INDIAN STATISTICAL INSTITUTE 624

QUESTION PAPERS

for

The Statistician's Diploma Examination

March & September 1956

Price Re. 1/-

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PART I (THEORETICAL)

Time : 4 Hours | Full Marks : 100

N.B. (a) Answers to the different groups are to be given in separate books.
(b) Figures in the margin indicate full marks for the questions.

GROUP A

Note : Answer Q-1 and *ANY TWO* from the rest.
Answers must be brief and to the point.

1. Let $f(x)$ be a function whose values at $x = 0, -1, -2, \ldots,$ are known. Let Δ be the usual difference operator defined as

$$\Delta f(x) = f(x+1) - f(x)$$

Derive the interpolation formula

$$f(x) = f(0) + x\,\Delta f(-1) + \frac{1}{2!}\,x(x+1)\Delta^2 f(-2) + \ldots$$

and briefly answer the following questions :—

(a) Has the above interpolation formula got any standard name ?

(b) If $f(x) = \sin \pi x$ then do you expect the above interpolation formula to be a good one ? If not, why not ?

(c) Give an expression for the remainder term when you stop at the nth term of the right hand expansion.

(d) If $f(x)$ were known to be a polynomial of degree not exceeding five then how many terms in the right hand expansion will give you an exact fit ?

(e) Discuss the suitability of using the above interpolation formula for finding the value of $f(x)$ at a positive x. (24)

2. An urn contains N tickets numbered 1 to N.

(a) If n tickets are drawn one by one and *without* replacement from the urn and if X be the sum total of the numbers drawn, then prove that $EX = \frac{n(N+1)}{2}$.

(b) If tickets are drawn one by one and *with* replacement from the urn *until some* ticket appears *twice* and if ν be the number of tickets drawn then prove that

$$P(2 \leq \nu \leq N+1) = 1 \text{ and that}$$

$$P(\nu = i+1) = \frac{N(N-1)\ldots(N-i+1)}{N^i} \cdot \frac{i}{N} \quad (i = 1, 2, \ldots, N)$$

(c) In the case of (b) indicate how you will find the modal value (mode) of ν. (18)

3. Let X be an arbitrary random variable. You are required to find the probability that $X>0$. Under *each* of the following assumptions about X find $P(X>0)$ or give an expression from which the probability can be computed. If in a particular case, you find that the required probability cannot be evaluated then state that specifically and give instead the best bounds (for the required probability) that you can find.

Assumptions :

(a) X^3 is a continuous random variable with median greater than zero.

(b) $2X+1$ is the sum of two independent standard normal variables.

(c) X has means -3 and variance 1.

(d) X is the ratio of two independent continuous random variables each with zero as its median.

(e) $X = Y-Z$ where Y and Z are independent and identically distributed continuous random variables.

(f) X does not take values less than -4 and $EX = -3$. (18)

4. Let (X,Y) be a pair of discrete random variables each taking the three values 0,1 and 2 with the following joint distribution.

Y \ X	0	1	2
0	.40	.15	.05
1	.10	.10	.10
2	.00	.05	.05

Below we give a number of statements which are either true or false. Indicate in each case whether the statement is true or false adding a remark or two in support of your conclusion.

(a) The probability that $X = Y$ is greater than the probability that $X \neq Y$.

(b) The marginal distributions of X and Y are the same.

(c) Given that $X = 0$, the condkitional probability that $Y = 2$ is nil and the conditional probability that $Y < 2$ is .50.

(d) The continued expectation of Y, given that $X = 1$, is $\frac{2}{3}$.

(e) X and Y are dependent random variables.

(f) X and Y are negatively correlated.

(g) The joint probability that $X = 0$ and $Y = 0$ is equal to the probability that $X+Y = 2$.

(h) The correlation ratio between X and Y is one. (18)

Group B

Answer ANY TWO questions

5. (a) Give illustrations of various tests depending on the distribution. (8)

(b) Consider a 2×2 table containing the frequencies with respect of two attributes obtained by one of the following two methods of sampling. The first is to draw a random sample from a population and then classify the individuals according to the two attributes. The second is to classify the whole population into two groups according to one attribute and take samples of fixed size from each group and classify the individuals in each sample with respect to the second attribute. Find the large and small sample (exact) tests of independence of attributes for different methods of sampling. (12)

6. (a) If x_1, x_2 and x_3 are three variates measured from their respective means as origin and of equal variances, find the coefficient of correlation between x_1+x_3 and x_2+x_3 in terms of r_{12}, r_{13} and r_{23} and show that it is equal to

$$\text{(i) } \frac{r_{12}+1}{2} \text{ if } r_{13} = r_{23} = 0 \text{ or (ii) } \frac{r_{12}+3}{4} \text{ if } r_{13} = r_{23} = 1. \qquad (10)$$

Comment on the results.

(b) If x_1, x_2 and x_3 are three variates measured from their respective means as origin and if e_1 is the expected value of x_1 for given values of x_2 and x_3 from the linear regression of x_1, on x_2 and x_3, prove that

$$\text{cov}(x_1, e_1) = \text{Var}(e_1) = \text{Var}(x_1) - \text{Var}(x^1 - e_1).$$

Mention the use of the correlation coefficient between x_1 and e_1. (10)

7. (a) Starting from the joint distribution of $\bar{x}$ and s^2 obtain the distribution of the t statistic

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

and show how it is used in testing the significance of an assigned mean value. (8)

(b) Suppose we want to test the linearity of regression of the height of the daughter on the heights of her father. How do you collect the data and what tests do you apply for examining the linearity of regression? Give full details of the computations involved in the test procedure without the actual proofs. (12)

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

Paper II (Theoretical)

Time : 4 Hours | Full Marks : 100

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

Group A

Answer ANY THREE of the following :—

1. Explain what is meant by a 'frame' for a sample survey, and the various types of defects to which it may be subject. Discuss how far the National Register of Citizens, prepared during the 1951 census, could be used as a frame for a socio-economic urban survey.

2. Explain how the official index number of industrial production is compiled, pointing out its limitations from the point of view of coverage, representativeness and treatment of seasonal items.

3. Describe the various methods of computing national income. Point out the appropriateness or otherwise of these methods, in estimating the contribution to national income in India from different sectors, taking into account the availability and quality of data in these sectors.

Why and how are estimates of national income for a series of years adjusted for price variations over time ?

4. (a) How are foreign trade statistics collected and compiled ?

(b) In order to prepare a working paper on the relationship between amount of investment and volume of employment in some of the organised industries in India like iron and steel, cotton-textiles, jute, coal etc., state the names of the official publications with sources, you will like to consult. Give a scheme of the report including sketch of the tables you will like to include.

Group B

Answer ANY THREE of the following :—

5. State the circumstances in which split-plot designs are recommended in practice. Prepare the layout of a simple split-plot design involving 2 irrigation practices, 3 varieties and 2 manures and indicate the structure of the analysis of variance, separating the error (or errors) against which the significance of the different comparisons has to be tested.

6. Explain what is meant by confounding and indicate its usefulness in the design of Agricultural Experiments. Given four factors at two levels each, show how you would prepare the design to test the main effects and the interactions (of lower order). Would you like to introduce confounding in such a design. If so, why ?

7. Explain clearly the meaning of the following terms, as understood in the context of Factory Analysis :—

(i) Specificity, (ii) Reliability, (iii) Communality, (iv) Uniqueness

8. A doubly heterozygous individual (*Aa, Bb*) is crossed with a doubly recessive one (*aa, bb*). Enumerate the different possible classes of offsprings and write down the probabilities for them. How do you test the following hypotheses on the basis of observed data :

(a) The character *A* is segregating in a 1 : 1 ratio

(b) The character *B* is segregating in a 1 : 1 ratio

(c) There is no linkage between *A* and *B*.

9. Sketch the outlines and main components of a fertility table and a mortality table and explain what data would be needed (and how you would look for them) to construct these tables. To what use can you put these tables ?

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PAPER III (THEORETICAL)

Time : 4 Hours Full Marks : 100

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

(b) Attempt ANY TWO questions from each group.

(c) All questions carry equal marks.

GROUP A

1. State clearly the properties of maximum likelihood estimates.

The probability density function for the three parameter logarithmic normal distributions is given below :—

$$f(x) = \frac{1}{(\sigma x-a)\sqrt{2\pi}}\, e^{-\left(\frac{1}{2}\right)\log^2 (x-a)/\beta}$$

Derive the maximum likelihood equations for estimating the population parameters. Give an outline of any method that should be followed for the numerical solution of the above equations.

2. What is the difference between 'point estimation' and 'interval estimation' in theoretical statistics ?

Give a brief description of the method of computing confidence intervals on the basis of a sample n in the following situations :—

(a) the mean of a normal population when the variance is known.

(b) the mean of a normal population when the variance is unknown.

(c) the ratio of means of two normal populations.

3. Give the definition of a sufficient statistics. State a necessary and sufficient criterion, in terms of the density function of a chance variable X, for the existence of a sufficient statistics.

Let the density function of X be given by $p_\theta(x)$. Y is an unbiased estimate of θ and T, a sufficient statistic for θ. Show that if

$Z = E(Y/T)$, the Z is also unbiased for θ and Var $Z \leq$ Var Y.

$X_1, \ldots, X_n$ are independent rectangular random variables in $(0, \theta)$. Obtain a sufficient statistic for θ.

Group B

4. (a) Describe the likelihood ratio test and state some of its desirable properties

(b) X is a random variable which assumes the values 0, ± 1, ± 2 with probabilities as follows :—

$X =$	-2,	$+2$,	-1,	$+1$,	0
H_0 :	$\frac{1}{2}\alpha$	$\frac{1}{2}\alpha$	$\frac{1}{2}-\alpha$	$\frac{1}{2}-\alpha$,	α
H :	$c\theta$	$(1-\theta)c$	$\frac{1-c}{1-\alpha}(\frac{1}{2}-\alpha)$,	$\frac{1-c}{1-\alpha}(\frac{1}{2}-\alpha)$,	$\alpha . \frac{1-c}{1-\alpha}$

Here α and c are constants such that $0 < \alpha < \frac{1}{2}$,

$\frac{\alpha}{2-\alpha} < c < \alpha$, and θ is a parameter, $0 \leq \theta \leq 1$.

Derive the likelihood ratio test for testing the simple hypothesis H against the composite alternative H_0 at the level of significance α.

(c) Show that in this example one can do better than the L.R. test by simply rejecting H_0 with probability α without observing X at all.

5. (a) State and prove Neyman and Pearson lemma useful in the derivation of a test for a specified alternative.

(b) Show that a uniformly most powerful test exists for testing the hypotheses that mean of a normal population is 4 against the alternatives exceeding 4.

(c) Develop the concept of locally most powerful tests and give your opinion about the use of such tests. Derive such a test for the above hypothesis when both sided alternatives are considered.

6. (a) Prove that in a bivariate normal population the regression of one variable on the other is linear.

(b) How do you test simultaneoulsy on the basis of sample whether the means of the two variables have assigned values ?

What do you think of the alternative method of testing for each mean separately?

(C) Propose a test for the hypothesis that the variances of the two variables are equal without making any assumption about their correlation.

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PAPER VI (PRACTICAL)

Time: 6 Hours | Full Marks: 100

N.B. (a) Figures in the margin indicate full marks.

(b) Use of calculating machines is permitted.

1. (a) Find $\int_0^{0.8} \cos x \, dx$ by quadrature from the following values:—

x	$\cos x$
0	1.000 000
0.1	0.995 004
0.2	0.980 067
0.3	0.955 336
0.4	0.921 061
0.5	0.877 583
0.6	0.825 336
0.7	0.764 842
0.8	0.696 707

(15)

(b) The values of q the rate of mortality for age x according to a standard table are given below:

Age	q_x
80	0.134
82	0.154
84	0.176
86	0.200
88	0.227

Find q_x for ages 81 and 85 by using appropriate interpolation formulae. (15)

2. Fit a *third degree* polynomial by the method of orthogonal polynomials to the following results of an experiment on the effect of time of mixing of dough on loaf volume. Examine the significance of the coefficient of the cubic term.

Mixing time in minutes	Loaf volume in cubic centimeters less 530
1.0	235
1.5	280
2.0	255
2.5	190
3.0	120
3.5	55
4.0	20
4.5	12

(20)

3. The following table shows the distribution of bacterial counts in 176 samples of milk. A represents the actual counts (x), B the logarithm of counts. The frequency distribution of log X in B might be regarded as approximately normal. Making use

of the normal curve fitted to this distribution, calculate the expected frequencies corresponding to the groupings in A and apply a test of goodness of fit.

A x	*A* frequency	*B* log x	*B* frequency
0—	26	2.4—	1
1,000—	26	2.6—	9
2,000—	37	2.8—	16
3,000—	21	3.0—	19
4,000—	14	3.2—	28
5,000—	9	3.4—	36
6,000—	10	3.6—	26
7,000—	7	3.8—	23
8,000—	4	4.0—	8
9,000—	4	4.2—	3
10,000—	4	4.4—	5
11,000—	—	4.6—	—
12,000—	1	4.8—	2
13,000—	1		
14,000—	2		
above 15,000	10		
	176		176

(30)

4. A machine is designed to produce articles of average length 4.200 cm. It is supposed that owing to wear and tear it is no longer giving satisfactory results. The data below gives the lengths of a sample of 100 articles 'produced at the beginning of installation and the lengths of a sample of 100 articles produced recently.

Length in cm. (x)	Number of articles of length x	
	at installation	now
(1)	(2)	(3)
4.195	1	–
4.196	1	1
4.197	2	6
4.198	11	12
4.199	21	18
4.200	26	26
4.201	15	20
4.202	10	10
4.203	6	5
4.204	6	–
4.205	1	2
Total	100	100

Test whether the machine was working according to specification at the begining of installation and whether a change has taken place in the mean performance due to possible wear and tear even if the specification was not maintained in the begining. (20)

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PAPER VII (PRACTICAL)

Time : 6 Hours Full Marks : 100

N.B. (a) Figures in the margin indicate full marks.

(b) Use of calculating machines is permitted.

1. In a proposed stratified sample survey with 5 strata let N_i, n_i and σ_i be the population size, the sample size and the standard deviation respectively in the ith stratum and let the total population and the total sample size be N and n respectively. It is known that the sampling variance V and the cost function C are given by

$$V = \frac{1}{N^2} \sum_i \frac{N_i^2 \sigma_i^2}{n_i}$$

$$\text{and} \quad C = a + \sum_i b_i n_i$$

where $a = 550$ Rs.

