INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1970

Paper I : Official Statistics and Descriptive Statistics (Theoretical)

Time : 4 hours Full marks : 100

Figures in the margin indicate full marks.

GROUP A

(Attempt any three questions from this group)

1. Describe the need for and the functions of a national statistical system with particular reference to India. (16)

2. Describe the sampling design adopted by the NSS for land utilisation survey and crop-cutting experiments in India. (16)

3. Write short notes on any two of the following :

   i) Central Statistical Organisation,

   ii) Indian Statistical Institute,

   iii) FAO or UNESCO,

   iv) Any State Statistical Bureau (16)

4. Give a brief account of any two of the following index numbers relating to India, mentioning the agencies responsible, item coverage and the base periods :

   i) Index Number of Industrial Production,

   ii) Index Number of Wholesale Prices,

   iii) Index Number of Agricultural Production,

   iv) Index Number of Consumer Prices - Working Class (8+8)=16

5. Describe the main sources of statistics on manufacturing industries in India and indicate the broad contents of the publications. Mention some of the uses of the data and comment on the limitations of the available statistics. (16)

6. Answer any two of the following questions :

   i) Supposing you are required to undertake a study of foreign trade of India, mention the official publications that you may have to consult and the types of data for which you will consult them.

   ii) What are the sources of statistical information on labour and employment in India? Give broadly the contents of these publications.

   iii) Describe the sources of statistics on education in India. Can you state broadly the contents of these publications?

   NEATNESS (2)

GROUP B

(Answer any three questions from this group)

7. (a) Define median of a set of observations.

   (b) If the median of a sample of n_1 observations be m_1 and that of a second sample of n_2 observations be m_2, show that the median of the combined sample of $n_1 + n_2$ observations will definitely lie between m_1 and m_2.

Please turn over

7. (c) (contd.) Show that the sum of absolute deviations of a set of observations from a number 'C' is the least when 'C' is taken to be the median of the set.

(d) Define standard deviation and range of a set of observations.

(e) Show that

$$\frac{R^2}{2n} \leq s^2 \leq \frac{R^2}{4},$$

where R is the range and s the standard deviation, n being the number of observations on which R and s are based. When do the equality signs hold ? (2+3+3+3+3)

8. (a) What is meant by the 'Gram Charlier series expansion' of the cumulative distribution function of a standardised variable ?

(b) Obtain the first few terms of the Gram-Charlier series expansion for the cumulative distribution function of a standardised gamma variable, involving moments up to that of the third order. Give necessary details. (8+8)

9. (a) Explain moment and cumulant generating functions.

(b) Show that for any symmetrical distribution all the central moments and the cumulants that exist (excepting k_1) vanish.

(c) Prove that all cumulants of a Poisson distribution are equal.

(d) Prove that for a normal distribution $k_r = 0$ for all $r > 2$. (4x4)

10. (a) Define multiple and partial correlation coefficients.

(b) Show that

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}\, r_{24.3}}{\sqrt{(1 - r^2_{14.3})(1 - r^2_{24.3})}}$$

(c) Find out the value of $r_{12.34}$ when all total correlations of the variables x_1, x_2, x_3 and x_4 are equal to r. (8+4+4)

11. What is a moving average? Discuss the uses of moving average in various stages of analysis of economic time series.

NEATNESS

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1970

Paper II : Probability Theory and Statistical Methods (Theoretical)

Time : 4 hours Full marks : 100

Figures in the margin indicate full marks.

GROUP A

(Attempt any four questions from this group.)

1. (a) Define the conditional probability $P(A/B)$ of the event A given that event B has occurred.

(b) r balls are distributed at random among n cells. Find the probability

i) that at least one cell is empty,

ii) that all the cells are occupied.

(c) A man selects 2 shoes at random out of 6 pairs of black and 4 pairs of brown shoes. Find the probability that he gets a pair (one shoe for the right foot and one for the left) of the same colour. (2+4+2+4)=12

2. X,Y are two random variables; the marginal density function of X is

$$\frac{1}{2} x^2 e^{-x}, \quad x \geq 0 ;$$

and the conditional density function of Y given $X = x$, is

$$\frac{2y}{x^2}, \quad 0 \leq y \leq x.$$

Find i) the marginal distribution of Y

ii) the conditional distribution of X given $Y = y$. (6+6)=12

3. (a) For the Poisson distribution find the mean, variance, and the third moment about the mean.

(b) Derive the Poisson limit to the Binomial distribution. (1+2+3+6)=12

4. (a) State and prove Tchebysheff's lemma.

(b) An unbiassed coin is tossed 2500 times, and m is the relative frequency of heads. Using the Central Limit Theorem show that

$$P\left\{ \left| m - \frac{1}{2} \right| \leq 0.0196 \right\} = 0.95$$

approximately. (5+7)=12

5. In sampling from a normal population show that :

i) the sample mean is distributed as a normal variable.

ii) the sample variance is distributed as a χ^2 variable.

iii) the distribution of the sample mean and variance are independent. (4+4+2)=10

NEATNESS (2)

Please turn over

GROUP B

(Answer any three questions from this group.)

6. (a) Explain the concepts of 'unbiasedness' and 'consistency' for a point estimate. Give an example of an estimate which is unbiased but not consistent and an estimate which is consistent but biased.

(b) In a sequence of n Bernoulli trials with success probability $p = 1-q$, r successes were observed. Obtain an unbiased estimate of pq and find its variance. Verify whether the estimate is consistent.

(6+10)=16

7. (a) Explain the methods of maximum likelihood and minimum χ^2 in estimation and their relative advantages and disadvantages.

(b) Let $x_1, x_2, \ldots, x_n$ be n independent random variables with a common probability density function:

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} \\ 0, \end{cases} \quad \text{if } 0 \le x \le \theta, \quad \theta > 0 \quad \text{otherwise}$$

Obtain the maximum likelihood estimate of θ.

(c) What is the estimator of θ by the method of moments in (b) ?

(8+5+3)=16

8. (a) Show that for the rectangular distribution

$$dF = \frac{dx}{\theta}, \quad 0 \le x \le \theta$$

and the confidence coefficient $(1-\alpha)$, the confidence limits for θ are R and R/ψ, where R is the sample range and ψ is given by

$$\psi^{n-1}\left\{n-(n-1)\psi\right\} = \alpha.$$

(b) Show that for a 2 x r contingency table

						Total
	a_1	a_2	...	a_i	...	a_r	a
	b_1	b_2		b_i		b_r	b
Total	n_1	n_2		n_i		n_r	n

the value of χ^2 for testing homogeneity is

$$x^2 = \sum_{i=1}^{r} w_i (p_i - p)^2 ,$$

where $p_i = \frac{a_i}{n_i}$, $p = \frac{a}{n}$, $w_i = \frac{n_i}{pq}$, $q = \frac{b}{n}$, $q_i = 1 - p_i$.

(8+8)=16

Please turn over

9. (a) What do you mean by non-parametric methods?
What advantages do they possess over parametric tests ?
Explain how you would use the run test for testing the homogeneity of two samples.

(b) For 20 paired observations without ties what numbers of positive differences would you consider significant at 5 % level.?

(c) If in a sample of 50, there are twenty runs above and below the median, is the sample random? (8+4+4)=16

10. (a) Given a random sample

$$(x_1, y_1), \cdots, (x_n, y_n)$$

from a bivariate normal population with means μ_1, μ_2, variance σ_1^2, σ_2^2 and correlation coefficient ρ (all unknown), describe how you would test

i) $\mu_1 = \mu_2$

ii) $\sigma_1^2 = \sigma_2^2$

iii) $\rho = 0$

(b) Independent random samples of sizes $n_1, n_2, \cdots, n_r$ respectively are available from r univariate normal populations with common variance 1 and unknown means $\mu_1, \mu_2, \cdots, \mu_r$. Derive the likelihood ratio test for the hypothesis $H_o : \mu_1 = \cdots = \mu_r$. (9+7)=16

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1971

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

Time : 4 hours | Full marks : 100

i) Figures in the margin indicate full marks.
ii) Use of calculating machines is not permitted.

GROUP A

(Attempt any three questions from this group)

1. (a) Describe the advantages of a sample survey in comparison with complete enumeration.

   (b) Define clearly the terms 'sampling unit' and 'sampling frame'.

   (c) Explain the use of random sampling numbers in the selection of a random sample.

   (6+4+4) = 16

2. (a) Explain the method of systematic sampling and discuss its advantages and disadvantages.

   (b) How would you estimate the sampling variance of the estimate of the population total based on a systematic sample?

   (c) Compare the efficiency of systematic sampling with that of simple random sampling.

   (6+5+5) = 16

3. (a) What is ratio method of estimation ? Obtain approximate expressions for the bias and variance of a ratio estimate. State clearly the assumptions involved.

   (b) Express the exact bias of a ratio estimate of the population mean of a variable Y (based on a simple random sample of n units selected with replacement) in terms of the covariance of the ratio $\hat{y} / \hat{x}$ and $\hat{x}$, where x is the auxiliary variable, $\hat{x}$ and $\hat{y}$ being unbiased estimates of the population mean of x and y.

   (c) Obtain an unbiased ratio type estimator of the population mean.

   (8+4+4) = 16

4. (a) If a sample of n units is selected with probability proportional to a given measure of size x and with replacement, suggest an unbiased estimate of the population total of the variable y. Derive its sampling variance and obtain also an unbiased estimate of this variance.

   (b) What are the 'non-sampling errors' in a sample survey and suggest steps for reducing them.

   (8+8) = 16

Please turn over

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

(i) Proportional and Optimum allocation.

(ii) Regression method of estimation.

(iii) Cluster sampling.

(iv) Double sampling

(v) Inter-penetrating sub-samples.

(16)

Neatness (2)

GROUP B

(Attempt any three questions from this group)

1. (a) Discuss the need for designing an experiment? What are the basi principles required to be followed?

(b) Cows of 4 different breeds are to be differentiated in respect o their average yield of milk. Describe in detail how an experim is to be designed to meet the requirements of the case.

(4+4+8) =

2. (a) Define a latin square. What is meant by a set of orthogonal latin squares?

(b) Explain in detail the analysis of a latin square design when an observation is missing.

(c) Calculate the loss of efficiency due to the 'missing' observatio

(2+2+8+4)

3. (a) What is meant by a connected design? Why is connectedness considered to be an essential criterion for incomplete block de Can square lattice design be considered a connected design?

(b) Outline the method for the intra-block analysis of a square lat design.

(c) Obtain the variance of the estimates of differences between eff of any two varieties of a square lattice design :

(i) when the varieties occur together in some block, and

(ii) when the varieties do not occur together in any block

(2+2+8+2+2) =

4. Define a balanced incomplete block design with parameters v, b, r, k, Establish the various well-known relationships among these parameters Derive the intra-block analysis of such a design and with the help of sums of squares which have to be calculated for this purpose, derive the inter-block analysis of the same design.

(2+4+8+4) = 16

5. (a) What is meant by a fractional replication of a factorial desig How is it related to confounding?

(b) Obtain the $\frac{1}{4}$ replicate of a 2^6 experiment retaining the main effects in 4 blocks of 8 plots each.

(c) Indicate the form of the analysis of variance table, appropria for this design.

(2+2+8+4) = 16

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1970

Paper IV : Applied Statistics (Theoretical)

Time : 4 hours [Any two halves of 50 marks each for groups A to E, are to be attempted.] Full marks : 100

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

(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

(Attempt any three questions from this group)

1. Give a brief description of the method you would adopt for the construction of index numbers of production and employment in manufacturing industries. If these two series of index numbers are given for different months over the last ten years, how will you study the changes in labour-productivity?

(16)

2. What do you mean by the trend of a time series? For what types of series do you calculate trends by (a) moving averages and by (b) polynomial fitting? If a polynomial trend is given, how do you use it for forecasting?

(16)

3. Explain the concept of income-elasticity of demand. How do you estimate it on the basis of cross-section data? Give some illustrations of the use of this elasticity figure.

(16)

4. Give a short description of an input-output table and indicate how it can be used for projection purposes. How do you estimate the change in employment when production targets for various sectors are known? How far are these estimates reliable?

(16)

5. Explain Pareto's law of income distribution. What types of income distribution are described by this law? How will you calculate the concentration ratio, after fitting a Pareto distribution?

(16)

Neatness : (2)

GROUP B : STATISTICAL QUALITY CONTROL

(Attempt any three questions from this group)

1. (a) Discuss the situation when 'modified control limits' may be used. How could you get some economic advantage in such situations?

Please turn over

Q.1 (contd.)

1. (b) Describe briefly a method of fitting a trend line and control limits when tool-wear is present.

(c) When can the 'Moving Average and Moving Range' charts be used in preference to the conventional $\bar{X}$ - R Charts. Give a few examples of such cases from any industry known to you.

(4+6) = 1

2. (a) Discuss the relative merits of attributes and variable inspection.

(b) State the acceptance criteria for single sampling acceptance plan for variables inspection for both the single and double specification limit cases, in which the standard deviation of the quality characteristic being not known, the mean range based on sub-samples has to be used.

(6+10) = 1

3. (a) Describe the precautions to be taken in designing and conducting a plant-scale experiment.

(b) Explain with the help of two examples from any industry known to you, the use of nested design in industrial experimentation. How would you analyse the results of such an experiment?

(6+10) = 1

4. Write short notes on any four of the following :

(a) Method of steepest ascent to determine optimum operating conditions.

(b) Group control charts.

(c) Rational sub-grouping.

(d) Simon's I.Q. charts.

(e) Narrow limit gauging.

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

(4 x 4) =

Neatness : (2)

GROUP C : STATISTICAL METHODS IN GENETICS

(No candidate available)

Please turn over

GROUP D : VITAL STATISTICS AND DEMOGRAPHY

(Attempt any three questions from this group)

1. You have been asked to estimate by the component method all-India female population for 1966, 1971, 1976 and 1981 (with 1961 as base year), by quinquennial age groups.

(a) What considerations would you take into account for projecting fertility and mortality rates?

(b) How will you make the actual estimates?

Explain in detail all the symbols you would use for your computations.

(6+10) = 16

2. (a) Enumerate briefly the different factors which influence fertility rates.

(b) Define gross and net reproduction rates. Examine their limitations as measures of the true reproductivity of a population.

(6+10) = 16

3. (a) Explain briefly the method adopted by the Census Actuary for the construction of Indian life tables.

(b) Establish the relationship

$$ {}_nq_x = \frac{2n\ {}_nm_x}{2 + n \cdot {}_nm_x}, $$

the symbols having their usual significance.

(12+4) = 16

4. Explain the concepts of stationary and stable populations. What is a quasi-stable population? How does it differ from a stable population?

Explain the significance of the equation,

$$ R_o = e^{rT} $$

defining carefully the meaning of all the symbols involved in the above equation.

(6+2+2+6) = 16

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

(a) Vital registration in India

(b) 'de-facto' and 'de-jure' population enumeration

(c) Standardisation of death rates.

(8+8) = 16

Neatness : (2)

GROUP E : EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Attempt question number 5 and any two from the others)

1. Describe the errors of measurement, substitution and prediction in psychological tests, giving appropriate mathematical formulations and derive the standard deviations of these errors. Explain the meaning of the symbols used in your mathematical expressions.

(16)

2. Describe in brief Spearman's two-factor theory and Thurstone's multiple-factor theory of factor analysis. How do you obtain factor loadings in each of the above two cases?

(16)

3. Briefly describe the computational procedures and properties of any two of the following methods of standardising raw scores :

(a) Percentile scores, (b) T-scores and (c) Z-scores.

(8+8) = 16

4. What do you understand by 'item validity' and 'item difficulty'? What are the item validity indices which involve the slope of the regression of item on test? How is item difficulty corrected for guessing?

(16)

5. (a) Write a short note on the effect of test length on the coefficient of validity.

(b) Prove the Spearman - Brown formula for the reliability of a test which is 'n' times the original length. Carefully state the assumptions in the derivation of the formula.

(c) Obtain a function of test reliability and test length which is invariant with respect to these two variables, for a given value of reliability of the original test.

(6+6+6) = 18

———oooOooo———

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1970

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 (24 marks)

(Attempt any three questions from this group)

(a) The cubes of a number of integers are given below :

n	n^3
7	343
9	729
11	1331
13	2197

Calculate the value of $(12.75)^3$ by using a suitable interpolation formula.

(b) Given

$$f(1)+f(2)+f(3) = 25,$$

$$f(4) = 29,$$

and $$f(5)+f(6) = 113$$

estimate the value of $f(10)$. (4+4)=8

For a certain country the death rates per 100,000 from Tuberculosis, are recorded below for a number of years. The death rate figures for 1914 and for 1915 however are not available. Estimate these values, stating clearly the underlying assumptions.

year	death rate
1911	159.0
1912	149.8
1913	148.7
1914	-
1915	-
1916	143.8
1917	147.2

(8)

Given the following table of values

x	262	263	264	265
log x	2.4183013	2.4190557	2.4216039	2.4232459

x	266	267	268
log x	2.4248816	2.4265113	2.4281318

determine the value of

$$\int_{262}^{268} \log x \, dx$$

by using a suitable quadrature formula. (8)

Please turn over

4. Solve the following set of simultaneous equations :

$$14x_1 + 5x_2 + 6x_3 + 8x_4 = 75$$
$$9x_1 + 19x_2 + 4x_3 + 4x_4 = 74$$
$$5x_1 + 8x_2 + 22x_3 + 5x_4 = 103$$
$$9x_1 + 13x_2 + 4x_3 + 18x_4 = 113$$

GROUP B

(Attempt question no.1 and any two from the others)

5. In a newspaper account, on the incidence of influenza among tubercular persons living in the same family, the following paragraph appeared :

"Exactly a fifth of the 1,00,000 inhabitants showed signs of tuberculosis and no fewer than 5,000 among them, had an attack of influenza, but among them only 1,000 lived in infected houses. In contrast with this, 1/15th of the tubercular persons who did not have influenza were still exposed to infection. Altogether 21,000 were attacked by influenza and 41,000 were exposed to the risk of infection, but the number having an attack of influenza but not of tuberculosis and living in houses where no other cases of influenza occurred, was only 2,000."

Present the above information in an appropriate tabular form and comment after making the necessary calculations.

6. (a) In the course of an experiment, 15 mosquitoes were put in each of 120 jars and subjected to a dose of D.D.T. After 4 hours the number alive in each jar was counted and the following frequency distribution was obtained :

no. of mosquitoes alive	0	1	2	3	4	5	6	7	8
frequency (no. of jars)	2	12	14	22	28	17	13	10	2

Find the corresponding expected frequencies on the assumption that each mosquito has a common probability of survival and test for the goodness of fit.

(b) If 150 mosquitoes were put in a jar and the dose of D.D.T. had been so adjusted that the common probability of survival after four hours for a mosquito, could be assumed the same as in (a), estimate the probability that at least a third of the mosquitoes would be found alive after 4 hours ? (16+4)=20

7. Examine the consistency of the following correlation matrix :

	x_1	x_2	x_3
x_1	1	0.01	0.33
x_2		1	0.01
x_3			1

(b) The solubility of nitrous oxide in nitrogen dioxide for temperature (T) ranging from 263 to 283 degrees absolute was determined. The reciprocal temperature is expressed in terms of (1000/T). A number of independent determinations were made at each temperature and the solubility results expressed in terms of reciprocal temperature, were as follows :

	Reciprocal temperature values				
	3.801	3.731	3.662	3.593	3.533
Solubility (percentage by weight)	1.20 1.33 1.52	1.21 1.27	1.11 1.04	0.81 0.82	0.65 0.59 0.63

Fit a linear regression equation of solubility on reciprocal temperature. Test whether the true regression deviates from linearity. (5+15)=20

For the following data on Sales (thousand rupees) of a firm,

i) fit a linear trend to the yearly totals

and ii) calculate seasonal indices by the ratio-to-trend method.

year	Quarter			
	I	II	III	IV
1960	35	45	41	39
1961	39	57	55	49
1962	44	63	59	53
1963	59	80	72	67
1964	84	96	90	87

(20)

GROUP C (26 marks)

(Attempt all questions from this group)

9. EITHER

With the help of the official publications supplied, obtain the number and value of nickel coins, minted at the Calcutta mint during the latest available three-year period. Collect also the corresponding figures for the Bombay mint. Arrange them in a suitable tabular form and make a brief statistical report on comparative fluctuations.

OR (14)

From the official publications supplied, compile neatly, for the five latest available years, figures on the total sanctioned strength of civil police force in India and the cost of maintaining the same. Comment briefly on the salient statistical features of the data.

10. From the official publications supplied, compile data on any three of the following :

i) Index number of value of exports (by sea, air, land) from India for six latest available months.

ii) Export of tea from India and Ceylon during the 3 latest available years.

iii) Distribution of total live-births for Calcutta by age of mother, during the latest available year.

iv) Number of newspapers and periodicals published during the 3 latest available years in India. (12)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – May 1970

Paper VI : Statistical Methods; Design & 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 any two questions from this group)

1. (a) A certain stimulus administered to each of 12 patients resulted in the following increase in blood pressure :

5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4 and 6.

Can it be concluded that the stimulus will on an average be accompanied by an increase in blood pressure ?

(5)

(b) The regression of saving (S) of a family over income (Y) may be expressed as $S = A + mY$ where A and m are constants. In a sample of 100 families, the variance of saving is one quarter of the variance of income and the coefficient of correlation between S and Y is found to be 0.4. Obtain the estimate of m. Given further that the variance of saving is 4.3, test for the significance of m, stating clearly necessary assumptions.

(2+2+2) = 6

(c) The following records the operating life in hours of several electric bulbs :

980, 992, 1005, 998, 1010, 1020, 1028, 1042

It is known that life -times are distributed as

$$f(t) = \frac{\alpha}{\theta} t^{\alpha-1} e^{-(t\alpha)/\theta}$$

Get the maximum likelihood estimate of θ, given $\alpha = 2.5$ and give some idea about its standard error.

(5)

(d) What is the smallest ratio (λ) of two variances (λ-1) that can be detected with power 0.95 by an F-test with two samples of size 10, the size of the test being 0.05?

(4)

2. (a) Five samples of seasoned mine-props were tested for maximum load. The means and standard deviations of the maximum load (in units of 1000 lbs. wt.) were as follows :

sample No. (i)	sample size (n_i)	$\bar{x}_i$ (mean)	s_i (S.D.)
1	20	42.0	10.10
2	24	52.0	12.00
3	30	65.5	5.45
4	25	51.8	9.54
5	31	73.5	10.26

Please turn over

Q.2(a) (contd.)

Test if the variation from sample to sample in the mean level of maximum load is significant. Is there a significant difference between sample 1 and sample 5?

