INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - May 1964

Paper I : Official Statistics and Descriptive Statistics
(Theoretical)

Time 4 hours — Full marks : 100

a) Figures in the margin indicate full marks.
b) Answer six questions in all - not exceeding three from any group.
c) Use of calculating machines is not permitted.

Group A

1.a) Write a critical note on the present position of statistics of live-stock and livestock-products in India. Give the names of the publications containing these statistics. [10]

b) What suggestions have you to offer for improving their reliability? [6]

2.a) Give a detailed account of the nature and scope of 'civil aviation' statistics, available at present in India. What are the publications you will consult for obtaining these statistics? [10]

b) Enumerate some of the important measures that you will adopt for ensuring comparability among statistics pertaining to different modes of transport. [6]

3. a) Describe the scope and coverage of the available foreign trade statistics of India. Also, suggest some possible improvements. [10]

b) What statistical devices have been adopted to ensure international comparability of foreign trade statistics? [6]

4. a) What does an index number of industrial production attempt to measure? [4]

b) Describe the scope and method of construction of the official and non-official indices of agricultural production in India.

What are their limitations, if any, and what steps would you suggest for their improvement? [8 + 4]

Group B

5.a) Define the concepts of product moment correlation coefficient between two random variables x and y, and of correlation ratio of y on x. [8]

b) If x and y are independent random variables show that
$r(x + y, x - y) = r^2(x, x + y) - r^2(y, x + y)$
where $r(u, v)$ denotes the coefficient of correlation between u and v. [8]

6.a) Show that, if np be a whole number, the mean number of successes in n independent trials with probability p of a success in any single trial, coincides with the mode. [6]

b) Eight mice are selected at random and they are divided into two groups of 4 each. Each mouse in group A is given a dose of certain poison, P, which is expected to kill one in four; each mouse in group B is given a dose of another poison Q, which is expected to kill one in two. Calculate the probability that there may be fewer deaths in group B than in group A. [10]

Please Turn Over

7.a) The standard deviation of two sets containing n_1 and n_2 members are σ_1 and σ_2 respectively, being measured from their respective means m_1 and m_2. If the two sets are grouped together as one set of $n_1 + n_2$ members, show that the standard deviation, σ, of this set measured from its mean, is given by

$$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2} + \frac{n_1 n_2}{(n_1 + n_2)^2}(m_1 - m_2)^2 \qquad [10]$$

b) Show that the mean deviation from the median is less than that measured from any other value. [6]

4.a) Prove that

$$E\left(\frac{\hat{P}\hat{Q}}{(n-1)}\right) = \frac{PQ}{n}$$

where $P\ (= 1 - Q)$ is the proportion of objects possessing a particular attribute A in an infinite population and $\hat{P}\ (= 1 - \hat{Q})$ is the proportion of objects possessing the attribute A in a random sample of n objects from the population concerned. [6]

b) Instead of ~~taking a sample in the above manner~~, if the objects of the sample are selected at random in succession ~~until a fixed number~~ m of objects possessing A is obtained in the sample, find the probability distribution of the size n of the sample and prove that

$$E\left(\frac{m-1}{n-1}\right) = P \qquad [10]$$

Neatness (for groups A and B). [4]

Please Turn Over

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - May 1964

Paper II : Probability Theory and Statistical Methods
(Theoretical)

Time : 4 hours Full marks : 100

a) Figures in the margin indicate full marks
b) Answer six questions in all - not exceeding three from any group.
c) Use of calculating machine is not permitted.

Group A

1.a) Each of the n urns contains x white and y black balls. One ball is transferred from the first urn into the second, then one ball from the second into the third urn and so on. Finally one ball is taken from the last urn; find the probability of its being black. [8]

b) There are n cards numbered 1, 2, upto n which are to be placed in n baskets similarly numbered. Find the probability that each card is placed in the wrong basket, if the placing of cards in the baskets is random. [8]

2. Define the distribution function of a random variable and obtain its various properties. If X and Y are independent uniform random variables in [0, 1], obtain the joint distribution function of X+Y and X-Y and find out whether they are independent. [16]

3. Define the moment generating function of a random variable and indicate its uses. Find this function for the random variable X with probability distribution

$$P(X = r) = \binom{n}{r} p^r q^{n-r}, \quad 0 < p < 1, \quad p+q=1,$$
$$r = 0, 1, \ldots, n.$$

and hence find the mean and variance of X.

Also find the covariance between

$$\frac{X}{n} \text{ and } \frac{n - X}{n}$$ [16]

4. Define the bivariate normal distribution and find the marginal distribution of each variate. Find also the conditional distribution of one variate when the other is fixed and hence find the mean and variance of this conditional distribution. (16)

5.a) State the weak law of large numbers and prove it for the case of a sequence of independent and identical random variables, when the variance is known to exist. [8]

b) Obtain the Poisson approximation to the Binomial distribution giving the conditions under which it is true. [8]

Group B

Show that the least square estimates of the parameters in a linear model are unbiased. How can one find confidence intervals for such parameters? [16]

Define a χ^2-distribution and find its mean and variance. Show that if X and Y follow independent χ^2-distributions with degrees of freedom m and n respectively then Z = X+Y, has a χ^2-distribution with m+n degrees of freedom. [16]

Please Turn Over

8. What is a maximum likelihood estimate and what are its properties ? Find the maximum likelihood estimate of the mean m of a Poisson population based on n independent observations from the population. [16]

9. Describe the method of evaluation of large sample standard errors of functions of sample moments and use them to calculate the standard error of the sample fourth central moment m_4. (16)

10. Explain the idea of regression of one variable on another and the connection between the regression coefficient and the correlation coefficient. How does this idea extend to the case of more than two variables. (16)

11. Discuss the following points in connection with Wilcoxon's test:

i) Statistical model, (ii) the null hypothesis and the alternative, iii) the test criterion, (iv) the critical region, (v) the distribution-free nature of the test.

Find the large sample distribution of Wilcoxon's test criterion under the null hypothesis. [16]

Neatness (for groups A and B). [4]

Please Turn Over

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - May 1964

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

Time : 4 hours — Full marks 100

i) Figures in the margin indicate full marks.

ii) Answer six questions in all - not exceeding three from any group.

iii) Use of calculating machines is not permitted

Group A

1.a) Comment on the use and importance of sample surveys and the principal steps involved in their planning and execution. [8]

b) Discuss the problems that arise in the construction of a frame. [3]

A city Directory, four years old, lists the addresses in order along each street, and gives the names of the persons living at each address. For a current interview survey of the people in the city, what are the deficiencies of this frame? Can they be remedied by the interviewers during the course of the field work? In using the directory, would you draw a list of addresses (dwelling places), or a list of persons? [5]

2.a) Define a probability sample. [2]

b) A sample of size n is drawn one by one with equal probability from a population containing N units in two different ways (i) with replacement, and (ii) without replacement. Show that in either case, the sample mean is an unbiased estimate of the population mean. Obtain the variance of the sample mean in each of the two cases and indicate how you will estimate it. [8]

c) A simple random sample of n villages is selected from a district containing N villages. Of these, m villages were found to have irrigation facilities. Give an unbiased estimate of the total number of villages in the district having irrigation facilities. Give the variance of the estimate. How will you estimate the variance? [6]

3. What is systematic sampling? Discuss the situations under which it can safely be recommended for adoption. [6]

In a population of N(= nk) units consisting of a linear trend, show that a systematic sample of size n is less precise than a stratified random sample with strata of size 2k and two units per stratum if $n > (4k + 2)/(k + 1)$. [10]

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

a) Interpenetrating samples
b) Regression method of estimation
c) Cost function in sample surveys
d) Cluster sampling. [16]

Please Turn Over

Group B

5. The Martindale Wear Tester machine consists of four rectangular brass plates on each of which is fastened an abrading surface consisting of special quality emery paper. Four weighted bushes into which test samples of fabric are fixed rest on the emery surfaces and a mechanical device moves the bushes over the surface of the emery, thus abrading the test specimens. The loss in weight after a given number of cycles is used as a criterion of resistance to abrasion. Five different rubber covered fabrics A, B, C, D and E were tested in five different runs of the machine according to the following scheme:

Run	Position in machine (brass plates)			
	1	2	3	4
1	A	E	C	D
2	E	A	B	C
3	D	F	A	B
4	C	L	E	A
5	B	C	D	E

a) identify the type of design and indicate briefly its advantages. [8]

b) describe the analysis appropriate for this design. [8]

6. Describe the analysis of a Latin Square experiment with two missing observations and provide expressions for the standard error of the estimated treatment difference. (You may assume that the observations lost relate to two different treatments). [10+6]

7. An experiment was conducted as follows to compare the abilities of two culture media to detect the presence of SALMONELLA organisms in specimens of faeces. Samples from each of 222 specimens were grown on each medium. For each sample a record was made as to whether presence (+) or absence (-) of the organism was found. The results are given below:

Medium A	Medium B	No. of specimens
+	+	7
+	-	5
-	+	0
-	-	210
	Total	222

i) give your comments on the suitability of the design adopted. [6]
ii) analyse the data. [10]

8. Write short notes on any two of the following

a) Role of randomization in planning of experiments.
b) Uniformity trials.
c) Multiple comparison procedures. [8 + 6]

Mentions (for groups A and B). [4]

Please Turn Over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1964

Paper IV : Applied Statistics (Theoretical)

Time : 4 hours Full marks : 100

a) Candidates will be required to answer questions from those two groups of subjects only for which they have registered their option.

b) Separate answer books are to be used for each of the two groups attempted.

c) Figures in the margin indicate full marks.

d) Use of calculating machines is not permitted.

GROUP A - ECONOMICS STATISTICS

(Answer any three questions from this group; for neatness 2 marks only)

1. Briefly describe the methods of calculation of index numbers of industrial production, and employment. How are these index numbers useful in the measurement of changes in labour productivity ? (10+6)

2. What do you mean by family budget data ? How will you use these data for the estimation of income elasticities of demand for some items, say, foodgrains, cloth, and entertainment ? How are these estimates of elasticities useful ? (3+6+5)

3. Clearly explain the following concepts :

(i) Gross national product (ii) Net national product (iii) National income at market price (iv) National income at factor cost (vi) National income at constant price (vi) National income at current price (vii) Domestic income (viii) Disposable income. (16)

4. How will you use concentration (Lorenz) curve for the purpose of studying whether inequality of income distribution has been reduced in the plan period in India ? Derive the form of this curve, when the income distribution follows Pareto's law. (9+7)

5. Write notes on any two of the following : (8+8)

i) business forecasting on the basis of time series

ii) assumptions in Cobb-Douglas production function

iii) input-output table

iv) statistics of manufacturing industries in India.

Please turn over

GROUP B - STATISTICAL QUALITY CONTROL

(Answer any three questions from this group; for neatness 2 marks only)

1. (a) Explain what is meant by the statement that a process is out of 'Statistical control'. What is meant by the term process capability ? (3+3)

(b) What are the main considerations involved in the selection of 'rational sub-groups' and how do they limit the ability of the control charts to detect significant causes of variation ? (5+5)

2. In the testing of a certain ballistic characteristic of ammunition it is desired that all values of this quality characteristic, fall between 1,680 and 1,800 units. $\bar{x}$ and R charts on the basis of samples of size 4 have been plotted from past samples. The R chart has shown excellent statistical control with an $\bar{R}$ of 34 units. The $\bar{x}$ chart has shown a decided lack of control. Past experience had shown that this is inevitable as the centering of the process is very difficult with respect to this quality characteristic.

It has been decided that it will be considerably more economical to test 16 items on an 'attribute' basis than 4 items on a 'variable' basis. In this connection it was decided to maintain a number defective chart. For this purpose the 'go', 'not-go' testing was adopted using special severe test limits. These are to be established at $\mu \pm G$ so that the appropriate 'p' (artificial defectives) is 0.15 for the "above limits" (corresponding to the 'not-go' limit) and is also 0.15 for the "below limits" (corresponding to the 'go' limit). Indicate :

a) with specification limits of 1680 and 1800 as stated, what would be the values of these special severe test limits ? (4)

b) with 16 items tested in each sample, where would the control limits be placed on the number defective chart ? (4)

c) if the mean of the process is at 1730 units, what would be the expected number of plotted points before a shift is detected ? [State clearly all the assumptions underlying your calculations] (8)

3. A vendor inspects his own product under a Dodge-Romig 2% AOQL plan, screening all rejected lots before shipment. The lot size is 1000 and the assumed process average 1% (one percent). The purchaser inspects the same lots under a Dodge-Romig 5% LTPD plan, returning all rejected lots to the vendor. The plan used is taken from the process average column headed "0.51 - 1.00". Single sampling is used.

a) what is the probability that a 2% defective lot will pass both the vendor's and the purchaser's inspection ? (8)

b) what is the probability that a 2% defective lot will pass the vendor's inspection and be rejected by the purchaser's ? (8)

Please turn over

L.T.P.D. Percent Defective = 5.0%

Lot Size	process average % 0.51 - 1.00			1.01 - 1.50		
	n	c	AOQL %	n	c	AOQL %
601 - 800	100	2	1.2	130	3	1.2
801 - 1000	105	2	1.2	155	4	1.4
1001 - 2000	130	3	1.4	180	5	1.6

A O Q L = 2%

Lot Size	process average % 0.81 - 1.20			1.21 - 1.60		
	n	c	p_t %	n	c	p_t %
601 - 800	65	2	8.0	65	2	8.0
801 - 1000	65	2	8.1	65	2	8.1
1001 - 2000	65	2	8.2	95	3	7.0

4. (a) Show how to determine the values of n and k for a variables sampling plan-known standard deviation when either an upper or lower specification is given and when AQL, α and LTPD, β are stated. (10)

(b) Show also how to obtain the o.c. curve of the sampling plan. (6)

GROUP C - STATISTICAL METHODS IN GENETICS

(Answer any two questions from this group: for neatness 2 marks only)

1. What data do you collect for examining whether a gene is sex linked or not ? Indicate the statistical analysis needed to establish such a linkage. Give a few examples of genes which are linked with the X chromosome. (24)

2. Write down the equation for estimating the linkage parameter from the intercross data $\frac{Ab}{aB} \times \frac{Ab}{ab}$. Suppose we have estimates of the same linkage parameter and standard errors, from two types of data. How do you examine whether the two estimates are consistent ? (24)

3. Describe the role of the discriminant function in plant selection. How is the discriminant function determined ? Give an expression for the genetic advance due to selection by the discriminant function. (24)

Please turn over

GROUP D - VITAL STATISTICS & DEMOGRAPHY

(Answer any three questions from this group; for neatness 2 marks only)

1. (a) Define a 'stationary population' and explain how its characteristics are determined by life-table functions. (10)

(b) It is stated that the mortality of a body of professional men, say, qualified doctors, is lighter than that of the general population, because the average age at death of qualified doctors is found to be higher than that of the general population.

Do you agree with this statement ? Give reasons for your answer. (6)

2. (a) In a sample survey of a population, for each selected household, the birth schedule gives, among other things, live births in the year preceding the date of survey, by sex, age of mother at birth and order of birth and the population schedule gives, among other things, sex, age and marital status of members. What addional information would you require in the population schedule to enable you to extract the age-parity specific fertility rates of married women ? (illegitimate fertility is to be ignored) (6)

(b) Define gross and net reproduction rates. Examine the appropriateness of each as an index of fertility, or as a measure of the true reproductivity of the population. (10)

3. (a) What is a mortality index and what is the necessity for constructing such an index ? (6)

(b) Derive two mortality indexes based respectively on direct and indirect standardisation, and indicate the circumstances in which the use of one or the other may be more appropriate. (10)

4. (a) Distinguish between estimation and projection of population. (4)

(b) Explain how the population by sex and age, 5 years after the last census, may be obtained approximately when full registration data are available. (8)

(c) How would you proceed when the registration data are not trustworthy and in what essential respect would your result be different in such a case ? (4)

5. Write notes on any two of the following : (8+8)

a) Demographic transition

b) Hospital morbidity statistics and their uses.

c) Vital registration in India.

Please turn over

GROUP E - EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Answer any two questions from this group; for neatness 2 marks only)

1. Explain the terms : common factors, factor loadings and communality as used in Factor analysis.

Show that under Thurstone's multiple factor theory, the rank of the reduced correlation matrix of scores is equal to the number of common factors. (24)

2. Define reliability and validity of a test. Describe and compare the following methods of assessing the reliability of a test :

i) Test-Retest method

ii) Parallel Form

iii) Split-Half technique

iv) Rational equivalence.

How would you assess the validity of a test ? (24)

3. A large number of applicants for admission to a course in statistics were given a selection test, and only those whose scores X in the test were greater than c, were selected. After one year of training, the selected candidates were given a test in statistics and their scores Y in this test are also available.

How would you use such data to estimate the correlation coefficient between the scores in the selection test and the scores in the statistics test for the population of all candidates ?

Explain clearly the assumptions underlying the procedure. (24)

ooooOoooo

Please Turn Over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1964

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

Time : 5 hours Full marks : 100

a) Figures in the margin indicate full marks.

b) Use of Calculating machines is permitted.

Group A

(Answer all questions)

1. The range of a continuous random variable is known to be (-1 to +1). The values of the cumulative distribution function (c.d.f.) of the random variable at two intermediate points -0.5 and 0.5 are given to be .37640 and .83934 respectively. Find (a) the value of the cumulative distribution function at 0.25 and (b) the value of the density function at 0.75. [1

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

x	f(x)
301	5.70711026
302	5.71042702
303	5.71373201
304	5.71702770
305	5.72031178
306	5.72358510

[1

Group B

(Answer three questions from this group of which question No. 3 is compulsory)

3. [Compulsory] The following table shows the age-distribution of the male populations of India and the United Kingdom in 1931.

Age-grouping	Population (lakhs)	
	India	U. K.
0 - 5	214	18
5 - 10	250	19
10 - 15	222	20
15 - 20	157	10
20 - 25	145	16
25 - 30	161	14
30 - 40	257	27
40 - 50	184	25
50 - 60	120	19
above 60	100	17

a) Draw the two frequency distributions on the same graph paper. [5]

b) Draw the cumulative age-distribution (ogive) curves of the two populations on the same graph paper. [5]

Please Turn Over

3(contd.)

c) Calculate the median ages and the semi-interquartile ranges of the two distributions from the ogive curves you have drawn. [4]

d) Write a short note comparing the age-distributions of India and the U.K. on the basis of (a), (b) and (c). [4]

4. 1000 groups of 12 students each were observed and the number of students in each group who habitually wear glasses recorded. The following table shows the data.

no. of students who wear glasses	0	1	2	3	4	5	6	7	8	9	10	11	12	total
no. of groups	6	44	124	209	232	196	116	52	16	4	1	0	0	1000

a) Estimate the probability that a student selected at random wears glasses, and on the basis of this estimated probability, calculate the binomial frequencies for the above data. [2 + 8]

b) Calculate the χ^2 value of goodness of fit of the binomial distribution and examine if this value is significant. [6]

5. The following table shows the yield in ounces of dry bark (X_1), height in inches (X_2) and girth in inches (X_3) of 15 cinchona plants.

serial no.	X_1	X_2	X_3	serial no.	X_1	X_2	X_3
1	42	21	3	9	10	6	3
2	25	7	2	10	20	4	4
3	10	5	1	11	27	8	4
4	44	10	4	12	13	7	3
5	56	13	6	13	49	12	5
6	30	12	3	14	27	6	3
7	32	13	4	15	55	16	7
8	25	5	2				

a) Obtain the equation of linear regression of X_1 on X_2 and X_3. [10]

b) From the above equation estimate the yield of dry bark from the values of X_2 and X_3 corresponding to plants with serial numbers 1, 3 and 13. [2]

c) How much of the variation in X_1 is explained by the regression and how much by deviations of observed values from regression? [4]

Please Turn Over

6. The following table shows yearly public investments under Central Government for the financial years 1901-1902 to 1920-21.

year	public investment under Central Government (crores Rs.)	year	public investment under Central Government (crores Rs.)
1901 - 1902	17.45	1911 - 1912	31.37
1902 - 1903	20.35	1912 - 1913	35.37
1903 - 1904	21.42	1913 - 1914	35.32
1904 - 1905	23.51	1914 - 1915	39.85
1905 - 1906	29.97	1915 - 1916	45.91
1906 - 1907	29.23	1916 - 1917	48.17
1907 - 1908	34.23	1917 - 1918	46.85
1908 - 1909	35.06	1918 - 1919	40.96
1909 - 1910	27.39	1919 - 1920	36.33
1910 - 1911	32.20	1920 - 1921	43.80

a) Plot the data on a graph paper. [4]

b) Fit a second degree polynomial trend to the data and draw the graph of this polynomial on the graph of the data, by plotting five selected points. [8]

c) Write brief comments on the nature of growth and fluctuations of governmental investments during the period indicated. [4]

Group C

(Answer all questions)

7. Select any three of the following items :

a) Foreign Trade, (b) Irrigation, (c) Industrial Production,

d) Land Utilisation Statistics, (e) Mining.

For each of the three items selected, write down

i) the names of publications from which you can get the required information in India,

ii) the office from which the publication is issued,

iii) periodicity of publication and

iv) the time lag in publication. [9]

8. Collect relevant information from the publications supplied in respect of changes in (i) per capita net national output in India and (ii) index number of industrial production in India for the latest available 10 consecutive years.

Comment on the salient features of data thus collected. [16]

Please Turn Over

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - May 1964

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.
ii) Use of calculating machines is permitted.

Group A

(Answer all questions from this group)

1.a) In a fertilizer experiment two sets of 10 plots of equal area, were selected at random. All the plots in the first set received fertiliser A, while all the plots in the second set received no fertiliser at all. The yields of wheat grown on these plots are given below:-

Yields (in lbs.) from plots with fertiliser A:	512, 432, 470, 533, 571, 482, 450, 396, 412, 475
Yields (in lbs.) from plots with no fertiliser:	317, 378, 575, 402, 411, 370, 501, 430, 472, 503

It is claimed that the fertiliser A is effective in increasing the yield per plot of wheat. Do you agree? Give reasons. [10]

b) The following table gives for a sample of married women, the level of education and marriage adjustment score:

		marriage - adjustment score			
		very low	low	high	very high
level of education	College	24	97	62	58
	High School	22	28	30	41
	Middle School	32	10	11	20

Can you conclude from the above, that higher the level of education, the greater is the degree of adjustment in marriage? [10]

2.a) Use any appropriate statistical test to examine that the following 4 samples come from the same population.

sample nos.			
I	II	III	IV
13	24	12	4
40	22	9	19
20	24	16	24
19	24	17	26
45	26	17	30
26	28	18	30
25	12	19	27
56	30	22	31
27	42	31	25
30	53	42	22
14	61	51	29
29	74	72	22
27	25	20	22
17	23	22	31

[10]

Please Turn Over

2.b) Four plots of land were divided into 5-sub plots each. For each plot, five treatments were assigned at random to the 5-sub plots. The yields from the subplots are given below. Test whether the 5 treatments are equally effective in respect of mean yield.

