INDIAN STATISTICAL INSTITUTE

QUESTION PAPERS

for

The Statistician's Diploma Examination

1952

Price Re. 1/-

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER I—THEORETICAL STATISTICS (GENERAL)

Time—4 hours Full Marks—100

Answers to the different groups are to be given in separate books.
All questions carry equal marks.

Attempt ANY THREE questions from each group.

GROUP "A"

1. (a) Calculate the real root of the equation

$$x^3+x-3=0$$

by inverse interpolation.

(b) Establish from first principles Weddle's rule of numerical integration.

$$\int_a^{a+6} f_x\,dx = \frac{3}{10}(f_a + 5f_{a+1} + f_{a+2} + 6f_{a+3} + f_{a+4} + f_{a+5} + f_{a+6})$$

2. R red and B black balls are arranged at random. What is the probability of finding r red and b black balls.

(a) in the first set of $(r+b)$ balls,
(b) in the last set of $(r+b)$ balls,
(c) in both the above two sets,
(d) in at least one of the above two sets ?

3. Write notes on the use and significance of the
(a) Law of large numbers
(b) Lexis Ratio

4. (a) Describe carefully a probability-model from which the binomial distribution can be generated.

(b) Show that the Poisson Exponential distribution can be regarded as a limiting case of the binomial distribution.

(c) The number x of alpha-particles emerging from a radio-active material is counted for a given period, say one minute. Do you think that the probability-distribution of x will be of the Poisson (Exponential) form ? State clearly the reasons for your answer.

GROUP "B"

1. Derive the sampling distribution of Student's t. Show that it has many more applications beyond that for which it was first introduced.

2. Show that the four frequency distributions—normal, t, χ^2 and z—which provide the greatest number of tests of significance in common use, are all closely related.

Write a short note on the wide class of problems for which z provides the appropriate test of significance.

3. What is the role of transformation of raw data in connection with analysis of variance? Discuss critically two well-known transformations, stating clearly the corrections under which each is appropriate.

4. (a) Discuss the uses of the multiple and partial correlation coefficients.

(b) In a p-variate population, with the variates $x_i (i=1, 2, \ldots\ldots p)$, denote the correlation coefficient between x_i and x_j by r_{ij} and the standard deviation of x by σ_j; further denote by R_{ij} the cofactor of r_{ij} in the determinant R of the elements r_{ij}. Then show that the partial regression coefficient is given by

$$b_{1j.\ 2,\ 3,\ \ldots\ldots\ldots (j-1),\ (j+1),\ \ldots\ldots p} = -\frac{R_{1j}}{R_{11}} \frac{\sigma_1}{\sigma_j}.$$

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER II—APPLIED STATISTICS (GENERAL)

Time—4 hours. Full Marks—100

Answers to the different groups are to be given in separate books.
All questions carry equal marks.

Attempt ANY THREE questions from each group.

GROUP "A"

1. Discuss the problem of statistical derivation of demand curves from time series data, pointing out some of the difficulties that have to be faced and indicating how they could be overcome.

2. Examine the concept of "index number of industrial production" and bring out its relationship with national income accounting.

Describe in general terms the method of construction of such an index.

3. Explain clearly what is meant by a cost of living index number. Sketch the lines on which you would proceed to prepare such an index and indicate some of the practical uses to which the index can be put where constructed. Mention some of the special difficulties that arise in the construction of the index in times of rapid changes of consumption pattern and indicate how they can be tackled.

4. Write a critical note on the official statistics available in this country relating to Inland Trade, pointing out their defects if any and suggesting lines of improvement if possible. Your answer should deal particularly with the following three points, viz., (a) method of collection of data, (b) system of compilation and (c) mode of publication of these statistics.

5. Explain the role of the "cost" and "variance" functions in planning an efficient sample survey.

You are required to estimate the area under a main crop grown in your province by using the sampling method. How will you design the survey and analyse its results?

GROUP "B"

1. Explain what is meant by a balanced incomplete block design.

In an agricultural field trial there are 6 varieties to be tested. The heterogeneity in the field does not permit a block size of more than 3 plots and the total number of plots available is 30. Construct a balanced incomplete block layout for testing the variety.

2. What are the main considerations used in the use and construction of confounded factorial designs? A, B, C, D are four factors at two levels each. In a

confounded design with these factors, if I, BC, ABD and ACD are the treatment combinations included in the principal block, what are the contents of the other blocks? What effects would be confounded in this design ?

3. Explain the terms :—Gross reproduction rate, Net reproduction rate and Stationary population.

What is the significance of a net reproduction rate less than one ? Name any country where this is so.

4. Give a brief account of the Indian Census. What are the special features of the 1951 Census as compared to the previous census from the point of view of the data collected and the processing thereof.

Indicate briefly how Life Tables are constructed from census data.

5. What are the essential features of a mental test which distinguishes it from everyday judgment.

Explain the various methods of combining scores in different tests in order to assess the position in the tests as a whole.

6. Describe the role played in genetic and statistical considerations in the development of plant selection as an objective and systematic technique.

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER III—STATISTICAL INFERENCE.

Time—4 hours Full Marks—100

Answers to the different groups are to be given in separate books.
All questions carry equal marks.

Answer three questions from Group "A" and one from Group "B".

GROUP "A"

1. A large manufacturing firm produces 100,000 bulbs every day and you as the statistician of the firm are asked to make a dependable statement every day about the proportion π of defective bulbs. You take a sample of 400 bulbs every day and observe the proportion of defective bulbs in the sample. Explain clearly how you can set up a confidence interval (approximately) for π.

Suppose it is known to you that the true proportion π varies from day to day and that it can lie anywhere between .01 and .02 all values being equally likely. How will you proceed to utilise the information given by the sample? Sketch a note on the method that you are using here.

2. The height distribution of women in a State is known to be normal with known s. d. σ and unknown mean θ. It is stated that the average height is 5 ft. You are to test the above statement by taking a random sample of 100 women. Clearly indicate how to proceed, explaining incidentally the following terms

(i) critical region, (ii) power function, (iii) uniformly best one-sided test and (iv) uniformly unbiassed critical region.

Suppose I want to test the above hypothesis on the basis of the observed proportion p of women who are above 5 ft. Obtain an approx. unbiassed critical region of size 5%. Give an expression (approx.) for the power function of the above test and also that of the uniformly best unbiassed test.

3. Give your comments on any three of the following statements (which may be true or false):—

(i) Maximum-likelihood estimators are always unbiassed but need not be always consistent.