The numerical values of the constants in the equations given above are as follows :—

Stratum No. i	N_i	σ_i	b_i in Rs.
1	37800	28	4
2	52600	18	3
3	8200	27	1
4	41600	21	3
5	28800	16	2

(a) Find the optimum values of n_i and the corresponding values of V for $C = 5000$ Rs.

(b) For the same value of the total sample size n as derived in (a), find the values of n_i when the allocation is proportional to $N_i \sigma_i$ only. Also find the corresponding values of V and C and compare them with those obtained in (a). (20)

2. An experiment was carried out to investigate the effect on wheat yield of applying a fixed amount of sulphate of ammonia at different times during growth. The experiment was designed as a 6×6 Latin square and the following results were obtained.

B 98	E 111	D 114	C 118	A 68	F 130
D 113	B 89	E 115	A 79	F 96	C 129
A 79	D 107	B 101	F 104	C 103	E 118
F 100	A 78	C 106	D 113	E 103	B 98
E 105	C 107	F 103	B 90	D 113	A 72
C 102	F 97	A 70	E 104	B 93	D 121

The figures in the table are yields in *lb* per $\frac{1}{100}$th acre plot, while the letters indicate the following treatments :—

A	No sulphate of ammonia applied
B	Sulphate of ammonia applied on November 5
C	-do- -do- -do- January 23
D	-do- -do- -do- March 13
E	-do- -do- -do- April 24
F	-do- -do- -do- May 25

(a) Give the analysis of variance as applicable to the above data to test treatment difference.

(b) Test the effect of the application of sulphate of ammonia by comparing the control (A) with the mean of the other treatments.

(c) Are there significant differences among B,C,D,E,F corresponding to different dates of application of the manure ?

(d) If C is true, is there any evidence of an optimum data of application ? Find that date from the response time curve by graphical methods. (35)

3. (a) Efficiency of operation of a cheese plant may be expressed by the yield of cheese which is made from 100 pounds of milk. Two observations taken each day from a production plant for a period covering nearly two months of April and May are as follows :—

Date		Cheese Yield sample/100 lbs. milk	
		I	II
April	1	10.48	11.17
	2	11.37	11.02
	5	10.81	11.81
	7	10.28	10.52
	9	10.66	10.60
	11	10.86	10.71
	13	10.89	10.58
	15	10.93	11.04
	17	10.83	10.74
	19	10.80	10.29
	21	11.11	11.75
	23	10.80	10.56
	25	11.14	10.41
	27	10.63	10.03
	29	11.60	11.12
May	1	10.09	10.62
	3	10.25	10.99
	5	10.56	10.96
	7	10.10	10.69
	9	10.63	10.50
	11	10.40	10.05
	13	10.43	10.36
	15	10.72	10.81
	17	11.03	11.51
	19	11.06	10.31
	21	10.44	10.42

(a) By setting up proper control charts examine whether the process can be regarded as under control.

(b) Give possible explanations for points going beyond the control limits. (25)

4. The matrix of inter-correlations of scores in 4 tests T_1, T_2, T_3 and T_4 is given below.

Correlation Matrix

	T_1	T_2	T_3	T_4
T_1	1.00	0.72	0.63	0.54
T_2		1.00	0.56	0.48
T_3			1.00	0.42
T_4				1.00

(a) Apply the centroid method to determine the first factor. Examine whether a single factor is sufficient to explain the intercorrelations.

(b) Also obtain the estimating equation for the first factor in terms of observed scores T_1, T_2, T_3, T_4. (20)

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PAPERS IV AND V : STATISTICAL QUALITY CONTROL (THEORETICAL)

Time : 4 Hours | Full Marks : 100

N.B. (a) Attempt ANY FOUR questions.

(b) All questions carry equal marks.

1. Under what conditions in Industry you would consider it economical to use (i) the *p*-chart and (ii) *c*-chart, respectively. Examine the basis and the approximations involved in the use of these charts.

As a routine, everyday, in a manufacturing plant all units produced are inspected for defectives. The number produced varies from day to day and is large. Show how you would plot a suitable control chart.

2. Lot quality is specified by proportion defective. An individual product is defective if quality $x > v$; otherwise is non-defective. A sampling scheme is specified as :

'Sampling size N,

$x + ks \leqq v$—Accept

$x + ks > v$—Reject'

where x and s are the mean and the standard deviation of the sample.

Stating the assumptions clearly, derive a formula to obtain points on the *OC*-curve for this plan. Use this to obtain the variables sampling plan when, $AQL = p_1$, and producer's risk $= \alpha$; and $LTPD = p_2$ and consumer's risk $= \beta$; and a product is defective if quality $x > v$ and otherwise non-defective.

3. Define 'Sequential probability Ratio Test', and explain how that concept can be used in Acceptance Sampling.

Develop a sequential test procedure for controlling the minimum level of quality in a process whose spread is steady and known. Obtain the five points of the *OC*-curve and the corresponding five points of the *ASN*-curve for this plan.

Devise a simple graphical procedure to instal this sequential plan for use by shop-personnel.

4. Differentiate clearly between 'Tolerance limits' and 'Confidence limits'.

From a controlled process a Random sample of N units give measures for a quality as $x_1, x_2, x_3, \ldots, x_N$. For large N, obtain a simple approximate formula for constructing tolerance limits L_1 and L_2 such that at least a given proportion Y of the entire production would be within L_1 and L_2 with a confidence coefficient of β.

5. (a) Discuss the situations when 'modified control limits' may be used.

Describe briefly a method of fitting a trend-line and 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 known to you.

6. Write short notes on :—

(i) Indifference quality
(ii) Chance causes and assignable causes
(iii) Simon's Ia charts
(iv) Reduced and Tightened inspection
(v) Theory of runs.

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PAPERS IV AND V : ECONOMIC STATISTICS (THEORETICAL)

Time : 4 Hours | Full Marks : 100

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

1. What is an index number of business activity ? How does it differ from an index number of industrial production ? Describe in detail, the procedure you will adopt to construct an index number of business activity for India.

2. Suppose that you are required to compare the change in real wages of textile workers with that of iron and steel industry workers of India in the post-independence years. Describe in short, the construction of the index numbers that you will use for this purpose, and the procedure you will adopt for the above comparison.

3. Why is it necessary to decompose a time series into a number of components ? Suppose that monthly cheque clearance in India during the period 1940–1955 and annual production of building materials in India during the period 1930–1955, are

given. Describe how you will decompose these two time series. How will you interpret the movements of the different components and use them for forecasting ?

4. What is an autoregressive time series ? How does an autoregressive scheme help in forecasting ? Taking a second order autoregressive series, show that its correlogram is a damped harmonic one.

5. Explain what is meant by price elasticity and income elasticity of demand. Give an account of the data needed and the methods to be used for estimating these elasticities. Describe some of their uses.

6. What is social accounting ? How does it help in the computation of national income ? What are its advantages over the traditional way of computing national income ?

7. Give a critical review of statistics of trade (domestic and foreign) available in India. Describe some of the uses of this type of data.

8. Suppose that you are asked to study the socio-economic conditins of rural and urban areas in your State. What type of information will you collect and how will you use the data for the above study ?

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PAPERS IV AND V : ACTUARIAL STATISTICS (THEORETICAL)

Time : 4 Hours | Full Marks : 100

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

1. (a) Explain the terms

(i) Force of interest δ

(ii) Effective rate of interest i

(iii) Normal rate of interest $i_{(m)}$

(b) Show by general reasoning that

(i) $i_{(p)}\, a^{(p)}_{\overline{n}|} = i_{(q)}\, a^{(q)}_{\overline{n}|} = \delta \bar{a}_{\overline{n}|}$

(ii) $2a^{(2)}_{\overline{n}|} + 2(1+i)^{\frac{1}{4}}\, a^{(2)}_{\overline{n}|} = 4a^{(4)}_{\overline{n}|}$.

(c) If $f(t)$ be the value of unity at the end of t years accumulated at the varying rate of interest δ_t find expressions for $\bar{s}_{\overline{n}|}$ and $\bar{a}_{\overline{n}|}$.

(d) You are given the equation $a_{\overline{n}|} = \beta$ where n and β are constants. It is required to find the effective rate of interest j that will satisfy the equation. You have decided to use method of successive approximation (2 trials).

Show that if σ be the value of j obtained by using the expansion of $a_{\overline{n}|}$ and ρ be the value obtained by the expansion of $\frac{1}{a_{\overline{n}|}}$ then $\rho > \sigma$.

2. (a) Describe the theory and assumptions underlying

(i) Census Method

(ii) Policy Year Method used in the construction of mortality rates.

(b) Obtain the central Exposed to risk formula by Policy Year Method.

(c) National life tables are constructed from census data. State

(i) the principal sources of error in census data

(ii) the other sources of error that might vitiate your result obtained for the death rates for the age group -0-1.

3. (a) Explain the theory underlying the summation method of graduation. Indicate what points you should bear in mind in choosing your operator and the operand.

(b) Obtain a suitable operand for the operator.

$$\frac{[4]\,[5]\,[6]}{4.5.6}$$

Criticise your formula.

4. Two mortality tables are such that $\mu_x^{II} = \mu_x^{I} + \lambda$ where λ is a constant.

(a) Criticise the following statements :—

(i) Constant additional mortality would need only a constant additional amount per annum to meet the extra cost. Hence

$$\bar{P}_x^{II} = \bar{P}_x^{I} + p$$

where $\bar{P}_x^{I}$ and $\bar{P}_x^{II}$ are the annual premiums payable continuously for whole life assurances at age x and p is a constant.

(ii) The constant addition to the force of mortality will have the effect of producing every year throughout the life of the assured an additional claim of λ per unit of insurance.

Therefore $\bar{A}_x^{II} = \bar{A}_x^{I} + \lambda\,\bar{a}_x^{I} = 1-(\delta-\lambda)a\,\bar{a}_x^{I}$

$$\text{Hence } \bar{P}_x^{II} = \frac{\bar{A}_x^{II}}{(1-\bar{A}_x^{II})/\delta} = \frac{\bar{A}_x^{I} + \lambda\,\bar{a}_x^{I}}{\frac{\delta-\lambda}{\delta}\,\bar{a}_x^{I}} = \frac{\delta}{\delta-\lambda}\,\bar{P}_x^{I} + \frac{\lambda\delta}{\delta-\lambda}$$

This proves conclusively that the constant addition to the premium under table I cannot be justified if the group of lives be subject to mortality table II.

(b) Show how the constant λ affects the premium P_x^{II} and obtain an expression for $\bar{P}_x^{II}$ in terms of $\bar{P}_x^{I}$.

5. (a) If two mortality tables I and II produce the same policy values at rates of interest i_1 and i_2, i.e., if

$$_t\bar{V}_x(\mu^I, \delta_1) = {_t\bar{V}_x}(\mu^{II}, \delta_2)$$

then show that

$$\mu_x^{II} \doteq \mu_x^I + \delta_1 - \delta_2 + \frac{r}{\bar{a}_x(\mu^I, \delta_1)}$$

where r is a constant and δ_1 and δ_2 are the force of interest corresponding to I_1 and I_2. Will the mortality tables I and II give the same values for

(i) Annual premium $\bar{P}_x$

(ii) Single premium $\bar{A}_x$

(iii) Annuity $\bar{a}_x$

when the tables are used in conjunction with the rates of interest I_1 and I_2.

(b) If Makeham's Law holds for μ_x then it is stated that the joint life annuity $a_{xy} = a_{ww}$

Obtain an expression for w in terms of x and y.

6. (a) Establish the following relation by general reasoning and press it as an integral

$$\bar{A}3_{xyz_1} = \bar{A}2_{xyz_1} - \delta \bar{a}^2_{yz)x}$$

(b) Write down the expressions for

$$\bar{A}3{:}4_{\overset{2\;1}{w:xyz}}, \quad A3_{\overset{2\;1}{wxyz_1}} \text{ and } a^{0(m)}_{\overline{yz}|x}$$

Explain the difference between the first two expressions i.e.,

$$A3:4_{\overset{2\;1}{w:xyz}} \text{ and } A3_{\overset{2\;1}{w\,xyz}}$$

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

Papers IV and V : Design of Experiments—Applied (Theoretical)

Time : 4 Hours | Full Marks : 100

N.B. (a) Attempt ANY FIVE questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is not permitted.

1. Describe briefly the various methods available for increasing the accuracy of agricultural field experiments.

2. What are the reasons for the common practice of repeating the same experiment at a number of different places and/or in a number of years? Discuss the difficulties which usually beset the statistician in making a combined analysis of these experiments and in interpreting the results.

3. (a) If V_1 and V_2 are the experimental error variances associated with main-plots and sub-plots respectively, both calculated on a sub-plot basis, in a split-plot experiment, explain the reason why it is expected that V_1 will be greater than V_2.

(b) In a split-plot design, there are p main plots for the variants $a_1, \ldots, a_p$ of a factor A in each of r randomised blocks and q sub-plots in each main plot for the variants $b_1, \ldots, b_q$ of a factor B. If V_1 and V_2 are the main-plot and sub-plot error variances, both estimated in units of a single sub-plot, prove that the variance of the difference between any two treatment means of the type $(a_i b_k - a_j b_k)$ is given by $\frac{2}{qr}\left\{V_1+(q-1)V_2\right\}$. What test of significance will you use for any of these treatment differences?

4. What are the parameters of a balanced incomplete block design? State and prove the necessary relations and inequalities connecting them.

Derive the expression for the efficiency factor of this design. What is the condition under which the efficiency of this design (using intra-block information only) will be greater than that of the corresponding complete block designs?

5. Explain the usefulness of 'confounding' and 'partial confounding' in factorial experiments and illustrate your answers with suitable examples.

Give a brief outline of the general structure of analysis of confounded designs.

6. Explain the principal of recovery of inter-block information for incomplete block designs. Illustrate the method in the case of:—

(1) balanced incomplete block designs

(2) Simple square lattice.

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PAPER IV AND V : SAMPLE SURVEYS—THEORY (THEORETICAL)

Time : 4 Hours | Full Marks : 100

(a) Attempt ANY FOUR questions.

(b) Treatment should be mathematical wherever possible.

(c) All questions carry equal marks.

1. How will you estimate the percentage literate in different age-groups from a random sample of individuals? State with reasons whether your estimates are strictly unbiased or not. Give the necessary theory for the estimation of the standard errors of the estimated percentages. Do you consider it likely that for the same sample size stratification by sex with proportionate allocation would have provided estimates with substantially lower sampling error? Give reasons.