Plots \ Treatments	A	B	C	D	E
1	306	352	304	295	497
2	288	397	442	263	415
3	307	322	434	310	467
4	268	308	404	166	428

[10]

Group B

(Answer any two questions from this group)

3. The table below gives the yield of wheat as observed in an experiment carried out in a 4 × 4 Latin Square. The four manurial treatments are denoted by numbers 1, 2, 3 and 4 in parentheses.

Yields in a 4 × 4 Latin square experiment.

(2)	(3)	(4)	(1)
425	442	540	340
(4)	(1)	(2)	(3)
384	512	490	408
(3)	(4)	(1)	(2)
506	508	536	600
(1)	(2)	(3)	(4)
451	568	499	347

Analyse the data and write a brief report on your findings. [15]

4. The following data relate to the initial weights (in pounds) and growth rates (in pounds per week) of 15 pigs, classified according to pen and type of feed given.

Examine the differences between the three types of feed A, B, and C in their effect on the growth rates of pigs, correcting for the difference in the initial weights of pigs.

Initial weight and growth rate of pigs.

pen	type of feed	initial weight (in lbs.) x	growth rate in (lbs.per week) y
I	A	48	9.94
	B	43	10.00
	C	48	9.75
III	C	33	7.63
	A	35	9.32
	B	41	9.34
V	B	37	9.67
	A	32	8.82
	C	30	8.57
II	B	32	9.24
	C	28	8.63
	A	32	9.48
IV	C	50	10.37
	A	48	10.56
	B	46	9.68

[You may like to make use of the following computations:

$\Sigma x = 508$, $\Sigma y = 141.03$,

Σx^2 [illegible]

(contd. from Group B)

5. Draw up a balanced confounded design for a factorial experiment involving three factors A, B, C each at three levels 0, 1, 2 using blocks of 9 plots, in 4 replications and with partial confounding of only the triple factor interaction. [15]

Group C

(Answer both the questions from this group)

6. A simple sample of n = 50 households was drawn without replacement from a village in which there are in all N = 250 households. It was found that amongst the sampled households, there were only 8 households each possessing a transistor radio. These households had respectively 3, 5, 3, 4, 7, 4, 4 and 5 members.

Estimate (a) the total number of households in the village possessing transistor radios (b) the total number of persons in these households; and calculate the standard errors of these estimates. [15]

7. For a certain survey, it was been decided to use a stratified sampling design, using 5 strata, sampling within each stratum being with replacement and with equal probabilities. Let N_i, n_i and σ_i denote respectively the population size, sample size and the standard deviation of the characteristic under study in the i-th stratum. Let c_o = Rs.500 be the overhead cost for the survey, c_i the cost for sampling and processing a single unit from the i-th stratum and C = Rs.10,000 the total cost fixed for the survey. The numerical values of these constants are given below.

stratum number (i)	N_i	σ_i	c_i (Rs.)
1	37800	28.5	3.50
2	52600	18.6	2.75
3	82000	27.6	2.25
4	41600	21.2	3.00
5	20900	16.8	2.50

Determine the optimum values of n_i's in the sense of minimising the variance of the customary estimate of the population total. [15]

Please Turn Over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1964

Paper VII : Applied Statistics (Practical)

Time : 5 hours Full marks : 100

a) Candidates will be required to answer questions from those two groups of subjects only for which they have registered their option.

b) Separate answer books are to be used for each of these two groups. attempted.

c) Figures in the margin indicate full marks.

d) Use of Calculating Machines is permitted.

GROUP A : ECONOMIC STATISTICS

(Answer all questions)

1. Average weekly wages (in dollars) in manufacturing industries of U.S. along with the consumer price index numbers are given below for 14 months. Compare the change in money wage with that in real wage from November 1955 to December 1956. (10)

month	weekly wage	consumer price index
1955 November	79.52	115.0
December	79.71	114.7
1956 January	78.55	114.6
February	78.17	114.6
March	78.78	114.7
April	78.99	114.9
May	79.00	115.4
June	79.19	116.2
July	79.00	117.0
August	79.79	116.8
September	81.40	117.1
October	82.21	117.7
November	82.42	117.8
December	84.05	118.0

2. On the basis of monthly sales (in million dollars) of a certain commodity for the years 1951 to 1955, the following calculations were made. (15)

Trend : $y = 25.74 + 0.455\ t$ where origin is at January 1951

t = time unit (one month)

y = monthly sales.

Seasonal Index :

month	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
index	79.6	76.6	94.9	97.5	105.5	97.9	85.1	89.2	102.1	121.9	113.4	136.3

Estimate monthly sales for 1956.

Please turn over

3. Index numbers of demand for agricultural products (y), prices of agricultural products (x_1) and national income (x_2) are given below for the years 1950 to 1959. Obtain the elasticities of demand for agricultural products with respect to price and income. What will be the change in demand, if income increases by 5 p.c. and price rises by 2 p.c. ? (20+5)

Year	y	x_1	x_2
1950	102	89	90
1951	98	99	96
1952	100	100	100
1953	105	91	98
1954	117	93	105
1955	120	72	96
1956	120	75	100
1957	127	91	113
1958	118	91	114
1959	134	96	125

GROUP B : STATISTICAL QUALITY CONTROL

(Answer any two questions)

1. (a) The following table gives the number of missing rivets noted at aircraft final inspection :

Air plane number	Number of missing rivets	Air plane number	Number of missing rivets	Air plane number	Number of missing rivets
1	8	11	23	21	10
2	16	12	16	22	22
3	14	13	9	23	7
4	19	14	25	24	28
5	11	15	15	25	9
6	15	16	9		
7	8	17	9		
8	11	18	14		
9	21	19	11		
10	12	20	9		

Compute trial control limits and plot a control chart for number of defects per unit. What would be the achievable standard value of the number of defects per unit ? (16)

(b) A control chart for defects per unit uses probability limits corresponding to probabilities of 0.975 and 0.025. The central line on the control chart is at 2.0. Determine the correct position of the upper and lower control limits when n = 5, i.e. 5 units are inspected at a time. (9)

Please turn over

2. The lot size N is 2000 in a certain AOQL inspection procedure. The desired AOQL of 1.0% (one percent) can be obtained with any one of three single sampling plans. These are n = 36, c = 0; n = 80, c = 1; and n = 130, c = 2. Which plan will involve the minimum total inspection, considering both Sampling inspection and screening of rejected lots, if a large number of lots of 0.25% defectives are submitted for acceptance ? (25)

3. (a) In a certain variables acceptance plan, samples of 4 are tested from each lot. The results are plotted on an $\bar{x}$, R control chart with control lines drawn at the usual 3σ limits. If the average of a sample of 4 lies within the control limits on $\bar{x}$ chart and the range also lies within the control limits on the R chart, a lot is accepted; otherwise rejected.

Assume that after a long period of satisfactory lots, a lot is submitted that is 10% defective. Assume for each lot a Normal distribution for the variable characteristic with a σ' equal to that used in computing the control limits. For the 10% defective lot the average may be assumed to have shifted to such a high value that the highest 10% of the frequency distribution is above the upper specification limit. Compute the probability that this defective lot will be accepted ?

(b) A sample of 5 bearings is taken from an automatic grinder and their diameters measured. The results are 1.003, 1.001, 0.996, 0.997, 0.998. The process is considered satisfactory if 99% of the output lies within the range 1 ± 0.005. On the basis of this sample, decide whether the grinder needs adjustment. (9)

oooOooo

Please turn over

GROUP C - STATISTICAL METHODS IN GENETICS

(Answer all questions from this group)

8. The following data relate to sex of children born to mothers of different (O, A, B, AB) blood groups.

Sex of child	Blood group of mother			
	O	A	B	AB
♂	200	192	156	60
♀	160	180	150	75

Examine

(a) whether the over-all sex ratio is 1 : 1,

(b) whether the sex ratio is different for different blood groups of mothers, and

(c) whether the figures for ♂ and ♀ children of mothers with O blood group as given in the above table show a ratio consistent with some previous data giving 400 ♂ and 350 ♀ children for mothers of O group. (25)

9. The mean values, standard deviations and correlations of three characteristics y, x_1, x_2 in a population are as follows:

	y	x_1	x_2
mean	200	35	46
s.d.	6	3	2

Correlations

	y	x_1	x_2
y	1	.50	.50
x_1	.60	1	.40
x_2	.50	.40	1

(a) Compute the linear function of x_1 and x_2 useful in selecting individuals for the characteristic y.

(b) How should the selection be done on the basis of x_1, x_2 (i.e., what value should the linear function determined in (a), exceed for selecting an individual) if, in the long run, 25 % of the available individuals are selected ? Assume normality of distribution.

(c) What is the mean value of y in the selected group.? (25)

oooOooo

GROUP D : VITAL STATISTICS & DEMOGRAPHY

(Answer any two questions)

1. The following data are taken relating to a certain European Community :

Table 1 : Population of women at ages 15 - 44

Age group	Year A		Year B	
	No. of married women	No. of non-married women	No. of married women	No. of non-married women
15 - 19	1,000	8,000	1,000	7,500
20 - 24	5,200	4,000	5,700	3,400
25 - 29	7,100	1,900	7,200	1,700
30 - 34	7,900	700	7,600	700
35 - 39	7,800	600	7,500	600
40 - 44	8,000	500	7,600	600
Total	37,000	15,600	36,600	14,500
Total Population	270,000		250,000	

Table 2 : Fertility rates by age of women, & number of births (for live births only)

Age group	Year A		Year B	
	legitimate	illegitimate	legitimate	illegitimate
15 - 19	.100	.005	.090	.004
20 - 24	.250	.003	.240	.002
25 - 29	.150		.140	-
30 - 34	.090	-	.090	-
35 - 39	.060	-	.060	-
40 - 44	.020	-	.020	-
No. of birth	3,804	54	3,752	37

Considering the year A to be the standard year, calculate indexes to show the relative fertility experiences of the two years and examine the inadequacy of the comparison of the respective crude birth rates. (25)

Please turn over

... of India (in lakhs) according ... consuses :

Year	Population (in lakhs)
1931	2,790
1941	3,187
1951	3,611
1961	4,392

Estimate the expected population of India in 1971, on the basis of the growth equation :

$$P_t = P_0 \cdot e^{\delta_t \cdot t}$$

where P_t and P_o are the populations at times t and o respectively, and $\delta_t = \alpha + \beta t$ is the momentary rate of growth at time 't'. Comment on the reasonableness or otherwise of this estimate from other considerations. (25)

3. The following table gives particulars about White Females, United States, 1950 :

Age group (year)	Age-specific fertility rate	Proportion surviving from birth to mid-point of age-group
10 - 14	.0004	.96928
15 - 19	.0700	.96683
20 - 24	.1904	.96338
25 - 29	.1651	.95915
30 - 34	.1026	.95387
35 - 39	.0514	.94658
40 - 44	.0145	.93569
45 - 49	.0010	.91912

Calculate the average length of the generation of the white females on the basis of the above data. (25)

Please turn over

GROUP E : EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Answer all questions)

1. The following table gives the ranks assigned to various moral items, by employers and employees.

Items	Employer's Ranking	Employees' Ranking
1	1	7
2	2	3
3	3	1
4	4	5
5	5	8
6	6	4
7	7	6
8	8	2

Indicate whether there is any relationship between the ranks assigned by the two groups. Comment on the results obtained by using product moment Correlation Coefficient. (16)

2. The Correlation matrix for 5 test variables is given below :

	1	2	3	4	5
1	-	-	-	-	-
2	.50	-	-	-	-
3	.41	.58	-	-	-
4	.30	.44	.54	-	-
5	.21	.34	.57	.62	-

Assuming that the tests may be described by two common factors, obtain (i) the factor matrix (ii) the first factor residuals. On the basis of the first factor residuals can you justify the above assumption. ? (1

3. The following table represents the judgements of 95 members of a wellknown symphony orchestra, concerning preferences for music of eight composers. Compute the scale values for the eight composers represented. The table gives the values of p_{ij} where p_{ij} is the proportion of times where composer j is preferred to composer i.

Value of p_{ij}

i \ j	2	3	4	5	6	7	8
1	.242	.842	.916	.884	.589	.411	.126
2		.895	.968	.989	.842	.758	.189
3			.853	.547	.232	.126	.053
4				.189	.126	.053	.032
5					.189	.084	.042
6						.326	.116
7							.147

(1

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - May 1964

Paper VIII: Subjects of Specialisation - I

Time: 4 hours Full marks: 100

i) Candidates will be required to answer questions from that group only for which they have registered their options.

ii) Figures in the margin indicate full marks.

Group A : Economic Statistics -(Econometrics)

(Answer any five questions)

1. a) What is the problem of identification in the estimation of simultaneous linear relationships? How would you proceed if you find one of the relationships is (i) under-identified, (ii) over-identified? [14]

b) Examine whether the following model is identified or not and describe how you would estimate the parameters involved.

$$D = \alpha_1 P + \alpha_2 y + u_D$$
$$S = \beta_1 P + \beta_2 w + u_S$$
$$D = S.$$

Here the endogenous variables D, S and P are the demand, supply and price variables and y and w are the exogenous variables income and weather respectively. [6]

2. Write a note on the applications of the lognormal distribution in economics and the related problems of estimation. When is the bivariate lognormal distribution needed and how are the parameters estimated? [20]

3. a) Describe the production-function with constant elasticity of substitution. Is it any improvement on the Cobb-Douglas production-function? [10]

b) What are the methodological problems involved in fitting a production-function to time-series data? [13]

4. a) Write a note on the generalised least squares estimation of parameters in single-equation-models. What are their properties? How are they needed in the statistical analysis of economic data? [12]

b) If $X_t = a + bt + \varepsilon_t$

where ε_t has mean zero and forms the first order autoregressive process given by

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t,$$

derive the generalised estimate of b. [8]

5. a) What are the measures of inequality of income and wealth distribution? How would you estimate them? [10]

b) How would you set about studying changes in income-distribution over time in India? What are the methodological and practical problems involved? [15]

6. Examine in detail the problems of formulation and of estimation of demand relationships from cross-section and/or time-series data. [20]

7.a) What are the open and closed Leontief's models of the economy? How are these models made dynamic? [10]

b) Examine the limitations and difficulties in constructing and using the input-output tables in the context of an underdeveloped or developing country. [10]

oooOooo

Group B - Techno-Commercial Statistics - (Statistical Quality Control)

(Answer any four questions)

1. Discuss briefly the problems ordinarily faced in starting and maintaining control charts in a manufacturing plant, with special reference to the statistical principles useful in dealing with them. [25]

2. What is narrow limit gauging? Examine the advantages of and the problems involved in the use of narrow limit gauging, for control of a measurable characteristic. [25]

3.a) Describe the salient features of the Mil. Std. 105 tables. How would you select a suitable plan from these tables for normal inspection? [20]

b) Examine the following statement. 'The Mil Std tables are more appropriate when the supply is regular and continuous whereas the Dodge Romig plans are more suitable when supply is occasional or sporadic.' [5]

4. Describe the statistical problems involved in determining the optimum operating conditions of a process affected by several variables. [25]

5. Write notes on any three of the following:

a) Role of standardisation in quality control

b) Specifications and process capabilities

c) Continuous sampling inspection plans

d) Cumulative sum charts

e) OC curve and indifference quality [25]

oooOooo

Please turn over

Group C - Biometric Methods

(Answer any four questions)

1. The distribution of the number of children per married couple can be described approximately by the formula

$$p(x) = \begin{cases} 1 - \alpha & \text{for } x = 0 \\ \alpha (1 - \theta)\, \theta^{x-1} & \text{for } x = 1, 2, \ldots \text{ ad inf.} \end{cases}$$

where $0 < \alpha < 1$, $0 < \theta < 1$ are unknown parameters and $p(x)$ denotes the proportion of couples with exactly x children. Only some couples, forming a proportion, say β of all couples are capable of producing albino children, and the chance is γ that any child born to such a couple will be an albino. The other couples are incapable of producing albino children.

(a) Work out the probability that a couple selected at random will have x children of whom y are albinoes. $x \geq y = 0, 1, 2, \ldots\ldots$. State clearly any extra assumptions that you may use in computing this probability.

(b) If a simple random sample of n couples is drawn with replacement from the population of all couples, and for each sampled couple the number of children x, and the number of albino children y are noted, how will you estimate the parameters α, β, γ and θ ? (25)

2. It has been suggested that the cranial capacity (C) of a skull can be predicted from the glabella - occipital length (L), the maximum parietal breadth (B) and the basio bregmatic height (H'), by means of a formula of the type

$$C = \alpha L^{\beta_1} B^{\beta_2} H'^{\beta_3}$$

where α, β_1, β_2 and β_3 are certain parameters.

Measurements on C, L, B and H' are available on $n = 26$ male skulls.

(a) How will you estimate the parameters ?

(b) It has been suggested that a simpler formula

$$C = \alpha L B H'$$

may do as well. How will you examine this point ?

(c) In any case, since C is a volume, dimensional considerations suggest that $\beta_1 + \beta_2 + \beta_3 = 3$. How will you test this hypothesis ?

Describe in detail all the assumptions that are needed to validate the statistical techniques adopted.

Please turn over

3. For a mixture of two Normal populations with possibly different means but same variance, work out the first four central moments.

Describe how you will estimate the four parameters namely the two means, the common variance and the mixture-ratio, on the basis of a large sample drawn from such a population. (

4. The lengths of the ten fingers are available for a large number of adult human males. Formulate some hypothesis of symmetry of these measurements which it may be of interest to examine on the basis of these data. Describe the test procedures that you would adopt. (

5. Describe Fisher's technique of discriminant function for classifying, on the basis of multiple measurements, an individual into one of two given populations.
Work out the probability of mis-classification by this technique when the apriori probability is $\frac{1}{2}$ that the individual belongs to either of two given multivariate Normal populations, with a common dispersion matrix. (

oooOooo

Group D - Design & Analysis of Experiments

(Answer any four questions)

1. Assuming that the yield of a plot is built up additively of the effects of the plot and of the treatment applied on it, work out the expected values of the mean squares for the different components in the analysis of variance for a randomised block experiment. (

Describe how to estimate the increase in efficiency due to the use of a randomised block design instead of a completely randomised one, from the results of randomised blocks experiment. (

2. Define a Partially Balanced Incomplete Block design with two associate classes. (5)

Examine whether the design given below (involving 12 treatments in 8 blocks of 3 plots each) can be considered as Partially Balanced with two associate classes.

(1, 2, 3); (4, 5, 6); (7, 8, 9); (10, 11, 12);
(4, 7, 10); (1, 8, 11); (2, 5, 12); (3, 6, 9)

If the above design is Partially Balanced write down all the parameters of the design; if it is not, demonstrate just which requirement is not fulfilled. (1

Write down computational instructions for the intra-block analysis of the results of an experiment with the above design. (1

Please turn over

3. Give a design for a $\frac{1}{2}$ replicate experiment involving 6 factors, each at 2 levels, using blocks of 8 plots. (10)

Describe the attractive features of the design chosen by you and explain how the experimental results are to be analysed. (15)

4. Describe the method of steepest ascent for locating the peak of a response surface. (10)

Show that prior knowledge of the slopes of the response surface (in different directions) is necessary for using this method. (8)

How do you propose to determine the slopes ? (7)

5. An astrologer claims that given the exact time of birth of a newborn baby, he can predict whether the baby will live for at least one year or die before attaining the age of one. Design an experiment to test his claim. Explain carefully how you will use the fundamental principles of randomisation, replication, and error control in designing this experiment. (25)

oooOooo

Group (E) : Sample Surveys -(Theoretical Aspects)

(Answer any four questions)

1. Define the following terms

a) Purposive Sampling

b) Simple Random Sampling

c) Response Errors

d) Interpenetrating Subsamples

e) Correlogram [25]

2.a) How is the population mean estimated when the sample is drawn one by one with varying probabilities without replacement?

b) How is the variance of the estimator of population mean estimated for the above scheme?

c) Examine if this estimate of variance could be negative.

d) Describe two schemes of sampling with varying probabilities without replacement for which positive definite estimators of the variance (of the estimate of mean) are available.

e) Describe a method of systematic selection with varying probabilities. [2

3.a) Show that under certain assumptions the regression estimator is better than the ratio estimator of the population mean.

b) How would you estimate the regression coefficient for a combined regression estimator when stratified simple random sampling is used?

c) How is double sampling used for regression estimation when the population mean of the auxiliary variate is not known in advance? [25

Group (E) (contd.)