(8)

(b) Rutherford and Geiger (1910) obtained the following distribution of the number (x) of α -particles emitted from a disc in 7.5 sec. Fit a suitable distribution to the observed frequencies and test for th goodness of fit.

x	f	x	f
0	57	7	139
1	203	8	45
2	383	9	27
3	525	10	10
4	532	11	4
5	408	12	2
6	273	$\geq$ 13	0

(7)

(c) Test if the age-distribution of white fish in Lake Wabamn, Alberta changed significantly between 1957 and 1958 from the catches in these two years classified below in age-groups :

	Age (Years)					
year	3-4	5	6	7	8	$\geq$9
1957	6	15	10	38	62	26
1958	10	12	9	22	36	5

(5)

3.(a) Use the method of least square to fit the curve $y = ax^2 + \frac{b}{x}$ to the following data :

x	1	2	3	4	5	6
y	-1.51	0.09	3.88	7.66	15.32	26.08

(10)

(b) Construct a sequential test for the mean μ of a Poisson distribution to test $\mu = 1.8$ against $\mu = 2.2$ taking $\alpha = 0.04$ and $\beta = 0.02$, where α, β have their usual significance. Find the operating characteristic of the test for $\mu = 2.0$. If the following are the values of x observed on the first 10 occasions, what inference, if any, can you make regarding μ ?

n (occasion)	1	2	3	4	5	6	7	8	9	10
x_n (value of x)	1	0	2	0	1	2	2	1	0	1

(10)

Please turn over

GROUP B

(Attempt any two questions from this group)

1. EITHER

Construct the lay out (with necessary randomization) of a 2^4 factorial design in 8-plot blocks in two replicates, confounding the highest order interaction in both the replicates.

OR

Construct the symmetric balanced incomplete block (SBIB) design with

$$v = b = 7, \quad r = k = 4, \quad \lambda = 2.$$

(8)

2. EITHER

The following are the yields of 25 varieties of soya-beans laid out in a simple lattice design.

Replicate 1

(1) 6	(2) 7	(3) 5	(4) 8	(5) 6
(6) 16	(7) 12	(8) 12	(9) 13	(10) 8
(11) 17	(12) 7	(13) 7	(14) 9	(15) 14
(16) 18	(17) 16	(18) 13	(19) 13	(20) 14
(21) 14	(22) 15	(23) 11	(24) 14	(25) 14

Replicate 2

(1) 24	(6) 13	(11) 24	(16) 11	(21) 8
(2) 21	(7) 11	(12) 14	(17) 11	(22) 23
(3) 16	(8) 4	(13) 12	(18) 12	(23) 12
(4) 17	(9) 10	(14) 30	(19) 9	(24) 23
(5) 15	(10) 15	(15) 22	(20) 16	(25) 19

[Each row represents a block, the figure within brackets the treatment and the other figure the yield.]

Analyse the above data.

OR

An experiment on sugarcane conducted in 5 randomised blocks gave the following values of number of plants per plot (x) and weight of cane in kg. (y). The data on number of plants provide a basis for error control through analysis of covariance. The three treatments are manures – nitrogen (N), phosphorous (P) and potash (K) in appropriate doses.

block	N		P		K	
	x	y	x	y	x	y
1	41	122	41	81	42	80
2	40	120	50	80	38	82
3	38	138	46	79	54	65
4	41	121	42	75	40	58
5	39	126	48	83	45	76

Analyse the data. In case of significance, judge the individual treatment differences with a view to finding the best treatment.

Is the measurement of the concomitant variable worthwhile?

(22)

GROUP C

(Attempt any two questions from this group) .

3.(a) The following data relate to the number of households (X) in each of 44 villages in a region. Select from this list 5 villages with replacement and with probability proportional to X.

sl.no.	X	sl.no.	X	sl.no.	X	sl.no.	X
1	20	12	38	23	49	34	43
2	27	13	41	24	20	35	42
3	51	14	63	25	27	36	20
4	18	15	43	26	103	37	21
5	73	16	72	27	207	38	30
6	20	17	35	28	21	39	20
7	81	18	19	29	29	40	70
8	118	19	113	30	32	41	35
9	43	20	128	31	43	42	39
10	105	21	31	32	51	43	20
11	181	22	37	33	36	44	63

(b) Assuming that the population figures for the 5 villages so selected in a particular sample are as follows, estimate the total population and the standard error of the estimate :

sample village (serial no.)	4	8	23	27	43
X	18	118	49	207	20
population	101	517	238	913	87

(7+8) = 1

4. The table below presents the summary of data obtained from complete census of all 2010 farms in a region. The farms are stratified into 7 strata. The number of farms (N_i), the population value of strata means ($\bar{Y}_{Ni}$) and those of strata standard deviations (S_i) for different strata are given below:

stratum	farm size (in acres)	number of farms (N_i)	average area under wheat ($\bar{Y}_{Ni}$)	standard deviation of area under wheat
1	0 – 40	394	5.4	8.3
2	41 – 80	461	16.3	13.3
3	81 – 120	301	24.3	15.1
4	121 – 160	334	34.5	19.8
5	161 – 200	169	42.1	24.5
6	201 – 240	113	50.1	28.0
7	More than 240	148	63.8	35.2

Compute the sampling variance of the estimated area under wheat the region for a sample of 150 farms

(1) if the sample is simple random without replacement and without stratification; and

(2) if the sample is stratified random without replacement within each stratum, with allocation proportional to N

(7+8)

Please turn over

5. From a region containing 1000 villages and of total cultivable area of 9,65,832 bighas, a simple random sample of 64 villages were drawn. The frequency table below, shows the total cultivable area and the area under wheat in these villages.

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

Obtain the ratio and regression estimates of the total area under wheat in the region and the standard error of the estimates.

(5+10) = 15

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1969

Paper VII : Applied Statistics (Practical)

Time : 3 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 options.

(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

(Attempt all questions from this group)

A series of Index Numbers of production and employment in industries for the years 1948-56 are given. Study the changes in production, employment and labour productivity on the basis of these figures.

year	Index numbers	
	production	employment
1948	107.0	100.8
1949	110.3	101.7
1950	119.0	102.6
1951	119.0	103.3
1952	118.0	102.8
1953	128.3	103.8
1954	130.7	104.8
1955	133.7	106.6
1956	140.3	107.0

(12)

Per capita demand in pounds, for meat (Y), retail price in cents per pound (X_1) and per capita disposable income in dollars (X_2), are given below for the years 1928-41. Calculate the elasticities of demand with respect to price and income.

year	Y (Pounds)	X_1 (Cents)	X_2 (Dollars)
1928	69.6	30.3	682
1929	67.0	29.1	684
1930	68.4	23.7	515
1931	70.7	15.6	390
1932	69.6	13.9	361
1933	63.1	18.8	411
1934	48.4	27.4	459
1935	55.1	26.9	517
1936	55.8	27.7	551
1937	58.2	24.8	600
1938	61.7	22.2	539
1939	73.5	19.3	576
1940	68.4	24.7	697
1941	65.8	25.3	552

(14)

Please turn over

3. For the income distribution shown below, draw a Lorenz curve and calculate the concentration ratio.

income-grade (Rs. thousands)	earners (number)	total income (Rs Crores)
below 3	3,779	0.3
3 - 5	1,60,190	189.3
5 - 7.5	3,33,776	203.8
7.5 - 10	1,52,320	132.0
10 - 12.5	82,316	92.2
12.5 - 15	56,774	77.6
15 - 17.5	35,283	57.2
17.5 - 20	23,663	44.1
20 - 25	20,504	59.0
25 - 30	15,227	41.6

income-grade (Rs. thousands)	earners (number)	tot inc (Rs C
30 - 40	15,252	5
40 - 50	7,592	3
50 - 60	3,690	2
60 - 70	2,216	1
70 - 100	2,535	2
100 - 200	1,618	2
200 - 300	327	
300 - 400	123	
400 - 500	67	
over 500	112	1

NEATNESS

GROUP B : STATISTICAL QUALITY CONTROL

(Attempt any two questions from this group)

1. (a) It is desired to control the mean weight of a cigarette at 1.21 gm. A 10 % increase in the weight is to be detected with a probability of 0.95. On the other hand the probability of wrongly deciding that the mean weight has increased should not be more than 0.01. Determine the sample size and control limits given that standard deviation of the process is 0.04 gm.

(b) Obtain 95 % confidence limits for the lot standard deviation, given that in a sample of 10 packets of margarine the standard deviation of percentage moisture, was 4.3.

(c) Construct 99 % confidence limits for average number of defects per item in lot, given that in a sample of 20 items inspected a total number of 22 defects were noted.

(d) A process is in control at $\mu = 18$ and $\sigma = 2$. A control chart based on a subgroup size of $n = 4$ is set up using 3σ limits. If a shift in the mean of 2.5 units occurs, what is the probability of the next $\bar{X}$ falling outside the control limits. (4+4+4+6).

2. The upper specification limit for an engineering product is 60 units In lot by lot inspection of this product for proportion defective, the AQL and the LTPD are specified respectively as 0.001 and 0.08 with the producer's and consumer's risks respectively as 0.05 and 0.10.

(a) Design an acceptance criterion for single sampling by variables, using the sample mean and the sample standard deviation.

(b) Devise a single sampling plan by attributes for the same conditions.

(c) Which of the above two plans you would prefer if the cost of inspection by measurement per unit, is twice that of examination by attributes. (6+6+6).

Please turn over

3. An experiment was conducted in a rubber factory with a view to reduce buckling in hose pipes. The factors inducing buckling are Coil length (C), sheet length (S) and degree of tension (T) each at three levels. For each combination, 4 coils were observed and the total number of buckles in the four coils is given below.

Tension (degree)	Coil length	Sheet length S_1	S_2	S_3
T_1	C_1	2	4	1
	C_2	6	4	4
	C_3	8	5	2
T_2	C_1	2	0	3
	C_2	1	0	0
	C_3	1	2	0
T_3	C_1	5	5	0
	C_2	6	0	5
	C_3	2	0	1

Comment upon the experiment and the results, making any analysis that you deem appropriate. (24)

NEATNESS (2)

GROUP C : STATISTICAL METHODS IN GENETICS

(No Candidate Available)

GROUP D : VITAL STATISTICS AND DEMOGRAPHY

(Attempt any two questions from this group)

1. (a) A country had (in terms of thousand) 775 births in 1964, 800 births in 1965, 820 births in 1966, 847 births in 1967 and 390 births in 1968. In the Census taken on 1st January 1969, 3310 thousand children aged 0-4 years were enumerated.

Find the percentage of under enumeration of children aged 0-4 years making use of the following extract from the life table, having a radix of 10,000.

x	l_x
0	9,105
1	8,550
2	8,305
3	8,102
4	8,023

(b) Given that the complete expectation of life at ages 30 and 31 for a particular group are 21.39 and 20.91 years respectively and that the number living at age 30 is 41,176 find the number that attains the age 31.

Please turn over

1. (c) (contd.) Below is given a skeleton life-table :

age x	no. of survivors at age x l_x	life table population in the age group (x to x+10) $10L_x$	life table population above age x T_x
5	681	6,350	-
15	590	5,670	-
25	543	5,052	-
35	461	4,116	10,833

Calculate

i) the complete expectation of life at age 15;

ii) the probability that out of two persons aged 15, at least one will die before reaching the age 35. (8+8+8)=24

2. The following table shows the growth of sunflower plants in a certain locality. Fit a logistic curve to the data.

Growth of Sun Flower plants

days	mean height (in cm.)	days	mean height (in cm.)
7	17.9	49	205.5
14	30.3	56	228.3
21	67.8	63	247.1
28	98.1	70	250.5
35	131.0	77	253.8
42	169.5	84	254.5

(24)

3. The following table gives the female specific fertility rates and the proportion of females surviving from birth to mid-point of the age group :

age group (in years)	age specific fertility rate	proportion surviving (from birth to mid-point of age group)
15 - 19	.0553	.5790
20 - 24	.1220	.5555
25 - 29	.1146	.5250
30 - 34	.0957	.4840
35 - 39	.0731	.4390
40 - 44	.0342	.3865
45 - 49	.0104	.3370

On the basis of the above data, calculate the true rate of natural increase (r) in a stable population. (24

NEATNESS (5

Please turn over

GROUP E : EDUCATIONAL & PSYCHOLOGICAL STATISTICS

(Attempt questions 1 and 2, and any two from the rest)

1. What is the estimated ultimate limit of the predictive effectiveness of a test for an external criterion with which it now correlates 0.55 and has a reliability coefficient of 0.85? (10)

2. The frequencies of the responses to an aptitude test item and later probationary status in job of 500 persons, are presented below :

performance during period of probation	item responses	
	correct	incorrect
Satisfactory	361	59
Not satisfactory	39	41

Estimate the coefficient of correlation between the two dichotomously expressed variables. (10)

3. The following is a 'centroid' factor loading matrix.

Variable	Common factor coefficient C_1	C_2	Communality h^2
1. height	.830	.300	.846
2. arm-span	.818	.460	.889
3. length of forearm	.777	.470	.825
4. length of lower leg	.798	.401	.798
5. weight	.786	.500	.868
6. bitrochantric diameter	.672	.438	.661
7. chest-girth	.594	.444	.550
8. chest-width	.547	.333	.529

Obtain the rotated factor loading matrix, after suitable orthogonal rotations of the centroid factor loading matrix. (14)

4. General Science examination marks and aptitude scores of 30 pupils are presented below. Obtain the rank-order correlation coefficient between the two sets of scores.

sl. no.	general science	aptitude test	sl. no.	general science	aptitude test
1	130	20	16	178	35
2	132	24	17	172	30
3	152	28	18	165	28
4	142	23	19	160	27
5	184	37	20	148	25
6	190	32	21	180	34
7	150	25	22	149	25
8	170	23	23	138	36
9	181	29	24	167	29
10	161	35	25	162	27
11	175	32	26	145	23
12	135	22	27	150	29
13	147	24	28	160	30
14	162	26	29	172	31
15	136	21	30	154	30

(14)

5. The scores of 15 boys and 15 girls on an aptitude test are given below. Sex is the dichotomous variable x and the aptitude test is the quantitative variable y. Compute a coefficient of correlation between x and y.

Write a brief note on the coefficient of correlation you make use of and comment on the data furnished.

boys	aptitude test	girls	aptitude test
1	50	1	33
2	42	2	32
3	32	3	35
4	43	4	25
5	25	5	30
6	60	6	47
7	45	7	57
8	40	8	42
9	35	9	43
10	47	10	35
11	62	11	55
12	56	12	40
13	52	13	35
14	48	14	43
15	40	15	45

NEATNESS

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1970

Paper VIII : Subjects of Specialisation - I

Time : 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 - Special Paper I)

(Answer any five questions from this group)

1. What are the assumptions necessary to set up tests of significance and confidence intervals based on least squares estimators $\hat{\beta}_i$ coefficients in a linear regression model:

$$y_t = \sum_{i=1}^{k} \beta_i x_{it} + u_t \qquad t = 1, \ldots, n.$$

Describe the procedure for setting up the confidence interval. (5+15)=20

2. (a) Show that the prices for the goods relative to wages in a Leontief System, can be obtained through the coefficients for the labour inputs.

(b) Derive national income from the input-output table for an economy. (12+8)=20

3. Describe Pareto's equation for income distribution. Discuss its limitations.

Derive the expression for the Lorenz measure of inequality for the distribution. (10+10)=20

4. In the following, the observed values of the variables X and Y are assumed to contain errors of measurement

$$X = x + u$$

$$Y = y + v$$

where x and y are the true values and u and v are the errors of observation. Suppose that the true values are connected by the exact relation

$$y = \alpha + \beta x .$$

Can we use the least squares method to estimate the parameters in the linear model derived from above relating X and Y?

Discuss your answer fully stating necessary assumptions. (20)

5. "A model in which all variables are present in all relations is statistically a hopeless problem".

Discuss in the light of this statement why a priori exclusion of variables is a necessary condition for identification in a simultaneous equation system. (20)

Please turn over

6. Discuss the problem of multicollinearity between two regressors.

Can you suggest an appropriate method for estimating the associated regression coefficients ?

What are the conditions necessary for such a method to be valid ? (10+10

7. Discuss two different algebraic forms of Engel curves and describe the methods of estimation of the coefficient for income elasticity of demand by fitting each curve to family budget data.

Discuss the difference in the properties of the coefficients derived from the two curves. (15+5

8. For a Cobb-Douglas production function

$$x = A n^{\alpha} k^{\beta}$$

where 'x' is output, 'n' labour input, 'k' capital input and 'A' is a constant,

i) Obtain the elasticities of output with respect to labour and capital inputs;

ii) Show that the sum of the exponents α and β indicates the degree of 'returns to scale' in production ;

iii) Obtain relations between average and marginal productivities of labour and capital; and

iv) Show that the wage share of the value of output is constant when real wage equals marginal productivity (in a competitive market).

GROUP B : TECHNO-COMMERCIAL STATISTICS

(Statistical Quality Control - Special Paper I)

(Attempt <u>any five questions</u> from this group)

1. (a) Describe a situation involving order of production as a basis for rational sub-grouping. Illustrate also with suitable examples how the order of production is not always a sufficient basis for sub-grouping.

(b) Discuss the considerations that prevail in the choice of the frequency and size of the sub-groups.

(c) Show how to construct the power curve of the $\bar{x}$ chart, when it is assumed that the process standard deviation does not change.

(d) When would you recommend the use of 'modified control limits' for an $\bar{x}$ chart? Explain how such limits are constructed and used in practice. Indicate some possible alternative courses of action available in such a situation. (+ + +8)=20

Please turn over

2. (a) Distinguish between specification limits and tolerance limits.

(b) Derive the necessary expressions to construct tolerance limits such that for the following two cases, it can be asserted with 100β per cent confidence that they will include at least 100 α per cent of the population (where α and β are preassigned numbers between 0 and 1):

i) a normal parent population

ii) a parent population with an unknown form (assume a continuous probability density function). (4+6)=20

3. (a) Distinguish between the Type A and Type B operating characteristic curves of an acceptance sampling inspection plan, for attribute characteristics.

(b) For an inspector, the errors of misclassification have been found to be as follows :

i) the probability of classifying a non-defective item as defective $= p_1$

ii) the probability of classifying a defective item as non-defective $= p_2$

Show that if π is the proportion of defectives in submitted product the effective proportion defective $\pi_{eff.} = p_1 + \pi (1 - p_1 - p_2)$.

(c) If $p_1 = 0$, what would be the effect of the inspector's error on the operating characteristic curve of the sampling plan ?

(d) Explain the basis for the following statement :

"the long term process average outgoing percent defective, may rarely be expected to exceed one-half the AOQL value associated with the inspection plan in use". (5+5+5+5)=20

4. (a) Derive expressions for the probability of acceptance and the average amount of inspection for a double-sampling attributes inspection plan, assuming that the rejected lots are subjected to a 100 per cent screening inspection.

(b) For a double-sampling attributes inspection plan, show that the average sample number, when inspection is curtailed on the second sample is given by the following expression:

$$ASN = n_1 + \sum_{m=c_1+1}^{c_2} P\,(m \text{ defectives in } n_1)\left[n_2 P(c_2 - m \text{ or less defectives in } n_2) + \frac{c_2 - m + 1}{\pi} P(c_2 - m + 2 \text{ or more defectives in } n_2 + 1)\right] \quad \text{for all values of } \pi,$$

where π is the fraction defective in presented lots

n_1 = size of the first sample

n_2 = size of the second sample

c_1 = acceptance number for the first sample

c_2 = acceptance number for combined samples

and $P(\ldots)$ = probability of the event specified in the bracket. (5+5+10)=20

5. (a) Discuss the advantages and limitations of acceptance sampling by variables as compared to attribute inspection.

(b) For the MIL-STD-414 Sampling procedures for inspection by variables for percent defective, what is the acceptability criterion for the variability unknown Range method for the double specification case, when different AQL values are specified for Upper and Lower specification limits ? Summarise neatly the steps.

(c) What are the criteria used for installing tightened and reduced inspections? Explain the steps involved and clearly state how the appropriate sampling plans are to be obtained from the published tables.

(d) What are the conditions under which mixed variables and attributes sampling plans are to be used? How are the plans to be selected and what are the criteria of acceptability?

(4+5+5+6)=20

6. (a) If n_c points of the cube of type $(\pm 1, \pm 1, \ldots \pm 1)$ are used along with $n_s = 2k$ points of the star of type $(\pm \alpha, 0, \ldots 0)$, $(0, \pm \alpha, \ldots 0) \ldots (0, 0, \ldots \pm \alpha)$, to form a central composite design, prove that

$$\alpha = \sqrt{n_c} \quad \text{for rotatability.}$$

(b) Explain how to obtain an orthogonal blocking arrangement for a central composite design for k = 4 factors? Indicate the experimental combinations to be carried out in each block.

(c) An experimenter decided to explore the relationship between the yield of a chemical product and two independent variables, and it was expected that the response function might be approximated by the second order model :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 + \epsilon$$

where the observed response -

y = purity of a chemical intermediate

x_1 = temperature of the reacting mass

x_2 = rate of addition of reagents

The factor levels chosen are :-

x_1 (temp) : 15°c, 30°c, 45°c, 60°c, 75°c

x_2 (rate) : 1 hr, 2 hrs, 3 hrs.

Experimental material available was sufficient to run only 10 trials.

Suggest a suitable second order design and give the lay-out of the experimental plan.

(3+6+6)=20

Please turn over

7. Write short notes on any four of the following :

a) Mid-Range chart

b) Cumulative Sum charts

c) Distribution free tolerance limits

d) Lot-quality estimation

e) Bulk sampling

f) Variation flow analysis. (5+5+5+5)=20

GROUP C : BIOMETRIC METHODS

(Special Paper I)

No candidate available

GROUP D : DESIGN AND ANALYSIS OF EXPERIMENTS

(Statistical Aspects - Special Paper I)

No candidate available

GROUP E : SAMPLE SURVEYS

(Theoretical Aspects - Special Paper I)

(Attempt four questions from this group)

1. (a) With a view to study a rare attribute it is decided to continue drawing a random sample of units without replacement, until a pre-fixed number of 'm' units possessing the rare attribute are found. If the finite population correction can be ignored, show that the probability that the total sample required is of size 'n' is

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

where P is the proportion of the rare attribute in the population.

(b) Also show that the average size of the total sample is $\frac{m}{P}$, and that $p = \frac{m-1}{n-1}$ is an unbiased estimator of P.

(c) Suggest a suitable sampling design and indicate the estimation procedure for estimating the survival rate of a dynamic population, stating clearly the assumptions under which your method holds good.

(8+8+9)=25

Please turn over

2. (a) If a cluster of elements is to constitute a sampling unit, what are the considerations involved in the choice of such a sampling unit.

(b) A population consists of M clusters of N units each. The variance within clusters (s_w^2) follows the empirical law $s_w^2 = AN^g$ where A and g are positive constants. Obtain the optimum size of the cluster, if for a simple random sample of m clusters the cost function is of the form $C = c_1 mN + c_2\sqrt{m}$

(c) If the clusters are of unequal size, show that the sample mean becomes a ratio estimate. Compare its precision with that of a simple random sample of the same expected number of units.

(d) Can a cluster sample be more efficient than a simple random sample of the same expected number of units ? Explain the situation with an example. (5+7+8+4)=25

3. (a) Enumerate the alternative ways of utilising ancillary information on an auxiliary variable correlated with a character under study in large scale sample surveys.

(b) Taking any of these ways as an example, show how the procedure can be extended to a case where information is available on two auxiliary variable related to the principal character under study.

