INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

Paper I : Theoretical Statistics (General).

me : 4 hours Full Marks—100

i. (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. Given a lot of size m containing s items of a specified kind. If items are to drawn without replacement until i of the s items have been drawn, show that on o average $\frac{i(m+1)}{(S+1)}$ drawings will be necessary.

In order to check a box of 144 screws, screws were drawn until 4 good screws re obtained. In a particular case only 20 drawings were necessary. Estimate e number of good screws in the box and find the standard error of the estimate.

2. In a series of independent trials, an even E has the constant probability p. hat is the probability of having a run of r E's in n trials ?

Find the mathematical expectation of $m-np$ where m is the number of ccesses in a series of n independent trials with constant probability of success p.

3. (i) The function $1/N$ is tabulated at unit intervals from 1 to 1000. Find the ssible error in linear interpolation of this function when $N=650$.

(ii) Using Everett's formula for interpolation, or otherwise, derive the quadra- ire formula :

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} y(x)dx = y_0 + \tfrac{1}{4}\delta^2 y_0 - \frac{17}{5760}\,\delta^4 y_0 + R.$$

4. (i) Show that the sum of two independent variates is normally distributed, thé two variables are normally distributed.

(ii) Writing

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt, \qquad x > 0$$

show that

$$1-F(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\left\{\frac{1}{x} + \frac{1}{x} + \frac{2}{x} + \ldots\right\}$$

GROUP B

5. (a) If u and v possess independent χ^2-distributions with n_1 and n_2 degrees of freedom respectively, obtain the distribution of

$$\frac{n_2}{n_1}\,\frac{u}{v}.$$

(b) Given two random samples of sizes n_1 and n_2 from two normal populations with the same variance, draw up the analysis of variance table for testing the equality of the means of the two populations. What is the relation between the F-statistic obtained in this case and the corresponding t-statistic?

6. (a) Given a $p \times q$ contingency tables for two variates, explain how, by restricting yourself to samples of a particular type, you will test the independence of the two variates. State the conditions, if any, under which your test would be valid.

(b) The entries in a $2 \times t$ contingency table are:

a_1	a_2	a_t	a
b_1	b_2	b_t	b
n_1	n_2	n_t	n

If $p_i = a_i/n_i$, $p = a/n$, $w_i = n_i/pq$, $q = b/n$,

Show that χ^2 for testing homogeneity is

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

7. (a) What are the advantages of Fisher's transformation of the correlation coefficient from r to Z?

(b) A number of correlation coefficients between two variates based on different sample sizes are available. How will you test whether these correlations have arisen from the same population? Give a scheme of computations for this purpose. If it is found that these correlations have come from the same population, what is the best estimate of the population correlation coefficient?

8. (a) What does the multiple correlation coefficient between y and $x_1, x_2, \ldots x_p$ measure? Explain how this coefficient may be used for prediction purposes. With the plausible assumption that there is a close association between the yield of paddy for a particular area and the meteorological factors over that area, indicate what records you would ask for and how you would proceed to construct the actual relationship.

(b) If $r_{13} = \gamma$, $r_{23} = -\gamma$, show that r_{12} will lie between -1 and $1 - 2\gamma^2$.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

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 three questions from each group.

GROUP A

1. It is desired to make a rapid assessment of the economic position of the Handloom industry in the country. Bring out the relative merits of a well-designed sample study in this context.

What are the problems that would arise at the various stages of such a survey and explain briefly how you will proceed to tackle them.

2. Discuss the consequences of Partition on the comparability over time of Indian Official Statistics with particular reference to foreign trade and industrial production.

If you are asked to determine the extent of seasonal variation in the output of the Indian Sugar industry, to what publications will you turn for obtaining the basic material for the purpose ? Do you think any preliminary scrutiny is necessary or desirable before the data are actually analysed ? If so, give details.

3. Name all the official series of index numbers now in existence and point out their limitations. In what way could the different series be improved ?

Describe in detail the CAPITAL index of industrial activity and suggest suitable modifications in its constitution with a view to its improvement.

4. What are the principal elements that go to make up the current and capital accounts of Balance of Payments; and how are they computed ?

Point out the gaps in the published estimates of India's balance of payments and make proposals, if any, for filling them in some order of priorities of your choice

5. Trace the historical development of research work in the field of correlation of economic time series in the United Kingdom, the United States of America, and India respectively with special reference to wheather conditions and crop yields.

GROUP B

6. Describe the logistic curve and discuss various methods of fitting it to population data. For what purposes is this curve usually fitted to such data.

7. Describe the important components of a Life Table and explain the uses of this table.

Define Expectation of Life and describe a method of calculating it from census data.

8. Describe various methods of determining reliability of test scores. Discuss the relative advantages and disadvantages of these methods.

9. Explain the theory of multiple-factor analysis in psychometry. Discuss in particular the single-factor methods.

10. Describe the major blood-groups and blood-types of man and explain their modes of inheritance and practical importance.

In a case of disputed parentage two babies were of type MN and N, respectively. Their mothers also were of types MN and N, but it was uncertain to which mother either baby belonged. The husband of woman MN was of type N; the husband of woman N was of type M. To which mother did the type N baby belong.

11. Describe the various statistically developed methods for increasing the accuracy of agricultural fields experiments.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPER III : STATISTICAL INFERENCE.

Time : 4 hours Full marks—100

N.B. (a) Attempt any 6 questions.
(b) All questions carry equal marks.

1. Describe the method of maximum likelihood for estimating unknown parameters from a given body of data. What is the justification for this method ?

Show that if $x_1, x_2, \ldots x_n$ is a sample of n independent observations from a Poisson distribution with mean λ, the maximum likelihood estimate of λ is $(x_1+x_2+\ldots+x_n)/n$

2. Let $x_1, x_2, \ldots x_n$ be n independent observations each having the density $f(x,\theta)\,dx \ldots -\infty<x<\infty$,

where θ is an unknown parameter, and let

$$L=\sum_{i=1}^{n} \log f(x_i, \theta.)$$

Show that if $t(x_1, x_2, \ldots x_n)$ is an unbiased estimate of θ, then

$$E\theta\ (t-\theta)^2 > 1/E\theta\left(\frac{\partial L}{\partial \theta}\right)^2$$

Use this result to show that in samples from a normal population the sample mean is the efficient estimate of the population mean.

3. Define a sufficient statistic. State whether or not the following statements are correct, and give reasons for your answer in each case :

(i) Every function of a sufficient statistic is also a sufficient statistic.

(ii) If T is a sufficient statistic, and T is a function of a statistic U, then U is a sufficient statistic.

(iii) A sufficient statistic contains all the available information concerning the unknown parameters.

4. Suppose that 100 crop cutting experiments were carried out in a certain district to determine the mean yield of paddy (in maunds per acre) for the district during the season in question. Let the (unknown) mean yield be denoted by μ. It is assumed that the sampling scheme of the survey was such that the 100 observations obtained could be regarded as a sample of independent observations from a normal distribution with mean μ and some unknown variance σ^2.