(ii) For a normal population with unknown mean and variance the sample mean is sufficient for the population mean but the sample variance is not so for the population variance.

(iii) Tchebyscheff's Lemma explains why we *must* take unbiassedness and minimum variance as our criteria for choosing good estimators.

(iv) An estimator t is said to be a consistent estimator if its bias tends to zero as the size of the sample on which t is based tends to infinity.

(v) If t_1 and t_2 be independent statistics then the amount of information in (t_1, t_2) is the product of the two separate informations. But if they are dependent then the amount of information is zero.

4. Write short notes on any three of the following :—

(i) The Law of Succession (ii) Similar Regions (iii) Efficient estimators (iv) The Method of moments.

Group "B"

1. Show that in a $(p+q)$ variate normal distribution the conditional distribution of p of the variates given the other q is also normal. Find the mean values and the variances and covariance of the p variates from this conditional distribution.

How do you interpret the correlation of any two variables in this conditional distribution ? What is the sample estimate of any one such correlation. How do you test the hypothesis that this correlation coefficient is zero in the population ?

2. In a randomised block design with two varieties and 10 blocks, three measurements were taken on the plants from each plot. They were (i) the yield (ii) total weight of bark and (iii) total weight of dry branches excluding the bark.

(a) How do you test the difference between the varieties by considering all these measurements simultaneously.

(b) Suppose the market values of these commodities (i), (ii) and (iii) mentioned above are Rs. 3/-, Rs. 2/- and Re. 1/- per unit. How do you determine whether one is economically a better variety

(c) If we are told that measurements (ii) and (iii) are not likely to be different on the average for the two varieties, how do you utilize this information in testing for the differences in yield ?

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER VI—PRACTICAL

Time—6 hours | Full Marks—100

Figures in the margin indicate full marks.

Use of calculating machines is permitted.

1. Making use of Simpson's one-third rule with interval of differencing equal to 0.1 estimate the value of π from the following formula :—

$$\frac{\pi}{4} = \int_0^1 \frac{dx}{1+x^2}$$

(In your calculations you are to retain figures correct to sixth decimal place). 10

2. The combined heights in inches of 11 pairs of brothers (x) and sisters (y) are shown in the following table :—

Pair No.	1	2	3	4	5	6	7	8	9	10	11
Combined height (inches) $x+y$.	140	132	131	130	135	133	135	137	138	124	128

Estimate the correlation coefficient between the heights of brothers and sisters and test for its significance making use of the following additional information :—

	Mean height (inches)	Standard deviation (inches)
Brother	69	2.59
Sister	64	2.54

Note :—Square of the standard deviation is equal to the sum of squares of deviation from the mean divided by 11. (15)

3. The following table gives the observed distribution of correlation coefficient (r) from random samples of size $n=8$ from a bivariate population. Test whether the sample s have been drawn from a bivariate normal population with correlation coefficient $\rho=0$, in which case the sampling distribution of r is known to be :—

$$f(r)\,dr = \frac{\left|\frac{n-1}{2}\right|}{\left|\frac{1}{2}\right| \left|\frac{n-2}{2}\right|} (1-r^2)^{\frac{n-4}{2}}\,dr.$$

Values of r	Frequency	Values of r	Frequency
— 1 to —.825	2	.075 to .225	85
—.825 to —.675	27	.225 to .375	98.5
—.675 to —.525	44	.375 to .525	65
—.525 to —.375	60	.525 to .675	37.5
—.375 to —.225	96	.675 to .825	14.5
—.225 to —.075	114.5	.825 to 1	3
—.075 to +.075	103		

(20)

4. Fit a curve of the form $y=a+bx+cx^2+dx^3$ to the data of the following table. Test whether a second degree parabola is a better fit than a straight line and whether the third degree polynomial is a better fit than a second degree parabola.

Commercial failures in the U.S.

Year (x)	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919
No. of failures (thousands) (y).	12.6	13.4	15.5	16.0	18.3	22.2	17.0	13.9	10.0	6.5

(20)

5. The following results were obtained from 100 sets of observations on x_1, x_2 and x_3.

x_1 = Stature in inch ; x_2 = Weight in lbs ; x_3 = basal metabolism per unit time.

n = No. of observations		Mean	Standard Deviation	Correlation coefficients with		
				x_1	x_2	x_3
x_1	100	177.0	7.6	1	0.57	0.61
x_2	100	64.1	10.3	0.57	1	0.80
x_3	100	1631.7	204.7	0.61	0.80	1

Note :—Square of the standard deviation is euqal to the sum of squares of deviations from the mean divided by n.

Set up the regression equation for estimating x_3 from x_1 and x_2. Test whether the partial regression coefficient of x_3 on x_1 is significantly different from 6 and whether the partial regression coefficient of x_3 on x_2 is significantly different from 13. (15)

6. The table below gives the frequency distribution of head breadth of 1000 males. What type of Pearson's system of frequency curves can fit this distribution ? Can you consider Normal curve as a good approximation ? Fit a normal curve to this distribution and test the goodness of fit.

x	f
5.45—5.55	3
5.55—5.65	12
5.65—5.75	43
5.75—5.85	80
5.85—5.95	131
5.95—6.05	236
6.05—6.15	185
6.15—6.25	142
6.25—6.35	99
6.35—6.45	37
6.45—6.55	15
6.55—6.65	12
6.65—6.75	3
6.75—6.85	2
TOTAL	1000

Make use of the following additional information :—

If $\dfrac{x-6.10}{0.1} = \delta$

$\Sigma f\delta = -385$

$\Sigma f\delta^3 = 4103$

$\Sigma f\delta^3 = -2017$

$\Sigma f\delta^4 = 55061$

(20)

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER VII—PRACTICAL

Time—6 hours Full Marks—100

Figures in the margin indicate full marks.

Use of Calculating Machines is permitted.

Credit will be given for neatness and proper arrangement of calculations.