2. Comment on the superiority or otherwise of cluster sampling over simple random sampling in terms of intra-cluster correlation. (Derive the necessary formula in this connection). State the considerations involved in the choice of the size of the sampling unit and explain carefully how the optimum size of the sampling unit may be determined by a pilot enquiry.

3. Discuss carefully various aspects of the problem of sampling on two successive occasions with particular reference to the question of determination of optimum degree of 'matching' the samples on the two occasions when it is desired to estimate the (1) means and/or (2) change of a given character. Explain to what extent moderate departure from the optimum degree of matching is likely to affect the precision of the estimates.

4. A population is divided into h strata with M_i first-stage units in the i-th stratum ($i = 1, 2, \ldots, h$). Each first-stage unit contans N second-stage units. A random sample of m first-stage units is selected from each stratum and then a random sample of n second-stage units is taken up for investigation in each of the selected first-stage units. How will you obtain an unbiased estimate of the population total of a character from this sample? Derive a formula for the estimation (from this sample) of the difference between the sampling variance of this estimated total and that of a linear unbiased estimate of the same population total which could have been obtained from an unstratified random sample of hm first-stage units with n second-stage units taken up for investigation within each.

5. Discuss carefully the relative merits of systematic sampling when compared with simple random sampling. What methods do you suggest for estimating the variance of the sample mean from a given systmematic sampling?

Find an expression for the variance of the sample mean in a two-stage plan in which m first stage units are selected wth replacement with probability proportional to the size N_i (where N_i = number of second stage units in the i-th first-stage unit) and then a systematic sample of the same size n is chosen from each of the selected first-stage units. (You may for the sake of simplicity assume that all the N_i's are multiples of n). How will you obtain an unbiased estimate of the sampling variance from your sample?

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

(a) Uses of inter-penetrating samples

(b) Sampling from a highly skewed population

(c) Stratification after sampling.

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

Papers IV and V: Sample Surveys—applied (Theoretical)

Time: 4 Hours　　　　Full Marks: 100

(a) Attempt ANY FIVE questions.

(b) All questions carry equal marks.

1. You are required to estimate the incidence of sugarcane borers (top, stem and root) from a given infested plot. Discuss how you would propose to take samples. The same cane in a clump may not be infested with all the three types of borers. (sugarcane setts are planted in rows.)

2. Explain the meaning, trace the importance, and state the inter-connections (if any) of (i) Pilot surveys, (ii) Cost function, (iii) Variance function and (iv) Optimum allocation, as understood in the context of large scale sample surveys.

3. It is decided to take a 0.5 per cent sample in a family budget enquiry to be conducted in a big city of your state. Discuss in this context briefly the problem of

(i) finding out sources for the frame,

(ii) selecting samples,

(iii) training the field personnel,

(iv) exercising a check over the field work, and

(v) scrutinising the field records.

4. Discuss how you would propose to organise the field work for price collection in connection with the execution of a scheme of constructing cost of living index numbers for the different strata of population of a big city. (The city has more than 50 markets of various sizes).

5. Discuss the problem of enumerating the number of beggars and vagrants in a big city. If you are asked to carry out a sample survey for an assessment of their economic conditions, how would you propose to take samples? Discuss briefly the difficulties associated with such a problem where the population under study is not stationary.

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

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

7. How does a defective sampling frame contribute to bias in estimation? Illustrate this with reference to an example.

8. Discuss how you would like to design a survey to estimate the yield per acre of a cash crop of your state in the absence of village maps. State in which of the stages of the procedure you suspect a bias in the estimation of acre-yeilds in such a situation.

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

Papers IV and V : Stochastic Processes (Theoretical)

Time : 4 Hours   Full Marks : 100

(a) Attempt ANY FIVE questions.

(b) All questions carry equal marks.

1. (a) Consider a random walk on the points $x = 0, 1, 2, \ldots, a$, with absorbing barriers at $x = 0$ and $x = a$. The probability of a step to the right is p and that of a step to the left is $q = 1-p$. Obtain a recurrence relation for $u_{z,n}$ = the probability that, starting at the point $x = z$ (with $0 < z < a$), the random walk ends with the nth step at the barrier $x = 0$.

Find a recurrence relation satisfied by $u_{z,n}$ and hence or otherwise obtain the generating function $P_z(t) = \Sigma\, u_{z,n} t^n$.

(b) In the above problem, if $w_{z,n}(x)$ be the probability that the nth step takes the particle to the point x, obtain the difference equation satisfied by $w_{z,n}(x)$ and state the boundary conditions.

2. (a) What is a Markov chain? What is meant by 'recurrent' "transient', 'periodic' and 'ergodic' states in a Markov chain? State a criterion for determining from transition probabilities the classes to which a state belongs.

(b) Two urns A and B contain N balls each. Of these $2N$ balls, N are white and N black. A 'trial' consists in drawing simultaneously one ball from each urn and putting it into the other urn.

Derive the conditional probability $p_{jk}^{(n)}$ of urn A containing exactly k black balls at the end of the nth trial, given that it contained exactly j black balls before the 1st trial. Show that as $n \to \infty$, $p_{jk}^{(n)}$ the probability of getting exactly k black balls if N balls are selected at random from a collection of N black and N white balls.

3. (a) In a certain physical process, the conditional probability of a change occurring in the time-interval $(t, t+h)$ given that n changes have occurred in $(0,t)$ is $h\{\lambda + \epsilon(h)\}$, where $\epsilon(h) \to 0$ as $h \to 0$, whatever n and t may be; and the probability of more than one change in $(t, t+h)$ is of smaller order of magnitude than h.

Show that the probability of exactly n changes occurring in $(0,t)$ is $e^{-\lambda t} \lambda^n/n!$

(b) Assuming that an electron travelling a distance h through the atmosphere has a probability $h\{\lambda + \epsilon(h)\}$ of being converted into two electrons, where $\epsilon(h) \to 0$ as $h \to 0$, show that the probability of there being exactly n free electrons at a distance x from the origin of the journey of the 1st electron is $e^{-\lambda t}(1-e^{-\lambda t})^{n-1}$, $n = 1, 2, 3, \ldots$

4. Consider a group of m automatic machines which are serviced by one repairman. As soon as a machine goes out of order, it is attended to by the repairman, unless he is already waiting on another machine, in which case the machine which has just broken down awaits its turn to be repaired. The machines are supposed to work independently of one another, and the time required to repair any one machine is supposed to be entirely independent of the states of the other machines. The working characteristics of each machine are determined by two parameters λ and μ, which are respectively the same for all machines and which have the following significance :—

If a machine is in working condition at time t, the probability that it will break down during the interval $(t, t+h)$ is $h\{\lambda+\epsilon(h)\}$ where $\epsilon(h) \to 0$ as $h \to 0$; on the other hand, if the machine is being repaired at time t, the probability that it will return to working condition in the interval $(t, t+h)$ is $h\{\mu+\epsilon(h)\}$.

Show that $P_n(t)$, the probability that at time t n machines are not working, satisfies $P'_n(t) = -\{(m-n)\lambda+\mu\}P_n(t)+(m-n+1)\ P_{n-1}(t)+\mu P_{n+1}(t)$, $1<n<m$.

Find the limit of $P_n(t)$ as $t \to \alpha$, and show that in the limit, the expected number of machines in the waiting line is $m-(\lambda+\mu)\ (1-p_0)/\lambda$, where p_0 is the probability of the repairman being idle.

5. (a) In connection with a continuous parameter process $X(t)$, what is meant by (1) continuity in probability, (2) continuity in mean square, (3) strict stationarity, (4) wide-sense stationarity ?

Briefly explain the relation, if any, between (1) and (2), and also between (3) and (4). Can you give examples to illustrate your statments ?

(b) In each of the following cases, discuss whether there can exist a mean-square-continuous and wide-sense-stationary process whose correlation function is $f(t)$, where

(i) $f(t) = e^{-|t|}$, $-\alpha<t<\alpha$,

(ii) $f(t) = \sum_{j=1}^{k} a_j e^{2\pi i t b_j}$, $-\alpha<t<\alpha$,

where $a_j>0$ for $j = 1, 2, \ldots, k$, and $b_1, b_2, \ldots, b_k$ are distinct real numbers. State the results you make use of.

6. Derive the Chapman-Kolmogorov equation satisfied by the conditional probability function of a continuous-parameter stochastic process, and also the back ward equation.

Discuss the significance of these equations in probability theory in general and Markov chain in particular.

7. (a) What is meant by a 'process with independent increments' ? What can you say about the probability distributions connected with such processes, and their relation to infinitely divisible laws ?

Discuss in detail Brownian motion and the stochastic process used to describe it and other similar phenomena.

(b) $X(t)$ is a stochastic process defined for all $t > 0$, and such that $X(o) = o$, with probability 1 $E\{X(t)\} = 0$ and $E\{X^2(t)\} < \infty$. For any t and $h > 0$, any positive integer n, and any $t_1, t_2, \ldots t_n$ in $[0, t]$, $X(t+h) - e^{ah}X(t)$ is stochastically independent of $X(t_1), \ldots X(t_n)$, where a is a certain non-zero constant.

Obtain the correlation function of the process. What can you say about the orthogonality or otherwise of the in crements of $X(t)$?

8. What is a stochastic difference equation ? Derive an expression for the correlation sequence of a sequence of random variables satisfying a stochastic difference equation with stochastically independent 'disturbances' on the right hand side.

Discuss the problem of least—squares estimation of the parameters ; with special reference to the problem of 'super-imposed' or observational errors.

9. Write a short essay on the application of the theory of stochastic process to problems in *one* of the following fields :—

(a) Public health statistics.

(b) Population growth

(c) Tele-communication engineering.

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

Papers VIII and IX : Statistical Quality Control (Practical)

Time : 4 Hours — Full Marks : 100

(a) Attempt ANY TWO questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. The number of surface defects observed in the final inspection of galvanised sheets of given area are given below :—

Sheet Number	No. of surface defects	Sheet Number	No. of surface defects
1	15	13	8
2	0	14	24
3	13	15	16
4	20	16	9
5	11	17	10
6	15	18	15
7	7	19	12
8	11	20	9
9	22	21	10
10	12	22	23
11	24	23	7
12	16	24	28
		25	8

(a) Analyse the data by the c-chart. And offer your suggestions for the future.

(b) How would you analyse the data if sheets numbered 4, 9, 11, 14, 22 and 24 are twice the area of the others ? Construct the appropriate control chart; and write down your comments.

2. A two-sided specification for a measurable quality x of a product is as follows : the product is non-defective if $25 \leq x \leq 35$; otherwise is defective.

Discuss a Two-sided plan of sampling Inspection by Variables. And draw the Acceptance Region based on Sample Mean $\bar{x}$ and sample S.D. s, when $AQL = 0.15$, $LTPD = 0.30$, producer's risk $(\alpha) = 0.01$ and consumer's rick $(\beta) = 0.02$. What is the sample size in this case ?

Show also the Acceptance Region if the Two-sided Test is replaced by the appropriate pair of one-sided tests. When can you use the pair of one-sided tests as a sufficient approximation to the Two-sided test considered above ?

3. The overall length of a certain product is specified to be 30 ± 1 units. 20 samples of 5 each are taken in the order of production. Their Averages and Ranges are given below :—

Sample No.	Average	Range	Sample No.	Average	Range
1	30.70	1.1	11	30.79	0.7
2	30.23	0.8	12	30.21	0.3
3	30.17	0.9	13	30.55	1.2
4	30.45	0.5	14	30.21	0.8
5	30.46	0.4	15	30.04	0.5
6	30.33	1.0	16	30.72	1.3
7	30.40	0.9	17	30.82	0.7
8	30.08	0.2	18	30.47	0.4
9	30.44	0.3	19	30.61	0.5
10	30.51	0.6	20	30.76	0.7

(a) Plot an $\bar{x}-R$-chart. Check for control.

(b) Estimate the per cent defective of the process when in control at the level indicated by the above data. Can the per cent defective be reduced by changing the process average ?

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PAPERS VIII AND IX : ECONOMIC STATISTICS (PRACTICAL)

Time : 4 Hours | Full Marks : 100

(a) Answer ALL questions.

(b) Figures in the margin indicate full marks.

(c) Use of calculating machines is permitted.

1. The amount of demand deposit (in billion dollars) in USA is given below for the period 1925–49 :

Year	Demand deposit	Year	Demand deposit
1925	20.6	1940	33.8
1926	20.1	1941	37.3
1927	21.4	1942	46.8
1928	20.7	1943	58.1
1929	20.8	1944	64.1
1930	19.4	1945	70.6
1931	16.3	1946	80.0
1932	15.1	1947	83.2
1933	14.4	1948	81.9
1934	17.5	1949	81.8
1935	21.0		
1936	24.2		
1937	22.8		
1938	25.1		
1939	28.8		

Analyse the series so as to study the important changes that have occurred in the period. Also estimate the demand deposits for the years 1950 and 1951. (45)

2. Per capita annual consumption of butter in kilograms, retail price in crown per kilogram of butter, per capita national income in crown and consumer price index of Sweden are given below for the period 1921–39 :

Year	Consumption of butter	Price of butter	National income	Consumer price index
1921	12.10	4.62	909	241
1922	12.63	3.29	732	195
1923	13.46	3.12	719	177
1924	14.12	3.03	743	174
1925	14.94	2.94	756	176
1926	15.34	2.61	772	172
1927	15.65	2.52	780	171
1928	17.04	2.46	798	171
1929	17.62	2.31	842	169
1930	18.04	2.13	860	164
1931	18.44	1.99	816	159
1932	18.85	1.88	753	156
1933	18.77	1.94	726	153
1934	19.11	2.16	782	154
1935	19.91	2.09	858	156
1936	20.38	2.06	907	158
1937	20.44	2.30	994	162
1938	20.20	2.42	1062	166
1939	20.44	2.53	1114	171

Use a suitable method to obtain the price elasticity and income elasticity of demand for butter. Also explain the significance of the estimates of elasticities. (55)

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

Papers VIII and IX : Design of Experiments—Applied (Practical)

Time : 4 Hours | Full Marks : 100

(a) Attempt ANY TWO questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. To study the effect of seed rate and nitrogen on different varieties of wheat, an experiment was tried in 3 confounded design with two replications in six blocks of 9 plots each. Denoting the varieties as V_1, V_2, V_3 the three seed rates as s_1, s_2, s_3 and the three levels of nitrogen as N_0, N_1, N_2 the plan of layout and yield per plot in lbs. are recorded in the table below Analysis the data statistically and summarise the results.