4.a) Given the strata, how would you allocate the sample size to different strata to estimate the population mean? Give reasons.

b) If there are only two strata, how would you distribute your sample over them, to estimate the difference between the strata means? Give reasons.

c) Describe stratification after sample selection.

d) Describe a self-weighting stratified two-stage design. [20]

5.a) What is a problem of non-response?

b) Derive the expression for the optimum fraction of the non-respondents (in a mailed questionnaire survey) to be interviewed.

c) Describe Politz and Simon's scheme of tackling the problem of not-at-homes. [20]

Group (G) Statistical Inference

(Answer any five questions)

1.a) Define 'unbiasedness' and 'consistency' of an estimator. [5]

b) T_n is a biased estimator of a parameter θ, based on a random sample of size n, and $E(T_n) = \theta + b_n$. If $b_n \to 0$ and $V(T_n) \to 0$ as $n \to \infty$, show that T_n is a consistent estimator of θ. [5]

c) In order to estimate the proportion θ of a certain characteristic in a large population, a random sample of size n is taken and the number of individuals with the characteristic counted. This counting involves an error which may be considered constant for every given n, and is of the order $\sqrt{n}$. Examine if the sample proportion is a consistent estimator of θ. [5]

2.a) Examine if the distribution

$$f(x; p, \alpha) = \frac{\alpha^p}{\Gamma(p)} e^{-\alpha x} x^{p-1}, \quad x \geq 0$$

admits of a sufficient estimator of α. [6]

b) Show that if t is any unbiased estimator and T a sufficient statistic for a parameter θ, then $E(t / T)$ is an unbiased estimator which has a smaller variance than that of t. [6]

c) $x_1, \ldots, x_n$ is a random sample from a normal population $N(\mu, 1)$.

i) Show that $t = \frac{1}{n} \sum x_i^2$ is an unbiased estimator of $\mu^2 + 1$. [3]

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

Please Turn Over

3.a) Describe an iterative method for solving the maximum likelihood equations for two parameters and for obtaining the variances and covariances of the estimators. [15]

b) For a normal distribution $N(\mu, \sigma^2)$, μ is known but σ^2 is not known. Show that the sample variance is not the most efficient estimator of σ^2, and obtain the maximum likelihood estimator of σ^2. [15]

4.a) Show, with proofs, how to construct a most powerful test for a simple hypothesis against a simple alternative. [7]

b) Is this test 'unbiased'? [5]

c) For the normal distribution $N(0, \sigma^2)$, derive a uniformly most powerful test for the hypothesis $\sigma^2 = \sigma_0^2$ against alternatives of the type $\sigma^2 < \sigma_0^2$. [8]

5.a) Define a 'composite hypothesis'. [5]

b) Describe the likelihood-ratio method for constructing tests for composite hypotheses. [7]

c) There are k normal populations $N(\mu_1, \sigma^2)$, ..., $N(\mu_k, \sigma^2)$. A random sample of size n has been taken from each of the k populations. Obtain the likelihood ratio test for the hypothesis $\mu_1 = \ldots = \mu_k$ [8]

6.a) How are shortest confidence intervals for a parameter defined by Neyman? [6]

b) How are these intervals related to best critical regions for testing hypotheses regarding the parameter? [6]

c) Illustrate the above by considering the problem of finding a confidence interval for the mean of a normal population, whose variance is not known. [8]

7. Write a short essay on Wald's theory of decision functions, explaining Bayes and minimax procedures, randomised decision function admissible and complete classes. [20]

Please turn over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1964

Paper IX : Subjects of Specialisation - II

Time : 4 hours Full marks : 100

a) Candidates will be required to answer questions from that group only for which they have registered their options.

b) Figures in the margin indicate full marks.

Group A - Economic Statistics (Indian Economics and Economics of Planning)

(Answer Six questions in all - not exceeding three from any Section)

Section I

1. What are the main features of the Community Development Projects and National Extension Service of India ? Examine their usefulness for raising agricultural productivity. What changes in their organization will help in increasing agricultural production faster ? (16)

2. Outline the main arguments for setting up of public enterprises in India. Also give a brief account of the important industrial enterprises started by the government in recent years. (16)

3. Describe the present system of Industrial finance in India. Give suggestions for improving it and also indicate what, at present, is being done in this regard. (16)

4. Provide an analytic description of the existing tax structure in India. What changes would you like to make in it with a view to (a) making it more progressive, and (b) providing more resources for economic development ? (16)

5. Give a sketch of, and critically comment on, India's balance of payments problems since the end of Second World War. To what extent India's foreign exchange difficulties are a consequence of Plans. (16)

Section II

6. Critically examine the thesis that the emergence of agricultural surpluses is a necessary pre-requisite for industrialization of over-populated areas. Account for the smallness of marketable surplus in Indian agriculture. (16)

7. What, in your opinion, is the main strategy of Indian Planning ? Illustrate your answer with examples from Indian Five Year Plans. (16)

Please turn over.

8. In what sense the shortage of foreign exchange can act as a bottleneck in the process of economic development ? How far and in what manner internal savings and foreign exchange are substitutes for each other. (15)

9. Give a concise statement of the Harrod-Domar model of growth and compare it with the Mahalanobis model. (15)

10. What are the main determinants of the structure of production ? How can you project the structure of production in the future ? (15)

For Neatness. (1)

oooOooo

Group B : Techno-commercial Statistics - Operations Research/ Elementary of Book-keeping & Accountancy, Statistical Methods in Business.

I. Operations Research

i) Use separate answer book for this group.

ii) Answer any four questions.

1. (a) State and formulate the assignment problem of linear programming. Show that if in an assignment problem, we add a constant to every element of a row of the 'effectiveness matrix', then an assignment which minimises the total effectiveness in the original matrix, also minimises the total effectiveness in the modified matrix.

(b) State Konig's result regarding a matrix containing two types of elements - zero's & nonzero's and the lines covering the zero's. Explain how this result can be used to provide a method of solving the problem of assignment. (5+12)

2. (a) State the warehouse problem of programming and formulate it in mathematical form.

(b) Explain the principle of dynamic programming with reference to the above problem. How would you solve the problem by dynamic programming as described by you ? (7+10)

3. You are given the mortality rates for a certain type of electronic equipment.

week (i)	1	2	3	4	5
probability of a new piece failing during week (i)	p_1	p_2	p_3	p_4	p_5

where, $p_i > 0$ $\quad \sum_{1}^{5} p_i = 1$

We start with N new pieces. It is proposed to replace the piece as they fail and at the end of a fixed interval all the N pieces are group-replaced. At what intervals should the group replacement take place ? You are given c_1 the cost of individual replacement and c_2 the cost per item of group replacement. ($c_1 > c_2$).

Explain how this problem can be solved. (17)

Please turn over

4. (a) Explain the problem of Inventory control in the case of deterministic demand and derive the economic lot size formula in the case of constant demand rate R per unit time, with unit set up cost C_3, unit holding cost C_1, and unit shortage cost C_2.

(b) Explain how linear programming can be used in solving production scheduling problems. (10)

5. (a) For case of a queue with k service channels, each exponential, with mean service rate μ and Poisson arrivals with mean rate λ, derive the recurrence relations for the steady state probabilities p_n, the probability that there are n units in the system at any instant of time. Solve for p_n.

(b) For the case of a single channel k = 1 (in (a) above), derive the distribution of the waiting time of an arrival.

(c) Define Erlangian service time distribution. What is its usefulness and importance in queueing problems ? (6+5+6)

For NEATNESS (2)

Group B : I (alternative) - Elements of Book-keeping & Accountancy

i) Use separate answer book for this group.

ii) Answer any four questions of which question no. 1 is compulsory.

1. (Compulsory) P and Q are partners in a Firm sharing Profits & Losses as P 80% and Q 20%. Their Trial Balance as on 31st December 1963 is given below.

	Dr.	Cr.
Closing Stock	30,000	
Wages	12,000	
Bad Debts	300	
Materials consumed	70,000	
Bills Payable		10,500
Investments	15,000	
Discounts	350	600
Bills Receivable	20,000	
Carriage Inward	7,500	
Carriage Outward	2,700	
Return Inward	1,500	
Debtors & Creditors	60,000	25,000
P-Capital Account		50,000
Q-Capital Account		40,000
Plant & Machinery	59,000	
Interest on Overdraft	3,500	
Commission		450
Bad Debt Reserve		1,500
Bank Overdraft		20,000
Cash in hand	950	
Sales		1,57,250
Salary	17,500	
P-Drawings	3,000	
Q-Drawings	2,000	
Rs.	3,05,300	Rs. 3,05,300

You are required to prepare the Trading and Profit & Loss Account for the year ended 31st December 1963 and the Balance Sheet as at

Please turn over

that date, after taking into account the following adjustments :

a) Bad Debt Reserve to be increased to Rs. 3,000.

b) Provision for Discount on Creditors to be made 5% on Creditors.

c) Interest on Capital is to be payable at 5% p.a.

d) Plant & Machinery is to be depreciated by 10%.

e) Salary account includes the following :- Rent 500
Stationery 250

f) Audit fee unpaid Rs. 1,000. (22)

2. Enter the following Cash and Bank transactions in a Single Column Cash Book, where an Account with the Bank, is kept in the Ledger :-

			Rs. nP
September	1	Balance of Cash	250.00
		Balance at Bank	3,750.00
"	3	Cash Sales	190.00
"	5	Deposited into Bank from Office Cash	120.00
"	7	Received from M. Chatterji in Cash	170.00
"	9	Bank Interest credited	5.00
"	11	Paid Roy Bros. by cheque	250.00
"	13	Paid carriage	10.00
"	15	Paid salary by cheque	250.00
"	17	Received from Mukherjee Bros. by cheque and deposited same into Bank	250.00
"	19	Withdrew from Bank for Office Cash	500.00
"	21	Paid wages in Cash	75.00
"	23	Cash Sales	600.00
"	25	Deposited cash into Bank	500.00
"	27	Sold goods and received cheque which is deposited into Bank	100.00
"	30	Purchased goods by cheque	190.00

(14)

3. On examining the Bank Pass Book of X Co. Ltd., it is found that the balance shown on 31st March 1955 the close of the Company's financial year, differs from the Bank balance Rs. 23,650 shown by the Cash Book on that date. From a detailed comparison of the entries it is found that :-

a) Rs. 2,060 is entered on the Cash Book as paid into Bank on 31st March 1955, but not credited by the Bank until the following day.

Please turn over

b) Bank charges Rs. 70 on 31st March 1955, are not entered in the Cash Book.

c) A bill for Rs. 5,500 discounted with the Bank is entered in the Cash Book without recording the discount charge of Rs. 270.

d) Cheques totalling Rs. 16,720 were issued by the Company and duly recorded in the Cash Book before 31st March, 1955, but had not been presented at the Bank for payment until after that date.

e) On 25th March 1955, a debtor paid Rs. 1,000 into the Company's Bank in settlement of his Account in the Company's Sales Ledger, but no entry was made in the Cash Book of the Company in respect of this.

f) No entry has been made in the Cash Book to record the dishonour on 15th March 1955, of a Cheque for Rs. 550 received from B. Chatak.

You are asked to show the necessary corrections in the Cash Book of X Co. Ltd. and to prepare a statement reconciling the Cash Book Balance as amended with the Balance which should appear in the Bank Pass Book. (14)

4. In taking out a Trial Balance, a Book-keeper finds that he is out by Rs. 41-10-8 as excess debit. Being desirous of closing his books, he places the difference to a newly opened Suspense Account which is carried forward. In the next period he discovers that (a) a credit item of Rs. 83-9-11 has been debited to a Personal Account as Rs. 38-11-9; (b) a sum of Rs. 62-10-9 written off Fixtures as depreciation, has not been posted to Depreciation Account; (c) Rs. 1,000 paid for Furniture purchased had been charged to the ordinary Purchase Account; (d) a discount received of Rs.14-5-6 (e) the total of the Inward Returns has been added Rupee One short; and (f) an item of Sale for Rs. 68 was posted as 86 in the sales Account.

Give the correcting journal entries and prepare the Suspense Account. (14)

5. (a) Why Stock-in-trade is being included in Financial Accounts ?

(b) In what way correct valuation of Closing Stock has direct bearing upon the correct ascertainment of the net profit of a trading concern ?

(c) What should be the correct method of valuation of Stock-in-trade ? (3+5+6)

6. Write notes on :-

a) Contingent Liability.

b) Goodwill.

c) Accommodation Bill.

d) Work-in-Progress. (14)

Please turn over

11. Statistical Methods in Business

i) Use separate answer book for this group.

ii) Answer any two questions.

1. Explain the six major steps involved in setting up and control of a job. Explain the terms, job standardisation, job review and analysis, job evaluation. (15)

2. (a) Suppose you are given data regarding monthly sales for a manufacturing company for the years 1953 to 1963. How will you use this data for forecasting the sales in the year 1964 for each month January to December ?

(b) The market demand for electric refrigerators in the U.S. has been analysed in the following equation

$$Y = 2312.8 + 34.4\ x_1 + 35.6\ x_2 + 2024.3\ x_3$$

y = number (in thousands) of refrigerators sold.

x_1 = disposable personal income in 1939 dollars.

x_2 = change in disposable personal income from the previous year.

x_3 = time (1925 = 1)

Explain the meaning of the terms and variables used, how this equation has been set up and what type of data has been used. (7+8)

3. (a) What are the important uses of sampling in Economics and Business ?

(b) Explain the term "Market research". What are the various types of marketing research activities ?

(c) Explain the term "consumer-panel" and its use in market research. (5+6+4)

oooOooo

Group D - Design & Analysis of Experiments (Combinatorial Aspects)

(Answer any four questions)

1. Prove that where p is a prime number, the class of residues modulo p, form a finite field. Give the addition and multiplication table for a finite field with 5 elements.

Explain how you will construct a finite field of order p^n, where p is a prime number and n a positive integer. Construct a field with 9 elements.

Mention some applications of finite fields in Statistics. (25)

2. Define a balanced incomplete block design (BIBD) and obtain the relations and inequalities amongst its parameters. Give five positive numbers v, b, r, k and λ satisfying these relations such that no BIBD with those parameters exists. Substantiate your answer with a proof.

Show how a BIBD with r = k can be used to obtain two more BIBD's. State whether this process is reversible. (25)

Please turn over

3. Define an orthogonal array of strength 2 and indicate the relation between such an array of index unity and a set of mutually orthogonal Latin Squares. Prove that the existence of an array $s^2, s, 1, 2$ implies the existence of $s - 1$ mutually orthogonal Latin Squares of order s. (25)

4. Define intra block sub-group in a confounded symmetric factorial experiment. Obtain the intra block sub group in a 2^6 experiment with factors A, B, C, D, E and F, confounding the 3-factor interactions ABC, BDE and CDF. Develop the intra block to obtain the complete design. What other interactions are confounded in this design. (25)

5. What is generalised interaction. Construct two replicates of a 3^3 design involving 3 factors A, B, C at three levels each, confounding the effects ABC, and BC^2, in the first replicate and AB and ABC in the second replicate. (25)

6. Define a rotatable design and state the necessary conditions that a second order rotatable design has to satisfy. Construct a rotatable design for a 3^{10} experiment, using the balanced incomplete block design with parameters.

$$v = 10, \quad b = 15, \quad r = 6, \quad k = 4, \text{ and } \lambda = 2.$$ (25)

Group E - Sample Surveys (Organizational Aspects)

(Answer any four questions)

1. Discuss the questions that require consideration at the planning stage of sample surveys.

   How will you resolve these questions if you are asked to prepare a design for a repetitive sample survey for construction of annual indices of unemployment in a big city of India. (25)

2. You are required to prepare a budget for a sample survey for estimating annual production of major crops of your State. Enumerate the different stages of work and estimate the work-load of each stage. Then prepare the budget-estimate showing the details of recurring and non-recurring items of expenditure. (25)

3. (a) Write a note on the supervisory functions relating to different stages of work in connection with a statistical surveys. (15)

   (b) How will you control the quality of primary data collected by the field-staff ? (10)

Please turn over

4. (a) Describe the various operations generally involved in the processing of data. (19)

(b) How will you control the quality of the work of processing of data ? (6)

5. (a) Discuss the various uses of interpenetrating sub-samples. (13)

(b) Comment on the following :- (6+6).

i) The census house list cannot be used as a frame for surveys of human population, as a family is sometimes found to have occupied more than one house in the list.

ii) The smaller the reference period, the better is the estimate of the incidence of unemployment.

6. Write notes on any three of the following :- (25)

a) Role of public relations and propaganda in statistical surveys.

b) Schedules and Questionnaires.

c) Coding.

d) Master-sample.

e) Report-writing.

oooOooo

Group F : Techniques of Computation (Practical)

i) Time allowed five hours for this practical group.

ii) Answer any four questions.

iii) Use of calculating machines is permitted.

1. A function $f(x)$ and its derivative $f'(x)$ are tabulated below :

x	$f(x)$	$f'(x)$
1.0	0.24197	0.24197
1.2	0.19419	0.23302
1.4	0.14973	0.20962
1.6	0.11092	0.17747

Use the full data to evaluate $f(1.1)$ and compare this with the value that is obtained when the information on $f'(x)$ is ignored. (25)

Please turn over

2. Solve numerically the differential equation

$$\frac{dy}{dx} = x - y^2$$

and tabulate y (correct to five places of decimals) as a function of x, for $x = 0.0(0.1)0.4$.

It is known that $y = 0$ when $x = 0$ (25)

3. Find to five significant figures all the roots of the equation

$$x^3 - x - 4 = 0 \qquad (25)$$

4. Evaluate

$$I = \int_1^2 \frac{dx}{1 + x^2}$$

using Simpson's one-third rule with

(a) $h = 1/8$

(b) $h = 1/10$

(c) Show that if E_a and E_b are the errors in evaluations I(a) and I(b) as obtained in 4(a) and 4(b) respectively

E_a / E_b is expected to be nearly equal to $(\frac{1}{8})^4 / (\frac{1}{10})^4$

(d) Hence show that

$$\frac{E_a - E_b}{E_b} \approx \frac{I(a) - I(b)}{E_b} \approx \frac{(\frac{1}{8})^4 - (\frac{1}{10})^4}{(\frac{1}{10})^4}$$

(e) Use this result to estimate E_b and correct your evaluation I(b). (25)

5. For the matrix

$$D = \begin{pmatrix} 18.75 & 8.48 & 6.84 & 30.30 \\ 8.48 & 29.04 & 8.78 & 44.10 \\ 6.84 & 8.78 & 28.86 & 36.29 \\ 30.30 & 44.10 & 36.29 & 126.92 \end{pmatrix}$$

a) Obtain a matrix B such that

$B D B' = I$ (the unit matrix)

b) Find the inverse of B. (25)

6. For the matrix D in question 5, obtain the latent root with the maximum absolute value and a latent vector corresponding to this root. (25)

oooOooo

Please turn over

Group G - Statistical Inference

(Answer any four questions)

1. Describe the Sequential Probability Ratio Test (SPRT) procedure for testing a simple hypothesis against a simple alternative. (10)

Obtain a suitable approximation to the Operating Characteristic function of the SPRT procedure in such a case. (10)

What are the good features of the SPRT procedure ? Is it, in any sense, better than other sequential test procedures ? (5)

2. Describe Stein's two sample procedure for obtaining a confidence interval (of fixed confidence coefficient and fixed length) for the mean of a Normal population with unknown variance.

Prove that the procedure is valid. (25)

3. Explain the term 'statistical tolerance limits' . (5)

Calculate the probability that the maximum and the minimum in a sample of n from a continuous population, will cover atleast a proportion x of the population. (10)

Calculate the expectation and the variance of the proportion covered by the sample range. (10)

4. Describe the Wald-Wolfowitz Run test for determining whether two samples are from the same continuous population. (10)

What are the good properties, if any, of this test ? (5)

Work out the asymptotic mean of the test statistic, under the assumption that the samples are from the same population. (10)

5. Obtain the likelihood ratio test for determining, on the basis of a sample of size n drawn from a multivariate Normal population, whether specified groups of variates are mutually independent.

Write down, without proof, the asymptotic distribution, of a suitable transformation of the likelihood ratio criterion, under the null hypothesis being tested. (25)

ooooOoooo

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - May 1964
Paper X - Subjects of Specialisation - III

Time : 5 hours Full marks :100

i) Candidates will be required to answer questions from that group only for which they have registered their options.

ii) Figures in the margin indicate full marks.
iii) Use of calculating machines is permitted.

Group A : Economic Statistics

(Attempt any three questions)

1. Given the economy with the following activities :

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
Commodity 1	+ 1.0	+ .8	+ 1.0	- .1	- .5	- .8	- .2
Commodity 2	- .1	- .2	- .7	+ 1.0	+ 1.0	+ 1.0	+ 1.0
Labour	-12.5	- 6	- 5	- 3.3	- 5	- 4	- 10
Capital	- 1.5	- 1.6	- .3	- 2.0	- 1.5	- .2	- 1.4

a) Assume :

i) the availability of 1500 units of labour,

ii) that each commodity has a value of 1.0.

Determine (algebraically or graphically) the minimum amount of capital needed to produce a net output valued at 100.

b) What are the returns to capital and labour in part (a)?

Note : (+) signs indicate outputs and (-) signs indicate inputs. (32)

2. You are given 23 annual observations on production, employment, Capital, and time trend for the U.S. 1929-41, 1946-55 . Estimate a macro-economic production function of the form :

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \epsilon_t$$

where

Y = real GNP, in billions of dollars

$X_1 = 1$

X_2 = employment, in millions of persons

X_3 = real capital, in billions of dollars

X_4 = time in years, measured from 1929 = 1 ;

this acts as a proxy for smooth productivity trend.

ϵ = random variable, normally distributed with zero mean, constant variance, independent time-wise, and independent of the X's

Please Turn Over

The observations are presented in matrix form: $X = X_{ti}$ and $Y = (Y_t)$. Thus in 1932, for example, $X_1 = 1$, $X_2 = 34$, $X_3 = 40$, $X_4 = 4$ and $Y = 98$.

$X =$

X_{1t}	X_{2t}	X_{3t}	X_{4t}
1	47	54	1
1	43	59	2
1	39	57	3
1	34	48	4
1	34	36	5
1	36	24	6
1	38	19	7
1	41	18	8
1	42	22	9
1	37	24	10
1	40	23	11
1	42	27	12
1	47	36	13
1	51	9	18
1	53	25	19
1	53	39	20
1	50	51	21
1	52	62	22
1	54	75	23
1	54	94	24
1	55	108	25
1	52	118	26
1	54	124	27

$Y =$

Y_t
142
127
118
98
94
102
116
128
140
131
143
157
182
209
214
225
221
243
257
265
276
271
291

Write out the estimated production relationship with sampling errors in parentheses below the respective coefficients. Indicate how you would test the hypotheses:

a) that there are constant returns to scale, and

b) that the time rate of growth of productivity per annum is some pre-assigned value.

[32]

3. Assume the following unprecedented model for the economy: Consumption is a linear function of current national Income, Investment is a linear function of current and one period lagged national income, and, by definition, national income equals the sum of investment, consumption, and government expenditures. Government expenditures are exogenous, while the consumption and investment equations are subject to temporally independent errors of behaviour.

i) Express the model in symbols.

ii) Using the rank (necessary and sufficient) conditions, show that the consumption and investment relations are identified for all values of the structural coefficients.

iii) Using the following sample observations for the above model,

Please Turn Over

estimate the coefficients of the consumption relation by the limited information method.

year	consumption	national income	Government expenditure
0	7	10	2
1	8	11	2
2	9	13	3
3	6	9	3
4	4	7	2

(5)

Note: You should use all the predetermined variables in the system while estimating the consumption relation in (2.11).

