , and between X_2 and Y eliminating X_1 . [7]
- b) Obtain the equation of linear regression of Y on X_1 and X_2 . [5]
- c) From the equation in (b), estimate the value of Y from X_1 and X_2 for serial numbers 1, 5, 10 and 15 and comment on the results. [3]
7. The following table shows the month-wise average number of eggs per hen in a region for the years 1930-1942.

Year	Average no. of eggs per hen											
	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1930	7.9	9.9	15.4	17.5	17.3	14.9	13.6	11.0	9.4	7.5	5.9	6.4
1939	8.0	8.7	14.9	17.0	17.0	14.6	13.2	11.7	9.3	7.4	6.0	6.8
1940	7.2	9.0	14.4	16.5	17.0	14.8	13.4	11.0	9.7	7.9	6.2	6.8
1941	7.0	8.7	14.2	16.0	17.2	14.5	13.0	11.5	10.0	8.1	6.0	6.5
1942	8.1	8.5	15.3	17.2	17.5	15.1	13.5	11.0	9.8	7.9	6.2	6.9

- a) Calculate the seasonal indices for each month. [10]
- b) Obtain the values of average number of eggs for every month of the year 1940, corrected for the seasonal effect. [5]

Group C

(Answer any one question from this group)

- 8.a) Name the publications in which any three of the following are available in India, the authority or organization issuing them, the periodicity with which they are issued and the break downs by which the informations are available:-
- i) Statistics of civil aviation. [3]
 - ii) Public health statistics. [3]
 - iii) Statistics of agricultural implements. [3]
 - iv) Irrigation statistics. [3]
 - v) Statistics of manufacturing industries. [3]
- b) From the publications supplied to you, collect the data in respect of any three of the following, presenting them in neat tabular forms, with proper headings and foot-notes, if necessary. Mention the source from which you have compiled the data.
- i) Index numbers of industrial profits in India for any five consecutive years. [5½]
 - ii) Statistics of employment in plantations in India for any five consecutive years. [5½]
 - iii) Prices of gold and silver in India for the latest twelve consecutive months for which these are available. [5½]
 - iv) Statistics of balance of payment of India on current account currency area-wise for four quarters in any year. [5½]
 - v) National income statistics for India for any one year. [5 1/3]
- 9.a) Collect relevant information from the publications supplied in respect of passenger traffic and earnings, and freight traffic and earnings, of railways for the latest twelve months for which these are available and write a critical note on the salient features of the data. Give the source from which you have compiled the data. [15]
- b) Comment on the method of collection and presentation of data, as also the time lag in publication of any three of the following publications issued by the Government of India:-
- i) Estimates of Area and Production of Principal Crops. [4]
 - ii) Monthly Production of Selected Industries in India. [4]
 - iii) Census of Indian Manufactures. [4]
 - iv) Bulletin of Agricultural Prices. [4]
 - v) Indian Agricultural Statistics. [4]

000000

Please turn over.

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination, November 1963

Paper VI : Statistical Methods, Design & Analysis of Experiments and Sample Surveys (Practical).

Time : 5 hours

Full marks : 100

- 1) Figures in the margin indicate full marks for each question.
 ii) Use of calculating machines is permitted.

GROUP(Answer all questions from this group)

1. In an investigation on the fat requirements of ruminants at the Indian Veterinary Research Institute, the influence of fat percentage on the total intake of digestion nutrients is being studied. The following data on the fat percentage (x) in ration and the total digestive nutrient (y) expressed as logarithms were recorded.

y	x	y	x
3.1959	0.94	3.2902	3.92
3.4320	1.93	3.2467	6.21
3.1637	1.04	3.0820	6.15
3.2138	4.06	3.1461	7.11
3.3705	3.25		
3.2997	3.27		
3.1322	8.74		
3.2201	9.23		
3.0166	9.93		
3.1090	1.69		
3.2565	1.72		
3.0913	1.74		
3.2618	4.02		
3.2401	3.97		

Obtain the regression equation of y on x and test the significance of the regression coefficient. (20)

2. (a) A fertiliser mixing machine is set to give 10 lbs of nitrate for every 100 lbs. of fertiliser. Ten 10 pound bags are examined. The percentages of nitrate in these bags are as follows

9.2, 12.3, 10.6, 9.1, 11.4, 11.7, 12.2, 9.5, 8.7, 10.1.

Is there any reason to believe that the mean is not equal to 10%?

- (b) Gram seed is treated chemically in an attempt to reduce the coincidence of blight. Of 100 plants from treated seeds, there were 20 which showed blight. Of the 70 plants from untreated seeds, 25 showed blight. Is this difference significant?

- (c) A series of experiments were carried out in 8 different regions to find the effectiveness of an insecticide. An F-test was used to test the hypothesis of no effectiveness, and the F-values were obtained as 2.76, 3.81, 8.45, 5.62, 1.23, 6.45, 0.96, 1.23 the degrees of freedom for each F-being 5 and 20. Obtain a combined test for the hypothesis and state the result. (20)

GROUP B

(Answer any two questions from this group)

3. In order to compare rubber treads prepared in four different ways (treatments A, B, C, D), four motor tyres (blocks 1, 2, 3, 4) were made, each in three different sections (plots) each section having its tread made in a particular way. The tyres were run on a test car and the loss of rubber on each section was estimated after the equivalent of a run of 10,000 miles. These are given below.

Treatment	Tyres			
	1	2	3	4
A	238	196	254	-
B	250	213	-	312
C	279	-	334	421
D	-	308	367	412

Analyse the data to find the best treatment.

(15)

4. The following table gives the results of an experiment to compare the yields of 4 varieties of wheat (M, S, V, T) carried out in 6 randomised blocks with one plot missing.

Analyse the data and give your comments.

Blocks	Varities	M	S	V	T
1		81.0	105.4	119.7	*
2		80.7	82.3	80.4	87.2
3		146.6	142.0	150.7	191.5
4		100.4	115.5	112.2	147.7
5		82.3	77.3	78.4	131.3
6		103.1	105.1	116.5	139.9

* missing

(15)

5. Draw up a scheme for a factorial experiment involving 5 factors each at 2 levels, using a maximum of 8 plots per block and a maximum of 5 replications, such that main effects and first order interactions are not confounded in any replication, and no interaction is totally confounded in all replications and, if possible, there is balanced loss of information on all partially confounded interactions. Name the interactions confounded in each replication and give the key block for any one replication.

(15)

GROUP C

(Answer both the questions)

6. A population has been divided into 5 strata for purposes of sampling, and crude estimates of the number of units (N_i), mean per unit (\bar{u}_i), and the standard deviation of the units (σ_i) within each stratum are given below.

stratum	number of units	mean per unit	standard deviation
	N_i	\bar{u}_i	σ_i
1	1200	75	15
2	600	125	20
3	2400	25	5
4	800	100	15
5	1800	50	10

The mean of the population is to be estimated on the basis of a sample of size $n = 1000$.

Compare the variance of the estimate for the following schemes of sampling :

- (i) simple sampling from the whole population
- (ii) stratified simple sampling with equal allocation
- (iii) stratified simple sampling with proportionate allocation
- (iv) stratified simple sampling with optimum allocation according to Neyman's formula. (20)

7. The results of 100 throws of an unbiased coin are given below, (where T and H stand respectively for "tail" and "head")

THTH	HPTH	THHT	THHH
HPTH	TTHT	HHTT	HTHT
HHTT	TTTT	THHT	THHT
THHT	HHTH	THHT	THHT
HHTT	TTTT	HHTT	HTTT

Using these results, draw a simple random sample of 2 villages with replacement from a list of 25 villages, numbered serially from 6132 to 6154 (both inclusive).