Suppose that you are supplied only with the following information concerning the results of the suvey : 30 observations were less than or equal to 10 maunds

per acre, 60 were greater than 10 but less than or equal to 20 maunds per acre, and 10 were greater than 20 maunds per acre. Describe how you would estimate μ without actually carrying out the computations.

5. Give a very brief account of the main ideas of the Neyman-Pearson theory of testing statistical hypotheses.

Let, $x_1, x_2, \ldots . x_n$ be a sample of n independent observations from a normal population with mean θ and variance σ^2, both θ and σ^2 being unknown. Let H_0 be the hypothesis that θ is zero. State without proof the optimum properties of (i) the one-sided t test of H_0, and (ii) the symmetrical two-sided t test of H_0. State which of these two tests you consider to be the more useful one, and give your reasons.

6. . Show that the maximum of a sample of n observations independently drawn from a rectangular population $(0, \theta)$ is a sufficient statistic for θ.

Construct an unbiased estimate with minimum variance for the parameter . Give also a consistent estimate of θ^2.

7. In an experiment with four different type of flours and three methods of manufacture, two characteristics (variates) of bread are measured. If for each combination of flour and method three samples are taken, how do you analyse the data to detect differences in flour types, methods and possibly interaction between the two.

If differences are noted what further considerations are necessary to select the best combination of flour and method, best in the sense of providing maximum satisfaction to the consumer ?

8. Describe how a discriminant function for classification is constructed in the case of multiple measurements. What are its uses ?

How is the discriminant function useful in tests of significance ?

9. Explain the concept and significance of intraclass correlation.

For k members in a family find the limits within which the intraclass correlation lies.

How do you test the hypothesis on the basis of computed intra-class correlation from a given sample whether the population value is zero.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

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. The following data relate to domestic demand for motor fuel (monthly average in thousands of barrel) in the U.S.A. Estimate the missing figures for 1925 and 1927. (10)

1922—10659, 1923—13062, 1924—15417, 1926—21818, 1928—27374.

2. The frequency distribution of life in hours of certain kind of electric bulbs is given below. Fit an appropriate Pearson's curve to this data and test the goodness of fit. Also draw the histogram and the fitted curve on a graph paper. (The appropriate formula for the constants will be supplied on request). (30)

Life in hours	frequency
501—540	44
541—580	189
581—620	161
621—660	73
661—700	24
701—740	7
741—780	2
	500

Note : Make use of the following data.

Mean = 590.34

μ_3 = 1909.57

μ_4 = 58071.68

μ_4 = 13864427.52

3. The numbers of live male and female births (in thousands) in England and Wales for the years 1910—1919 were as shown below : (15)

Year	male births (in thousands)	Female births (in thousands)
1910	457	440
1911	449	432
1912	445	428
1913	449	433
1914	447	432
1915	415	399
1916	402	383
1917	341	327
1918	339	324
1919	356	336

Do the above data support the conclusion that the sex-ratio remained the same throughout the period covered ?

4. (i) In a sample of 380 men from a certain city, 254 are found to be smokers. In another sample of 689 from a comparatively bigger city, 342 are smokers. Do

the data indicate that the cities are significantly different with respect to the prevalence of smoking. (4)

(ii) Sample readings for the rate of diffusion (in suitable units) of carbondioxide through two types of soils are given below. (5)

Through soil A : 20, 31, 18, 23, 23, 28, 20, 26, 27, 20, 12, 17.

Through soil B : 10, 30, 32, 28, 15, 26, 35, 18, 25, 27, 35, 34.

On the available evidence do you infer that the rates of diffusion for the two soil types are different ?

(iii) The following correlation coefficients between weight of ears of wheat as harvested and weight of grains of wheat after threshing and cleaning, were obtained from the results of a crop-cutting experiments in the district of Shahabad of Bihar, the district being divided into 6 sub-blocks. Investigate if the correlation differs from sub-block to sub-block. (8)

Sub-block No.	1	2	3	4	5	6
Number of samples	274	179	233	54	255	50
Correlation coefficient.	0.9055	0.8710	0.9210	0.9365	0.9508	0.9780

(iv) The following table gives the fundamental constants obtained from the measurements of ($n=450$) eggs. (8)

	correlation coefficient with			Mean	S.D.
	x_1	x_2	y		
length in m.m. (x_1)	1.0000	0.5751	0.5797	56.3222	2.3862
bulk in c.c. (x_2)	0.5751	1.0000	0.9804	51.8400	4.2438
weight in gm. (y)	0.5797	0.9804	1.0000	55.2400	4.5923

Note : Square of S.D. is equal to the sum of squares of deviations from the mean divided by n.

(a) Calculate the multiple correlation coefficient $Ry.x_1x_2$ and test its significance.

(b) Calculate $ryx_1.x_2$ and test its significance.

5. Table below gives the frequency distribution of marks in mathematics obtained by students at the Matriculate examination in three successive years. Assuming that the distributions of marks in the three years are normal with a common standard

deviation, test whether there is significant difference in the average marks of students in the three years. (20)

Range of marks	Frequency			
	1937	1938	1939	total
0—10	4	7	20	31
10—20	22	25	57	104
20—30	77	109	140	326
30—40	178	181	202	561
40—50	191	179	135	505
50—60	105	129	94	328
60—70	43	73	67	183
70—80	26	32	40	98
80—90	14	30	22	66
90–100	7	15	29	51
Total	667	780	806	2253

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

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.
(c) Credit will be given for neatness and proper arrangement of caulations.

1. Collect relevant statistics from official publications and write a short report on the present food situation in India. (15)

2. Using the data tabulated below compute an index-number of prices of agricultural commodities in 1948 with 1939 prices as base. (15)

Commodity	Unit	Quantity produced in 1939 (in million units)	Price per unit (in dollars)	
			1939	1948
(1)	(2)	(3)	(4.1)	(4.2)
Corn	Bu	2679	2.62	5.72
Cotton	Lb	5705	.71	2.20
Hay	Ton (sh)	76.59	40.3	85.65
Wheat	Bu	952.1	4.26	9.57
Oats	Bu	1107	1.40	3.25
Potatoes	Bu	297.3	3.16	6.92
Sugar	Lb	4371	.20	.53
Barley	Bu	131.1	2.43	4.12
Tobacco	Lb	1444	.78	2.51
Flax seed	Bu	6.77	8.76	17.52
Rye	Bu	78.7	2.66	4.18
Rice	Bu	42.69	5.33	9.83

3. Samples of four articles are taken and tested for quality every hour in a factory producing articles of specified quality. As past samples show, the production process was under control for a long time with deviations from specification having a mean +0.01 and standard deviation 0.06. A new series of observations are tabulated below :

Sample No.	Daviation from specification			
1	−1.59	−0.54	−0.21	−0.60
2	+0.32	+1.55	−0.74	+1.82
3	+0.27	−0.73	+0.58	−1.17
4	+0.08	−0.46	+0.70	+1.13
5	+1.74	−0.20	+1.44	+1.62
6	+2.35	+1.97	−0.78	−1.07
7	−0.67	+0.94	−0.06	+1.21
8	+0.22	−1.41	+0.08	+1.27
9	−0.48	+0.42	−0.53	−0.20
10	+0.84	−0.76	+0.88	+0.88
11	+0.31	+0.20	−0.40	+1.29
12	+1.09	−0.17	+0.22	+0.63
13	+0.22	−0.40	+1.29	+0.54
14	−0.21	+1.08	+0.48	+0.48
15	+2.17	+1.26	−0.28	+1.03

Apply the control-chart technique to analyse the material and report on the peculiarities of the production process.