4. From the data given in the table below :

Food Expenditure by Income Class and Family Size

Income class(₹)	single member	two member family	three member family	four member family	five or more member family
under-1000	349	614	1587	1027	360
1000- 1999	577	730	898	867	1082
2000- 2999	809	944	1077	1240	1380
3000- 3999	820	1098	1266	1315	1560
4000- 4999	1400	1287	1450	1533	1770
5000- 5999	1276	1511	1602	1668	2007
6000- 7499	1660	1487	1790	1910	2050
7500 and over	2100	2250	2320	2425	2968

estimate the relation :

$$\log E_f = \alpha + \beta \log Y + \gamma \log N + \mu$$

where :

E_f = expenditure on food,

Y = total income,

N = number of person in family,

μ = error term with usual assumption.

Comment on your results. (32)

For Neatness (4)

oooOooo

Please turn over

Group B. Techno-Commercial Statistics - Statistical Quality Control, Operations Research/Elements of Book-keeping Accountancy, Statistical Methods in Business.

(Use separate answer books for each of the sub-groups of Group B)

I. Statistical Quality Control

(Answer any two questions) from this sub-group)

1.a) Data were collected for the rejections that can be seen at the knock-off stage for castings. The number of bad castings obtained out of 100 inspected for each cast, are noted below.

9	14	7	4	6	5	2	1
11	7	4	7	0	6	4	4
4	4	5	6	7	7	0	
8	0	2	3	13	3	5	

Examine the data with a view to set-up a control chart. [10]

b) After the chart was set up. The process was observed to be well under control for a long period with an average of 4 percent and it was decided to reduce inspection for control. A sample of size 25 was decided upon. Obtain the control limit for the revised chart. At a later date the examination of the control chart gave the following number of defectives per sample over the last 20 samples.

0,	0,	2,	3,	0,	2,	1,	1,	1,	0
0,	1,	2,	1,	1,	2,	4,	4,	3,	3

The quality control-in-charge, decided on the basis of the data that the process has to be checked. Do you agree? [15]

2. An experiment has been carried out on an assembly operation to determine the best method by measuring the time to perform the operations. Altogether 5 objects, A, B, C, D and E were assembled by 3 methods X, Y and Z on each of 3 days, Wednesday, Thursday and Friday, giving a total of 45 observations. The actual average times for 10 trials were:

		A	B	C	D	E
	Wednesday	17	13	17	19	25
X	Thursday	16	13	17	18	24
	Friday	16	13	17	18	23
	Wednesday	15	16	16	21	22
Y	Thursday	14	15	16	19	21
	Friday	15	15	17	19	20
	Wednesday	17	16	17	18	22
Z	Thursday	17	13	16	17	22
	Friday	17	14	16	17	21

Make a suitable analysis of the data. Suggest which method of assembly is suitable for which type of object. [25]

3. Select n and k' for a known-sigma variables sampling inspection plan for one-sided specification so that the O.C. curve of the plan will be similar to that of the single sampling attributed plan n = 15, c = 2 with this plan what will be the probability of acceptance of a 5 percent defective lot? [20 + 5]

Please Turn Over

II. Operations Research

(Answer any two questions from this sub-group)

1. Table below gives the quantity of resources required for unit amount of each product and the net profit per unit, on each product. The units are not specified here.

resources	products			available resources
	potato	corn	beans	
Land	1	1	1	120
Labour	2	3	5	600
Capital	99	38	20	2000
Net profit per unit amount of the product	83	72	27	

From the data noted above, find the best allocation of land, labour and capital for a farming enterprise. [15]

2. Solve the following transportation problem, given the following cost matrix, supplies and demands.

origin \ destination →	A	B	C	D	E	supplies
↓ O_1	10	6	14	6	16	6
O_2	10	12	24	10	14	8
O_3	4	16	6	8	16	4
O_4	11	8	12	7	12	16
demands	6	6	12	4	6	

[15]

3.a) A newspaper boy buys papers for 10 nayapaise each and sells them at 14 nayapaise. He cannot return unsold newspapers. Daily demand has the following distribution

no. of customers	95	96	97	98	99	100
probability	.10	.15	.25	.27	.14	.09

How many papers should the boy order each day to maximise his average profit? [10]

b) A book-binder has one printing press, one binding machine and the manuscripts of 5 books. The time required to perform the printing and binding operations for each book are given below.

book	time taken (in minutes) for printing	binding
1	30	80
2	120	100
3	50	90
4	20	60
5	90	30

Please Turn Over

Contd. II. - 3(b)

Determine the order in which the books have to be processed in order to minimise the total time required to turn out the books. [5]

II(alternative) : Elements of Book Keepings and Accountancy

(Answer both the questions from this sub-group)

1. EITHER

C. starts a business on 1st February 1960. His transactions for the month of February are given below. You are asked to record them in suitable books of accounts and to extract a trial balance as on 29th February 1960.

1st February : Paid in Rs.5,000 and opened an account with the Bank of India Limited.

2nd February : Purchased Furniture for Rs.230 from Calcutta Furnishers ; Purchased goods by a Cheque for Rs.420.

3rd February : Withdrawn from Bank Rs.20 : Purchased goods from A. Traders Ltd. Rs.250 ; Sold goods to B. Company Rs.300 less 2 percent discount and paid Calcutta Furnishers Ltd. a Cheque for Rs.100 on a/c.

8th February : Purchased goods from S.Co., Ltd., Rs.1,175 less 2½ percent discount. Sold goods for cash deposited into Bank Rs.125 ; Purchased Stationery for cash Rs.15.10.

9th February : Withdrawn from Bank Rs.50 ; Paid by cheque A. Traders Ltd., in full settlement of their bill at a cash discount of 5 percent.

10th February : Purchased Machinery and Plant from Bengal Machinery Works Rs.1,500.

15th February : Paid by Cheque Bengal Machinery Works in full settlement of their dues ; Paid Calcutta Furnishers a cheque for Rs.125 in full settlement of their bill.

16th February : Purchased postage stamp Rs.10 and paid wages to Office Peon Rs.7.00

19th February : Sold goods to C. Ltd. Rs.925 less 2 percent. Sold goods for cash Rs.206 and deposited the same into Bank next morning.

23rd February : Received a Cheque for Rs.200 only on a/c. from B. Company.

25th February : Received a cheque from C. Ltd., in settlement of one bill at a cash discount of 3 percent.

27th February : Purchased goods from A. Traders Ltd. Rs.1,520.

28th February : Returned goods to A. Traders Ltd. Rs.47.00 Purchased stationery for cash Rs.19.00 Paid by a Cheque, a printing bill for Rs.150.00. [22]

Please Turn Over

1. OR

Pass necessary adjustment entries at the Closing date for the following items :-

1. Closing Stock Rs.50,000

2. Depreciation at 5 percent on Furniture at book value of Rs.10,000.

3. Outstanding Insurance Premium Rs.400.

4. Balance of Sundry Debtors' A/c. was Rs.85,000 and that of Reserve for Doubtful Debts A/c. was Rs.2,000, a sum of Rs.5000 is to be written off as bad. Reserve for Discount on Debtors and Reserve for Doubtful Debts to be provided at 5 percent each.

5. Subscriptions received in advance, amount to Rs.400.

6. Interest and Rent receivable outstanding to the extent of Rs.200 and Rs.100 respectively.

7. Balance of Sundry creditors a/c was Rs.50,000 Reserve for Discount on Creditor's to be provided at 5 percent per annum.

8. After the compilation of Profit and Loss a/c. there is a credit balance of Rs.30,000. The Managing Director is to receive a commission of 5 percent before changing such commission.

9. A is entitled to an Interest of 5 percent per annum on his Capital standing in the Ledger at Rs.10,000.

10. After the Profit and Loss a/c. has been prepared, it is found that there is a credit balance of Rs.21,000, on which the Managing Director is to receive a commission at 5 percent after changing such commission. [22]

2. EITHER

The Swadeshi Industries Ltd., removed their works to a more suitable premises, and given below are some of their transactions :-

a) A sum of Rs.4,750 was expended on dismantling, removing and reinstalling plant, machinery and fixtures.

b) The removal of stock from the old works to the new site, cost Rs.500.

c) Plant and machinery which stood in the books at Rs.75,000 included a machine at a book value of Rs.1,500. This being obsolete was sold off at Rs.500 and was replaced by a new machine costing Rs.2,400.

d) The freight and carriage on the new machine amounted to Rs.150, and erection charges cost Rs.275.

e) A sum of Rs.1,200 was spent on painting the new factory.

State with reasons which items of expenditure would be charged to capital and which to revenue. [8]

Please Turn Over

2. OR

Enter up the Returns Inward Book of the Carpet Suppliers Ltd. Open Ledger Accounts including Returns Inward Account and post the items:

August 2: Returned by Banerjee Stores Ltd. 1 Seamless Carpet 10 ft. by 9 ft. at Rs..70/-.

August 12: Returned by John and Cooper Ltd. 1 Heavy Turkey Carpet 12 ft. by 9 ft. at Rs.200/-. 1 Seamless Carpet (10 ft by 9 ft.)at Rs. 70/-.

August 19: Returned by Modern Furnishers $1\frac{2}{3}$ yd yds. plain Hair Carpeting at Rs.5/- a yd. [8]

III: Statistical methods in Business

(Answer any two questions)

1. A job was performed by 30 workmen using method I and by 40 workmen using method II with the following results.

Time (in minutes)	50	51	52	53	54	55	56	57	58	59	60	total
No. of workmen finishing the job by method I	1	3	5	4	7	5	3	1	1	0	0	30
No. of workmen finishing the job by method II	0	1	2	5	8	9	6	3	3	1	2	40

Give 95 percent confidence limits for the average saving in time that can be expected from the use of method I instead of method II. [10]

2. A study of the market for various commodities among 8000 readers of Collier's Magazine, revealed the following distribution of the sample households by size of household, as compared with corresponding Census Estimates for all U.S. households. Examine whether the Collier sample provides a representative picture of the size-of-household distribution of all U.S. households.

Table showing relative distributions of 8000 Collier families and all U.S. families by size of households.

persons in household	Collier sample	U.S. families (percent)
1	576	10.0
2	2368	29.8
3	1968	24.2
4	1592	18.0
5	800	10.0
6	376	4.5
7	152	1.7
8 or more	168	1.8
total	8000	100

[10]

Please Turn Over

3. The following table gives data regarding test scores made by salesmen on an intelligence test and their weekly sales.

salesman	test scores	sales (in hundreds of rupees
1	40	25
2	70	60
3	50	45
4	60	50
5	60	45
6	50	20
7	90	55
8	40	30
9	60	45
10	60	30

a) Examine whether the correlation between the test score and value of sales is significant.

b) Obtain the regression equation of the sales (y) on the test score (x).

c) Suppose a new salesman gets a test score 70. Estimate his weekly sales.

d) What is the standard error of the estimate in (c)?

[3 + 3 + 1 + 3]

Please turn over

Group D: Design and Analysis of Experiments

(Answer any three questions from this group)

1. A varietal trial on a certain crop involving three different varieties was arranged in a Latin Square. After the seedlings had appeared, it was decided to combine one more objective in the same investigation and to study simultaneously, the effect of a certain fertiliser (usually applied late) on the crop yield. Accordingly the experimental crop plots where the seedlings had already appeared, were divided into three parts each, and the fertilisers were applied at three levels (0, 2 and 4 cwt per acre). The assignment of the level to the part was done at random and independently, for each crop plot.

The plan and the yields are given below [the levels are indicated by (0), (1) and (2) respectively].

rows	columns 1		columns 2		columns 3	
1	(2) 118 (0) 100 (1) 105	V_3	(0) 111 (1) 130 (2) 157	V_1	(0) 117 (1) 114 (2) 161	V_2
2	(1) 108 (2) 126 (0) 70	V_2	(1) 124 (0) 96 (2) 121	V_3	(0) 61 (1) 91 (2) 97	V_1
3	(1) 90 (2) 100 (0) 62	V_1	(0) 80 (2) 94 (1) 82	V_2	(2) 109 (0) 63 (1) 70	V_3

Analyse the data to test for varietal differences. Do these varietal differences depend upon the level at which the fertiliser is applied? Is it possible to make some recommendations regarding the optimum level at which the fertiliser should be applied. [32]

2. An experiment was carried out to compare 9 strains of paddy, which are denoted by (i, j), $i = 1,2,3$; $j = 1,2,3$. The design adopted was an incomplete block design, in two replications, each consisting of three blocks with three plots per block. The table below gives the plan and the yield of grain in lb. per plot.

	Block 1	(3, 2) 11.3	(3, 3) 7.8	(3, 1) 7.5
Replication 1	Block 2	(2, 2) 12.0	(2, 3) 9.3	(2, 1) 8.3
	Block 3	(1, 2) 8.8	(1, 3) 9.6	(1, 1) 6.5
	Block 1	(2, 3) 6.3	(3, 3) 8.5	(1, 3) 8.0
Replication 2	Block 2	(1, 2) 13.8	(3, 2) 14.0	(2, 2) 13.3
	Block 3	(3, 1) 11.3	(1, 1) 8.9	(2, 1) 10.3

Please Turn Over

a) Write down the analysis of variance table (intra-block). [16]

b) Also find estimated variances for estimates of differences between any two treatment effects. [8]

c) It was discovered later that the 9 strains of paddy were actually obtained from a factorial structure, the numbers i and j being the levels of two factors A and B respectively. Split up the treatment sum of squares into components due to the main effects and the interactions and examine if the data indicate the presence of interactions. [8]

3. In a chemical experiment fourteen combinations of three factors were tried. Table below gives the yields and the levels x_1, x_2 and x_3 (on a suitable scale) for each combination.

x_1	x_2	x_3	yield	x_1	x_2	x_3	yield
-1	-1	1	45	$\sqrt{2}$	0	0	42
-1	1	-1	39	$-\sqrt{2}$	0	0	42
1	-1	-1	49	0	$\sqrt{2}$	0	43
1	1	1	43	0	$-\sqrt{2}$	0	48
0	0	0	43	0	0	$\sqrt{2}$	43
0	0	0	45	0	0	$-\sqrt{2}$	46
0	0	0	44	0	0	0	43

a) Fit a second degree polynomial regression equation to predict the yield, in terms of the levels of the factors. [14]

b) Write down the analysis of variance table writing down separately the expressions due to first and second degree terms. Give your comments on the analysis. [10]

c) Obtain the variance function to give the variance of the predicted yield at x_1, x_2, x_3. Have you any comments to offer? [5]

4.a) i) In a manurial experiment on groundnuts three factors are to be studied at four levels each. Interactions involving all the three factors are expected to be negligible, though it is thought desirable to provide for the possibility of confirming this from the results of the experiment. Past experience recommends blocks of sixteen plots each. Suggest an appropriate design using four replications. (Give full lay-out for any one replication and the key-blocks for the remaining ones). [12]

ii) How many replications would be required for a balanced scheme of partial confounding? [ ]

b) Write down three mutually orthogonal Latin Squares of order four indicating their method of construction. [12]

For Neatness. [ ]

oooOooo

Please turn over

Group E: Sample Surveys.

(Answer only four questions of which question No.1 is compulsory)

1. Study the items of information to be collected in the schedule given [Compulsory] below and prepare suitable scrutiny and tabulation programmes. [25]

Demographic and Activity Particulars of Household Members

1. serial no. of person	1	2	3	4	5	6
2. relation to head						
3. sex (m-1, f-2)						
4. age last birth day						
5. marital status (code)						
6. educational standard (code)						
7. whether worked for at least one day during the last 7 days (yes-1, no -2)						
8. if yes for (7), a) hours worked b) additional hours available for work						
9. if no for (7), whether had job or enterprise from which temporarily absent (yes-1, no-2)						

codes : marital status: never married-1, married-2, widowed-3, divorced-4, separated-5, educational standard: illiterate-1, literate but below primary-2, primary-3, middle-4, secondary-5, graduate and above in (i), agriculture-6, (ii) engineering-7, (iii) technology-8, (iv) medicine-9, (v) other subjects-0.

2. In a large scale socio-economic survey, a stratified uni-stage design is used where from each stratum a simple random sample of 10 villages is selected with replacement. The following table gives the number of households for each of the sample villages, for all the strata in a region.

stratum sr. no.	total no. of villages	number of households in sample villages									
		1	2	3	4	5	6	7	8	9	10
1	6411	43	84	90	0	10	44	0	124	13	0
2	4705	50	147	62	87	84	158	170	104	56	160
3	2558	228	262	110	232	139	178	334	0	63	220
4	14997	17	34	25	34	36	0	25	7	15	31

i) Estimate the total number of households and its sampling variance unbiasedly. [15]

ii) Examine whether there has been any gain due to the use of the strata, compared with unstratified simple random sampling with replacement. [10]

iii) Compare the efficiency of the present allocation with that of the optimum allocation. [5]

Please Turn Over

3. For estimating the total cultivated area in a region, a stratified unistage design is adopted, where from each stratum 2 independent sub-samples of 6 villages each are selected systematically, with random starts. In the following table are given the totals of geographical area and cultivated area for each of the 2 sub-samples in each of the 10 strata into which the region is divided.

stratum sr. no.	total no. of villages	total geographical area (in sq. miles)	total of sample villages: geographical area (in sq. miles) s.s.1	s.s.2	cultivated area (in acres) s.s.1	s.s.2
1	2044	2893	15.40	10.02	3641	2935
2	1304	2114	14.68	6.43	4633	3649
3	1265	2898	9.93	12.93	3050	5043
4	1252	2944	8.80	23.23	3490	4722
5	4264	6998	8.22	8.14	2619	3638
6	1598	2755	10.27	22.56	936	3652
7	810	3576	28.60	18.85	9596	6935
8	567	3281	46.23	30.95	9135	10024
9	500	3142	44.09	47.73	9772	12152
10	486	2779	16.04	15.81	8105	7690.
total	14090	33300	--	-	-	-

i) Estimate the total cultivated area in the region and its sampling variance unbiasedly. [10]

ii) Obtain an alternative estimate of the total cultivated area using the ratio method of estimation with the geographical area as the supplementary variate and compare its efficiency with that of the unbiased estimator. [10]

iii) Estimate (at least approximately) the bias if any, in the ratio estimator obtained in (ii), and correct the estimate in (ii) for its bias. [5]

4. Suppose it is required to estimate the total out-put of a population of 500 factories in a region with a margin of error of not more than 10 percent on either side of the true value with a confidence level of 95 percent. The coefficient of variation of the distribution of total output is known to be approximately 60 percent.

i) Assuming that the sample mean is approximately normally distributed, calculate the sample size required in case of simple random sampling without replacement. [12]

ii) Other things being the same, find out the required sample size if the number of factories in the population is 1000, 2500 or 5000 and comment briefly on the behaviour of the sample size. [8]

iii) Also find the increase in the sample size, if any, in case simple random sampling with replacement is used instead of simple random sampling without replacement.

5. To estimate the total number of persons in a region, a stratified two-stage design is used. In the first stage the villages are selected from each stratum with probability proportional to the previous census population with replacement, and at the second stage households are selected from each sample village in a linear systematic way with a random start. The sampling interval to be used in each village is so specified that the sample design becomes self-weighting with the constant multiplier 48000.

Please Turn Over

The following table gives the number of sample households and the total number of persons in the sample households for each sample village.

stratum	sample village	sample households number	sample households no.of persons	stratum	sample village	sample households number	sample households no.of persons
1	1	8	35	4	1	5	26
	2	7	40		2	10	30
	3	5	22		3	7	28
	4	6	32		4	8	29
2	1	5	16	5	1	1	6
	2	6	28		2	13	54
	3	2	8		3	0	0
	4	9	32		4	6	18
3	1	5	19	6	1	5	27
	2	9	35		2	4	20
	3	7	36		3	5	21
	4	6	32		4	11	47

i) Estimate the total population within the region and its sampling variance unbiasedly. [15]

ii) Also estimate unbiasedly the total number of households in the regions and its sampling variance [10]

Please turn over

Group C : Statistical Inference

(Answer any four questions from this group)

1. In a population under random mating the expected proportions of individuals with blood-groups O, A, B and AB are respectively r^2, p^2+2pr, q^2+2qr and $2pq$ where p, q and r are three 'gene frequencies' which characterise the population $(p+q+r = 1)$.

In a random sample of 400 individuals from a certain population, the blood group frequencies were :

blood group	O	A	B	AB
frequency	163	170	55	12

By approximate formulae, the following estimates were obtained :

$$p = 0.26450, \quad q = 0.09315, \quad r = 0.64235.$$

Use these as trial values and, by the method of scoring, obtain the maximum likelihood estimates of p, q and r (use only one iteration step). Also obtain the variances of your estimates. [10 + 7]

2. The following two samples were obtained from two independent normal distributions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$:

Sample 1 : 93, 89, 112, 8, 93, 11, 16, 32

Sample 2 : 87 56 47 33, 102, 17, 59, 87, 62, 54

a) If σ^2 is known to be 850, test the hypothesis $\mu_1 = \mu_2$ against the possibility of $\mu_1 > \mu_2$, a uniformly most powerful test. [6]

b) Perform the same test as in (a) if σ^2 is not known. [7]

c) Draw the power curve of the test in (a), for the whole range of possible values of $\mu_1 - \mu_2$. [12]

3. 25 students were selected at random from a population one by one and given a test. The scores, in the order in which the students were drawn, are shown below :

32, 54, 38, 44, 60, 50, 30, 43, 46, 41, 45, 46, 48

32, 40, 17, 30, 37, 37, 40, 35, 45, 42, 30, 40.

It is assumed that the scores come from a normal population whose variance is 87.

a) Use the sequential probability ratio test for the hypothesis that the mean population score is $\mu_0 = 40$, against an alternative mean value $\mu_1 = 44$, taking $\alpha = 0.10$ and $\beta = 0.30$. [1

b) If the whole sample is used for a fixed sample size test of the same null hypothesis, what would be the power of this test? [

Please Turn Over

c) It is known that if $f(x, \mu)$ denotes the probability density function of $N(\mu, \sigma^2)$ then, for any specified value μ, the unique solution of the equation (for h)

$$\int_{-\infty}^{\infty} \left[\frac{f(x, \mu_1)}{f(x, \mu_0)} \right]^h f(x, \mu)\, dx = 1$$

is given by $h = \dfrac{\mu_1 + \mu_0 - 2\mu}{\mu_1 - \mu_0}$.

Use this result to draw the OC curve of the test in (a). [10]

4.a) Weights of dry jute fibre (in tolas) were recorded for two samples of jute plants, each sample having been treated by a different chemical process:

Sample 1 : 0.50, 0.44, 0.10, 0.40, 0.35, 0.15, 0.25, 0.22, 0.24
0.25, 0.44, 0.29, 0.20, 0.13.