(c) Suppose 'm' independent interpenetrating sub-samples of the same size 'n' are selected according to the same sampling design and unbiased estimates $\{\hat{Y}_i\}$ and $\{\hat{X}_i\}$ (where i = 1, 2, ..., m) of the population totals Y and X of two characters are obtained. The following two ratio estimators of $R = Y/X$ can be constructed.

$$\hat{R}_1 = \frac{\hat{Y}}{\hat{X}} \left(\hat{Y} = \frac{1}{m}\sum_{i=1}^{m} \hat{Y}_i, \hat{X} = \frac{1}{m}\sum_{i=1}^{m} \hat{X}_i\right)$$

and $$\hat{R}_2 = \frac{1}{m}\sum_{i=1}^{m} Y_i, \; X = \frac{1}{m}\sum_{i=1}^{m} \frac{\hat{Y}_i}{\hat{X}_i}$$

Obtain expressions for estimating approximately the bias in them.

(d) Using $\hat{R}_1$ and $\hat{R}_2$ construct an estimator which is relatively free from bias. (5+5+10+5)=25

4. (a) What are the relative merits of systematic sampling when compared with simple random sampling ? Indicate how systematic sampling can be considered as a special case of 'cluster' sampling.

(b) Deduce the sampling variance of the mean of a systematic sample and indicate a procedure for getting an approximate estimate of the sampling variance from a single such sample.

Please turn over

4.(c) (contd.) A population consists of N = nk units arranged in n rows and k columns. The value Y_{ij} of the element in the ith row ($i = 1, 2, \ldots n$) and jth column ($j = 1, 2, \ldots, k$) is given by the linear law :

$$Y_{ij} = \mu + (ik + j - k)\lambda$$

where μ and λ are constants.

A systematic sample of n units is drawn by selecting all the units in a random column. A stratified sample is drawn by randomly selecting one unit from each row.

Compare the sampling variances of the means obtained from the two samples. (5+10+10)=25

5. (a) What is a self-weighting sampling design ? Indicate its merits and also its drawbacks, if any.

(b) Show how a stratified two-stage sampling design for estimating population total can be made self-weighting at field stage, by using 'probability proportional to size' sampling technique.

(c) Suppose a non-self-weighting design was adopted for estimating population total. Indicate how you can make the design self-weighting at the tabulation stage.

(d) Obtain an estimator of the increase in the sampling variance with the modification. (6+7+7+5)=25

6. Write critical notes on <u>any four</u> of the following :

(a) Method of double sampling for stratification.

(b) Multi-stage and multi-phase sampling designs.

(c) Weighted regression estimators.

(d) Sampling on successive occasions.

(e) Use of cost functions in planning sample surveys.

(f) Methods to control non-sampling errors in sample surveys.

(25)

Paper VIII : Group F : Techniques of Computation
(Numerical Analysis - Special Paper I)
No candidate available

Paper VIII : Group G : Statistical Inference
(General Theory - Special Paper I)
No candidate available

Paper VIII : Group H : Probability Theory
(Special Paper I)
No candidate available

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1970

Paper IX : Subjects of Specialisation - II (Theoretical)

Time : 4 hours Full marks : 100

(a) Candidates are required to answer questions from that group only for which they have registered their option.

(b) Figures in the margin indicate full marks.

Group A : Economic Statistics - (Indian Economics and Economics of Planning) - Special Paper II

Section I - Indian Economics

(Attempt any three questions from this section)

1. Analyse the major factors governing the changes in Indian agricultural output in the 1960's. What are the main obstacles against a rapid spread of the "Green Revolution"?

(16)

2. "The increase in Indian exports recorded in 1968-69 was entirely the result of industrial recession combined with massive government subsidies for exports and cannot be repeated in the immediate future". Examine this statement carefully.

(16)

3. Examine how far industrial workers have benefited from the labour policy pursued by the government of India, after the abolition of compulsory arbitration as a general policy.

(16)

4. Discuss how far the nationalisation of commercial banks can be expected to curb the growth of monopoly in India.

(16)

5. Discuss the relative merits of

(i) a progressive income tax,

(ii) a progressive land tax, and

(iii) the enforcement of unfavourable terms of trade for agriculture with a ceiling on the size of holdings,

as alternative means of taxing agriculture in India.

(16)

6. Analyse the trend of Indian industrial production since 1951, distinguishing clearly between capital goods and consumer goods industries.

Discuss in this connection the problem of choice of a suitable price index for deflating the money value of industrial production.

(16)

Please turn over

Section II – Economics of Planning

(Attempt any three questions from this section)

7. Show, with the help of the Mahalanobis model or any other multi-sector model of planning, how national income, investment and consumption behave over time, if the aim is to effect a permanent increase in the rate of growth of national income. (16)

8. Show, with the help of a theoretical model, or of a concrete case, how the existence of surplus labour in an economy might influence the objectives and the instruments of planning. (16)

9. Discuss the utility and limitations of shadow prices for planning India's foreign trade. (16)

10. Show how lags in the formulation and execution of plans would affect the benefits expected from a programme of import substitution in basic industries.

11. How would you modify the proposed outline of the Indian Fourth Five Year Plan so as to reduce India's dependence on foreign aid and foreign loans? (16)

12. Write notes on any two of the following :

    (a) the warranted rate of growth;

    (b) the optimum rate of saving;

    (c) the strategy of unbalanced growth;

    (d) a block-triangular input-output matrix. (8+8)

Neatness (Sections I and II) (4)

GROUP B : TECHNO-COMMERCIAL STATISTICS – (II)

Section I : Operations Research (70 marks)

Section I : (Alternative) : Elements of Book-keeping and Accountancy (70 marks)

Section II : Statistical Methods in Business (30 marks)

Section I : Operations Research

(a) Use a separate answer book for this section

(b) Attempt any four questions from this section

An item has five months selling period with the probability distribution (given overleaf) of selling prices in each month.

(Q.1 contd. on

Probability distribution of selling price in each month

price	month				
	1	2	3	4	5
4	.10	.05	.05	.05	.05
5	.10	.10	.15	.05	.25
6	.20	.15	.30	.35	.30
7	.30	.25	.15	.25	.20
8	.15	.20	.15	.15	.15
9	.10	.15	.10	.10	.05
10	.05	.10	.10	.05	.00

(a) Calculate the expected price for each month.

(b) Faced with these probability distributions for the price over the demand season, use a method of Dynamic programming to determine an optical selling-policy.

NB: [We would expect such an optical policy would consist of rules telling that if in a given month the actual price exceeded certain predetermined value, we should sell everything during that month.]

(4+13) = 17

2.(a) Discuss 'Duality' in Linear Programming and its uses.

(b) Describe the application of Linear Programming in production scheduling.

(10+7) = 17

3.(a) Describe the mathematical structure of transportation, transhipment and assignment problems.

(b) Examine critically the different methods of solving them.

(10+7) = 17

4.(a) Outline the computational problems of a generalized (n) job (m) machine scheduling problem.

(b) State the rule and the assumptions underlying the "Johnson's Rule" for (2) machine and (3) machine cases.

(c) Show how the Akers and Freedman's Graphical Approach to (2) ..ob (n) machine problem works.

(d) Relate the structure of (b) and (c) above to (a).

(4+4+5) = 17

5.(a) Indicate the difference between replenishment and reorder system in Inventory Control.

(b) If the demand as well as the lead-time follow an explicit random pattern, e.g. normal distribution, show how you would proceed to obtain the distribution of demand during the lead-time.

(c) If several products are to be processed in a single machine, how would the Economic Manufacturing Quantity vary from the case when the machines are unlimited.

(5+5+7) = 17

Please turn over

6. Suppose there are two activities with durations T_1 and T_2 to be begun simultaneously in a project. Indicate how the following probabilities can be calculated :

(i) $P(T_1 > T_2)$

(ii) $P\{\text{Max}(T_1, T_2) < t\}$

(iii) $P\{\text{Min}(T_1, T_2) < t\}$

and (iv) also the expected value of

$$E\{\text{Max}(T_1, T_2) - \text{Min}(T_1, T_2)\}$$

(4+4+4+3)

Neatness (2)

Paper IX : Group B (Alternative) : Elements of Book-keeping & Accountancy

(No candidate available)

Section II : Statistical Methods in Business

i) Use separate answer book for this section.

ii) Attempt any two questions from this section.

1. A factory has two machines both of which make 1/8 inch steel wa Thickness of independent samples from each machine were measur nearest thousanth of an inch, with the following results.

Machine I : .117, .130, .122, .139, .132

Machine II : .120, .127, .133, .115, .137, .12

Are the mean thicknesses significantly different?

(10)

2. A purchaser of dolls from a store observed that 8 out of 150 dolls manufactured by a company A were defective, while only 4 out of 130 manufactured by a company B were defective. Is it reasonable to suppose that the two percentages represent a co level of quality?

(10)

Please turn over

3. In an attempt to judge the ability of a certain forecaster to correctly forecast cloud coverage, 24 hours in advance, the following information was compiled from a series of independent trials.

category observed	category forecast			total
	cloudy	scattered	clear	
cloudy	32	28	3	63
scattered	8	55	24	87
clear	1	8	36	45
total	21	91	63	175

(a) Is there a significant degree of association between the forecast and observed categories of cloudiness?

(b) What observations would you make about the forecaster's ability?

(6+4) = 10

GROUP C : BIOMETRIC METHODS - Special Paper II
(Statistical methods in Genetics & Bio-assays)
(No candidate available)

GROUP D : DESIGN & ANALYSIS OF EXPERIMENTS - Special Paper II
(Combinatorial Aspects)

(Attempt any five questions from this group)

1. Define the term 'primitive mark' of a finite field. Find the primitive marks of GF (11).

Construct a GF (3^2) and provide the corresponding addition and multiplication tables.

(2+8) = 10

2. How would you construct a complete set of mutually orthogonal Latin Squares of side 's' where 's' is power of a prime?

Let $s = p_1 p_2$ where p_1 and p_2 are two distinct primes. Let r = minimum of $p_1 - 1$ and $p_2 - 1$. Then show that a set of r mutually orthogonal Latin Squares exists.

(10+10) = 20

Please turn over

3. Define (i) Balanced incomplete block design (BIBD)

(ii) Resolvable balanced incomplete block design (RBIBD)

Prove the following inequalities:

(i) $r \geq K$ (in the case of BIBD)

(ii) $r \geq k+\lambda$ (in the case of RBIBD)

If N is the incidence matrix of symmetrical BIBD, show that

$N^{-1} = (N' - \frac{\lambda}{r} E_{vv})/r -$ where N' is the transpose of N and E_{vv} is the matrix of order v x v with positive unit elements everywhere. Hence show that in a symmetrical BIBD any two blocks have λ varieties in common.

(4+8+8) =

4. When is an incomplete block design said to be partially balanced with two-associate classes? Establish the various relationships between the parameters of two-associate partially balanced incomplete block design (PBIBD).

Write the parameters of group divisible design.

(12+8) = 2

5. Show that the existence of symmetrical BIBD with the parameters $v = b$, $r = k$, λ implies the existence of the following BIB designs with the parameters v', b', r', k', λ' and v'', b'', r'', k'', λ'' where

(i) $v' = v-k$, $b' = b-1$, $r' = r$, $k' = k-\lambda'$, $\lambda' = \lambda$

(ii) $v'' = k$, $b'' = b-1$, $r'' = r-1$, $k'' = \lambda$, $\lambda'' = \lambda - 1$ 1.

Let 0, 1, 2, ..., 10 be the elements of GF (11). Show that the set of elements (1, 3, 4, 5, 9) forms a difference set. Hence construct (a) symmetrical BIBD with $v = b = 11$, $r = k = 5$, $\lambda = 2$

(10+10) = (b)

6. Explain the terms 'confounding' and 'partial confounding' using 2^3 symmetrical factorial design.

Construct a $(3^4, 3^2)$ design such that the main effects and first order interactions are not confounded.

Explain how EG (m, s) may be employed for the construction of confounded designs.

(4+8+8) = 2.

Define the term 'alias' in factorial experiments with fractional replication. (a)

Give a method of construction of 2^kth replicate of 2^m symmetrical factorial design. (b)

Construct half-replicate of 2^6 design involving 6 factors A, B, C, D & F with confounding, to reduce block size to 8, such that ABC = DEF, ABD = CEF, CD = ABEF, are confounded with the blocks. (c)

(3+9+8) =

Please turn over

8. (a) Let X_n be a square matrix of order n, $(n > 2)$, consisting of ± 1 as elements, such that $X_n X_n' = nI_n$ where I_n is the n^{th} order identity matrix.

Show that a necessary condition for the existence of X_n is that $n \equiv 0 \mod 4$.

Show also that the existence of X_n of order $n = 4t$ where t is +ve integer, implies the existence of symmetrical BIBD with the parameters $v = b = 4t + 3$, $r = k = 2t+1$, $\lambda = t$.

(b) Seven plants chosen for an experiment have each three leaves growing serially along the stem. Plan a scheme of allocation of seven treatments to the leaves such that each pair of treatment occurs once on the same plant and each treatment occurs once on each leaf position.

(12+8) = 20

GROUP E : SAMPLE SURVEYS - Special Paper II
(Organisational Aspects)

(Attempt any five questions from this group)

1. Suppose a consumption survey is to be conducted in a State during 1970-71.

(a) Discuss the general lay-out of the schedules and draft instructions for some important aspects of the schedule.

(b) Discuss the relative advantages and disadvantages of employing

(i) ad-hoc whole-time staff,

(ii) permanent whole-time staff, and

(iii) part-time services of the employees of different Government and non-Government organisations,

for conducting the field-work of the said survey.

(9+5+2+2+2) = 20

2. You are asked to conduct a sample-survey of the non-organised sector of manufacturing establishments in the urban areas of a State.

(a) What are the various problems which you should consider at the stage of planning the survey?

(b) What arrangements would you propose to make for public relations and propaganda?

(c) Draw up an organisation chart of field-staff you propose to employ for the purpose.

(12+4+4) = 20

3. (a) Discuss carefully the points that are to be taken into consideration in writing the report of a large-scale sample survey.

(b) Illustrate your answer with an example.

(12+8) = 2

4. (a) Describe a suitable method for evaluating differences between investigators in a crop survey conducted by direct physical observation, with details of its administrative implications.

(b) Describe briefly the structure of the budget of a large-scale sample-survey.

(c) Enumerate the different stages of compilation-work for construction of the consumer price-index.

(6+7+7) = 2

5. (a) Describe a suitable method for constructing the daily index of performance of a punching/verifying operator. How do you propose to use it for administrative control?

(b) Discuss the relative advantages and disadvantages of the alternative procedures indicated below :

(i) Centralised versus decentralised field-scrutiny,

(ii) Centralised versus decentralised training of field-staff.

(7+3+5+5) = 2

6. An un-employment survey was conducted in the urban areas of a State using census-blocks as first-stage sampling units and households, as second-stage units. Estimates are to be prepared for the percentages of employed and unemployed amongst males and females of the working force.

(a) Prepare a suitable punching design.

(b) Describe a method of interpenetrating sub-samples for scrutinising the data.

(c) How will you use dummy cards to ensure that the punched cards have been properly verified?

(d) Describe the various stages of processing work.

(e) Draw the flow-chart of processing work.

(5 x 4) =

7. Write notes on any four of the following :

(a) Border bias in crop-cutting experiments

(b) Bias due to sponsoring agency

(c) Completion-report on field-surveys.

(d) Quad

(e) The testing of schedules

(f) Methods of checking the final tables.

(4 x 5) =

GROUPS F (Techniques of Computation), G (Statistical Inference), & H (Probability Theory)

(No candidates available for these groups)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1970

Paper X - Subjects of Specialisation - III (Practical)

Time : 5 hours | Full marks : 100

GROUP A : ECONOMIC STATISTICS

(Practical - Special Paper III)

(Attempt any three questions from this group)

1. The following data are based on a family budget enquiry covering the whole of urban India during October 1953 - March 1954 :

monthly per capita total consumer expenditure (Rs.)	estimated percentage of population (to the nearest unit)	estimated average expenditure per capita per month (Rs.)	
		foodgrains	all items
0 - 8	8	3.78	6.24
8 - 11	12	4.45	9.36
11 - 13	9	5.25	11.02
13 - 15	9	6.00	14.01
15 - 18	11	6.83	16.27
18 - 21	10	6.94	18.98
21 - 24	8	7.40	22.59
24 - 28	8	6.71	25.64
28 - 34	5	8.27	30.67
34 - 43	8	8.44	38.15
43 - 55	5	7.36	48.70
55 -	7	8.60	80.33

Estimate the Engel elasticity for foodgrains consumption at the overall average of per capita total consumer expenditure, assuming that the Engel curve is of the form $y = \alpha + \beta \log x + \frac{\gamma}{x}$, where y = per capita foodgrains demand and x = per capita total consumer expenditure, and α, β, γ are constant.

[Use weighted least squares method, taking the percentages of population as weights.] (24+8)=32

2. The following measures of labour input (L), capital input (K) and production (P) in British industry, are given :

year	$\log_{10}L$	$\log_{10}K$	$\log_{10}P - 1$
1871	0.423	0.242	0.832
1876	0.447	0.300	0.863
1881	0.471	0.375	0.908
1886	0.503	0.387	0.908
1891	0.535	0.443	0.987
1896	0.564	0.491	1.017
1901	0.593	0.568	1.045
1906	0.631	0.620	1.093
1911	0.669	0.669	1.114

Fit a Cobb-Douglas production function and estimate the standard errors of the estimated exponents of labour and capital.

How are these standard errors altered if the same function is fitted, assuming constant returns to scale ? (14+6+10)=32

Please turn over

3. Consider the following model of a meat market

$$y_1 = a_1 y_2 + a_2 x_1 + u_1 \quad \text{(demand)}$$

$$y_1 = b_1 y_2 + b_2 x_2 + b_3 x_3 + u_2 \quad \text{(supply)}$$

where the endogenous variables are

y_1 = meat consumption per person in lbs

y_2 = meat price index;

and the predetermined variables are

x_1 = disposable income per person (£),

x_2 = unit cost of meat processing (index),

and x_3 = cost of agricultural production (index).

The following gives the corrected sum of squares and products matrix based on 23 yearly observations :

	y_1	y_2	x_1	x_2	x_3
y_1	1370	-352.4	3672	-536.5	983.9
y_2		1531.5	8355	850.3	1236
x_1			83434	3612	12205
x_2				2535	730.8
x_3					2627

Estimate the meat demand function by two-stage least squares and estimate the standard errors of the estimated coefficients $\hat{a}_1$ and $\hat{a}_2$. (20+12)

4. A firm can produce three different products in different proportions subject to the availability of two factor inputs. The requirements of these inputs per unit of output are as follows (in suitable units)

input type	product type 1	2	3
1	3	4	1
2	1	3	3

The total availabilities of the two factor inputs are 25 and 30 respectively. Find the optimal product mix (that is, the quantity to make of each product, to maximise profit) given that profit per unit of output is Re.8, Re.19 and Re.7, respectively for product types 1, 2, and 3. What is the maximum value of profit? (8+20+4)

Please turn over

5. (a) Examine, by quick graphical tests, whether the size distribution of population by per capita total consumer expenditure presented in Q 1 above, is approximately

i) of the loglogistic form or

ii) of the three-parameter lognormal form with threshold at Rs.4/=.

(b) The following is a condensed inter-industry transactions table for a certain economy:

Producing Sectors	Using sectors (S)	(A)	(I)	Final use	Total output
Services (S)	21	24	96	62	203
Agriculture (A)		29	118	103	250
Industry (I)		24	183	350	550
Primary Input	102	173	153	823	

[Note : All figures are in appropriate units.]

Compute the matrix of input-output coefficients, and find the total outputs of the three sectors if the final use vector changes to (50, 125, 400) (12+8+14)=32

NEATNESS (4)

GROUP B : TECHNO-COMMERCIAL STATISTICS
(Special Paper III - Practical)

I. Statistical Quality Control (50)

II. Operations Research/Elements of Book-keeping and Accountancy (30)

III. Statistical Methods in Business (20)

Section I. Statistical Quality Control

(Attempt any two questions from this section)

1. Certain types of resistive components for the electronic industry are produced by an automatic process, at the rate of 1200 components per hour. The resistance of these components must not exceed a specified value of 265 ohms. The components which have a resistance greater than 265 ohms are considered 'defectives' and have to be scrapped. The resistance of the components increase with the passage of time, due to deterioration of some of the parts of the process.

Samples of four components were taken from the process every quarter of an hour. When reset, the process produces components, initially, with a mean value of $\bar{x}$ = 250 ohms. The average and range values of these resistance measures for 25 samples are given below :

Please turn over

1. (contd.) Average and Range values of measures of resistance for 25 samples

Sample no.	Average (ohms)	Range (ohms)	Sample no.	Average (ohms)	Range (ohms)
1.	255.5	15	14	247.3	65
2	254.5	7	15	266.0	31
3	254.8	20	16	269.3	15
4	253.0	5	17	271.3	10
5	248.3	25	18	263.8	10
6	245.5	37	19	270.0	17
7	255.8	20	20	275.0	13
8	250.3	9	21	270.5	10
9	258.3	7	22	273.0	18
10	257.3	27	23	274.0	4
11	263.3	5	2	201.8	31
12	258.8	3	25	270.8	49
13	254.3	30			

(a) Test the linearity of the regression trend line.

(b) Work out a suitable control chart scheme and recommend the period after which the process is to be reset so that no defective is produced. (10+15).

2. Under an acceptance-rejection inspection scheme, items were selected one by one at random from a submitted lot and were classified as defectives or non-defectives. It is desired to develop an item by item sequential sampling inspection plan in the above case, such that,

i) lots containing 5 percent defectives would be accepted in 90 percent of cases, whereas

ii) lots containing 15 percent defectives would be accepted only in 20 percent of cases.

(a) Obtain the equations for the acceptance and rejection lines and hence prepare a table of acceptance and rejection numbers for sample sizes from 5 to 30, at intervals of 5.

(b) Sketch an approximate OC curve of the above plan.

(c) Under the above plan, suppose that the rejected lots are screened 100 percent and accepted. Plot an approximate AOQ curve and obtain the value of AOQL of the plan. (8+3+9).

3. Nine aluminium alloys were tested for their resistance to corrosion in a chemical plant atmosphere. Three sites in the factory were chosen and on each of them a plate made from each alloy was exposed for a year. The plates were then submitted to four observers, who assessed their condition visually and awarded marks to each from 0 to 10 according to the degree of resistance to corrosion. The observers worked independently and the plates were submitted to them in random order. The data as furnished are shown overleaf :

Please turn over

3. (contd.) Data on corrosion resistance for nine aluminium alloys

Site	Observer	Alloy								
		1	2	3	4	5	6	7	8	9
I	A	5	5	5	4	6	6	1	6	7
	B	4	6	5	4	5	3	1	5	7
	C	7	7	7	7	9	5	4	7	7
	D	6	5	4	5	7	6	3	6	7
II	A	8	7	7	7	5	4	5	4	5
	B	7	8	6	7	8	5	3	7	8
	C	9	8	8	8	8	6	7	8	8
	D	8	6	7	7	5	5	7	4	5
III	A	4	4	5	3	4	3	0	5	5
	B	3	5	3	2	5	2	0	4	5
	C	5	5	5	6	6	4	3	7	8
	D	3	5	7	2	3	3	1	6	6

Analyse the data and state clearly your conclusions. (25)

Section II. Operations Research

(Attempt any two questions from this section)

1. Consider the following linear programming problem

minimise $z = 3x_1 - 2x_2 + 4x_3$

subject to i) $3x_1 + 5x_2 + x_3 \geq 7$

ii) $6x_1 + x_2 + 3x_3 \geq 4$

iii) $7x_1 - 2x_2 - x_3 \leq 10$

iv) $x_1 - 2x_2 + 5x_3 \geq 3$

v) $4x_1 + 7x_2 - 2x_3 \geq 2$

vi) $x_1, x_2, x_3 \geq 0$

(a) Write the dual problem

(b) Solve the dual and hence solve the primal problem. (15)

2. There are three marketing areas and the yearly profit obtained in any area depends on the number of salesmen allotted. The results appear below.