4. In a two-stage sampling with replacement scheme the variance of the sample mean is (15)

$$V=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_1 n_2}$$

where n_1 is the number of first stage units and n_2 the number of second stage units per first stage unit. The cost of carrying out such a survey is

$$T=a+bn_1+cn_1n_2$$

The values of the constants σ_1, σ_2, a, b, c as estimated from a pilot asurvey are as follows:

$\sigma_1=6.2$, $\sigma_2=9.4$, $a=$Rs. 513, $b=$Rs. 6.4, $c=$Rs. 1.4.

Find the optimum value of n_1, n_2 when the cost of the survey is fixed at Rs. 2000.

5. The distribution of marks in a certain examination was found to be normal with 23% of the candidates scoring above 60 and 21% below 40. Find the mean and the standard deviation of the distribution of marks. Also compute the mean and the standard deviation of marks obtained by candidates scoring over 60. (15)

6. A fertilizer experiment involving three fertilizers N, P, K each at two levels was carried out in a latin-square arrangement. The layout and the yields are tabulated below. Analyse the data and write a report on your findings. (20)

YIELD OF WHEAT IN AN 8×8 LATIN-SQUARE FERTILIZER EXPERIMENT

P 18	N 12	NP 18	K 15	NK 11	O 11	NPK 19	PK 18
N 12	NK 7	PK 17	NPK 17	P 19	K 12	NP 19	O 15
NK 10	NP 17	N 10	P 18	O 9	NPK 16	PK 17	K 14
PK 18	K 12	NPK 14	O 12	N 11	NP 14	P 16	NK 16
NP 17	O 12	NK 13	N 11	PK 16	P 15	K 10	NPK 16
K 14	PK 18	O 12	NP 17	NPK 15	N 9	NK 8	P 20
NPK 19	P 18	K 11	PK 17	NP 17	NK 8	O 10	N 14
O 17	NPK 20	P 20	NK 16	K 16	PK 18	N 13	NP 23

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS IV & V (SPECIAL SUBJECT) : ECONOMIC STATISTICS (THEORETICAL).

Time : 4 hours Full marks—100

N.B. (a) Attempt any five questions.
(b) All questions carry equal marks.

1. In what respect and to what extent does the 1951 Census of India differ from the 1941 Census ? Argue the case for a quinquennial in place of the present decennial census.

2. Describe the main sources of Labour Statistics in India, mentioning some of the Acts under which they are collected. Are there any gaps and defects ?

3. Discuss briefly the problems, both theoretical and practical, that arise in connection with the construction of an index number of prices. What special difficulties have to be contended against during a period of price control and rationing ?

4. Obtain an equation for correcting an index number of cost of living for an income group other than the group earning the average income. What assumptions are to be made and how are the constants in the equation computed ?

5. (a) Why is the ordinary theory of errors not applicable to time series data ?

(b) How would you eliminate the "trend" from a time series when it is given (i) annually for fifty years and (ii) monthl for fivey years ?

(c) Describe the method of comparing two time series expressed in dis-similar units, say maunds and rupees, so far as their (i) long-period trend and (ii) short-period fluctuations are concerned.

6. How is a "Gallup Poll" taken ? Why do the predictions made by it sometimes prove false ?

7. Describe the importance of multi-purpose sample surveys at short intervals of a big country like India, mentioning the uses of necessary internal and external checks.

8. Describe the social accounting method of estimating national income. How far is it applicable to India ?

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

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

Time : 4 hours Full marks—100

N.B. Attempt question no. 1 and any three of the remaining questions.

1. (a) Describe the broad nature of preferences for particular digits in age records, experienced in the past censuses. What other factors are suspected to contribute to mis-statements of age in census returns ? How could the mis-statements be located, and what treatments could be applied to the original data to approximate to the real age distribution of the population in age-groups ? Make particular mention of any special features betrayed by the 15—25 age-group in the general run of population by age distribution, in 1911, 1921 and 1931 Indian censuses.

(b) Indicate (i) how the relative sex-proportions, in the three ranges of life, childhood, adultage and old age, were seen to vary in 1931 census record, and (ii) the probable reasons for the observed variation and their effect on future growth of population.

(c) What additional statistical information, besides the census count, is required to produce age-specific birth and death rates ? How far such statistics obtain in India ?

2. Explain what gross and net reproduction rates mean : What variants of the rates, to allow for sex compositions and marriage proportions, are current ? What effects changing patterns of marriage rates are likely to have on the emerging series of net reproduction rates ?

A statistician advances the view that for the purpose of determining population trends, analysis of fertility by 'cohorts' or by the size of average 'completed families' is more suitable. Discuss the view, and indicate what gaps still remain.

3. It is proposed to conduct an independent sample survey to estimate the specific fertility and mortality of the population in India in age-groups, for scientific forecasts of births deaths and future populations in age-groups.

Give your detailed comments on the proposal and list the items of information under the major events like births deaths and marriages that you should like to be collected for the purpose. How would you provide for the incidence of polygamy ?

4. Differential mortality and sickness experiences of (i) coalminers working underground and (ii) of others employed in coalmining industry including office and welfare workers, are proposed to be investigated separately in connection with a contributory benefit scheme.

(a) Discuss and suggest a definition of "sickness" that you would recommend. Do you anticipate that the character of the experience will change on inception of the proposed scheme ? Give your reasons.

(b) How far would you expect the differential mortality rates to reflect relative occuptional hazard ? Indicate the difficulty in comparing occupational mortality rates generally.

(c) Do you consider that standardised death-rates will be suitable for comparison of deaths by various causes of death ? What alternative method of presentation could you suggest ?

5. In a recent article in a technical journal, an economist estimated the adult population (ages 20 and above) at the end of the planning period in 1977 by assuming an average rate of population growth of 1.5% per annum, and applying it to 1951 census population aged 20 and above. He arrived at the growth rate adopted by arguing that the overall population increase during the decade 1941—51, as determined by the two census counts, was 13.4% ; this worked out to an average annual rate of increase of 1.34%, and a slightly higher rate was appropriate in view of improving public health standards etc.