- Describe the procedure adopted by you and show all the details of how the sample is selected. (10)

00000000

Please turn over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination, November 1963.

Paper VII : Applied Statistics (Practical)

Time : 5 hours

Full marks : 100.

- (a) Answer questions only from the two groups relating to the subjects you have opted for.
 (b) Use a separate answer book for each group.
 (c) Figures in the margin indicate full marks for each question.
 (d) Use of calculating machines is permitted.

GROUP A - ECONOMIC STATISTICS

(Answer all questions from this group)

1. Import of Coffee to U.S.A. in million pounds, and the per capita disposable income in dollars in U.S.A. are given below for the period 1931-45. Obtain the linear trends of the series and compare their movements. (20 + 5)

year	import of coffee	per capita disposable income	year	import of coffee	per capita disposable income
1931	1749	510	1939	2021	540
1932	1508	380	1940	2062	570
1933	1592	360	1941	2260	690
1934	1531	410	1942	1723	670
1935	1761	460	1943	2198	970
1936	1747	520	1944	2607	1060
1937	1707	550	1945	2716	1080
1938	1991	500			

2. Per capita monthly expenditure in Rupres on food, intoxicants and services are given below along with total expenditure for different income classes in rural areas of North Bihar in 1960. Calculate the income elasticities for the above items and account for the difference. (20 + 5)

Income (monthly) class	no. of persons in the sample	total expenditure (per capita)	expenditure on		
			food	intoxicants	services
below 5	50	4.76	3.86	0.03	0.06
5 - 10	303	8.85	6.93	0.11	0.12
10 - 15	177	13.33	9.48	0.16	0.32
15 - 20	140	17.21	11.59	0.24	0.58
20 - 30	123	23.83	15.00	0.28	0.93
30 - 50	42	38.31	21.40	0.29	2.03
over 50	20	54.22	25.33	0.21	4.21

Please turn over

GROUP B - STATISTICAL QUALITY CONTROL

(Answer any two questions from this group)

- 3.. Samples of 5 cigarettes, successively produced, were taken from a machine at regular intervals and weighed. The data for 20 samples are given below in some arbitrary units.

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

- (a) Examine for the stability of the process.
 (b) The specifications for the weight are 4 ± 3 .
 Estimate the proportion of rejection in the current productions.
 (c) What is the process capability for this machine ?

Note 1. for a sample of size 5 the control chart factors with usual meaning are
 $D_3 = 0$, $D_4 = 2.11$, $A_2 = 0.58$.

(25)

4. Draw the O C curve of the plan

$$N = 300, \quad n = 70, \quad C = 1.$$

Draw also the A S N curve for the above plan if the rejected lots are completely screened.

(25)

5. (a) A p-chart indicates that the current process average is 0.02. If 50 items are inspected each day, what is the probability of detecting a shift to 0.04 (i) on the first day after the shift (ii) by the end of the third day after the shift ?

(10)

Please turn over

- (b) Obtain the 95% confidence limits for the following :
- (i) average number of defects per unit in the process, given that in a sample of 8 units there were in all 18 defects.
 - (ii) the process mean, given that the mean and standard deviation (divisor $\sqrt{n-1}$) of a sample of 15 items were 12.1 and 1.6 respectively.
 - (iii) the percent defective in the process, given that there were 4 defectives in a sample of size 45.

(1)

GROUP D - VITAL STATISTICS AND DEMOGRAPHY

(Answer any two questions from this group)

6. The following table shows age-wise population of a group in the 1937 census (column 2). In order to test the accuracy of these figures, the population of the same group in the 1947 census, according to the age this latter population had in 1937, is tabulated. Also, 10-year survival ratios for this population are taken from a life-table.

Please turn over

Age	Census population 1947	Census population 1947 (by ages in 1937)	Survival ratio $\frac{L_x + 10}{L_x}$
0 - 4	1021900	1142332	
5 - 9	1107879	984033	.88036
10 - 14	1030949	677765	.88945
15 - 19	713185	685730	.90487
20 - 24	539659	620074	.90530
25 - 29	616659	659225	.88327
30 - 34	557875	569069	.84004
35 - 39	600423	428502	.78160
40 - 44	414768	421222	.73330
45 - 49	345106	171105	.70092
50 - 54	330313	252020	.66571
55 - 59	144706	83798	.61922
60 - 64	201702	107787	.55424
65 - 69	72281	23443	.46682
70 - 74	100850	34834	.38753

With a note on the accuracy of the 1937 census by

- (a) calculating the 1947/1937 ratio of cohort size, and
 (b) calculating the 1937 population from the 1947 population, using the survival ratios. (25)

7. The following table gives the female populations of England and Wales in 1954 in 5 yearly age-groups, together with the number of live births and survival factors.

Age	Female population (000)	Total live births	Survival factor
15 - 19	1399	30317	.9694
20 - 24	1422	188512	.9668
25 - 29	1521	205216	.9632
30 - 34	1756	147103	.9584
35 - 39	1451	61804	.9519
40 - 44	1689	19640	.9424
45 - 49	1667	1367	.9279

Proportion of male to female live births = 1 : 1.05

- (a) Calculate the fertility rate for each age-group.
 (b) Assuming that the fertility rate is the same at each year of age within an age-group, calculate the expected number of births to a woman passing through the age-group 15 - 49.
 (c) Calculate the total, gross and net reproduction rates. (25)

Please turn over

8. Obtain the equation of a suitable logistic curve to fit the population of the U.S.A. as given in the following table :

Year	Population
1800	5308
1810	7240
1820	9638
1830	12066
1840	17069
1850	23192
1860	31443
1870	39818
1880	50156
1890	62948
1900	75995
1910	91972
1920	105711
1930	122775
1940	131669

Also estimate the maximum value which this population is expected to attain.

(2)

GROUP E - EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Answer all questions from this group)

9. Twelve subjects were independently ranked by three judges in respect of the trait : ability to appreciate music. The results are given below :

Judges	Subjects											
	1	2	3	4	5	6	7	8	9	10	11	12
1	5	1	4	2	7	10	12	11	9	8	3	6
2	7	3	6	1	5	12	8	9	10	11	2	4
3	4	1	6	2	5	10	12	8	11	9	3	7

Examine if the rankings by the three judges can be regarded as concordant. If so, assign final ranks to the subjects by combining the three rankings in a suitable manner.

10. A test consisting of 5 items were applied on a large number of subjects and their scores (Y) on a reliable criterion were also recorded. The following table gives the values of

P_i = proportion of subjects answering the i^{th} item correctly.

P_{ij} = proportion of subjects answering both the i^{th} and the j^{th} items correctly.

Please turn over

μ_i = average criterion score for those subjects who answered the i^{th} item correctly.