No. of salesmen allotted in area	Profit obtained (in thousand rupees)		
	Area I	Area II	Area III
0	38	40	60
1	41	42	64
2	48	50	68
3	58	60	78
4	65	66	90
5	72	75	102

Suppose we have a total strength of 5 salesmen. Allot the salesmen to areas using dynamic programming, to maximize the profit. (15)

Please turn over

3. The following table gives the demand distribution for a certain spare part of an expensive equipment during its working life-time. Suppose that the cost of the spare is C = Rs.2000/ when purchased with the original equipment and that the loss suffered for each spare that is not available when needed is u = Rs.10,000/-. Further, assume that the spare, if unused, has a salvage value of v = Rs.1000/-.

Probabilities (p_d) of actually needing 'd' spare parts during the working life-time of the equipment are given as

d	P_d
0	0.923
1	0.040
2	0.020
3	0.010
4	0.005
5	0.002

Determine the optimal number of spare parts that should be purchased.

Section II (Alternative) : Elements of Book-keeping and Accountancy (Practical III)

(No candidate available)

Section III : Statistical Methods in Business

(Attempt question No.1 and any other from this Section)

1. A concern manufacturing ice-cream divides a working year (December to next November) into four quarters as follows :

1st quarter : December, January and February (worst quarter)

2nd quarter : March, April and May (trade picking up)

3rd quarter : June, July and August (peak months)

4th quarter : September, October and November (trade falling off)

Over a period of four successive years the sales (in thousands of Kg) were :

Year	Quarter			
	I	II	III	IV
1	1	2	5	2
2	1	3	6	2
3	1	3	6	4
4	2	2	8	4

(a) Plot the data and carry out a time series analysis for long term trend, seasonal variation and residual variation.

(b) Predict the sale for the four seasons in the year immediately following the fourth year. Assume that the uncertainty in each case is measured by 50 percent of trend value. (CS1)*

Please turn over

2. In a painting competition the various entries are ranked by two judges. Compute rank correlation coefficient to test whether there is significant agreement among the judges.

Entry	A	B	C	D	E	F	G	H	K	L
Judge X	5	2	6	8	1	7	4	9	3	10
Judge Y	1	7	8	10	4	5	3	8	2.	9

(1)

3. A company had four salesmen A, B, C and D each of whom, was sent for a week into three types of area : country area (C) Outskirts of a city (O) and shopping centre of a city (S). The sales figures (in Kg per week) were distributed as follows :

Area	Salesmen			
	A	B	C	D
C	30	70	30	30
O	80	50	0	70
S	100	60	80	80

Carry out an analysis to test the claim 'all the salesmen are equally good'. (6)

Group C : Biometric Methods
Paper III – Practical
(No candidate available)

Group D : Design and Analysis of Experiments
Paper III – Practical
(No candidate available)

GROUP E : SAMPLE SURVEYS
(Special Paper III – Practical)

(Attempt any three questions from this group)

1. An industrial area contains 300 factories from which a sample of 10 factories has been selected with probability proportional to the number of workers (x) and with replacement. The following table gives the output (y) of the sample factories in the year 1968–69 in thousand rupees.

sample factory	number of workers x	factory output 'y'	sample factory	number of workers x	factory output 'y'
(1)	(2)	(3)	(1)	(2)	(3)
1	47	153	6	60	174
2	10	28	7	35	101
3	25	65	8	20	82
4	19	44	9	75	200
5	12	31	10	23	52

Total number of workers in the Area = x = 12,000.

Please turn over

1. (a) (contd.) Estimate the total factory output unbiasedly and also obtain an estimate of its relative standard error.

(b) Estimate the efficiency of the above sampling scheme compared to simple random sampling without replacement. (5+12+12)=32

2. In order to estimate the average household size $\bar{H}$ $(=Y/X)$ in an urban area, a sample of 20 blocks was selected in the first stage, with probability proportional to 1961 census population and with replacement. From each such sample block, a sample of households was then selected linear systematically. The sampling intervals used for household selection came to be so specified that the design became self-weighting with the constant multiplier 4.0 for each sample household.

The following table gives the number of sample households (x) and number of persons (y) in them for each sample block.

sample block	sample household	number of persons in sample household	sample block	sample household	number of persons in sample household
(1)	(2)	(3)	(1)	(2)	(3)
1	7	33	11	3	14
2	1	6	12	8	45
3	2	8	13	3	33
4	10	43	14	2	9
5	8	30	15	9	48
6	4	13	16	5	23
7	5	27	17	1	7
8	6	27	18	4	25
9	2	11	19	7	33
10	1	3	20	3	16

(a) Estimate the average household size $\bar{H}(=Y/X)$

(b) Estimate the bias and relative standard error of the estimate in (a). (5+10+14)

3. The agricultural farms of an intensive cultivation area were divided into two strata by their sizes, and a simple random sample of farms was selected without replacement from each stratum. The table below gives the area under wheat for each sample farm in acres.

stratum	number of farms		wheat area in sample farms
	total	sample	
I	40	4	125, 108, 149, 135
II	130	10	37, 64, 41, 29, 59, 44, 71, 50, 43, 63

(a) Estimate unbiasedly the average wheat area per farm, and also the relative standard error of this estimate.

(b) Estimate the efficiency of stratified sampling compared to unstratified simple random sampling, without replacement. (4+14+10)

Please turn over

4. It is proposed to conduct an enquiry on small-scale manufacturing establishments owned and operated by households in the rural and urban areas of India, with a view to collect information on the following aspects :

i) nature of products

ii) man-days worked in 1969-70; man days lost with reasons

iii) details of household labour and hired labour, working in the establishment,

iv) capital assets,

v) availability of raw material, power, labour,

vi) storage, transport and marketing facilities,

vii) details of input and output

viii) loan

ix) problems of expansion etc. and any other important aspect.

Draw a suitable Schedule for this purpose and write a short-note on the concepts and definitions used in the schedule. (24+8)=32

5. A Radio Listeners' Preference Survey, is to be conducted in the urban areas of West Bengal with a view to collecting information on the listeners' likes and dislikes about the radio programmes and their suggestions about the same.

Prepare a suitable scheme for the survey giving details of sampling design, programme of field work, and requirements of field staff. Also prepare a budget indicating the expenditure on field work. The budget should take into consideration the following points.

(a) journey time

(b) time for preparation of sampling frame

(c) proper enquiry time for filling schedules

(d) scrutiny of 50 % schedules

(e) provision of adequate number of investigators, inspectors and supervisors.

(f) setting up of four zonal offices for controlling the field work

(g) salaries of field staff and office staff

(h) printing of schedules, forms etc.

(i) other items of contingent expenditure (16+16)=32

NEATNESS 14

GROUP F : TECHNIQUES OF COMPUTATION

(Special Paper III - Practical)

No candidate available

Please turn over

GROUP G : STATISTICAL INFERENCE

(Special Paper III - Practical)

(Attempt Question No.1 and any other three from the rest)

1. The following table gives the inter-correlations, means and standard deviations of three tests X_1, X_2, X_3 administered to a group of 164 students. The criterion C, called the 'quality criterion' was a measure of the excellence of the students' project work.

	Mean	Standard deviation	Inter-correlation			
			C	X_1	X_2	X_3
C	61.2	9.62	-	.365	.526	.518
X_1	58.3	7.20	-	-	.243	.260
X_2	63.4	5.25	-	-	-	.628
X_3	68.7	6.47	-	-	-	-

(a) Compute the multiple correlation coefficient of C on X_1, X_2, X_3, and test if it is significantly different from zero.

(b) Examine if the inclusion of X_3 has any significant effect on the multiple correlation coefficient.

(c) Build up a linear formula for estimating C on the basis of X_1, X_2 and X_3.

2. A gunner is shooting at a circular target from a given distance. If the point of hit is (x, y) with respect to a coordinate system (x - axis horizontal and y - axis vertical) with the origin at the centre of the target, it is known that x and y are independent and normally distributed with 'zero' as mean values and the same variance σ^2. Let R denote the distance of the actual point of hit from the centre of the target. For 10 independent shots fired, the values of R are :

4.7, 4.3, 3.6, 5.1, 4.0, 4.8, 3.9, 4.5, 4.2, 5.0 .

(a) Test the hypothesis H_0 : $(\sigma = 3.5)$ against H_1 : $(\sigma > 3.5)$

(b) If the mean value of the distribution of R be denoted by $E(R) = \theta$, test the hypothesis H_0 : $(\theta = 4.4)$ against H_1 : $(\theta > 4.4)$

(c) In (a), find that value of the alternative where the power of the test is 0.90. (25)

2. (a) X is a random variable with probability function $P(X = x) = k\theta^x$, $x = 0, 1, 2, 3, \ldots, (0 < \theta < 1)$. Find the most powerful test, based on X, of H_0 : $(\theta = \frac{1}{2})$ against H_1 : $(\theta > \frac{1}{2})$ for $\alpha = 0.0625$ and plot its power function.

Please turn over

2. (b) (contd.) X is a random variable with density

$$f_\theta(x) = k|x|\, e^{-\theta x^2}, \quad -\infty < x < \infty, \quad (\theta > 0).$$

Find the uniformly most powerful test, based on X_1 of $H_0 : (\theta = 1)$ against $H_1 : (\theta > 1)$ for $\alpha = 0.05$ and plot its power function.

(c) $X_1, X_2, \ldots, X_n$ are independent and have the same probability function $f_\theta(x) = P(X_i = x) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$, $(0 < \theta < 1)$. Find the value of n so that the most powerful test of $H_0 : (\theta = \frac{1}{2})$ against $H_1 : (\theta < \frac{1}{2})$ has approximately the size $\alpha = 0.10$ and the power = 0.80, at the alternative $\theta = \frac{1}{3}$. (25)

3. (a) A manufacturer of pens wants to reduce variations in the lengths of pens as much as possible. The lengths (in cm.) of 10 pens produced by a new process are

8.00, 7.72, 8.68, 8.12, 8.18
8.20, 7.92, 8.60, 8.02, 7.90.

In the past, the standard deviation of the lengths of pens has been 0.380 cm. Test if the new process can be considered as superior to the old.

(b) A manufacturer claims that no more than 6% of the items he produces are defective. In a sample of 20 items produced there were 4 defectives. Is the manufacturer's claim justified?

(c) The average daily number of motor car accidents in five major cities of India during April 1969, were found to be as follows :

City :	1	2	3	4	5
Average daily number :	18.5	14	11.5	12	15

Do you think that the traffic problem is equally acute in all the five cities above ?

[You may assume that the number of accidents in a month follows a Poisson distribution]

(d) The following table shows the number of men and women entrants to to government service in five Indian States, during six months of 1968.

State	Men	Women	Total
I	1636	161	1797
II	2195	223	2418
III	2130	193	2323
IV	1620	137	1763
V	2317	230	2547

Does the proportion of men and women differ significantly in the various States ? (25)

Please turn over

5. (a) The following is a random sample of size 20 from a population :

37.0, 31.4, 34.4, 33.3, 34.0, 31.6, 31.3, 34.6., 32.0, 31.0

36.2, 31.0, 33.5, 33.7, 33.4, 33.4, 32.1, 33.3, 32.3, 31.5

Test, at a level nearly 0.01, the hypothesis that the population median is 32. Find a nearly 99% confidence interval for the median.

(b) Construct 90% confidence intervals for the mean and variance for a normal distribution, given the following 14 observations from it :

4.28, 4.32, 4.32, 4.33, 4.28, 4.31, 4.29,

4.31, 4.37 4.38, 4.35, 4.29, 4.28, 4.32.

(c) One puppy is given diet 'A' and another diet 'B', in each of 30 pairs of puppies - each pair being taken from a single litter. If diet 'A' puppies fare better in 17 of the 30 pairs and fare worse in 10 pairs, is this sufficient evidence to assert that diet A is superior ?

(d) The correlation coefficient of two variables based on 30 pairs of observations is - 0.218. Test whether there is any correlation between the characters in the population. Find the 95% confidence interval for the true correlation coefficient.

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1970

Paper I : Official Statistics and Descriptive Statistics (Theoretical)

Time : 4 hours | Full marks : 100

(a) Figures in the margin indicate full marks

GROUP A

(Attempt any three questions from this group)

1. (a) Describe the rôle of the Central Statistical Organisation (C.S.O.) in respect of the Statistical System of India.

(b) Mention the important publications of C.S.O.

(12+4) = 16

2. Describe briefly the work in the field of Statistics of any two of the following :

(i) Any one specialised agency of the United Nations.

(ii) The Reserve Bank of India.

(iii) Any important non-governmental agency in India.

(8+8) = 16

3. (a) Write a note on the functions of the office of the Registrar General of India.

(b) Give a short account of the historical background of population census in India and the special features of the proposed 1971 Census.

(6+10) = 16

4. Mention the main publications relating to :

(i) Agricultural Statistics in India.

and (ii) Employment and Unemployment Statistics in India.

Indicate the coverage, contents and broad limitations of these publications.

(10+6) = 16

5. Indicate the type of Statistics that you will need for studying the rate of growth of India's national income during the 3 plan periods, and for comparing the same, with the rate of growth and investment in respect of any one of the following sectors :

(a) Agriculture,

(b) Manufacturing industries,

(c) Public Administration and Defence.

Mention the publications you would like to consult for your purpose.

(11+5) = 16

Neatness : (2)

Please turn over

GROUP B

(Attempt all the questions from this group)

6. (a) Prove that the mean deviation about the mean cannot exceed the standard deviation. Where are the two equal?

(b) Show that for n observations $x_1, x_2, \ldots, x_n$ the variance $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is also equal to

(i) $\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2/2n^2$

(ii) $\sum_{1}^{n} i(x_i - \bar{x}_i)^2/(i^2 - 1) \div n$,

where $\bar{x}_i = \frac{1}{i}(x_1 + x_2 + \ldots + x_i)$.

(5+11) = 16

7. If the variates x_1, x_2, x_3 have the multivariate normal distribution with the mean vector 'zero' and variance-covariance matrix V, obtain the conditional distributions of

(i) x_1 and x_2 given x_3

(ii) x_1 given x_2 and x_3.

Find these distributions when $V = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$

(4x4) = 16

OR

Obtain the moment-generating function of xy, where x and y follow the bivariate normal law given by

$$\text{Const. exp}\left[-\frac{1}{2}(1-\rho^2)\left\{\frac{x^2}{\sigma x^2} - 2\frac{\rho xy}{\sigma x\,\sigma y} + \frac{y^2}{\sigma y^2}\right\}\right]$$

Hence obtain the value of the r-th cumulant of xy.

(10+6) = 16

8. What does a cost of living index measure? How do you derive the weights required in constructing such an index? State the important uses of a cost of living index. Briefly describe the procedure followed in working out any important cost of living index.

(16)

OR

(a) Give a short description of the various components of a time series.

(b) Discuss the various methods of obtaining the trend component of time series.

(6+10) = 16

Neatness : (2)

Please turn over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1970

Paper II : Probability Theory and Statistical Methods (Theoretical)

Time : 4 hours Full marks : 100

Figures in the margin indicate full marks.

GROUP A

(Answer any four questions from this group)

1. (a) Define the conditional probability of an event given another event and show that it satisfies Kolmogorov's axioms.

(b) Suppose we distribute randomly r balls in n cells and we assume that each arrangement has probability n^{-r}, then find the probability that exactly m cells remain empty.

(5+7) = 12

2. (a) Give one example each of random phenomenon obeying Hypergeometric and Negative Binomial distributions.

(b) Calculate the variances for Hypergeometric and Gamma distributions.

(4+8) = 12

3. Given $f(x,y) = 8xy$, $0 \leq x \leq 1$, $0 \leq y \leq x$, find the conditional expectation of y given x.

Further, show that x and y are not statistically independent, even though $f(x,y)$ is the product of a function of x alone and a function of y alone in the region : $0 \leq x \leq 1$, $0 \leq y \leq x$.

(12)

4. (a) Prove the weak law of large numbers for a sequence of independent identically distributed random variables, with finite variance.

(b) Check whether the weak law of large numbers holds if the common density function is :

(i) $f(x) = \frac{\alpha^p}{\Gamma} e^{-\alpha x} x^{p-1}$, $0 \leq x \leq \infty$

(ii) $f(x) = \frac{1}{\pi} \frac{1}{(1+x)^2}$, $-\infty \leq x \leq \infty$

(5+4+3) = 12

5. If $X_1, \ldots, X_m$; $Y_1, \ldots, Y_n$ are independent random variables, the first m following a $N(\mu_x, \sigma_x^2)$ distribution and the rest obeying a $N(\mu_y, \sigma_y^2)$ distribution, then show that the moment generating function of the difference $(\sum_1^m X_i/m) - (\sum_1^n Y_j/n)$ (i.e. of the averages of the two groups of variables), is that of a normal distribution with mean $\mu_x - \mu_y$ and variance $(\sigma_x^2/m + \sigma_y^2/n)$.

(12)

Neatness : (2)

PAPER II – GROUP B

(Answer any three questions from this group)

6. (a) Explain the concepts of 'efficiency' and 'sufficiency' in the theory of estimation.

(b) Examine which of the following distributions admit a sufficient statistic for θ.

(i) $f(x,\theta) = (1+\theta)\, x^{\theta}$, $0 \leqslant x \leqslant 1$

(ii) $f(x,\theta) = \theta/(x+\theta-1)^2$, $1 \leqslant x \leqslant \infty$

(iii) $f(x,\theta) = \theta.(\sin x)^{\theta-1}\cos x$, $0 \leqslant x \leqslant \frac{\pi}{2}$

(7+9) = 16

7. (a) Explain briefly the terms (i) level of significance, (ii) size of the test and (iii) power of a test.

(b) Examine whether Uniformly Most Powerful (U.M.P.) tests exist, for testing H_0 against H_1 in the following cases, when n independent observations are drawn from the distributions :-

(i) $f(x,\theta) = \frac{1+\theta}{(x+\theta)2}$, $1 \leqslant x \leqslant \infty$

$H_0: \theta = \theta_0$, $H_1: \theta > \theta_0$

(ii) $f(x,\theta) = \theta\, e^{-\theta x}$,

$H_0: \theta = \theta_0$, $H_1: \theta \neq \theta_0$.

(8+8) = 16

8. (a) Describe a non-parametric method of setting confidence limits to a quantile of a continuous distribution.

(b) Write a short note on robustness of test procedures.

(8+8) = 16

9. (a) Describe the 'likelihood ratio approach' to the problem of hypothesis testing. Show that this leads to the t-test for testing student's hypothesis.

Obtain the likelihood ratio test for testing the equality of variances of k ($k > 2$) normal populations.

(b) Describe briefly the sequential probability ratio test.

(12+4) = 16

10. (a) Given k normal populations $N(\mu_i, \sigma^2)$ $(i=1,2,\ldots k)$, $k > 2$ with all parameters unknown, describe a method of testing the homogeneity of means.

(b) Describe an exact method of testing independence in a 2 x 2 contingency table.

(c) Show that for a r x n contingency table $\chi^2 \sim \sum_r \left[\frac{N_1 N_2 (a_{1r}/N_1 - a_{2r}/N_2)^2}{(a_{1r} + a_{2r})}\right]$

where a_{1r} and a_{2r} are frequencies in the r-th column and N_1, N_2 are the marginal frequencies of the two rows.

(5+5+6) = 16

Neatness :

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1970

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

Time : 4 hours Full marks : 100

Figures in the margin indicate full marks

GROUP A

(Answer any three questions from this group)

1. (a) Define the following terms :

(i) Sample
(ii) Elementary unit
(iii) Sampling unit
(iv) Sampling error

(b) In simple random sampling with replacement, obtain the probability of the i-th unit of the population being included in the sample.

(c) Obtain the standard error of the mean of the sample selected by the method of simple random sampling without replacement.

(6+2+8) = 16

2. (a) Explain the purpose of stratification in sample surveys.

(b) What are the factors that are taken into consideration in allocating the sample size to different strata.

(c) Obtain the expressions for gain in precision due to stratification under proportional allocation, and Neyman allocation.

(4+4+8) = 16

3. (a) Define the regression estimate of the population mean and derive its expectation and variance, when the regression coefficient is known.

(b) Obtain the approximate expressions for its expectation and variance, when the regression coefficient is not known and is estimated from the sample.

(c) State under what conditions, you would recommend the use of the regression estimate.

(4+8+4) = 16

4. Give a short account of the scope, design and organisation of the Indian National Sample Survey, with special reference to any recently conducted round.

(16)

5. Write notes on any two of the following :

(i) Sampling with varying probabilities.
(ii) Systematic sampling.
(iii) Sub-sampling.
(iv) Non-sampling errors.

(8+8) = 16

Neatness : (2)

(X5)

Please turn over

PAPER III GROUP B

(Answer any three questions from this group)

1. (a) How do size and shape of plots affect the results of an experiment?

   (b) State how the fundamental principles of design are incorporated in a randomised block experiment. Compare a randomised block experiment with a completely randomised experiment.

   (c) What is the difference between a (complete) three factor experiment and a Latin square experiment?

   (6+8+2) = 16

2. (a) Explain the different methods of controlling heterogeneity in experimental designs.

   (b) Give the intra-block analysis of a Simple Lattice design and derive the expression for the average sampling variance of estimated elementary treatment contrasts.

   (8+8) = 16

3. (a) Define the terms 'main effects' and 'interaction effects' for a 2^n experiment.

   (b) Why is confounding used even at the cost of loss of information on the confounded effects?

   (c) A 2^4 experiment is to be carried out in blocks of 4 plots each, without confounding any main effect. Give a suitable design in 4 replications partially confounding every interaction in some replicate or the other. Indicate what effects are confounded in each replicate and give an outline of the analysis of the experiment.

   (6+2+8) = 16

4. (a) Give in detail the analysis of a completely randomised experiment with one concomitant variable.

   (b) Discuss the efficiency of the above experiment relative to the completely randomised experiment in which the concomitant variable is ignored.

   (10+6) = 15

5. Write notes on any three of the following :

   (a) Missing-plot technique

   (b) Fractional replication

   (c) Inter-block and intra-block analysis

   (d) Youden square

   (16)

Neatness : (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1970

Paper IV : Applied Statistics (Theoretical)

Time : 4 hours

Full marks : 100
(For two groups combined)

(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 : ECONOMIC STATISTICS - 50 marks

(Attempt any three questions from this group)

1. What type of data would you need for the calculation of a consumer price index number for the salaried employees of the Govt. of India working at New Delhi and drawing not more than Rs.1000 per month, as basic pay? Given the data, how will you calculate the index number and use it for adjustment of pay, so as to maintain a fixed real income?

(16)

2. What do you mean by seasonal fluctuations in a time series? Give illustrations of at least five types of series data, where this fluctuation is important. Given the time series, how will you calculate seasonal indices and use them for forecasting? If the seasonal pattern changes over time, how will you take account of this?