Rep. I			Rep. II		
1	2	3	4	5	6
$N_2V_3S_2$ 28	$N_0V_2S_1$ 27	$N_0V_1S_1$ 27	$N_1V_1S_2$ 16	$N_0V_1S_2$ 27	$N_0V_2S_2$ 34
$N_2V_2S_1$ 36	$N_0V_3S_2$ 37	$N_0V_2S_2$ 26	$N_0V_2S_3$ 30	$N_2V_1S_1$ 30	$N_1V_1S_1$ 35
$N_1V_2S_2$ 27	$N_1V_3S_1$ 17	$N_2V_2S_3$ 29	$N_1V_2S_1$ 38	$N_0V_2S_1$ 32	$N_0V_3S_1$ 41
$N_2V_1S_3$ 25	$N_1V_2S_3$ 29	$N_1V_3S_2$ 23	$N_0V_3S_2$ 30	$N_2V_2S_3$ 35	$N_1V_2S_3$ 26
$N_0V_1S_2$ 29	$N_2V_1S_1$ 30	$N_2V_3S_1$ 30	$N_1V_3S_3$ 26	$N_2V_3S_2$ 29	$N_0V_1S_3$ 33
$N_1V_3S_3$ 29	$N_1V_1S_2$ 21	$N_1V_2S_1$ 32	$N_0V_1S_1$ 27	$N_1V_1S_3$ 33	$N_2V_1S_1$ 33
$N_0V_3S_1$ 27	$N_0V_1S_3$ 24	$N_2V_1S_2$ 23	$N_2V_1S_3$ 20	$N_0V_3S_3$ 38	$N_1V_3S_2$ 29
$N_0V_2S_3$ 25	$N_2V_2S_2$ 30	$N_1V_1S_3$ 18	$N_2V_3S_1$ 33	$N_1V_2S_2$ 21	$N_2V_2S_1$ 33
$N_1V_1S_1$ 26	$N_2V_3S_3$ 31	$N_0V_3S_3$ 32	$N_2V_2S_2$ 27	$N_1V_3S_1$ 32	$N_2V_3S_3$ 36

Question 2.

Replication I									Block total
(1) 24	(2) 18	(3) 25	(4) 22	(5) 24	(6) 23	(7) 21	(8) 25	(9) 17	199
(10) 28	(11) 26	(12) 27	(13) 23	(14) 27	(15) 26	(16) 27	(17) 24	(18) 24	232
(19) 28	(20) 29	(21) 24	(22) 26	(23) 25	(24) 25	(25) 26	(26) 25	(27) 26	234
(28) 22	(29) 20	(30) 27	(31) 21	(32) 24	(33) 25	(34) 21	(35) 23	(36) 23	206
(27) 24	(38) 23	(39) 24	(40) 24	(41) 24	(42) 23	(43) 22	(44) 24	(45) 25	213
(46) 23	(47) 23	(48) 21	(49) 23	(50) 21	(51) 22	(52) 22	(53) 20	(54) 22	197
(55) 24	(56) 28	(57) 28	(58) 27	(59) 26	(60) 17	(61) 25	(62) 23	(63) 21	219
(64) 21	(65) 21	(66) 23	(67) 22	(68) 20	(69) 19	(70) 20	(71) 17	(72) 17	180
(73) 21	(74) 20	(75) 18	(76) 15	(77) 20	(78) 17	(79) 18	(80) 20	(81) 17	166
Replication II									
(1) 20	(10) 24	(19) 18	(28) 13	(37) 18	(46) 14	(55) 18	(64) 18	(73) 12	155
(2) 21	(11) 25	(20) 24	(29) 23	(38) 21	(47) 24	(56) 29	(65) 23	(74) 23	213
(3) 24	(12) 25	(21) 22	(30) 28	(39) 23	(48) 23	(57) 25	(66) 17	(75) 16	203
(4) 13	(13) 18	(22) 20	(31) 13	(40) 22	(49) 21	(58) 23	(67) 21	(76) 13	164
(5) 23	(14) 24	(23) 24	(32) 24	(41) 22	(50) 25	(59) 22	(68) 20	(77) 21	205
(6) 21	(15) 25	(24) 22	(33) 25	(42) 24	(51) 20	(60) 18	(69) 20	(78) 16	191
(7) 20	(16) 24	(25) 23	(34) 25	(43) 20	(52) 22	(61) 23	(70) 20	(79) 13	190
(8) 22	(17) 27	(26) 24	(35) 21	(44) 24	(53) 23	(62) 21	(71) 18	(80) 22	202
(9) 17	(18) 24	(27) 24	(36) 23	(45) 25	(54) 22	(63) 20	(72) 16	(81) 17	188

N.B. The figures in brackets are treatment (seed origin) numbers.

The data tabulated above give the mean height (in ft.) of trees in each plot of a forest genetics experiment involving 81 seed origins (treatments) of a particular tree species laid out in a simple lattice design with two replications in blocks of 9 plots.

The analysis of variance appropriate for these data has been partially done in the table given below. Complete the table and calculate the adjusted mean height for each treatment (seed origin) and also the standard errors for comparison of these adjusted means.

Source of variation	Degrees of freedom	Sum of squares	Mean square
Replications	1	112.50	
Blocks within replications (unadjusted)	16	783.32	
Treatments (adjusted)	80		
Intra-block error	64		
Total	161	1990.70	
Treatments (unadjusted)	80	1489.20	
Blocks within replications (adjusted)	16		

3. A balanced incomplete block design was laid out to compare the yields of 21 barley varieties. There were 21 blocks of 5 plots each, each variety being replicated 5 times. The layout plan and data are given below, the upper figure denoting the variety (numbered 1–21) and the lower figure being the yield of the plot, the block totals being also given against each block.

Carry out the statistical analysis of the data below, testing the significance of the differences among varieties. Obtain the adjusted variety means and calculate the S.E. of the difference between any two adjusted variety means. What is the gain in information from the recovery of inter-block information (excluding losses due to inaccuracy of weighting)?

Block No.						Total
	5	4	20	16	18	
(1)	796	1165	737	800	629	4327
	6	10	18	1	7	
(2)	824	686	781	700	579	3550
	9	13	21	18	11	
(3)	773	736	824	920	698	3951
	19	13	6	16	12	
(4)	677	892	866	615	627	3677

(5)	5 645	7 512	17 501	9 628	19 622	2908
(6)	4 746	17 641	6 867	21 731	15 481	3466
(7)	8 522	1 510	9 563	16 465	15 379	2439
(8)	4 743	12 758	14 499	10 453	9 473	2926
(9)	11 628	10 565	2 579	16 476	17 402	2650
(10)	10 488	15 473	5 322	13 242	3 534	2059
(11)	15 491	11 566	7 612	20 655	12 720	3044
(12)	14 622	11 543	6 641	5 805	8 944	3555
(13)	8 782	10 511	20 766	19 764	21 565	3388
(14)	13 593	14 607	20 624	17 518	1 549	2891
(15)	2 578	14 632	15 707	19 813	18 880	3610
(16)	8 868	12 658	17 652	3 1029	18 757	3964
(17)	1 453	3 686	4 922	11 670	19 775	3506
(18)	7 532	14 653	3 872	21 765	16 506	3418
(19)	1 458	5 337	21 642	12 945	2 748	3130
(20)	9 455	6 493	3 649	20 614	2 457	2668
(21)	2 472	8 501	4 488	13 480	7 580	1521

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

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

Time : 4 Hours | Full Marks : 100

Use of calculating machines is permitted

1. You are given the Census (1951) handbook of a district.

Making use of the village statistics for any suitable region (say a tehsil) make an assessment of the relative efficiencies of 3 or 4 different sampling schemes, with village as the sampling unit, for estimating the total cultivated area of the region.

[You can think of simple random sampling with sample unbiased estimate or ratio estimate or selection with probability proportional to size with (geographic) area or population as supplementary variate, etc. etc.].

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PAPERS VIII AND IX : SAMPLE SURVEYS—APPLIED (PRACTICAL)

Time : 4 Hours | Full Marks : 100

(a) Attempt ANY THREE questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.
(d) Mathematical or statistical tables will be supplied as demanded.

1. Draw up a complete schedule (questionnaire) for a family budget enquiry the main purpose of which is to determine 'weights' for the construction of cost of living index numbers. Explanatory notes may be furnished wherever thought necessary.

2. You are required to carry out a survey of refugees (displaced persons) in your state with a view to ascertaining the extent of rehabilitation needed and in this connection it is necessary to prepare the sampling frame from the first phase survey which has necessarily to be very rapid. Draw up a schedule for the purpose and prepare instructions for field staff who would be filling up the schedule.

3. Prepare a budget estimate of the cost for field work in connection with the economic survey of small industries in the rural areas of a district of your State. Advantage may be taken in this connection of the following hints :—

(i) The district may be taken to have three thousand villages in it
(ii) The size of the schedule and the distribution of the small industries may be taken to be such that a village would need on average 4 investigator days inclusive of the time spent in journey.
(iii) The budget estimate should have provision for supervisory staff.
(iv) There should be provision against printing of forms, stationery etc.

Each item of cost should be accompanied with a note of justification.

4. Out of one thousand bales of greasy wool from Argentina landed in Boston, corresponding weights of only 20 bales are available as shown in the table below:

(i) Find from the data given in the table the ratio estimate of the total weight of the thousand bales in Boston, given that the weight of the thousand bales in Argentina was 447,000 kgs.

(ii) Find also the coefficient of variation of 'f' from the formula,

$$(c.v.f.)^2 = \frac{M-m}{M} \cdot \frac{1}{m(m-1)} \sum_{1}^{m} \frac{(x - fy_j)^2}{x}$$

where $f = \sum_{1}^{m} x_j / \sum_{1}^{m} y_j$, $M = 1000$, and $n = 20$.

(iii) Select also a sample of 10 bales at random from the data and find $(c.v.f.)^2$ as in (ii).

TABLES OF WEIGHTS

Twenty Bales of Greasy Wool from Argentina landed in Boston

Bale Number	Y_j 'Marked' weight Buenos Aires (kgs)	X_j Weight as determined by the Bureau of Customs, Boston (lbs)
(1)	(2)	(3)
1366	447	980
1367	445	978
1368	446	975
1369	448	982
1370	449	986
1371	447	978
1372	447	981
1373	449	982
1374	445	978
1375	448	980
1376	449	988
1377	448	981
1378	449	983
1379	446	977
1380	447	979
1381	449	987
1382	447	979
1383	448	979
1384	447	979
1385	449	983

STATISTICIAN'S DIPLOMA EXAMINATION, MARCH 1956

PAPERS VIII AND IX : ACTUARIAL STATISTICAL (PRACTICAL)

Time : 4 Hours | Full Marks : 100

(a) Attempt ANY THREE questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. 'A' purchases an annuity of Rs. 600/- p.a. payable for ten years, the purchase price being Rs. 4551/-. Each annual payment is regarded, for tax purposes, to comprise of two elements (a) part repayment of principal and (b) interest on the principal outstanding immediately after the payment of the previous instalment. Tax is payable at the rate of 5 as. per rupee on the interest income alone.

(i) Ignoring tax, calculate the rate of interest 'A' is earning on the transaction.

(ii) Calculate the amount of tax 'A' will have to pay each year for the next 10 years, assuming that 'A' has no other income.

Given

$100i$	$(a^i_{\overline{10}\rceil})^{-1}$	$100i$	$(a^i_{\overline{10}\rceil})^{-1}$	$100i$	$(a^i_{\overline{10}\rceil})^{-1}$
3	.117231	$4\frac{1}{2}$	.126379	6	.135868
$3\frac{1}{2}$	.120241	5	.129505	$6\frac{1}{2}$	.139105
4	.123291	$5\frac{1}{2}$	.132668	7	.142377

2. A large industrial concern issues on the 1st January 1957, 1,00,000 debentures of Rs. 1000 each, on the following terms :—

(i) The debenture holder will receive interest at the rate of 5 p.c. per annum on the nominal value so long as the debenture remains unredeemed.

(ii) 5 per cent of the debentures remaining unredeemed at the beginning of any year will be redeemed at the end of the year at 102 but no redemption will take place during the first ten years.

All debentures which are not redeemed before 1st January 1976 will be redeemed at par on the 31st December 1976.

You are asked to calculate the issue price on the assumption that

(a) No tax is payable on capital gain but interest earnings will be subject to the standard rate of tax of 5 as. 4 ps. per rupee.

(b) the market rate of interest is $3\frac{1}{2}$ per cent net.

(c) the chance of redemption of a debenture at the end of a year is the same for all debentures in that year.

3. On the assumption that the mortality table follows Makeham's Law, obtain the graduated values of q_x for $x \times 40$ to $x = 50$ from the following data.

Age	Exposed to risk	Deaths	Age	Exposed to risk	Deaths
35	1,98,956	533	45	2,49,481	1,306
36	2,08,804	607	46	2,48,700	1,320
37	2,17,862	664	47	2,48,776	1,437
38	2,26,441	800	48	2,48,900	1,503
39	2,34,839	778	49	2,46,567	1,569
40	2,37,821	878	50	2,32,218	1,804
41	2,40,529	976	51	2,21,270	1,867
42	2,42,781	1,093	52	2,16,477	1,821
43	2,47,137	1,129	53	2,13,176	2,100
44	2,50,560	1,233	54	2,06,567	2,119

Details of the working should be shown.

4. Given the values of q_x for $x = 30$ to $x = 59$, calculate the values of $P_{30:\overline{25}|}$ and $P_{35:\overline{20}|}$ at 3½ per cent interest.

x	$q_x \times 10^5$	x	$q_x \times 10^5$	x	$q_x \times 10^5$
30	241	40	388	50	764
31	246	41	413	51	831
32	253	42	439	52	906
33	262	43	466	53	990
34	273	44	495	54	1084
35	286	45	527	55	1190
36	302	46	563	56	1311
37	320	47	604	57	1450
38	341	48	651	58	1608
39	364	49	704	59	1783

Working schedule should be shown.

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

Paper I : Theoretical Statistics (General)

Time : 4 Hours | Full Marks : 100

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

(b) All questions carry equal marks.

Group A

(Attempt any *three* questions)

1. (a) An unbiassed coin is tossed until k heads are obtained. Find the probability distribution of the number of tosses required.

