INDIAN STATISTICAL INSTITUTE

QUESTION PAPERS

for

The Statistician's Diploma Examination

August 1955

Price Re. 1/-

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPER I : THEORETICAL STATISTICS (GENERAL)

Time : 4 hours Full Marks : 100

N.B. (a) Answers to the different groups are to be given in separate books.
(b) Attempt ANY THREE questions from each group.
(c) All questions carry equal marks.

GROUP A

1. What is a Central difference formula for interpolation ? Discuss the relative accuracy of different Central difference formulae.

Taking fifth order differences of U_x to be constant and given u_0, u_1, u_2, u_3, u_4, u_5, prove that

$$U_{\frac{1}{2}} = \tfrac{1}{2}c + \frac{25(c-b)+3(a-c)}{256}$$

where $a = u_0+u_5$, $b = u_1+u_4$, $c = u_2+u_3$.

2. Starting from the traditional definition of mathematical probability as the ratio of the number of favourable cases to the number of equally likely cases, develop a theory of probability in continuum, clearly explaining just how and where the generalisations are made. P is a fixed point on a given line AB. Points X and Y are taken at random on AP and PB respectively. If $AP = a$, and $PB = b$, what is the probability that AX, XY, and BY can form a triangle ?

3. Explain clearly the importance of the 'Normal Distribution' in Statistics.

Give a few examples of sampling distributions which tend to normality with increasing sample size. Give also an example where this does not happen.

In a population which is exactly normal, 31 per cent of the individuals are under 45, and 8 per cent are over 64. Find the mean and standard deviation of the population.

4. 'One of the functions of Statistics is the provision of techniques for making inductive inferences and for measuring the degree of uncertainty of such inferences'.

Discuss the above statement.

GROUP B

5. (a) If $x_1, x_2, \ldots x_n$ follow the probability law

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

then show that $i = \sum_1^n \frac{(x_i-\bar{x})^2}{\sigma^2}$ is distributed as χ^2. Derive a test of significance of the difference of an observed standard deviation and a specified population value.

(b) Suppose that you are told that an epidemic of cholera develops according to a particular law. Briefly explain how you would verify this statistically.

6. (a) If x, y and z represent three characteristics of an individual, what is meant by the partial correlation between x and y and the total correlation between them. If the partial correlation coefficient $\gamma_{xy.z}$ is positive, will the total correlation coefficient γ_{xy} be also positive? Give reasons for your answer.

(b) If the age, height and weight of each N children in a school are given, obtain a prediction equation for age in terms of height and weight.

If the multiple correlation of age on height and weight is high, can you infer that the predicted values of age will be very reliable?

7. (a) What is the z-transformation of the correlation coefficient γ? Explain its uses (i) in testing the equality of several correlation coefficients and (ii) in combining them into a single coefficient when the first test is favourable.

(b) A random sample of n paddy fields were treated by two different fertilizers in two successive years. The records of yields of paddy are to be studied to decide whether any reduction in the variability between fields has resulted by changing the fertilizer used. State with suitable comments the appropriate statistical test which you would recommend for this purpose.

8. (a) Analyse the total variance in a Randomized Blocks experiment. Stating the assumptions made, show how it is possible to test for the significance of (i) the differences amongst all the varieties, and (ii) the difference between two particular varieties.

(b) Find the expression for the standard error of the second moment m_2 of a sample. Hence derive the standard error of $\sqrt{m_2}$. How do these expressions get simplified when the population is normal?

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPER II : APPLIED STATISTICS (GENERAL)

Time : 4 Hours — Full Marks : 100

N.B. (a) Answers to the different groups are to be given in separate books.
(b) Attempt ANY THREE questions from each group.
(c) All questions carry equal marks.

GROUP A

1. Explain the various steps you would take in framing a questionnaire for a field survey.

Prepare a short schedule with suitable instructions for collection of information by the interview method regarding the propagation of news in the rural areas of your State—sources of news, main news items of interest, etc.

2. Under what conditions would you employ the ratio method of estimation in a sample survey and why ?

Outline a sampling design for an enquiry to estimate the total consumption of cereals by the rural population of your State, and discuss how far population census data could be fruitfully utilised for purposes of estimation.

3. (a) Describe the official method of collection of statistics of acreage under crops in (a) temporarily settled, (b) permanently settled and (c) unsurveyed areas of India. To what extent do you consider these statistics to be reliable ? Suggest ways of improving their accuracy.

(b) Describe briefly the statutory census of Manufacturing Industries in India. What are the limitations of the data collected ?

4. (a) Explain the meaning and uses of Cost of Living Index Numbers. How are they constructed for the working class in the Capital City of any Indian State ?

(b) Write a short account on the method of computing 'National Income' in India adopted by the National Income Committee of the Government of India.

5. Indicate how family budget data could be utilised for studying the demand variations for any particular item of consumpton, due to changes in the size of income, pointing out clearly the assumptions made.

State to what extent does the information collected in the National Sample Survey serve the purpose of family budget data.

Group B

6. What are balanced incomplete block designs ? What does this balance mean and how is it attained ? State the fundamental parametric relations. Indicate the structure of the analysis of variance and mention the important steps in the calculations involved.

7. What is meant by an uniformity trial and what is its purpose ? Indicate the important steps that should be taken in preparing the lay-out for such a trial, stating the precautions which you would take to ensure that the yields do not get mixed up ?

Describe in brief what analysis you would perform on the data.

8. Trace the gradual development of the process of factor analysis. Discuss in particular the salient points in (a) Two factor theory, (b) Sampling theory, (c) Multiple factor theory.

9. (a) Describe the various measures of fertility in common use. Show that under certain conditions, to be stated, a very fertile community can remain stationery in size for a long period of time.

(b) What is a logistic curve ? Explain to what extent it is useful for population projection in India.

10. 'Linkage is an association of genetic factors which alters the normal ratios resulting from segregation and independent assortment'—Explain clearly what is ment by this statement.

In a back cross involving two factors, suppose one parent is doubly heterozygous, *Aa*, *Bb* and the other doubly recessive *aa*, *bb*. In this case, how does each of the two factors segregate according to the Mendelian expectation, and what statistical test do you propose to apply to detect linkage ?

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPER III : STATISTICAL INFERENCE

Time : 4 hours Full Marks : 100

N.B. (a) Answers to the different groups are to be given in separate books.
(b) Attempt ANY THREE questions from each group.
(c) All questions carry equal marks.

GROUP A

1. What is a consistent estimate ? Show that if both the bias of an estimate and its variance approach zero with increase in sample size, then the estimate will be consistent.

Show that maximum likelihood estimates are consistent under certain conditions to be stated. Give an example where such an estimate is not consistent.

2. State and prove Markof's theorem on best unbiased linear estimates. How is this theorem modified for the case where the ratios of the variance of the variates are known but are not all equal to unity ? Give its use in analysis of variance.

3. (a) State and prove the Cramer-Rao theorem on the lower bound to the variance of unbiased estimates. Give its analogue in the theory of multi-parameter estimation.

(b) When is a statistic said to be sufficient for a parameter θ ?

If an unbiased estimate and a sufficient statistic exist for θ, show that the best unbiased estimate of θ is an explicit function of the sufficient statistic.

4. Define asymptotic efficiency of estimates. Show that if θ and ϕ are jointly asymptotically normally distributed and if θ is most efficient asymptotically the coefficient of correlation between θ and $\theta-\phi$ approaches zero. State the corresponding result for small samples. Show that the median based on $(2m+1)$ independent observations from the distribution with density function

$$\tfrac{1}{2}\exp(-|x-\theta|), -\infty<x<+\infty$$

is most efficient asymptotically and obtain its asymptotic intrinsic accuracy.

5. Describe briefly Neyman's theory of confidence intervals.

Obtain from first principles the confidence limits for θ from a sample from a rectangular distribution in the rage $(0,\theta)$.

Show how it is possible to obtain an exact confidence interval for the population proportion π from the observed proportion p in a sample of size n, the confidence coefficient being α.

GROUP B

6. What is meant by a null hypothesis ? What is the meaning of the 5 per cent and 1 per cent significance levels ?

In testing whether a coin is unbiased, we should have the results of not less than 4 or 5 tosses. Do you agree ? Give reasons.

Develop a test for the equality of two proportions of success on the basis of two independent samples of sizes n_1 and n_2 with the observed number of successes r_1 and r_2 respectively. Ensure that the test procedure is independent of the unknown common proportion.

7. What is a likelihood ratio criterion ? Are you intuitively satisfied with its use ? Try to support your intuition to the extent possibly by a process of reasoning which makes fullest use of the properties of the criterion.

Derive the likelihood ratio test for the hypothesis that the means of two normal populations are the same when the common standard deviation is unknown, the information available being two random samples drawn from these two populations.

What is the test for equality of variances when the means are unknown and presumably different ?

8. Write a critical note on the role of multivariate tests in scientific investigations with particular reference to the two situations described below :—

(1) Suppose that two populations are distinguishable on the basis of two samples on a single character by using a 't' test. If measurements on more than one character are available, what test would you use ? Is it necessary that this test should show significance when the 't' test for one of the characters shows significance ?

(2) It has been pointed out that in some situations the test for the correlation between the variables y and x_1 showed significance while the test for multiple correlation of y on x_1 and x_2 was not significant. Do you consider this as a drawback of the test for multiple correlation ? Do you suggest that if you want to test the dependence of y on the set of variables (x_1, x_2) you would test for the individual correlations separately rather than the multiple correlation of y on x_1 and x_2 ?

9. Derive the analysis of dispersion test for the hypothesis that each of the four correlated variables x_1, x_2, x_3, x_4 has linear regression with another variable t observations on (x_1, x_2, x_3, x_4) for each of k different values of t being given. You may first consider the appropriate analysis of variance test for a single variable x and then generalise.