Draft a brief note for publication in the journal criticizing the crude method of population forecasting adopted in the article and outline a more scientific but practical line of approach in estimating the adult population in working ages in India at the end of the planning period.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS IV & V (SPECIAL SUBJECT) : PROBIT ANALYSIS (THEORY).

Time : 4 hours Full marks—100

Attempt all questions.

1. (a) Explain the method of direct assay and point out its advantages and disadvantages from theoretical and experimental points of view.

(b) How do you calculate relative potency from data on direct assay and find the confidence limits.

(c) Discuss the possibilities and methods of eliminating concomitant variation in obtaining relative potency in the case of direct assays.

2. (a) What are the considerations leading to a logarithmic transformation of the dose in the analysis of a biological assay.

(b) Explain any method of obtaining the probit regression line mentioning and justifying if possible the assumptions on which the calculations are based.

3. Write an account of the various choices of dosage response curves and methods of estimation employed giving your own views on the subject.

4. Explain the principles of the design of assays based on quantal responses with special reference to (a) choice of doses, (b) balancing of inherent differences in experimental material, (c) sequential sampling and (d) control charts of a routine character during experimentation.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS IV & V (SPECIAL SUBJECT) : STATISTICAL QUALITY CONTROL.

Time : 4 hours Full marks—100

N.B. Attempt any four questions.

1. Show how to set up $\bar{X}$ and R charts in statistical quality control. Explain their use in industry.

In a factory, certain tolerence limits have been specified on the weight of items produced by a machine. Each item is automatically weighed as soon as it comes out of the machine. The operator of the machine has been instructed to adjust the machine as soon as an item is found to have its weight beyond specifications. Preliminary observations have shown that the machine does not produce more than six consecutive items within specifications and after the machine has been adjusted, it gives at least 4 consecutive items within specifications. State, giving reasons, how you would select rational sub-groups in order to estimate the natural limits of variation of the machine. Would you include observations falling beyond specification limits in your computations ? You are allowed to take any number of observations but you are not allowed to interfere in the working of the machine in any other way.

2. What considerations should influence the selection of an acceptance sampling plan ?

Of the 3 sampling plans given below, which would you prefer and why ?

(i) $N = 100$, $n = 5$, $c = 0$
(ii) $N = 500$, $n = 25$, $c = 0$
(iii) $N = 1000$, $n = 50$, $c = 0$

"A view commonly held by producers is that if the consumer is willing to tolerate a stated percentage of defectives in the lot, he should be willing to tolerate the same percentage in the sample." Comment, giving reasons.

3. Describe the three important types of control charts generally used in industry, giving the fields of applicability of each type.

4. Explain the construction of Dodge and Roming table for (1) single sampling and (2) doubling sampling.

Compare the relative merits of single sampling and double sampling.

5. Derive a sequential test procedure for standard deviations σ_1 and σ_2, when the true mean is known, the populations being normal. Draw a rough sketch of the OC and ASN curves.

6. Write short notes on :

(1) Average outgoing quality limit
(2) Modifiedcontrol charts
(3) theory of runs
(4) Producer's and consumer's risks.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS IV & V (SPECIAL SUBJECT) : PSYCHOLOGY & EDUCATION (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. Give a critical account of the uses and limitations of the factorial methods as applied to the Psychological and Educational problems.

2. (a) In what respects the 'Multiple Factor' theory can be regarded as an improvement over the 'Two Factor' theory ?

(b) Describe briefly the 'Principal Components' method of factoring and indicat eits advantages over the 'Centroid' method of solution.

3. (a) What is 'g' ? How is this estimated for an individual from his scores in several tests ?

(b) For a pattern given by $Z_j = a_{j1} F_1 + a_{j2} F_2 + a_j T_j$

$j = 1, 2, \ldots . n$, the number of tests, derive equations for the estimation of common factors F_1 and F_2, by the method of minimisation of the sum of square of the unique factors.

4. (a) State the general principles of the selection and arrangement of items for the construction of a test of Intelligence.

(b) Explain the procedure you will adopt to standardize an Indian language adaptation of the Binet-Simon scale.

5. (a) Why it is necessary to convert scores in an examination into so-called standard scores ? Define percentile score and T-score and discuss their relative advantages.

(b) Marks of a certain student A, for two subjects lie at the 75th and 45th percentile, while those for another student B, lie at the 60th percentile. Comment on their comparative performances.

6. (a) Define 'reliability' and 'validity' of tests, and describe briefly the available statistical methods of estimating them, with special reference to a suitable psychological test.

(b) Increasing the reliability of a test improves its validity only when the improved reliability means an increase in variance contributed by factors that the test has in common with the criterion—Explain this remark by stating relations that are known to exist between validity, test length, and reliability ; also give a suitable illustration of the situation.

7. (a) Describe the suitability of Kendall's τ-coefficient as a measure of rank correlation stating its advantages of application over Spearman's ρ.

(b) How would you measure the agreement in $\binom{n}{2}$ preferences between pairs of n objects provided by each of m observers ? Describe the significance test for the statistic proposed for this problem.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS IV & V (SPECIAL SUBJECT) : MATHEMATICAL THEORY OF SAMPLING DISTRIBUTION.

Time : 4 hours — Full marks—100

N.B. (a) Attempt any four questions.
(b) All questions carry equal marks.

1. Show that the ratio of two independent normal variables has frequency function

$$f(v) = \frac{1}{\sqrt{(2\pi)}} \frac{m_1\sigma_1^2 + m_1\sigma_2^2 v}{(\sigma_1^2+\sigma_2^2v^2)^{1/2}} \exp\left\{-\tfrac{1}{2}\frac{(m_1-m_2v)^2}{\sigma_1^2+\sigma_2^2v^2}\right\}$$

where m_1, σ_1 are the mean and the standard deviation of the first variate, m_2, σ_2 those of the second variate, and it is assumed that m_2 is so large compared with σ_2 that the range of the second variate is effective positive.

2. A stochastic variable is said to be distributed in a Cauchy's distribution with a parameter 'a' when its probability density function is given by

$$f(x) = \frac{1}{\pi} \frac{a}{a^2+x^2} \quad (-\infty<x<\infty)$$

Let (x_k), $(k=1, 2, \ldots. n)$ be a set of mutually independent stochastic variables, each x_k being assumed to be distributed in a Cauchy's distribution with a parameter a_k, $(k=1, 2, \ldots. n)$ respectively.

Find the distribution functior of their sum $s_n=x_1+x_2+\ldots.x_n$.

3. Let x_1 and x_2 be two independent and normally distributed stochastic variables. If x is defined as the greater of the values $|x_1|, |x_2|$ that is,

$$x = max\ (|x_1|, |x_2|)$$

find the mean value of x as well as that of x^2.