(b) Prove that the sum of any number of independent Poisson variables is a Poisson variable.

2. (a) Write a short note on the problem of inverse interpolation explaining the different methods available, the assumptions involved and the situations in which they are useful.

(b) Derive Simpson's Rule for numerical integration. Use of formula to evaluate the integral

$$\int_0^1 x^n \, dx$$

For what values of n will Simpson's Rule give the exact value for the integral ?

3. In an army there are N soldiers. *It is known* that the average height of the soldiers is 65 inches and that the standard deviation of the distribution of heights is 2 inches. Briefly answer each of the following four questions :

(a) What is the coefficient of variation of the distribution of heights ? Can you make (using only the informations given) any statement about the magnitude of the fourth central moment of the distribution ?

(b) If you assume that the distribution of heights is approximately normal then what can you say about the proportion of soldiers who are more than 5 ft. 9 inches tall ? What is the fourth central moment then ?

(c) If you assume the distribution to be only symmetric (but do not make the drastic assumption of normality) then can you say anything about the proportion of soldiers who are more than 5 ft. 9 inches tall ? (Hint : use Chebyscheff's Inequality).

(d) If you change your unit of measurement from inches to ft. then what is the standard deviation of the average heights of 100 soliders chosen at random (with replacement) from the army ?

4. Carefully examine each of the following five statements (a) to (e) and say whether the statement is true or false adding in each case a few lines to substantiate your conclusion.

If the correlation coefficient between the random variables X and Y be negative then it means that

(a) the correlation coefficient between $-X$ and $-Y$ is positive.

(b) $E(XY) < E(X)E(Y)$

(c) the regression coefficient of X on Y and that of Y and X are both negative.

(d) the regression of Y on X is linear and the line is downward sloping.

(e) Y^2 and X^2 are dependent random variables.

GROUP B

(Attempt any *two* from this group)

5. (a) What is meant by standard error? Illustrate its use in tests of significance and in obtaining confidence intervals.

(b) Describe the method of testing an observed r from a small sample against a hypothetical ρ of the population and show how correlation coefficients obtained from a number of independent samples can be tested for homogeneity and if homogeneous can be combined into a single one as an estimate of the population correlation.

(c) How is the method mentioned in (b) extended in partial correlation coefficients?

6. (a) If u and v follow chi-square distributions which are independent with y_1 and y_2 degrees of freedom respectively, obtain the distribution of $\frac{u/y_1}{v/y_2}$

(b) Show how this distribution can be used to test the hypothesis of equality of variances from sample variances based upon two random samples from a normal population.

(c) Explain how the variance ratio distribution is useful in problems of analysis of variance. Give some examples.

(d) If in tests of analysis of variance, where the error variance is used in the denominator, the variance ratio turns out to be small, how do you test by using the existing (percentage points) tables of variance ratio whether the observed ratio is significantly small.

7. (a) A factory-manager, wishing to compare the performances, measured in certain units, of m different types of a machine in a production process, employs mn persons and assigns n persons at random to each machine for a specified time and collects the results. Obtain a systematic method of analysing the data of this experiment explaining clearly the assumptions and the theory involved.

(b) Why will this experiment be considered as poorly designed? How will you re-design it? Explain clearly the method of analysis of the data obtained from the re-designed experiment.

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPER II : THEORETICAL

Time : 4 Hours Full Marks : 10

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

(b) All questions carry equal marks.

GROUP A

(Answer any *three* questions)

1. What are the advantages of stratification in a sample survey ?

2. Examine critically the scope and method of construction of the official index of wholesale prices. Point out its limitations and the steps that might be taken towards its improvement.

3. What are the various components of a time series ? If you are given the annual rainfall data of an area over a long period, explain how you will analyse them for any cycles.

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

5. (a) Name the official publications that give information on the occupational pattern and the availability of technical personnel in India. Give a scheme of tables you will include for preparing a technical report on the topic.

(b) Write a note on the availability of statistics on employment.

GROUP B

(Answer any *three* questions)

6. (a) What roles are played by 'randomization' and 'replication' in the 'Design of Experiments' ?

(b) In a varietal trial with a large number of varieties, what types of design would you recommend in practice, and why ? For one such design, indicate the important steps of analysis.

7. Explain the concepts of reliability and validity of test scores. Mention a few available methods of estimating the reliability of a set of items.

What are standardised scores ? Why is it necessary to transform the scores this way ?

8. Give an account of the factors that influence the rate of growth of population. Give full account of any method which you would like to adopt to fit a growth curve and predict the future trends.

9. (i) In the wings of butterfly, black is dominant over white. What ratio of 'black' to 'white' do you expect to find in an F_2, and how do you propose to test the significance of the difference of the observed ratios from the expected ?

(ii) Describe how the O, A, B, AB blood groups are inherited. Write down the expected frequencies of these four phenotypes in terms of gene frequencies of O, A and B under the conditions of random mating.

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

Paper III : Theoretical

Time : 4 Hours | Full Marks : 100

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

(b) All questions carry equal marks.

Group A

(Answer any *two* questions)

1. What is Bayes' theorem ? What are the difficulties encountered in its practical use ?

Explain an iterative method by which maximum likelihood equations can be solved.

2. State some of the properties which a sample estimator should possess.

Illustrate each property by an appropriate example, giving in each case a sample estimator which does, and a sample estimator which does not possess the property.

3. (a) Write a critical note on the criterion of minimum variance in the theory of estimation.

(b) Given independent estimates $T_1, T_2, \ldots, T_k$ all unbiased for the same parameter θ but with different variances $c_1, \ldots, c_k$ independent of θ, obtain the linear compound of $T_1, \ldots, T_k$ which is unbiased for θ and has a minimum variance.

Group B

(Answer any *three* questions)

4. What are similar regions with reference to testing a composite hypothesis ? Give at least two illustrations of tests based on similar regions and explain how they are derived. Details of proofs are not needed.

5. What is meant by *bias* in a statistical test and why is it desirable to use an unbiased test wherever possible ?

Show that in the case of a simple hypothesis involving a parameter θ, the interior of critical region of a locally most powerful test is defined by

$$p''(\theta_0) > k_1 p'(\theta_0) + k_2 p(\theta_0),$$

where the null hypothesis specifies $\theta = \theta_0$ and k_1, k_2 are so chosen as to make the region unbiased and of a desired size.

Obtain the unbiased locally most powerful test of the hypothesis that the variance of a normal population is σ^2. Show that this test is also uniformly most powerful everywhere.

6. (a) From a sample from a k-variate normal population, explain how you will test the hypothesis that the means of the varieties are $\mu_1, \mu_2, \ldots, \mu_k$.

(b) What is Mahalanobis' D^2 statistic ? Show how it is useful in providing a test criterion for the hypothesis of equality of corresponding means in two p-variate normal populations having the same covariance matrix.

7. What is a likelihood ratio test ? What is its status amongst other tests in respect of power ?

Show how the likelihood ratio method is used to test the following hypotheses and indicate how the test procedure is carried out :—

(i) In a bi-variate normal distribution, the variances of x and y are equal

(ii) In a k-variate normal distribution, one of the variates x_1 is independent of the set $x_2, x_3, \ldots, x_k$.

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPER VI : PRACTICAL

Time : 6 Hours — Full Marks : 100

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

1. (a) In an interview for selection of a candidate for filling a post, examiners A and B gave marks to each of the candidates (full marks being 100). The marks given to a random sample of 10 candidates are as follows :—

Candidate No.		1	2	3	4	5	6	7	8	9	10
Marked by	A	27	35	31	47	23	35	36	31	37	37
	B	16	43	27	57	36	50	41	44	59	30

(i) Test whether the variances of scores given by A and B are significantly different from one another.

(ii) Test whether the average of scores given by A and B are significantly different from one another. (10)

(b) The following table gives the sample size, average and standard error of the average of stature for a number of castes included in a major anthropometric survey. Construct the analysis of variance table for testing differences in stature between castes. Test the variance ratio for significance and comment on the results.

	Sample size	Average	± standard error
Brahmin	200	165.3	± .7235
Kayastha	300	165.2	± .5234
Muslim	200	164.1	± .6893
Garo	350	162.8	± .4831

(15)

2. The sums of squares and products of three measurements x_1, x_2, x_3 taken on 60 circular cuts of 1' radius from jute growing plots are given below :—

$$n = 60;\quad \Sigma x_1 = 213; \qquad \Sigma x_2 = 480, \qquad \Sigma x_3 = 4890;$$

$$\Sigma x_1^2 = 1003, \qquad \Sigma x_2^2 = 5020, \qquad \Sigma x_3^2 = 4127287,$$

$$\Sigma x_1 x_2 = 2020, \qquad \Sigma x_1 x_3 = 18345, \qquad \Sigma x_2 x_3 = 3820.$$

(i) Determine the linear regression equation for predicting x_1 from x_2 and x_3. (15)

(ii) Test whether it is worthwhile to include x_3 in addition to x_2 in the linear regression equation of x_1 on x_2 and x_3. (8)

3. (a) The values of t for different values of the integral

$$P = 2\int_t^\infty p_{10}(t)\,dt$$ are given below. Find by a suitable interpolation method the value of P for $t = 1.0$ correct to three places of decimals.

P	t
0.1	1.812
0.2	1.372
0.3	1.093
0.4	0.879
0.5	0.700
0.6	0.542

(15)

(b) Evaluate $\int_1^3 e^{-x}x^3\,dx$ using the value of e^{-x} given below :—

x	e^{-x}
0	1.000000
0.5	0.606531
1.0	0.367879
1.5	0.223130
2.0	0.135335
2.5	0.082085
3.0	0.049787

(10)

4. In a radio listener's survey for music preference, the adult persons interviewed were classified thus :—

	Young (Below 30)		Middle-aged (30 to 45)		Old (above 45)	
	Male	Female	Male	Female	Male	Female
Prefer film music to classical music	29	20	16	17	10	8
Prefer classical music to film music	9	14	12	13	27	30
No preference	12	13	24	20	7	16

(a) Test for association between (i) music preference and age of listener, and (ii) music preference and sex of listener. (15)

(b) Write a report giving your recommendations about the nature of music to be broadcast. Do you need any supplementary information in order to make more definite recommendations ? (12)

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

Paper VII : Practical

Time : 6 Hours | Full Marks : 100

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

1. The plan and yields of a 2^4 experiment on beans are given below. The yields are given in pounds. The manurial factors employed are :—

Phosphate (p) : none, 0.6 Cwt. P_2O_5 per acre.

Potash (k) : none, 1.0 Cwt. K_2O per acre.

Nitrate (n) : none, 0.4 Cwt. N per acre.

Dung (d) : none, 10 tons per acre.

Block I	p	k	d	npk
	47	57	55	38
	dnk	dnp	dpk	n
	43	40	57	44

Block II	dp	nk	dk	pk
	52	46	45	53
	dnpk	(1)	dn	np
	46	60	43	52
Block III	npk	d	p	dnk
	45	44	41	36
	n	dnp	k	dpk
	49	54	52	46
Block IV	nk	dp	(1)	np
	45	54	50	41
	pk	dk	dnpk	dn
	58	54	56	44

Answer the following questions :—

(1) Identify the confounded interaction.

(2) Calculate main effects and unconfounded interactions and comment on them.

(3) Prepare a table of main effects and two-factor interactions and their standard errors.

(4) Is there reason to suppose that confounding has increased the precision of the expriment ? If you think so, estimate the gain in precision.

(25)

2. Given

q_{20}	q_{21}	q_{23}	q_{24}	q_{25}	q_{26}	q_{27}	q_{28}	q_{20}
.003	.004	.004	.004	.004	.004	.004	.004	.005

and radix l_{20} = 999, 999, prepare the life table given both l_x and d_x upto age 30. From this table answer the following questions :—

What is the probability that

(a) a person aged 22 will survive age 25 ?

(b) a person aged 23 will die between the ages of 27–29 ?

(c) three persons aged 21, 22, 23 respectively will survive four years ?

(25)

3. It is proposed to carry out a sample survey by interviewing sample households with a view to studying the consumer expenditure pattern of agricultural and non-agricultural households in the rural sector of a State. The following particulars are available :—

(i) Number of investigators available = 60

(ii) Number of households which an investigator can survey *per day* (including journey between villages and within villages) = 2 households *per day.*

(iii) Data collection is to be completed in a month (Monday the 1st October to 31st October), Sundays being the only non-working days.

(iv) The following particulars are available for the 80 tehsils of the State :—

Tehsil	Population in 1951 (000)	Agricultural population in 1951 (000)
(1)	(2)	(3)
1	160	52
2	141	77
3	112	87
4	90	66
5	100	80
6	106	85
7	136	113
8	131	88
9	93	64
10	168	129
11	99	68
12	132	93
13	94	76
14	85	68
15	105	82
16	75	57
17	70	58
18	28	18
19	127	101
20	79	53
21	90	74
22	91	65
23	186	127
24	131	105
25	70	67
26	126	93
27	96	59
28	92	75
29	126	108
30	94	73
31	135	116
32	95	79
33	108	86
34	74	55
35	86	77
36	49	21
37	96	36
38	80	27
39	102	75
40	85	58
41	73	60
42	75	54
43	98	74
44	172	122
45	170	125
46	135	97
47	125	88
48	65	41
49	101	91
50	108	94
51	132	108
52	117	100
53	70	53
54	138	85
55	125	83
56	49	40
57	70	54
58	146	68
59	63	36
60	92	69
61	87	69
62	97	68
63	132	102
64	34	26
65	141	110
66	70	52
67	53	25
68	86	70
67	70	65
70	120	104
71	96	49
72	119	93
73	82	51
74	41	20
75	20	10
76	64	47
77	65	52
78	72	41
79	38	27
80	60	39

(v) The average size of a household according to 1951 census is equal to 5 nearly.

(vi) List of villages along with their total population in 1951 in each tehsil will be available in the District headquarters of the State.

(a) What sampling design do you recommend and why?

(b) Draw the requisite number of tehsils according to the sampling design and write a brief set of instructions to the investigators (with specimens of necessary schedules) explaining how he can reach the sample households. (25)

4. The following figures are available for the breaking strength (in lbs) of cotton fabric strips each 3 yds. long and 2 inches wide. Each sample is taken from a separate roll of fabric and it may be assumed that the order of the sample and the order within the sample represents the order in which they were manufactured.