Sample 2 : 0.05, 0.14, 0.63, 0.39, 0.43, 0.45, 0.03, 0.44, 0.28
0.27, 1.00, 0.48, 0.02, 0.42, 0.65, 0.34.

Test if the two samples can be said to have come from the same population by using

i) the run test; [5]
ii) Wilcoxon's test [5]

b) Find the smallest size of sample which should be drawn from an unspecified population so that the probability is 95 percent that at least 90 percent values of the population lie between the minimum and maximum of the sample. [5]

c) The following are 15 pairs of observations on stature (in cms.) and weight (in lbs.):

(146, 83), (159, 92), (150, 70), (146, 74), (145, 62),
(158, 106), (153, 73), (157, 85), (162, 98), (152, 70),
(162, 95), (152, 70), (158, 91), (149, 80), (159, 76),
(158, 88), (152, 78), (144, 62), (155, 98), (156, 86).

Use (i) a parametric test, under specified assumptions, and [5]
(ii) a non-parametric test [5]
to examine the association between stature and weight.

5. The means of three biometric characters X_1, X_2 and X_3 and the matrix of pooled variances and covariances are shown below, for two groups of female locusts in two different phases of development:

Please Turn Over

Mean values

	X_1	X_2	X_3
phase 1 (n = 20)	25.00	7.81	10.77
phase 2 (n = 72)	28.35	7.41	10.75

Pooled variances and covariances (90 d.f.)

	X_1	X_2	X_3
X_1	4.7350	0.5622	1.4685
X_2		0.1431	0.2174
X_3			0.5702

a) Test if the population mean values in phase 1 can be taken to be 25, 7 and 10. [10]

b) Test if there is a difference in the corresponding mean values in the two phases. [10]

c) A new female locust is known either to be in phase 1 or in phase 2, and its measurements on the three characters are 26.05, 7.20 and 10.80. In which of the two phases would you place the new locust? [5]

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1964

Paper I - Official Statistics and Descriptive Statistics (Theoretical)

Time: 4 hours Full marks: 100

a) Figures in the margin indicate full marks.
b) Answer three questions from each group.
c) Use of calculating machines is not permitted.

GROUP A

1(a) Describe briefly the organizational set-up of the National Sample Survey (N.S.S). [4]

(b) Enumerate the subjects taken up for survey in the eighteenth round of the N.S.S., indicating their importance in the context of the present national emergency. [4]

(c) Discuss briefly the various aspects of the sample design of the eighteenth round of the N.S.S., with emphasis on the sampling frame, interpenetrating sub-samples, and self-weighting nature of the design for household enquiries. [8]

2.(a) Give a list of the sectors for which employment statistics are currently available in India. Give the names of the publications containing these statistics. [8]

(b) Indicate the limitations of the unemployment statistics provided by the Employment Exchanges, for purposes of trend study? [8]

3.(a) What is the importance of vital statistics? Give the sources of information on Indian vital statistics. [8]

(b) Critically examine the scope, coverage and limitations of these statistics and give your suggestions for improving their reliability. [8]

4.(a) What are the sources of statistics relating to motor transport (road)? What are the main heads under which these statistics are available? [10]

(b) What are the special publications that arise in the collection of statistics regarding movement of goods by road, in developing countries? [6]

Footnotes. [2]

GROUP B

5. Describe with suitable examples, the various types of column and bar charts used in graphical presentation of statistical data. [16]

6.(a) From a sample of n observations, the arithmetic mean and variance are calculated. It is then found that one of the values, x_1, is wrong, and should be replaced by x_1'. Show that the adjustment to the variance to correct this error is

$$\frac{1}{n} \sum (x_1' - x_1)\left(x_1' + x_1 - \frac{x_1' - x_1 + 2T}{n}\right),$$

where T is the total of the original results.

Hence or otherwise derive the formula for the variance, when there is no error but only an additional observation is included. [8]

Please Turn Over

6.(b) Prove the inequality Arithmetic Mean $\geq$ Geometric Mean $\geq$ Harmonic mean. [8]

7. Define correlation coefficient between two random variables x and y, and correlation ratio of y on x.

The variates x and y have their mean values zero and correlation zero and the same variance σ^2. Show that for a given α,

$$u = (x \cos\alpha + y \sin\alpha) \quad \text{and}$$
$$v = (x \sin\alpha - y \cos\alpha)$$

have the same variance σ^2 and zero correlation.

If the variances of x and y are different and are given by σ_x^2 and σ_y^2 respectively show that the correlation coefficient between the two variates u and v is given by

$$r_{uv} = \frac{\sigma_x^2 - \sigma_y^2}{[(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_x^2\sigma_y^2 \operatorname{cosec}^2 2\alpha]}$$

[16]

8.(a) Define the multiple correlation coefficient $R_{0.12}$ of the variable X_0 on variables X_1 and X_2 and the partial correlation coefficient $r_{02.1}$ between X_0 and X_2, eliminating X_1.

(b) Establish the relation

$$(1 - R_{0.12}^2) = (1 - r_{01}^2)(1 - r_{02.1}^2) \qquad [16]$$

Neatness. [2]

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - November 1964
Paper II : Probability Theory and Statistical Methods (Theoretical)

Time : 4 hours — Full marks 100

a) Figures in the margin indicate full marks
b) Answer any three questions from Group A and four from Group B.
c) Use of calculating machine is not permitted.

GROUP A

1. N letters are randomly placed in N correctly addressed envelopes. Prove that the probability that exactly k letters are placed in correct envelopes is given by

$$\frac{1}{k!} \sum_{j=0}^{N-k} (-1)^j \frac{1}{j!}, \quad k = 0, 1, \ldots N.$$

Show that the generating function of the frequency distribution of the random variable k is given by

$$P(s) = \sum_{r=0}^{N} (s-1)^r / r! \qquad [16]$$

2. Define conditional and marginal distributions.

A discrete variable X follows the Poisson law

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots .$$

If λ itself is a random variable having the density function

$$g(\lambda) = \frac{1}{\Gamma k} \lambda^{k-1} e^{-\lambda} \text{ where } 0 < \lambda < \infty$$

and $k > 0$; then find the unconditional distribution of X. [16]

3. If X_i's $i = 1, 2, \ldots, n$, are ~~xxxxxxxxxx~~ independent normal random variables $N(0, 1)$, find the distribution of $\sum_1^n X_i^2$.
Prove the additive property of two independent chi-squares and find the mean and variance of chi-square with n degrees of freedom. [16]

4.(a) State the Central Limit theorem in any form and indicate its uses.

(b) Obtain the normal approximation to the Binomial distribution. [16]

5.(a) Find the mean and variance of a Poisson random variable.

If X_1 and X_2 are two independent Poisson random variables with parameters λ_1 and λ_2, then find the condition on a_1 and a_2 so that $a_1 X_1 + a_2 X_2$ is also a Poisson random variable, where a_1 and a_2 are any real numbers.

(b) Find the distribution of the sample mean from a random sample of size two, from a Cauchy population. [16]

Contd.

GROUP B

6. Set up an analysis of variance table for a two-way classification with equal number of observations in a cell. How can you test for the interaction between these two ways of classification? [12]

7. What is the principle of likelihood ratio tests ?

Derive two well-known tests from this principle. [12]

8. Define an admissible test and estimate and a minimax test and estimate. Are the minimax tests in general unique? What is meant by a complete class of tests? [12]

9. Define an (i) unbiased test and (ii) a uniformly most powerful test. For a sample of n from the exponential population

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad \text{when } x \geq 0 \quad (\theta > 0)$$

$$= 0 \quad \text{otherwise,}$$

find the most powerful test of the hypothesis $\theta = 1$ against the alternative $\theta = 2$ and examine if it is uniformly most powerful for the alternatives $\theta > 1$. [12]

10. What is a sufficient statistic? Show by examples how the problems of estimation or testing of hypothesis are simplified when a sufficient statistic exists. [12]

Mentions [2]

Please turn over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1964

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

Time : 4 hours. Full marks : 100

a) Figures in the margin indicate full marks.
b) Answer three questions from each group.
c) Use of calculating machines is not permitted.

GROUP A

1.(a) Describe briefly any statewise large-scale sample survey for estimation of yield of crops, and indicate how that is an improvement over the old official procedure of annawari estimation. Give details regarding the design of the survey, information collected and statistical analysis of the data. [12]

(b) What are the various aspects that are usually studied at the pilot stage before plans for a statewise yield survey on any crop, are finalized ? [4]

2.(a) Discuss the efficiency of systematic sampling as compared to simple random sampling. [8]

(b) An area contains three compact communities, viz., Hindus, Muslims and Christians. There is an up-to-date directory in which the persons in a house are listed in the following order: husband, wife, children (by age), others. Houses are listed in order along streets. The average number of persons per house is five.

The choice is between a systematic sample of every fifth person in the directory and a 20 percent simple random sample. For which of the following variables do you expect the systematic sample to be more precise? (a) population of people of Hindu descent, (b) proportion of males, (c) proportion of children. Give reasons. [8]

3.(a) What are the principal steps involved in the choice of a sample size in a sample survey ? [6]

(b) A survey is to be made of the prevalence of the common diseases in a large population. For any disease that affects at least 1 percent of the individuals in the population, it is desired to estimate the total number of cases, with a coefficient of variation of not more than 20 percent. (i) What size of simple random sample is needed, assuming that the presence of the disease can be recognized without mistakes ? (ii) What size is needed if total cases are wanted separately for males and females, with the same precision ? [10]

4. Write critical notes on any two of the following :-

(a) Non-sampling errors.
(b) Ratio method of estimation
(c) Sampling without replacement
(d) Variance function in sample surveys [16]

Neatness [2]

Please Turn Over

5. An experiment was carried out on 20 individuals to test if there are significant differences in the diastolic blood pressures of the right and left arm. Two consecutive observations, one on each arm, were to be made on each individual and since the order of measurement was reported to have a significant influence on its numerical value, the individuals were randomly divided into two groups of equal size. In group I the left arm measurements were taken first followed by measurement on the right arm, while the reverse order was used for individuals in group II. Values recorded were the differences in first and second measurement in each case (altogether twenty in number).

a) Describe the appropriate analysis for this experiment to test if the difference in blood pressure between the two arms is significant. Also suggest a suitable estimate for the average difference. [6]

b) Show how the same data could be used to test if the 'order' effect is actually present. [4]

c) Since the blood pressure is known to be dependent on age, it is felt desirable that the individuals be first stratified into relatively homogeneous group strata in respect of age, and the experiment repeated separately and independently within each stratum. Describe the modifications in analysis that will be required. [6]

6.(a) Describe the Fisher-Yates procedure for obtaining a random layout for a 5×5 Latin Square Experiment. [8]

(b) Write down the analysis of variance for such an experiment. [8]

7.(a) Explain clearly the principle of 'confounding' both 'total' and 'partial', in factorial experiments. [5]

(b) Given below is the plan for a 3^2 factorial experiment in two replications, involving blocks of size 3.

Replication 1

Block	1	(00)	(11)	(22)
	2	(10)	(21)	(02)
	3	(20)	(01)	(12)

Replication 2

Block	4	(00)	(21)	(12)
	5	(10)	(01)	(22)
	6	(20)	(11)	(02)

(i) Identify the effects confounded in each replication. [5]

(ii) Indicate the analysis for this experiment. [5]

8. Discuss any two of the following methods for increasing accuracy of an experiment. [8+8]

a) Proper choice of treatments and experimental material.

b) Use of concomitant observations.

c) Planned grouping of experimental units.

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1964

Paper IV : Applied Statistics (Theoretical)

Time: 4 hours. Full marks: 100

a) Candidates are required to answer questions from the two groups of subjects for which they have registered their option.

b) Separate answer books are to be used for each of the two groups attempted.

c) Figures in the margin indicate full marks.

d) Use of calculating machines is not permitted.

GROUP A - ECONOMIC STATISTICS

(Answer any three questions from this group, for neatness 2 marks only)

1. How will you calculate the index number of prices received from sale of agricultural products, and of prices paid for purchasing materials required for agriculture, by farmers? What changes in the prices of agricultural products are necessary so as to maintain parity between prices received and prices paid by farmers?

[10 + 6]

2. What do you mean by the trend of an economic time series ? How will you obtain this component? Is it always possible to estimate accurately the trend for the future?

[3 + 10 + 3]

3. What is a production function? For the Cobb-Douglous production function, derive the productivities of the factors of production, and the law of return. What type of data are required for estimating the parameters of this function?

[3 + 8 + 5]

4. What determining variables will you include to the estimate the demand function for cotton textiles in India, on the basis of time series data? How will you obtain the function and the elasticities of demand with respect to the determining variables?

[4 + 12]

5. Write notes on any two of the following :

i) net national product, by industrial origin

ii) Lorenz curve

iii) input-output table

iv) statistics of industrial employment and wages in India.

[8 + 8]

Neatness [2]

Please turn over

GROUP B - STATISTICAL QUALITY CONTROL

(Answer any three questions from this group, for maximum 2 marks only)

1. The specification limits on a measurable characteristic are 100 ± 30 units. The process is such that in the course of production, the process mean is likely to shift, but the process variability remains almost stable. 10 samples of size 5 each, taken at random from the process have given the following values of sample range,

9.8, 4.5, 6.7, 11.4, 12.3, 3.6, 8.4, 10.4, 7.6 and 12.9

a) estimate the process capability [4]

b) find out the interval within which the process mean can shift, such that the proportion of items outside specification is ≤ 0.0027. [5]

Technical action for resetting the process level is quite expensive. Hence the management would prefer to leave the process alone as long as the process mean is such that the proportion of items out of specification does not exceed 0.0027.

c) explain how to set up a control chart for $\bar{X}$ $(n = 4)$ in such a case [7]

2. A process is under statistical control with $\mu = 1000$ and $\sigma = 40$ with respect to an item measurement. An $\bar{X}$ chart for the item measurement is being maintained with control limits, drawn at

$$\mu \pm 3\sigma/\sqrt{n} \quad \text{where } n, \text{ the sample size is 4.}$$

The Average Run Length (A.R.L) of the control chart is defined to be the average number of sample points to be plotted on the control chart, before a point crosses either of the control limits. Thus for instance, if the very first point plotted violated the limits, the run length is 1 and if the second plotted point violates the limits, the run length is 2 etc.

a) Find A. R. L. when $\mu = 1000$ and $\sigma = 40$. [5]

b) Find A.R.L. when μ increases by 2 percent of its previous value, while σ remains the same. (modified value of $\mu = 1020$; σ remains at 40). [5]

c) Find A.R.L. when μ remains the same as in (a) but σ increases by 10 percent of its previous value in (a). i.e. μ remains at 1000, modified value of $\sigma = 44$. [6]

3.a) It is desired to determine the O.C curve for the sampling plan $n = 50$, $c = 1$ and lot size $N = 600$. Give the detailed formula and insert all the constants (the numerical answer is not wanted) for determining :

i) the exact value of the probability of acceptance, for a lot of quality of $p = 0.10$ [4]

ii) an approximation to (i) using Binomial [3]

iii) an approximation to (i) using Poisson [3]

b) Explain the following terms in respect of acceptance sampling plans

i) AQL (ii) AOQL (iii) Tightened inspection. [2 + 2 +

Please turn over

4. Specifications require that a certain quality characteristic of a manufactured product have a minimum value of 200 units. This characteristic can only be tested by a destructive test. The product is made in batches of several thousand. The past practice regarding acceptance inspection has been to test 4 articles from each lot. If all the 4 articles met the quality specification of 200, the batch was accepted. If 2 or more failed, the batch was rejected. If 1 failed, a second sample of 4 was taken. With no failures in the Second Sample, the batch was accepted, otherwise rejected.

a) What is the probability that a batch containing 5 percent defectives will be accepted by this procedure ? [7]

b) Control charts for $\bar{X}$, R have been plotted from the first samples. These charts indicate that range stays in statistical control, even though the average shifts from batch to batch. The estimate of standard deviation is 10 units. The suggestion is made that acceptance decision be based on the average value computed from a Single Sample of 4, with batch acceptance if the average is 210 or more. Assuming that the standard deviation continues to be at 10 units and assuming a normal distribution of the quality characteristic within the batch, what is the probability of acceptance of a 5 percent defective lot? For what lot quality will the probability of acceptance be 0.10 ? [9]

5. Write short notes on

a) group control charts

b) control by gauging

c) salient features of Mil. Std. 105 A plans. [5 + 5 + 6]

Nor bicas [2]

GROUP C - STATISTICAL METHODS IN GENETICS

(Answer any three questions from this group, for neatness 2 marks only)

1. Under simple dominance and linkage, work out the theoretical frequencies of four phenotypes AB, Ab, aB and ab, in the F_2 generation, assuming that each character A/B is controlled by a diallele system.

Obtain the likelihood equation for estimating the linkage ratio, on the basis of such data. [16]

2. Write a short note on the O - A - B system of classification of human blood and the mode of its inheritance. Describe how blood-typing is useful in

i) blood transfusion
ii) cases of doubtful paternity. [16]

3. If a rare malady is a recessive character, controlled genetically by sex linked dialleles under simple dominance, show that the males (heterogametic sex) are at a disadvantage in respect of inheritance of the malady.

Show that under pan-mixia, the proportion of females having the malady equals the square of the proportion of males having the malady. [16]

Please turn over

4. Consider a breeding programme in which all recessives in each generation are prevented from propagation, and pan-mixia holds in the residual population. If the character is controlled by a single pair of genes and r_n denotes the frequency of the recessive gene in the n-th generation, show that

$$r_{n+1} = \frac{r_n}{1 + r_n} \cdot \qquad [16]$$

Neatness [2]

GROUP D - VITAL STATISTICS AND DEMOGRAPHY

(Answer any three questions from this group, for neatness 2 marks only)

1.a) Explain what is meant by a 'Stable population'. [5]

b) Show how the momentary rate of increase of a stable population can be expressed approximately in terms of the average length of a generation and the net reproduction rate. [10]

2.a) Describe the principal sources of demographic data in India, and their limitations and relative importance in demographic studies. [6]

b) Explain briefly the method used by the census actuary in India for constructing Indian Life tables, and also why the method differs from the normal method used, say, in England and Wales. [10]

3.a) Define a_x as a life-table function and show how, in an actual population, an approximate value of a_x can be obtained by relating the deaths at age x last birth day in a period of n years to n times the mean population at age x, last birth day in that period. [2 + 4]

b) Criticise the adoption of the infant mortality rate as an approximation to the rate of mortality in the first year of life (q_0). [4]

c) What are the implications of the infant mortality rate and what considerations should be borne in mind while making comparisons with this index ? [5]

4.a) Examine the defects of the 'crude marriage rate' (defined by the ratio of the total number of marriages in a given year to the mean total population of the year) and obtain the correct definition of the 'marriage rate' for the purpose of comparing marriage experiences. [...]

b) Explain why,

i) with increasing age, the 'proportion married rises to a peak value and then falls, being effected at the older ages chiefly by mortality. [...]

ii) the 'proportion ever-married' rises with increasing age but tends to be constant at older ages, unaffected by mortality. [...]

iii) marriage rates are subject to wider secular fluctuations than proportions married. [...]

Please turn over

5. Write notes on <u>any two</u> of the following: [8+8]

a) Differential fertility.

b) Health Surveys and their importance and limitations.

c) Stability of the age-structure of the Indian population.

<u>Neatness</u> [2]

GROUP E - <u>EDUCATIONAL AND PSYCHOLOGICAL STATISTICS</u>

(Answer <u>any three</u> questions from this group, <u>for neatness 2 marks only</u>)

1. Write a critical note on Wilk's $L_{s/e}$ test. Discuss its importance and use in the theory of construction of psychological tests. [15]

2. Obtain the matrix equation for estimating factors in the following form

$$\bar{F} = M' \bar{R}^{-1} Z$$

where $\bar{F}$ is the Column Vector of all the factor estimates, and M, R and Z have the usual significance. Give an outline of the method which you propose to adopt, for the numerical evaluation of factors. [16]

3. State the reasons why 'Rotation of Axis' is necessary in factor analysis? Explain with diagram the technique of rotation for a three factor case. [16]

4. Give the operational definition of 'equivalence' in item analysis. How do you test this 'equivalence'? How do you calculate the reliability coefficient of the test from the two sub-tests formed on the basis of 'equivalent items'? [16]

<u>Neatness</u> [

Please turn over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1964

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

Time : 5 hours Full marks : 100

a) Figures in the margin indicate full marks.

b) Use of Calculating machines is permitted.

Group A

(Answer all questions)

1. Given that $f(x)$ is a polynomial in x and that $f(0) = 10$, $f(10) = 15$, $f(20) = 24$ and $\int_0^{10} f(x)\,dx = 120$, calculate $f(x)$ at $x = 1, 2, 3, 4$ and 5, as accurately as possible. (10)

2. The following values of l_x, the number of persons at age x in any year of time are available. Estimate μ_x (the force of mortality) $= -\frac{1}{l_x}\frac{dl_x}{dx}$ for $x = 50$.

Age x	50	51	52	53
l_x	73,499	72,724	71,753	70,599

(7)

3. Using the data of the following table compute the integral $\int_{0.5}^{1.1} xy\,dx$

x	0.5	0.6	0.7	0.8	0.9	1.0	1.1
y	0.4804	0.5669	0.6490	0.7262	0.7965	0.8605	0.9281

(7)

4. Select any three of the following items and mention in each case the publications in which you can get the required information in India indicating the office from which the publications are issued, periodicity of publications and the time lag in publication.

(i) Agricultural Production, (ii) National Income, (iii) Education, (iv) Vital Statistics, (v) Trade Unions. (9)

5. From the official publications supplied collect data on total cropped area and total area under irrigation in India classified by different methods of irrigation for latest available 6 consecutive years and comment briefly on the salient features of the data. (10)

Please turn over

Group B

(Answer three questions in all, number 6 being compulsory)

6. The following table shows the frequency distribution of heights (in inches) of 300 students.

height in inches	frequency	height in inches	frequency
62 -	1	70 -	48
63 -	2	71 -	42
64 -	1	72 -	35
65 -	4	73 -	21
66 -	12	74 -	14
67 -	31	75 -	8
68 -	31	76 -	2
69 -	47	77 -	1
		Total	300

a) Calculate β_1 and β_2 of the above distribution and (without making any tests) suggest if a normal distribution would fit the data. (12)

b) Assuming a normal distribution with the mean and variance as calculated from the data, find the expected frequencies in the class-intervals 63-65, 69-70, and 75 and above. (6)

7. A biological organism was observed under microscope, and the following table shows the numbers of slide-squares on which 0, 1, 2, ... organisms were found.