10. Discuss the principles behind the theory of testing hypotheses according to Neyman and Pearson.

State and prove a set of sufficient conditions for the existence of similar regions for testing a composite hypothesis having one unspecified parameter and show how it is possible to select the best critical region for the test from amongst the similar regions.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

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. 2500 screws produced by two lathes in a certain factory were found on measurement to have a mean length of .999 inch and a standard deviation of .003 inch. The lengths of the 1500 screws produced by the first lathe are shown in the following frequency table :—

Length of screw to nearest thousandth of an inch	Number of screws produced by the first lathe
.992	1
.993	5
.994	48
.995	161
.996	301
.997	481
.998	308
.999	170
1.000	37
1.001	7
1.002	1
Total	1500

Calculate the mean length and standard deviation of length of screw for each lathe. (Do not apply Sheppard's adjustment). (10)

2 The results of estimation of carbon content (per cent) of 60 specimens of clay by each of three different methods, A, B and B' are given below. Method A is an accurate reference method for the determination of carbon and the values obtained by method A are denoted by x. Methods B and B' are approximate methods both of which give values (denoted by y and y' respectively) which are roughly proportional to the correct one. Estimate the constant of proportionality b in the equation $y = bx$ assuming that x is virtually free from error and the error in y (standard deviation) is (i) constant over the range of measurement, (ii) proportional to the value of x and (iii) proportional to the square root of x. Also supply estimates of the constant of proportionality b' in the equation $y' = b'x$ under assumptions similar to (i), (ii) and (iii) above.

The estimation may be made by the method of least squares using appropriate weights.

Specimen No.	Method A	Method B	Method B′	Specimen No.	Method A	Method B	Method B′
	x	y	y'		x	y	y'
1	0.74	0.55	1.09	31	0.12	0.09	0.12
2	0.13	0.00	0.06	32	1.50	1.13	2.36
3	0.15	0.10	0.00	33	1.31	0.98	2.14
4	1.47	1.17	2.33	34	0.31	0.15	0.08
5	0.52	0.38	0.50	35	0.14	0.15	0.00
6	0.27	0.22	0.53	36	2.98	2.18	4.53
7	0.99	0.70	0.65	37	6.84	4.51	9.94
8	0.53	0.49	0.72	38	2.15	1.59	3.68
9	1.53	1.17	2.46	39	1.35	1.15	1.84
10	0.87	0.70	1.54	40	0.40	0.33	0.97
11	0.28	0.12	0.70	41	4.18	3.24	6.14
12	0.27	0.16	0.00	42	0.22	0.21	0.52
13	3.07	2.11	4.82	43	0.38	0.24	0.40
14	0.25	0.18	0.30	44	0.24	0.18	0.46
15	0.25	0.16	0.64	45	1.79	1.24	2.80
16	0.28	0.28	0.78	46	0.58	0.36	2.09
17	2.54	2.32	4.08	47	6.55	5.58	9.68
18	1.43	1.40	2.80	48	0.33	0.18	0.56
19	2.74	2.37	3.93	49	0.48	0.31	1.09
20	6.08	5.39	8.22	50	0.26	0.17	0.01
21	0.75	0.64	0.28	51	0.17	0.11	0.00
22	0.16	0.11	0.35	52	0.20	0.13	0.53
23	5.06	4.51	7.49	53	0.41	0.41	0.68
24	0.86	0.80	1.41	54	0.37	0.31	0.67
25	0.16	0.12	0.00	55	0.98	0.65	0.87
26	0.19	0.14	0.18	56	3.90	3.67	6.55
27	0.13	0.09	0.28	57	2.05	1.69	3.88
28	0.25	0.14	0.63	58	1.30	1.09	2.21
29	0.19	0.22	0.44	59	0.26	0.26	0.36
30	0.79	0.23	0.80	60	0.72	0.50	1.61

(25)

3. The average monthly production of finished steel in India during the years 1948 to 1954 is given below :—

Year	1948	1949	1950	1951	1952	1953	1954
Finished Steel (000 tons)	71.3	77.4	82.7	87.6	89.7	82.3	101.0

Fit a polynomial of the third degree by the method of orthogonal polynomials to study the trend of production of finished steel in India. Test by the method of analysis of variance whether it is worthwhile to retain the third degree terms in the polynomials. (15)

4. The distribution of 176 coils of a sheet metal is as follows :—

Thickness (in m.m.)	Frequency	Thickness (in m.m.)	Frequency
565.5	4	605.5	28
570.5	14	610.5	14
575.5	14	615.5	—
580.5	8	620.5	2
585.5	10	625.5	—
590.5	20	630.5	2
595.5	24		
600.5	36	Total	176

Find the frequency curve that best fits the data. Test for goodness of fit. (25)

5. (a) In an experimental study to build up a formula for predicting the yield of dry bark from concomitant plant characters, the following total correlation coefficients were obtained from a sample of 32 observations.

x_1 = yield of bark in ounces per plant, x_2 = height of plant in inches, x_3 = girth of plant (in inches) at 6 inches above ground.

$r_{12} = 0.367$, $r_{13} = 0.684$, $r_{23} = 0.321$

$s_1 = 14.25$ $s_2 = 56.67$ $s_3 = 1.03$

$x_1 = 21.68$ $x_2 = 166.4$ $x_3 = 3.14$

Estimate the prediction equation and also find $R_{1\cdot23}$ and $r_{12\cdot3}$. Test the significance of $R_{1\cdot23}$ and $b_{13\cdot2}$

(b) Test whether an observed correlation coefficient of -0.41 obtained from 67 pairs of observations differ significantly from -0.60. State your conclusions. (25)

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

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. Samples of 5 pieces of a certain part of a stove were taken every 15–20 minutes in order of production from a certain production process and over-all length are .830 inch

±.010 inch. From the data given below, examine whether the process is under control and comment on your findings. (20)

Group No.	a	b	c	d	e
1	.833	.835	.834	.833	.833
2	.836	.834	.835	.837	.836
3	.832	.837	.833	.837	.830
4	.835	.836	.836	.834	.833
5	.835	.835	.832	.832	.833
6	.833	.836	.833	.837	.835
7	.834	.834	.834	.834	.835
8	.831	.832	.834	.832	.831
9	.834	.832	.833	.832	.832
10	.834	.832	.832	.832	.833
11	.835	.833	.834	.833	.831
12	.837	.833	.833	.834	.834
13	.835	.836	.835	.837	.836

2. By collecting relevant statistics from official publications examine the progress in attaining the targets set in the following Agricultural and Industrial sectors in the first five year plan :— (15)

Food grains, Jute, Cotton, Oil seeds, Sugar, Steel, Cement, Aluminium, Coal, Textiles, Heavy Chemicals, Diesel Engines, Fertilizers and Power.

3. The results of a randomised block experiment with brussels sprouts are summarised below. Analyse these data as fully as you can, paying attention to immediate and cumulative effects of the various fertilizers. (25)

	Block			
	I	II	III	IV
R_0	27.5	37.5	26.0	21.5
R_1	37.0	20.0	19.5	25.5
R_2	25.0	41.0	20.0	27.0
M_0	17.5	34.5	23.5	23.5
M_1	26.5	34.0	17.0	30.0
M_2	13.5	41.0	20.0	24.4
N_0	17.0	44.5	17.0	20.5
N_1	12.0	32.5	20.5	30.0
N_2	22.5	35.0	15,5	28.0
S_0	15.5	21.5	19.5	28.0
S_1	25.0	26.0	30.0	24.5
S_2	41.0	36.5	25.5	30.0

Notes : The yield is given in lbs. of saleable sprouts per 0.024174 acres. Nitrogen may be applied in four ways either through rape dust (R), poultry manure (M), sulphate of ammonia (N), or soot (S).

R_0 means 0.8 cut. nitrogen per acre applied in 1934 and no nigroten in 1935.
R_1 means 0.4 cut. nitrogen per acre applied both in 1934 and 1935.
R_2 means no nitrogen in 1934 and 0.8 cwt. per acre in 1935.
The subscripts to M, N and S have similar interpretations.

4. The following gives the number of workers in all the 60 establishments in a particular district for two consecutive years.

1953	1954	1953	1954	1953	1954	1953	1954	1953	1954
16	20	21	25	61	73	99	135	44	40
26	36	63	63	87	63	43	39	63	63
38	50	58	48	93	102	45	61	71	92
51	70	47	59	129	159	37	28	82	74
76	104	39	51	35	41	64	88	84	108
81	113	53	53	43	25	52	72	72	93
68	80	54	34	11	8	31	37	50	45
96	87	30	33	86	94	54	54	30	42
75	61	62	86	79	107	19	20	34	26
82	98	70	56	51	61	95	131	60	69
112	136	59	79	56	46	75	75	79	72
20	24	32	36	27	19	36	24	85	101

For each of the schemes of stratification specified below compute the standard errors of the two estimates of the average number of workers per establishment in 1954 for a sample of size 20 with (a) proportional and (b) optimum (for 1953) allocations.

Schemes of Stratification

Stratum 1 : Establishments with workers numbering $\leq W$ in 1953.
Stratum 2 : Remaining establishments.

W being fixed in such a manner that (1) equal number of establishments fall in each stratum, (2) range of variation of number of workers (1953) in an establishment for stratum 1 (approximately) equals that for stratum 2.

Also compare the performances of the above 4 methods with a scheme in which W is fixed in a manner which (approximately) equalises the total number of workers in 1953 and then an equal number of sample establishments are selected from each stratum. (25)

5. From the following data of monthly wholesale farm price index 1943–47, obtain measures of seasonal variations :— (15)

	Jan.	Feb.	Mar.	Apr.	May	Jun.
1943	101.9	102.6	103.4	103.7	104.1	103.8
	July	Aug.	Sep.	Oct.	Nov.	Dec.
	103.2	103.1	103.1	103.0	102.9	103.2

1944	Jan.	Feb.	Mar.	Apr.	May	Jun.
	103.3	103.6	103.8	103.9	104.0	104.3
	July	Aug.	Sep.	Oct.	Nov.	Dec.
	104.1	103.9	104.0	104.1	104.4	104.7
1945	Jan.	Feb.	Mar.	Apr.	May	Jun.
	104.9	105.2	105.3	10.57	106.0	106.1
	July	Aug.	Sep.	Oct.	Nov.	Dec.
	105.9	105.7	105.2	105.9	106.8	107.1
1946	Jan.	Feb.	Mar.	Apr.	May	Jun.
	107.1	107.7	108.9	110.2	110.0	112.9
	July	Aug.	Sep.	Oct.	Nov.	Dec.
	124.7	129.1	124.0	134.1	139.7	140.9
1947	Jan.	Feb.	Mar.	Apr.	May	Jun.
	141.5	144.5	149.5	147.7	147.1	147.6
	July	Aug.	Sep.	Oct.	Nov.	Dec.
	150.6	153.6	157.4	158.5	159.5	163.2

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

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. It is required to compare the changes in prices of commodities sold by farmers with the changes in prices of commodities purchased by them. Describe the calculation of suitable index numbers which may serve the above purpose.