The average criterion score is $\mu = 16.8$ and the standard deviation of the criterion score is $\sigma = 3.1$

Let X denote the total number of items correctly answered by a subject. Compute the mean and the variance of X and the covariance of X and Y . Hence obtain the correlation coefficient between X and Y .

i \ j	values of P_{ij}					P_i	μ_i
	1	2	3	4	5		
1	-	-	-	-	-	0.75	12.3
2	0.67	-	-	-	-	0.70	14.9
3	0.55	0.52	-	-	-	0.63	18.2
4	0.41	0.36	0.28	-	-	0.47	19.0
5	0.52	0.48	0.40	0.20	-	0.58	16.9

(20)

11. Five thousand candidates appeared in a certain examination in which there were three tests T_1 , T_2 and T_3 , the maximum marks in each test being 100. The means, standard deviations and the correlation coefficients of the scores obtained in those tests are given below. Assuming that the scores in the three tests are jointly normally distributed, find the number of candidates securing at least 150 in the aggregate.

test	mean	standard deviation	inter correlations		
			T_1	T_2	T_3
T_1	39.46	6.21	-	0.29	0.38
T_2	52.31	9.40		-	0.43
T_3	45.26	8.60			-

(15)

ooooo

Please turn over

INDIAN STATISTICAL INSTITUTE
 Statistician's Diploma Examination - November 1963
 Paper VIII: Subjects of Specialisation - I

Time: 4 hours.

Full marks 100.

- 1) Answer questions only from the section below on the subject you have opted for.
 ii) Figures in the margin indicate full marks for each question.

(a) ECONOMIC STATISTICS - Econometrics

(Answer any five questions)

- 1.a) If $\pi = \sum_{i=1}^{m+n} p_i q_i$, the profit is to be maximized subject to the production-function

$$F(q_1, \dots, q_{m+n}) = 0$$

when the markets for the inputs and outputs are perfectly competitive, show (i) the rate of product transformation for every pair of outputs is equal to the ratio of their prices and, (ii) the rate of technical substitution for every pair of inputs is equal to the ratio of their prices provided the second order conditions are satisfied. Here q_1, \dots, q_m are outputs and are ≥ 0 and $x_r = -q_{m+r}$ $r=1, \dots, n$ are inputs and are ≥ 0 .

Write down the second order conditions for maximization of π subject to the production-function.

- b) When $m=1$ and the production-function is $q_1 = a \frac{\sum_{r=1}^n b_r x_r}{\sum_{r=1}^n x_r}$ $b_r \geq 0$, show that when $\sum b_r < 1$ the second order conditions are satisfied. [20]
2. If the long-run cost-functions of the i th firm producing q_i is given by

$$C_i = \sum_{r=1}^n a_{ir} q_i^2 \quad \begin{matrix} a_{ir} < 0 & i \neq r \\ a_{ir} > 0 & i = r \end{matrix}$$

$$i = 1, \dots, n$$

then show that the effect of the external economies indicated by $a_{ir} < 0$ $i \neq r$ is only to reduce the aggregate cost, the aggregate supply-function remaining unchanged. Construct an example when external economies lead to an aggregate supply function with a negative slope. [20]

3. Write a detailed note on the technique of input-output analysis and its applications in economics. [20]
4. Explain with illustrations how econometric analysis is useful for economic policy-making. State also the limitations of econometric analysis. [20]
5. Define and describe the different types of structural relationships in the context of economic analysis and define economic models in terms of these structural relationships. Formulate as an illustration a stochastic economic model with more than one structural relationship, describing in detail the types of variables, relationships and errors involved. Explain very briefly with reference to this model why the single equation method of estimation may fail when there are more than one structural relationships in the model. [20]

Please Turn Over

6. Explain the multi-collinearity problem and the lunch-map analysis developed by Frisch to tackle it. Are there any other methods of tackling the problem? [20]

7. What are k-class estimators in the context of estimation of simultaneous linear relationships? How are they related to the two-stage least squares estimator and the estimators obtained by the limited information method?

Explain also the two-stage least-squares method and the three-stage least-squares method of estimation of simultaneous relationships. [20]

8. Define a linear autoregressive-process of the k^{th} order. If a first order linear autoregressive scheme is given by

$$x_t - \rho x_{t-1} = \epsilon_t, \quad t \text{ taking only integral values,}$$

$$\text{where } E(x_t) = E(x_{t-1}) = 0, \text{ and } |\rho| < 1;$$

obtain an estimate of ρ and work out its asymptotic variance.

Discuss the relevance of linear autoregressive schemes to economic analysis. [20]

9. It is observed that the estimates of the own-price elasticities of demand come out to be unduly very large (and negative) when the algorithm used for estimation is

$$X = a Y^b P_1^c P_2^d$$

where X is the percapita demand for cloth adjusted for price in an income-group in an U.S.S. round, Y the corresponding percapita income, P_1 is the price-index of cloth and P_2 , the price-index of food and the data cover the different income-groups (cross-sectionally) and the first nine rounds of the U.S.S. survey (temporally). What explanation would you give for this phenomenon and how would you set about further analysis, taking into account the nature of U.S.S. data, the features of estimation from cross-section and time-series data and the appropriateness or otherwise of the form of the demand relationship? [20]

10. Interpret the production-function relationship

$$Y_t = a b^t c^0 L_t^d$$

where Y_t stands for output at time t , C_t and L_t for the corresponding capital and labour at time t ; a , b , c and d are constants. Explain in detail how you would estimate the parameters a , b , c and d using only time-series data or both time-series and cross-section data and give the formulae of the estimators.

Does the above production-function relationship take a realistic account of technical progress? If it does not, then indicate a more realistic form of the production-function relationship to take account of technical progress. [20]

11. Why is a log-normal distribution usually found to fit the distribution of income better than a normal distribution? Can you suggest a socio-economic explanation. Do you think a normal distribution may give a better fit to the distribution of income in a group which is comparatively homogeneous with respect to occupation or some other relevant characteristic than to the distribution of income in the total population? Derive the estimates of the parameters of the log-normal distribution. [20]

000000

(b) TECHNIC-COMMERCIAL STATISTICS - Statistical Quality Control.

(Answer any four questions)

1. It is known that the life of electric lamps is distributed exponentially with the density function

$$f(x) = \frac{1}{\theta} e^{-x/\theta} \quad \theta > 0; 0 \leq x < \infty$$

n bulbs were sampled and tested for life and the test terminated... just after m bulbs had burnt out. The lengths of the lives of these m bulbs were recorded. Obtain by the method of maximum likelihood or otherwise an expression for estimating the average life and the variance of the estimate. [20]

2. Under what circumstances would you advocate the use of an AOQL sampling plan?

Comment on the statement 'the AOQL value will never be exceeded in practice'.

Indicate the derivation of the AOQL function for a double sampling attribute plan. [20]

3. What is the difference between direct measurement and indirect measurement of a characteristic? Give examples.

Briefly explain the problems faced in the use of an indirect measurement to control the characteristic and the statistical tools that may be gainfully employed in these situations. [20]

4. Write short notes on:

- Use of variance components analysis in quality control.
- Experiments to determine optimum conditions.
- Unknown sigma variables plan for acceptance. [20]

- 5.a) Comment upon the salient points of the Dodge Romig and Mil-Std. attribute sampling plans. [15]

- b) Discuss the role of control charts in acceptance inspection. [10]

000000

Plan a: Turn Over

(d) DESIGN AND ANALYSIS OF EXPERIMENTS - Statistical Aspects
(Answer any five questions)

- Describe the various techniques adopted to increase the efficiency of an experiment giving suitable examples, in the choice of treatments, experimental material, allocation to different treatments, concomitant variables etc. [20]
- With reference to a randomised block design, state carefully the hypothesis that is tested by the criterion of 'ratio of variance due to treatments to variance due to error'. Give a justification of the criterion employed taking into account only the randomisation aspect. [20]
- What are the situations in which a split plot design is used? Suggest a split plot design for an experiment involving 5 irrigation practices, 4 levels of manure and 5 varieties, the main aim of the investigation being to suggest suitable manures for given combinations of irrigation practice and variety. Draw up the analysis of variance for analysing the experimental results indicating the sources of variation and the degrees of freedom. [20]
- What are the situations suitable for elimination of effects due to concomitant variables? Discuss the relative advantages and disadvantages of the following types of experiments to determine whether diet A is better than diet B for fattening pigs. Indicate the statistical analysis in each case.