4. Let (x_k) be a random sample of size n from a population whose frequency density function $f(x)$ is continuous and has a continuous derivative $f'(x)$ in some neighbourhood of the quantile of order p of the distribution, i.e., the root (assumed unique) of the equation

$$\int_{-\infty}^{\xi} f(x)dx = p$$

where $0<p<1$

Find the asymptotic distribution of the sample quantile when n becomes infinite

5. Find the first two moments of a stochastic variable w defined by

$$w = ax_1^2 + bx_2^2$$

where x_1^2 and x_2^2 are assumed to be independently distributed according to the chi-square; distributions with the degrees of freedom f_1 and f_2 respectively, a and b being assigned positive constants.

Show that the distribution of w is approximately of the Type III form.

6. Obtain the distribution of the sample regression coefficients, in a random sample of size n from a bivariate normal distribution.

7. Find the distribution function of the ratio defined by

$$O = \frac{x_1^2 + x_2^2 + \ldots\ldots + x_m^2}{x_{m+1}^2 + x_{m+2}^2 + \ldots + x_n^2}$$

where $x_1, x_2, \ldots x_m, x_{m+1}, x_{m+2} \ldots x_{n-1}$ and x_n are independent normal variates with means $a_1, a_2, \ldots a_{m+1}, a_{m+2} \ldots a_{n-1}$ and a_n respectively and common variance unity.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS IV & V (SPECIAL SUBJECT) : SAMPLE SURVEYS (THEORY).

Time : 4 hours Full marks—100

N.B. Attempt any five questions.

1. Express the variance of n observations in a sequence in terms of the auto-correlation coefficients.

Use this result to deduce the variance of the mean of random, stratified and systematic samples.

How would you compare the efficiences of the three types of sampling when

$$\rho_u = e^{-\lambda|u|}$$

2. Explain the non-response problem in sample surveys.

How would you obtain an unbiassed estimate when this problem is serious in a survey.

Calculate the variance of the unbiassed estimate in the above case.

3. Explain the various aspects of sampling on two or more successive occasions.

Derive the estimate of the mean and its variance when sampling is carried out on two successive occasions.

4. In a multistage sampling survey scheme obtain the expectations of the different stage variances in the analysis of variance table. Derive the variance of the mean for finite sampling for the general case.

5. A single section of length 'a' of a line possesses the attribute A and is represented by the function

$$f(x) = 1 \quad (0 \leqslant x \leqslant a);$$
$$f(x) = 0 \quad \text{elsewhere.}$$

Obtain the variance of the totals of random and systematic samples by taking one sample each at intervals of length d randomly located with reference to the section possessing the attribute A for the following cases

$$a < d \quad \text{and} \quad a > d$$

6. Explain the term Interpenetrating checks (I.P.C.). Discuss critically the advantages and disadvantages of I.P.C.

7. Write short notes on —

(a) Balanced samples
(b) Sampling with probability proportional to size of unit.
(c) Multiple stratification.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS IV & (SPECIAL SUBJECT) : SAMPLE SURVEYS (APPLIED).

Time : 4 hours Full marks—100

N.B. (a) Attempt any five questions.
(b) All questions carry equal marks.

1. Describe and discuss the different types of "frames" suitable for censuses or surveys of human population.

2. Discuss in general terms the different sampling methods for market research and public opinion surveys. What is your opinion regarding the "quota method" of sampling ? What were the Possible causes for the failure of public opinion polls to predict the result of the 1948 Presidential election in U.S.A. ?

3. Try either (a) or (b).

(a) Give a general description of the "first round" of the "National Sample Survey" (India), and indicate briefly the desing of the survey.

(b) Discuss the present position of agricultural statistics in India with regard to the estimation of acreage and yields of different crops by sampling methods. Do you wholly agree with the review of this topic made by the "United Nations Sub-commission on Statistical Sampling" at its fifth session in December 1951 ?

4. Describe and discuss with suitable illustrations the following alternative methods of sampling on successive occasions (i.e., periodical surveys) :—

(i) a new sample is selected on each occasion without regard to previous samples;

(ii) the survey is repeated on the same sample;

(iii) part of the sample is replaced on each occasion, the remainder being retained;

(iv) a re-survey of a sub-sample of the original sample is made.

5. Distinguish between the different types of bias arising from the following causes :—(i) faulty method of selecting the sample, (ii) faulty method of estimation, and (iii) observational and computational errors.

Describe with suitable illustrations in what ways faulty methods of selection of the sample may give rise to bias, and how to avoid such bias.

6. Write a note on "random sampling numbers series", including the following points :—definition, methods of construction, tests for randomness (of the series), and uses of the series in drawing random samples from various types of population (hypothetical, existent etc.).

7. (a) Suggest a method for obtaining a random sample of words from the English language by the use of random sampling numbers and a dictionary.

(b) Consider a possible source of bias in replies to the following enquiries :— (i) persons are asked to state how often they attended a place of entertainment during the previous year; and (ii) persons are asked to state how many days have elapsed since they last attended a place of entertainment. Also consider how far the answers to (ii) may be used as a check on the answers to (i).

(c) Criticize the following plan for a proposed survey of social workers :— "Since it has not been possible to provide as much publicity for the study as is desirable, it is believed necessary to send out questionnaries to all social workers in the state so as to insure an adequate number of replies within a relatively short period of time. ..... No field recalls are to be made but we plan to send out a short note about a week after the original mailing to all those to whom the original questionnarire was sent, to remind them to send in their replies."

8. Suppose that in a sample survey for yield estimation for wheat the ultimate sampling unit is to be a rectangular "cut" of size 30 ft. by 20 ft. in a wheat-plot. Write out instructions for the field-enumerator indicating clearly (i) how to locate the "cut" randomly inside a wheat-plot, and (ii) how the boundaries of the "cut" should be demarcated with the help of any suitable appliance (e.g., rope, chain, tape, wooden rod, etc.). Emphasize particularly the precautions to be taken, assuming that the field-enumerator has not much previous experience of this type of work, or of the use of random sampling numbers.

Do you think that there is a possibility of bias if the "cut"-size is taken to be too large or too small ?

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPER IV & V (SPECIAL SUBJECT) : DESIGN OF EXPERIMENTS—APPLIED (THEORETICAL).

Time : 4 hours Full marks—100

N.B. (a) Attempt any five questions.
(b) All questions carry equal marks.

1. (a) Define inter-block information in incomplete block designs. Show that it can be extracted for every treatment comparison in a balanced incomplete block design. Hence or otherwise show that $\gamma > K$ in the latter design.

(b) Derive the expression for the efficiency factor of a balanced incomplete block design. Under what conditions will the efficiency of this design, using only intra-block information, exceed that of the corresponding complete block design.

2. (a) Explain the parameters of a partially balanced incomplete block design and prove the well-known relations between them.

(b) Give a few examples of partially balanced incomplete block designs having two associate classes and involving only two replications of each treatment. Also calculate the parameters for these examples.