1. The following table gives the Electric Power Production (Per calendar day) of the United States of America during the years 1921-1930 :—

ELECTRIC POWER PRODUCTION 1921-1930

(Average per calendar day ; millions of killowatt hours)

Years	January-March	April-June	July-September	October-December
1921	112	107	109	120
1922	123	124	130	145
1923	153	149	148	159
1924	165	155	153	171
1925	177	172	178	195
1926	200	194	199	216
1927	220	215	214	228
1928	235	230	238	256
1929	263	261	266	276
1930	271	262	256	260

Seperate out the trend and seasonal components of the series and estimate the electric power production of U.S.A. in the four quarters of 1931. (15)

2. Samples of five were taken every hour in a factory producing aluminium plates of specified thickness for use in an aeroplane engine. The thickness of the sample plates were measured and the deviations from specified thickness are tabuled below :—

Sample No.	Thickness of plates (deviation from specification in m.m.)				
1	− .54	− .21	− .60	−1.59	− .60
2	+ .22	+1.45	− .84	+1.71	+ .93
3	+ .17	−1.27	+ .48	− .83	+ .14
4	− .12	− .66	+ .50	+ .93	+ .81
5	+1.54	− .40	+1.24	+1.42	+ .77
6	+2.05	+1.67	−1.08	−1.37	+ .45
7	− .97	+ .64	− .36	+ .91	+1.26
8	− .18	−1.81	− .32	+ .87	+ .25
9	− .88	+ .02	− .93	− .60	+ .61
10	+ .34	−1.26	+ .38	+ .38	− .32
11	+ .92	+2.17	+1.17	+ .56	+1.87
12	+ .31	+ .20	+1.20	+ .03	− .40
13	+ .22	+1.09	+ .63	− .17	+ .54
14	− .21	+1.08	+ .48	+ .48	+ .37
15	+1.26	+1.03	− .28	+2.17	+ .38

Prepare a control chart for the average of the samples, the expected proportion to lie outside the inner limit being 10 per 100 and that to lie outside the outer limit being 1 per 100.

Write a note on the features of the production process brought out by the control chart. (15)

3. A weaving experiment was done to determine the warp-breakage rates of four warps numbered 1, 2, 3, 4 each of which had been treated differently. The warps were woven simultaneously in four looms A, B, C, D and the total weaving time was divided into four periods. At the end of each period the warps were interchanged between the looms according to the plan of the table below, so that by the end of the experiment each warp had spent one period in each loom. The warp-breakage rates are given in the table below:

	Warp no.			
	1	2	3	4
Weaving period	A	C	D	B
1	5.52	2.87	9.76	6.69
	C	D	B	A
2	6.02	6.25	5.14	9.16
	D	B	A	C
3	8.90	2.91	5.77	6.53
	B	A	C	D
4	6.09	5.07	2.83	9.77

Analyse the data to detect the effect of looms, weaving periods and warp-treatments on warp-breakage rates. (15)

4. Two teachers T_1 and T_2 have rated a large group of pupils for "Social Responsibility" on a five-point scale. A rating of 1 means that the trait is possessed in marked degree, a rating of 5 that it is almost if not entirely absent and ratings of 2, 3 and 4 indicate intermediate degrees. The percentage of children assigned each rating is as follows:—

	Teachers	
	T_1	T_2
Rating		
1	10%	20%
2	15%	40%
3	50%	20%
4	20%	10%
5	5%	10%

Calculating combined ratings, compare for "Social Responsibility" five children C_1, C_2, C_3, C_4, C_5 whose ratings by the two teachers are given below:—

	Teachers	
	T_1	T_2
Child		
C_1	1	5
C_2	2	4
C_3	3	3
C_4	4	2
C_5	5	1

(10)

5. The following production figures for Indian industries were collected from official publications. Compute an index of Industrial Production for May 1949 with base 1948=100.

Items:	Units	Production		Weights
		Monthly average	May, 1949	**
(1)	(2)	(3)	(4)	(5)
Indian Mill Cotton consumption ..	400 lbs. bale.	365323	374022	47
Jute Manufactures ..	tons	87367	86687	10
Coal ..	1000 tons	2477.5	2595.7	5
Electric energy sold to industries	Millions K.W.H.	200.2	211.9	5
Steel Ingots ..	Tons	104334	121297	8
Pig Iron ..	Ton	119076	140488	7
Paper ..	Ton	8159	0000	1
Matches ..	Cases	53324	55266	2
Paints ..	X	*	*	1
Sulphuric Acid ..	Ton	78667	7200	1
Cement ..	Ton	126352	169734	3
Sugar ..	Ton	83333	17096	10

*Production figures for paints were not available. The production ratio was estimated to be 108 for May 1949 in comparison with the monthly average production in 1948.

**Given in "Eastern Economist". (15)

6. From vital statistics and census records the following central mortality rates m_x at each age x were obtained for $x=21$ to $x=30$.

Age $x=$	21	22	23	24	25	26	27	28	29	30
Central rate of mortality per thousand at age x										
$m_x =$	3.27	2.84	4.07	3.59	3.66	3.74	3.81	3.90	4.00	4.12

Taking a radix of 10,000 at age 21, complete the life-table from age 21 to age 30.

Estimate the complete expectation of life at age 30, when it is given that the complete expectation of life at age 21 is 45.224 years. (15)

7. Collect relevant statistics from official publications and write an one-page note on agricultural, mineral and industrial production in India during the years 1950 and 1951. (15)

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPERS IV & V—(SPECIAL SUBJECT)

STATISTICAL QUALITY CONTROL

Time—4 hours Full Marks—100

All questions carry equal marks.

Attempt ANY FOUR questions.

1. Derive the Poisson distribution as a limiting case of the Binomial as n approaches infinity and np remains constant.

Explain how to set up control chart for defects and its uses in industry.

2. Explain the use of Operating Characteristic curve in sampling. What is the effect of inspection errors on the OC curve ?

It is decided to use the sequential probability ratio test to decide whether the mean m of a Poisson distribution is 3 or 6. Derive an expression for the OC curve.

3. Prove that the probability is unity that the sequential probability ratio test procedure will eventually terminate.

Compare the single, double and sequential sampling inspection plans. Show that among plans of the three types having almost the same OC cruve, the sequential plan requires the least amount of inspection.

4. Obtain an expression for the relationship between range and standard deviation.

Explain the use of standard deviation in control chart for variables. In what cases would you advocate the use of range in place of standard deviation and why ?

5. Write short notes on :—

(i) Probability limits versus 3σ limits

(ii) Multiple sampling

(iii) Simon Charts

(iv) Sampling inspection as an acceptance program.

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPERS IV & V—(SPECIAL SUBJECT)

MATHEMATICAL THEORY OF SAMPLING DISTRIBUTIONS.

Time—4 hours Full Marks—100

All questions carry equal marks.

Attempt ANY FOUR questions.

1. State the Inversion Theorem which expresses probability density function of a stochastic variable in terms of its characteristic functions.