(16)

3. Indicate the nature of data you would need if you are required to obtain the demand function for food grains on the basis of time series data. What form of the demand function will you choose?

Briefly describe the method of estimating the parameters of the demand function and the elasticities of demand with respect to price and income.

(16)

4. What is a production function? For the Cobb-Douglas production function obtain the conditions for increasing, constant and diminishing returns. What type of data do you need to estimate this function? How will you estimate the parameters of the function? How will you estimate the elasticities of output with respect to the factors of production?

(16)

5. What is national income? Briefly describe the methods of calculation of national income of a country in a given year. Given a series of national income of India in different years starting from 1948-49, how will you study the change in the standard of living over the period?

(16)

Neatness : (2)

Please turn over

PAPER IV - GROUP B : STATISTICAL QUALITY CONTROL - 50 marks

(Attempt any three questions from this group)

1. (a) Explain the terms :

(i) "Tolerance limits",

(ii) "Confidence limits",

(iii) "Specification limits", and

(iv) "Control limits".

(b) Describe a method of obtaining tolerance limits for the case of normal distribution with known standard deviation. Derive the necessary expressions.

(6+10) = 16

2. (a) The average number of visual defects in a specified size of cloth has been found from past records to be 2.8. It is desired to establish control at this level, so that there is 1 in 100 chance of the control limit being reached or exceeded. Determine this control limit stating clearly your assumptions.

What is the chance of this limit being reached or exceeded when the average number of defects increases to 3.6.

(b) Samples of size n are drawn from each of 3 machines at regular intervals to study a measurable characteristic. After k samples are drawn, it is desired to test whether the variabilities of the 3 machines are the same. Suggest a quick and ready method of doing it, by making use of the 'range's.

(9+7) = 16

3. (a) Describe clearly any scheme of inspection of a continuous production process to ensure a prescribed limit on the out-going quality.

(b) Describe clearly any one of the published sampling plans that are available for industrial use, for inspection by measurement.

(8+8) = 16

4. (a) Explain the different methods of controlling heterogeneity of experimental material in experimental design, indicating the structure of the designs used for this purpose. Indicate also, how you will test if the control of heterogeneity has been effective in those different designs.

(b) Four treatments are to be compared using the experimental material consisting of sixteen units symbolically arranged in two-way array of four rows and four columns. Prepare a lay-out for carrying out the experiment. Describe the randomisation procedure adopted by you.

(8+8) = 16

Neatness : (2)

Please turn over

PAPER IV-GROUP C : STATISTICAL METHODS IN GENETICS - 50 marks

(Attempt any three questions from this group)

1. Gene pairs A-a and B-b are suspected to be linked. Both genes exhibit dominance (of A over a and B over b). Data on F_2 from double heterozygote in repulsion phase, gave the following frequencies :

Phenotypic class :	AB	Ab	aB	ab
Frequency :	a	b	c	d

Discuss the relative efficiency of the product-ratio estimate of the recombination fraction relatively to the maximum likelihood estimate.

(16)

2. Explain the usefulness and limitations of the concept of Wright's coefficient of in-breeding.

How intense is the in-breeding of animals inbred between full brother and sister for three generations?

Which other concepts of in-breeding do you know of?

(16)

3. With the help of a suitable model explain the components of observed variation in a population.

Explain the procedure of estimating the components of variation, with the help of half-sib and full-sib data.

(16)

4. Discuss the inheritance of Rh factor in man.

(16)

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

(i) Methods of ascertainment of inherited abnormalities in man

(ii) Equilibrium under mutation and selection

(iii) Assortative mating.

(16)

Neatness : (2)

PAPER IV - GROUP D : VITAL STATISTICS AND DEMOGRAPHY - 50 marks

(Attempt any three questions from this group)

1. Define the following terms : (i) child-woman ratio, (ii) crude birth rate, (iii) total fertility rate, (iv) general fertility rate, (v) gross reproduction rate, (vi) net reproduction rate.

Indicate under what conditions each of these can be considered a suitable measure of the fertility of a population.

(16)

Please turn over

PAPER IV - GROUP D : VITAL STATISTICS & DEMOGRAPHY (contd.)

2. (a) Explain the concept of 'stationary population'.

In a population known to be stationary, the complete expectation of life at birth is 40 years. Determine the level of crude birth and death rates in the population.

(b) Describe King's method of constructing an abridged life-table.

(6+10) = 16

3. Describe in details the component method of population projection.

(16)

4. Write notes on any two of the following :

(a) Sources of demographic data in India.

(b) Stable population.

(c) Measures of morbidity.

(8+8) = 16

Neatness : (2)

PAPER IV - GROUP E : EDUCATIONAL AND PSYCHOLOGICAL STATISTICS - 52 marks

(Attempt any three questions from this group)

1. Derive the basic equations in test theory from a definition of random error.

(16)

2. (a) Derive the equations for the effect of test length on reliability.

(b) Find a function of test reliability that is invariant with respect to changes in test length.

(8+8) = 16

3. Solve the problem of equating test Z (given to Group B) to test Y (given to Group A) by means of an equating test X or a set of k equating tests X_g, g = 1, ..., k.

(16)

4. Show the relationship of item parameters to test variance, reliability and validity.

(16)

5. Using hypothetical data, show the differences, if any, between the centroid and principal component methods of factor analysis.

(16)

Neatness : (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination – November 1970

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

Time : 5 hours Full marks : 100

(i) Figures in the margin indicate full marks.

(ii) Use of calculating machines is permitted

GROUP A (25 marks)

(Answer any two questions from this group)

1. Using the data furnished in the following table, find Cosh (0.7357) and Cosh (0.7384)

X	Cosh X
0.735	1.2824037
0.736	1.2832074
0.737	1.2841023
0.738	1.2849985
0.739	1.2857159
0.740	1.2865247
0.741	1.2873348
0.742	1.2881461

(12½)

2. (a) From the table provided below,

x :	1.2	1.3	1.4	1.5	1.6	1.7
f(x) :	81.82	79.75	78.73	78.62	79.35	80.36

Find (i) f (1.45), (ii) the value of x, in the given range, for which f(x) is a minimum and (iii) the minimum value of f(x) in this range of x.

(b) Compute the approximate value of

$$\int_0^6 \frac{dx}{(1+x)^2}$$

by (i) Simpson's One-third Rule and (ii) Weddle's Rule.

Compare their accuracies. (7+5½) = 12½

3. Find the inverse of the following correlation matrix :

$$\begin{vmatrix} 1.000 & & & \\ 0.243 & 1.000 & & \\ 0.266 & 0.628 & 1.000 & \\ 0.307 & 0.553 & 0.489 & 1.000 \end{vmatrix}$$

(12½)

PAPER V - GROUP B

(Answer all questions from this group)

4. Examine the three series of figures given in the following table and state whether they may be regarded as mutually compatible. Point out the inconsistencies, if any.

crop	area under crop (000 acres)	total yield (000 tons)	yield rate (lbs. per acre)
Rice	74,424	24,209	728.63
Maize	9,325	2,944	707.70
Wheat	26,842	8,530	7125.9
Barley	7,099	2,786	780.17
Jowar	43,456	9,092	468.65
Bajra	27,350	3,555	401.16

(1 ton = 2,240 lbs.)

(7)

5. Answer any three from (a) - (d)

(a) If x and y are independenty and normally distributed with 'zero' means and 'Unit' standard deviations

(i) calculate the probability that $x^2 + y^2 < 1$, and

(ii) determine the radius of the circle (with its centre at the origin) which has the probability 0.95 that the point (x,y) will fall inside it.

(7)

(b) Assuming that during a process of firing shots at a target, the horizonal distance (measured from the centre line of the point at which an aimed shot hits) is distributed with a standard deviation of 2 ft., find

(i) how many out of 200 shots fired are likely to miss the target if the target is 10 ft. wide & sufficiently high

(ii) how many shots would need to be fired, if 50 or more of these shots are to liewithin 3 ft. of the centre line, with a probability of 0.05.

(7)

(c) The results of a particular examination are shown below in a summary form :

result	percentage of candidates
(i) passed with distinction	13
(ii) passed without distinction	46
(iii) failed	41
Total	100

Given that a candidate gets plucked if he obtains 40 marks (out of 100), and that he must secure at least 80 marks in order to pass with distinction, determine the mean and standard deviation of the distribution of marks, assuming this to be normal.

5. (d) (contd.) Assume that customers enter a store at the rate of 120 persons per hour.

(i) What is the probability that during a 2-minute interval no one will enter the store?

(ii) What time interval is such that the probability is $\frac{1}{4}$ that no one will enter the store during that interval?

[Assume that the number of arrivals in any fixed interval of time follows a Poisson distribution.]

(7)

6. EITHER

The pressure of a gas (p) and its volume (v) are related by an equation of the form $pv^b = b$. In a certain experiment the following values were obtained

p	.5	1	1.5	2	2.5	3
v	1.62	1	.75	.62	.52	.46

Determine a and b by applying the method of least squares on the logarithmic equation.

(12)

OR

In an experiment on wheat, fertilizers were applied at various levels, with resulting yields as follows :

Fertilizer level (x)	0	5	10	15	20	25	30
Yield (y) in appropriate units	27.1	32.1	35.0	36.2	37.2	38.1	36.7

Fit the equation $y = 40.0 - \exp(a + bx)$ to the given data.

(12)

7. The following table gives the yield-rate of rice in West Bengal for a number of years.

Determine the trend values by means of moving averages of an appropriate period.

year	yield of rice (maunds per acre)	year	yield of rice (maunds per acre)
1946-47	9.36	1952-53	13.53
1947-48	9.88	1953-54	13.48
1948-49	9.75	1954-55	10.40
1949-50	10.26	1955-56	11.11
1950-51	10.86	1956-57	10.81
1951-52	9.98		

(13)

Please turn over

PAPER V - GROUP C (25 marks)

(Attempt both the questions)

From the Official Publications placed at your disposal, furnish information with regard to questions 8 and 9. Wherever possible the data collected should be arranged neatly in suitable tabular form.

8. EITHER Figures for 3 latest years for the Corporations of Bombay, Calcutta and Madras, under the following heads :

(a) Total Tax Revenue, (b) Total ordinary income, (c) Total wages and salaries paid to all employees, and (d) Total Revenue Expenditure. (13)

OR Figures for 2 latest years, under the following heads :

(a) Estimated Mid-year population,

(b) Index of Agricultural Production,

(c) Index of Mineral Production,

(d) General index of Industrial Production,

(e) Index number of Wholesale prices,

(f) Index number of value of imports, and

(g) Index number of value of exports.

/The available years should be noted as also the nature of the years, the base years and any other specially important remarks./

9. Attempt any four of the following :

(a) Figures of expectation of life (for both males and females) at (i) birth, (ii) age 20, (iii) age 60, for the latest available year.

(b) Index number of Industrial profits for the latest available year (separately for) (i) Public Limited Companies and (ii) PrivateLtd.Cos.

(c) Total capital-at-charge and total staff (all classes combined) employed in All Indian Railways for the two latest available years.

(d) All-India Average Consumer price index numbers for Industrial workers under general & group heads, for any two latest months.

(e) Production of Sewing and Knitting machines for any 3 latest months' total figures and figures for West Bengal.

(f) Monthly average production of Crude Steel for the latest two available years for the following countries : India, Japan, U.S.A. and U.S.S.R.

(g) Recorded foreign exchange rates (quoted in terms of units of foreign currency to Rs.100/- in Indian money) of the latest two months, for the following countries : India, Malayasia, Japan, West Germany.

/NB: Care should be taken always to mention

(i) years and months, units and any other important remarks considered specially necessary.

(ii) at every stage the page, the table number and the complete name of the official publication with its year, as consulted by you./

(4x3) = 12

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1971

Paper VI : Statistical Methods; Design & 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 any two questions from this group)

1. (a) In each of ten pairs of rats, one received protein from raw peanuts while the other received it from roasted peanuts. The following table gives the gain in weights :

feed	gain in weights of ten pairs of rats									
raw peanuts	61	60	56	63	56	63	59	56	44	61
roasted peanuts	55	54	47	39	51	61	57	54	62	58

Test whether or not roasting the peanuts had any effect on their protein value. If there is evidence of a significant effect, compute a 90% confidence interval for the mean difference.

(b) The correlation coefficients between wing size and tongue length in bees, based on four independent samples each of size 40, were reported as 0.731, 0.354, 0.690 and 0.740. Test the hypothesis that these are samples from populations having a common correlation coefficient.

(10+10) = 20

2. (a) The following two-way table gives the classification of 1000 school boys, according to their general ability and their mathematical ability. Examine whether there is any association between the two abilities :

		mathematical ability		
		good	fair	poor
general ability	good	44	22	4
	fair	265	257	178
	poor	41	91	98

(b) Construct a sequential acceptance-rejection chart for testing the binomial probability $p = 0.5$ against the alternative $p = 0.7$, given the strengths $\alpha = 0.1$, $\beta = 0.2$. The following table gives the results of a sequence trials (only 10 trials are given), x_n being the total number of successes upto and including the n-th trial.

n :	1	2	3	4	5	6	7	8	9	10
x_n :	0	0	1	1	2	3	3	4	4	4

(c) Would you, on the basis of above results, advise sampling to be continued? If not, what will be your decision?

(10+10) = 20

Please turn over

3. (a) One hundred rats were exposed to a poisonous gas and the number of survivors, observed after certain periods of time, are given below :

exposure time (minutes)	5	10	15	20	25
number of survivors	72	53	39	30	20

Assume that the density function of the survival time t (measured in minutes) of rats exposed to the gas, is given by

$$f(t,\theta) = \theta^{-1} \exp(-t/\theta) \text{ for } 0 \leqslant t \leqslant \infty \quad , \theta > 0$$

where θ is an unknown parameter.

Estimate θ and calculate the expected frequencies.

Test the goodness of fit.

(b) To compare four treatments, field experiments were conducted in five experimental stations. Data from the different stations were analysed separately. The exact levels (upper tail probabilities) corresponding to the computed values of the analysis of variance F-statistic, came out as (i) .04970, (ii) .001463, (iii) .007243, (iv) .1337 and (v) .1236, in the five stations respectively. Combine the evidences of the experiments carried out at the different stations.

(15+5) = 20

PAPER VI - GROUP B

(Attempt any two questions from this group)

4. The following table relates to data from an agricultural experiment conducted in 4 randomised blocks of 5 plots each. For each plot in a block the topmost figure (in brackets) represent the treatment, the next figure below, current yield and the third figure at the bottom, the yield on a previous occasion.

Block I	(1)	(2)	(3)	(4)	(5)
	26	28	25	21	14
	46	28	40	39	19
Block II	(2)	(3)	(4)	(1)	(5)
	8	9	4	7	9
	14	33	18	14	36
Block III	(1)	(5)	(2)	(3)	(4)
	21	19	22	4	11
	50	43	41	22	42
Block IV	(2)	(1)	(4)	(5)	(3)
	19	20	15	2	2
	29	35	45	8	11

(a) Analyse the data ignoring the yield figures for the previous occasion and draw suitable conclusions.

(b) If the figures for the previous occasion be considered, would you change the conclusions that you have drawn in (a)?

(5+10) = 15

5. An agricultural experiment was conducted to study the effects of three manures A, B and C each at three levels, on the yield of wheat. It was decided to examine all combinations of the levels of these factors and the whole experiment was repeated to give sufficient precision. The levels of each manure correspond to equi-spaced doses.

yields of wheat

	A_0			A_1			A_2		
	B_0	B_1	B_2	B_0	B_1	B_2	B_0	B_1	B_2
C_0	74	13	69	112	46	130	71	56	125
	85	12	115	148	52	117	75	47	79
C_1	211	119	190	166	218	220	201	216	227
	181	145	164	288	204	142	216	230	265
C_2	74	147	195	47	146	198	90	102	164
	75	104	183	65	124	165	60	70	114

Analyse the data to test for the main effects and interactions.

Also test for the linear and the quadratic components of the significant main effects.

(11+4) = 15

6. The following data relate to a Latin Square experiment with 5 treatments A, B, C, D, E.

A 220	B 98	C 149	D 92	E 282
B 158	C 228	D 168	E 238	A 74
C 279	D 183	E 278	A 176	B X
D 104	E 295	A 54	B 232	C 213
E 242	A 60	B 188	C 187	D 99

The figure of B in third row is missing. Analyse the data to test for treatment differences.

(15)

PAPER VI - GROUP C

(Attempt any one question from this group)

7. The following data relate to area and yield of paddy for 30 villages :

village (sl.no.)	cultivated area x(acres)	paddy yield y(maunds)	village (sl.no.)	cultivated area x(acres)	paddy yield y(maunds)
1	870	8521	16	1220	11913
2	883	8554	17	1260	11617
3	894	8783	18	1280	12105
4	901	8862	19	1295	12428
5	914	7325	20	1318	12477
6	973	8887	21	1330	12421
7	995	9569	22	1340	12786
8	1031	9598	23	1367	12786
9	1043	10316	24	1398	13040
10	1054	8663	25	1415	13221
11	1089	9562	26	1437	13340
12	1121	10512	27	1455	13118
13	1152	10874	28	1469	12966
14	1189	11580	29	1491	12804
15	1214	11631	30	1512	13985

7 (d+) Select a sample of 10 villages using the method of sampling with probability proportional to size and with replacement, taking area as a size-measure. On the basis of this sample obtain an unbiased estimate of the total yield of paddy in all these 30 villages.

Estimate from the sample the variance of the estimate.

Also, using the population records, determine the true value of this sampling variance.

(8+1+8+13) = 31

There are 2357 small agricultural farms in a certain country. A sample survey of these farms was undertaken to estimate, the average number of cattle per farm. The farms were divided into five strata in respect of the area xxxxxxxxxx of the farm, and from each stratum a specified number of farms was sampled on a simple random sampling basis, without replacement. In the table below, N_i stands for the number of farms in the i-th stratum, n_i stands for the number of farms sampled in the i-th stratum, $\bar{y}_i$ stands for the mean and s_i^2 stands for the variance (divisor n_i) of the number of cattle in the sample from the i-th stratum (i = 1,2,3,4 and 5).

area of farm (acres)	stratum no. (i)	no. of farms in stratum (N_i)	no. of farms in sample (n_i)	mean no. of cattle per sample farm ($\bar{y}_i$)	variance of no. of cattle per sample farm (s_i^2)
less than 15	1	724	61	4.246	27.548
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 over simple random sampling.

(c) An alternative sampling plan would be to divide the population into two strata consisting of farms with area in the 'less than 50' range and '51-100' range respectively and to take a simple random sample of size 61+55+46 from the first stratum and another of size 29+9 from the second. On the basis of the above data, judge whether this would result in substantial loss in efficiency, compared to the above plan.

(13+13+13) = 39

Please turn over

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1970

Paper VII : Applied Statistics (Practical)

Time : 6 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 options.

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

(c) Figures in the margin indicate full marks.

(d) Use of calculating machines is permitted.

GROUP A : ECONOMIC STATISTICS

(Attempt all questions from this group)

1. EITHER

Estimated figures of population (in millions) and production of cloth (in thousand metres) are given below for India during the years 1951 to 1969. Compare the trends of total production of cloth and per-capita production of cloth. Obtain the cyclic component of the series of production figures of cloth and examine whether marked cyclical fluctuation is noticed.

Also make a forecast of the production of cloth for 1970 and 1971.

year	population (in millions)	production of cloth (000 metres)	year	population (in millions)	production of cloth (000 metres)
1951	363	4740	1961	443	7073
1952	369	5318	1962	453	6972
1953	376	5871	1963	464	7209
1954	383	6582	1964	476	7719
1955	390	6278	1965	487	7643
1956	398	6515	1966	499	7336
1957	406	6673	1967	511	7277
1958	414	6473	1968	524	7896
1959	423	6578	1969	536	7706
1960	433	6629			

(20)

OR

Consumer expenditure in rupees per household for a period of one year on food and clothing, are given below along with the total expenditure for the agricultural labour-households in the Community Project areas of India. Find the elasticities of expenditures on food and clothing with respect to total expenditure.

Estimate the expenditures on these items for an increase of 7 per cent in the total expenditure. (20)

annual per-capita expenditure (rupees)	number of agricultural labour households	average size of household	total expenditure of households (rupees)	expenditure (Rs.) on food	expenditure (Rs.) on clothing
0 - 50	63	5.6	234.3	187.9	4.5
51 - 100	587	5.3	426.0	349.5	10.6
101 - 150	892	4.5	553.8	444.2	24.6
151 - 200	527	4.0	700.8	536.6	49.3
201 - 250	288	3.3	731.3	537.1	61.5
251 - 300	148	3.0	827.2	579.7	92.0
301 - 350	93	2.5	824.2	563.9	91.0
351 and above	122	3.2	1476.3	778.5	244.7

(Turn) Please turn over

2. Index numbers (base 1929 = 100) of real output, man-hours of labour employed and capital, are given below for the period 1940–57 for U.S.A. Fit a Cobb-Douglas production function and interpret the elasticities of output with respect to labour and capital. (3

Index numbers of output, labour employed and capital: USA (1940–57)

year	real output	man hours of labour	capital used	year	real output	man hours of labour	capital used
1940	110.1	93.9	97.2	1949	161.3	135.3	129.5
1941	134.8	103.2	100.5	1950	178.6	137.1	125.2
1942	152.6	114.9	103.2	1951	191.1	112.9	131.0
1943	171.3	120.9	103.3	1952	199.9	114.6	135.3
1944	183.1	133.0	102.1	1953	205.7	115.1	141.3
1945	181.0	124.7	100.5	1954	202.3	110.7	145.5
1946	163.3	109.3	102.6	1955	216.9	114.4	149.8
1947	160.2	108.3	108.9	1956	221.8	116.1	155.1
1948	163.6	108.9	115.6	1957	224.3	114.9	160.4

PAPER VII – GROUP B : STATISTICAL QUALITY CONTROL

(Attempt any two questions from this group)

1. Carbon content is an important characteristic for the steel castings required for high pressure valves. 3 test bars are taken from each heat for 20 days in a month during September 1970. The carbon determinations are given below :

Heat No.	Carbon % in test bar			Heat No.	Carbon % in test bar		
	1	2	3		1	2	3
1	2.407	2.406	2.408	11	2.405	2.401	2.404
2	2.408	2.410	2.403	12	2.402	2.397	2.397
3	2.406	2.406	2.411	13	2.411	2.413	2.406
4	2.409	2.412	2.413	14	2.409	2.416	2.413
5	2.405	2.409	2.412	15	2.403	2.408	2.405
6	2.408	2.406	2.405	16	2.411	2.412	2.407
7	2.403	2.404	2.408	17	2.408	2.413	2.405
8	2.406	2.402	2.407	18	2.407	2.413	2.413
9	2.404	2.405	2.403	19	2.411	2.409	2.412
10	2.402	2.402	2.405	20	2.411	2.410	2.412

(a) Carry out a control chart analysis to adjudge the stability of the process and establish standards.

(b) If the minimum carbon content be specified at 2.400% and the maximum at 2.413%, what percentage of castings would fail to meet these specifications?

(c) It was found that the limits in (b) were too tight and the functional requirements could be satisfied with minimum carbon of 2.395% and the maximum at 2.441%.