Experiment 1 : Twenty young ones are divided at random into two groups of ten each. Diet A is given to one group and diet B to another. The weight of each animal is recorded at the end of the experiment.

Experiment 2 : The groups are made as in experiment 1, but for each animal both the initial and final weights are available.

Experiment 3 : To each young one chosen for diet A, another young one with approximately the same initial weight is chosen for diet B. Only the final weights are recorded. [20]
- Define a linked block design and indicate its uses. Obtain intra block estimates of variational differences and their standard errors. [20]
- Describe how factorial experiments are useful in exploring a response surface to find the optimum combination of factors. What is a rotatable design of the first order and what is the advantage in using such a design. [20]

000000

(e) SAMPLE SURVEYS - Theoretical Aspects
(attempt any four questions)

- Define 'sampling with probability proportional to size'.
 - Derive an expression for the relative efficiency of sampling with probability proportional to size as compared with stratification by size and the use of a variable sampling fraction.
 - Let a random sample be taken with probability proportional to size and the sample (after selection) be stratified into size-groups. Consider the analysis of variance between and within size-groups of the values of the ratio (observation/size) obtained from the sample. Show that under certain conditions (to be stated by you) the relative efficiency required in (b) can be estimated by the ratio of within size-groups to total mean square. [5]

- 2.(a) Define 'cluster sampling' and discuss its advantages.
- (b) What are the considerations that lead to the optimum choice of cluster size?
- (c) Stating clearly and discussing critically the assumptions that are involved, derive a suitable formula for the determination of optimum cluster size.
- (d) In a certain survey, for estimating the area under wheat, clusters were formed by grouping 4 consecutive survey numbers. Use the following data to estimate the relative efficiency of cluster sampling.
- Between clusters mean square estimated with 9 d.f. = A (acres)²
Within clusters mean square estimated with 30 d.f. = B (acres)² .
- 3.(a) What is systematic sampling?
- (b) Distinguish between circular and ordinary systematic sampling.
- (c) What are the advantages and disadvantages of systematic sampling?
- (d) Compare systematic, simple random (without replacement) and stratified sampling for a population with a linear trend.
- (e) Show that systematic sampling can be regarded as equivalent to simple random sampling with replacement when the order in which units are arranged can be taken to be a random one.
- 4.(a) For two stage sampling (simple random at each stage) suggest an estimate of the population mean and derive an expression for its variance. Assume that each first stage unit contains the same number of second stage units.
- (b) For the analysis of variance of the observed second stage values, compute the expectations of mean squares - both between and within first stage units.
- (c) Hence or otherwise, show how the analysis of variance table could be used to study consequences of changing the sampling fraction at each stage.
5. Write short notes on any three of the following:
- (a) The Yates and Grundy estimate of the variance of Horvitz Thompson estimator.
- (b) Line Sampling
- (c) Non Sampling Errors
- (d) Lahiri's method of selection with probability proportional to size.
- (e) Ratio versus Regression estimates.

oooOooo

Please turn over

Page 01

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination, November 1963

Paper IX : Subjects of Specialisation - II

Time : 4 hours

Full marks : 100

- 1) Answer questions only from the section below on the subject of specialisation you have opted for.
- ii) Figures in the margin indicate full marks for each question.

(a) ECONOMIC STATISTICS - Indian Economics

(Answer any three questions from each group.
Neatness carries 4 marks)

GROUP A

1. What are the reasons for low productivity of agriculture in India? Examine critically the "package programme" aiming at raising agricultural production to the target set in the third five year plan. (16)
2. Examine the pattern of industrial location in India. Is the present policy of dispersal of industries for removing disparity between the states, economically sound? (16)
3. Examine critically the importance of a national minimum wage in Indian industries. What factors would you take into consideration to fix it? (16)
4. Give an account of the organisation of the banking system in India. Do you think that the present structure is suitable for the current financial problems? (16)
5. Analyse the effects of foreign aid on India's future balance of payments. What suggestions would you offer for reorganising the foreign trade of India to meet the situation? (16)
6. Write notes on any two of the following : (16)
 - i) abolition of expenditure tax in the current budget of India
 - ii) holding the price line in India
 - iii) progressive change in income distribution in India
 - iv) experience of co-operative organisations in India.

GROUP B

7. Distinguish between average and marginal capital-output ratios. How are these ratios useful in planning? Indicate the limitations of these ratios with regard to statistical measurement and use in planning. (16)

Please turn over

8. Give an account of Dornar's model of economic growth and deduce the growth curve. Is the model applicable to situations in under-developed countries? (16)
9. What is the matrix of technical coefficients? What method will you adopt to use the matrix of technical coefficients for deriving future production targets in a changing economy? (16)
10. What is linear programming? How will you use this technique to check whether the allocation of resources in the third five-year plan of India is optimum? (16)
11. Examine the investment patterns in the three five-year plans of India in the light of the difficulties faced. What type of investment allocation the planners should, in your opinion, have made in the third plan so as not to be postured by bottlenecks? (16)
12. Write critical notes on any two of the following: (16)
- statistical information available for planning in India.
 - organisation and function of the Planning Commission of India.
 - perspective planning in India.
 - investment in human capital in Indian planning.

ooo0ooo

(b) TECHNO-COMMERCIAL STATISTICS - Operations Research (Or Elementary of Bookkeeping & Accountancy) & Statistical Methods in Business.

(i) Operations Research

(Answer any four questions. Neatness carries 2 marks).

(Use separate answer book for these questions)

- Discuss the role and use of quantitative models in operational research. (17)
- What are the problems involved in inventory management? What are the usual models of inventory management and what are the data needed for these models? (17)
- Work out the expression for the discounted value K_n of all future costs associated with a policy of replacing equipment after every n periods, when the initial cost of new equipment is A , the costs incurred in the periods 1, 2, 3, of equal length are C_1, C_2, C_3, \dots respectively and the cost of money is $100 r\%$ per period. Thus or otherwise deduce the following rules for determining n corresponding to the minimum of K_n :
 - if the cost for the next period is less than the weighted average of previous costs, do not replace, and

Please turn over

- ii) replace if the cost for the next period is greater than the weighted average of previous costs, the weights being given by the discount factors applied to the costs in each period.

Discuss how the cost equation and the replacement rules can be used in actual practice. (17)

4. For the case of one servicing station, Poisson arrivals with mean arrival rate λ and exponential service time with mean service rate μ , where n is the number of units in the system, show that P_n , the (limiting) probability that there are n units in the system (at time t or any other time) satisfy the following equations

$$P_1 = \frac{\lambda}{\mu} P_0$$

$$\text{and } P_n = \frac{\lambda + (n-1)\mu}{n\mu} P_{n-1} = \frac{\lambda}{n\mu} P_{n-2}, \quad n > 1.$$

Obtain the solution of the equations.

Obtain also the distribution of the waiting time and the distribution of the total time an arrival spends in the system.

Thus or otherwise obtain average queue-length, average length of non-empty queues, average waiting time of an arrival and average time an arrival spends in the system. (17)

5. What is the dual of a linear programming problem? What are the complementary relationships between a linear programming problem and its dual? What is the interpretation and use of the dual programming problem? Explain with the help of illustrations. (17)
6. Explain in brief the dynamic programming approach and pose the following as a dynamic programming problem and solve.