2. Explain the significance of cost of living index. Show that

$$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} > I_0 \text{ and } \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} < I_1$$

where p_0, p_1, q_0 and q_1 have their usual meanings, the summation is over all items of consumption, and I_0 and I_1 are 'true' cost of living indices corresponding respectively to the situations 0 and 1.

3. What is meant by the trend of a time series ? Give an account of different methods of obtaining the trend of a time series.

4. Define serial correlation. How can the serial correlations in a time series be used to determine the type of the series ?

5. Explain what is structural relationship among economic variables. Taking some suitable variables, describe how the structural relation among them can be obtained from time series data.

6. Explain the concept of national income. Distinguish between gross national product, net national product, personal income and disposable income. How will you calculate the contribution of the sectors, agriculture and trend to the national income of India, using latest available information ?

7. Give an account of the types of information collected in a family budget enquiry. What are the uses of this type of data ?

8. Give a critical review of price statistics available in India. Describe some of the uses of this type of data.

9. You are asked to prepare a plan for the expansion of primary education in India such that all persons between the ages of 5 and 14 will get the facility of primary education in 1965-66. Describe your plan with the details and sources of information needed.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPERS IV AND V : VITAL STATISTICS AND POPULATION STUDIES (THEORETICAL)

Time : 4 Hours — Full Marks : 100

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

1. Describe the growth of India's population as judged from Census data and discuss the underlying factors behind it. Explain the demographic bases on which you would work out projections of India's population upto 2000 A.D.

2. Define Gross reproduction rate and explain the method used for its computation. What are the limitations of the use of Gross reproduction rate of study trends in fertility ? What other methods would you suggest for studying fertility trends ?

3. Give the outline of a plan for estimating by means of a sample survey the fertility and mortality rates of the population of an Indian State and their variations with economic and social factors. Indicate the types of schedules that would be used and the items on which information would be obtained. What steps would you take to ensure the accuracy of the data ?

4. Show that any population, if subjected continually to unchanging birth and death rates will attain a stable age distribution. What will be its rate of increase after attaining the stable state ?

5. Write notes on the following :—
   (a) Demographic information available from the National Sample Survey.
   (b) International Statistical classification of causes of death.
   (c) Methods of construction of abridged life tables.
   (d) Infant mortality rate and improved methods for calculating it.
   (e) Reserve-survival method for the estimation of the birth rate.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

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. Compare and contrast the uses of Analysis of Variance with the uses of $\bar{X}$ and R charts in process control by variables. Discuss the assumptions involved in both cases.

State, giving reasons, whether control chart can replace the usual tests of significance.

2. Explain the use of multivariate quality control in industry. What are its advantages over keeping conventional control charts for each characteristic separately ?

Explain the use of Hotelling's T^2 in multivariate quality control.

3. Write a mathematical note on the 'Range in random samples', emphasising the conditions under which the use of mean range to estimate the population standard deviation is justified.

Explain the use of range in control charts.

4. Explain the use of operating characteristic curve in sampling. What is the effect of inspection errors on the *OC* curve ?

The following plan is adopted for inspecting products on a conveyor belt :

Every kth item is inspected until one defective item is found. Then 100 per cent inspection is resorted to until i items have been inspected without finding a defecting item, after which again every kth item is inspected. Assuming that production varies in a random manner and the items are taken as they come, derive expressions for :—

(i) Average number of pieces inspected following the finding of a defect,
(ii) Average number of pieces passed under the sampling procedure before a defect is found,
(iii) Average fraction of total produced units inspected in the long run,
(iv) Average fraction of produced units accepted without inspection.

5. Compare the single, double and sequential sampling plans.

Design a sequential procedure to test whether the proportions of defectives from two processes are the same. Mark the Acceptance, Continuance and the Rejection regions with instructions for installing the sequential plan.

6. Write short notes on the following :.
(1) Group control charts
(2) Control limits and specification limits
(3) Factors affecting the selection of an acceptance sampling plan
(4) Rational sub-groups.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

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

Time : 4 Hours — Full Marks : 100

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

1. X_1, X_2, X_3 are independent random variables uniformly distributed in (0,1). Find the frequency function of X_1+X_2 and $X_1+X_2+X_3$ respectively. Show by induction that the sum of n such independent random variables has the frequency function

$$\frac{1}{(n-1)}\left\{ X^{n-1} - \overset{n}{(1)}(X-1)^{n-1} + \overset{n}{(2)}(X-2)^{n-1} - , \ldots \right\} (0<x<n)$$

2. (a) Find the distribution of the median of a sample of n independent and identically distributed random variables and show that for large n, it is approximately normal.

(b) Find the median of the distribution whose frequency function is given by

$$f(X, \theta) = \frac{a}{\theta}\left(\frac{\theta}{X}\right)^{a+1} \text{ if } X > \theta, = 0 \text{ if } X < \theta \ (a > 0, \theta > 0)$$

If $X_1, \ldots, X_n$ is a sample of n independent observations from this poulation, find the value of θ, say $\theta\ (X_1, \ldots, X_n)$ which maximizes the joint frequency function of the sample.

Prove that $\theta\ (X_1, \ldots, X_n)$ coverages to θ in probability as $n \to \infty$.

3. (a) Write a short note on the important properties of the characteristic functions which make it a useful tool in distribution theory.

(b) $X_1, \ldots, X_n$ are n independent and identically distributed random variables. Give an example to show that their arithmetic mean $\bar{x}$ need not always be asymptotically normally distributed. Prove, however, that $\bar{x}$ is asymptotically normally distributed under appropriate and fairly general conditions.

4. (a) The components X, Y, Z of the velocity v of a molecule with respect to a system of rectangular axes are independent and normally distributed with mean zero and variance σ^2. Find the frequency function of the absolute value of v.

(b) Show that if n independent normal $[N(0, \sigma)]$ variates $X_1, \ldots, X_n$ are transformed to $Y_1, \ldots, Y_n$ by means of an orthogonal transformation, the new variates are also independent and $N(0, \sigma)$.

(c) Suppose $C = (c_{ij})$ is an orthogonal matrix and $Y_1, \ldots, Y_p$ $(p<n)$ are linear functions given by $Y_i = \sum_{j=1}^{n} c_{ij} x_j$, [where the x's are as in (b) above].

Derive the distribution of the quantity $\sum_{1}^{n} x_j^2 - \sum_{1}^{p} y_i^2$. How is this result useful in obtaining the distribution of the sample variance s^2 of sample from a normal $N(0,\sigma)$ population ?

5. Explain clearly how Hotelling's T^2 is a generalisation to the multivariate case of Students' t.

Obtain the distribution of T^2 when the population mean vector is zero.

6. If $x_1, x_2, \ldots, x_n$ are independent random variables with common mean and variance, obtain the distribution of the ratio

$$\frac{(\bar{x}-\mu)\sqrt{n}}{s} \text{ where } \bar{x} = \frac{\Sigma x_i}{n},\ s^2 = \Sigma(x_i - \bar{x})^2/\ n-1$$

Let $z_j = (x_j - \bar{x})/s$. Show that z_j is distributed as

$$(n-1)t_{n-2}\{n^2n-2+n(n-2)\}^{-\frac{1}{2}}$$

where t_{n-2} has the distribution of t with $n-2$ degrees of freedom.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

Papers IV and V : Actuarial Statistics (Theoretical)

Time : 4 Hours — Full Marks : 100

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

1. (i) What type of experience has supplied the basis of the Indian assured life mortality table in common use 0(25-35) ?

Describe the main features in regard to initial selection and regional and religious group variations found in the analysis of this experience. Indicate the relative levels of Indian and English assured life mortality.

(ii) A scheme is to be drawn up for introducing industrial assurance on country-wise basis in India, for the benefit of the lower income classes. Specify with reasons, the mortality basis that you will recommend for adoption.

2. (i) Discuss the comparative advantages and disadvantages of the summation method of graduation. For what properties will you test a graduated series ?

(ii) Work out a suitable operand for the operator $\frac{[5]^2\,[7]}{350}$ in a formula for graduation by summation, with no second difference error. Calculate the range, the wave cutting index and the smoothing index of the formula. Do you envisage any difficulty in applying the χ^2 test to a graduation by the formula ?

3. (i) Put $\bar{A}^{1}_{xyz:\overline{\omega}_3}$ in the form of integrals and reduce it to contingent assurances payable on second death. Justify the final result directly by general reasoning.

(ii) Do the following expressions represent the same contingent assurances If not, explain what different contracts they stand for :—

(a) $\int_0^\infty v^t(1-{}_tp_y)(1-{}_tp_z){}_tp_x\,dt - \delta\int_0^\infty v^t(1-{}_tp_y)(1-{}_tp_z){}_tp_x\bar{a}_{x+t}\,dt$

(b) $\int_0^\infty v^t(1-{}_tp_y)(1-{}_tp_z)\bar{A}_{x+t}\,dt$

(c) $\int_0^\infty v^t\,{}_tp_x\{(1-{}_tp_y){}_tp_z\mu_{x+t}+(1-{}_tp_z){}_tp_y\mu_{y+t}\}\bar{A}_{x+t}\,dt$

(d) $\int_0^\infty v^t(1-{}_tp_y)(1-{}_tp_z){}_tp_x\mu_{x+t}\,dt$

Give an approximate integration formula for evaluation of expression (d) along with working sheet headings.