Sample No.	Breaking strength in lbs.	Sample No.	Breaking strength in lbs.
1	177	10	164
	176		181
	183		170
	180		158
2	182	11	163
	188		154
	164		166
	184		174
3	180	12	178
	168		178
	178		173
	175		158
4	176	13	156
	162		166
	178		164
	168		177
5	158	14	183
	175		178
	157		181
	174		170
6	165	15	142
	167		158
	160		171
	172		163
7	172	16	176
	152		177
	172		156
	154		167
8	163	17	167
	175		175
	161		171
	154		166
9	166	18	166
	174		167
	166		160
	173		178

Sample No.	Breaking strength in lbs.
19	157
	167
	161
	174
20	182
	172
	160
	165
21	176
	172

Sample No.	Breaking strength in lbs.
	186
	182
22	184
	172
	182
	178
23	182
	174
	174
	166

(a) Analyse the data to draw control charts for the mean and the range. State which, if any of the rolls from which the samples have been drawn, you would reject as being 'out of control'.

(b) Suppose that the consumer has set the specification that the breaking strength of the fabric shall *nowhere* exceed 190 lbs. or be less than 150 lbs. Modify your control limits for the sampling mean according to this requirement and see if this modification makes any difference to the rolls which you would accept or reject.

(c) It was known subsequently that the mean breaking strength of the rolls—11th onwards, had decreased by an amount equal to one standard deviation, standard deviation itself remaining unchanged. How is the risk of rejecting a good roll by the control chart you have already drawn in (a) changed ? (25)

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPERS IV AND V : STATISTICAL QUALITY CONTROL (THEORETICAL)

Time : 4 Hours — Full Marks : 100

(a) Attempt ANY FOUR questions.

(b) All questions carry equal marks.

1. (a) Explain, 'Why the Control Chart Works' to a shop foreman in non-mathematical terms, indicate the experiments and demonstrations that you may perform to make your points clear and convincing.

(b) Also, draw up a suitable form of directions of the foreman for installing the use of $\overline{X}-R$ charts.

2. Define the concept of 'Average Amount of Inspection' with reference to Double Sampling and Multiple Sampling (by attributes). A Multiple Sampling Scheme for fraction defective is operated as follows :—

(i) Each sample is of size n.

(ii) The inspection terminates with a decision of acceptance or rejection, at the end of or before the sth sample.

(iii) The entire first sample is inspected always.

(iv) If a decision 'to accept' is reached during a sample, the inspection is continued to the end of that sample.

(v) If a decision 'to reject' is reached during a later (than the first) sample, the inspection is curtailed immediately.

Obtain a formula for the Average Amount of Inspection in this case.

3. In acceptance inspection of lot per cent defective, by Single Sampling Plans by variables with one-sided criterion, discuss :—

(i) how the estimate of the process average from the combined records of Inspection can be used in comparison with AQL to offer criteria for reducing or tightening or continuing normal inspection.

(ii) the savings in the Amount of Inspection (N) yielded by a knowledge of the standard deviation of the quality (x) distribution. (Derive the necessary expressions, stating the assumptions).

4. Write a mathematical note on : 'The use of Sequential Analysis (by attributes and variables) in Industrial Experimentation and Inspection.'

5. Write short notes on :—(i) Risks of producers and buyers. (ii) Group control charts, (iii) The OC–Curve, (iv) The Shainin Lot Plot Method, (v) Cost aspects of Quality Decisions.

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPERS IV AND V : SAMPLE SURVEYS—THEORY (THEORETICAL)

Time : 4 Hours Full Marks : 100

(a) Attempt ANY FOUR questions.

(b) All questions carry equal marks

(c) Treatment should be mathematical wherever possible.

1. Describe the various uses of 'double sampling'. Why is it also called 'two-phase' sampling? Give the details of this method (supplying proof wherever necessary) as it is used in problems of non-response.

Further work out the formula for the standard error of the estimate when the auxiliary variate (x) is used to give a regression estimate of the mean of the variate under enquiry (y) is double sampling.

2. What are the merits of a multi-stage sample as compared against uni-stage sample?

Discuss in its various aspects the case of p-stage sampling when the sample units are of equal size and the sample-sizes in the various stages are equal for the same stage. (Actual formulae are to be worked out for $p = 4$, while those for the general case are to be indicated without proof, the 'finite population correlations' may be ignored).

3. Work out the estimate of the population mean, its standard error and the estimate of the standard error for the following two cases of two-stage sampling :—(i) when the sample-sizes in the second stage are equal and the 'finite population corrections' are to be used, and (ii) when the sample-sizes in the second-stage (within each first-stage, unit) are unequal and the 'finite population corrections' may be ignored. (The units may be assumed to be of equal size in both the cases).

4. Discuss in detail (supplying the necessary mathematical deductions) a suitable sampling plan for estimating the acreage under a crop in a state of India. Illustrate your answer with a concrete case. Further discuss how to collect the preliminary information for planning such a sample survey in an optimum manner.

5. (a) Write a note on the general principles of design of a sample survey using suitable illustrations.

(b) Write a note on the frames, pointing out specially the difficulties in their construction.

6. (a) Write a general note (witthout detailed proofs) on the various ways in which the supplementary information regarding the values of an auxiliary variate (x) may be utilised in estimating the average (or total) of the variate under enquiry (y).

(b) The population mean is to be estimated by the mean of a simple random sample, whose size, n, is fairly large, but the sampling fraction is negligibly small. Assume that the loss $l(z)$, incurred by an error of amount z in the estimate is given by $l(z) = \lambda z^2$ and that the cost function for the sample is $C_0 + C_1 n$, where λ, C_0, C_1 are constants. Find out the value of n which will minimise cost plus (expected) loss. (Assume the knowledge of any other constant which will be necessary).

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPERS IV AND V: SAMPLE SURVEYS—APPLIED (THEORETICAL)

Time: 4 Hours | Full Marks: 100

(a) Attempt ANY FIVE questions.

(b) All questions carry equal marks.

1. What are the various types of defect to which a sampling frame may be subject? Explain how each such defect occurs.

2. Write short notes on:—(i) Investigator bias, (ii) Cost function, (iii) Non-response, (iv) Double sampling, and (v) Master sample.

3. In respect of an area damaged by cyclone, you are required to submit to Government within about 10 days of the cyclone an estimate of the damage to house property. The area involved may be taken to consist of 20 police stations for each of which will be available only 20 primary field workers. It is further known that on average each police station has 50 villages, and each village 125 families. How do you propose to take samples?

4. What are exploratory surveys? State, with reference to an actual survey how the experiment gained in such an exploratory survey was tutilised in subsequent planning.

5. What are sampling and non-sampling errors? How do you propose to estimate the former and control the latter?

6. What details of information you think it should be proper to incorporate in a report on sample surveys. A broad outline of the main features is only needed.

7. Indicate the important steps that you would take in training the field staff who would be entrusted with the responsibility of conducting a family budget enquiry.

8. How do you propose to control the accuracy of field work of a large scale sample survey, and what should be your arrangements for receiving the data from the field?

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPERS IV AND V: DESIGN OF EXPERIMENTS—APPLIED (THEORETICAL)

Time: 4 Hours | Full Marks: 100

(a) Attempt ANY FIVE questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is not permitted.

1. Explain the characteristics of a good experimental design. What is meant by the *efficiency* of a design?

2. Discuss the covariance method of estimating missing values in a designed experiment. Apply that method for analysis of data of a randomized block experiment involving one missing plot. Derive the expression for the standard error of the difference in effects between the treatment involving the missing value and any other treatment.

3. Write notes on any *two* of the following :—

(a) Fractional replication.

(b) Intra-and inter-group balanced designs

(c) Interaction as error.

4. (a) In what situations will you recommend (i) a split-plot design, (ii) strip plot design for a factorial experiment. Discuss the advantages and disadvantages of both the designs.

(b) A split-plot design is arranged in r randomised blocks with p main plots and q sub-plots in each main plot. The main plots are allotted to p variants of a factor A and the sub-plots to q variants of a factor B. Give the structure of the analysis of variance table and show how you will test the significance of the two main effects and the interaction of the factors. Also, obtain the standard errors for comparisons between pairs of treatment combinations.

5. (a) Write a critical note explaining the nature and utility of the concept of inter-block information.

(b) Explain the lay-out of a simple square lattice and comment on the scope of its utility. Also obtain the estimates of treatment contrasts and their standard errors taking into account both the intra-block and inter-block information.

6. (a) Dscuss the advantages of a factorial experiment. Explain the principles of total and partial confounding in factorial experiments using as illustration a 2^5 factorial experiment conducted in 4 blocks of 8 plots each per replication

(b) It is proposed to study the effects of two dates of sowing, two levels of nitrogen and three spacings on two varieties of cotton.

How will you plan this experiment ? Give the structure of the analysis of variance for the design you finally decide upon. Draw the lay-out plan.

7. (a) Explain the combinatorial parameters : $v, l, r, b, \lambda_i, n_i$ and p^i_{jk} of partially balanced incomplete block designs and write down the relationships they satisfy. Discuss the situations in which a paratially balances design would be useful.

(b) Show that the following arrangement of 12 treatments in 9 blocks of 4 plots each with 3 replications of each treatment is a p.b.i.b. design.

Blocks	I	II	III	IV	V	VI	VII	VIII	IX
	1	7	6	1	11	10	1	9	5
	2	10	11	7	5	9	11	2	3
	3	5	9	6	2	3	10	7	6
	4	4	4	8	8	8	12	12	12

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1958

PAPERS IV AND V: MATHEMATICAL THEORY OF SAMPLING DISTRIBUTION (THEORETICAL)

Time : 4 Hours Full Marks : 100

(a) Attempt ANY FOUR questions.

(b) All questions carry equal marks.

1. (a) Define a probability generating function. How will you obtain the moment generating function from the P.G.F.?

If $\phi(n)$ is the P.G.F. for the binomial distribution of index n, obtain the relation between $\phi(n)$ and $\phi(n+1)$ by using the relation between $p(n,r)$, $p(n+1, r)$ etc., where $p(n, r)$ stands for the probability of r successes in n trials.

(b) If the rate of change in the number of particles in a given space at any instant is proportional to λ find the P.G.F. for the number of particles at time t.

(c) Using characteristic functions, obtain the distribution of the mean of samples from a Cauchy distribution.

2. (a) If x_1, x_2 and x_3 are three independent consecutive random variables from a rectangular population (0—1), obtain the probability that $x_1 > x_2 < x_3$.

(b) Defining a trough by three consecutive observations where $x_1 > x_2 < x_3$, find the first three moments of the distribution of the number of troughts in n consecutive observations from a rectangular population (0.—1).

3. (a) Derive the distribution of the median of a sample of n independent observations from a normal population.

(b) Obtain the distribution of the ratio of variances of two samples from a normal population.

4. (a) Obtain the sampling cumulant $k(2^21)$ for Fisher's k-statistic.

(b) If $M(x, y, z)$ is the moment generating function for the joint distribution of three variables establish the relationship between k_{222} and the corrected moments.

5. Explain the relationship between Hotellings T^2 and Professor Mahalanobis's D^2-statistic.

Obtain the distribution of Hotelling's T^2.

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

Papers IV and V : Actuarial Statistics (Theoretical)

(a) All questions carry equal marks.
(b) Use of calculating machines is not permitted.
(c) Attempt any ONE question from 'Group A' and TWO from each of the 'Groups B and C'. Altogether FIVE questions are to be answered.

Group A

1. (a) What is (i) effective rate of interest, (ii) force of interest ?

Express mathematically the relation between the two.

(b) The Revenue Account of an Insurance Company shows that the fund increased from A at the beginning of the year to B at the end of the year, the net interest earnings were I. Find the expression for the effective rate of interest earned during the period.

Obtain approximate expressions for δ and i.

2. A foreign Government has issued a loan C redeemable by a cumulative sinking fund. The loan bears interest at the rate of g per annum payable p times and is redeemable at par by annual drawings spread over n years. Show that the market value of the loan is

$$C\,\frac{a_{\overline{n}|}^{i}}{a_{\overline{n}|}^{g}} + Cg\,\frac{s_{\overline{1}|}^{(p)}-1}{i-g}\left(1-\frac{a_{\overline{n}|}^{i}}{a_{\overline{n}|}^{g}}\right)$$

where i is the effective market rate of interest.

Interpret the meaning of the first and second terms of the above expressions.

Group B

3. The mortality experience of a group of people belonging to pension fund is to be examined over a period from 1st March 1950 to 31st August 1954, each life being traced from 1st March 1950 or subsequent entry until 31st August 1954 or earlier death or withdrawal. A card is available for each life showing

(a) date of entry and nearest age at that time
(b) date of death or withdrawal
(c) cause of exit

Explain how you would proceed and give the exposed to risk formula you would adopt. How would you classify the following cases :—

(i) at entry into experience
(ii) at exit

Date of entry	Nearest age at entry	Cause of exit	Date of exit
13th May 1937	32	—	—
10th August 1942	26	Death	13th February 1953
5th March 1950	35	Withdrawal	5th December 1950
8th October 1949	42	—	—
14th April 1946	22	Death	18th November 1954

4. (a) What is the purpose of Graduation ?

(b) What are the merits and demerits of the following methods of graduation?

(i) Graphical method
(ii) Formula method
(iii) Summation method

(c) What are the tests you would apply to find out whether your graduated figures are reasonably satisfactory ?

(d) What advantages do we gain by graduating mortality rates on the assumption that they follow Makeham's Law ?

5. (a) In a discussion on mortality investigation it was remarked that the comparison of the values of the expectation of life at birth e_0 would give sufficiently accurate picture of the relative mortality rates of two countries. Comment on the above observation.

(b) If you are asked to compare the mortality rates of two countries what methods would you apply and why ?

(c) What do you understand by mortality index figures ? Explain the different indices that are used for the comparison of natural and sectional mortality rates.

Group C

6. (a) A life annuitant wants to convert his life annuity into an annuity certain for a fixed term of years. He argues that the insurance company should give him an annuity certain for a term equal to his expectation of life since on the average he is expected to live till the end of the term. If the company agrees to the proposition the company will neither gain nor lose in the transaction.

As an executive of the company, do you think you can agree to his proposition ? If so, why ? If not, why not ?