A dealer places an order with his wholesaler on the first of each month and obtains delivery one month later. The cost of holding inventory is C_1 per unit per month and the cost of shortage is C_2 unit per month, shortages being carried over from one month to the next. If the monthly demand X is a random variable with density-function $p(x)$, find the policy that minimises the long-term average costs per month. (17)

(ii) Elements of Bookkeeping & Accountancy

(Use separate answer book for these questions)

7. You are given a Trial Balance as at 31 October 1963 drawn by the Accountant of X. You are required to prepare the Trading and Profit and Loss account for the year ended 31 October 1963 and the Balance Sheet as at that date, after making necessary adjustments for providing for a few reserves and for rectifying a few errors found in the Trial Balance as given below.

(a) Debit of Rs. 500 in the Drawing account should be charged to furniture and fittings.

(b) Materials purchased for the erection of building of the value of Rs. 1000/- have been charged to purchase account.

(c) Rs. 500/- paid in cash by X was credited to Sales account.

(d) Depreciation has to be provided on Building @ 2½% on Furniture and Fittings @ 6%.

(e) Bad debts of Rs. 100/- should be written off.

(f) Bad debt reserve at 5% on Sundry Debtors and 2½% discount on Sundry Creditors are to be provided.

(g) Adjustments have to be made for salaries unpaid Rs. 200/-; Rates prepaid Rs. 60/-; Insurance prepaid Rs. 100/-; carriage inwards unpaid Rs. 100/-.

Trial Balance as at 31.10.1963		
	Rs.	Rs.
Capital A/C		50,000
Drawings account	1,500	
Sale		7,40,000
Purchases	6,99,200	
Salaries	1,200	
Carriage Inwards	400	
Carriage Outwards	500	
Lighting	300	
Rates & Insurance	400	
Discount		500
Furniture	6,000	
Sundry Debtors	8,000	
Sundry Creditors		20,000
Petty cash	50	
Cash at Bank	1,700	
Stock at 31.10.1963	61,250	
Building	30,000	
	<u>8,10,500</u>	<u>8,10,500</u>

Please turn over.

8. EITHER

- (a) The Managing Director of a company is to receive a commission of ten percent on the net profits after the payment of such commission. Find the amount of his commission if the net profit before charging his commission in a year amounts to Rs. 24,200/-.
- (b) The rate of marine insurance premium is 7% on the declared value of the cargo, plus 2% commission and 1% expenses. A person consigns goods valuing Rs. 9,225/-. For how much should it be insured so that in case of loss the full cost of the cargo together with all insurance expenses may be recovered?
- (c) Discount on a Bill of Rs. 528.75 NP payable at the end of five years is Rs. 78.75 NP. What is the rate of interest? (5X3=15)

OR

In taking out the Trial Balance of a trading concern, you find that it does not agree and you post the difference to suspense account in the first place. On closer scrutiny, the errors making for the difference in the Trial Balance are detected. These are given below. You are required to rectify the errors by means of journal entries and to post the suspense account.

- (a) one item of sales return for Rs. 500/- was posted to the credit of 'Sales A/C'.
- (b) the purchase day-book was undercast by Rs. 300/-
- (c) one item of sale for Rs. 5.08 to 'A' was entered correctly in the Sales Day-book but was posted to the Credit of 'A' in the Sales Ledger as Rs. 8.05. (15)

9. EITHER

Briefly note the differences between a 'bearer' cheque; an 'order' cheque; a crossed cheque; a 'specially crossed' cheque; a crossed cheque marked 'not negotiable'; and a crossed cheque marked 'account payee only'.

- (a) which of these forms, in your view, affords the greatest protection against fraud?
- (b) can a cheque crossed 'not negotiable' be negotiated by endorsement?
- (c) What other protections against fraud or forgery are usually resorted to by big establishments in issuing cheques. (20)

OR

Give some of the important causes which may account for an abnormal increase or decrease in the ratio of gross profit to the turn over. (20)

(iii) Statistical Methods in Business

(Answer any two questions. Use separate answer book for these questions)

10. Put yourself in the position of a market research technician attached to a firm manufacturing washing soap and detergents, who has to conduct a sample survey in a city (take any Indian city for illustrative purposes) to study the habits and attitudes of consumers with respect to washing soap and detergents. Describe in detail how you would proceed with the survey. (15)

Please turn over

11. Write a detailed note on the statistical methods of business forecasting. (15)
12. "The purpose of job-evaluation is to make an analytical study of the contents and functions of all jobs in order to develop a description of their characteristics, to appraise these characteristics according to their relative importance and value, and to establish a sound basis for setting equitable rates and rate changes. Describe in detail the procedures that are followed in job-evaluation for the above purpose. (15)

(d) DESIGN & ANALYSIS OF EXPERIMENTS - Combinatorial Aspects

(Answer any four questions)

1. Define a finite field of order p^n where p is a prime. Construct $GF(2^3)$ and utilise it to construct a complete set of mutually orthogonal Latin Squares of order 8. (6+10)
2. Define a balanced incomplete block design BIBD and state and prove any theorem on difference sets to construct a BIBD. (15)
- Show that if v is a prime power of the form $4t+3$ where t is a positive integer, one can construct a BIBD with parameters
- $$v = b = 4t+3, \quad r = k = 2t+1, \quad \lambda = t$$
- with the help of a difference set. (12)
3. Define a partially balanced incomplete block design (PBIBD) with two associate classes and obtain the relations among its parameters. (10)
- Construct a PBIBD for 36 treatments with 3 replications, two associate classes and block size 6 and obtain its remaining parameters. (9)
- Indicate how, from this design, we can construct a BIBD with parameters
- $$v = b = 36, \quad r = k = 15, \quad \lambda = 6. \quad (7)$$
4. Define an orthogonal array of strength t and index λ in S distinct symbols. (4)
- Show how you can construct such an array of strength 2 and index 1 for 20 symbols with 5 constraints. (2)
- Give a construction for the array $[S^t, S+1, S, t]$ where $S = p^n$, p a prime and $S > t$. (12)
5. Show that the maximum number of points in $PG(2, S)$, where S is a prime power, such that no three points are collinear can not exceed $S+2$. Prove that this number is attained when $S = 2^n$. What is the importance of the above result in the theory of confounding? (11+3)

Please turn over

6. Explain the principle of generalised interaction in symmetrical factorial designs of the type S^m where S is a prime power and explain how it leads to a general method of construction for confounded designs of the type (S^m, S^k) . (11 + 14)

oooOooo

(e) SAMPLE SURVEYS - Organizational Aspects

(Answer any four questions)

1. Write a critical note on the various types of non-sampling errors that arise in large-scale sample surveys in India and the steps to be taken to minimize them. Discuss how scrutiny of data should be arranged to ensure accuracy and reliability of the data collected. (25)
 2. Write a detailed note on processing and tabulation of data of large scale sample surveys, discussing both manual and mechanical tabulation. Illustrate by an example how the machine operation in the tabulation section should be programmed so as to minimize the use of available resources. (25)
 3. You are required to estimate the cost of production of milk in your State. What sampling unit would you use and what type of frame would you choose? What items of information would you like to include and what ultimate results would you like to present? (25)
 4. 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? Describe the organization of the field work and draw up a set of instructions to field workers and supervisors. (25)
 5. (a) Discuss possible defects in a sampling frame. How will you scrutinize an actual frame to detect and assess the magnitude of such errors? (8)
(b) Suggest a possible frame for conducting a working class family-budget enquiry in a big city for determining weights for cost of living indices. How will you make a choice among the alternatives available? (9)
(c) A frame contains some overlapping area sampling units. The object of the survey is to estimate population total of a particular characteristic. It is suggested that the defect of the frame can be rectified by either of the following two methods :-
 - 1) Boundaries of all the overlapping sampling units may be arbitrarily defined before drawing the sample.
 - 11) As the boundaries of the sampled units only are necessary for purposes of estimation, it is sufficient if boundaries of these units only are arbitrarily defined.
- Give your critical comments on these two methods. (8)