4. (i) Express $(I\bar{A})_x : \overline{n}|$ in terms of $\bar{a}_x : \overline{n}|$ and $(I\bar{a})_x : \overline{n}|$

(ii) A small extra premium is charged for non-participating endowment assurance policies for a certain class of under-average lives. Will a high extra premium be necessary for with profit endowment assurance policies to cover the extra risk for the reversionary bonus part ?

If the bonus loading in the level premium is calculated by the formula

$$b.\ \frac{(I\bar{A})_x : \overline{n}|}{\bar{a}_x : \overline{n}|}$$

and the probabilities of survival ${}_tp'_x$ of the class of under-average lives be related to the corresponding probabilities ${}_tp_x$ for the normal class such that ${}_tp'_x = {}_tp_x - \epsilon_{x+t}$, develop a criterion to judge when an extra in the bonus loading is required. In particular, examine if any extra in the bonus loading is required when (a) $\epsilon_{x+t} = \epsilon$, a constant for all values of t and (b) $\epsilon_{x+t} = \epsilon_{x+t}{}_tp_x$

5. (i) Establish mathematically a relationship between $t+1V_{xy} : \overline{n}|$ and $tV_{xy} : \overline{n}|$ and give the expected death strain. What is the altered relationship in the last year of the term, when $t = n-1$?

(ii) Suggest an approximate value of $t+\frac{1}{k}V$ to meet the case where the valuation date is only a fraction of a year ahead of the policy anniversay.

6. (i) Give the value of the increasing continuous annuity $(I\bar{a})_{\overline{n}|}$ in terms of the nominal rate of interest j per annum convertible m times a year. Derive from it or otherwise, the value in terms of the force of interest .

(ii) Show that $a_{t\overline{x}|} > a_x$ where a_x is the curtate expectation of life.

7. (i) If the weighted average period of payment be taken as the equated time of payment for a number of amounts due at different times demonstrate whether such arrangement will be in favour of the creditor or the debtor.

(ii) A practical rule for determining the number of years in which a sum of money will double itself at compound interest i per annum says that $.70 \div i$ gives the number. Establish the rule and also give a first approximation of the error involved. For what rate of interest this rule gives an exact result ?

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1958

PAPERS IV AND V : STATISTICAL MECHANICS (THEORETICAL)

Time : 4 Hours | Full Marks : 100

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

1. Deduce Planck's formula for radiation.

Hence obtain Stefan-Boltymann law and the expression for Stefan's constant. Deduce Rayleigh-Jeans law.

2. Obtain the expressions for the entropy and the specific heat of Einstein's crystal and show that it yields Dulong-Petit law.

Consider the behaviour of the specific heat at low temperatures for Einstein's crsytsl and discuss how it differs from Debye's crystl in this respect.

3. Derive Fermi-Dirac Statistics stating clearly the postulates behind it.

Obtain the expressions for total energy and specific heat for a strongly degenerate gas.

4. Establish Clausius' Virial Theorem. Hence or otherwise deduce the equation of state for Var der Waal's gas.

5. Obtain, for Brownian motion, Einstein's formula as

$$\xi^2 = \frac{kT}{3\pi\eta a}\ \tau$$

where ξ^2 = mean square displacement along x-axis at time τ, K = Boltymann's Constant, T = temperature, η = coefficient of viscosity, a = radius of Brownian particle.

Discuss the importance of the kinetic-theoretical explanation of Brownian motion in development of molecular physics.

6. State and prove Boltymann's H-Theorem. Find out relation between Boltymann's H-function and the entropy and then deduce an expression for the entropy for a monotonic idea gas.

7. Find out the probable error of any function of independent quantities whose probable errors are known.

If a quantity Q be given by

$$Q = f(q_1, \ldots, q_k),$$

where $q_1, \ldots, q_k$ are observable quantities, show how to find the allowable probable errors in q_i's when a prescribed probable error is given for Q.

8. Establish Bose-Einstein statistics. How do you propose to explain physically the hypothesis of the coming of a particle in the phase-cell, occupied by n particles ?

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPERS IV AND V : GENETICS (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 permitted.

1. The following segregations were observed of the recessive factor ivory as opposed to yellow in the flower colour of the plant antirrhinum majus.

Backcross families	Family No.	Yellow	Ivory
(Ii × ii)	1	33	12
	2	38	23
	3	27	15

F_2 (Four yellow plants from each of families 1 and 2 above were self pollinated and gave).

Family No.	Yellow	Ivory
1a	27	2
2a	15	12
3a	24	6
4a	25	5
2a	44	16
2b	21	7
2c	23	7
2d	22	6

Show whether the joint segregation disagrees with simple expectation from a single gene pair with yellow completely dominant over ivory.

To account for the significant discrepancy from the simple expectation that would be revealed by a proper test, a further hypothesis that a recessive lethal gene (l) is linked with the gene for flower colour is proposed. How would you propose to test this hypothesis ? (Do as much numerical work as you can).

2. Explain the role of discriminant function is selection for the improvement of plants and animals. Illustrate your answer with a concrete case, showing the type of data required and the method of their analysis.

Explain any cases of actual application of the discriminant function for this purpose and the broad conclusions reached there regarding the efficiency of this technique.

3. In an experiment to study of quantitative character (abdominal chaetae number) in *drosophila*, cultures were raised of two parental strains O and B, their F_1 and F_2 and biparental progenies from F_2. Each culture was raised in a separate bottle. There were five cultures of each parent strain and of the F_1, twelve of F_2 and thirtyone biparental progenies each in its own bottle. The following statistics were calculated from the observations made on the flies:

Variance of F_2
Mean variance within biparental progenies
Covariance between F_2 parents and their biparental progenies
Variance of biparental progeny means
Variance within bottles } with each parent and F_1
Variance between bottles }

Explain the procedure for estimating fixable and non-fixable components of heritable variation and the non-heritable variation from the above statistics. Explain also how the data can be used to detect possible linkage among genes controlling this character.

4. Describe Bernsteins' method of estimation of gene frequencies as applied to the ABO blood group system in man. Compare it with the method of maximum likelihood.

In a random sample of the British population the following phenotypic distribution was observed for this character:

Blood Group	O	A	B	AB
Frequency	202	179	35	6

Illustrate your answer with the help of these data. (Do as much numerical work as you can).

5. 'Any F_2 contains twice as much information about the recombination fraction as a backcross, but the limitations of classification result in a certain loss which in the case of completely dominant genes in close repulsion may amount to an extremely large proportion of the whole.'

Explain the concept of information and on that basis discuss the above statement critically.

6. Comment on the following statement: 'In estimating linkage, when linkage is close, progenies obtained by self-fertilizing heterozygotes in coupling one of nearly equivalent value with backcross progenies. The advantage of the latter in that case lies in the opportunity it affords of eliminating and of evaluating differential viability and of detecting any difference there may be in the recombination fraction in the male and female gama to genesis'.

7. A stock of plants initially consisted of p_0 of the type AA, q_0 of the type Aa and γ_0 of the type aa, $(p_0+q_0+\gamma_0=1)$. With continued self-fertilisation, what will these proportions be after n generations? What will the approximate proportions be if n is at all large?

8. In the auto tetraploid species Lythrum salicaria plants with three types of flowers, with a long style, with a medium style and with a short style, occur. The factor for mid is M which is absent in the Long plants, the genotype for the latter being represented by m_4. The factor for short is S and all Long and Mid genotypes have their counterparts in the short plants. The frequencies of different genotypes in a population of L. Salicaria are shown below :—

	m_4	Mm_3	M_2m_2	M_3m	M_4
		b'	c'	d'	e'
s	a' long	Mid			
		b	c	d	e
S	a	Short			

Fertilization is possible only among unlike phenotypes. In the gametic output a fraction s will be produced by a process known as double reduction. These gametes consist of two homologous genes or duplicates derived from genes on sister meiotic chromatids. Express the relative genotypic and phenotypic composition of the next generation in terms of the initial frequencies and s.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

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

Time : 4 Hours — Full Marks : 100

N.B. (a) Attempt ANY TWO questions from each group.
(b) Figures in the margin indicate full marks.

GROUP A

1. (a) Explain under what circumstances biases may be introduced in the estimates by (i) selection of units nearest to randomly selected points, (ii) substitution of one random household by another randomly selected one.

(b) It is desired to estimate the 'total population' (to be suitably defined) of an urban area from a sample of houses by collecting information about household size for all the households occupying the sample houses, the survey being conducted over a period of one week. Examine critically as to which of the following is preferable : household size being defined as the total number of persons who have passed (i) the night previous to the date of enquiry in the premises occupied by the household, or (ii) at least four nights out of the last seven nights. Can you suggest any better definition ? (30)

2. What items of information would you collect from each sample household in a two-stage stratified sampling scheme (with towns and villages as first-stage units and households as second-stage ones) in order to obtain an estimate of the total number

of agricultural holdings jointly operated by two or more households, for a given State. What difference would it make if you leave out the towns ? Also establish the formula which you will use in estimtating the S.E. of your estimate. (The details of the sample design are at your choice). (30)

3. It is desired to carry out this year a sample survey in the rural areas of India with the following sampling structure : A two-stage stratified scheme with districts as strata, a fixed number of sample villages per district selected with replacement with probability proportional to population (1951) and then a sample of household from the selected villages, the list of households being constructed by the investigator immediately before the enquiry proper. It is considered desirable for operational convenience if the investigator is asked to draw a systematic sample of households, the random start and sampling interval being already given to him. In what manner will you fix these two (sampling interval and random start) so that a self-weighting sample may be obtained ? Do you think an *exact* self-weighting scheme can be evolved ? Also explain how you will estimate the S.E. of an estimated total from your data. (30)

GROUP B

4. Write a note on the method of double sampling (1) with regression estimates or (2) with stratification, dealing particularly with the problem of optimum allocation. (20)

5. Write a note on optimal allocation of sample in a stratified simple random sampling scheme when population totals of more than one variate are to be estimated. Restrict yourself to the case of minimisation of cost when the population totals are desired to be estimated with given precision. (20)

6 Write notes on :— (20)

(a) Bias of a ratio estimate

(b) Variance of a double ratio estimate.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

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

Time : 4 Hours Full Marks : 100

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

1. (a) Discuss the likely defects in a 'frame' used for a sample survey.
(b) Describe suitable 'frames' for sample surveys :
(i) for a study of the agricultural holdings in the rural areas of India, and (ii) for estimating the yield per acre of wheat in any state of India.