No.of organisms	0	1	2	3	4	5	6	7
No.of squares	60	135	142	94	46	15	4	1

a) assuming a Poisson distribution, obtain the expected frequencies. (8)

b) test the goodness of fit of the Poisson distribution. (8)

8. The following table shows the selection test scores (x) and final examination scores (y) of 80 students :

y \ x	60-	80-	100-	120-	140-	160-	180-	200-	220-	240-	260-	280-
60 -				1								
80 -												
100 -	2		1									
120 -												
140 -				1			1					
160 -				5	3	1						
180 -					2	4	2					
200 -						3	4	7	2			
220 -							2	4	7	4		
240 -								2	7	3	1	
260 -									1	4	4	
280 -									2			

Please turn over

a) calculate the coefficient of correlation between x and y. (7)

b) suppose the linear regression line of y on x is obtained; calculate the variance of the deviations of y from the line of regression. (3)

c) calculate the correlation ratio of y on x and comment on the linearity of regression. (4)

9. Monthly data on the end-of-month stocks of a commodity were obtained from January 1952 to December 1958, and the trend removed after computing a centred 12-month moving average. The adjusted values of stock as percentages of trend values are shown in the following table :

Month	Year						
	1952	1953	1954	1955	1956	1957	1958
January		106.1	108.6	104.6	101.2	104.2	102.5
February		94.8	96.7	91.7	88.3	94.2	92.4
March		86.6	87.7	83.5	78.1	85.6	83.3
April		77.0	80.5	78.2	71.0	83.6	78.6
May		70.5	76.7	72.4	67.3	79.1	73.2
June		73.1	76.8	79.9	76.0	82.8	75.5
July	88.3	87.0	86.0	89.0	90.1	91.6	
August	105.2	104.6	105.9	111.7	111.9	110.2	
September	119.1	123.5	123.9	124.4	126.1	124.7	
October	127.9	130.6	126.8	128.2	129.2	126.6	
November	124.9	126.3	123.9	117.0	124.6	123.1	
December	116.1	122.1	117.4	114.6	115.6	114.6	

a) calculate the seasonal indices. (10)

b) with the help of these indices, obtain the monthly figures for 1957 from which trend and seasonal effects have been removed, and write a note on the results. (6)

oooOooo

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1964

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

Time: 5 hours. Full marks: 100

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

GROUP A

(Answer all questions)

1.a) From the data given below, estimate the parameters A and B in the model:

$$Y = AX^B$$

by fitting a straight line in log X to log Y by the method of least squares.

X	Y	X	Y	X	Y
10	1.62	60	0.62	110	0.47
20	1.60	70	0.52	120	0.46
30	1.52	80	0.51	130	0.44
40	1.43	90	0.51	140	0.42
50	0.75	100	0.49	150	0.40

[10]

b) A sample of size 5, from the population with frequency function

$$\frac{1}{\pi} \cdot \frac{1}{1 + (x - a)^2}$$

gives the following values 12.16, 12.98, 13.60, 15.62 and 12.45. Find by successive approximation the maximum likelihood estimate of the parameter 'a'. Take 12.2 as trial value and go through two cycles of iteration. [10]

2.a) Twenty individuals are chosen at random from a population and their weights (in lbs.) are found to be

190	122	141	151	171
132	128	196	111	96
78	84	123	142	175
140	162	123	117	140

Do these data agree with the contention that the mean weight of the population is 145 lbs.? Assume that the weights are normally distributed in the population. [5]

b) The heights in inches and weights in lbs., of 9 individuals are given below. Calculate the correlation coefficient and test if it is significantly different from zero.

height (in inches)	63	63½	67	66	68	69	70	71	70
weight (in lbs.)	104	105	117	115	119	120	140	142	170

[5]

Please turn over

2.c) In a certain experiment on 10 varieties of wheat, 50 plots of the same size are randomly allocated, five plots to each variety, and the following yields in lbs., per plot are obtained.

Varieties	1	2	3	4	5	6	7	8	9	10
	7	7	14	11	9	6	9	8	12	9
	8	9	13	10	9	7	13	13	11	11
	7	6	16	11	12	5	12	11	11	12
	9	14	17	12	8	4	7	12	11	12
	10	13	14	13	7	8	17	14	11	13

Test if the varieties are significantly different in respect of average yield. [10]

GROUP B

(Answer any two questions)

3. The following table gives the results of an experiment involving 4 treatments (indicated by numbers within parentheses) in 4 blocks each of 3 plots, with a balanced layout.

Blocks	Treatments	and	yields
	(1)	(3)	(2)
1	138	154	96
	(4)	(2)	(1)
2	212	113	120
	(1)	(4)	(3)
3	179	321	234
	(2)	(3)	(4)
4	208	267	312

Carry out an analysis of the intra-block information and write brief report on your findings. [15]

4. A completely randomised design in which whole-plots were split into sub-plots was used to test 6 varieties of corn. Planting was done on 7 different dates. On each date, 2 whole plots each of 6 sub-plots were selected at random and planted. Within each whole plot, the varieties were allocated to the sub-plots at random.

a) show the structure of the analysis of variance appropriate for this design, indicating the degrees of freedom for each component. [6]

b) given the following mean squares, test for the significance of the main effects of the date of sowing and the varieties and the interaction between the two. [7]

Please turn over

Source of variation	Mean square
1. dates	69.74
2. varieties	54.10
3. dates × varieties	17.05
4. residual variation between whole plots	37.52
5. residual variation between sub-plots within whole plots	16.84

5. To study the effects of glass type (I, II) and phosphor type (A,B,C) on the brightness of a TV tube screen, an experiment was conducted using 3 replications of each of the combinations of glass type and phosphor type. The measured variable was the current in micro-amperes necessary to produce a certain brightness, the larger this current, the poorer the tube screen characteristics. The results are given below; analyse the data and comment on the effects of the two factors.

Current in micro-amperes necessary to produce standard brightness.

Glass type	Phosphor type A	B	C
I	280	300	270
	290	310	285
	285	295	290
II	230	260	220
	235	240	225
	240	235	230

[15]

GROUP C : SAMPLE SURVEYS

(Answer both the questions)

6. The results of 100 throws of an unbiased coin are given below, where T and H stand respectively for 'tail' and 'head'.

THHTH	HTHTT	TTTHT	HTHHT
HHTTT	HTTTH	THHHH	TTTTH
HHTHT	HTTHH	HTTTH	THHTH
HTHTT	THTTH	HTTTT	HTTHT
TTHHT	HTHTT	HHHTT	HTTHT

Using these results, draw a random sample of 2 units with replacement, from a list of 17 units numbered serially from 2136 to 2152 inclusive. Describe fully the procedure adopted by you. [10]

7. There are 2357 small agricultural farms (area of each less than 100 acres) in a certain country. A sample survey of these farms were undertaken to estimate, amongst other things, the average number of cattle per farm. The farms were divided into 5 strata in respect of the area of the farm and from each stratum a specified number of farms were sampled, on a simple random sampling basis, without replacement. In the table below, we denote by N_i the number of

Please turn over

farms in the i-th stratum, the number of farms sampled in the i-th stratum is denoted by n_i, and $\bar{y}_i$ and s_i^2 denote respectively the mean and the variance (divisor n_i) of the number of cattle in the sample from the i-th stratum; i = 1, 2, 3, 4, 5.

area of farm (acres)	stratum number (i)	number of farms in stratum N_i	sampled n_i	mean number of cattle per sampled farm $\bar{y}_i$	variance of number of cattle per sampled farm s_i^2
less than 15	1	724	61	4.246	27.546
16 - 30	2	648	55	11.636	56.738
31 - 50	3	560	46	15.957	71.697
51 - 75	4	344	29	23.586	192.326
76 -100	5	81	9	29.667	334.922

a) estimate the average number of cattle per farm in the whole country and calculate the standard error of the estimate.

b) estimate the gain in efficiency due to stratification.

Please turn over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1964

Paper VII : Applied Statistics (Practical)

Time : 5 hours. Full marks : 100

a) Candidates are required to answer questions from the two groups for which they have registered their option.

b) Separate answer books are to be used for each of the two groups attempted.

c) Figures in the margin indicate full marks.

d) Use of Calculating Machine is permitted.

GROUP A : ECONOMIC STATISTICS

(Answer all questions)

1. Value of imports (in million dollars) of consumption and capital goods into ECAFE region, are given below for the years 1951-61. Compare the trends in these two types of imports. [12]

year	consumption goods	capital goods
1951	4031	1378
1952	4185	1501
1953	3829	1443
1954	3405	1514
1955	3245	1718
1956	3540	2200
1957	3953	2915
1958	3529	2377
1959	3268	2517
1960	3728	3233
1961	3853	3548

2. Average per capita monthly total expenditure and expenditures (in Rs.) on food grains, and clothing are given below for different income groups in a rural area. Obtain income elasticities for these items. What will be the change in expenditure on food grains and clothing, if income rises by 5 p.c. ? [17 + 5]

monthly income (in Rs.)	no. of persons	total expenditure	expenditure on food grains	expenditure on clothing
below 5	58	4.76	3.15	0.27
5 - 10	303	8.85	5.65	0.63
10 - 15	177	13.33	7.31	1.31
15 - 20	140	17.21	8.72	1.77
20 - 30	123	23.83	9.92	2.77
30 - 50	42	38.31	12.33	4.78
over 50	28	54.22	13.01	7.65

3. Distribution of assessed income in India is given below for the years 1951 and 1960. Analyse the data and examine if the inequality of income distribution has increased. [10]

Please turn over

Annual income (Rs.)	1951 no. of assessees (thousand)	1951 assessed income (Rs. crores)	1960 no. of assessees (thousand)	1960 assessed income (Rs. crores)
below 10,000	371.2	174.2	643.0	358.7
- 20,000	61.8	84.1	155.2	212.7
- 70,000	32.7	108.2	79.1	270.9
-100,000	2.5	20.5	6.5	53.9
-200,000	2.4	33.3	4.6	61.3
over 200,000	1.8	154.4	2.6	234.5

GROUP B : STATISTICAL QUALITY CONTROL

(Answer any two questions)

1. The following 40 readings represent a pilot run, on the production of a temperature control device. They give the 'on' temperature at which a thermostatically controlled switch operates at a given setting.

Device number	Temperature	Device number	Temperature
1	67.6	21	67.7
2	67.1	22	67.4
3	67.6	23	67.3
4	67.5	24	67.6
5	67.3	25	67.5
6	67.5	26	67.1
7	67.4	27	67.7
8	67.5	28	67.9
9	67.6	29	67.2
10	67.6	30	67.5
11	68.0	31	67.7
12	67.7	32	67.2
13	67.5	33	67.8
14	67.6	34	67.8
15	67.6	35	67.6
16	67.5	36	67.6
17	67.1	37	67.6
18	67.6	38	67.8
19	67.4	39	67.7
20	67.6	40	67.9

Check these for control. Make an estimate of σ'. What conclusions, if any, can you reach regarding the prospect that the process can meet specifications of 67.5 ± 0.5? [25]

2. Where destructive testing is involved, a common acceptance procedure is to test one article from a lot, passing the lot if the article is satisfactory and taking a second sample of one if the article fails the test. If the second article also fails, the lot is rejected, if the second article proves satisfactory, the lot is accepted. Plot the O.C. curve for this plan, assuming that submitted lot is large enough, so that the selection of a defective Article on the first sample does not make any appreciable change in the probability of getting a defective article on the second sample.

Comment on the quality protection given by this plan? [25]

Please turn over

3.a) A manufacturer wishes to produce electric fuses with no more than 1 percent defective. He checks quality every so often by taking a sample of 10 fuses from the line; if one or more of the 10 are defective, the manufacturing process is halted, and a search is made for an assignable cause.

i) how often will he needlessly halt production, when defects are running at 0.75 percent ?

ii) how often will he fail to halt production, when defects are running at 2 percent? [8 + 8]

b) A machine produces parts which on an average contain 1 percent defective, when these parts are put in batches without further examination, percentage defective would naturally vary from batch to batch. What should be the size of each batch to ensure that only one batch in a hundred would contain more than 2 percent defectives? [9]

GROUP C : STATISTICAL METHODS IN GENETICS
(Answer all questions)

1. Estimate the O - A - B gene frequencies from the observed frequency distribution of O, A, B and AB blood group classes.

blood group	O	A	B	AB
frequency	70	45	68	15

Test, whether the expectations are according to Bernstein's theory.. [25]

4. The following data relate to 4 back cross tests $\frac{AB}{ab} \times \frac{ab}{ab}$.

Test	frequency			
	AB	Ab	aB	ab
1	50	75	100	65
2	52	80	105	60
3	48	70	90	45
4	40	63	85	40

Examine whether the four tests provide consistent estimates of the linkage parameter. [25]

Please turn over

GROUP D : VITAL STATISTICS AND DEMOGRAPHY
(Answer any two questions)

1. The mortality experiences in a given Calendar year of two populations A and B consisting of men aged 70-74, are as follows:

Age x	population A			population B		
	numbers exposed to risk at age x	deaths at age x	q_x	numbers exposed to risk at age x	deaths at age x	q_x
70	2,000	64	.032	3,000	81	.027
71	1,200	42	.035	3,500	112	.032
72	1,400	56	.040	3,000	117	.039
73	1,800	81	.045	4,000	184	.046
74	1,600	80	.050	2,500	140	.056
Total	8,000	323		16,000	634	

Express the mortality of Population B at ages 70-74 as a single percentage ratio of that of Population A by means of the following:

i) Indirect standardisation
ii) Direct standardisation
iii) Comparative mortality index.

Discuss the differences between the results and the reasons for these differences. [25]

2. The following table shows the population of England and Wales, enumerated at every census from 1801 to 1911.

year	census population (in thousands)	year	census population (in thousands)
1801	8,893	1861	20,066
1810	10,164	1871	22,712
1821	12,000	1881	25,974
1831	13,897	1891	29,003
1841	15,914	1901	32,528
1851	17,928	1911	36,070

Fit a logistic curve to the Census Populations in the above table, and estimate the maximum expected population of England and Wales. [25]

3. The following table may be taken to be characteristic of the Indian population in respect of females:

age group	age-specific marital fertility rate	proportion married	life-table survivors (l_0 = 1,000)
15 - 19	.198	.551	2904
20 - 24	.268	.830	2792
25 - 29	.237	.947	2647
30 - 34	.190	.907	2446
35 - 39	.141	.827	2207
40 - 44	.087	.736	1958

Calculate the net female reproduction rate, ignoring illegitimate fertility.
(Proportion of female to male births is 100 : 106) [25]

GROUP E : EDUCATIONAL AND PSYCHOLOGICAL STATISTICS
(Answer all questions)

1. Subjects were given an attitude test before and after viewing a motion picture designed to influence their attitudes favourably. A high score indicates a favourable attitude and a low score an unfavourable attitude.

subject	pre-test	post-test
1	2.6	2.5
2	4.6	5.7
3	8.9	9.3
4	5.5	6.7
5	1.9	1.5
6	6.2	7.8
7	4.6	4.7
8	5.6	5.9
9	6.9	7.3
10	6.6	7.0

Can you conclude that the motion picture resulted in a significant mean change in attitude? [15]

2. The following table gives the matrix of factor loadings of a 6 × 6 correlation matrix, using the 'Centroid process'.

	I	II	III
1	.542	.612	.074
2	.629	.341	- .348
3	.529	- .492	.191
4	.281	- .182	- .550
5	.628	.143	.274
6	.429	- .424	.359

By rotating the axes, obtain an orthogonal simple structure of the matrix of loadings. Comment on the method adopted. [20]

3. The following table gives the split test scores of 25 students.

students	split scores odd (X)	even (Y)
1	227	226
2	124	111
3	210	237
4	178	161
5	192	188
6	104	93
7	191	201
8	140	160
9	125	123
10	141	157
11	171	178
12	168	182
13	129	118
14	192	222
15	176	171
16	172	180
17	215	224
18	102	144
19	177	196
20	109	125
21	146	150
22	100	184
23	179	193
24	141	131
25	141	135

i) examine whether the two halves of the test have the same mean and standard deviation,

ii) determine the reliability coefficient using product moment correlation coefficient,

iii) calculate the reliability coefficient of the test by the method of maximum likelihood estimate.

Write notes on the results obtained. [20]

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1964

Paper VIII : Subjects of Specialisation -(I)

Time : 4 hours Full marks : 100

a) Candidates will be required to answer questions from that group only for which they have registered their options.

b) Figures in the margin indicate full marks.

Group A : Economic Statistics -(Econometrics)

(Answer any five questions)

1. Consider the economic system

$$d_t = \alpha_o + \alpha_1 p_t + u_t$$

$$s_t = \beta_o + \beta_1 p_t + v_t$$

$$d_t = s_t + w_t$$

where

d_t = quantity demanded at time t

p_t = price at time t

s_t = quantity supplied at time t

u_t, v_t, w_t = error terms

α_o, α_1, β_o, β_1 = constant parameters

Are the parameters of this system identifiable ? If they are not, what restrictions would you impose in order to obtain identifiability ? (2

2. Milton Friedman's permanent income hypothesis takes the simple form

$$y = y_P + y_T$$

$$c = c_P + c_T$$

$$c_P = k\, y_P$$

where

y = income

c = consumption

P and T are subscripts denoting the permanent and transitory components

k = true marginal propensity to consume.

The theory assumes that transitory components are independent of their permanent counterparts and also that transitory consumption is uncorrelated with transitory income. The main objective is to estimate the true marginal propensity k.

Describe how you would proceed to estimate the marginal propensity given a sample (c_i, y_i) $(i = 1,2,\ldots, N)$ of household incomes and consumption expenditures.

Please turn over

Discuss briefly the nature of the estimate you would obtain and compare it with the ordinary least squares estimate. (2)

3. "One of the central problems in economic planning has been the matching of demand and supply, at least of consumer goods. This serves the dual purpose of eliminating speculation by minimising price fluctuation, and of mitigating the undesirable effects of trade cycle. Due to increase in income resulting from planned efforts, the consumer demand is normally expected to rise. For a realistic fixation of output targets, and proper mobilisation of productive forces we need a precise knowledge of increase in consumer demand, emanating from a given increase in per capita income."

Formulate the above problem from an econometric point of view. State what information you would need in order to estimate the increase in demand.

Show also that a reduction in inequality of income distribution gives rise to a an increase in demand for essential consumer goods and services.

State clearly all your assumptions. (20)

4. Outline some of the major criticisms against the use of Cobb-Douglas production function. Describe how you would use such production functions for estimating the total capital employed in Indian manufacturing industries. (4)

5. Write short notes on the following :

i) Durbin-Watson test

ii) Fractile Graphical Analysis

iii) Multi-collinearity. (2

6. The log-logistic distribution of income (x) is characterised by the equation

$$\log \frac{F(x)}{1-F(x)} = \alpha + \beta x$$

where $F(x)$ is the cumulative distribution function.

Work out the mean and variance for the log-logistic distribution. Show that the Lorenz measure of concentration for this distribution is given by the reciprocal of β.

In what respects does this distribution differ from the log-normal distribution ? (2

7. Arrow, Chenery, Minhas and Solow have proposed a new class of production functions devised in the following manner.

Please turn over

Suppose it were an empirical fact that the relation between the wage rate W (stated in terms of the relevant output) and output per man-hour $\left(\frac{Q}{L}\right)$ is of the form

$$(1) \qquad W = A\left(\frac{Q}{L}\right)^{\beta}$$

where A, β are some real constants

Suppose, further, that the underlying production of the economic activity in question is of the form

$$(2) \qquad Q = L\, F\left(\frac{K}{L}, 1\right),$$

i.e. it is homogeneous of degree one, where $\left(\frac{K}{L}\right)$ represents capital per man-hour.

In addition it is assumed that the product and labour markets are both competitive so that at factor employment equilibrium

$$(3) \qquad \frac{\partial Q}{\partial L} = W$$

There is, of course, the further unstated but implicit assumption, viz. that the observations relating to (1) refer to an equilibrium situation.

Using the above three relations obtain an explicit form for the production function. Also, work out the marginal productivity of labour and capital as well as the elasticity of substitution between capital and labour. (20)

oooo0oooo

Group B : Techno-commercial Statistics
(Statistical Quality Control)

(Answer any four questions)

1.(a) Examine the use of 'rational subgroups' and 'control limits' in operating a control chart. (10)

(b) When would you recommend the use of 'modified' control limit in an $\bar{X}$ - R chart ? Explain how such limits can be constructed and used in practice. (15)

2. (a) Discuss how you would examine the capabilities of a process in respect of a measurable characteristic using variance component analysis. (12)

(b) Describe briefly the factors that have to be taken into consideration in suggesting remedial action when process capability fails to meet specification limits. (13)

3. Derive, the sequential probability ratio test for the 'proportion defective' from the point of view of acceptance sampling. Explain how a sequential procedure can be set up for testing the superiority of an experimental method over a standard method when the result of each application of the method is classified either as a success or a failure. (25)

Please turn over

4. Explain the terms AOQL and LTPD in respect of an acceptance sampling plan. When would you advocate the use of an AOQL plan in preference to an LTPD plan ? Describe briefly the salient features of any AOQL plan you are familiar with. (25)

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

a) Use of order statistics for estimating lot/process quality.

b) Interval estimates of lot quality.

c) Control charts for moving averages and ranges.

d) Hamilton's standard lot plot method of acceptance sampling by variables.

e) Rotatable designs. (25)

oooo0oooo

Group D : Design & Analysis of Experiments

(Answer any four questions)

1. Describe the Latin Square layout and the randomisation procedure appropriate for such a layout. Give your critical comments on the advantages and disadvantages of such layout.

In an experiment involving t treatments, laid out in the form of t X t Latin Square, it was discovered at the stage of analysis of the results, that two of the treatments were identical. What modifications would you make in the analysis ?

2. Describe a Balanced Incomplete Block design and write down (without proof) the relations connecting the parameters of the design. (10)

If the yield of one plot in such an experiment is missing, how would you analyse the results ? Work out the variance of the estimated difference in the effects of two treatments in such a case. (15)

What is meant by partial confounding in a factorial experiment ? Why is this taken recourse to ? (5)

Give a balanced scheme of partial confounding for a factorial experiment involving 3 factors each at 3 levels using blocks of 9 plots each. (10)

Describe how you would analyse the results of such an experiment. (15)

Please turn over

4. What is a 'rotatable design' for exploration of a response surface ? What are the advantages for such a design ? (10)

Obtain necessary conditions for rotatability for a design for exploring a response surface of the second degree in two factors. (15)

5. An experiment is to be conducted to assess which of two methods of memorizing a poem takes less time : (1) reading a line repeatedly and memorizing it before proceeding to the next line (2) memorizing stanza by stanza (instead of line by line as in method 1).