Please turn over

6. Write short notes on any three of the following :-

- a) Centralized and decentralized training. (2 $\frac{1}{2}$)
- b) Destruction of records (2 $\frac{1}{2}$)
- c) The flow chart in data processing. (2 $\frac{1}{2}$)
- d) Interpenetrating sub-samples in scrutiny and tabulation. (2 $\frac{1}{2}$)
- e) Punching and verification. (2 $\frac{1}{2}$)

oooOooo

Please turn over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1963

Paper X : Subjects of Specialisation - III

Time : 5 hours.

Full marks:100

- (i) Answer questions only from the section below on the subject you have opted for.
- (ii) Figures in the margin indicate full marks for each question.
- (iii) Use of Calculating machine is permitted.
- (a) ECONOMIC STATISTICS - Practical
(Answer any four questions)

1. The following table gives the agricultural income-series for India for the years 1921-40. Fit a logistic curve to the series.

Year	Agr. income (Rs. Abn)	Year	Agr. income (Rs. Abn)
1921	38.67	1931	45.98
1922	45.37	1932	45.40
1923	46.90	1933	45.23
1924	39.70	1934	45.64
1925	42.36	1935	44.48
1926	42.88	1936	43.67
1927	43.62	1937	43.87
1928	43.39	1938	46.73
1929	44.50	1939	42.64
1930	45.34	1940	44.93

[Note: When the logistic curve is given by

$$X_t = \frac{K}{1 + be^{-at}}$$

$$\frac{1}{X_t} = \frac{1 - e^{-a}}{K} + e^{-a} \cdot \frac{1}{X_{t-1}}$$

and b may be estimated by the formula of Rhodes

$$\log_e b = a \frac{N+1}{2} + \frac{1}{a} \sum_{t=1}^N \log_e \left[\frac{K}{X_t} - 1 \right]$$

where N is the total number of observations.]

Work out the proportion of variance explained both in the reciprocal series and the original series. Interpret the results and comment. Examine if a linear trend fits the data. [25]

2. The following table gives international data on labour inputs and wages for two industries. Fit the relationship

$$\frac{V}{L} = a W^b$$

to the data for the two industries. Here V is value added in thousands of U.S.dollars, L is labour input in man-years and W is wage rate (total labour cost divided by L).

Obtain the estimates of b for the two industries and test if each is different from 1. Interpret the results and comment.

(Note on the data:- The data on wage-payments for different countries include varying proportions of non-wage benefits. The data on employment are not corrected for intercountry differences in the number of

hours worked per year or the age and sex composition of the labour force). Also obtain the estimates of b when W is treated as a dependent variable and V/L as the independent variable. Compare the two sets of estimates and comment.

No.	Country (year of data)	Percent industry		Pricer/iron and steel	
		V/L	βW	V/L	βW
1	U.S.A. (1954)	0.6752	4216	0.1266	4387
2	Canada (1954)	0.6635	3910	0.1416	3769
3	New Zealand (1955/56)	0.6686	2503	0.2340	2190
4	Australia (1955/56)	0.2562	2071	0.2330	2306
5	Denmark (1954)	0.2634	1730	0.2173	1655
6	Norway (1954)	0.2831	1569	0.2913	1560
7	United Kingdom (1951)	0.2394	1431	0.4503	1224
8	Columbia (1953)	0.2040	1227	0.4199	1168
9	Mexico (1951)	0.2709	630	0.4728	594
10	Argentina (1950)	0.5443	501	0.7920	591
11	Iraq (1954)			1.2066	210
12	Japan (1953)	0.1693	1025	0.7526	664
13	India (1953)	0.8090	319	0.9060	450

3. The following two relationships give the supply and the demand relationships. Verify that both are just-identified and estimate first the coefficients in the reduced form of equations and then estimate the structural coefficients, from the table of sums of squares and products of deviations of the variables from their means, given below. Work out also the proportions of variance explained for the two reduced form equations.

The supply and demand relationships are

$$Y_1 = \gamma y_2 + \beta_1 X_1$$

$$Y_1 = \gamma' y_2 + \beta_3 X_3$$

respectively,

where Y_1 = per capita consumption (or sales) of meat (pounds per year).

y_2 = retail price of meat (index, 1935-39 = 100)

X_1 = cost of processing meat (index, 1935-39 = 100)

X_3 = per capita real disposable income (dollars per year)

The variables are measured from their means. The sample consists of annual data, 1919-1941.

Table of sums of squares and products of deviations from the means.

	Y_1	Y_2	X_1	X_3	Arithmetic mean
Y_1	1369.54	-352.55	-536.48	3671.91	166.19
Y_2		1581.49	850.33	8354.59	92.34
X_1			2534.00	3611.72	80.42
X_3				8343.65	495.57

4. The following table gives the distribution of income (U.S.G. data) in the 4 rural groups in India. Test if the distribution pattern changes from group to group. Compare the means and coefficients of variation and compare these results with the test results. Comment.

per capita monthly expenditure class	Rural areas - number of persons			
	percentage in rural areas			
	1	2	3	4
0 - 8	12.57	12.10	22.35	8.34
8 - 11	17.93	13.66	25.55	15.60
11 - 13	12.81	10.62	14.80	17.90
13 - 15	11.22	14.45	7.34	5.96
15 - 18	14.24	11.11	7.60	12.21
18 - 21	10.40	12.61	5.15	11.04
21 - 24	7.65	8.03	3.23	6.69
24 - 28	4.60	4.52	5.19	7.78
28 - 34	3.02	6.17	4.22	5.12
34 - 43	3.56	3.59	2.12	2.03
43 - 55	.87	2.71	1.31	4.00
55 -	1.65	.83	1.14	1.65
Total	100.00	100.00	100.00	100.00
Total number of persons	1221	1012	1152	987

(Note: Rural areas 1, 2, 3, 4 are respectively, North India, East India, South India and West India.)

[25]

5. Solve the following linear programming problem either by working out all the corner solutions and then spotting the optimum one or by any other method that is convenient for you. Write down the dual programming problem and, using the solution of the original linear programming problem, solve the dual programming problem. Comment on the results.
- Three processes are available for the production of corn. The input coefficients for unit level of production of corn (100 tons) are given below.

Inputs	Units	input coefficients		
		Process I	Process II	Process III
Labour	man-months	25	5	4
Land	acres	50	100	125
Tractors	tractor-months	20	3½	0

The restrictions on the resources are that there are 10 man-months of labour available and 110 acres of land available.

The price of corn is given at ₹ 10 per ton. Maximize the receipts. Comment on the results.

[25]

6. Below is given the transaction-matrix for the United States for the year 1947.

- Work out (1) total gross outputs;
 (2) gross national product at factor prices;
 (3) total resources available at market prices;
 (4) input coefficients for the different industries;
 (5) indirect taxes net subsidies
 total gross-output for different industries;
 (6) personal consumption
 total gross-output for different items;
 (7)

- (8) Exports for different industries
Imports
- (9) Exports for different industries
total gross outputs

and carry out any other analysis that is relevant and comment on the results.