2. Describe the use of punched-card methods for the analysis of data of a large-scale sample survey, explaining the functions of the different machines in any of the system—Hollerith, IMB, and Powers Samas.

3. Discuss the problem of controlling the numerical accuracy in the analysis of data of large-scale sample surveys by the method of duplicate computation and other suitable methods.

4. Discuss the problems of organisation, selection, training and supervision of the enumerators for conducting a sample survey of public health in the rural areas of your State.

5. A large- scale sample survey has to be undertaken for estimating the acreage under wheat in a state of India using square-shaped sample units located randomly; the primary enumeration will, however, be made not directly for the sample units, but for smaller-sized revenue plots. Describe clearly in this connection (i) how the sample-units will be located randomly first on maps of different scales and then on the ground, (ii) how the actual enumeration will be made, and (iii) how the primary data will be converted to readings for sample-units considering the cases when revenue plots lies either partly or wholly inside any sample-unit.

6. Discuss the following statement quoted from the 'Final Report of the National Income Committee, February 1954'—issued by the Department of Economic Affairs, Ministry of Finance, Government of India :

'Within the last decade crop-cutting experiments of the Indian Council of Agricultural Research (ICAR) and the Indian Statistical Institute have made considerable progress and now (*i.e.*, in 1952-53) for the major cereal crops of rice and wheat, yield estimates are made as a result of ICAR crop-cutting experiments over bulk of the area under these crops.'

Also describe briefly the present position in India regarding the crop-cutting surveys for the major commercial crops.

7. Discuss how you will conduct a pilot survey for collecting relevant information on cost and variance functions, to enable you later to plan family budget inquiry in a large city in India on a sample basis.

8. (a) Consider the problem of counting the number of people in a crowded street by any sampling method. (The people are, of course, moving).

(b) Discuss the merits (and demerits) of aerial survey photographs as a 'frame' for agricultural and land utilisation surveys.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

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. Develop the principle of analysis of Latin square designs for estimating treatment effects and testing their difference. Enumerate the advantages and disadvantages of the Latin Square design.

2. Show that in a factorial design of the 3^3 type, the 26 d.f. for treatment comparisons can be partitioned into 13 mutually independent sets of 2 d.f. Show how to arrange the experiment and to analyse the data when 4 replications are available and the sets of 3-factor interactions are partially confounded.

3. For testing the difference between 4 treatments (t_1, t_2, t_3, t_4) the lay-out was in blocks of two treatments as follows with randomisation within each block :—

Replication 1		Replication 2		Replication 3	
t_1	t_2	t_2	t_1	t_1	t_2
t_3	t_4	t_4	t_3	t_3	t_4

The whole set of the three replications was repeated 10 times giving altogether 30 replications. Give a plan for the analysis of the corresponding yield data, and calculate the efficiency of this design relative to a randomised block design with the same experimental units. Does the efficiency change if inter-block information is also taken into account ?

4. In a varietal experiment with 8 varieties (1, 2, . . . , 8) a design is constructed using 8 blocks of size 3 as follows :

(124), (167), (135), (258), (378), (468), (236) and (457)

Find out the parameter of the design and develop, in general terms a method of calculating varietal effects eliminating the effect of blocks.

5. Suppose the t treatments are tested on a random sample of v varieties in randomised blocks of v plots using r blocks, and that the yields Y_{ijk} of the plot ijk (i = block number, j = treatment number, k = variety number) may be expressed in usual notations by the model

$$Y_{ijk} = \mu + b_i + t_j + v_k + (vt)_{kj} + \epsilon_{ijk}$$

Explain the nature of the terms, involved in the model and the relationship between them, if any. Exhibit the structure of the analysis of variance with expectation of the mean squares specifying the conditions under which the 'interaction' or the 'error' sum of squares should be used in the test of significance for varietal differences.

6. Bring out the essential features of the analysis of covariance technique and its practical usefulness in agronomic experiments.

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

(a) Latin square design with 1 missing plot,

(b) Split-plot design,

(c) Perennial crop experiment,

(d) Additivity of the sum of squares in Analysis of Variance.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

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

Time : 4 Hours Full Marks : 100

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

(b) All questions carry equal marks.

1. Show that in a symmetrical factorial design involving four factors each at s levels and in which each replication is laid out in blocks of s^2-plots, a complete balance over the three-factor interactions can be achieved in $(s-1)^2$ replications.

2. Give an account of the principles of fractional replication of a 2^n factorial experiment. What is the smallest fraction to which one can go if the main effects and two factor interactions are not to be confounded ?

3. Describe a method of constructing the hypercube (m, s, t, d). Illustrate the method for the hypercube (4,3,3,3).

4. Give a method of constructing the series of balanced incomplete block designs having the parameters :—

$$v = 6t+1, \quad b = t(6t+1), \quad r = 4t, \quad k = 4, \quad \lambda = 2,$$ where $6t+1$ is a prime or power of a prime. Illustrate the method for the case $t = 3$.

5. Show that affinine resolvable incomplete block designs for which $b = v+r-1$ are always balanced.

6. Explain the parameters of the first and second kind of p.b.i.b. designs and prove the relationships among them. Show that from every p.b.i.b. design having $k>2$, p.b.i.b. design with $k = 2$ and $r<(v-1)$ can be constructed.

7. Show that for every two-associate p.b.i.b. design for which $r<k$, the following condition should hold good :—

$$(r-\lambda_1)(r-\lambda_2)+(\lambda_1-\lambda_2)\{(r-\lambda_1)p_{1s}^1 - (r-\lambda_2)p_{1s}^1\} = 0$$

Verify this in the case of the simple square lattice.

8. Define the *group divisible* type of p.b.i.b. designs having two associate classes. Show that there are three sub-types among them satisfying one or the other of the following conditions :—

(a) $r = \lambda_1 ; \; rk - v\lambda_2 > 0$

(b) $r > \lambda_1 ; \; rk - v\lambda_2 = 0$

(c) $r > \lambda_1 ; \; rk - v\lambda_2 > 0$

What is the connection between this design and the confounded design for the asymmetrical factorial experiment involving two factors ?

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPERS VIII AND IX : ECONOMIC STATISTICS (PRACTICAL)

Time : 4 Hours Full Marks : 100

N.B. (a) Attempt Question No. 1 and ANY TWO of the rest.
(b) Use of calculating machines is permitted.

1. Average monthly expenditure per household in 1951 on food articles in the rural areas of North India and the prices of these articles in 1951 and in September 1954 are given below :

Commodities	Exp. per household in 1951 (Rs.)	Prices per Seer in Rs.	
		1951	September 1954
Food grain	53.27	.40	.45
Pulses	7.11	.43	.38
Edible Oil	4.07	2.12	1.50
Vegetables	2.19	.44	.42
Milk	4.80	.75	.62
Meat and fish	1.41	1.50	1.62
Fruits	0.34	1.60	1.50
Salt	0.56	.07	.07
Species	2.23	2.40	2.00
Sugar	4.46	.70	.75

Calculate a suitable index showing the change in food price in September 1954 compared to 1951.

2. Calculate the seasonal index and comment on the seasonal variation, if there is any, for the following series showng monthly raisings of coal (in lakh tons) in India during 1946–52.

Year	1946	1947	1948	1949	1950	1951	1952
Month							
January	15.5	18.4	18.8	16.3	21.7	20.1	21.0
February	17.3	20.0	21.0	18.9	22.8	21.7	23.6
March	18.0	20.0	18.8	19.2	22.4	21.1	22.5
April	17.0	18.1	18.5	18.6	22.4	20.5	21.4
May	18.0	18.2	18.2	17.6	20.5	20.4	21.2
June	16.4	16.3	16.6	17.5	19.7	19.0	19.8
July	14.8	15.6	14.7	17.9	18.6	18.9	20.9
August	15.3	13.9	15.5	17.0	20.2	20.1	20.5
September	17.2	16.3	17.6	21.3	23.7	23.9	23.1
October	17.9	16.6	16.4	19.4	18.4	20.1	22.4
November	15.7	17.1	12.8	16.8	19.4	18.3	21.4
December	18.0	18.1	16.2	21.8	22.3	21.4	22.9

3. Distribution of persons by per capita monthly income and percentage consumption of cereals in different groups are given below for rural and urban areas of West Bengal.

Per Capita income (Rs.)	Number of persons (thousand)		Consumption of cereals (p.c.)	
	Rural	Urban	Rural	Urban
1— 7	463	61	2.64	0.86
8—10	915	38	5.22	0.29
11—13	1098	151	6.26	2.08
14—16	2320	387	13.23	5.05
17—19	1668	475	9.51	5.43
20—24	2320	912	13.28	9.47
25—29	2855	1200	16.28	15.71
30—39	2996	802	17.09	12.61
40—49	1366	698	7.79	10.85
50 and above	1526	1022	8.70	37.65

Draw Lorenz curve showing the concentration of consumption of cereals and study the difference between rural and urban areas, if there is any.

4. The following table gives per capita annual expenditure in 1955-56 (as estimated) and the income elasticities of various items in urban India.

Item	Per Capita Exp. (Rs.)	Income elasticity
Food grain	82.28	0.52
Pulses	12.84	0.74
Vegetables	14.02	0.75
Milk and milk products	36.00	1.35
Meat, egg and fish	12.84	1.02
Sugar	11.32	0.03
Tobacco	7.72	0.77
Clothing	24.44	0.89
Footwear	2.56	1.09
Fuel and light	25.72	0.73
Education	9.12	1.63
Medicine	11.04	1.52
Amusement	4.08	1.24
Conveyance	9.60	1.20
Ceremonials	23.02	1.49

Assuming that income will rise by 5 per cent in each year of the period 1956-57 to 1960-61, calculate the total expenditures on the above items in these five years in the urban area. The urban population in 1956-57 may be taken as 6.75 crores and a rise of 1.50 per cent per year may be assumed.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPERS VIII AND IX : VITAL STATISTICS AND POPULATION STUDIES (PRACTICAL)

Time : 4 Hours — Full Marks : 100

N.B. (a) Attempt ANY TWO questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.
(d) Census of India, Paper No. 2, 1954, Life Tables, 1951 Census will be provided for use at the examination.