3. Using the inter-block information approach, or otherwise, show that for a partially balanced incomplete block design with two associate classes and $K > \gamma$,

$$(\gamma - \lambda_1)(\gamma - \lambda_2) = (\lambda_2 - \lambda_1)\{(\gamma - \lambda_1)p_{12}^1 - (\gamma - \lambda_2)p_{11}^1\}$$

4. (a) Show how randomized lay-outs for an experiment involving p^2 varieties can be proposed for the following alternative designs :

(i) Triple lattices
(ii) Balanced lattices
(iii) Balanced lattice squares.

(b) Show the method of analysis for (iii).

5. Indicate the method of analysis of a split-plot experiment involving γ randomized blocks each of which has p main plots, and each main plot has q sub-plots, if accidentally either of the following situations arises :

(i) one of the main plots is missing, or,
(ii) one of the sub-plots is missing.

6. (a) If σ_1 and σ_2 are the theoretical standard errors per main plot and sub-plot respectively, both calculated on a sub-plot basis, in a split-plot experiment, give reasons why it is expected that σ_1 will be greater than σ^2 .

(b) If there are q sub-plots per main plot show that the variance of the difference between means of any two treatments not appearing inside the same main plot involves the weighted average of σ_1^2 and σ_2^2, the weights being in the ratio $1 : (q-1)$.

(c) Discuss how you would test the significance of the difference referred to in (b) when σ_1^2 and σ_2^2 are replaced by their estimates obtained from the experimental data.

7. The actual design (before randomization) of a $\frac{1}{8}$ replicate of a factorial experiment involving six factors, a, b, c, d, e, f, each at two levels, arranged in 4 blocks of 8 plots each, is given below :

Block I	Block II	Block III	Block IV
(1)	*ac*	*ad*	*ae*
ab	*bc*	*bd*	*be*
acde	*de*	*ce*	*cd*
bcde	*abde*	*abce*	*abcd*
acdf	*df*	*cf*	*cdef*
bcdf	*abdf*	*abcf*	*abcdef*
ef	*acef*	*adef*	*af*
abef	*bcef*	*bdef*	*bf*

Determine which of the interactions have been confounded and which main effects and interactions form *aliases* while being estimated from this design.

Assuming three-factor and higher interactions to be non-existent, write down the partition of the degrees of freedom in the table of analysis of variance against each ascribable source of variation.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS VIII & IX : ECONOMICS STATISTICS (PRACTICAL).

Time : 4 hours — Full marks—50

N.B. (a) Attempt any four questions.
(b) All questions carry equal marks.
(c) Use of Calculating machines and statistical and mathematical tables is permitted.

1. Fit a parabola by the method of least squares ti the following series of SAUERBECK'S Index Number of Wholesale Prices and plot two curves for observed and graduated values on the same sheet of graph paper :—

Year	Index No.	Year	Index No.	Year	Index No.
1881	85	1885	72	1889	72
1882	84	1886	69	1890	72
1883	82	1887	68	1891	72
1884	76	1888	70		

2. The numbers (in hundreds) of letters posted in a certain city on each day in a period of five consecutive weeks were as follows :—

Week	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
First	18	161	170	154	143	161	76
Second	18	165	179	157	168	195	85
Third	21	162	169	153	139	185	82
Fourth	24	171	182	170	162	179	95
Fifth	27	172	190	180	170	202	120

Calculate the indices of average fluctuations from day to day within each week and state how they may be utilized in adjusting the raw data.

3. (a) In 183 amalgamations, the number of companies in each such amalgamation is distributed as follows :

No. of companies in each amalgamations	2—3	4—5	6—7	8—9	10—11	12—13	Total
No. of amalgamations	74	50	32	17	8	2	183

Obtain the number of amalgamations each with 4, 5, 8 and 9 companies respectively.

(b) Find the rank correlation coefficient between length of service and order of efficiency of twelve salesmen A to L as reported by their manager below:

Salesman	Years of Service	Order of Efficiency
A	5	6
B	2	12
C	10	1
D	8	9
E	6	8
F	4	5
G	12	2
H	2	10
I	7	3
J	5	7
K	9	4
L	3	11

4. (a) Out of a large number of dividend-paying companies listed in a Stock Exchange Year Book, a random sample of 400 companies is selected and the average dividend paid out by them works at Rs. 4.74. The selected companies are then divided into 40 sub-samples of ten companies each by some random process and the standard deviation of the average dividends of the 40 sub-samples is found to be Rs. 0.775. Calculate the limits between which the average dividend paid by all the companies shown in the Year Book is expected to be.

(b) In a sample of 1,000 workers in a certain factory the mean weekly wage is found to be Rs. 17.5 and the standard deviation is Rs. 2.5. In another sample of 800 workers in another factory the mean wage is Rs. 18/- and the standard deviation Rs. 2.7. Discuss whether the wages in the two factories have the same mean and the standard deviation.

5. Give that x_1, the supply of jute, x_2 the price of raw jute and x_3 the price of manufactured jute (in suitable units) are connected by the equation $x_1 = 5.42 - 0.34x_2 + 1.33x_3$ and further that $r_{13} = 0.78$ and $r_{12.3} = 0.70$, calculate the value of $R_{1.23}$ and describe its uses. (The symbols used have their usual meanings; r=correlation coefficient (total or partial); R=multiple correlation coefficient; X_1 is the expected value of x_1; b=partial regression coefficient).

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

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

Time : 4 hours — Full marks—100

N.B. (a) Attempt question no. 1 and any one of the remaining questions.

(b) Use of Calculating machines is permitted.

(c) Census of India 1931, Vol. I, Part 1, will be provided for use at the examination.

1. (a) Starting with $l_{12} = 74{,}879$, compute abridged life table values of l_{17}, l_{22}, l_{27}, l_{32}, l_{42}, l_{47} and l_{52} 5 for a population of females in the child-bearing age-range (particulars of which are given below), using the approximate formula l_{x+n}

$$= \left(1 - \frac{Dx.x+n-1}{Px.x+n-1+\frac{1}{2}Dx.x+n-1}\right) l_x,$$

where l_x denotes the life-tables numbers living at age x, $Px.x+n-1$ the population actually enumerated between ages x to $x+n$ and $Dx.x.+n-1$ the related deaths in the age-group during a year.

	population, P	deaths, D	Female births, females
12—16	295,039	3,390	14.35
17—21	278,547	3,342	68.24
22—26	251,190	3,768	75.71
27—36	421,235	8,531	47.05
37—41	153,211	3,677	35.60
42—46	149,703	4,496	24.49
47—51	126,858	4,742	0.87

(b) Assuming that the formula $L_{x.x+n-1} = \frac{n}{2}(l_x + l_{x+n})$ holds for all the age-groups, where $L_{x.x+n-1}$ is the life-table population living between ages x and $x+n$, work out the net reproduction rate for the population. The rates of female births per 1000 females given above, are appropriate for the respective age-groups as a whole.