The rth cumulant of a stochastic variable is

(i) $k_r = Ch^r$, $(c>0)$

(ii) $k_r = n\,(r-1)\,!$

Find the characteristic functions and using the Inversion Theorem obtain the corresponding probability functions.

Show that the second distribution tends to a normal distribution as $n \to \infty$

2. If $p(x)$, $p(y)$ be the probability density functions (p, d, f.) of two independent variables x, y $(a \leqslant x \leqslant b \quad c \leqslant y \leqslant d)$, show how you would obtain analytically the $p.d.f$'s of $x-y$ and $\frac{x}{y}$.

Given $p(x) = 6x\ (1-x)$, $0 \leqslant x \leqslant 1$

$p(y) = 1$, , $0 \leqslant y \leqslant 1$

find the distributions of $ax-by$ and $\frac{ax}{by}$, where a and b are positive constants. Deduce the distribution of $\frac{x^2}{y^2}$.

3. (a) A population has probability density function $f(x)$. A sample of x observations is drawn from the population and arranged in order of magnitude $x_1, x_2, \ldots\ldots x_n$. Find the distribution of the area under $f(x)$ between x_1 and x_2.

(b) Find the sampling distribution of the mean of a sample drawn from a rectangular distribution.

$$dF = dx, \quad 0 \leqslant x \leqslant 1$$

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

$$\frac{(x-\mu)\sqrt{n}}{s}$$

where $\qquad x = \Sigma x_i/n\ ,\ s^2 = \Sigma(x_i-x)^2/n-1$

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

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

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

5. s_1^2, s_2^2 are two estimated variances calculated from two independent samples of n_1 and n_2 observations respectively from two normal populations with unknown means. Obtain the distribution of the ratio s_1^2/s_2^2. Show how this distribution could be used to test the hypothesis that the two populations have equal variances and also to determine the power function of the test.

Find the likelihood criterion λ for testing the hypothesis on the basis of equal samples of size n and derive the distribution of $\lambda^{2/n}$.

If there are k samples, what is the criterion you would employ for testing the equality of the variances? Indicate the nature of its distribution and the method of handling it.

6. If $x_1, x_2, \ldots\ldots x_n$ are independent normal variates with means $a_1, a_2, \ldots a_n$ respectively and common variance unity, show that the distribution of

$$\chi'^2 = \sum_{i=1}^{n} x_i^2$$

given by

$$p(\chi'^2)d\chi'^2 = \frac{e^{-\frac{\chi'^2}{2}} e^{-\lambda/2}}{2^{n/2}} \sum_{j=0}^{\infty} \frac{(\chi'^2)^{\frac{n}{2}+j-1} \lambda^j}{\left|\left(\frac{n}{2}+j\right) 2^j j!} d\chi'^2$$

where

$$\lambda = \sum_{i=1}^{n} a_i^2$$

Find the expressions for the first two moments of χ'^2. Using these two moments, fit a χ^2 distribution to the above distribution.

7. Obtain the distribution of the sample correlation coefficient between two normal variables x and y when the population correlation is ρ.

Find also the distributed of b the coefficient of linear regression of y and x.

8. Define the D^2-statistic to test the hypothesis that the means in two multivariate normal populations having the same covariances are the same and obtain its distribution.

Show that this statistic is invariant for any linear transformation of variables.

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPERS IV & V—(SPECIAL SUBJECT)
SAMPLE SURVEYS (APPLIED)

Time—4 hours Full Marks—100

All questions carry equal marks.

Attempt ANY FIVE questions.

1. Discuss, with suitable examples, how the clear and unambiguous definition of the sampling units in a sample survey demands the existence or construction of some form of "frame".

Also describe in what ways such a "frame" may be defective.

2. Write a note on the preparation of Reports on Sampling Surveys, examining the recommendations of the "United National Sub-Commission of Statistical Sampling" in this respect.

3. Bring out clearly the difference between *bias* and *sampling error* in the estimate derived from a sample survey. Discuss the various ways in which such a *bias* may arise: e.g., faulty selection or location of the sample, faulty demarcation of the sampling unit, faulty methods of enumerating and processing the data, faulty methods of statistical estimation, etc.

4. Describe, with suitable examples, the uses of the "cost and variance functions" in planning a sample survey.

5. Discuss the problems of organisation, selection, training and supervision of the field investigators in a large-scale sample survey.

6. Discuss the problems of organisation, selection, training and supervision of the computing staff for processing the data of a large-scale sample survey.

7. A large-scale sample survey has to be undertaken for estimating *either* the acreage under, *or* the yield per acre of, wheat in an Indian province. Suppose that you are put in charge of all the preparatory work in this connection before the actual field-work starts; e.g., preparing the detailed design (including randomisation, if necessary) record forms, instruction-sheets for the investigators, etc. Describe briefly, but clearly, the tasks you have to perform in this connection.

8. Discuss critically some recent developments, either in India or abroad, in the technique of large-scale sample surveys which seem to you novel and useful.

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPERS IV & V—(SPECIAL SUBJECT)

PROBIT ANALYSIS

Time—4 hours Full Marks—100

Attempt all the questions.

All questions carry equal marks.

1. What are the principles underlying a biological assay ? Describe the statistical procedure of comparing the relative efficacy of two insecticides on the basis of data giving the percentage kill at various doses of the two drugs.

Supposing that the underlying tolerance distribution is unknown and no suitable transformation could be found to make the probit regression lines straight how do you compare the efficacy of two drugs given that each drug has been used in 4 different doses (103, 203, 303 and 403) and number exposed and killed are available in all cases.

2. Two drugs A and B have their ED50, estimated on large samples, as 5 m.g. and 6 m.g. Using this information only, can you determine the $ED50$ of a preparation obtained by mixing A and B in an equal ratio ? Design an experiment to determine the optimum ratio in which the two drugs are to be mixed to have the minimum $ED50$. How do you analyse the records of such an experiment ? Is it always possible to reduce $ED50$ below that of drug A by combining it with B ?

3. Assuming that the tolerance distribution is of the form Const. $e^{\alpha(x-\theta)}dx$ $(0,^{\infty})$ determine the estimates of α and θ from the four observations.

Dosage	x_1	x_2	x_3	x_4
Number exposed	n_1	n_2	n_3	n_4
Number killed	r_1	r_2	r_3	r_4

Also find the standard errors of estimates. Explain clearly the method of estimation you employ and its advantages over other methods of estimation.

4 Examine how the probit technique can be used to grade the items by their difficulty in an intelligence test consisting of several items. What data are needed for this purpose and how do you collect them ?