1. The Wage distribution of the Indian female population according to the 1951 Census (corrected for enumeration bias) is given below. Assuming the Life Table of 1941–50 and an annual number of 7 million female births to operate continually, there being no migration, calculate the female population in 1981 below 100 years of age by quinquennial age groups.

Age	No. in 00's	Age	No. in 00's
0— 4	237,066	40—44	89,264
5— 9	223,504	45—49	74,887
10—14	195,610	50—54	61,161
15—19	174,229	55—59	48,477
20—24	157,670	60—64	36,630
25—29	141,624	65—69	24,767
30—34	124,200	70	3,689
35—39	105,865	Residual 70	36,477
			1,735,120

2. Using appropriate formulae for interpolation, work out from the data given under Question 1, the Indian female population at individual ages 0, 2 etc., to 9 in 1951.

3. A group of 382 married women whose age at survey was 33 years or below had given birth to 876 live children. These women are classified in Table I according to their ages at survey and age at marriage. The births are classified in Table II according to the age of the mother at the time of birth.

The period of exposure during which a live birth could occur is given by the following scheme for each age starting from the age at marriage denoted by x to the age at survey. (It is assumed that a live birth cannot occur within 9 months of marriage) :—

Exposure at age x is zero for women aged x at survey and $\frac{1}{2}$ years for women aged $x+1$ or more at survey

Exposure at age $x+1$ is $\frac{111}{384}$ years for women aged $x+1$ at survey and $\frac{23}{32}$ years for women aged $x+2$ or more at survey.

Exposure at age $x+i$ is $\frac{1}{2}$ year for women aged $x+i$ at survey and 1 year for women aged $x+i+1$ at survey, for $i = 2,3$, etc.

Show how the risk of having a live birth varies with the age of the women by calculating the number of live birth for 100 years of exposure for each age between 17 and 33 years.

TABLE I : FREQUENCY TABLE OF WOMEN'S AGE AT SURVEY AND AGE AT MARRIAGE

Marriage at age	Age at survey																	Total
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	
16	4	3	10	11	11	11	9	5	10	11	8	7	14	5	3	5	17	10	154
17		2	6	4	13	10	7	9	5	9	7	7	6	6	7	1	3	2	104
18			8	14	12	14	16	7	6	6	7	6	6	5	10	4	0	3	124
Total	4	5	24	29	36	35	32	21	21	26	22	20	26	16	20	10	20	15	382

TABLE II : NUMBER OF LIVE BIRTHS ACCORDING TO THE AGE OF THE MOTHER AT BIRTH OF CHILD

Age of mother	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	Total
No. of Live Births	56	95	128	109	76	70	49	53	34	31	24	21	16	10	8	4	2	786

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPERS VIII AND IX : STATISTICAL QUALITY CONTROL (PRACTICAL)

Time : 4 Hours Full Marks : 100

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

1. The following table gives the averages and ranges of sub-groups of size 4 of test records of copper content in commercial brass sheets.

Subgroup No.	Average	Range	Subgroup No.	Average	Range
1	11.10	0.6	16	11.45	1.3
2	11.70	1.2	17	11.55	1.6
3	11.35	1.0	18	9.98	0.9
4	11.25	1.0	19	10.78	1.2
5	11.40	2.0	20	11.23	0.7
6	11.00	0.6	21	10.93	1.7
7	11.20	1.0	22	11.50	2.7
8	11.35	1.2	23	10.78	0.7
9	11.50	2.0	24	10.05	1.1
10	10.88	1.1	25	11.48	2.9
11	10.85	1.0	26	10.80	0.4
12	11.53	1.2	27	12.20	2.0
13	11.15	0.8	28	11.88	1.5
14	11.28	1.0	29	11.23	0.8
15	11.00	0.8	30	11.30	0.6

Test if the process is under statistical control.

A minimum of 9 per cent in any sheet is the market specification of this charactoristic. Excess of 0.1 per cent results in a loss of Rs. 8000/- per annum to the factory. Estimate how much saving can be effected by maintaining statistical control at a proper level so as to satisfy the market specification.

2. A brand of bricks is statistically controlled with average crushing strength at 28 tons and S.D. as 4 tons. A consumer is willing to run a risk of 5 per cent of receiving lots of coefficient of variation, $V_0 > 0.2$. The producer wishes to run no greater risk than 1 per cent of having lots rejected. How many bricks should be tested and what is the coefficient of variation v which divides the acceptable from unacceptable lots ?

3. A consumer receives batches containing 10,000 components each. Generally 9 out of 10 batches are free from defects, and 1 out of 10 batches contains 20 per cent defectives. If 20 per cent defective batch is accepted, the loss incurred by putting the defective components on production line is Rupees 3000 for 10 batches. If the inspection cost is half anna per item, find the number of items inspected per batch to minimise the total of inspection costs and consumer costs. It is assumed that acceptance number is zero.

4. Observations were taken simultaneously on 3 machines expected to give similar production. The following table gives the averages and ranges for samples of size 4.

Sample No.	Macine I		Machine II		Machine III	
	X	R	X	R	X	R
1	51.5	7.5	45.0	5.0	45.7	3.0
2	46.6	12.5	48.1	9.0	47.5	4.0
3	46.6	6.5	43.4	7.0	44.4	5.5
4	44.1	3.0	44.1	7.0	46.4	4.5
5	43.6	8.5	42.7	6.0	42.7	4.0
6	45.7	5.5	43.1	1.5	43.4	6.0
7	45.4	5.5	41.2	3.5	42.0	5.0
8	44.0	3.5	41.6	4.5	44.0	5.5
9	41.8	4.5	41.2	3.5	44.6	3.0
10	43.7	5.5	41.2	3.5	46.0	2.0
11	43.8	4.0	42.8	5.5	42.0	7.5
12	42.6	7.5	42.6	5.0	41.0	6.0
13	40.1	4.0	39.4	2.0	40.4	2.5
14	41.1	5.5	40.5	6.5	41.2	5.5
15	41.4	4.5	41.9	3.5	40.1	3.5
16	42.1	4.0	41.6	3.0	45.6	3.5
17	42.9	4.0	42.7	6.0	44.7	3.5
18	41.5	5.0	41.0	6.5	40.5	1.5
19	44.1	5.0	48.9	3.5	45.4	5.0
20	46.9	8.5	47.9	6.5	47.1	6.0
21	45.9	4.5	47.9	6.5	47.1	6.0
22	44.9	6.9	46.9	2.5	46.4	5.0

The sums and sums of squares for all the 88 observations on each machine are given below :—

	Σx	Σx^2
M/c I	3885.0	172520.0
M/c II	3808.5	165743.25
M/c III	3863.0	170295.75

Analyse the data by (1) control chart method and (2) Analysis of variance methods and discuss fully the conclusions arrived at in both cases.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

Papers VIII and IX : Mathematical Theory of Sampling Distributions (Practical)

Time : 4 Hours — Full Marks : 100

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

1. Take 15 random samples of 5 observations (x) each from a normal population with mean $\mu = 1$ and s.d. $\sigma = 1$. Calculate for each sample the statistic

$$t' = \frac{2\bar{x}}{s}$$

where $\bar{x}$ and s are the mean and standard deviation.

Compare the mean value of t' with its theoretical value given by

$$\frac{\Gamma\left(\frac{p-1}{2}\right)\sqrt{p}}{\Gamma\left(\frac{p}{2}\right)\sqrt{2}}\frac{\mu}{\sigma},$$

p in the present case being 4.

2. Using a table of random numbers and Biometrika Table Vol. I (Table 7), obtain a sample of 20 from a x^2 population defined by

$$p(x^2) = \frac{e^{-\frac{x^2}{2}}(x^2)^3}{2^6\Gamma(6)}$$

Find how many in the sample exceed the 20 per cent point on this distribution.

3. Fifteen observations on (x,y,z) are given below:

x	y	z
40	11	14
34	13	30
39	25	33
40	22	37
36	13	27
30	18	19
37	14	25
38	24	31
24	16	26
34	20	30
41	32	35
35	16	23
36	18	21
36	13	30
41	33	39

Under the Hypothesis $\mu_x = 35$, $\mu_y = 15$, $\mu_z = 25$,
calculate Hotellings' T.
Find the probability of obtaining a value of T^2 equal to or greater than this observed value.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

Papers VIII and IX : Probit Analysis (Practical)

Time : 4 Hours Full Marks : 100

N.B. (a) Attempt Question 1 and either Question 2 or Question 3.
(b) Figures in the margin indicate full marks.
(c) Use of calculating machines is permitted.
(d) Special credit will be given for neat graphs and tabular arrangement of computations.

1. With the logarithm (base 10) of the concentration in gm/cc as dose metameter, the regression equation of the probit (Y) of the percentage killed on the dose metameter (X) for two insecticides A and B are as follows:—

$$A: \quad Y = 4.2135 + 0.7256X$$
$$B: \quad Y = 4.5784 + 0.7256X$$

The actions of the two insecticides are similar in the sense that in a mixture of the two, one acts merely as a dilution of the other.

Draw a graph from which one can read the percentage of insects killed when a dilution of λ gms/c.c. of a mixture of A and B in the ratio $1 : r$ is used for value of λ and r in the ranges $\lambda = 1$ to 25 and $r = 1$ to 10. (40)

2. The following table gives the results of a biological assay to examine the effects of an insecticide at different levels of concentration. Using the logarithm of the concentration as dose metameter estimate the parameters of the tolerance distribution. (60)

What concentration of this insecticide would you prescribe to ensure an average kill of 90 per cent in a large batch of insects with a natural morality rate of 20 per cent?