(b) Prove that $(\bar{I}\bar{A})_x = \bar{a}_x - \delta(\overline{Ia})_x$

7. Deduce by general reasoning the benefits represented by the following expression, and verify your conclusion by reducing the expression to its simplest form in terms of joint life annuity functions :—

$$\int_0^\infty a_{\overline{t|}}\; {}_tp_{abc}(\mu_{a+t}+\mu_{b+}\mu_{b+t}+\mu_{c+t})(1-{}_tp_{xyz})\,dt +$$

$$\int_0^\infty (a_{\overline{t|}} + v^t\bar{a}_{x+t:y+t:z+t})\;{}_tp_{xyz}\;{}_tp_{abc}(\mu_{x+t}+\mu_{y+t}+\mu_{z+t})\,dt.$$

8. Obtain an expression for $\bar{A}\,{}^{3}_{xyz}$ and show that the above contingent assurance can be expressed in terms of contingent assurance depending on first deaths only.

To what extent your computation work of the above function will be simplified if the mortality rates follow

(i) Makeham's Law
(ii) Gompertz Law ?

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPERS IV AND V : ECONOMIC STATISTICS (THEORETICAL)

Time : 4 Hours | Full Marks : 100

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

1. If you are required to study the changes in the volume of exports and imports and the prices of exported and imported articles, over the last decade in India, what type of index numbers will you use ? Give an outline of the methods of construction of such index numbers.

2. Define serial correlation and show how you will use the observed serial correlations of a time series to find out the scheme generating the series.

3. Explain what is meant by the structural relations among a set of economic variables. How do they differ from regression relations obtained on the basis of time series data ? Taking demand and supply relations, show why it is not possible to obtain unbiassed estimates of the parameters of these relations, on the basis of time series data, by applying least square method to each relation separately.

4. Describe the method that you will adopt for obtaining demand relation for food grains in India on the basis of time series data.

5. Explain clearly the concepts of gross national product, national income at market price, national income at factor price, personal income and disposable income. How will you calculate the national income generated in the manufacturing industries of India in any recent year ?

6. It is expected that there will be more urbanisation in the second five year plan period because of industrial development. If you are asked to prepare estimates of number of houses by types that will be needed in urban areas, what type of data will you collect and how will you prepare your estimates ?

7. If you are asked to study the causes of the present rises in food prices in India, how will you prepare the scheme of your study ? Indicate the nature and source of data that you may require for the purpose.

8. If you are required to prepare time series—showing annual production of jute, cotton, wheat, rice, cotton textile and sugar and factory employment from 1930 onwards relating to the present boundary of India, what procedure will you adopt ?

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPERS IV AND V : THEORY OF INFERENCE (THEORETICAL)

Time : 4 Hours Full Marks : 100

(a) Attempt ANY FOUR questions.

(b) All questions carry equal marks.

1. Give a general method of obtaining Bayes' estimates. Show that when the average risk associated with a Bayes' estimate is constant over the parameter space, it is a minimax estimate.

Hence or otherwise obtain the minimax estimate of 'p' of a binomial distribution when loss function is given by the mean square difference.

2. Show that for a fixed sample size, there does not exist a test for student's hypotheses whose power is independent of the variance.

State and prove Stein's two sample result on fixed length confidence intervals for the mean of a normal population with unknown variance.

3. State and prove Wald's sequential procedure for testing the mean of a normal population with unknown variance.

4. Sketch the proof of the optimum property of sequential probability ratio test.

Two gamblers A and B start playing with rupees 'a' and 'b'. Winner at each game wins Re. 1/- from the loser. The probability of A winning any game is p. Find the probability that A will be ruined and the average number of games before the play comes to a close which will be when either players loses all this money.

5. Give a general procedure for testing a composite hypotheses against a sample alternative. Hence or otherwise find the most powerful test for testing

$$H_0: \xi < -\xi_0, \xi > \xi_0 \quad \text{against} \quad H_1: \xi = 0$$

where the observations are normally distributed with unknown mean ξ and variance unity. Show that this test is consistent.

6. How will you proceed to test a newly prepared table of random numbers for randomness ?

7. An observation is known to come from one of the two known normal multivariate distributions. How will you classify the observation such that error of mis-classification are the minimum. How will you modify this procedure in case the parameters of the normal distributions are unknown but having the same dispersion matrix and can be inferred from samples. Generalise this to k populations.

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1958

PAPERS VIII AND IX : STATISTICAL QUALITY CONTROL (PRACTICAL)

Time : 4 Hours | Full Marks : 100

(a) Attempt ANY TWO questions.

(b) Use of calculating machines is permitted.

1. Construct a p-chart using the following data for 25 days. Assuming that assignable causes have been found for all points outside the control limits, project a suitable p-chart for use for further production.

Number of defectives found (per sample of 200) in the order of production	6,	6,	6,	5,	0,	0,	6,
	1,	4,	0,	1,	3,	2,	4,
	15,	7,	1,	3,	1,	4,	0,
	4,	4,	4,	15.			

Also draw the OC-curve for this projected p-chart, calculating at least 11 points on the curve. (Give reasons for any approximations you may use in your calculations)

2. A manufacturer of complete rounds of ammunition submitted his products for acceptance inspections in lots of 1000 each. Muzzle velocities were determined for random samples of *five* from 25 consecutive lots. The $\bar{X}$, R for these samples were (in the order of production):—

$\bar{X}$: Average muzzle velocity ft./sec.	1710,	1711,	1713,	1718,	1735,	1739,	1723,
	1741,	1738,	1725,	1731,	1721,	1719,	1735,
	1741,	1783,	1777,	1794,	1773,	1780,	1798,
	1789,	1788,	1799,	1807.			

R : 42, 40, 39, 26, 10, 25, 14, 15, 11, 31, 25, 19, 43, 30,
17, 51, 9, 15, 37, 54, 15, 29, 39, 30, 44.

Lots were rejected, if sample range (R) exceeded 70, or if sample mean ($\bar{X}$) was less than 1715 ft./sec. or more than 1785 ft./sec. Otherwise lots were accepted.

(i) Make out the rejected lots by the sampling scheme.

(ii) Plot the $\bar{X}$, R-charts and analyse, identify if there were more than one Universe (i.e., grand lots).

(iii) Give estimates of the Means and Standard deviations of the grand lots so identified.

(iv) Obtain the probabilities of acceptance, by the above sampling scheme of lots of 1000 each from those grand lots.

(v) Write a critical review of your findings and comment on the sampling scheme used.

3. A standard gun is to be compared with an experimental gun on the basis of hits on a designated target under specified conditions. An odds ratio of 3 or more is set on the criterion of superiority for the experimental gun, on odds ratio of 1.2 or less as the criterion of superiority for the standard gun. That is, if the ratio of hits to misses for the experimental gun is three or more times as large as for the standard gun, it is important to decide in favour of the experimental gun, but if the ratio is

1.2 or less times as large for the experimental as for the standard gun, it is important to decide in favour of the standard. α (the risk of deciding the experimental is superior when in fact the standard is superior) is to be 0.02 and β (the risk of deciding the standard is superior when in fact the experimental is superior) is to be 0.05.

(a) Develop a sequential testing programme and set up

(i) a graphic procedure and

(ii) a tabular procedure (draw the graph and present the table).

(b) Obtain the approximate Average Sample Numbers.

(c) In an actual experiment, the following results are obtained :—

N— the cumulative number of pairs fired

n— the cumulative number of pairs favourable to one gun or the other

E— the cumulative number of pairs favourable to the experimental gun.

N :	27	40	50	51	52	53	77	84	92	99	101	103	107	114	125	126
n :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
E :	1	2	2	3	3	4	5	5	5	6	6	6	7	7	7	7

What would be your conclusion on the basis of the above results ? At what stage would you reach that conclusion ?

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPERS VIII AND IX : THEORIES OF INFERENCE (PRACTICAL)

Time : 4 Hours   Full Marks : 100

(a) Attempt ANY TWO questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. Examine the following table of numbers for randomness (Asyptotic results may be used).

6	6	4	4	2	1	6	6	4	6	5	8	6	5	6	2	6	8	1	5
2	2	1	5	8	6	2	6	6	3	7	5	4	1	9	9	5	8	4	2
4	4	1	4	5	1	2	3	2	2	3	8	8	8	5	7	9	5	6	7
9	1	6	1	1	9	6	1	2	6	7	2	9	3	4	8	9	8	5	7
7	3	9	0	8	4	4	3	8	0	9	4	3	6	4	5	5	6	6	9
0	0	6	1	5	6	7	3	1	2	2	3	9	8	5	5	8	5	1	1
1	8	7	8	2	3	0	8	9	3	3	5	2	8	8	6	9	0	2	9
5	0	4	7	4	8	8	9	6	4	5	8	8	9	7	5	8	3	8	5
8	3	8	7	6	6	7	0	2	4	3	1	6	6	5	6	2	1	4	8
6	8	0	7	6	5	9	3	7	3	5	2	1	6	5	6	3	6	5	3
2	0	1	3	3	9	3	5	9	1	2	5	7	1	3	4	6	2	3	3
3	2	6	8	9	2	3	3	9	8	7	5	6	6	9	9	4	2	1	4
7	5	2	9	1	3	8	6	8	3	5	4	7	7	2	7	0	6	9	4
6	7	0	5	1	3	2	2	5	2	4	4	0	5	9	4	6	4	8	5
1	3	8	3	2	7	0	2	7	9	6	4	6	4	7	2	2	8	5	4
7	8	4	5	0	5	1	3	1	5	9	1	4	1	8	4	9	3	7	7
4	3	4	3	3	5	8	2	8	8	3	3	6	9	9	6	7	2	3	6
7	1	6	2	4	6	6	7	8	7	8	1	3	0	3	7	3	4	3	9
3	1	8	8	7	1	4	4	9	1	1	4	8	8	4	7	8	9	2	3
1	3	6	1	7	8	7	1	3	2	7	6	9	5	6	2	8	7	4	5
1	3	8	3	2	7	9	2	7	9	6	4	6	4	7	2	2	8	5	4
7	8	4	5	6	5	1	3	1	5	9	1	4	1	8	4	9	3	7	7
4	3	4	3	3	5	8	2	8	8	3	3	6	0	9	6	7	2	3	6
7	1	6	2	4	6	6	7	8	7	8	1	3	0	3	7	3	4	3	9
3	1	8	8	7	1	4	4	9	1	1	4	8	8	4	7	8	9	2	3
1	3	6	1	7	8	7	1	3	2	7	6	9	5	6	2	8	7	4	5

2. The number of cysts per fish in a commercial catch of white fish is known to follow the distribution whose probabiity generating function is given by $q^{-k}\left(1-\frac{xp}{q}\right)^{-k}$ where p, k are parameters and $q = 1+p$. It is known that k is constant from one batch to another but the mean number of cysts per fish varys from batch to batch. Set up a sequential procedure to test (H_0) that the mean is .2 against the hypotheses (H_1) that the mean is .6. You can also assume that the variance under H_1 is 2. The Type I and type II errors may be fixed each at 5 per cent.

Find the OC as well as the ASN curve and comment on the plan.

3. Find the size of the sample required to ensure that 99 per cent of the population lies within the sample range with probability 95 per cent

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

Papers VIII and IX : Design of Experiments—Applied (Practical)

Time : 4 Hours | Full Marks : 100

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

1. An experiment was carried out in a 8×8 simple lattice design to compare the yields of 64 strains of paddy. The lay-out plan and the yield of grain in ounces per plot are given below.

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

Replication I

Block 1	(11) 5.0	(12) 11.3	(15) 12.3	(13) 8.3	(9) 11.8	(16) 7.8	(10) 10.0	(14) 7.0
Block 2	(50) 17.3	(56) 5.5	(55) 9.5	(49) 6.8	(51) 5.0	(52) 7.5	(54) 8.3	(53) 10.8
Block 3	(26) 10.8	(31) 12.0	(27) 6.8	(25) 7.0	(30) 6.8	(32) 9.5	(25) 10.3	(28) 11.5
Block 4	(20) 10.3	(22) 10.3	(23) 7.0	(19) 7.8	(18) 9.3	(21) 8.8	(24) 7.8	(17) 7.3
Block 5	(47) 10.3	(41) 12.0	(42) 6.0	(45) 10.5	(48) 1.5	(46) 7.8	(44) 6.8	(43) 6.8
Block 6	(40) 6.5	(35) 6.3	(39) 4.0	(37) 11.0	(34) 5.5	(33) 8.0	(38) 7.3	(36) 5.8
Block 7	(62) 11.0	(57) 10.0	(60) 9.0	(64) 16.8	(61) 10.8	(63) 9.9	(58) 11.8	(59) 8.0
Block 8	(6) 7.0	(1) 6.5	(4) 10.5	(3) 7.0	(7) 8.8	(5) 14.0	(2) 10.5	(8) 8.3

Replication II

Block 9	(64) 8.3	(56) 6.8	(48) 1.5	(8) 8.3	(24) 7.8	(32) 9.0	(40) 6.5	(16) 9.8
Block 10	(62) 11.0	(46) 19.3	(38) 9.5	(6) 9.5	(30) 6.8	(22) 11.0	(54) 8.3	(14) 7.0
Block 11	(53) 11.3	(45) 10.5	(37) 11.0	(61) 7.0	(13) 7.5	(5) 14.0	(29) 6.3	(21) 8.8
Block 12	(3) 9.3	(59) 8.0	(51) 5.0	(43) 11.5	(11) 6.0	(27) 6.8	(19) 12.3	(35) 14.3
Block 13	(31) 12.0	(7) 13.3	(63) 9.9	(23) 7.3	(15) 12.3	(55) 9.5	(39) 7.0	(47) 12.5
Block 14	(44) 6.8	(20) 10.3	(4) 10.5	(12) 11.5	(28) 7.5	(52) 8.3	(60) 9.0	(36) 5.8
Block 15	(1) 9.8	(25) 10.3	(17) 9.8	(9) 11.8	(57) 10.0	(41) 7.0	(33) 9.0	(49) 6.8
Block 16	(26) 10.0	(34) 5.5	(2) 10.5	(58) 9.5	(42) 6.0	(18) 9.3	(50) 11.3	(10) 10.3

2. An unreplicated factorial experiment was performed to determine the effects of 4 different levels of each of 3 factors (temperature, time of reaction and alkali per cent) and their interactions on the yield per cent of pulp recovered from a cellulosic raw material. Analyse the data and test the significance of main effects and two-factor interactions assuming the three-factor interaction to be non-existent.

Split up the variance of each main effect into linear, quadratic and cubic components and test the significance of each.