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPERS IV & V—(SPECIAL SUBJECT)

PSYCHOLOGY & EDUCATION.

Time—4 hours Full Marks—100

All questions carry equal marks.
Attempt ANY FOUR questions.

1. Explain clearly the main purpose of individual and group tests.

Suppose it is required to recruit candidates for a certain vocation; explain in details the method of procedure which you would adopt.

2. What is meant by parallel tests? Suppose $\| x_{ik} \|$ is a score matrix where n represents the number of individuals and k represents the number of parallel tests. Show that the variance due to interaction between persons and tests is equal to the variance of the error of measurement.

3. Starting with the fundamental equation $S = FP$ where S is the score matrix, F is the factor matrix and P is the population matrix, show that the evaluation of factors ultimately depend on the solution of a matrix equation.

Explain clearly how with the "Centroid Method" it is possible to estimate the factor loadings.

4. Suppose that there are t tests and that X_i denotes the score of any one person in the ith test, then it is assumed that

$$x_i = \lambda_i f + \mu_i g + \ldots + v_i h + \tau_i s_i \qquad (i = 1, 2, \ldots t)$$

where f, g, h ..... represents the person's measures in one or more general or group factors and S_i is the person's specifice ability in the ith test.

Find the estimates of $\lambda_i, \mu_i, \ldots \tau_i$ by using the Method of "Maximum Likelihood".

5. Obtain the expressions for T and P in the case of two rankings. How these expressions are modified when ties exist.

For the case of m rankings of n individuals, obtain an expression for the general relationship between them.

6. Write short notes on any two of the following:—

(i) Standardisation of test
(ii) Validity and Reliability Coefficient
(iii) Binet Scale.

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPERS IV & V—(SPECIAL SUBJECT)

DESIGN OF EXPERIMENTS (APPLIED)

Time—4 hours Full Marks—100

All questions carry equal marks.

Attempt ANY FIVE questions.

1. (a) The following diagram shows the lay-out plan of a completely randomized design for testing the effect of 8 treatments (0, 1, 2, 3, 4, 5, 6, 7) on the yield of a crop, 0 denoting control or no treatment :—

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

Set up the table of analysis of variance for this design, giving the degrees of freedom, sums of squares, etc., and explain how you obtain the different sums of squares. Also develop formulae for the standard errors of differences between the treatment means.

(b) Point out the defects in the above design. Give two improved lay-outs which are both free from these defects and discuss their relative merits and demerits.

2. A 5×5 Latin-square experiment was carried out to test the effect of five manurial treatments A, B, C, D and E, of which A was control (or no manure), on the yield

of sugarcane. The lay-out plan and the yields of plant cane in maunds per plot of 1/10th of an acre are given in the table below :—

PLAN AND YIELDS IN MDS.

Row	Column				
	I	II	III	IV	V
I	A 28	E 46	D 36	C 32	B 41
II	D 44	B 42	A 30	E 44	C 43
III	B 49	A 33	C 40	D 41	E 47
IV	C 37	D 38	E 47	B 45	A 35
V	E 47	C 37	B 46	A 40	D 37

The analysis of variance of the yield data is shown in the following table :—

Analysis of Variance in (mds.)^2 per plot.

Source of variation	D. F.	Sum of Squares	Mean Square
Rows	4	89.2	22.3
Columns	4	10.0	2.5
Treatments	4	555.6	138.9
Error	12	145.2	12.1
Total :—	24	800.0	

(a) Test the significance of variation due to treatments. Arrange the mean yields for the treatments in the descending order of magnitude, indicating by the conventional bar notation the significance at the 5 per cent level of probability of the differences in mean yields among the five manurial treatments.

(b) If a single observation is missing, obtain expressions for (i) estimating its value and (ii) estimating the standard error of the difference between the mean of the treatment with a missing value and the mean of any other treatment.

(c) Suppose that in the above experiment, the observation for treatment D in the second row (or first column) is missing. What would be its estimated value ?

3. Define partially balanced incomplete block designs and distinguish these from the generalised partially balanced incomplete block designs. For the partially balanced designs involving two and three associates develop the analysis of variance, the formuale for the variances of the differences between two estimated varietal means and the efficiency factors of the designs compared to randomised block designs.

4. (a) "No aphorism is more frequently repeated in connection with field trials than that we must ask Nature few questions, or, ideally, one question at a time.

This view is wholly mistaken, for Nature will best respond to a logical and carefully thought out questionnaire. Indeed, if we ask her a single question she will often refuse to answer until some other topic has been discussed." Discuss, bringing out clearly the advantages of factorial experimentation in comparison with simple single-factor experiments, both when all the factors are independent and when these are not independent.

(b) In a 2^4 factorial experiment with 'a', 'b', and 'c', as the three factors, each at two levels, define the various main effects and interaction. If four replications are adopted, give the structure of analysis of variance and outline briefly an actual method of computation of the various main effects and interactions.

5. (a) What do you understand by the term "partial confounding" in factorial designs, and how do you distinguish it from "complete confounding"? Explain, illustrating your remarks with the help of the 2^3 confounded designs.

(b) A 2^4 experiment in blocks of 8 plots is given in four replications so that a quarter of the information on each of the three-factor interactions is lost. Give the lay-out of the experiment and set up a table giving the estimates and the sums of squares for the three-factor interactions.

6. What is the physical significance of the concept of orthogonality of two effects in a design? When is a design said to be orthogonal? Give examples of the situations in which (i) a design laid out to be orthogonal becomes non-orthogonal (ii) an experiment may be suitably planned to be orthogonal though commonly non-orthogonal and (iii) two or more effects are inevitably like to be non-orthogonal and consequently a more complicated procedure of analysis has necessarily to be employed to separate the effects.

The following are the relevant extracts from an experiment in which five doses of Ammonium Sulphate had been tested with other factors:—

Dose (N lbs./acre)	Total yields of 24 plots.
0	311.94
30	374.46
60	363.78
90	330.12
120	305.76

Error Mean Square based on 80 d. f. 12.88

Separate the main effect of N into linear, quadratic, cubic, etc. components and test their significance.

7. What are the practical situations under which it becomes necessary to use strip-plot designs in field experiments? What are their advantages and disadvantages over the simple split-plot and randomised block designs? Give the structure of analysis of variance for a 4×3 design (4 strips p_0, p_1, p_2, p_3, across 3 main-plot treatments n_0, n_1, n_2,) with 6 replications, explain how you would obtain the different sums of squares and develop formulae for calculating the standard errors of the comparisons $(n_1 p_1 - n_2 p_1)$ and $(n_1 p_1 - n_1 p_2)$.