2. Employees in a certain government are compulsorily retired and pensioned off on attaining age 55 : voluntary retirements on full pension at exact ages 53 and 54 are also permitted. The numbers retired during the last 7 years at the various ages are given below :

Year	Age 53 (voluntary)	Age 54 (voluntary)	Age 55 (compulsory)
1946	57	126	2113
1947	32	341	1958
1948	25	289	2071
1949	22	73	2065
1950	18	103	1997
1951	20	85	982
1952	19	64	2060

Estimate the number of pensioners between ages 60 to 65 on first January 1958, assuming that the following life-tables applied to them :

Voluntary retirements		Compulsory retirements	
Age,		Age,	
53	21,397	55	14,013
54	20,425	56	14,091
55	19,476	..	..
..	..	..	..
..	..	..	..
60	14,033	60	10,084
61	14,069	61	10,254
62	13,221	62	9,543
63	12,389	63	8,851
64	11,573	64	8,178
65	10,773	65	7,524

Assume that deaths and retirements are uniformly distributed.

3. The existing population distribution in age-groups, in a certain constituency in an industrial area, is given below :

Age (last birth day)	Population
16—25	4,453
26—35	3,604
36—45	2,694
46—55	1,751
56—60	517
61.65	392
66—70	259
71.75	126
76—80	38
81 & above	21

For a general election with adult franchise (each individual aged 21 or above eligible to vote) exactly five years' hence, estimate the number of voters in the constituency, assuming that 1931 Census Punjab life table for males applies to the population. The combined net effect of migration and immigration could be ignored.

State the assumptions that you may have to make in working out the estimate.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS VIII & IX : PROBIT ANALYSIS (PRACTICAL).

Time : 4 hours Full marks—50

Attempt any one question

1. The following data give direct response measurements for standard and test preparations together with the weights of animals used. Estimate the relative potency correcting for the body weight and find its standard error.

Dose of standard preparation						Dose of test preparation					
.25		.50		1.00		0.025		0.05		0.10	
Response (R)	Weight (W)	(R)	(W)	(R)	(W)	(R)	(W)	(R)	(W)	(R)	(W)
35	47	50	48	140	55	30	50	50	48	80	53
40	50	100	50	80	47	35	48	45	46	65	45
45	42	75	46	80	48	40	51	75	52	70	51
60	48	80	50	55	41	60	49	50	49	120	55

2. Assuming that the tolerance distribution is of the form

Const. $e^{-\alpha(x-\theta)}dx$ $(0<x<\infty)$, obtain the estimates of α and θ from the observations given below.

Find the standard errors of estimates. Also estimate E.D. 50 with its standard error.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS VIII & IX : STATISTICAL QUALITY CONTROL (PRACTICAL).

Time : 4 hours Full marks—50

Attempt all questions.

1. The head office of a big industrial concern receives payroll checks from its various branches at the end of each month. The payrolls are prepared on sheets of 30 lines and in each sheet there are about 300 computations. It is necessary to investigate into the accuracy of these computations so that at least very bad offices can be improved. Prepare a graphical plan of action for a sequential sampling plan which would give only a 5% chance of condemning an office averaging 5 or fewer errors per sheet and 2% chance of approving an office averaging 6 or more errors per sheet.

2. Results of 60 observations on end breaks in 100 yard-lengths of yarn are given below. Plot a suitable control chart and give your comments.

Obs. No.	Breaks per 1000 yards	Obs. No.	Breaks per 1000 yards	Obs. No.	Breaks per 1000 yards	Ob. No.	Breaks por 1000 yards	Obs. No.	Breaks per 1000 yards	Obs. No.	Breaks per 1000 yards
1	7	11	5	21	4	31	2	41	3	51	3
2	9	12	4	22	7	32	4	42	0	52	4
3	9	13	1	23	5	33	5	43	1	53	2
4	4	14	6	24	6	34	2	44	2	54	3
5	7	15	5	25	5	35	2	45	3	55	3
6	13	16	6	26	5	36	2	46	4	56	4
7	5	17	5	27	5	37	4	47	4	57	10
8	6	18	6	28	4	38	2	48	4	58	11
9	2	19	7	29	4	39	3	49	3	59	7
10	4	20	2	30	1	40	3	50	6	60	7

3. A manufacturer producing rolled metal sheets wants to run a risk not more than .02 of a lot to be rejected. His product is under statistical control, the process average for Rockwell hardness number being 72 and variance 28.94. A buyer does not want to run a risk of more than .01 of receiving a lot with average hardness number less than 68. How many sheets should be tested frim each lot and what is the sample average which will differentiate between acceptable and rejected lots ?

If the buyer wants to run a risk of .01 of receiving lots with more than 3% of the items having hardness number less than 60, how many sheets should be tested from each lot and what should be the sample average differentiating between acceptable and non-acceptable lots ?

4. In a factory it was known that tool wear and trends due to other causes exsited. But it was uneconomical to attempt by frequent re-settings and replacements to correct for the consequent departures from prefect statistical control. It was necessary to maintain a control of the process that was sufficient but short of perfect control. Tolerances of 6.150 ,0.200 inches were found to be satisfactory for individual items. The following table gives the averages and ranges of 25 samples each of size 4. Set up a modified control chart which may be used by the factory as a basis for corrective action.

Sample No.	$\bar{X}$	R	Sample No.	$\bar{X}$	R	Sample No.	$\bar{X}$	R
1	6.090	0.080	10	6.034	0.096	19	6.288	0.134
2	6.018	0.119	11	6.129	0.122	20	6.307	0.091
3	6.139	0.157	12	6.226	0.117	21	6.234	0.125
4	6.084	0.106	13	6.187	0.103	22	6.264	0.121
5	6.055	0.121	14	6.214	0.132	23	6.245	0.138
6	6.068	0.135	15	6.199	0.109	24	6.270	0.085
7	6.177	0.098	16	6.101	0.099	25	6.295	0.156
8	6.153	0.124	17	6.160	0.113			
9	6.041	0.092	18	6.118	0.128			

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS VIII & IX : MATHEMATICAL THEORY OF SAMPLE DISTRIBUTIONS (PRACTICAL).

Time : 4 hours Full marks—50

N.B. (a) Attempt any 2 question.
(b) All questions carry equal marks.
(c) Use of Calculating machines is permitted.

1. Conduct the following experiment with the aid of a table of random numbers arranged in digits of four.

Obtain 10 samples of size 5 from the normal population with mean 1.50 and the variance 2.25.