How would you modify your control scheme to take advantage of the new specification limits?

Comment on the necessity of such a scheme to control the carbon content.

(15+4+5) = 24

Please turn over

Group B (contd.)

2. The minimum requirements of tensile strength of a forged item is 82 kg/mm^2. It is desired to instal a suitable sampling scheme with the following properties.

AQL = 2 percent and producer's risk = 0.01
LTPD = 6 percent and consumer's risk = 0.05

Specify a suitable single sampling variable plan and the corresponding attribute plan. Clearly set down the plan of action in each case. Under what conditions would you prefer to use the variable plan in preference to the attribute plan?

(24)

3. Four trials were conducted in each of 8 machines in a Randomised block set up to examine the effects of two process factors, speed(S) and feed (F), on the production of Timing cover. Two levels of S (s_0 and s_1) and two levels of F (f_0 and f_1) were tried in all the four possible combinations. The average productions per hour are given below.

F	S	
	s_0	s_1
f_0	38.25	47.50
f_1	45.75	52.50

Analysis of variance of the experimental result gave 10.532 as the mean square due to error.

a) Compute the sums of squares due to treatments and analyse it into three components (i) due to the main effect of S (ii) due to the main effect of F and (iii) due to the interaction between S and F.

b) Draw up a table showing the structure of the complete analysis of variance; the degrees of freedom for the different components and the mean square due to as many of these components as are possible to obtain from the data supplied to you. Test the significance of the mean squares computed in (a).

c) Given that the ratio of the benefit of extra productivity per hour and the cost of running at prescribed levels of S and F is 8 : 10, what is your recommendation about the best combination of the two process factors?

d) Write a short report (for non-statisticians) on your findings.

(8+6+5+5) = 24

Neatness : (2)

(Please turn over)

PAPER VII - GROUP C : STATISTICAL METHODS IN GENETICS

(Attempt any two questions from this group)

1. The following table gives the number of plants in F_2 and F_3 generations, raised from a cross between two pure strains of cotton and segregating for the characteristics governed by gene pairs A - a and B - b. (A is dominant over a and B over b).

generation	Phenotypic Class			
	AB	Ab	aB	ab
F_2	17[illegible]	23	2[illegible]	44
F_3	28[illegible]	24	3[illegible]	7[illegible]

Test for the segregation ratios and linkage. If linkage is present test for the homogeneity of the recombination fraction as estimated from the two sets of data.

Also obtain a combined estimate, if valid.

(24)

2. Three characters,

(i) boll per plant

(ii) seeds per boll

(iii) lint per seed

were used as a basis for building up a selection index for improving yield of cotton (y). The estimates of genotypic and phenotypic variances and covariances obtained from an experiment were as follows:

Estimates of genotypic variances and covariances

g_{11}	g_{12}	g_{13}	g_{1y}
0.273	0.659	-0.110	0.272
	g_{22}	g_{23}	g_{2y}
	1.512	-0.101	0.712
		g_{33}	g_{3y}
		0.055	0.025
			g_{yy}
			0.613

Estimates of phenotypic variances and covariances

t_{11}	t_{12}	t_{13}	t_{1y}
1.023	0.555	-0.083	1.098
	t_{22}	t_{23}	t_{2y}
	2.157	-0.120	0.915
		t_{33}	t_{3y}
		0.066	0.077
			t_{yy}
			2.772

Obtain the selection index (linear discriminant function) and compare the genetic advance expected by its use with that from straight selection.

(24)

3. Using maximum likelihood method, estimate the gene frequencies from the following observed distribution of O-A-B blood group classes :

Blood group	O	A	B	AB
Frequency	70	43	68	15

Test whether the data conform with Bernstein's theory.

(24)

Neatness : (2)

Paper X - GROUP F : TECHNIQUES OF COMPUTATION (Paper III: Practical)

Numerical computation with the help of Unit Record Machine

(Attempt all the stages of the given question)

In the operations to be carried out, you will be dealing with two files of cards. They are identified by 78F1 and 78F3 punched in columns 1-4, of the respective files of cards.

The card designs of 78F1 and 78F3 cards are as follows:

Card-design 78F1

Serial No.	Item	Card columns	Remarks
1	design index number	1 - 4	punched 78F1
2	block serial number	5 - 7	
3	stratum	8 - 9	
4	sub-sample	10	
5	sub-round	11	
6	sample household	12 - 14	
7	party	15	
8	'couple' reference	16 - 17	
9	month of birth	18 - 21	
10	age last birthday	22 - 23	
11	month of marriage	24 - 28	
12	age at marriage	29 - 30	
13	month of birth of husband	31 - 34	
14	wife's age at consummation of marriage	35 - 36	
15	other socio-economic particulars	37 - 47	
16	social and economic characteristics	48 - 71	
17	residential history	72 - 80	

Card design 78F3

Serial No.	Item	Card columns	Remarks
1	design index number	1 - 4	punched 78F3
2	block serial number	5 - 7	
3	stratum	8 - 9	
4	sub-sample	10	
5	sub-round	11	
6	sample household	12 - 14	
7	party	15	
8	'couple' reference	16 - 17	
9	status code	18 - 19	
10	number of children desired	20 - 21	
11	" " sons "	22 - 23	
12	" " daughters "	24 - 25	
13	attitude code	26 - 27	
14	number of sons born	28 - 29	
15	" " daughters born	30 - 31	
16	other details of family size	32 - 55	
17	particulars of family planning methods used	56 - 71	

Please turn over

The various stages of operations required to be carried out are indicated below:

STAGE 1

(i) Take 78F1 cards and sort them by columns (5-7) × (12-14) × (16-17) [Note (5-7) indicates cols. 5 to 7 & so on]

and note down the total volume of cards handled in the answer sheet provided.

(ii) Take 78F3 cards and repeat the same operations as advised for 78F1 cards.

(iii) Note down the number of cards replaced (if any) during sorting and hand over the damaged cards (if any) to the Examiner after their replacement.

(5+3+5+3+4) = 20

STAGE 2

(i) Wire a control panel of a collator machine for carrying out the operations under (ii) below.

(ii) Take the sorted packs of cards in stage 1. Merge them on a collator machine with selection of unmatched cases, considering the arrangement of sorting in stage 1 and making 78F1 cards leading.

(iii) Note down the number of 78F1 cards and 78F2 cards selected in (ii) above, on the answer sheet.

(iv) Write down your name and the number of cards selected, at the back of the selected 78F1 and 78F3 cards separately and hand them over to the Examiner.

(8+8+8+6) = 30

STAGE 3

(i) Take the merged deck of cards in (ii) of stage 2 and prepare panel chart/an Accounting Machine for printing the following particulars:

Serial number	Item	Remarks
1	design index number	
2	block serial number	
3	sub-sample	
4	sample household	
5	couple reference	
6	age last birthday	from 78F1
7	age at marriage	- do -
8	age at consummation of marriage	- do -
9	number of sons desired	from 78F3
10	number of daughters desired	- do -
11	number of sons born	- do -
12	number of daughters born	- do -

(ii) Obtain the Print.

(iii) Write table-heading etc. on the machine print and hand over the tabulated sheet, 78F1 cards and 78F3 cards to the Examiner.

(iv) Put your signatures on the Performance sheet provided and produce it at the end of every stage for countersignature. Hand over finally to the Examiner at the conclusion of the examination.

(2+5+12+6+5) = 30

N.B. For the practical examination the Examiner will be represented by the Supervisor/Supervisors.

Please turn over

PAPER VII - GROUP D : VITAL STATISTICS AND DEMOGRAPHY

(Attempt any two questions from this group)

1. You are given the following data from an occupational mortality investigation :

age group	standard population		occupation X		occupation Y	
	population at risk	death rate	population at risk	death rate	population at risk	death rate
15 - 24	270,000	.001	[illegible]	.005	12,000	.001
25 - 34	310,000	.002	16,000	.002	20,000	.001
35 - 44	350,000	.003	33,000	.002	23,000	.001
45 - 54	320,000	.008	50,000	.007	30,000	.005
55 - 64	250,000	.022	28,000	.021	28,000	.021

Express the mortality of occupation X at ages 15 - 64 as a single percentage-ratio of that of occupation Y by means of the following :

(a) standardised mortality ratio.

(b) comparative mortality figure.

(c) comparative mortality index.

(25)

2. Starting with $l_{15} = 100,000$, construct an abridged life table for a female population in the child-bearing age range from the table given below :

[You may take $e^{0}_{50} = 20.2$ years]

age group	number of females enumerated in 1961 (000's)	number of female deaths registered during the three year period (1960-62)
15 - 20	761	2008
20 - 25	777	4190
25 - 30	721	4485
30 - 35	637	5002
35 - 40	560	6621
40 - 45	582	9552
45 - 50	597	15370

(25)

Please turn over

Group D (contd.)

3. Below is given an extract of a life-table for a female population.
[Radix of the life-table l_0 = 10,000]

age group x to x+5	life table population in the age group x to x+5 $_5L_x$	age group x to x+5	life table population in the age group x to x+5 $_5L_x$
0 - 5	48039	45 - 50	43484
5 - 10	47556	50 - 55	41961
10 - 15	47379	55 - 60	39807
15 - 20	47171	60 - 65	36748
20 - 25	46852	65 - 70	32475
25 - 30	46442	70 - 75	26642
30 - 35	45952	75 - 80	19229
35 - 40	45346	80 +	* 17200
40 - 45	44562		

* Denotes life table population aged 80 and above.

Given r = .00042, compute the stable age distribution for the population by quinquennial age groups.

[r is the true rate of natural increase for the stable population]

(25)

PAPER VII - GROUP E : EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Attempt any three questions from this group)

1. Find the reliability of a test from the following data : (16).

subjects	score on 1st test	score on retest	subjects	score on 1st test	score on re-test
A	8.5	8.25	K	6.5	7.25
B	6.75	8.25	L	7.75	8.5
C	6.25	7.75	M	9.0	8.0
D	6.0	5.50	N	7.5	8.25
E	7.0	7.25	O	7.0	7.75
F	9.25	9.25	P	7.75	8.25
G	5.5	5.75	Q	5.75	5.75
H	9.25	8.25	R	8.75	9.5
I	7.75	6.75	S	8.75	8.5
J	5.0	5.25	T	9.0	8.0

Please turn over

Group E (contd.)

2. Find the validity of a clerical proficiency test from the test scores and criterion ratings of a group of workers given below :

test scores	criterion ratings: semi-skilled and unskilled workers	criterion ratings: skilled workers
51 – 52	0	1
49 – 50	1	5
47 – 48	0	7
45 – 46	3	8
43 – 44	2	4
41 – 42	2	8
39 – 40	1	9
37 – 38	6	7
35 – 36	7	2
33 – 34	6	3
31 – 32	4	2
29 – 30	8	4
27 – 28	6	1
25 – 26	9	2
23 – 24	3	–
21 – 22	1	–
19 – 20	6	–
17 – 18	4	–
15 – 16	1	–

(16)

3. From the correlation matrix given below, extract three centroid factors :

Test	1	2	3	4	5	6
1		.63	.00	.00	.27	.09
2			.32	.36	.21	.09
3				.72	.27	.24
4					.00	.00
5						.72
6						

(16)

4. The raw scores on a test are given below. Compute percentile scores, standard scores ($\bar{X}$ = 50, SD = 10) and normalized scores.

37	43	27	44	27	27	26	31	35	42	50
35	43	36	26	50	47	30	26	32	32	38
36	21	24	40	36	35	38	36	38	21	17
26	35	22	18	50	30	38	50	16	45	8
34	26	34	23	41	27	39	41	30	23	33
22	31	36	40	54	24	22	8	33	42	41
41	31	34	36	32	20	22	34	41		

(16)

Please turn over

Group E (contd.)

5. From the following item-score matrix obtained from a test, find the mean and standard deviation of test-scores, item difficulty and average item discrimination.

Also estimate the correlation of total score with criterion.

Item-score matrix

	Items											
examinees	a	b	c	d	e	f	g	h	i	j	k	l
1	1	0	1	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	1	0	0	0	0	0	0
3	1	1	1	1	0	0	0	0	0	0	0	0
4	1	1	0	1	1	0	0	1	0	0	0	0
5	1	1	1	1	1	0	0	0	0	0	0	0
6	1	1	1	0	1	1	1	0	0	0	0	0
7	1	1	1	1	1	1	1	0	0	0	0	0
8	1	1	1	1	0	1	1	1	1	1	0	0
9	1	1	1	1	1	1	1	1	1	1	1	0
10	1	1	1	1	1	1	1	1	1	1	1	1

(16)

Neatness : (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1970

Paper VIII : Subjects of Specialisation - (I) [Grs. A to H]

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 - Special Paper I

(No candidate available)

Paper VIII-GROUP B : TECHNO-COMMERCIAL STATISTICS - Paper I (SQC)

(Answer any five questions from this group)

1.(a) Discuss the difference between Inspection and Quality Control.

(b) Distinguish between 'action limits' and 'warning limits' in a control chart. What are the advantages in having warning limits?

(c) Explain the procedure for installing the cumulative sum (cusum) control charts.

(4+8+8) = 20

2.(a) Discuss the situations when 'modified limits' may be used.

Describe briefly a method of fitting a trend-line and of determination of control limits, when tool-wear is present.

(b) When can the 'Moving Average and Moving Range' charts be used in preference to the conventional $\bar{X}$, R charts. Give a few examples of such cases, from any industry you know.

(4+8+8) = 20

3.(a) What are the difficulties encountered in the use of 'lot by lot' acceptance sampling plans for continuous production?

(b) Explain Dodge's CSP - 1 procedure for continuous production.

(c) Establish the following results for CSP-1 plan (standard notation).

(i) the average number of pieces inspected in a 100% screening sequence following the finding of a defect,

$$\text{is } u = \frac{1 - q^i}{p\, q^i}$$

(ii) The average number of pieces passed under the sampling procedure before a defect is found

$$\text{is } v = \frac{1}{fp}$$

Please turn over

3.(c) (iii) The average fraction of total production inspected,

$$\text{is } F = \frac{u + fv}{u+v}$$

What is the physical interpretation of the CC curve of this plan?

(3+3+14) = 20

4.(a) Develop a sequential test procedure for controlling the mean of a process with known standard deviation when the quality characteristic is normally distributed.

(b) Devise a simple graphical procedure for use when inspection is on a routine basis.

(12 + 8) = 20

5.(a) Compare and contrast the salient features of Dodge-Romig and Mil-Std-105-D Sampling inspection tables for attribute inspection.

(b) Mention any procedure you may know of, by means of which an AQL value can be selected on a rational basis, for use in Mil-Std tables.

(c) In Mil-Std-414 for the known σ case with one-sided upper specification limit u, the following formula is used to estimate the fraction defective for a given $\bar{X}$.

$$\hat{p} = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\frac{n}{n-1}}\frac{(u-\bar{x})}{\sigma}}^{\infty} e^{-x^2/2}\, dx$$

Discuss the theoretical basis of this formula.

(8+6+6) = 20

6.(a) For a continuous quality characteristic, the ordered statistics for a sample of size n are $x_1, x_2, \ldots\ldots x_n$. Show that the probability that the tolerance range $[x_1, x_n]$ includes atleast 100 α% of the population is given by the expression:

$$\frac{1}{\beta(n-1,2)} \int_{\alpha}^{1} x^{n-2} (1-x)\, dx$$

(b) Write a short note on the statistical principles that you can use in problems involving the combination of tolerances of mating parts.

(12 + 8) = 20

7.(a) What are the various costs of carrying out a company's quality mission? Classify the costs and clearly define the various components in each category.

(b) How would you proceed to develop systems for reducing avoidable quality losses?

(12 + 8) = 20

8.(a) Describe the precautions needed in designing and conducting plant scale experiments.

(b) Explain with the help of one or two examples from any industry known to you, the use of 'nested design' in industrial experimentation. How would you analyse the results of such an experiment?

(10 + 10) = 20

Paper VIII- GROUP C : BIOMETRIC METHODS - Special Paper I
(No candidate available)

Paper VIII- GROUP D : Design and Analysis of Experiments
(Statistical Aspects) - Special Paper I

(Answer any five questions from this group)

1. Let $\underline{\eta}$, X and $\underline{\beta}$ be $N \times 1$, $N \times L$ and $L \times 1$ matrices respectively, and rank X be equal to $L \leq N$. Let $\underline{Y}$ be a random vector of order $N \times 1$ such that $E(\underline{Y}) = \underline{\eta} = X\underline{\beta}$ and dispersion matrix of $\underline{Y}$, $D(\underline{Y}) = V\sigma^2$ where X and V are known matrices, with V positive definite of order $N \times N$ and symmetric. $\underline{\beta}$ and σ^2 are unknown parameters. Obtain,

(i) the best unbiased linear estimator $\hat{\underline{\beta}}$ of $\underline{\beta}$;

(ii) the dispersion matrix $V(\hat{\underline{\beta}})$ of $\hat{\underline{\beta}}$;

(iii) $E(\underline{Y} - X\hat{\underline{\beta}})'(\underline{Y} - X\hat{\underline{\beta}})$;

(iv) $V(\underline{Y} - X\hat{\underline{\beta}})$.

If the model $\underline{\eta} = X\underline{\beta}$ is inadequate and the true model is $\underline{\eta} = X\underline{\beta} + X_1\underline{\beta}_1$ where X_1 is a known matrix of order $N \times L_1$ and β_1 a $L_1 \times 1$ vector of unknown parameters, compute afresh

$E(\underline{Y} - X\hat{\underline{\beta}})'(\underline{Y} - X\hat{\underline{\beta}})$ where $\hat{\underline{\beta}}$ is same as in (i) above.

(4+4+4+4+4) = 20

2. In a randomized block experiment, originally planned with v varieties and r replications, it is found later that there is not enough material of variety 1 and excess of material of variety 2. Variety 2 is therefore applied twice in blocks 1, 2, ..., r_1 and only once in the remaining blocks. Variety 1 is applied only in blocks $(r_1 + 1)$, (r_1+2), ..., r.

(i) Write down the C - matrix of the design.

(ii) Obtain the variances of different treatment comparisons.

(iii) Indicate how you would carry out the analysis of variance of this design.

(4+8+8) = 20

3. Let $\lambda_1, \ldots, \lambda_{v-1}$ be the non - zero latent roots of the matrix of coefficients of the reduced normal equations for intra-block estimates of treatment effects in an incomplete block design in which v treatments are tested in b blocks of k plots each, and each treatment is replicated r times. Show that the average variance of all elementary treatment comparisons is given by,

$$\frac{2\sigma^2}{(v-1)} \sum_{i=1}^{v-1} \frac{1}{\lambda_i} ,$$

where σ^2 is the variance of the yield of a plot.

Please turn over

3. (contd) Hence show that the design with maximum efficiency with respect to this average variance as the criterion is a Balanced Incomplete Block design, if such a design exists.

(10 + 10) = 20

4. A balanced incomplete block design, with parameters v,b,r,k,λ is modified to include a new treatment by increasing the block size to (k+1) plots, so that the new treatment occurs once in each of the b blocks.

(i) Show how you would analyse this experiment ?

(ii) Prove that the intra - block variance of the difference between effects of any two of the original set of v treatments is

$$\frac{2\,(k+1)\,\sigma^2}{v\lambda + r}$$

where σ^2 is the intra-block variance for blocks of (k+1) plots.

(12 + 8) = 20

5. The actual design (before randomization) of a factorial experiment involving six factors A, B, C, D, E, F each at two levels (zero and non-zero) arranged in 4 blocks of 8 plots each, is given below:

Block 1	Block 2	Block 3	Block 4
(1)	ae	ad	ae
ab	be	bd	be
acde	de	ce	cd
bcde	abde	abce	abcd
acdf	df	cf	cdef
bcdf	abdf	abcf	abcd
ef	aef	adef	af
abef	bef	bdef	bf

Identify the confounded interactions and determine the 'alias' sets.

Assuming that three factor and higher order interactions are not existent, write down the partitioning of the degrees of freedom in the table of analysis of variance, against each ascribable source of variation.

(6+6+8) = 20

6.(a) Explain the layout of a simple square lattice design and comment on the scope of its utility.

(b) Obtain expressions for the intra-block estimates of treatment contrasts and their standard errors.

(c) Calculate the efficiency factor of the design.

(d) Discuss the connection of this type of design with a Partially Balanced Incomplete Block Design, identifying the primary and secondary parameters appropriately.

(4+6+6+4) = 20

(N>0) Please Turn Over

7.(a) Discuss the advantages and disadvantages of a split-plot design.

(b) A split-plot design with p whole-plot treatments and q sub-plot treatments arranged in randomized blocks with r replications. Give the complete analysis of variance of the design on a sub-plot basis. Give the standard errors of the various types of treatment comparisons.

(4+8+8) = 20

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

(i) Uses of orthogonal latin squares in design of experiments.

(ii) Problem of balancing in confounded factorial designs.

(iii) Weighing Designs.

(iv) Youden square designs.

(10 + 10) = 20

.....

Paper VIII-GROUP E : SAMPLE SURVEYS (THEORETICAL ASPECTS) - Sel.Paper I

(Answer any five questions from this group)

1. From a tehsil consisting of N villages, a sample of n villages is to be drawn with probability proportional to the area of the village and with replacement. The aim is to estimate Y, the total area under paddy in the tehsil.

(a) Give a method of selection that ensures the required probabilities and show how exactly it is ensured.

(b) Give an unbiased estimator $\hat{Y}$, say, of Y and prove its unbiasedness.

(c) Derive an unbiased estimator of $V(\hat{Y})$, the variance of $\hat{Y}$.

(5+5+10) = 20

2. Suppose, as in question 1, a sample of size n is drawn from a population of size N, with probabilities proportional to the size-measures and with replacement. Let Y_i be the value of the variable of interest to us and p_i the probability of selection, for the ith unit $(1 \leq i \leq N)$.

Let n' be the number of distinct units in the sample.

(a) Find $E(n')$ and $V(n')$.

(b) Prove that the estimator :

$$\Sigma^* \frac{Y_i}{1 - (1 - p_i)^n}$$ is unbiased for Y.

(where the sum Σ^* is over the distinct units i that belong to the sample)

(4+6+10) = 20

(531)

Please turn over

3. A population is divided into two strata, these being formed on the basis of some administrative convenience. A simple random sample, with replacement, of size n_1 is to be drawn from the first stratum, while from the second stratum we intend to draw a simple random sample, without replacement, of size n_2. The cost of sampling a unit is C_i for the ith stratum (i = 1,2).

(a) Find the optimum values of n_1 and n_2 that minimise the variance of the usual estimator of the population total.

(b) Suppose a stratified sample has been drawn as indicated above with some values of n_1 and n_2 (not necessarily the optimum values). As an afterthought one is interested in knowing whether this stratified sample is better than an unstratified simple random sample without replacement of size m.

(i) What should be the value of m if this unstratified method is to be comparable to the stratified method, as regards the cost?

(ii) Estimate, on the basis of the stratified sample drawn, the gain (or loss) that would have resulted, had we adopted this unstratified sampling method with the value of m obtained in (i) above.

(3+4+10) = 20

4. To estimate the total rural indebtedness in a district it is proposed to adopt a 2-stage sampling method with villages as the primary stage units (p.s.u's) and households (h.h's) as the secondary stage units (s.s.u's). A sample of n villages are to be drawn with probabilities proportional to the population as per the latest available census figures and with replacement. From each selected village a sample of n households are to be selected by simple random sampling without replacement.