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPERS IV & V—(SPECIAL SUBJECT)

ECONOMIC STATISTICS.

Time—4 hours Full Marks—100

All questions carry equal marks.

Attempt ANY FIVE questions.

1. Do you think that the centralization of official statistics is an improvement on "departmentalization"? Discuss the present trend of official statistical organization in India.

2. Describe the main sources of foreign trade statistics in India and their chief features. Are there any gaps and defects? How can they be remedied?

3. How is India's Balance of Payments computed? Can you suggest any improvement?

4. Consider the validity of the following formulae for index number of prices, either singly or in combination:—

$$\frac{\Sigma p_n q_0}{\Sigma p_0 q_c} \text{ and } \frac{\Sigma p_n q_n}{\Sigma p_0 q_n}$$

where p_n, q_n, p_0, q_0 have their usual meanings.

5. Describe briefly the method of compiling an index number of production. How can a seasonal industry such as the sugar industry be included in such an index even though it is monthly?

6. Explain how income elasticity of demand may be estimated from family budget data. What adjustments are necessary in the cost of living index number for different income levels?

7. What are statistical demand curves? Discuss some of the pit-falls in their derivation. How can they be avoided?

8. Show the relationship between the moving average and the least square methods of graduation. Illustrate by deriving a formula for a seven- point weighed moving average from a least square cubic fitted to seven consecutive points.

9. Discuss some of the difficulties that have to be contended against in connection with correlation in time series.

10. Distinguish between individual income and social income. Explain how social accounting method helps in the estimation of the national income of a country.

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER IV & V (SPECIAL SUBJECT)

GENETICS

Time—4 hours Full Marks—100

All questions carry equal marks.

Use of Calculating Machines is permitted.

Attempt ANY FIVE questions.

1. How does study of linkage furnish support to the chromosomal theory of inheritance ?

Stating the assumptions involved, derive Haldane's formulae for relative distances between loci in terms of corresponding linkage values. What is the modification suggested by Kosambi and what is the resulting formula ?

2. A double heterozygote for the factors A-a and B-b is crossed with a stock known to be recessive for B-b and segregating for A-a. Calculate the amount of information regarding linkage value which can accrue from a progeny of the test-cross in the cases when (i) complete classification of the progeny is possible (ii) both A-a and B-b exhibit incomplete dominance but the classification is otherwise not completed and (iii) both the factors show complete dominance and classification is not completed. Show also that in the last case, if the double heterozygote is in repulsion phase, it is profitable to make this cross in preference to raising F_2's by intercrossing if the linkage value p is less than $\frac{1}{4}$.

3. In the course of an investigation into partial sex-linkage of a dominant defect in man, the following record was collected of the progeny of normal mothers and affected fathers :—

Source.	No. affected		No. Normal	
	sons	daughters	sons	daughters
Defect inherited from father's father	120	64	36	60
Defect inherited from father's mother	24	26	54	40

If the defect is known to be attributable to a single factor, test for the existence of linkage with sex and proceed to estimate the linkage value and its standard error, indicating the general procedure you adopt.

4. Discuss the distinct types of crosses, the records of which would provide information regarding sex-linkage of a human trait when the trait in its abnormal form is known to be (i) dominant (ii) recessive. Indicate which of Fisher's statistics would be useful in each case. Taking any one of these statistics outline the method of derivation of the estimate of linkage value and the variance of the estimate when the nature of ascertainment of the families containing affected progeny is not known.

5. Stating clearly the underlying assumptions show how the regression of progeny means on parental values may be taken as an estimate of the "genetic" fraction of the total variation in the parental population. Discuss the types of bias the estimate is subject to. Suggest a method of separation of genetic component of variation from the non-fixable part of heritable variation, in the case where the parental generation has been raised from a cross between pure strains.

6. Selection in a certain plant population is to be made on the basis of a number of observable characters. Develop a suitable discriminant function for selection, stating clearly the assumptions involved and discuss their validity. How would you obtain an idea of the relative superiority of the discriminant function adopted over the use of any single characteristic for selection from the data you would need for developing the discriminant function ?

7. The progeny of a mating $Aa \times aa$ are subjected to continued brother-sister mating in successive generations. If x_n denotes the frequency of heterozygotes in the nth generation in which sib-mating is practised, show that

$$x_n = \frac{2}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{4}\right)^n - \frac{(1-\sqrt{5})^n}{4}\right]$$

Derive the corresponding expression for the case of two sex-linked allelomorphs.

8. If a_1, a_2, a_3, a_4 are the four allelomorphs at a locus of an autotetraploid, give the expected proportions of different types of gametes produced, if the mean frequencies of equational separation and non-disjunction at the first division are respectively a and e for both male and female gametogenesis. Derive the expected segregations of the progeny of a simplex individual when (a) cross with (a) nulliplex and (b) selfed, in terms of the simplex index of separation.

In tetraploid tomatoes, plants simplex for a factor A-a were crossed with nulliplex and selfed. The progeny segregated as follows :—

	(A)	(a)
Backcross	600	480
Selfed	4000	1200

It is suspected that the dominant containing genotypes are equally viable but the nulliplex has a viability u relatively to them. Estimate the index of separation and the value of relative viability.

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER VIII—STATISTICAL QUALITY CONTROL (PRACTICAL)

Time—3 hours Full Marks 50

Attempt all the questions.

Figures in the margin indicate full marks.

Use of calculating machines, control chart tables and sampling tables is permitted.