Calculate the statistic t defined by

$$t=\sqrt{5}(\bar{x}-2)/s$$

for each sample $O.(x_1, x_2, \ldots\ldots x_5)$,

where

$$\bar{x}=(x_1+x_2+\ldots\ldots+x_5)/5$$

$$s=\sqrt{\frac{1}{4}\sum_{i=1}^{5}(x_i-\bar{x})^2}$$

Obtain from the t tables the 10 probabilities of the t's exceeding the calculated values. How do you combine these probabilities in practice to judge the significance of 10 calculated t's simultaneously ?

2. Conduct the following sampling experiment with the aid of a table of random members arranged in digits of five.

Obtain a table of 100 random numbers which may be considered as distributed according to the Poisson distribution with the mean 2. Test the goodness of fit for the empirical distribution.

Given $e^{-2}=0.13533528$

3. Obtain a table of 50 random numbers which may be considered as distributed according to the binomial distributions with the individual probabilities.

$${}^5C_r\,(0.20)^r\,(1-0.20)^{5-r},$$

$$r=0,\ 1,\ 2,\ \ldots,\ 5.$$

Construct the histogram for the distribution of these 50 numbers.

STATISTICIAN'S DIPOLMA EXAMINATION, 1953.

PAPERS VIII & IX : SAMPLE SURVEYS THEORY (PRACTICAL).

Time : 4 hours — Full marks—50

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

1. Below are given the data of crop-cutting experiments on maize for the Shahdol district in Vindhya Pradesh conducted during 1952-53. The district was stratified Kanungo-circle wise From each stratum two villages were selected at random and in each selected village tow fields were so ected at random and from each of the selected fields c.c. experiments on a plot randomly located in the field of size 1/80 acre was conducted. Yields were noted in chhataks per plot (1/80 acre). For each circle the area under maize in 52-53 is also given. Calculate the estimate of total production and its standard error. Also compare the readings of villages number 1 with those of villages number 2.

Name of Circle	Village		Area in acres	Name if Circle	Village		Ajea in acres
	1	2			1	2	
1. Beohari	NR*	NR	1699	9. Eastern	367	339	3505
	NR	NR			378	328	
2. Jaisinghpur	90	346	2767	10. Western	NR	144	7210
	102	166			NR	197	
3. Amdih	NR	NR	1931	11. Sohagpur	236	180	4975
	NR	NR			48	208	
4. Baronda	87	85	1577	12. Burhar	241	195	3778
	69	99			198	149	
5. Amarpur	82	40	1778	13. Jaitpur	0	139	1924
	195	236			189	121	
6. Chandia	206	110	6669	14. Kotura	182	146	1457
	176	53			86	183	
7. Singhwara	60	58	369	15. Dhurwasin	32	112	1456
	72	67			148	160	
8. Manpur	241	120	2875	16. Anupur	163	169	2030
	28	34			113	189	

(NR*—the experimental results were not received).

2. The percentage solids-not-fat in two successive months for all the 16 cows in a herd which were in their 2-6 months of lactation in one, at least, of these months were observed to be as follows :

Cow	1	2	3	4	5	6	7	8
November	8.82	8.94	9.86	8.90	9.00	9.13	8.90	9.02
December					8.98	8.66	8.68	8.86
Cow	9	10	11	12	13	14	15	16
November	9.46	9.52	9.28	9.22				
December	9.30	9.50	9.13	9.32	9.38	8.78	9.10	9.04

Estimate the change in percentage solids-not-fat between the two months, the mean percentage for December, the revised mean percentage for November together with the standard errors for all these estimates.

STATISTICIAN'S DIPLOMA EXAMINATION, 1953.

PAPERS VIII & IX : SAMPLE SURVEYS, APPLIED (PRACTICAL).

Time : 4 hours — Full marks—50

N.B. (a) Figures in the margin indicate full marks.
(b) Use of Calculating machines is permitted.
(c) Use of Barlow's Tables, Logarithmic Tables, Random Sampling Numbers is permitted.

1. In a stratified sample to estimate the acreage under a crop in a region, it is given that the variance, V, of the estimate, and the total cost involved W, are of the following forms :— (16)

$$V = \sum_1 \left\{ A_i^2 \, a_i \Big/ (n_i \, x_i^{g_i}) \right\},$$

$$W = b + \sum_1 n_i \, (c_i + d_i \, x_i),$$

where A_i is the geographical area, n_i = number of sampling units and x_i = size of the sample unit in the i-th stratum, and the other letters represent some characteristic constants. The number of strata is only three and the numerical values of the constants are given below :

b = Rs. 5,000/-

	A_i	a_i	g_i	c_i	d_i (per acre)
	sq. miles			Rs.	Rs. as.
1	7,000	0.09	0.4	30	3 8
2	9,000	0.12	0.5	36	4 0
3	8,000	0.11	0.6	32	3 0

Find out the optimum values of n_i's and x_i's for W = Rs. 50,000/-, and find out the corresponding value of V.

2. You are given a rectangle defined by the rectangular cartesian co-ordinates $0 \leq x \leq 90$ metres, and $0 \leq y \leq 40$ metres. Select four random "points" in this rectangle with their co-ordinates given correct to metres only, with the help of "random sampling numbers". Call these four "points" A, B, C, and D. Suppose you have to connect these four "points" by a "path" in such a number that the journey from any "point" to the next "point" in the path will be a straight line. (Thus any "path" will consist of three straight lines, e.g., B to A, A to D, and D to C). (14)

Enumerate all the possible "paths" connecting the four "points" selected by you and work out the length of each such path. (You are not to use a graph paper in answering this question).

3. Try any two of the following: (a), (b) and (c). (10+10)

(a) Draw up a questionnaire form for an investigation into the radio-listeners' preference as regards the radio programme of any station (e.g., Calcutta, Bombay, Madras, Delhi, etc.). (The form has to be drawn in English).

(b) Suppose that there is a pond exactly in the shape of hemisphere with a radius of 18 feet. (Obviously the upper surface of the pond forms the plane boundary and the bottom forms the curved-surface boundary). You are required to locate six "points" at random inside the pond in order to sample the water for bacteriological examination. A depth of two feet from the upper surface of the pond is to be excluded from sampling. (You are allowed to use "random sampling numbers", and you are to indicate the selected points by any convenient system of co-opernates correct to, say, tenth of a foot. Indicate clearly the procedure of selection).

(c) Attach the following weights to the 26 letters of the English alphabet :— The five vowels (a, e, i, o, u) will each have a weight $=w_1$ the two semi-vowels (w, y) will each have a weight$=w_2$, and the remaining 21 letters (consonants) will each have a weight$=w_3$. It is given that

$$w_1 : w_2 : w_3 = 2 : 3 : 1$$

Further for each letter (say the first letter) the "capital" and "small" symbols (like A and a) are to be considered separate, there being thus 52 symbols in all. The weight attached to a "capital" symbol will be half the weight of the corresponding "small" symbol.

Select, with replacement, six symbols out of the above mentioned 52, with the probabilities of selection proportional to the weights. (You are allowed the use of "random sampling numbers'. Indicate clearly the procedure of selection).