The students in 3 boys' school and 2 girls' schools in the city are available for the experiment. Three poems, of about the same length written in the same metre, have been selected for the trial. It has been estimated that an average student takes about 30 minutes to memorize a poem of this length. It has been decided to use about 100 students in all, each spending not more than one hour. Particulars of age, family back-ground, performance in school tests etc for all the students in the 5 schools are available.

Give a suitable design for the experiment, explaining briefly how the fundamental techniques of randomisation, replication and error control are utilized in the design. Describe how the results are to be analyzed. (25)

ooooOoooo

Group E : Sample Surveys - (Theoretical Aspects)
(Answer any four questions)

1. Explain the following terms

a) Probability sampling
b) Area sampling
c) Self-weighting designs
d) Stratification
e) Double sampling. (25)

2. i) Describe Lahiri's method of selection with probabilities proportional to sizes.
ii) Show that this method achieves the desired probabilities.
iii) How is the population mean estimated when sample is selected with probabilities proportional to sizes with replacement ?
iv) How would you estimate the variance of the estimator of population mean, for such a design ?
v) Describe briefly the advantage of selection with varying probabilities. (25)

Please turn over

3. i) What are the circumstances under which ratio-estimate of population mean is superior to the simple unbiased estimator ?

ii) Derive the expression for the bias of ratio-estimator.

iii) How would you estimate difference between two ratios ?

iv) How would you estimate the variance of the estimator of this difference ?

v) Describe Midzuno's method of selecting a sample with probability proportional to the sum of sizes. (25)

4. i) Derive the expression for optimum matching for estimating population mean on the second occassion in a repeated survey.

ii) Show that a close approximation to optimum matching may be achieved even though the parameters entering the above expression are unknown.

iii) Given a sample with partial replacement, how would you estimate the difference between the means on the two occasions ? (25)

5. i) What are the sources of errors in a survey ?

ii) What are interpenetrating subsamples ?

iii) Describe the use of interpenetrating subsamples in the analysis of different sources of errors. (25)

ooooOoooo

Group C : Statistical Inference

(Answer any five questions)

1. (a) If t_1 and t_2 are two unbiased minimum variance estimators of a parameter θ, show that the coefficient of correlation between them is 1.

(b) Prove that $P(t_1 \neq t_2) = 0$, and hence comment on the uniqueness of the unbiased minimum variance estimator.

(c) Suppose t is an unbiased estimator of a function $g(\theta)$ of θ. Find an expression for the minimum possible variance of t. (20)

2. (a) Let $x_1, \ldots\ldots, x_n$ be a random sample from a population whose probability density function $f(x;\theta)$ depends on a parameter θ. State, without proof, the properties of consistency, asymptotic efficiency and normality of the maximum likelihood estimator of θ. You are required to give accurate statements of the relevant results, and the conditions under which these results hold.

Please turn over

(b) Consider a population characterised by the probability density function

$$f(x;\theta) = \frac{1}{\Gamma(\theta)} x^{\theta-1} e^{-x}, \quad (x > 0, \ \theta > 0).$$

i) calculate the lowest bound of variance of unbiased estimators of θ.

ii) find an unbiased estimator of θ by the method of moments and verify that it has a variance larger than the value in (i).

iii) obtain the likelihood equation and examine if the estimator in (ii) is a solution of this equation.

(2?)

3. (a) State precise definitions of "sufficient statistic" and "complete statistic" for the parameter θ of a probability density function $f(x;\theta)$.

(b) Given that unbiased estimators t of θ exist, prove that

i) if t_0 is a sufficient statistic, it is possible to obtain an unbiased estimator of θ as a function of t_0, whose variance is not greater than the variance of t;

ii) if t_0 is sufficient and complete, there is a unique unbiased estimator with minimum variance.

(c) Write a note on how sufficient and complete statistics are useful in the construction of tests of hypotheses.

(2?)

4. Independent random variables $y_1, \ldots, y_n$ have a common variance σ^2 and their expectations are of the form

$$E(y_i) = \sum_{j=1}^{k} a_{ij} \pi_j, \quad i = 1,\ldots, n,$$

where a_{ij} are known constants and π_j unknown parameters.

(a) Define an "unbiased linear estimate" of a linear function $p_1\pi_1+\ldots+p_k\pi_k$ with known coefficients $p_1, \ldots, p_k$, and derive necessary and sufficient conditions that such estimates would exist.

(b) Derive the "normal equations" for the minimum variance unbiased linear estimation of $p_1\pi_1 + \ldots + p_k\pi_k$.

(c) Assuming that $y_1,\ldots, y_n$ are normally distributed, obtain a test for the hypothesis

(2?)

$$H_0 : \sum_{j=1}^{k} h_{ij} \pi_j = \theta_i, \quad i = 1,\ldots, s,$$

when it is known that $\pi_1,\ldots, \pi_k$ obey the restrictions

$$R_0 : \sum_{j=1}^{k} r_{ij} \pi_j = \gamma_i, \quad i = 1,\ldots, s.$$

Please turn over

5. (a) State and prove the Neyman-Pearson lemma.

(b) Use this lemma to show how locally most powerful unbiased tests can be obtained in a single parameter case.

(c) Derive a locally most powerful unbiased test for the hypothesis $\sigma^2 = \sigma_0^2$ regarding a normal population $N(0, \sigma^2)$. (2[illegible])

6. (a) Define "similar regions".

(b) Give a set of sufficient conditions on a probability density function with two parameters in order that similar regions may exist for testing a hypothesis regarding one of the parameters. Provide necessary proofs.

(c) For a normal distribution $N(\mu, \sigma^2)$ show how you would construct similar regions for testing the hypothesis $\mu = \mu_0$ and then choose a uniformly most powerful test for a specified class of alternatives. (2[illegible])

7. The distribution of a random variable is known to be one of a finite number of alternatives. Also, on the basis of a random sample of size n, one of a finite number of decisions has to be taken.

(a) Define

i) loss and risk functions;

ii) randomised decision functions.

(b) Define 'Bayes', "minimax" and "admissible" decision functions in the above situation and, using the supporting hyperplane theorem for convex sets, show that every admissible decision function is also a Bayes decision function. (2[illegible])

oooOooo

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1964

Paper IV: Subjects of Specialisation -(II)

Time: 4 hours Full marks: 100

a) Candidates are required to answer questions from that group only for which they have registered their options.

b) Figures in the margin indicate full marks.

GROUP A : ECONOMIC STATISTICS (Indian Economics & Economics of Planning)

(Answer three questions from each section)

Section I

1. 'Development of agriculture and development of Industry in India, to a very large extent are not competitive with each other'. Give your considered views on this thesis. [16]

2. Discuss the main benefits that India has enjoyed from her membership in I.M.F. and G A T T. Outline the restrictive constraints that operate on the exchange and commercial policies of India in consequence of her membership in these organizations. [16]

3. Analyse India's national income by its components. Discuss the difficulties in the way of satisfactory estimation of various components.

4. What are the main defects of joint-stock banking in India? How have these been sought to be removed through legislation in the recent past? Can you suggest some further steps for removing specific defects? [16]

5. Discuss the main features of the development of steel industry in India, in recent years. Also examine critically the present distribution policy of the government with regard to this industry. [16]

Section II

6. What in your opinion should be the important features of a long term planning model for India? Give the sketch of an optimization model incorporating such features into it. [16]

7. What do we mean by a Matrix of input coefficients? List the main factors which affect the stability of these coefficients. Can you demonstrate the usefulness of an input - output matrix for purposes of measuring sectional interdependence? [16]

8. 'India should plan for a structure of production which can ensure a faster rate of growth, but the socialist goal of equalitarianism ' should be sought through proper tax and public expenditure policies. Comment on this prescription. [16]

9. Compare and contrast different models of economic growth in their treatment of the question of 'input' substitution. Assess the realism as well as the operational significance of the notion of factor substitution in economic development [16]

Please turn over

10. Write critical notes on any two of the following:

i) Marginal capital - output ratios and their use in investment allocation.

ii) Import substitution as development strategy.

iii) The concept of commodity balances in Russian Planning.

iv) Regional planning in India. [16]

Neatness (sections I and II). [4]

GROUP B : TECHNO-COMMERCIAL STATISTICS - OPERATIONS RESEARCH/ELEMENTS OF BOOK-KEEPING AND ACCOUNTANCY/STATISTICAL METHODS IN BUSINESS.

Section I: Operations Research

i) Use separate answer book for this group.

ii) Answer four questions of which question no. 1 is compulsory.

1.a) Give a definition of 'O. R.' [3]

b) Give mathematical formulations of the following classical problems in 'O.R.'

i) Travelling Salesman's problem

ii) Newspaper-boy problem

iii) Job Sequencing

iv) Zero-Sum Two person game [4 + 4 + 4 + 4]

2.a) Explain the different types of costs that will, in general, occur in inventory control analysis. [6]

b) Briefly describe the different components that make up the inventory carrying cost. [3]

c) Explain the following terms in the subject of inventory control

i) lead time
ii) back-log
iii) deterministic and random models [3 + 3 + 2]

3.a) State and formulate the transportation problem as a problem of linear programming; also formulate its dual. (4 + 5)

b) Formulate the following problem as a linear programming problem. (you need not solve it). [4 + 5]

A blender of whisky imports three grades A, B and C. He mixes them according to recipes which specify the maximum or minimum percentages of grades A and C in each blend.

Please turn over

3.b) contd.

Specification of blends

	blend	specification	price per bottle
	Blue Dot	not less than 60 percent of A and not more than 20 percent of C	Rs.6.80
	Highland Fling	not more than 60 percent of C and not less than 15 percent of A	Rs.5.70
c)	Old Frenzy	not more than 50 percent of C	Rs.4.50

Availability and cost of ingredients

whisky	maximum number of bottles available per day	cost per bottle
A	2000	Rs. 7.00
B	2500	Rs. 5.00
C	1200	Rs. 4.00

It is required to obtain the production policy that will maximize the profit. [8]

4. A large population of bulbs is subject to a given mortality (or life) curve, for a very long period. All failures are immediately replaced by new ones and there are no other entries or exits. Show that the age distribution ultimately becomes stable and that the number of failures (or replacements) per unit time, becomes constant. [17]

5. Buses are scheduled to pass a certain corner every 15 minutes but actually the arrival of a bus varies normally about its scheduled arrival time with a standard deviation of 3 minutes. People wishing to board the bus arrive in a Poisson fashion, with mean arrival rate of 4 persons per hour and the number of empty seats on the bus has a Poisson distribution with mean 3/2. If no standing is permitted, explain in detail, how you can find the average waiting time of an arrival by Monte Carlo method. [17]

Section II: Statistical Methods in business

1. EITHER

What are the objectives and advantages of merit rating? Explain different methods of merit rating. [15]

OR

What are the different job characteristics to be considered in job evaluation? Explain with examples? What is the method of ranking in job evaluation and what are its draw-backs? [15]

2. EITHER

Discuss the importance of forecasting in business management. Suggest Economic Indicators and indexes which a radio manufacturing organization might develop as an aid in forecasting its business and indicate how these would be utilized. [15]

Please turn over

2. CR

Suppose that you are a statistician in a firm which manufactures hair oils. You are asked to conduct a consumer survey by the firm. Explain clearly the procedure which you will adopt, giving as many relevant details as possible. [15]

Section I (alternative) - Elements of Book-keeping and Accountancy

(i) Use separate answer book for this group.
(ii) Answer all the questions.

1. JUNIOR ACCOUNTANTS' CLUB
Receipts and Payments Account for the year ended 31st December, 1963

Opening Balances:	Rs.		Rs.
i) Cash in hand	250	(includes Rs.400 for	
ii) Balance at Bank	20,550	1964)	6,000
Subscriptions:		Telephone Charges	540
1962	500	Stamp and Stationery	600
1963	20,000	Travelling	150
1964	750	Meeting Expenses	500
Hall rent	1,250	Rent	5,400
Interest on securities	1,000	Library	3,000
Donations	10,000	Donation for Political agitation	5,000
Telephone receipts	50	Closing Balances:	
		i) Cash in hand	310
		ii) Balance at Bank	32,600
	Rs. 54,350		Rs. 54,350

The following additional information is available:-

a) The Association holds 3 percent G.P. Notes amounting to Rs.40,000

b) The Library Account stood at Rs.20,000 on 1st January 1963

c) The donations of Rs.10,000 are to be funded for a prize to be awarded by the Association.

d) The outstanding liabilities are:

(i) Rent Rs.300; (ii) Printers bill Rs.150.

You are required to prepare an Income and Expenditure Account for the year ended 31st December, 1963 and a Balance Sheet as at that date. [24]

2. KICHER

During an interim audit at 31st December 1960 (final accounts made up to a later date), you are instructed by your principal to reconcile the bank balance shown in the books of account with the balance shown in the Bank Pass Book from the following particulars. Prepare a Reconciliation Statement:- Balance per Pass Book 31st December, 1960, overdrawn Rs.1,026.12 n.p. Cheques drawn on 31st December but not cleared till January 1961: Rs.12.00 Rs.1,021.00, Rs.98.20 and Rs.112.15.

Please turn over

2.contd.

Bank overdraft interest, 20th December, 1960, not entered in Cash Book Rs.151.61.

Sum received on 30th December, 1960, but not lodged in bank till 3rd January 1961 Rs.20,100.00.

Cheque book Rs.5.00, 10th November 1960 entered in the Cash book twice in error and another Cheque Book debited in Bank Pass Book in August, 1960. Rs.1.00 not entered in the Cash Book.

Bill Receivable due on 29th December, 1960, was passed to the Bank for collection on 28th December, 1960 and was entered in the Cash Book forthwith. Whereas the proceeds were not credited in the Bank Pass Book till 1st January 1961, Rs.250.00.

Bill stamps amounting to Rs 1.80 had been debited in the Pass Book, no entries having been made in the Cash Book.

Chamber of Commerce subscription paid by bankers order on 1st December 1960 had not been entered in the Cash Book Rs.40.00.

Note: Assume that you do not alter the Cash Book Balance at 31st December 1960, all correcting entries being made in January 1961. [24]

2. OR

A trader's Book-Keeper has agreed a Trial Balance and drafted the Trading and Profit and Loss Account and the Balance Sheet. You discover the following errors:-

i) Sales on Approval amounting to Rs.100 have been included in the Sales Account, Rs.75 of these goods were returned. No record of the return was made in the books, but the returned goods were included in Stock at their cost price of Rs.50.

ii) A Cheque for Rs.250 received for a loss of stock sustained by fire has been paid by the proprietor into his private Bank A/c and not recorded in the books.

iii) Purchased goods amounting to Rs.200 are included in Stock but the invoice was dated forward and is not entered for the period under review.

iv) There were three compensating errors, viz., Discounts received were undercast Rs.5.00, Debit side of a Sales Ledger Account was overcast Rs.10.00 and a payment of Rs.15.00 for Legal Expenses had not been posted from the Cash Book.

State the effect of each of these errors, and summarise the alterations to be made in the Accounts as originally drafted. (24)

3.EITHER

The Gross Profit of a trading concern has come down from 15 percent to 10 percent in an accounting year. How would you justify the short fall of Gross Profit in comparison with the previous period? [10]

OR

There is a Consumers' Stores in your locality. Prescribe suitable books with rulings which will be sufficient to keep the details of its business transactions. [10]

Please turn over

4. EITHER

How human errors can be eliminated in recording books of accounts? What are the common types of errors which you may come across in checking up the final trial balance of a business concern? [10]

OR

Write short notes on :-

a) Deferred Revenue Expenditure
b) Watered capital
c) Current Assets. [10]

Neatness - [2]

GROUP D : DESIGN AND ANALYSIS OF EXPERIMENTS

(Answer any four questions)

1. Define a partially balanced incomplete block design with two associate classes and state the relations among the parameters. Find out the parameters of second kind for such a design when the association scheme is (i) Triangular and
(ii) L_3

Construct a group divisible design based on a balanced incomplete block design with $\lambda = 1$. [25]

2. Explain the method of difference sets in the construction of a balanced incomplete block design and construct one of the following designs

i) $v = b = 13, \quad r = k = 4, \quad \lambda = 1$
ii) $v = b = 19, \quad r = k = 9, \quad \lambda = 4.$ [25]

3.a) Define a finite projective plane and show how it can be constructed with the help of a finite field.

b) Define a Hadamard matrix and show how to obtain another Hadamard matrix with the help of two given Hadamard matrices. [25]

4. Defind 'balanced' in symmetrical factorial experiments. Show that complete balance can be achieved over the 2, 3, ..., m factor interactions in a (s^m, s^{m-1}) design (i.e. for a design in which the complete replication consists of s^{m-1} blocks of s each) in $(s-1)^{m-1}$ replications and where the main effects are unconfounded. [25]

5. Show that if the alias sub-group involves only interactions of the $(t + k - 1)$th and higher orders $(t > k)$, the alias of an interaction of order $(k-1)$ would be an interaction of order $(t-1)$ or more [25]

6. Define a rotatable design and explain its uses.

Give a method of constructing rotatable design of second order in three dimensions. [25]

Please turn over

GROUP E. SAMPLE SURVEYS - ORGANIZATIONAL ASPECTS

(Answer any four questions)

1. Describe the various operations involved in the collection of data through sample surveys. [25]

2.a) Explain the basic principles that should be observed at the time of processing of data. [15]

b) Describe in short the functions of the principal punch-card equipments. [10]

3. You are asked to conduct an economic survey of small industrial establishments of your state.

a) describe the various phases of the work subsequent to the planning stage of the survey, [15]

b) how will you control numerical accuracy in analysing the data? [10]

4. You are required to conduct a family budget enquiry in your state for determining the weights for construction of cost of living indices of the non-agricultural class of population.

a) draw up the necessary schedules for the survey, [5]

b) draft the instructions for the field workers explaining the concepts and definitions of the major items of the schedules, [10]

c) draw up the blank tables in which the data will be presented, [5]

d) prepare the necessary punch-card designs for mechanising the analysis work, [5]

5.a) Discuss the merits and demerits of the methods of mail-enquiry and interview. [15]

b) You are asked to conduct a sample survey for estimating the number of vagrants of different categories in a big city of India. What will be your frame? Discuss the possible sources of non-sampling errors that may affect the results of the survey. How will you minimise these errors? [12]

c) Write a critical note on the types of non-sampling errors that may arise when decisions on sample design are based on sample results. [7]

6. Write notes on any three of the following. [25]

a) follow up of non-response cases

b) optimum design

c) stock-taking of resources

d) grid sampling.

Please turn over

GROUP G. STATISTICAL INFERENCE

(Answer any four questions)

1. Derive the Sequential Probability Ratio Test procedure for testing the hypothesis that the mean of a Normal population is μ_0 against the alternative that the mean is μ_1, when the standard deviation of the population is known to be unity. Work out the Average Sample Number required by the procedure if the test is to be of strength (α, β). (25)

2. Write a note on Hoeffding's U-statistics and its limiting distribution. Express the sample variance as a U-statistic and hence derive its limiting distribution (after normalisation) as the sample size tends to infinity. (25)

3. Let y and z be respectively the r-th and the s-th $r < s$ order statistics in a sample of size n from a population with cumulative distribution function $F(x)$ which has a continuous density. Let G be defined by $F(G) = p$ where $0 < p < 1$. Work out the probability that $y < G < z$, and show how this result could be used in setting up a confidence interval for G. Discuss in particular the case $r = 1$, $s = n$, $p = \frac{1}{2}$. (25)

4. Write short notes on the following non-parametric test procedures.

   i) Wilcoxon's test for two samples
   ii) Sign test
   iii) Kendall's Rank correlation test for independence. (25)

5. How will you test, on the basis of samples from two p-variate Normal populations known to have a common dispersion matrix, the hypothesis that the two mean vectors are equal when the common dispersion matrix is (i) known and (ii) unknown and is to be estimated. In each case work out the sampling distribution of the test statistic when the hypothesis is true. (25)

6. Derive the likelihood ratio test for the hypothesis that two multivariate Normal populations have a common dispersion matrix. Can you suggest any other suitable test procedure ? (25)

Please turn over

INDIAN STATISTICAL INSTITUTE
Statistician's Diploma Examination - November 1964
Paper X - Subjects of Specialisation - (III)

Time : 5 hours . . . . . Full marks :100

i) Candidates are required to answer questions from that group only for which they have registered their options.

ii) Figures in the margin indicate full marks.

iii) Use of calculating machine is permitted.

GROUP A : ECONOMIC STATISTICS - III

(Answer any three questions)

1. A notorious model of the meat market is :

Demand : $\beta_{11} Y_1 + \beta_{12} Y_2 + \gamma_{11} Z_1 = \mu_1$

Supply : $\beta_{21} Y_1 + \beta_{22} Y_2 + \gamma_{22} Z_2 + \gamma_{23} Z_3 = \mu_2$

The variables are :

Y_1 : meat consumption per capita (pounds)

Y_2 : retail price of meat

Z_1 : per capita disposable income (tens of $)

Z_2 : unit cost of processing meat (index)

Z_3 : cost of agricultural production (index)

(All variables are expressed as deviations from means.)

The sample consists of 23 annual observations. The master moment matrix $M_{XX} = [m_{x_i x_j}] = [\frac{1}{T} \sum_{t=1}^{T} x_{it} x_{jt}]$ has been computed :

	Y_1	Y_2	Z_1	Z_2	Z_3
Y_1	59.545	-15.328	15.965	-23.325	42.777
Y_2		68.760	36.324	36.971	53.729
Z_1			34.276	15.703	53.064
Z_2				110.209	31.773
Z_3					114.217

Obtain :

a) mechanical least square estimates for the demand equation, assuming Y_1 as the dependent and Y_2 and Z_1 as explanatory variables, and

b) limited information maximum likelihood estimates for the supply equation. Estimate sampling errors. [32]

Please turn over

2. The following table gives the frequency distribution of income in the United States [1954].

income ($ 000)	percentage of families
less than 1	6.1
1 - 2	10.6
2 - 3	12.3
3 - 4	14.4
4 - 5	14.8
5 - 6	12.6
6 - 7½	12.3
7½ - 10	9.6
10 - 15	4.3
more than 15	3.0

Fit a Pareto distribution to the above data. Estimate the parameters. Use this information to derive the concentration coefficient and the mean level of income. [32]

3. Given an economy with the following possible production activities:

	A_1	A_2	A_3	A_4	A_5	A_6
Commodity 1	+ .8	+ 1.0	+ 1.0	- .5	- .8	- .2
Commodity 2	- .2	- .1	- .7	+ 1.0	+ 1.0	+ 1.0
Labour	- 6	-12.5	- 5	- 5	- 4	- 10
Capital	-1.6	- 1.5	- .3	- 1.5	- .2	- 1.4

a) determine the most efficient combination of activities for producing any combination of the two commodities if labour is the only limiting factor and only 750 units of labour are available.

b) determine the relative prices of the two commodities and of one unit capital in part (a).