1. It is desired to control the weight per unit length of material produced by four similar machines. The following table gives the values of averages and ranges in sub-samples of 2 for the four machines. Plot a control chart and give your comments. (17)

Sub-group No.	Machine No.	$\bar{x}$	R	Sub-group No.	Machine No.	$\bar{x}$	R	Sub-group No.	Machine No.	$\bar{x}$	R
1	1	53.50	1	7	1	48.00	6	13	1	49.50	3
	2	52.00	2		2	51.50	1		2	49.00	4
	3	49.50	5		3	49.50	5		3	50.00	6
	4	40.00	10		4	51.00	2		4	51.00	0
2	1	48.50	3	8	1	50.00	0	14	1	50.00	0
	2	50.00	8		2	53.00	2		2	52.00	0
	3	48.00	0		3	50.00	2		3	48.50	1
	4	45.00	10		4	50.50	1		4	53.00	6
3	1	48.00	10	9	1	50.00	0	15	1	51.50	1
	2	53.50	7		2	49.50	1		2	49.50	1
	3	52.50	5		3	51.50	1		3	48.50	1
	4	52.50	3		4	52.50	3		4	49.00	6
4	1	50.50	3	10	1	47.00	2	16	1	59.00	6
	2	51.00	2		2	55.00	0		2	52.50	1
	3	50.50	3		3	52.50	3		3	50.00	0
	4	53.50	5		4	47.00	6		4	50.50	3
5	1	53.00	2	11	1	54.50	7	17	1	51.50	1
	2	51.50	5		2	53.50	5		2	49.50	5
	3	49.50	3		3	51.00	6		3	55.50	5
	4	52.00	0		4	50.00	8		4	48.50	1
6	1	53.00	2	12	1	51.50	3	18	1	47.50	9
	2	53.50	5		2	53.00	2		2	48.50	1
	3	51.00	0		3	49.00	6		3	50.50	1
	4	49.50	3		4	52.00	0		4	51.00	2

2. On a spinning frame of 200 spindles, end breaks in yarn are observed for 15-minute intervals. Results of 40 observations are given below :—

Observation No.	End breaks	Observation No.	End breaks	Observation No.	End breaks	Observation No.	End breaks	Observation No.	End breaks
1	13	9	13	17	7	25	8	33	12
2	8	10	11	18	10	26	4	34	3
3	10	11	5	19	12	27	7	35	5
4	4	12	10	20	10	28	8	36	10
5	13	13	7	21	12	29	11	37	10
6	6	14	14	22	8	30	12	38	8
7	10	15	5	23	13	31	11	39	9
8	13	16	4	24	5	32	13	40	14

Plot a suitable control chart for the data and state whether breaks are under statistical control. (15)

3. Try either (a) or (b) (18)

(a) An article is submitted for inspection in lots of 1,000 and its estimated production is about 100,000 a year. In the recent past, it has averaged about 1.0% defective and has shown fairly good statistical control. The loss incurred by putting a defective component on production line is Rs. 1.45. Unit labour cost of inspecting the part is Rs. 0.018 under 100% inspection and Rs. 0.03 under a 2% AOQL double sampling plan. The plan selected is as follows :—

$$N=1{,}000 \qquad n_1=36, \qquad c_1=0 \qquad n_2=59 \qquad c_2=3.$$

It is estimated that 100% inspection manages to remove 80% of the defectives.

Compare the annual cost of 100% inspection and AOQL inspection.

(b) For a brand of bricks, $\bar{X}$ has been estimated as 28.1 and σ as 4.1. A consumer is willing to run a risk of 5 chances in 100 of receiving lots of coefficient of variation $V_s=0.2$. The producer whose statistically controlled output is presumed to be characterised by the above mentioned values of $\bar{x}$ and σ wishes to run no greater risk than a 1 in 100 of having lots rejected. How many bricks should be tested and what is the sample coefficient of variation v which divides acceptable from unacceptable lots ?

If the sample size is fixed by an industrial agreement, at what level must the producer control the quality of his product and what shall be the value of v which separates acceptable from non-acceptable lots in order that the consumer shall run a 1% risk of obtaining lots whose quality is given by a coefficient of variation $V_s=0.2$ and the producer a 5% chance of having lots rejected ?

STATISTICIAN'S DIPLOMA EXAMINATION—1952.

PAPER VIII—MATHEMATICAL THEORY OF SAMPLING DISTRIBUTION (PRACTICAL)

Time—3 hours Full Marks—50

Attempt any two questions.

All questions carry equal marks.
Use of calculating machine is permitted.

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

Take 50 consecutive sets of 5 four-digit numbers and considering them as samples from a rectangular distribution with range 0 to 1, find the range in each sample. Construct a histogram for the distribution of the sample range and sketch the frequency curve.

(Mention the tables of random numbers you use and the pages from which the numbers are taken).

2. In the following contingency table with fixed marginal totals the probability of getting a individuals in the first cell is given by the hypergeometric term

$$\frac{m!\,n!\,r!\,s!}{a!\,b!\,c!\,d!\,N!}$$

a	c	m
b	d	n
r	s	N

It has been suggested that this probability is well approximated by the area of a normal curve with mean $\frac{rm}{N}$ and $s.d. = \sqrt{\frac{mnrs}{N^2(N-1)}}$ between the ordinates at $a \pm \frac{1}{2}$

Test the adequacy of this approximation by computing for the case $m=n=10$, $r=5$, $s=15$, both the hypergeometic probabilities for all the possible values of a and their normal approximations.

3. Obtain a sample of 15 from the Cauchy population

$$dF = \frac{dx}{\pi(1+x^2)}, \quad (-\infty < x < \infty)$$

and find its mean.

Test whether the sample you have drawn is a random sample.

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER VIII—SAMPLE SURVEYS, THEORY (PRACTICAL)

Time—3 hours Full Marks—50

Attempt all the questions.

Use of Calculating machines is permitted

All questions carry equal marks.

1. A survey was conducted in 190 villages in Cuttack for estimating the total area under winter paddy under the following scheme. The whole area was divided into 31697 "clusters" of 8 continuous survey numbers each, and 13 villages were selected with probability proportional to the total number of clusters in the villages with replacement. From each village four Clusters were selected and were completely enumerated for the area under winter paddy. The data obtained are as shown below :—

Village Nos.	Area in acres (cents)			
1	549	1154	1002	86
2	136	329	0	337
3	98	522	938	0
4	610	298	357	188
5	109	296	915	105
6	652	0	79	190
7	510	94	605	682
8	0	497	0	0
9	105	410	107	161
10	213	126	575	561
11	1382	177	574	0
12	0	0	673	368
13	0	682	161	173

Obtain the total estimate and the error of the area under winter paddy in Cuttack. Find also the number of villages required to be sampled for estimating the total area with 5% precision with 2 and 4 clusters per village.

2. The following data give the yields of green weight (G.W.) of jute for plots of 1/160 acre sampled from twelve villages having three samples per village. The fibre weight was obtained for each of the samples in six of the villages only.

3. Attempt *either* (A) or (B).

(A) Draw random samples, each of size 5, from the following three populations —

(i) The fifty-two letters of the English alphabet, considering the capital letters and small letters as separate;

(ii) A number of iron balls whose weight are distributed as follows :—

Weight (Milligram)	Frequency
25.8	45
25.9	49
26.0	67
26.1	42
26.2	28

(iii) A number of persons, whose classification according to provinces and professions are shown below (imaginary data) :—

Province/ Profession	Bengalee	Bihari	Assamese	Punjabee
Lawyer	53	12	19	15
Doctor	28	52	23	19
Engineer	64	23	72	21
Teacher	14	25	33	38

(B) Draw a random sample of ten words from the words occurring in this question paper, ignoring the numerals, symbols and the words occurring in the headings and the tables.