[25]

Industry	Output by Industries										Total Sectoral Output	Exports Capital formation	Total Sectoral Investment			
	1	2	3	4	5	6	7	8	9	10						
1. Agriculture and food	-	•	3	19	27	1	•	•	•	49	4	103	36	-13	13	314
2. Coal and power	3	-	3	8	2	2	3	2	1	49	•	72	4	•	2	1
3. Building, building materials	9	3	-	4	1	5	4	10	3	67	14	121	5	159	56	24
4. Chemicals and rubber	30	5	17	-	12	6	11	12	2	28	20	142	17	3	4	91
5. Textiles and clothing	3	•	4	5	-	2	•	4	1	2	11	31	13	1	3	135
6. Paper, printing and steel	6	•	4	5	4	-	•	3	3	41	26	92	4	4	3	38
7. Metal working	•	1	22	2	•	2	-	47	18	2	14	100	8	1	•	•
8. Dyeing	2	1	15	•	1	1	1	-	5	19	44	05	34	127	17	51
9. Metal goods	6	•	10	2	1	1	•	20	-	4	15	60	6	0	2	18
10. Services	80	4	67	23	14	12	13	16	5	-	70	364	35	40	63	915
Total allocated	22	4	22	36	21	21	15	39	14	82	-	269	•	-8	•	•
Total	160	17	150	90	82	52	48	154	51	343	219	1393	136	29	162	1569
Imports	28	•	3	9	5	8	6	1	1	5	1	66	-	•	13	113
** Govt.	21	11	15	17	12	10	8	18	5	139	22	278	0	0	3	315
*** Private	258	51	109	94	84	72	55	143	38	871	18	1372	8	2	301	21
Total Gross Imports																
Total Gross Output																

Low than 100 million
as indirect taxes net subsidies
present to industry factors.

(Total gross outputs and total gross inputs may not tally because of rounding off errors)

(b) INDUSTRIAL-COMMERCIAL STATISTICS - PracticalGROUP A - Statistical Quality Control

(Answer any two questions from this group.
Use separate answer book for this group)

1. To a process which has been giving 30 percent defectives, some changes have been made to improve the process. It is desired to know whether the changes have been beneficial.

The risk of assuming an improvement when the process has not changed, and of not detecting a change in the incidence of defectives when it is 25 percent, are both sought to be kept at 0.05.

Workout a suitable sequential scheme. Prepare a graphical layout to operate the scheme.

Draw the OC curve of the scheme. [25]

2. The following data show the results of an experiment conducted in an attempt to validate a production standard. Three machines producing the same item, three operators and four time study men are involved. Data are given in hundredths of a minute over two minutes.

Machines	Operator	Time study men			
		1	2	3	4
1	A	56	27	28	14
	B	84	25	6	14
	C	73	20	6	20
2	A	34	22	26	11
	B	31	25	53	8
	C	11	17	45	17
3	A	78	25	62	14
	B	22	115	14	11
	C	64	56	22	48

Determine which sources of variation are significant and estimate their magnitudes.

Based on the outcome of your analysis, can you suggest any other experiment which might be conducted? [25]

3. The specifications for the thickness of the coating of an enamelware are, in certain arbitrary units, 50 ± 10 .
Based on the specifications, set up a trial control chart for averages and ranges.

The first fifteen samples collected every fifteen minutes, gave the following averages and ranges.

Sample No.	\bar{X}	R	Sample No.	\bar{X}	R
1	52	2	9	48	1
2	52	4	10	49	3
3	54	6	11	50	2
4	55	5	12	51	3
5	56	3	13	52	1
6	57	2	14	52	4
** 7	59	3	15	52	1
8	48	1			

** Process adjusted.

Plot the data on your control chart. Give your recommendations for the future control of the process making whatever further analysis you consider necessary for this purpose. [25]

Please Turn Over

GROUP B - Operations Research

(use separate answer-script for this group.)

(Answer any two questions.)

4. Using simplex or any other method, maximize

$$Z = 5X - 2Y + 3Z$$

subject to

$$-(2X + 2Y - Z) \leq -2$$

$$3X - 4Y \leq 3$$

$$Y + 3Z \leq 5$$

and $X \geq 0$, $Y \geq 0$ and $Z \geq 0$.

5. The following table gives the length of time during which a number
- n
- of customers were observed to be present in the department during a total period of four hours. Test if the distribution is geometrical. Work out the mean and variance of the variable
- n
- . Comment.

n	Time in minutes	n	Time in minutes
0	16.8	7	9.6
1	35.5	8	5.8
2	52.0	> 8	9.6
3	49.0	Total	240.0
4	29.3		or 4 hours.
5	18.8		
6	13.6		

6. The following table gives information regarding the quantity of a certain goods demanded from a stock-keeping unit of a department store. Examine whether a Poisson distribution can be used to generate demand for the stock-keeping unit. Comment.

Units demanded per week	No. of weeks
0	5
1	3
2	4
3	4
4	2
5	2
6	1
7	0
8	2
9	1
10	1
Total	25

GROUP C - Statistical Methods in Business

(Use separate answer-script for this group)

(Answer any two questions)

7. The following table gives the industrial average monthly wage-rates obtained through the wage survey and the point values obtained by a company, of ten key jobs. Fit a linear regression to the data to obtain the relationship between the industrial rate (as the dependent variable) and the point value (as the independent variable). Does the linear regression give a good fit? Interpret the results.

Please Turn Over

Job No.	Industrial average monthly rate.	Point value
1	100	82
2	105	115
3	130	120
4	148	135
5	155	164
6	160	100
7	195	225
8	255	240
9	270	275
10	280	310

[10]

8. The following table gives the percentage distribution of men and women using different brands of toilet soap. The total number of men and women interviewed are also given. Using the χ^2 -test or otherwise, determine if men and women differ in their preference for the different brands. Comment on the results.

Toilet soaps used by men and women		
Brand	percentage of women using the brand	percentage of men using the brand
A	20.3	15.5
B	16.6	15.1
C	15.9	18.8
D	10.5	7.9
E	6.0	15.4
Rest	31.4	27.3
Total	100.0	100.0
Total number	5000	5000

9. The following table gives percentage errors of forecaster of yearly company sales based on three alternative methods.

Examine which method of forecast is to be preferred most. Comment.

year	percentage errors of forecasts based on actual annual investment.	percentage errors based on the level of the previous years' sales.	percentage errors based on fourth quarterly investment of the previous year
1946	-2.2	-	-
1947	11.4	4.0	-
1948	6.1	19.9	7.3
1949	1.0	11.4	4.2
1950	0.2	10.4	9.3
1951	6.7	0.5	2.9
1952	0.0	6.7	1.9

000000

[11]

Please turn over

(d) DESIGN AND ANALYSIS OF EXPERIMENTS - Practical

(Answer any three questions)

1. Sixteen passengers on a ship discover that they are an exceptionally representative body. Four are Englishmen, four are Scots, four are Irish and four are Welsh. There are also four each of four different ages, 35, 45, 55 and 65 and no two of the same age are of the same nationality. By profession also four are lawyers, four soldiers, four doctors and four clergymen and no two of the same profession are of the same age or of the same nationality.