(a) Give an unbiased estimator $\hat{Y}$, say, of the total indebtedness of the rural population of the district.

(b) Obtain an unbiased estimate of the variance of $\hat{Y}$.

(c) Give an estimate of the proportion of the h.h.'s in the district whose debt is at least Rs.100. Is it an unbiased estimator ? If so prove the same and if not, give an approximate expression for the bias of this estimator.

(3+9+8) = 20

5. A population consisted of K strata. A stratified sample is intended to be taken by adopting simple random sampling without replacement, of size n_i from the ith stratum, i=1,2, ..., K. The object is to estimate Y, the total of the variable y.

(a) Find the optimum values of the n_i's that minimise the variance of $\hat{Y}$, an unbiased estimator of Y you intend to use, for a fixed value n of the total sample size.

(b) It is recognised that the optimum values of n_i involve a knowledge of some population parameters which are functions of the y's. Since these are unknown beforehand it is intended to substitute in their place the corresponding values of an auxiliary variable x which are available beforehand.

5. (contd.) If the relationship between $\mathcal{Y}$'s and $\mathcal{X}$'s in the ith stratum can be expressed as a model given by

$$E(Y_{ij} \mid X_{ij}) = \alpha_i X_{ij}$$

$$V(Y_{ij} \mid X_{ij}) = \sigma_i^2 X_{ij}^2$$

$$Co\ (Y_{ij}; Y_{ij'} \mid X_{ij}; X_{ij'}) = 0, \quad \text{for } j \neq j'$$

where X_{ij} and Y_{ij} are the values of $\mathcal{X}$ and $\mathcal{Y}$ for the jth unit of ith stratum and α_i and σ_i^2 are constants ($1 \leq i \leq k$), examine the validity or otherwise of the substitution method given above in the light of the model assumed.

(6 + 14) = 20

6. Comment critically on the following statements with suitable mathematical arguments:

(a) As an estimator of the population ratio $\frac{Y}{X}$, where X and Y are the population totals of the $\mathcal{X}$ and $\mathcal{Y}$ variables, the ratio of the sample means is better than the mean of the sample ratios. (Assume that the sampling is simple random sampling with replacement).

(b) To estimate the population total Y the ratio estimator is better than the regression estimator if the regression of $\mathcal{Y}$ on the auxiliary variable $\mathcal{X}$ has a small intercept on the $\mathcal{Y}$-axis.

(10+10) = 20

7. Write short notes on any four of the following:

(a) Circular systematic sampling.

(b) Self-weighting designs.

(c) Varying probability sampling methods.

(d) Combined ratio estimators versus separate ratio estimators in stratified sampling.

(e) Sources of non-sampling errors.

(5+5+5+5)=20

.....

Paper VIII : GROUP F : TECHNIQUES OF COMPUTATION
(Numerical Analysis - Special Paper I)
(No candidate available)

Paper VIII-GROUP G : STATISTICAL INFERENCE (GENERAL THEORY) - Spl.Paper I

(Answer any five questions from this group)

1.(a) State and prove the Cramer - Rao inequality.

(b) Explain how the inequality can be used to find minimum-variance unbiased estimates.

(c) Let $X_1, \ldots, X_n$ be independent. Observations from the Poisson distribution with mean θ. Obtain the minimum-variance unbiased estimate of θ.

(8+4+8) = 20

2.(a) Define the terms (i) admissible decision function, (ii) minimax decision function, (iii) complete class of decision functions.

(b) Show that if a minimax decision function is unique then it is admissible.

(c) Suppose that X is a single observation from the density

$$f(x,\theta) = \frac{1}{\theta^p \sqrt{(p)}} e^{-x/\theta} x^{p-1}, \quad 0 < x < \infty,$$

where p is a known positive number and θ is an unknown positive parameter. Suppose we want to estimate θ with squared error as the loss function. By considering estimates of the form k X, show that the usual estimate X/p is not admissible.

(6+6+8) = 20

3.(a) Explain the maximum likelihood method of estimation of parameters.

(b) Independent observations $X_1, \ldots, X_n$ are available from the density

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

where a is a real parameter and θ is a real positive parameter. Obtain the maximum likelihood estimates of a and θ.

(c) Give an example of a maximum likelihood estimate which is not consistent.

(5+9+6) = 20

4. Suppose that n independent observations are available from the normal distribution with mean μ and variance 1.

(a) Construct the usual $100(1-\alpha)\%$ confidence interval for μ. Justify the facts that the interval is based on $\bar{X}$ and that it is symmetric about $\bar{X}$.

(b) Obtain the minimum - variance unbiased estimate of μ^2.

(12 + 8) = 20

5.(a) Define the terms :

(i) unbiased test, (ii) similar test, (iii) test with Neyman structure.

Please turn over

5.(b) State and prove any theorem you know connecting similar tests and tests with Neyman structure.

(c) Given n independent observations from a normal distribution with unknown mean μ and unknown variance σ^2, construct a Uniformly Most Powerful (U.M.P.) unbiased test of level α for testing

$$H_0 : \mu \leq \mu_0 \quad \text{against} \quad H_1 : \mu > \mu_0$$

(6+7+7) = 20

6.(a) Let X have the probability density f given by

$$f(x, \theta) = \frac{1}{\beta(\theta, \theta)} \left[x(1-x)\right]^{\theta-1}, \qquad 0 < x < 1$$

where θ is an unknown positive parameter.

Show that, for testing (at level α)

$$H_0 : \theta = 1 \quad \text{against} \quad H_1 : \theta > 1,$$

a U.M.P. test exists. Find such a test and find its power when $\theta = 2$.

(b) Suppose $\left\{T_n\right\}$ is a sequence of estimates of θ.

Show that if $E(T_n) \longrightarrow \theta$ and $\text{Var}(T_n) \longrightarrow 0$, then T_n is consistent for θ.

(12 + 8) = 20

.....

PAPER II : PROBABILITY THEORY
Special Paper I - BASIC PROBABILITY
(No candidate available)

.....

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1970

Paper IX : Subjects of Specialisation - II (Theoretical)

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 and Economics of Planning)
-Special Paper II

Section I - Indian Economics

(Answer any three questions from this section)

1. Indicate clearly the methods you would adopt for controlling the growth of population in India, if such control is considered desirable. (16)

2. Are small farms uniformly more efficient than large farms in India ? Discuss the factors influencing the relative productivity of large and small farms. (16)

3. Does the industrial licensing policy pursued by the Government of India, effectively limit the degree of monopoly in Indian industry? Will the recent modifications in this policy make it more effective for the control of monopoly? Give reasons for your answer. (16)

4. In what ways can commercial banks help finance agricultural production and investment in India ? (16)

5. Discuss the main features of changes in real wages of labour employed in modern industries in India since independence. (16)

6. Discuss how you would try to assess the incidence of indirect taxation on the rich and the poor, and on the urban and the rural populations in India. (16)

Section II - Economics of Planning

(Answer any three questions from this section)

7. How can the planner planning for a finite time horizon, take account of the fact that life will go on after the last year of the plan period has elapsed ? (16)

8. Discuss how the required amounts of investment can be calculated for the purpose of grafting on to a multi-period, input-output model of planning. (16)

(N30) Please turn over

9. How can financial and physical balances be made consistent in a practical planning exercise ? (16)

10. Discuss the relative merits of import substitution and export promotion as strategies for increasing India's foreign exchange earnings in the near future. Are these strategies necessarily exclusive of each other ? (16)

11. Assess the total impact of PL 480 aid on India's economy. (16)

12. Write notes on any two of the following :

(a) the social rate of discount;

(b) the Raj-Sen model of planning;

(c) the strategy of balanced growth;

(d) the Mahalanobis 2-sector model.

(8+8) = 16

Neatness (Sections I and II) (4)

.....

Paper IX - GROUP B : TECHNO-COMMERCIAL STATISTICS
(Special Paper - II)

Section I : OPERATIONS RESEARCH (70 marks)

Section I (Alternative): ELEMENTS OF BOOK-KEEPING AND ACCOUNTANCY (70 marks)

Section II: STATISTICAL METHODS IN BUSINESS (30 marks)

Section I : OPERATIONS RESEARCH

(i) Use a separate answer book for this section.

(ii) Attempt question 1 and any other two questions from this section.

1.(a) Explain the simplex method of solving a linear programming problem with the help of the following example:

Maximize $Z = 3x + 4y + 5z$

subject to the restrictions

$$2x + 3z \leq 8$$

$$5y + 2z \leq 10$$

$$3x + 4y + 2z \leq 15$$

$$x, y \text{ and } z \geq 0.$$

(b) Write the dual problem of the above and also its optimal feasible solution.

1.(c) Formulate mathematically the following problem as a linear programming problem (DON'T SOLVE THE PROBLEM):

An agriculturist has 3 farms and wants to grow 4 crops A, B, C and D on them. The land to be allotted to different crops on each farm is to be determined so as to maximize the total profit of the output from the farms. The relevant data are given below:

Table I

Farm	Usable acreage	Water available in acre feet
1	400	1,500
2	300	2,000
3	300	900

Table II

Crop	Maximum Crop acreage	Water requirements in acre feet per crop acre	Profit per crop acre
A	500	5	450
B	700	4	300
C	400	3	250
D	300	1	200

(14+8+8) = 30

2.(a) Describe briefly various methods of solving the transportion problem for material stored at m sources and required at n destinations, assuming the total available material is just sufficient to satisfy the total requirements.

(b) Study the following production-cum-transportion problem:

Plant	Depots				Available (Regular production)
	1	2	3	4	
I	25	17	25	24	300
II	15	10	18	24	500
III	16	20	8	13	600
Required	300	300	500	500	1,400 / 1,600

Here the regular production quantities are not sufficient to meet the total requirements. It is possible, however, to increase production of each plant by 50 per cent, by paying overtime. The costs shown in the above table are unit costs of production-cum-transportation set-up. When on overtime, these unit costs are increased by 10, 15 and 20 for the plants I, II and III respectively.

Please turn over

The company's objective is not to have overtime production at any plant unless its regular time production is fully used. Set up a suitable initial transportation problem tableau by adding rows and or columns to the above tableau which will ensure company's objective (DON'T SOLVE THE PROBLEM FOR OPTIMUM SOLUTION).

(12 + 8) = 20

3. Develop an economic lot-size formula in the following circumstances:

(a) Orders for replenishment can only be placed on the first day of a calendar month.

(b) If an order is placed, delivery is immediate.

(c) Demand is r per month and arises only on the 15th of each month.

(d) Carrying costs are C_1 per unit per month.

(e) No shortages are permitted.

(f) Set up costs are C_2.

(20)

4.(a) Develop a multi-item deterministic inventory model for determining economic order quantities when (i) the production is assumed to be instantaneous, (ii) the shortages are not allowed and (iii) total value of average inventory of all stocked items does not exceed a specified value (say I Rupees).

(b) For the following data, determine approximately the economic order quantities when the total value of average inventory levels of three products, is ≤ Rs.10,000. (YOU MAY USE SIMPLE LINEAR INTERPOLATION FOR APPROXIMATE CALCULATIONS).

Costs	Product 1	Product 2	Product 3
Holding cost* (%)	20	20	20
Set-up cost (Rs)	50	40	60
Cost per unit (Rs)	6	7	5
Yearly Demand rate	10000	12000	7500

*as per cent rate of yearly interest on the value of average inventory.

(8 + 9) = 17

5.(a) Starting from first principles, obtain the following expressions for a single server queue with Poisson arrivals, with mean arrival rate λ and exponential service time with mean service rate μ.

(i) $\phi_T(n) = P\{n \text{ services in time } T \text{ given that servicing is going on throughout } T\}$

$= (\mu T)^n \cdot e^{-\mu T} / n!$

(ii) P_n = Probability that there are n units in the system

$$= (1 - \frac{\lambda}{\mu})(\frac{\lambda}{\mu})^n \quad , \quad n \geq 0$$

(iii) $\psi(w)$ = Probability law for waiting time w of a new arrival

$$= \begin{cases} 1 - \lambda/\mu & \text{for } w = 0 \\ (1 - \lambda/\mu)\lambda \, e^{(\lambda-\mu)w} & , \; w > 0 \end{cases}$$

5.(b) Assuming a Poisson arrival with average rate of 15 per hour at a single ticket-booking counter, at what average rate the counter clerk should work to ensure a probability of 0.90 that a new customer will not have to wait longer than 12 minutes? (Assume that the length of service by the clerk has an exponential distribution).

(4+5+6+5) = 20

...

IX Group B: Section I (Alternative): ELEMENTS OF BOOK-KEEPING AND ACCOUNTANCY

(a) Use separate answer book for this section

(b) Attempt question 1 and any other three from this group.

1. The following is the Trial Balance of XYZ company on 31.3.70

	Dr.	Cr.
	Rs.	Rs.
Cash in hand	1,080	
Cash at Bank	5,260	
Purchase account	81,350	
Sales account		107,580
Return Inwards account	1,360	
Return Outwards account		1,000
Wages account	20,980	
Fuel and Power account	9,460	
Carriage on Sales account	6,400	
Carriage on Purchases account	4,080	
Stock on 1 April 1969.	11,510	
Buildings account	60,000	
Freehold land account	20,000	
Machinery account	40,000	
Patents account	15,000	
Salary account	30,000	
Sundry Expenses account	6,000	
Insurance account	1,200	
Drawings account	10,490	
Capital account		1,42,000
Sundry debtors	20,000	
Sundry creditors		12,000
	3,53,100	3,53,160

Prepare a Trading and Profit and Loss account and a Balance sheet as at 31.3.70 after making the following adjustments:

(1) Stock in hand as at 31.3.70 Rs. 13,600

(2) Depreciation: Machinery @ 10% and patent @ 20%.

(3) Salary for the month of March '70, was unpaid by Rs. 3,000

(4) Insurance includes a premium of Rs. 340/- on a policy expiring on September 1970.

(5) Wages include Rs. 4000/-, for creation of a Shed in Building.

(6) Provision for Bad and Doubtful debts to be created to the extent of 5% on Sundry debtors.

(25).

2. A receives three promissory notes from B dated 1st January 1970 for 3 months. The first one is for Rs. 1000/-, the second is for Rs. 500/- and the third for Rs. 250/-. The first is endorsed to C, the second is discounted with a Banker for Rs. 420/- and the third bill is cashed on maturity. Journalise the transactions and also show the entry if all these bills are dishonoured.

(15)

3.(a) What is an accommodation Bill? Illustrate.

(b) What is a 'Del Credere Commission' in consignment?

(c) How consignment differs from joint venture ?

(15)

4. A trial balance of ABC company was drawn, and it was found to be arithmetically accurate. The book-keeper during posting came to make the following mistakes,

(a) A bill for Rs.85/- received from a customer was credited to Machinery account.

(b) Discount of Rs.54/- allowed to customers had been debited to Reserve account.

(c) Repairs to engine for Rs. 130/- and cost of new boiler for Rs.850/- had been charged to Machinery account.

(d) A sum of Rs. 500/- for 3 years advertisement, had been carried to Printing and Advertising.

What errors of fact or principle had the book-keeper committed? How would you rectify them and what would be their results if unremedied ?

(15)

5. What is a Petty Cash book. Draw a multi-column petty cash book along with entries of five imaginary accounts therein.. What is meant by imprest system of Petty cash ?

(15)

6. From the following particulars show the various accounts in the Sales ledger and General ledgers of XY & Company.

	Balance as on 1.1.69 (1)	Credits during January (2)	Cash received during January (3)	Discount allowed (4)	Bills Received (5)
A	570	1370	555	15	-
B	700	3450	830	-	320
C	1140	780	950	25	-
D	350	650	-	-	100
E	690	430	670	20	-

(15)

IX Group D - Section II : STATISTICAL METHODS IN BUSINESS

(i) Use separate answer book for this section.

(ii) Attempt any two questions from this section.

1. The packaging machinery in a dry-food firm is known to pour dry food in boxes with standard deviation of 0.6 ounce. Constant checks on the net weights in the boxes are made in order to maintain the adjustment of the machinery that controls the net contents. Two samples taken on two different dates yield the following information:

1st sample	2nd sample
$n_1 = 30$	$n_2 = 35$
$\bar{x}_1 = 14.7$ ounces	$\bar{x}_2 = 17.9$ ounces

Using 5% level of significance, test the hypotheses that

(a) on the first date, the machine was adjusted to fill 16 ounces.

(b) on the second date, the machine was adjusted to fill 18 ounces.

(c) there was no change in the adjustment between the two dates.

(5+4+6) = 15

2. A survey was conducted in Calcutta on the voters' reactions towards various policies of the Government before the general election of 1967. Some of the results are given below:

Satisfied with the Government's	Men (n = 420)	Women (n = 450)
Foreign policy	20%	35%
Labour policy	32%	34%
Defence policy	21%	53%

(a) Assuming the samples are random, do you think there is a significant difference between the reactions of men and women towards each of the policies mentioned above ?

(b) What over-all conclusions can you draw from your computations?

(9 + 6) = 15

3.(a) How would you proceed to estimate the demand for a cooking gas in a given town ?

(b) What do you understand by time series and its trend ? How will time series analysis help you to forecast demand for a certain commodity.

(7 + 8) = 15

...

IX Group C : BIOMETRIC METHODS - Special Paper II
(Statistical methods in Genetics and Bio-assays)
(No candidate available)

IX GROUP D : DESIGN AND ANALYSIS OF EXPERIMENTS - Special Paper II
(Combinatorial Aspects)

(Answer any five questions from this group)

1.(a) Define a projective plane and show its relationship with complete sets of mutually orthogonal latin squares.

(b) Construct a pair of mutually orthogonal latin squares of order 12.

(10 + 10) = 20

2.(a) Define an orthogonal array (N, k, s, t) of index λ. For an orthogonal array (λs^2, k, s, 2) of index λ, show that

$$k \leq \left[\frac{\lambda s^2 - 1}{s - 1} \right],$$ where (x) is the greatest integer contained in x.

(b). Give a geometric method of constructing the orthogonal array $(s^2, s+1, s, 2)$ of index 1, when s is a prime or a prime power.

(4+8+8) = 20

3.(a) Define an affine resolvable balanced incomplete block design. Show that a resolvable balanced incomplete block design is affine resolvable if $b = v + r - 1$. Find the parameters of affine resolvable balanced incomplete block designs in terms of two unknowns.

(b) Define a complement of a balanced incomplete block design and derive its parameters. Show that every triplet of treatments occur together the same number of times, in 2b blocks formed by a balanced incomplete block design and its complement.

(2+6+4+1+2+5) = 20

4.(a) Define a partially balanced incomplete block design and derive all parametric relations. Determine the characteristic roots along with their multiplicities of $N N^1$, where N is the incidence matrix of a two associate class partially balanced incomplete block design.

(b) Write the parameters of a triangular association scheme.

(4+6+6+4) = 20

5.(a) Describe the method of differences for constructing group divisible designs.

(b) Determine the nature and parameters of the design obtained after omitting one complete replication of a resolvable balanced incomplete block design.

(c) In a semi-regular group divisible design, show that every block has the same number of treatments from each group of the association scheme.

(8+6+6) = 20

6.(a) Describe the concept of confounding in factorial designs. When do we say a factorial design is completely balanced? Show that the total relative loss of information in a connected confounded design in b blocks of r replications is $\frac{b}{r} - 1$.

6.(b) Construct a $(3^4, 3^3)$ experiment confounding the degrees of freedom carried by the pencils P (1,C,C,C), P (C,1,1,C) and P (C,1,C,1). Detect the other confounded pencils.

$(4+2+6+6+2) = 20$

7.(a) Describe the concept of fractional replication. Construct a $\frac{1}{2^2}$ replicate of a 2^6 experiment with the identity relationship I = ABCD = ABEF = CDEF and find the aliases sets.

(b) Give the method of constructing a second order rotatable design using a Balanced Incomplete Block Design with $r = 3\lambda$.

$(4+6+4+6) = 20$

8. Write short notes on any three of the following:

(a) Projective geometries and their applications in design of experiments,

(b) Block structure of balanced incomplete block designs,

(c) Generalized interaction,

(d) Hadamard matrices and their construction from Balanced Incomplete Block Designs,

(e) Youden Square construction.

$(6\frac{2}{3} + 6\frac{2}{3} + 6\frac{2}{3}) = 20$

.....

GROUP E : SAMPLE SURVEYS - Special Paper II
(Organisational Aspects)

(Answer any five questions from this group)

1.(a) Discuss the importance of (i) field scrutiny, (ii) pre-punching scrutiny and (iii) post Machine scrutiny in a sample survey.

(b) Why is it necessary to inspect periodically the field work undertaken by investigators? Discuss relative merits and demerits of intimated and surprise inspections.

$(3+3+3+4+7) = 20$

2. A survey of Goods Traffic in the mechanised and non-mechanised sectors, is proposed to be undertaken in a certain State during the period July 1971 to June 1972. Spell out the various stages of work that would be involved, upto the writing out of the report. Indicate clearly the items on which information should be collected. Draft a few paragraphs which you would like to be included in the instructions about the "objectives" and "concepts and definitions" of the survey.

$(5+7+3+5) = 20$

3.(a) Discuss the circumstances under which you would prefer

(i) Mechanical Tabulation

(ii) Manual Tabulation

(iii) Combination of manual and mechanical Tabulation.

(b) Describe a suitable method for evaluating daily performance of a computor, engaged in compilation and computational work.

(d) Draft a set of proforma together with brief instructions for the Monthly Progress Report in respect of the following categories of field workers:

(i) Investigators,

(ii) Inspectors,

and (iii) Field Scrutineers.

(3+3+2+3+3+3+3) = 20

4. A survey of Distributive Trade for estimating 'trade margins' by commodity groups, is to be conducted. Draft the following:

(i) Schedule for the Survey,

(ii) Field scrutiny Programme for the proposed schedule,

and (iii) Various compilation and computation sheets, assuming that the tabulation is to be undertaken manually.

(9+5+6) = 20

5. A land-holding survey is proposed to be undertaken in a State, which would throw up the following broad results:

(i) Operational holdings by size-groups,

(ii) Cropping pattern by sizes of operational holdings,

(iii) Operational holdings by irrigation status.

The sample is to be confined to only 2000 villages and a scheme of interpenetrating sub-samples, is to be adopted. Estimates would be required for the various regional levels in addition to the State as a whole.

You are required to draw up realistic estimates of a Budget for this purpose.

[Basis of calculation of your estimates should be very clearly indicated at each stage].

(12 + 8) = 20

6.(a) Discuss the relative merits and demerits of the following alternative forms of presenting a sample survey report relating to economic conditions of a section of the society:

(i) Report giving only statistical tables with brief description of the survey.

(ii) Report giving detailed economic analysis of the results.

Please turn over

6.(b) Give the headings for the various chapters and Appendices of the proposed report, relating to the survey indicated in question 5 above.

(c) If the resources available do not permit calculation of standard errors of various estimates, how can the reliability of the estimates be brought out in the Report?

(5+5+0+4) = 20

7. Write short notes on any four of the following:

(i) Pre-pilot and Pilot survey.

(ii) Role of Refresher Training to field workers.

(iii) Sample versus complete checking of compilation and computational work.