Note: (+) signs indicate outputs and and (-)signs indicate inputs. [32]

4. Consider the model:

$$Y_i = \alpha + \sqrt{\beta}\ X_i + \mu_i$$

where X_i is non-stochastic

and μ_i independently and identically distributed with mean zero, over a set of 12 observations.

The sums of squares and cross-products (around the respective means are:

	Y	X
Y	250	150
X		100

Derive and compute an unbiased estimate of β. [32]

Monitoring [4]

GROUP B. TECHNO-COMMERCIAL STATISTICS -
STATISTICAL QUALITY CONTROL,
OPERATIONS RESEARCH/ELEMENTS OF BOOK-KEEPING &
ACCOUNTANCY, STATISTICAL METHODS IN BUSINESS.

(Separate answer books are to be used for each of the sub-groups of B)

I. Statistical Quality Control

(Answer any two questions from this sub-group)

1. Determine the equations of the rejection and acceptance lines for an item-by-item sequential sampling inspection plan, in which

A Q L = 0.05	Producer's risk = 0.10
L.T.P.D = 0.15	Consumer's risk = 0.20

Prepare an item-by-item table of acceptance and rejection numbers for values of n from 1 to 50. [16]

Plot an approximate O.C curve for the above plan. [9]

2. In an investigation of the eccentricity of helical gear, the measurements were taken on a few randomly chosen products at three main operations namely turning, milling and grinding. The data are given below.

Turning	Milling	Grinding
0.0020	0.0080	0.0110
0.0005	0.0030	0.0100
0.0025	0.0040	0.0050
0.0050	0.0050	0.0100
0.0020	0.0120	0.0040
0.0015	0.0100	0.0090
0.0050	0.0060	0.0040
0.0020	0.0045	0.0040
0.0020	0.0070	0.0120
0.0010	0.0070	0.0050

i) can the variations in the eccentricity at the turning and the milling operations be considered different?

ii) is there any evidence that the variations at the milling and grinding operations are different? [25]

3. In Aluminium rolling, ingots were cast with two different heading metal temperatures, high and low. There were 30 ingots cast at the high temperature and 20 ingots cast at the low temperature. After casting, ingots were hot rolled and cut into two halves. The heading half of each ingot was again sheared into 7 pieces. A count was made of the pieces rejected due to blisters. The results are shown in the table below.

	Rejected	Good	Total
High temperature	39	171	210
Low temperature	40	100	140
Total	79	271	350

Is there any evidence that temperature has an effect on the occurrence of blisters? (25)

II - Operations Research

(Answer both the questions)

1.a) Using simplex method, find the non-negative values of x_1, x_2, x_3 that will maximize the expression

$$N \quad Z = 3x_1 + 5x_2 + 4x_3$$

subject to the restriction

$$\begin{aligned} 2x_1 + 3x_2 &\leq 8 \\ 2x_2 + 5x_3 &\leq 10 \\ 3x_1 + 2x_2 + 4x_3 &\leq 15 \end{aligned}$$

[8]

b) From the optimal simplex tableau obtained in (a), write down the optimal values of the dual variables, and verify the 'duality property' that the optimal values of the original and the dual problems are the same. [8]

2. EITHER

A national car-rental service has a surplus of one car in each of the cities 1, 2, 3, 4, 5, 6 and a deficit of one car in each of the cities 7, 8, 9, 10, 11, 12. The distances between cities with a surplus and cities with a deficit, are displayed in the matrix below. How should the cars be dispatched so as to minimize the total mileage travelled?

	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}
C_1	4	7	4	5	2	5
C_2	2	3	5	6	8	5
C_3	3	4	6	5	3	3
C_4	4	5	5	5	4	4
C_5	3	4	4	3	3	3
C_6	8	4	4	6	5	3

[14]

OR

A man is engaged in buying and selling identical items, each of which requires considerable storage space. He operates a warehouse of capacity 500 items. He can order on the 15th day of each month, at cost prices shown below, for delivery on the first of the following month. During a month, he can sell any amount up to total stock on hand, in the beginning of the month at market sale prices given below. If he starts the year with 200 items in stock, how much should he plan to purchase and sell each month for the next half-year, so as to maximize his profit.

Month (i)	Jan.	Feb.	March	April	May	June
Cost price (Rs.) c_i per unit	150	155	165	160	160	155
Sales price (Rs.) p_i per unit	165	165	165	175	170	155

[14]

Please turn over

II (alternative) : Elements of Book Keeping and Accountancy

(Answer both the questions)

EITHER

Pass necessary adjustment entries at the Closing date for the following items :

i) Closing Stock Rs.50,000

ii) Depreciation at the rate of 5 percent on Furniture at Book value of Rs.10,000.

iii) Outstanding Insurance Premium Rs.400.

iv) Interest and Rent Receivable outstanding to the extent of Rs.200 and Rs.100 respectively.

v) Balance of Sundry Debtors A/c., was Rs.85,000 and that the reserve for Doubtful Debts A/c., was Rs.2,000. A sum of Rs.5,000 is to be written off as bad. Reserve for Discount on Debtors and Reserve for Doubtful Debts are to be provided at the rate of 5 percent each. [14]

OR

Show in a tabular form (1) what balance each will usually show, Debit or Credit; (2) where and in which side of Trading A/c. Profit and Loss A/c., and Balance Sheet, each of the following will appear :-

(i) income tax paid, (ii) sales tax paid, (iii) returns outward (iv) carriage outward, (v) defence bonds, (vi) donation towards defence fund, (vii) advance from customers (viii) interest on loan given. [14]

EITHER

State with reasons, whether the following items of expenditure are capital or revenue :

i) wages of workmen employed for setting up new machinery ,

ii) legal expenses incurred in connection with raising a loan for purposes of business ,

iii) carriage paid on goods purchased,

iv) payment of damages for breach of a contract to supply certain goods,

v) compensation paid to a retrenched employee for loss of employment,

vi) the cost of demolishing an old building for building a new one in its place,

vii) compensation received from Eastern Railway for compulsory acquisition of land for the purpose of electrification ,

viii) cost of replacing a worn out part of a flat,

ix) purchase of patent rights,

x) cost of installing heating and ventilating apparatus. [16]

2. CR

Rule a form of Petty Cash Book containing analysis columns for the following kinds of petty expenditure :-

stationery, postages, telegrams, carriage, travelling expenses, cleaning, lighting, firing,

together with a column for miscellaneous payments required to be posted direct to the ledger. Enter therein the following transactions bringing down the balance on hand at the conclusion, and showing how the necessary double entry is effected throughout.

1963		
May 1.	Received cheque from the chief cashier for starting balance and cashing same.	550.00
	1. Paid for postages	0.63
	2. Paid for telegrams to Bombay	2.50
	3. Paid for electric light bill upto March	5.06
	4. Paid for cleaning charges	1.12
	5. Paid for charges for fire	0.60
	6. Paid for railway fare to Asansol	3.00
	7. Paid for samples sent to Ray and Co.(carriage charge)	0.63
	8. Paid for postages	0.60
	9. Paid for telegram to Madras	1.00
	10. Paid for new filing cabinet (Debit Office furniture account)	20.00
	11. Paid for bus fares	0.16
	15. Paid for typewriting paper	0.80
	20. Paid for new typewriter (Debit Office furniture A/c)	425.00
	20. Received from chief cashier - cash	150.00
	21. Paid for postages	7.13
	22. Paid for telegrams to Delhi	3.14
	31. Paid for railway fare to Dacca	14.45 [16]

Please turn over

III . Statistical methods in Business

(Answer both the questions)

1. The books of a firm show the inventory to be valued at Rs.1,27,461. There are 1000 inventory items. To confirm the reasonableness of the rupee value of the inventory, a random sample of 12 items was selected and the values of the items checked and corrected if necessary. The following table gives the book values and the audited (correct) values of the items in the sample.

Item No.	book value in rupees (X)	audited value in rupees (Y)
1	117	117
2	96	96
3	213	213
4	290	247
5	260	260
6	93	73
7	198	165
8	81	81
9	83	75
10	327	327
11	110	110
12	87	77

a) using the ratio method of estimation, get an estimate for the total value of the inventory,

b) find the standard error of your estimate. [5 + 5]

2. EITHER

The following data show weekly production in units for ten employees before change (X_1) and after change (X_2) in the production technique.

employee	weekly production in units	
	before change (X_1)	after change (X_2)
A	24	26
B	26	26
C	20	22
D	21	24
E	23	24
F	30	30
G	32	32
H	25	26
I	23	24
J	23	25

Test whether there is any significant change in average production due to the changes in the production technique. [10]

Please turn over

2. CR

The table below gives individual tardiness rates (calculated as the number of times tardy per hundred starts) for 310 employees of a banking organisation.

age group	tardiness rates				
	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20
20 - 30	10	14	18	22	9
30 - 40	19	25	33	26	11
40 - 50	21	16	24	42	20

Test whether the tardiness rate is independent of the age group. [10]

GROUP D: DESIGN AND ANALYSIS OF EXPERIMENTS

(Answer any three questions)

1. You are given below the data from a 2^3 factorial experiment on soybeans, in two 4×4 quasi - Latin squares. The factors are limestone (A), phosphorus (B) and potash (C) and in each case the lower level implies 'no application'.

Square I				Square II			
c	(1)	abc	ab	c	a	abc	b
58	62	60	59	62	65	69	63
a	ac	b	bc	(1)	ac	bc	ab
66	61	60	63	62	63	59	69
abc	bc	(1)	a	abc	b	(1)	ac
64	61	63	59	66	57	64	67
b	ab	ac	c	ab	bc	a	c
58	62	67	67	66	53	58	58

a) identify the effects that are confounded with the rows and columns of the squares, [4]

b) estimate the main effects and the interactions and obtain the expressions for the variances for your estimates, [10]

c) test the individual main effects and interactions and write a small report on your findings. [10]

2. In an experiment to compare four treatments A, B, C, D four tyres were built up each in three parts using three, of the treatments, one for each part. The compositions of tyres together with relative wear values obtained after a suitable length of run, are given below.

Treatments	Tyres			
	1	2	3	4
A	238	196	254	-
B	239	214	-	312
C	279	-	330	419
D	-	308	365	411

Please turn over

a) analyse the data using within tyres comparisons only. [16]

b) obtain an estimate of between tyres variation giving your comments, if any, [8]

c) it was found later that the measure of relative wear on the part of tyre 2 built by using treatment A was obtained by a different procedure and it was considered desirable to ignore it. Examine the hypothesis of equality of treatment effects. [6]

Henderson conducted an experiment in retail grocery stores to determine the effect on McIntosh apple sales (in pounds) due to packaging the apples in 4-pound (treatment A), 6 pound (treatment B) and 8-pound (treatment C) polythene bags. Six grocery stores in central New York were used. Bags of particular types were put on sale in different stores in the three periods as indicated in the table below.

Period	Department Stores						Total for periods
	1	2	3	4	5	6	
first week	72(A)	102(B)	64(C)	164(A)	208(B)	96(C)	706
second week	90(B)	82(C)	72(A)	156(C)	156(A)	112(B)	668
third week	72(C)	92(A)	80(B)	209(B)	178(C)	90(A)	721
Total for stores	234	276	216	529	542	298	2095

Treatment	Treatment total
A	646
B	801
C	648

sum of squares of the observations = 283797
sum of squares of period totals = 1464501
sum of squares of store totals = 839997

The letter in any cell (period X store) indicates the type of bags put on sale and the figures, the actual total sale in pounds during the week. It is assumed that the type of bags used in one week may have some effect on the sales during that week and the next but not thereafter.

Analyse the data and test for treatment differences (test separately for direct and residual effects). You may use the computations supplied. [32]

Give a second order rotatable design involving three quantitative factors. Write down the analysis of variance table giving algebraic expressions for the sum of squares due to the first and the second order terms. Obtain an expression for the variance of the estimated response at specified levels of the factors.

[12 + 12 + 8]

Neatness [4]

Please turn over

GROUP E : SAMPLE SURVEYS (Practical)

(Answer four questions, of which no. 1 is compulsory)

1. Study the information to be collected in the schedule given below and prepare suitable scrutiny and tabulation programmes. [25]

Fertility history of couples

	marital status*	educational standard**	age (years) at formal marriage	age (years) at effective marriage	age (years) at present	age (years) at termination of marriage	marriage duration
husband							
wife							

serial number of children	sex m-1 f-1	interval since previous birth	interval since marriage	age of mother	age of child (years) at present	age of child (years) at death

* never married-1, married-2, widowed-3, divorced-4, separated-5

** illiterate-1, literate but below primary-2, primary-3, middle-4, secondary-5, graduate and above in (i) agriculture-6, (ii) engineering-7, (iii) technology-8, (iv) medicine-9, (v) other subjects-0.

2. To estimate the total cultivated area in Dindigal Taluk of Madurai District, a sample of 10 villages was drawn from the 120 villages in that taluk, with probability proportional to the 1951 census population and with replacement. The table below gives the 1951 census population and cultivated area (in acres) for the 10 sample villages.

serial number of sample village	population (1951 census)	cultivated area (in acres)
1	5511	4824
2	865	1124
3	2535	1648
4	3523	3613
5	8368	3678
6	7357	1566
7	5131	4051
8	4654	3060
9	1146	509
10	1165	2013
	(total population 415149)	

i) estimate the total cultivated area and its sampling variance unbiasedly, [10]

ii) how many sample villages are to be selected to ensure a coefficient of variation of 10 percent ? [5]

iii) use this sample to examine whether there has been any gain in selecting the villages with probability proportional to the 1951 census population, instead of with equal probability. [10]

3. The table below furnishes complete enumeration data on length (X) of strip and volume (Y) of timber for each strip, in 3 blocks of the Black Mountain Forest in California.

Please turn over

block no.	strip no.	x	y	block no.	strip no.	x	y	block no.	strip no.	x	y
1	1	12	762	2	1	9	471	3	1	6	165
	2	12	651		2	9	426		2	6	224
	3	12	461		3	9	448		3	6	192
	4	12	521		4	9	402		4	6	161
	5	12	653		5	9	372		5	5	104
	6	12	544		6	9	372		6	5	94
	7	12	542		7	9	411		7	5	102
	8	12	590		8	9	323		8	5	115
	9	11	533		9	9	381		9	4	110
	10	11	517		10	9	430		10	4	109
	11	11	520		11	9	434		11	4	83
	12	11	539		12	9	394		12	4	36
	13	10	509		13	9	543		13	4	61
	14	10	449		14	9	607		14	4	92
	15	10	492		15	8	416		15	4	74
	16	10	498		16	8	326		16	4	64

(x-length of strip, y - volume of timber)

i) examine the behaviour of the sampling variance of estimates of volume of timber based on systematic samples for different sample sizes (n = 2, 3, 4, 6, 8, 12). [12]

ii) compare the efficiency of systematic sampling with those of simple random sampling with and without replacement, for the sample sizes considered in (i), [8]

iii) also study the efficiency of sampling the strips with probability proportional to the length of the strips with replacement. [5]

Suppose it is required to estimate the rate of incidence of a particular disease in a region having a population of 2500 persons within 5 percent of the true value in the sense that if the sample estimate is k percent, the population proportion is to lie between (k-5)percent and (k+5) percent, with 95 percent confidence.

i) determine the sample size required in case of simple random sampling with replacement, assuming that the sample proportion is normally distributed and that the proportion is likely to be about 60 percent, [10]

ii) suppose the region is divided into two sub-regions having a population of 1000 and 1500 such that the proportions of persons having the specified disease in the two sub-regions are likely to be about 30 percent and 70 percent respectively. Determine the total sample size required for estimating the over-all population proportion with the specified margin of error, if the sampling is done separately from the 2 sub-regions. [15]

5. To study the living conditions of the working class population residing in an industrial region, a stratified two-stage design is proposed where the factories are to be selected with probability proportional to the number of workers obtained in an earlier period with replacement in the first stage and at the second stage a sample of workers is to be selected linear systematically from each sample factory using the current list of workers. The following table gives the number of workers used in the selection of factories and the current number of workers for each of the sample factories.

Please turn over

stra-tum	total no.of workers	sample factory no.	no.of workers frame*	no.of workers current
1	5896	1	99	163
		2	523	465
		3	110	64
		4	741	829
2	23093	1	4200	3504
		2	3187	2527
		3	2215	2186
		4	5322	5285
3	15800	1	2697	2039
		2	4667	6255
		3	1423	1150
		4	1064	1150
4	10774	1	90	91
		2	618	416
		3	150	131
		4	266	282

* frame for selecting factories.

i) determine the constant multiplier for the region as a whole that is to be used in making the sample design self-weighting, so as to ensure a total sample size of about 200 workers; [12]

ii) specify the sampling intervals to be used in each sample factory so as to make the design self-weighting with the constant multiplier determined in (i); [8]

iii) also find approximately the number of workers that would be selected from each sample factory thereby determining the total sample size actually achieved. [5]

GROUP C: STATISTICAL INFERENCE (PRACTICAL)

(Answer any four questions)

1.a) To test the effect of the presence of coal in the sand used for making concrete, 5 different proportions of coal were decided upon, and 4 concrete cylinders were made, containing each of these proportions of coal. The following table shows the breaking strength of these cylinders.

	percentage of coal				
	0	0.05	0.1	0.5	1.0
Breaking strength in 10 lb./sq.inch	169	155	162	172	153
	158	144	145	155	154
	174	164	151	143	156
	168	154	160	144	152

Test whether variability in breaking strength differs significantly with varying proportion of coal present in concrete.

Please turn over

1.b) The following table shows scores in two tests of 15 candidates.

Candidate	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Score in test 1(X)	15	18	8	16	54	95	22	69	75	8	53	26	21	9	18
Score in test 2(Y)	29	16	18	20	20	34	10	31	39	32	35	14	27	10	11

Assuming that the scores X and Y come from a bivariate normal distribution, test whether their variances are equal. [25]

2.a) Three sets of observations from bivariate normal distribution yielded the following coefficients of correlation.

set no.	sample size	coefficient of correlation
1	35	0.62
2	40	0.45
3	25	0.70

Test whether the samples come from populations with the same coefficients of correlation.

b) In the above example, a hypothesis is made that the population from which the first sample was drawn has a coefficient of correlation 0.50.

i) test this hypothesis

ii) draw the power curve of your test by calculating its power for the values 0.10, 0.25, 0.40, 0.50, 0.60, 0.75, 0.90 of the coefficient of correlation. (25)

3.a) Variables X and Y denote age in years and chest expansion in inches respectively. Two groups of children were observed and the following computations made:

group no.	size of group	mean		sums of squares and products (corrected)		
		X	Y	S_{XX}	S_{XY}	S_{YY}
1	12	13.481	1.560	3.416	0.234	8.625
2	16	13.573	3.062	3.759	0.263	13.632

Examine if the linear regression of Y on X in the two groups are (i) identical, (ii) parallel.

b) Each of 10 pots were measured twice for volume by four investigators A, B, C, D, thus yielding 80 volume measurements. The total sum of squares (uncorrected) of these values is 100269. The following table shows the sum of the two measurements taken on each pot.

pots	investigators				pots	investigators			
	A	B	C	D		A	B	C	D
1	78	61	69	90	6	14	35	22	26
2	82	82	75	85	7	102	102	96	101
3	78	66	76	96	8	60	81	72	58
4	24	36	41	39	9	36	59	56	47
5	44	50	49	37	10	90	109	94	97

Please turn over

The sum of squares (uncorrected) of these pair-sums is 198359.

Make out an analysis of variance table showing variation due to pots, investigators, interaction between pots and investigators and error, and write your comments. (25)

4.a) In the following table

col. 1 shows upper limits of class-intervals of an observed distribution,

col. 2 the observed class frequencies and

col. 3 expected class frequencies obtained from an assumed normal distribution.

upper class limits (1)	observed class frequencies (2)	expected class frequencies (3)	upper class limits (1)	observed class frequencies (2)	expected class frequencies (3)
20.5	2	0.6	32.5	102	113.7
22.5	7	3.9	34.5	67	70.2
24.5	15	16.4	36.5	36	29.3
26.5	49	47.6	38.5	5	8.4
28.5	89	93.2	40.5	1	1.8
30.5	138	125.9			
			Total	511	511.0

Calculate the Kolmogorov statistic (D_N) to test whether the assumed normal distribution gives a good fit and examine its significance with the help of the following table which gives for large values of sample size N, pairs of numbers λ and α which satisfy the relation $P\left\{ D_N > \lambda \right\} \leq \alpha$.

α :	0.20	0.10	0.05	0.01
λ :	$1.07\, N^{-\frac{1}{2}}$	$1.22\, N^{-\frac{1}{2}}$	$1.36\, N^{-\frac{1}{2}}$	$1.63\, N^{-\frac{1}{2}}$

b) The observed values of X_1 and X_2 from two paired samples is given below. No assumptions can be made regarding the form of the distributions of X_1 and X_2.

X_1	15	19	31	36	10	11	19	15	10	16
X_2	19	30	26	8	10	6	17	13	22	8

Is it reasonable to assume that the medians of the populations from which the values of X_1 and X_2 arose are equal ? (25)

5. In a certain experiment 18 rabbits each received a high dose of insulin and 18 other rabbits each a low dose. The blood sugar was measured at 1, 2 and 3 hours after each dose and these readings are denoted by X_1, X_2 and X_3. The difference of the means of X_i in the two samples is denoted by d_i $(i = 1,2,3)$. The following table shows the values of d_i and the pooled 'within' sums of squares and products.

Please turn over

	d_1	d_2	d_3
	7.594	19.73	25.64
	X_1	X_2	X_3
X_1	2677	1278	1814
X_2		2358	1966
X_3			3223

a) Find the coefficients of X_1, X_2 and X_3 in the linear function which would best discriminate between the two underlying populations.

b) Calculate the D^2 distance between the two populations and test its significance. (25)