Temperature °C		138				143			
Time of reaction (hours)		1	2	3	4	1	2	3	4
Alkali per cent	12	51.1	48.8	46.1	42.0	46.0	42.1	39.1	37.0
	15	49.3	48.1	45.2	39.1	43.2	40.0	37.0	35.1
	18	48.1	46.2	43.1	38.2	41.6	38.5	36.0	34.2
	21	47.4	45.3	42.2	37.1	39.8	36.2	34.8	33.1
Temperature °C		**148**				**153**			
Time of reaction (hours)		**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
	12	43.0	40.4	38.1	36.0	38.4	36.4	34.1	32.0
	15	41.2	38.1	36.4	34.2	36.7	33.8	32.2	30.2
	18	40.1	37.2	35.7	33.1	35.4	32.6	31.3	29.3
	21	39.2	36.1	34.5	32.4	34.7	32.1	30.4	29.3

3. (a) Analyse the data given below of yield in ounces for each plot of a randomized block experiment involving 8 strains of *mung*.

Strain / Block	1	2	3	4	5	6	7	8
I	17.5	27.0	22.0	10.5	21.0	25.0	14.5	11.5
II	20.0	15.0	21.0	11.0	22.0	27.0	15.0	18.0
III	15.0	20.5	13.5	19.0	5.5	12.0	16.0	13.0
IV	31.5	19.5	24.5	19.0	24.5	29.0	17.5	11.0
V	20.0	25.0	26.0	24.0	26.0	31.5	14.0	16.5
VI	21.0	20.5	20.5	16.5	24.5	22.0	14.0	15.5

(b) The experimenter reports that the plots allotted to strains 5 and 6 in block III were badly damaged by birds etc., thus depressing the yields of these plots considerably. Analyse the data treating these two plots as missing. Comment on the differences you find in the results compared to the previous analysis in regard to (i) error variance, (ii) variance between strains and (iii) test of significance of differences between strains 5 and 6 and between any of them and any of the remaining six strains.

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPERS VIII AND IX: SAMPLE SURVEYS—APPLIED (PRACTICAL)

Time: 4 Hours | Full Marks: 100

(a) Attempt ANY THREE questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. You are required to draw up a simple schedule for the purpose of estimaing the damage done by cyclone to house property from a rapid survey of households. Show the proforma for such a schedule and write out the instructions for the field staff who would fill up the schedule.

2. Draw up a schedule, as complete as possible, for the economic survey of small scale and cottage industries. All the important features, which should be common to all types of such industries, should be incorporated in the schedule.

3. Prepare the first year's budget estimate for the cost of field work on crop cutting experiments for 10 *Rabi* crops of your State. The following points should be taken into consideration in the preparation of the budget estimate:—

(i) One investigator (primary field worker) can take two cuts a day.

(ii) There are 240 police stations in the State, for each of which the number of cuts per crop has to be 30.

(iii) Labour charges should be provided at the rate of Re. 1/- (one) per cut.

(iv) There should be provision for supervisory staff, printing of forms, and stationery.

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

(vi) Provision need not be made for the crop cutting accessories or any other contingent expenditure.

4. (i) If the values in the table given below are taken to represent measurements on a random sample of 20 objects selected from a batch of such objects with a sampling fraction of 1/25, estimate the mean measurement of the batch, the total of all measurements of the batch, and the standard error of the estimated mean.

(1)	(2)	(3)	(4)
6.2	8.0	8.2	11.0
13.8	12.0	8.7	10.3
8.0	10.7	8.5	14.6
7.6	9.1	10.1	8.0
10.3	10.4	9.3	9.0

(ii) If the number in the batch is known to be 507, what should you suggest for the estimate of the total?

(iii) If the above measurements in the four columns are taken to have been drawn from four strata consisting respectively of 125, 130, 140 and 112 objects, into which the batch might be supposed to have been divided, how will you estimate the total of measurements?

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

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

Time : 4 Hours | Full Marks : 100

(a) Attempt either (a) or (b) of Question 1 and either (a) or (b) of Question 2.

(b) Figures in the margin indicate full marks.

(c) Use of calculating machines is permitted.

1. (a) The following table shows the results of a sampling experiment to estimate the yield-rate of wheat in a certain region, the details of which are described below :— (80)

TABLE 1(a) : YIELD FIGURES FOR WHEAT PER UNIT AREA IN SOME SUITABLE UNITS

	District I				District II			District II	
1st line Set 1	47	48	75	105	93	58	76	77	60
1st line Set 2	63	51	71	82	84	78	57	74	88
2nd line Set 1	67	45	75	97	75	68	70	44	84
2nd line Set 2	55	46	85	80	80	83	78	57	97

	District IV												
1st line Set 1	66	93	55	127	84	80	81	93	21	84	87	79	90
1st line Set 2	73	70	56	106	80	86	107	106	63	51	67	79	117
2nd line Set 1	64	80	83	84	63	88	135	71	50	82	135	71	112
2nd line Set 2	73	67	60	98	89	110	82	83	29	80	114	80	122

	District V										
1st line Set 1	29	45	57	69	78	59	68	97	60	65	81
1st line Set 2	21	39	63	55	100	59	56	88	53	59	94
2nd line Set 1	20	60	46	21	90	58	74	109	43	49	93
2nd line Set 2	31	57	66	40	51	53	61	95	48	71	92

Five districts were chosen subjectively and a number of farms were chosen in each district, selection of farms in each district was not random, the farms being taken in the neighbourhood of the centres at which the investigators were located. One or two fields were chosen at random in each chosen farm. Each chosen field was traversed in the direction of the rows along two lines chosen at random. Two sets of unit area were taken randomly in each line. The yield of grain (in some suitable unit) is shown in the Table above. The four values in each column refer to the values in the four sets of a field. Fields in the same farm are bracketted together at the top, the absence of such a bracket means that only one field was chosen from the farm.

(i) Work out the analysis of variance table from the data supplied.

(ii) Estimate the relative gain in efficiency by taking four lines randomly from each field (instead of only two) with two sets per line as before. (You may ignore the 'finite population corrections' and the change in cost).

(iii) Further work out roughly the 95 per cent confidence limits of the estimate of the relative gain in efficiency in (ii) above.

(b) The following table shows the stratification of all the farms in a region by farm size, and the average acres of corn (maize) per farm and the standard deviations (in acres of corn in the farms). All these are *population* values.

TABLE 1(b): STRATIFICATION BY FARM SIZE ETC.

Farm size (acres)	Number of farms	Average corn acres	Standard deviation (corn acres)
0— 40	394	5.4	8.3
41— 80	461	16.3	13.3
81—120	391	24.3	15.1
121—160	334	34.5	19.8
161—200	169	42.1	24.5
201—240	113	50.1	26.0
241—	148	63.8	35.2

If one takes a sample of 100 farms (for estimating the average corn acres in the whole region), compute the sample sizes in each stratum under (i) proportional allocation, and (ii) optimum allocation. Compare the precisions of these methods with that of simple random sampling.

If the 7 strata are to be combined into 2 strata, show that the best point of division is at 120 acres. What is the relative precision of 2 strata to 7 strata under proportional allocation ?

2. (a) The following table shows the results of a one-dimensional systematic sample. (The variate is area of green land in a given unit). (20)

TABLE 2(a): RESULTS OF A SYSTEMATIC SAMPLE

Serial No.	1	2	3	4	5	6	7	8	9	10
Green land	0.0	0.9	0.0	0.0	0.3	0.1	0.5	3.1	2.8	2.7
Serial No.	11	12	13	14	15	16	17	18	19	20
Green land	2.8	2.6	2.3	3.5	2.4	3.8	4.1	4.9	6.0	5.4
Serial No.	21	22	23	24	25	26	27	28	29	
Green land	2.3	2.9	2.1	6.3	8.2	5.4	6.6	6.6	4.1	

Estimate roughly the standard error of the sample mean by the method of successive differences.

Estimate further the standard error of the sample mean assuming the sample to be a simple random one (ignoring the 'finite population corrections'). Comment on the results you obtain.

(b) In connection with planning a sample survey for estimating the proportion of land under jute in a particular region, it was decided to study first the relation between s, the size (i.e., area) of the sample-unit and the corresponding variance, V_s. For this purpose a pilot survey was undertaken in which separate and (independent) random samples were taken for different sizes, s, of the sample-unit and the estimates, V_s, for the corresponding variance were calculated, the values of which are shown in the following table. (In the table, n denotes the sample-size on which the estimate, V_s, is based).

TABLE 2(b): RELATION BETWEEN VARIANCE AND SIZE

s Sample-unit size (acres)	n Sample-size	V_s Estimate of variance
1.00	1756	0.1119
2.25	1377	0.0813
4.00	1161	0.0659
6.25	951	0.0577
9.00	832	0.0505
12.25	732	0.0455
16.00	618	0.0419
25.00	476	0.0398
36.00	365	0.0342

From inspection of the results it was suggested that the relation between V_s and s ('variance function') may be of the form

$$V_s = \frac{a}{s^g}, \text{ where } a \text{ and } g \text{ are constants.}$$

Estimate the values of a and g from the data supplied and test for the suitability of the suggested form for the variance function. (You may ignore the 'finite population corrections').

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

PAPERS VIII AND IX: ECONOMIC STATISTICS (PRACTICAL)

Time: 4 Hours | Full Marks: 100

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

1. The table below gives the values of machinery used and output produced per worker in some manufacturing industries of India in 1950, the factories within an industry being classified according to size, as reflected by the average number of workers employed per day. (20)

dustry	Size-class	Per worker value of	
		Machinery (Rs.)	Output (Rs.)
Sugar	1000 and above	853	10609
	250—999	1835	10606
	below 250	1142	8819
Paints and varnishes	250—299	1206	21455
	60—249	1054	8014
	below 60	1022	2651
Soap	500—1999	1666	32211
	40— 499	728	9284
	below 40	954	9709
Cement	1000 and above	2740	9643
	250—999	4912	11420
Paper and paper board	1000 and above	3341	7444
	500—999	11396	7053
	below 500	2247	2801
Cotton Textile	1500 and above	685	5059
	500—1499	1204	5049
	below 500	1686	4051
Jute Textile	2500 and above	530	5272
	1000—2499	454	5032
	below 1000	971	5396

Write a note on capital-intensiveness and capital-output ratio of these industries.

2. The following table gives the indices of volume of imports, import prices and home employment (with trend eliminated) in the United Kingdom diuring 1924–38. (60)

Year	Import	Home Employment	Import Price
1924	95.5	103.3	105.6
1925	100.0	102.9	103.9
1926	104.3	95.9	98.2
1927	105.4	104.6	99.2
1928	99.2	103.3	105.8
1929	103.4	104.1	105.1
1930	102.4	98.6	93.9
1931	104.5	93.7	83.4
1932	90.5	91.9	91.2
1933	91.7	94.0	94.3
1934	95.8	95.4	96.8
1935	96.0	98.5	100.4
1936	101.9	102.5	102.9
1937	107.8	106.6	110.0
1938	102.0	104.9	108.0

Obtain the elasticities of import with respect to home employment and import price and explain the significance of these elasticities.

3. Prepare a schedule that you may like to use for a survey of unemployment in your State. (20)

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1950

PAPERS VIII AND IX: MATHEMATICAL THEORY OF SAMPLING DISTRIBUTION (PRACTICAL)

Time : 4 Hours | Full Marks : 100

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

1. (a) Take a random sample of 30 observations from a known normal population. Calculate the mean and the variance of the sample. Find the probability of obtaining the mean and the variance equal to or greater than the observed values.

(b) By examination $z_i = (x_i - \bar{x})/s$ find the number of observations of your sample which is greater than or equal to 20 per cent value of z_i.

2. Draw five random samples of 10 observations in each sample from a bivariate normal population with the following parameters.

$$m_x = 5, \; m_y = 0$$

$$\sigma_x = 1, \; \sigma_y = 1$$

$$\rho = .5$$

Determine the regression coefficients of y and x for each of the samples and find the probability that the regression coefficients will be equal to or greater than the observed values.

3. Ten observations on x, y and z are given below :—

x	y	z
54	21	19
60	23	35
56	35	38
57	32	42
54	23	32
61	28	24
56	24	29
58	34	36
61	26	31
54	30	40

Assuming the hypothesis $\mu_x = 55$, $\mu_y = 25$ and $\mu_z = 30$, calculate Hotelling's T^2 and find the probability of $T^2 <$ the observed value.

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1956

Papers VIII and IX : Actuarial Statistics (Practical)

Time : 4 Hours Full Marks : 100

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

1. The payment under 2½ per cent U.P. Zamindary Bond of the face value of Rs. 100 is an annual payment of Rs. 4/- per annum payable for 40 years. The annual payment is made to cover interest on the outstanding capital and part repayment of the capital. The interest portion of the payment is subject to the Union Government's incometax at the rate of 5 as. 4 ps. to the rupee.

What price would you pay for the Bond with 40 payments outstanding if you want to realise 4½ per cent net on the whole transaction

2. An industrial company issued a debenture repayable at 102 in 40 years from now carrying interest at the rate of 3½ per cent per annum payable half-yearly. The debenture holders have recently consented to the issue of a further loan to rank *pari passu* with the earlier loan in return for an increase in the interest rate to 4 per cent per annum. The stock is currently quoted at 95 x.d. What is the market rate of interest ?

The Company now asks that instead of this increase in interest rate the debenture holders should accept a reduction in the term of the debentures. Calculate the minimum reduction in term that would be acceptable to an investor (ignore incometax).

3. On the assumption that mortality table follows Makeham's Law, obtain the graduated values of q_x from the following grouped data for ages $x = 35$ to $x = 45$.

Age Group	Exposed to risk	Deaths
21—25	28,054	34
26—30	72,537	82
31—35	1,06,584	124
36—40	1,62,900	260
41—45	1,96,020	489
46—50	1,87,428	849
51—55	1,44,102	1,156
56—60	1,00,225	1,303
61—65	54,746	1,183

4. Give the values of q_x for $x = 30$ to $x = 55$. Calculate the values of $A_{30:30:\overline{25|}}$, $a_{30:30:\overline{25|}}$ and $P_{30:30:\overline{25|}}$ at 3½ per cent.

x	$q_x \times 10^5$	x	$q_x \times 10^5$	x	$q_x \times 10^5$
30	116	41	208	51	671
31	118	42	231	52	750
32	120	43	259	53	837
33	123	44	292	54	931
34	127	45	330	55	1035
35	132	46	372	56	1148
36	139	47	420	57	1272
37	148	48	474	58	1408
38	158	49	534	59	1557
39	171	50	599	60	1720
40	188				