(iv) Work Audit.

(v) Employment of field investigators at monthly versus piece rates.

(vi) Flow charts.

(4 X 5) = 20

*****

GROUP F : TECHNIQUES OF COMPUTATION
Special Paper II (Practical)
(No candidate available)

GROUP G : STATISTICAL INFERENCE - Special Paper II
(Statistical Inference - Special Topics)

(Answer any five questions from this group)

1.(a) Define Sequential Probability Ratio Test (SPRT) and give an example of a test which is sequential but not a SPRT.

(b) Show that under certain conditions (to be stated), a SPRT terminates with probability one.

(c) Obtain the formula for the O.C. function of a SPRT for a simple Hypothesis against a simple alternative, neglecting 'the excess over the boundaries'.

2.(a) Define the two-sample Mann-Whitney test statistic V, show that this test can equivalently be based on $p = \sum_{j=1}^{n_2} s_j$, the sum of the ranks in the second sample and obtain the relation between p and V.

(b) Frame the single sample location problem and for this case, show that i $(x_1, \ldots, x_n)$ (= number of positive x_i's) is a sufficient statistic. Hence, or otherwise, show that the sign test is Uniformly Most Powerful (UMP) for the one-sided location problem.

2.(c) If $f(\underset{\sim}{x})$ is sufficient (θ) for the class of measures $\{P_{\theta\eta} \mid (\theta,\eta) \in \Omega \times H\}$, then show that there is a UMP test for testing,

Hypothesis : $\theta = \theta_0$, $\eta \in H$ against alternative : $\theta = \theta_1$, $\eta \in H$.

3.(a) Considering the case of a single sample define U-statistics with illustrations.

(b) State and prove Hoeffding's limiting theorem on U-statistics.

(c) Let $U^{(1)}$, $U^{(2)}$ be two U-statistics based on the symmetric kernels, $f(x_1, \ldots, x_{m_1})$ and $g(x_1, \ldots, x_{m_2})$ respectively. Prove that the covariance of $U^{(1)}$ and $U^{(2)}$ is given by

$$\mathrm{Cov}\,(U^{(1)}, U^{(2)}) = \binom{n}{m_{12}} \sum_{k=1}^{m_{12}} \binom{m_{12}}{k} \binom{n - m_{12}}{m_{12} - k} \zeta_k^{(1,2)}$$

where $m_{12} = \min(m_1, m_2)$ and $\zeta_k^{(1,2)}$ is the covariance between $f_k(X_1, \ldots, X_k)$ and $g_k(X_1, \ldots, X_k)$ where $f_k(x_1, \ldots, x_k) = E\, f(x_1, \ldots, x_k, X_{k+1}, \ldots, X_{m_1})$ and g_k is defined similarly.

4.(a) What is a rank test? What are its special advantages?

(b) Distinguish between a most powerful and locally most powerful rank test, against a specific parametric alternative.

(c) Derive the locally most powerful rank test for the two-sample problem against location alternatives.

5.(a) Let the column vectors of $\underset{\sim}{X} : p \times n$ be independently and identically distributed as $N(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$. Obtain likelihood ratio test procedures for testing :

(i) H_0 ($\sigma_{ij} = 0$ for $i \neq j$ and $\sigma_{ii} = \sigma^2$ (unknown for all $i,j = 1,2,\ldots,p$) against $H \neq H_0$,

and (ii) H_0 ($\sigma_{ij} = 0$ for $i \neq j = 1,2,\ldots,p$) against $H \neq H_0$.

(b) Obtain the moments of the test statistics developed in (a).

(c) Obtain the null distributions for $p = 2$ of the test statistics for the two cases mentioned in (a).

6. If π_1 and π_2 are two multivariate normal distributions differing only in means, show that

(i) the criterion for discriminating π_1 and π_2 is a linear function $L(\underset{\sim}{x})$, depending on $\underset{\sim}{x}$, mean vectors $\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2$, and covariance matrix $\underset{\sim}{\Sigma}$;

Please turn over

6. (contd) (ii) $V\{L(\underset{\sim}{x})\} = E\{L(\underset{\sim}{x})|\pi_1\} - E\{L(\underset{\sim}{x}) | \pi_2\}$

and (iii) If the procedure is so formulated that the probability of wrong classification is same under π_1 and π_2, then show that this common value is a function of $\sqrt{V\{L(\underset{\sim}{x})\}}$.

Since $L(\underset{\sim}{x})$ depends on $\underset{\sim}{\mu}_1$, $\underset{\sim}{\mu}_2$ and $\underset{\sim}{\Sigma}$, the unknown parameters, show how you will estimate $L(\underset{\sim}{x})$ on the basis of independent samples from π_1, π_2. Obtain the bias, if any, of your estimate. If the samples from π_1 and π_2 are equal, how can you ensure equal probability of wrong classification here? Give a test based on your estimate for testing H_0 ($\pi_1 = \pi_2$) against $H \neq H_0$.

7. Let the column vectors of $\underset{\sim}{X} : p \times n$ be independently distributed as normals with $E(\underset{\sim}{X}') = \underset{\sim}{A}\underset{\sim}{\xi}$ and with common covariance matrix $\underset{\sim}{\Sigma} : p \times p$ where $\underset{\sim}{A} : n \times m$ is a known matrix, and $\underset{\sim}{\xi} : m \times p$ and $\underset{\sim}{\Sigma}$ are unknown matrices.

(a) Find the best linear unbiased estimate of $\underset{\sim}{\psi} = \underset{\sim}{C}\underset{\sim}{\xi}\underset{\sim}{M}$ where $\underset{\sim}{C}$ and $\underset{\sim}{M}$ are given matrices, with appropriate definition of bestness and with appropriate assumptions on the matrix $\underset{\sim}{C}$.

(b) Find the distribution of the estimate derived in (a).

(c) Develop Hotelling T-statistic for testing H_0 ($\underset{\sim}{\psi} = \underset{\sim}{\psi}_0$) against $H \neq H_0$. Hence or otherwise, give Hotelling T-statistic for testing equality of k normal populations, having common covariance matrix.

8. Write short notes on any three of the following:

(a) Sequential estimation.

(b) Optimality of SPRT.

(c) Tolerance limits.

(d) Principal components.

(e) Factor analysis.

.....

Paper IX - GROUP II : PROBABILITY THEORY
Special Paper II - Limit Distributions
(No candidate available)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1970

Paper X - Subjects of Specialisation - (Special Paper III)

(Practical)

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.

Paper X - GROUP A : ECONOMIC STATISTICS

Special Paper III - Practical

(No candidate available)

Paper X - GROUP B : TECHNO-COMMERCIAL STATISTICS (Special Paper III- Practical)

I : Statistical Quality Control (50)

II : Operations Research/Elements of Book-keeping and Accountancy (30)

III : Statistical Methods in Business (20)

Section I : Statistical Quality Control

(Answer any two questions from this section)

1. The following table gives the mean values and ranges in 30 samples (of size 4 each), of percentage carbon content in commercial metal sheets, as obtained from test records.

The data were collected for a continuous period of 30 days assuming that the percentage carbon content within a day is unlikely to show high variation from sheet to sheet.

Sample	Average	Range	Sample	Average	Range
1	11.10	0.6	16	11.45	1.3
2	11.70	1.2	17	11.55	1.6
3	11.35	1.0	18	9.98	0.4
4	11.25	1.0	19	10.78	1.2
5	11.40	2.0	20	11.23	0.7
6	11.00	0.6	21	10.93	1.7
7	11.20	1.0	22	11.50	2.7
8	11.35	1.2	23	10.78	0.7
9	11.50	2.0	24	10.95	1.1
10	10.88	1.1	25	11.48	2.8
11	10.85	1.0	26	10.80	0.4
12	11.53	1.2	27	12.20	2.0
13	11.15	0.8	28	11.88	1.5
14	11.28	1.0	29	11.23	0.8
15	11.00	0.8	30	11.30	0.6

1. (a) Test by means of control chart whether the process is under statistical control.

(b) It was further suggested to work in respect of only lower specification limit (9%), using the following type of control chart.

"Plot only the smallest observation in each sample of 4. Take action if any plotted point falls below the control limit L for the smallest observations. It is further given that the probability of unnecessary action is 0.025 and L is obtained for the average level of the process as obtained in (a)".

Sketch the O.C. curves of this control chart and the corresponding $\bar{X}$-chart and hence comment about their relative efficiencies. [Assume that the process variability remains stable at the level as obtained in (a)].

(10+15) = 25

2.(a) Construct three single sampling plans (plans 1,2 and 3) by attributes, having AQL (p_1), Producer's Risk (α), LTPD (p_2) and consumer's Risk (β) approximately as given below:

Plan	p_1(%)	α	p_2(%)	β
1	0.6	0.02	3	0.15
2	1.0	0.15	4	0.10
3	1.2	0.20	5	0.02

(b) Sketch on approximate AOQ curve for any one of the plans 1, 2 and 3 and hence read the value of AOQL (p_L). Use the lot size as 10,000 and assume that the rejected lots are given 100% inspection and any defective item found is replaced. How will the above plan with AOQL as p_L change if the acceptance number is doubled? (use the formula $P_L = y\left(\frac{1}{n} - \frac{1}{N}\right)$ of Dodge and Romig, where/symbols have the usual meaning).

(9+16) = 25

3. An experimenter decided to explore the relationship between the purity of a chemical intermediate (y) and two independent variables, temperature of the reacting mass (X_1) and rate of addition of reagents (X_2). It was expected that the response function might be approximated by the second order model. The design proposed was a Hexagon design, with four centre points for a total of N = 10 experiments. The following is the table of results:

Run No.	Factor levels: Temp	rate	Observation y	Matrix of independent variables: X_0	X_1	X_2	X_1^2	X_2^2	X_1X_2
4	15°C	2 hrs.	54.6	1	-1.00	0	1.00	0	0
2	60	3	78.7	1	0.50	0.87	0.25	0.75	0.43
1	60	1	78.9	1	0.50	-0.87	0.25	0.75	-0.43
5	45	2	90.4	1	0	0	0	0	0
3	45	2	87.5	1	0	0	0	0	0
8	30	3	71.7	1	-0.50	0.87	0.25	0.75	-0.43
6	30	1	59.8	1	-0.50	-0.87	0.25	0.75	0.43
10	75	2	91.0	1	1.00	0	1.00	0	0
9	45	2	91.0	1	0	0	0	0	0
7	45	2	88.3	1	0	0	0	0	0

$\Sigma y^2 = 64,304.29$ $\Sigma y = 791.9$

Please turn over

1. (contd.) (a) Analyse the experiment by fitting a second degree polynomial for the yield, in terms of X_1 and X_2.

(b) Examine whether a second order surface is adequate by setting up a suitable analysis of variance table.

(10 + 6) = 25

Section II - OPERATIONS RESEARCH

(Attempt question 1 and any other one from this section)

1. A large industrial organisation has entered into an agreement with the Government, to supply every month specified quantities of a certain product to four Government depots D_1, D_2, D_3 and D_4. They have three plants P_1, P_2 and P_3, where these products can be manufactured. The following information is given:

Plant	Capacity/month (units)	Cost/unit (Rs.)
P_1	6500	3
P_2	3500	4
P_3	7500	2

The unit transportation cost (in Rs.) from the factories to the depots and the requirement of each depot per month is as follows:

	P_1	P_2	P_3	Requirement/month (units)
D_1	10	7	1	3000
D_2	13	5	3	2000
D_3	15	10	8	5000
D_4	17	12	9	6000

Formulate this as a transportation problem and find out how much each plant should produce every month and how they should be shipped.

(18)

2. The following failure rates have been observed for a certain type of light bulbs:

End of week	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Prob. of failure to date	.05	.13	.25	.43	.68	.88	.98	1.00

The cost of replacing an individual failed bulb is Rs.1.25. The decision is made to replace all bulbs simultaneously at fixed intervals, and also to replace the bulb as they fail in service. If the cost of group replacement is 0.30 paise per bulb, what is the best interval between group replacement? At what group replacement price per bulb, would a policy of strictly individual replacement, become preferable to the adopted policy?

(12)

3. In a hospital clinic a doctor averages 4 minutes on each phase of the check-up and the distribution of time spent on each phase is approximately exponential. If each patient goes through four phases in the check up and if the arrivals of the patients to the hospital clinic are approximately poission at an average rate of three per hour.

(a) What is the average time spent by a patient waiting in the hospital clinic for the doctor attention?

(b) What is the total average time a patient has to spend in the hospital clinic?

(c) What is the average number of patients in the clinic?

(12)

Section III - STATISTICAL METHODS IN BUSINESS (20 marks)

(Attempt question 1 and any other one from this section)

1. The following table shows the percentage of yearly output of a certain commodity exported over a period of 10 years.

Year	Quarter ended			
	March	June	September	December
1	24.7	22.0	21.0	21.7
2	24.7	22.1	20.3	21.9
3	25.2	21.8	21.0	22.1
4	23.7	20.9	19.6	21.5
5	21.0	18.7	17.3	19.1
6	20.7	17.8	17.0	17.7
7	17.7	14.6	13.8	15.4
8	17.2	13.6	11.8	14.9
9	15.5	11.7	9.5	11.8
10	15.8	11.6	9.8	12.0

Find a trend line suitable for estimating annual export rate. Make a seasonal analysis and get seasonal factors expressed as the percentage of the current trend export rate and finally estimate the uncertainity of prediction under stable conditions, as a percentage of current trend value.

(12)

2. A manufacturer of brand A automobiles believed that his cars would last longer than the automobiles of competitors (say brands B and C). A survey was made of cars manufactured exactly 10 years ago. The results are summarised below:

Brand	Number of cars which ran without major overhaul for					
	0-1 yrs	2-3 yrs	4-5 yrs	6-7 yrs	8-9 yrs	No overhaul yet
A	15	37	55	108	118	60
B	5	13	190	205	386	400
C	55	160	300	106	55	40

Please turn over

2. (contd) What do these figures prove? State clearly the population from which samples have been drawn. What interpretations can and cannot be made from the data ?

(8)

3. At an industrial fair 10 different brands of products marked 1 to 10 were ranked by two different judges. The results were as in the table:

Product brands →	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Judge 1	9	4	3	7	2	1	5	8	10	6
Judge 2	7	6	4	9	2	3	8	5	10	1

Carry out an analysis to investigate the state of agreement or disagreement between the two judges.

(8)

....

Section II (Alternative): Elements of Book-keeping and Accountancy

(Special Paper III): Practical

(Answer any two questions from this section)

Full marks 30

1.(a) Explain with reference to their practical importance:

(i) Preliminary Expenses (iii) Deal account

(ii) Deferred revenue expenditure (iv) Promissory note

(v) Days of grace.

(b) What do you mean by Depreciation? Calculate @ 10% per year depreciation on an asset, assuming its value to be Rs.10,000/-, by using two methods- describing the practical importance of each of these methods.

2. The following figures appear at the Trial balance of PQR Company at at 31.12.69.

(a) (i): Sundry Debtors Rs. 20,600

(ii) Provision for doubtful Debts. Rs. 400.

Show the necessary accounts after writing off Rs.600/- as Bad debts. The provision for Doubtful debts is to be maintained at 5% on Sundry Debtors.

(b) Stock on 31.12.69 was valued at Rs. 14,600/-

A fire occurred on 25.12.69 in the godown, and stock worth Rs.5000/- was destroyed. It was fully insured and the Insurance Company also admitted the claim in full. Journalise.

Please turn over

2.(c) How do you show in your Profit & Loss account,

(i) Nominal account expenses prepaid?

(ii) Nominal account expenses outstanding? Make use of imaginery figures for this and indicate the basis of your calculations.

(15)

3.(a) What is "Secret Reserve"? Does it appear on the Balance Sheet? Explain what do you mean by Capital reserve and general reserve.

(b) The following are the Credit sale of ABC company for the month of January 1970.

(i) 1st January Rs.100/-. (ii) 15th January Rs.300/-;
(iii) 31st January Rs.1000/-.

Sundry Debtors as on 1.1.70 on the general Ledger was Rs.3000/-. Cash received against credit sale on this month is Rs.500/-.

Show the sales Day book and Sundry Debtors account, especially showing how the sales figures from the day book, is transferred from sales Day book to General Ledger.

(15)

.....

Group C : Biometric Methods
Paper III - Practical
(No candidate available)

.....

Group D : Design and Analysis of Experiments
Paper III - Practical

(Answer any three questions from this group)

1.(a) An experiment is to be conducted for studying the effects of ammonium sulphate, bone meal and farm-yard manure and their interactions. It was decided to take three levels of ammonium sulphate and two levels each of bone meal and farm-yard manure.

Prepare the layout plan of a suitable confounded design, complete with randomisation, assuming that the available resources are sufficient for just three replications.

(b) For a varietal trial on 31 strains of wheat, an area of about two acres is available. The soil being rather heterogeneous a block-size of not more than 1/10th of an acre can be recommended. If the plot size of 1/100th of an acre is taken, give the complete layout of the experiment.

(10+10) = 32

2. In order to study the effects of 7 'trace' elements [A,B,C,D,E,F and G] and potash (k), as also their interactions on the yield of paddy, two levels of each of these eight factors were taken and an

Please turn over

2. (contd.) experiment in $\frac{1}{2^3}$ replicate of a 2^8 factorial, arranged in 4 blocks of 8 plots each, was conducted. The layout plan and the yield of dried paddy in lbs. per plot of size 43.5' × 16.5', are given below:

Block I		Block II		Block III		Block IV	
Treatment	Yield	Treatment	Yield	Treatment	Yield	Treatment	Yield
adg	28.00	aeg	31.34	bcefk	28.22	beefk	36.08
cdef	32.81	abcdf	20.85	acfk	29.00	abgk	27.77
acek	33.28	adfgk	32.20	adefg	30.32	abcdo	32.24
bdk	33.13	ef	31.02	egk	32.05	adfk	34.20
abf	31.30	adek	30.45	cd	31.73	bdfg	31.73
abcdefgk	35.43	bek	38.13	abcdgk	28.81	acefg	33.58
fgk	34.15	bdeg	30.45	abe	28.46	cdegk	31.80
beeg	35.57	abefgk	33.64	befg	28.01	1	35.82

Identify the defining contrasts and the confounded interactions. Carry out also the analysis of variance, assuming the second and higher order interactions to be negligible. Hence interpret the results.

(8+26) = 32

3. An experiment with a strip-plot design, was carried out at a research station to study the effects of three row spacings (R_1, R_2, R_3) and three plant spacings (P_1, P_2, P_3) and their interactions, on the yield of Sesamum seeds. The yield data in gms per plot (of size 12' x 6') are given below:

Yield of Sesamum seeds in gms per plot

		Replications		
		R_1	R_2	R_3
I	P_1	147	128	80
	P_2	80	81	30
	P_3	100	34	66
II	P_1	86	114	158
	P_2	90	67	126
	P_3	96	68	115
III	P_1	236	105	187
	P_2	130	45	53
	P_3	161	101	102
IV	P_1	201	203	223
	P_2	198	135	230
	P_3	153	193	142
V	P_1	205	157	165
	P_2	206	124	273
	P_3	184	162	201
VI	P_1	323	213	227
	P_2	225	240	331
	P_3	284	150	273

Please turn over

3. Analyse the data and draw your conclusions.
(contd.) Find the standard error of the estimate of the difference between $R_2 P_1$ and $R_1 P_3$.

(20+6) = 32

4. In an experiment, it was desired to compare 10 varieties of Soyabeans. The actual lay out adopted and the yields in lbs. per plot, are given below:

Soyabean Experiment

Block I	(1)	(2)	(3)	(4)
	32	45	37	40
Block II	(6)	(1)	(7)	(5)
	88	15	43	85
Block III	(8)	(8)	(4)	(9)
	68	66	56	21
Block IV	(2)	(5)	(9)	(10)
	53	04	13	12
Block V	(10)	(8)	(3)	(7)
	18	60	35	40

(a) Identify the layout adopted and give the parameters of the design.

(b) Carry out the statistical analysis both with and without recovery of inter-block information and calculate the efficiency of the design over a Randomised Block Design (R.B.D.) plan with the same experimental material.

(c) Calculate the standard errors for comparing varietal defferences and report on your findings.

(6+18+8) = 32

Neatness 4

Group E : Sample Surveys

Paper III : Practical

(Answer any three questions from this group)

1. It is required to estimate the total population of a region divided into 3 geographical strata. Villages are selected by simple random sampling with replacement in each stratum. The relevant data are given below:

stratum number	number of villages	average village population	standard deviation	C_h* (in rupees)
(1)	(2)	(3)	(4)	(5)
1	2,000	480	560	40
2	1,500	950	810	70
3	1,000	1,250	1,100	90

(*C_h = cost of surveying one village in the h^{th} stratum.)

(a) Assuming a total budget of Rs.20,000, find the optimum sample sizes for the three strata.

(b) Find the proportional allocations for the same total sample size as obtained in (a).

(c) Obtain the cost-efficiency of scheme (a) as compared to scheme (b).

(d) Also calculate, the sampling variance in case of unstratified simple random sampling with replacement, for the same total sample size as in (a).

(8+4+12+8) = 32

2. In order to estimate the total number of children in the age group 5 - 14 years, a two-stage sampling scheme was adopted. The urban blocks (first-stage units) were selected with probability proportional to their 1961 census population and with replacement; the households (second-stage units) were selected systematically. The data are given below:

SAMPLE DATA for Question 2.

sample block	1961 census population	total number of households	number of sample households
(1)	(2)	(3)	(4)
1	1100	220	8
2	850	190	7
3	600	150	6
4	900	200	8
5	750	160	6
6	500	100	5

Please turn over

SD(N70)-X(10)

2. (contd.)

SAMPLE DATA for Question 2.(contd.)

sample block	no. of children in the sample households							
(1)	(5)							
1	2,	1,	3,	C,	1,	6,	4,	3
2	1,	3,	1,	1,	0,	5,	3,	-
3	1,	2,	1,	4,	2,	2,	-	-
4	5,	2,	1,	2,	0,	3,	C,	3
5	3,	4,	1,	1,	1,	2,	-	-
6	2,	1,	3,	5,	C,	-	-	-

City total : 1961 population = 200,CCO

total number of blocks = 300

total number of households = 46,CCC

(a) Estimate the total number of children in the city as well as the relative standard error of this estimate.

(b) Estimate the efficiency of this scheme compared to the case when households are selected directly by simple random sampling without replacement.

(20+12) = 32

3. For estimating the total cattle population of an area, a sample of 20 villages was selected from 1CCO villages in that area with simple random sampling without replacement. The table below gives the observed number of cattle (y) at present and also the number of cattle as counted in a previous census (x) for each sample village.

SAMPLE DATA

sample village	x	y	sample village	x	y	sample village	x	y
1	05	78	8	15	0	15	92	60
2	100	135	9	42	50	16	73	97
3	330	375	10	70	70	17	30	55
4	80	100	11	85	100	18	77	95
5	50	55	12	48	60	19	150	175
6	110	110	13	88	80	20	123	145
7	75	100	14	125	150			

total number of cattle in the previous census (= X) = 1CCC,CC

(a) Calculate the ratio estimate of total number of cattle.

(b) Calculate the bias and variance of the estimate in (a).

(c) Also obtain the regression estimate of the total number of cattle.

(8+10+8) = 32

Please turn over