N. B. In questions 2 and 3 above, the random selections are to be made with the help of "random sampling numbers"; all the steps starting from the "random sampling number" leading to the final selection are to be clearly indicated on the computation sheets.

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER VIII—PROBIT ANALYSIS (PRACTICAL)

Time—3 hours Full Marks—50

1. The following table gives log concentration of dose (x), the number of animals exposed (n) and the number killed (r) at each concentration.

x	n	r
Control	129	21
2.17	142	142
2.00	127	126
1.68	128	115
1.08	126	58

(a) Obtain the probit regression line and find the fiducial limits for its slope.

(b) Estimate E.D.50 and find its standard error.

2. For each concentration of a preparation replicated samples are available to determine the percentage of kill. Using this information, test whether the assumptions needed for the computation of the probit regression line are true.

Log	1st sample		2nd sample		3rd sample	
Concentration	n	r	n	r	n	r
2.17	100	95	100	92	100	89
2.00	100	75	100	70	100	80
1.68	100	55	100	57	100	50
1.08	100	43	100	40	100	45

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER VIII—ECONOMIC STATISTICS (PRACTICAL)

Time—3 hours. Full Marks—50

Attempt any three questions.

All questions carry equal marks.

Use of calculating machines is permitted.

1. The following table gives the exports of two varieties of cotton yarns and five varieties of cotton manufactures from the United Kingdom during 1938 and 1943 by quantities and by values:—

	Quantities		Values	
	1938	1943	1938	1943
Cotton Yarns	(mn lbs.)	(mn lbs.)	£(000)	1(000)
X	110.0	16.37	8,397	2,817
Y	12.03	2.818	1,278	531
Cotton Manufactures	(thous. sq. yds.)	(thous. sq. yds.)	£(000)	£(000)
A	235.3	69.32	3,841	2,854
B	421.6	83.15	7,776	3,319
C	281.9	102.9	7,152	5,805
D	368.8	90.86	10,628	6,381
E	78.58	28.08	2,564	1,922

Calculate a quantum index number Q for 1943 on 1938 as base, using 1938 values as weights, and another Q' for 1938 on 1943 as base, using 1943 values as weights.

Explain why $\frac{Q}{100}$ is not the reciprocal of $\frac{Q'}{100}$.

2. The following table gives the estimated personal expenditure on Fuel and Light in the United Kingdom:

	Quarterly £ (million)		
	1945	1946	1947
Quarter 1	78	84	92
2	62	64	70
3	56	61	63
4	71	82	85

Assuming a linear trend, calculate the seasonal index.

Forecast the expenditures for the first three quarters of 1948.

3. Fit (*a*) a straight line and (*b*) a parabola to the following series of cost of living index numbers by the method of least squares :—

Year	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935
Index No.	176	172	167½	166	164	158	147½	144	140	141	143

Compare the graduated values for each year, 1927 to 1933, derived from (a) the straight line and (b) the parabola.

4. (a) Assuming linear trends in the cases each of the two following series of index numbers, x for wholesale prices and y for retail prices, find the correlation coefficient for the fluctuations :—

year	1893	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905
x	266	295	324	306	291	275	281	283	296	283	366
y	107	86	99	116	101	102	92	97	84	100	116

(b) Fit a Pareto curve to the following data for a group of 20, 855 persons, where y is the number of persons with incomes over Rs. (x): —

x Rs. (000)	5	10	15	20	25	35	45	Total
y	11,554	4,143	2,114	1,327	889	507	321	20,855

Is the fit satisfactory ?

STATISTICIAN'S DIPLOMA EXAMINATION—1952

PAPER VIII—VITAL STATISTICS AND POPULATION STUDIES (PRACTICAL)

Time—3 hours Full Marks—50

Attempt all the questions.

Use of calculating machines & mathematical tables is permitted.

1. The number of registered male births for each year in the five year period 1936-40 and the census count for the age groups 0–1, 1–2, 2–3, and 4–5 taken at January 1941. (or at the close of 1940) for a certain community are given below. If you assume that the Indian life table for meals (1931 census report) is applicable to this community for the period 1936-40, calculate the expected number of male children in the age groups 0–1, 1–2, 2–3, 3–4, and 4–5. (There is neigther immigration nor emigration).

Male births		Census count	
Year		Age	
1936	18534	0–1	18122
1937	20387	1–2	18391
1938	19465	2–3	15614
1939	21911	3–4	15011
1940	18871	4–5	14500

Indian life table (males) 1931

Age X	Living at age X	Living between age X and $X+1$
0	100,000	85,443
1	75,126	71,352
2	68,230	66,174
3	64,380	63,037
4	61,856	60,954

If you assume that the census count is accurate calculate the average under registration of births for the period 1936-40.

2. The following table gives the age distribution and specific mortality rates for males for all India and for a certain industrial region :—

	Indian Males (1931)		The industrial region (Male 1931)	
Age group	Pop. (1000)	Death rate per 1000	Population (in units)	Death rate per 1000
0 — 1	5,349	291.0	5027	202.3
— 5	21,086	57.2	21402	5.3
—10	23,796	12.7	50	20.0
—15	21,573	8.5	41	0
—20	16,040	11.0	64	0
—30	31,781	15.2	75,403	5.4
—40	25,765	23.8	81,101	9.2
—50	17,485	34.7	72,011	13.7
—60	10,181	48.3	5,117	30.1
—70	4,905	73.1	Nil	—
Over 70	2,245	156.4	Nil	—

Compute the standardised death rate for males for the above industrial region keeping Indian (males) population of 1931 as the standard. Write down a sufficient

condition for the standarided death rate computed by the direct method to be equal to that computed by the indirect method.

3. Attempt either (a) or (b)

(a) In a health survey of a certain community, it was revealed that 1% of the population is suffering from malaria at any time. If the average duration of a malaria case is 7 days what is the annual incidence rate for malaria in that community ?

(b) If the attack of a certain disease results either in death or lasting immunity from that disease and if out of 500 people who were not attacked by that disease previously 200 came down with the disease before the end of two years of observation how many were expected to be attacked within the first year of observations.

———