Concentration gm/litre	Number of insects	
	Tested	Killed
(1)	(2)	(3)
Control	100	12
3.03	100	27
5.47	100	33
8.65	100	49
13.30	100	63
21.74	100	81
43.73	100	89

3. To examine the nature of the fall off with distance in the blast effect of explosive charges an experiment was carried out as follows :—

A given weight of charge was detonated at a certain distance from a standard disk of card board and a record made of whether the disk was perforated or not. This was done 16 times at each of the five distances shown in the first column of the table below and the second column shows the number of disks perforated. The same experiment was repeated with another charge (of the same weight) and the number of disks perforated out of 16 is given in the third column.

Use probit analysis techniques to compare the blast effects of the two charges. (60)

Distance (ft.)	Number of disks perforated out of 16	
	Charge I	Charge II
(1)	(2)	(3)
50	2	0
45	4	4
40	10	5
35	10	9
30	14	10

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPERS VIII AND IX : ANTHROPOMETRY (PRACTICAL)

Time : 4 Hours Full Marks : 100

N.B. (a) Attempt FIVE questions, question 10 being compulsory.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.

1. Draw and describe any two of the following instruments :—
Goniometer for Attachment, Diatograph, Osteometric Board, Palatometer, Sliding Calipers.

2. Define the following land marks :—
Inion, Alare, Euryon, Mastoidale and Vregma.

3. Measure the following (the subject to be provided) :—
Auricular Height, Physiognomic Facial Breadth, Nasal Height, Interocular Bredth and Height Tragus.

4. Describe the Frankfort Plane and indicate the advantages of using the same in craniometric study.

5. Find out the following on the skull (to be supplied) :—
Metopic Angle and Alveolar Profilile Angle.

6. Find out the following indices on the living (subject to be provided) : Radio-humeral, Orbito-nasal, broadth-height, and nasal.

7. Measure the following on the skull (to be supplied) :—
Transverse Cranial Arc, Basio-Bregmatic height and Height of ramus.

8. It is said that man's stature is variable during the day and that the range is half an inch, the maximum stature being observable when the man gets up in the morning and the minimum when he retires to bed. How would you study the variability ?

9. Compare the Rollet and Manouvriar formula for determination of lengths of limb bodes to the total stature of the same individual when living.

10. Two series of Rajputs measured in U.P. have the following mean values :

A	Mean	Standard Deviation
Max. Head Length (151)	193·09	5·72
Max. Head Breadth (151)	138·09	3·87
Max. Frontal Breadth (151)	105·85	3·01
B		
Max. Head Length (139)	192·58	6·63
Max. Head Breadth (139)	138·72	5·23
Max. Frontal Breadth (139)	103·98	3·97

The pooled correlation matrix of the three measurements is

	H.L.	H.B.	F.B.
H.L.	—	.543	.325
H.B.		—	416
F.B.			—

The number of individuals in series A is 151 and in B 139.

(a) Set out all the computations needed to compute the Mahalanobis D^2-statistic.

(b) Examine whether the observed value of D^2 indicates (by a test of significance) a real difference between the two series of Rajputs.

(c) Calculate the percentage error committed for discriminating an observed individual as belonging to series A or B.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

Papers VIII and IX : Actuarial Statistics (Practical)

Time : 4 Hours Full Marks : 100

N.B. (a) Attempt Question No. 1 and ANY TWO of the remaining questions.

(b) Figures in the margin indicate full marks.

(c) Use of calculating machines is permitted. (d) Census of India Paper No. 2 of 1954 (Life Tables) will be supplied for use at the examination.

1. Trial values of $p_{20:40}$ are required for a proposed scheme of industrial insurance on country-wide scale in India, on Census 1941–50 life table all-India males mortality basis, alternatively (i) without any rating up and (ii) with lives rated up by 5 years, and uniform 2.995 per cent per annum interest base.

Construct the necessary commutation columns and compute the required values of $p_{20:40}$. Also compute the bonus loading necessary in the level premium to provide for a uniform reversionary bonus of Rs. 10/- per Rs. 1000/- assured. (50)

2. A corporation wishes to raise a loan of Rs· 1,00,000/- on 1st September 1955 repayable in 10 years in equal half-yearly instalments covering both principal and interest, the first instalment to be paid on 1st April 1956.

Calculate the amount of the instalment on the basis of 5.99 per cent nominal rate of interest per annum with half-yearly rests and break up the first ten instalments in principal and interest components. (25)

3. On the basis of Census 1941–50 life table, all-India males, evaluate :—

(i) the average future working life time of an Indian male at age 20 on the assumptions that he enters work at age 20, continues work till 2 months before death if death occurs before age 50 and 4 months before death if death occurs between ages 50 and 60, finally retiring at age 60 if alive.

(ii) the forces of mortality at ages 30 and 50. (25)

4. A loan of Rs. 1,00,000/- repayable at 105 per cent by five level annual payments of principal of Rs. 21,000/- each, the first falling due exactly three years hence, carries interest at the rate of 5.99 per cent per annum payable half-yearly.

If the whole loan is bought now, just after payment of a half-yearly interest instalment, for Rs. 1,04,000/- and held till final redemption, what effective rate of interest will a purchaser earn on the whole transaction? (25)

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPERS VIII AND IX : STATISTICAL MECHANICS (PRACTICAL)

Time : 4 Hours Full Marks : 100

N.B. (a) Attempt ALL questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.

1. The equation of state for polytropic gas is given by

$$pv^{\gamma} = \text{a constant},$$

where p = pressure, v = volume, γ = a suitable constant.

Find the best estimate of γ from the following data :—

p (kg/cm)	0.5	1.0	1.5	2.0	2.5	3.0
v (litres)	1.62	1.00	0.75	0.62	0.52	0.46

2. The density d of air at 0° C and 760 m.m. Hg. is measured by finding the weight m grams of the air in a container having the volume $V cm^3$ at t° C. and p m.m. Hg :

$$d = \frac{m}{V}\,\frac{760}{p}\left(1 + \frac{t}{273}\right).$$

Find the value, the dispersion (i.e., standard error) and the relative dispersion of d, when

$m = 2.4875$ gm. $V = 980.3$ cm^3, $p = 741.5$ mm. Hg.

$\sigma_m = 0.0023$ gm. $\sigma_v = 1.3$ cm^3, $\sigma_p = 0.9$ mm Hg.

$t = 21.4$°C

$\sigma_t = 0.3$°C

(where σ stands for the standard error of the corresponding quantity).

3. The diameter D of molecules of a gas is given by

$$D = \frac{(3kmT)^{\frac{1}{2}}}{(4\pi\eta)^{\frac{1}{2}}}$$

where k = Boltymann's constant, m = molecular mass, T = temperature, η = co-efficient of viscosity. If it is desired to calculate D to 1 per cent, compute the allowable errors in k, m, T η.

4. The molal heat c_p (in calmot: at 1 atmes) of copper are given for diffeont temporatures as

T =	20°k	30°k]	40°k	50°k	60°k	70°k	80°k
c_p =	0.1	0.4	0.8	1.4	1.04	2.5	3.0

Using Planck's formulation of Norust postulate, calculate approximately the absolute entropy of copper at 300°k.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

Papers VIII and IX: Sample Surveys—Theory (Practical)

Time : 4 Hours — Full Marks : 100

N.B. (a) Answer either Question 1 or Question 2.
(b) Use of calculating machines is permitted.

1. Table 1

Sl. No. of village	Total culti-vated area	Area under Wheat		Sl. No. of village	Total culti-vated area	Area under wheat	
	1931	1936	1937		1931	1936	1937
1	401	75	52	18	186	45	27
2	634	163	149	19	1767	564	515
3	1194	326	289	20	604	238	249
4	1770	442	381	21	701	92	85
5	1060	254	278	22	524	247	221
6	827	125	111	23	571	134	133
7	1737	559	634	24	962	131	144
8	1060	254	278	25	407	129	103
9	360	101	112	26	715	192	179
10	946	359	355	27	845	663	330
11	470	109	99	28	1016	236	219
12	1625	481	498	29	184	73	62
13	827	125	111	30	282	62	79
14	96	5	6	31	194	71	60
15	1304	427	399	32	439	137	100
16	377	78	79	33	854	196	141
17	259	78	105	34	824	255	265

Table I above shows the total cultivated area in 1931 as also area under wheat in two consecutive years 1936 and 1937 for a sample of 34 villages in Lucknow sub-division. The villages were selected with replacement with probability proportional to cultivated area (1931). The total cultivated area in 1931 and the total area under wheat in 1936 for all the 170 villages in this sub-division were known to be 78,019 and 21,288 acres respectively. Estimate the area under wheat in 1937 by three suitable methods and calculate their respective standard errors. Comment on their relative efficiency.

2. (a) You are provided with the 1931 census handbooks of four districts. Draw a random sample of 30 *villages* from the region constituted by the three largest (as regards population) districts. Make the best use of this selection if you are now asked to change over to a selection of the villages from the entire covered by all the four given districts. (Try to use as labour saving a method as you can think of; and put down clearly with full details the various steps you take for this purpose Do NOT write anything on the Census Handbooks).

(b) The following table shows the distribution of the No. of saw-mills in various production size groups in the year 1942. The total production and the standard deviation of the production within each group are also shown.

Stratum No.	Production Class (M. 6d. ft.)	No. of Mills	Total Productions (M. 6d. ft.)	Standard Deviation
1	5000 and over	538	5,934,000	9,023
2	1000—4999	4,756	8,464,000	1,159
3	Under 1000	30,964	6,311,000	300

Find the optimum allocation of a sample of 2,500 mills for estimating the total production of all the mills. Calculate also the sampling variance for optimum allocation.

Also, with reference to the above data examine by means of suitable numerical calculations as to how far it is true that the use of approximate values of the standard deviations (as is common in practice) in arriving at the 'optimum' allocation is not likely to substantially increase the sampling variance over that attainable if the true deviations were really known.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

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

N.B. (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.

Time : 4 Hours Full Marks : 100

1. Suppose that you have to conduct a sample survey of the conditions of living and study of the college students in a large city of India. A ten per cent random sample of the students from each college in that city has to be drawn.

(a) Write down suitable instructions for the enumerators explaining the use of 'random sampling numbers' for selecting a sample of students from each college register.