It appears also that four are bachelors, four married, four widowed and four divorced and that no two of the same marital status are of the same profession or of the same age or of the same nationality. Finally, four are conservatives, four liberals, four socialists and four fascists and no two of the same political sympathisers are of the same marital status, or the same profession, or the same age, or the same nationality.

Three of the fascists are known to be an unmarried English lawyer of 65, a married Scots soldier of 55 and a widowed Irish doctor of 45. Specify the remaining fascist.

It is further given that the Irish Socialist is 35, the conservative of 45 is a Scotsman, and the Englishman of 55 is a clergyman. What do you know of the Welsh lawyer? [5]

2. With the object of improving the quality of a basic dyestuff, three stages of the manufacturing process were examined and the factors studied were as follows:

Stage	Factor	Levels
1	A. Temperature	(0) Low (1) High
	B. Quality of material	(0) and (1)
2	C. Reduction pressure	(0) Atmospheric (1) Increased
	D. Pressure	(0) Low (1) High
3	E. Vacuum	(0) Low (1) High

One measure of the quality is the 'shade' as assessed by a recording photo-electric spectrometer, the lower the value recorded, the better being the quality.

From technical considerations, it is surmised that A and B interact, and so might C, D, and E, but it was less likely that there would be interactions between A or B on the one hand and C, D or E on the other.

The results of the experiment are given below. Analyse the data and give your comments.

Levels of factors					Quality characteristic	Levels of factors					Quality characteristic
A	B	C	D	E		A	B	C	D	E	
0	0	0	0	0	201.5	0	0	0	1	1	255.5
1	0	0	0	1	178.0	1	0	0	1	0	240.5
0	1	0	0	1	183.5	0	1	0	1	0	268.5
1	1	0	0	0	176.0	1	1	0	1	1	244.0
0	0	1	0	1	188.5	0	0	1	1	0	274.0
1	0	1	0	0	178.5	1	0	1	1	1	257.5
0	1	1	0	0	174.5	0	1	1	1	1	256.0
1	1	1	0	1	196.5	1	1	1	1	0	274.5

Please Turn Over

3. The layout for a varietal trial involving 10 varieties in 5 blocks each of 4 plots is obtained from the following arrangement.

(1)	(2)	(3)	(4)
(1)	(5)	(6)	(7)
(2)	(5)	(8)	(9)
(3)	(6)	(8)	(10)
(4)	(7)	(9)	(10)

By taking the numbers within brackets as the treatments and the rows as blocks.

Verify that it is a two associate partially balanced design and write down the values of the parameters of the design.

Write down computational instructions for complete analysis (including recovery of inter block information) of the results of an experiment with this layout.

Calculate also the efficiency factor of the design.

4. To compare three varieties of barley A, B, and C, experiments were conducted in three localities selected at random within a region, using in each locality a Latin Square layout. The table below gives the yields and the layout of the experiment.

Analyse the data and prescribe the best variety if any.

Yields of three varieties of barley in bushels/acre

Locality 1			Locality 2			Locality 3		
B	C	A	C	B	A	A	C	B
44.6	41.2	29.2	63.9	55.4	47.5	24.0	42.1	27.5
A	B	C	A	C	B	B	A	C
25.0	39.1	31.9	52.2	63.8	53.4	25.5	24.7	42.5
C	A	B	B	A	C	C	B	A
36.6	26.8	45.5	56.8	46.9	64.8	46.7	33.3	33.6

CCCGCC

(e) SAMPLE SURVEYS - Practical

(Answer all questions)

1. The following table shows the population size estimates (x) of different groups of the population of the U.S.A. in 1950, and square of the sample coefficient of variation (v_x^2) of these estimates:

Sl. no.	Population group	Estimate (x)	Squared Coeff. of variation (v_x^2)
1.	Females, white, did not work in 1950, married, 35-54 years old.	8756600	.000732
2.	Females, white, usually worked full-time, married, clerical and sales workers.	2815800	.002550
3.	Females, white, usually worked full-time, widowed, divorced or never married.	7328300	.001056
4.	Females, white, did not work in 1950, widowed, divorced, or separated, 14-44 years old.	803500	.012322

Sl. No.	Population group	Estimate (x)	Squared coeff. of variation ($\frac{s^2}{x}$)
5.	Females, did not work in 1950, never married.	406600	.015744
6.	Males, white, did not work in 1950, 20-64 years old.	1613100	.002623
7.	Males, non-white, did not work during the year.	570000	.013633
8.	Males, white, usually worked full-time, worked 6-26 weeks in 1950.	3100200	.001357
9.	Males, non-white, did not work in 1950, 14-19 years old.	201400	.023140
10.	Females	56631400	.000294

- a) Plot the values $\frac{s^2}{x}$ against x on a graph.
- b) Estimate the constants of a relationship of the form $\frac{s^2}{x} = a + b/x$.
- c) Write your comments regarding the usefulness of relationships of the above type in sample surveys, particularly a population survey. [2]

2. The following is a list of the 28 municipalities of a particular region, with the number of farms and the number of cattle in each municipality.

Serial number of municipality	Number of farms	Number of cattle (000)	Serial number of municipality	Number of farms	Number of cattle (000)
1	175	15.7	15	78	5.2
2	135	10.3	16	80	9.0
3	361	36.0	17	120	10.7
4	142	9.5	18	128	9.3
5	91	8.2	19	207	15.6
6	329	30.3	20	158	15.0
7	185	19.6	21	95	7.4
8	404	35.4	22	350	20.5
9	180	15.7	23	133	9.9
10	181	10.5	24	138	12.1
11	91	10.0	25	160	17.3
12	104	16.1	26	273	20.6
13	236	20.6	27	85	6.6
14	84	6.9	28	137	11.4

- a) Draw a sample of 5 municipalities with probabilities proportional to the number of farms, with replacement,
- by the cumulative method;
 - by Lahiri's method.
- In each case the steps in the selection process should be written down in detail.
- b) Obtain unbiased estimates of the average number of cattle per farm in the whole region from the two samples you have drawn.
- c) Calculate the variance of these estimates from the values given for the whole population.
- d) Suppose you are given only the sample drawn in [(a)(i)] above, estimate the sampling variance of the unbiased estimate of the population mean number of cattle per farm. [2]

3. The following table shows the number/seedlings in every individual foot of sown bed which is 80 feet long, shown in a rectangular table form for convenience.

No.	No.	No.	No.	No.	No.	No.	No.
1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80
26	16	27	37	4	36	20	21
28	9	20	14	5	25	21	26
11	22	25	14	11	43	15	16
16	26	39	24	9	27	14	18
7	17	24	16	25	20	13	11
22	39	25	17	16	21	9	19
44	21	18	14	13	18	25	27
26	14	44	30	22	19	17	29
31	40	55	36	10	24	7	31
26	30	39	29	9	30	30	29

- a) Find the variance of the mean of a systematic sample consisting of every 10th foot.
- b) Compare this variance with the variance for
- a simple random sample of size 8
 - a stratified sample of size 8 with 2 units per stratum.
4. From a region containing 1000 villages and of total cultivable area 965830 bighas, a simple random sample of 66 villages were drawn. The following table shows the total cultivable area and the area under wheat in these villages.

Area under wheat (bighas)	Total cultivable area (bighas)				
	200	700	1200	1700	2200
100	12	6			
300	2	18	4	2	1
500		4	7	3	1
700				2	1
900				1	2

- a) Obtain the ratio and regression estimates of the total area under wheat in the region.
- b) Obtain estimates of the variances of the estimates you have calculated, and compare them with the variance of an estimate if, from the sample, only information regarding the area under wheat in each village was used.