(b) Draw up a questionnaire for this enquiry.

(The questionnaire and the instructions are to be given in English).

2. With the help of 'random sampling numbers' draw random samples, each of size 5, from the following four theoretical populations :—

(a) univariate population with the following probability density function (Cauchy's distribution) :

$$f(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x-\mu)^2}$$

where the variate x ranges from $-\infty$ to $+\infty$, and the parameters $\lambda = 3.8$ inches and $\mu = -2.1$ inches.

(b) A bivariate normal population in which the means of the two variates, x and y, are respectively 68 inches and 170 pounds, the standard deviations of x and y are respectively 3 inches and 7 pounds, and the correlation coefficient (between x and y) p is zero.

(c) Same as the population in (b), excepting the value of p, which is $+1$.

(d) Same as the population in (b), excepting the value of p, which is -1.

3. (a) With the help of 'random sampling numbers' construct a sequence of length 100 out of the three letters, A, B and C occurring with probabilities proportional to 3, 3 and 2 respectively. (By 'length' of a sequence is meant the number of times the letters occur in it, e.g., BACCABACBC has a length = 10).

(b) Make a statistical test comparing the observed frequencies of occurrence of A, B and C with their expected frequencies in the sequence constructed by you in (a) above.

(c) Break up the sequence constructed by you in (a) above into 80 pairs of adjacent letters. (For example, BACCABACBC is to be regarded as BA, CC, AB, AC, BC).

Make a statistical test comparing the observed frequencies of occurrence of the nine possible combinations (AA, AB, AC, BA, BB, etc.) with their expected frequencies.

4. Let us denote by N_i, n_i and σ_i the population size, the sample size and the standard deviation respectively in the i-th stratum for a stratified sample; the total population size, ΣN_i is denoted by N, and the total sample size, Σn_i, is denoted by n. The sampling variance V of the final estimate, and W, the total cost incurred in the sampling survey, are given by

$$V = \frac{1}{N^2} \Sigma \left(N_i^2 \sigma_i^2 / n_i \right)$$

and $W = a + \Sigma b_i n_i$.

The numerical values of the constants are given below :—

Stratum No. i	N_i	σ_i	b_i
		Kilograms	Rs. As.
1	37800	28.5	3 8
2	52600	18.6	2 12
3	82000	27.6	2 4
4	41600	21.2	3 0
5	28800	16.8	2 8

a = Rs. 500, W = Rs. 10000

(a) Find out the optimum values of n_i's and the corresponding value of V.

(b) With the same total sample-size n as in (a) above work out the optimum values of n_i and the corresponding V_i assuming that b_i has the same value in all the five strata.

(c) With the same total sample-size n as in (a) above work out the value of V if the n_i's are taken to be proportional to N_i's only.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPERS VIII AND IX : DESIGN OF EXPERIMENTS—APPLIED (PRACTICAL)

Time : 4 Hours — Full Marks : 100

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

1. The data in the table below are grain and straw yields for 8 mannurial treatments and 8 replications of each (the grain is shown in brackets).

Treatments	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
1	242(620)	321(646)	261(681)	317(644)	255(706)	331(615)	218(552)	295(726)
2	267(644)	382(745)	201(542)	316(711)	280(705)	285(637)	200(543)	309(646)
3	215(523)	330(713)	298(686)	381(688)	300(692)	294(612)	256(635)	284(748)
4	212(601)	292(693)	265(685)	255(714)	238(699)	300(697)	283(701)	324(746)
5	322(664)	370(693)	284(666)	323(516)	232(656)	393(663)	351(657)	363(693)
6	200(514)	261(637)	259(697)	361(710)	234(633)	258(595)	306(697)	376(712)
7	260(550)	318(708)	266(663)	340(673)	362(671)	400(626)	276(655)	385(671)
8	203(521)	275(661)	207(594)	331(730)	220(625)	266(644)	276(745)	328(747)

Test the significance of the gain yields adjusted for the straw weight. From the results obtained and from other characteristics of the data decide as to the suitability of straw yield for the adjustment of grain yields.

2. The table below gives the yields of oats in bushels per acre in a simple lattice experiment, with 25 varieties and 2 replications.

Replicate 1	Variety	11	12	13	14	15	21	22	23	24	25
	Yield	50.8	88.2	50.3	73.0	49.2	66.5	41.6	76.6	45.3	83.8
	Variety	31	32	33	34	35	41	42	43	44	45
	Yield	69.4	71.0	59.1	39.7	64.2	82.8	50.6	67.3	51.3	82.1
	Variety	51	52	53	54	55					
	Yield	68.1	72.3	65.7	46.7	79.7					
Replicate 2	Variety	11	21	31	41	51	12	22	32	42	52
	Yield	55.4	88.4	75.4	94.8	52.1	67.7	50.6	68.1	48.5	70.5
	Variety	13	23	33	43	53	14	24	34	44	54
	Yield	60.9	66.8	59.5	54.3	73.3	58.1	60.0	49.4	79.0	46.3
	Variety	15	25	35	45	55					
	Yield	49.6	62.7	72.2	52.7	69.7					

Analyse the data and find the correction to the variety totals.

3. The table below gives the yields of 6 varieties in 4 replications with 1 missing value :—

Block	Treatments					
	1	2	3	4	5	6
1	18.5	15.7	16.2	14.1	13.0	13.6
2	11.7	—	12.9	14.4	16.9	12.5
3	15.4	16.6	15.5	20.3	18.4	21.6
4	16.5	18.6	12.7	15.7	16.5	18.6

Analyse the data taking account of the non-orthogonality and compare the treatments 1 and 2, and 5 and 6.

4. The table below gives the yields of mangels (in pounds of dry matter per plot) for a partially confounded 2 factorial experiment on 32 plots as shown in the table :—

Replicate 1				Replicate 2			
ab	24.0	b	26.2	(1)	27.5	ab	27.0
abc	29.6	a	28.2	abc	25.9	c	30.8
(1)	30.0	bc	29.1	ac	27.4	bc	32.8
c	30.5	ac	25.2	b	29.6	a	33.2

Replicate 3				Replicate 4			
abc	30.2	b	31.0	(1)	31.7	a	33.5
a	32.4	ab	34.4	bc	29.8	abc	33.2
bc	31.9	c	35.2	ac	32.5	b	35.6
(1)	31.6	ac	34.2	ab	29.2	c	32.4

Make an analysis of the results and obtain treatment means corrected for block effect.

5. The data given below is from a $3 \times 3 \times 2$ design of the factors A, B, C where A represents fertilisers applied in the row of tung trees at rates 0 lbs. 200 lbs. and 400. lbs, B represents no meal, tung meal (650 lbs.) and cotton seed meal (650 lbs.), C represents side dressing fertiliser with two alternatives, none and 200 lbs. per acre respectively.

The data shown (within brackets) are mean heights in centimeters of 12 tung trees of on each plot. Each replication contains 3 blocks of 6 units as shown :—

Replication 1

100(80)	200(78)	000(38)
210(86)	010(55)	110(73)
020(70)	120(82)	220(75)
201(74)	001(67)	101(78)
011(82)	111(67)	211(51)
121(86)	221(57)	021(66)

Replication 2

200(136)	000(43)	100(89)
010(56)	110(81)	210(87)
120(64)	220(90)	020(68)
101(95)	201(81)	001(91)
211(76)	011(61)	111(94)

Replication 3

100(86)	200(73)	100(66)
010(79)	110(76)	210(85)
220(73)	020(97)	120(101)
201(97)	001(116)	101(117)
111(79)	211(86)	011(106)
021(113)	121(81)	221(102)

Replication 4

200(88)	000(53)	100(81)
110(107)	210(66)	010(58)
020(70)	120(92)	220(56)
101(79)	201(88)	001(90)
011(92)	111(109)	211(68)
221(96)	021(95)	121(67)

The object of the experiment was to discover the type of fertiliser that would stimulate early growth so that the young trees would be ready for transplanting to commercial orchards by early autumn.

Now analyse and interpart the data.

STATISTICIAN'S DIPLOMA EXAMINATION, AUGUST 1955

PAPER VIII AND IX : DESIGN OF EXPERIMENTS—CONSTRUCTION OF DESIGNS (PRACTICAL)

Time : 4 Hours — Full Marks : 100

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

(b) All questions carry equal marks.

1. Using Galois fields, construct one complete set of orthogonal squares of side 8. Show that you can present not more than six of them with ordered directrices.

2. Construct one 9×9 quasi-Latin square for a $3 \times 3 \times 3$ factorial experiment confounding only three-factor interactions between the rows and the columns. Identify sets of degrees of freedom, which are confounded.

3. (a) Construct a 1/2-replication of a 2^6 experiment in which the six-factor interaction is sacrificed.

(b) Divide this 1/2-replication into 4 blocks of 8 plots confounding the following interactions :

ABC	=	DEF
ABD	=	CEF
CD	=	ABEF

Give the break-down of the individual degrees of freedom in the analysis of variance for the two cases taking interactions of higher order than two-factor as non-existant and hence identifying them as experimental error.

4. Construct the b.i.b. design $v = b = 21$, $r = k = 5$, $\lambda = 1$ and the corresponding Yonden quare.

5. Construct a p.b.i.b. design for $v = 15$, $r = 5$, $k = 3$, $b = 25$, $\lambda_1 = 0$, $\lambda_2 = 1$, $n_1 = 4$, $n_2 = 10$ and $p'_{12} = 0$.

6. Show that the following design involving 14 treatments A, B, ..., G, a, b, ..., g in blocks of two plots each, is a three-associate p.b.i.b. design, by deriving the values of all the parameters and verifying that they satisfy all the conditions for a p.b.i.b. design including the crucial one :

$$n_i p^i_{jk} = n_j p^j_{ik} = n_k p^k_{ij}$$

Design

Aa,	*Cb*,	*Ac*,	*Bd*,	*Ae*,	*Bf*,	*Eg*
Ba,	*Db*,	*Cc*,	*Dd*,	*De*,	*Cf*,	*Fg*
Ea,	*Eb*,	*Fc*,	*Fd*,	*Ge*,	*Gf*,	*Gg*