

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, PART I—JULY 1947

FIRST PAPER

Time allowed 1 hour.

Maximum marks 100.

(It is not necessary to answer all the questions in order to secure full marks. Attempt as many as you like remembering that the questions do not carry equal marks and credit will be given according to the quality of the answer.)

1. (a) Two dice are thrown n times in succession. Find the least value of n so that the probability of throwing a double six exceeds $1/2$.

(b) Balls are taken one by one out of an urn containing a white and b black balls, until the first white ball is drawn. Show that the expectation of the number of black balls preceding the first white ball is $b/(a+1)$.

2. The means of two samples of sizes n_1 and n_2 are respectively m_1 and m_2 and their standard deviations are respectively σ_1 and σ_2 . Show that the standard deviation of the sample obtained by pooling together the two samples is given by

$$\sigma^2 = \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n_1 + n_2} + \frac{n_1 n_2}{(n_1 + n_2)^2} (m_1 - m_2)^2$$

In a class of 50 students, a defective student obtains 30 marks below the average mark obtained by the other students. Show that the standard deviation of the marks obtained by all the students cannot be less than 5.04. If this standard deviation is actually 6.4, calculate the standard deviation, when the marks of the defective student are left out.

3. (a) Define Pearsonian beta-coefficients. Calculate these for the distribution

$$p(x) = \text{Constant } x^2(1-x)^2 \quad 0 \leq x \leq 1.$$

(b) Assign the following distributions to one of Pearson's types :

$$dF = K_0 z^{-2/2} z^{n-1} dz$$

$$dF = C dx \left(1 + \frac{x^2}{n-1} \right)^{-n/2}$$

$$dF = K(1-y^2) dy$$

$$dF = Kt^{n-3} (1-t)^{\frac{N-p-2}{2}} dt$$

4. (a) For the binomial distribution $p(x) = {}^n C_x p^x (1-p)^{n-x}$ show that the mode is the positive integral value of x for which $np - q < x \leq np + p$ where $q = 1 - p$. Determine n and p if the mean is 90 and the standard deviation is 3.

(b) A distribution is known to be normal. The quartiles are 8.64 and 14.32. Calculate the mean and the standard deviation.

5. (a) R is the correlation coefficient between x and y . The variables are transformed into u and v given by $u = Ax + B$ and $v = Cy + D$. Show that the correlation coefficient between u and v is also equal to R . Establish the inequality relations between the correlation coefficient and the correlation ratio between two variables.

(b) x and y are two variables following the probability law

$$p(x,y) = 12x(x-y)^2 \quad 0 \leq x \leq 1 \quad \text{and} \quad 0 \leq y \leq 1.$$

Examine whether x and y are independent. Calculate the correlation coefficient and the regression equation of x and y .

6. Explain and illustrate the use of the concept of regression.

Work out the formula necessary to fit a parabolic regression curve to a given body of data relating to two characters x and y (more than one value of y corresponding to every x being available and the regression of x on y being wanted).

How should you test whether the deviation from parabolic regression is significant?

7. (a) Write a critical note on the concept of the random sampling distribution of a statistic and discuss its use in 'tests of significance'.

(b) A sample x_1, x_2, \dots, x_n is drawn from a normal population with mean m , and standard deviation σ . x' is an independent observation from the same normal population. Find the distribution of the Statistic

$$\frac{\bar{x} - x'}{\sigma} \sqrt{\frac{n}{n+1}}$$

$$\text{where} \quad n\bar{x} = \sum_{i=1}^n x_i, \quad (n-1)s^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

Illustrate the answer to the part (a) by explaining how the distribution can be used to test the hypothesis that a new observation belongs to the same normal population from which a previous sample of size n is known.

8. Give an account of the method of testing hypothesis based on large sample theory. Indicate the tests appropriate to the following:—

(i) The correlation coefficient R is 0.32 in a sample of size 400. Can you infer that the variables are non-correlated?

(ii) Two samples give 8.6 and 12.3 as means and 2.4 and 4.1 as standard deviations respectively. If the samples are of size 121 and 144, can you infer that the means are the same?

9. Obtain the distribution of the sum of squares of n standard independent normal variates, subject to k linear restrictions. How does this enable you to derive a suitable test for association in a contingency table. Explain the nature of Yates' correction for continuity, when using the statistic, for testing association in a 2×2 contingency table.

10. Here are two samples

(i) 0, 7, 6, 5, 2, 0, -2, -3.

(ii) 6, 4, 3, 1, 0, -1.

Test the hypotheses:

(a) That the first sample is from a normal population with zero mean.

(b) That the second is from a normal population with zero mean.

(c) That the two samples are from the same normal population with zero mean.

(d) That the samples are from the same normal population.

11. Bring out the essential nature of the numerator s_1^2 and the denominator s_2^2 in the ratio whose half logarithm gives the well known z-statistic

$$z = \frac{1}{2} \log_e \frac{s_1^2}{s_2^2}.$$

There are three normal populations with means m_1, m_2 and m_1, m_2 respectively. Two independent observations

$$\begin{array}{cc} x_1 & x_1' \\ x_2 & x_2' \\ x_3 & x_3' \end{array}$$

are available from each population. Show that the four linear functions

$$x_1 - x_1', x_2 - x_2', x_3 - x_3' \text{ and } x_1 + x_1' + x_2 + x_2' - x_3 - x_3'$$

are mutually orthogonal and that their expectations are independent of m_1 and m_2 .

How does this fact enable you to derive an estimate for σ^2 ?

Obtain a test of the hypothesis $m_1 = m_2 = 0$.

12. Write explanatory notes on the following :-

- | | |
|-------------------------------|-------------------------------|
| (i) Law of small numbers | (iv) Spurious correlation |
| (ii) Fiducial limits | (v) Efficiency of an estimate |
| (iii) Contingency coefficient | (vi) Method of Moments |

STATISTICIAN'S DIPLOMA EXAMINATION, PART I—JULY 1947

SECOND PAPER

Time allowed 4 hours.

Maximum Marks 100.

(It is not necessary to answer all the questions in order to secure full marks. Attempt as many as you like remembering that the questions do not carry equal marks and credit will be given according to the quality of the answer.)

- What are the chief components of an economic time series and how could you decompose such a series. Give a general exposition of the method, commenting on the adequacy or otherwise of the statistical procedure employed in this connexion.
- How would you set about constructing an Index of General Business Conditions in India. Explain what you would include as significant, where you would get your data and what use could be made of the Index when constructed.
- Define income elasticity and point out its difference with price elasticity. State the general method of evaluating the former statistically from family budget data.
- Give a short account of Indian Official Statistics relating to trade. Show how you would work out India's balance of trade with the help of these statistics.
- Write notes on any two of the following :-

(a) Lorenz Curve ;	(c) Fixed base versus chain Index ;
(b) Indifference Curve ;	(d) Logistic Curve.
- To what use, if any, can you put (i) vital statistics, (ii) Marriage Register, and (iii) periodic census figures, either singly or collectively, for preparation of (a) a fertility table, and (b) a mortality table ? State clearly the assumptions on which the tables are prepared from these books and justify these assumptions having regard to the object of these tables. Only a sketch need be given.

7. Explain the genetical structure underlying human blood groups with reference to both the systems (O, A, B, AB) and (M, N, MN). How does this help in laying down correct rules for blood transfusion and in some elucidation of the question of parenthood in doubtful cases?

8. Suppose you wanted to assess by a sample survey public opinion on the issue of the partition of Bengal. What questionnaire would you frame so as to quickly get at all shades of opinion, and how would you plan your enquiry and the analysis of your data?

9. What items of information, broadly speaking are published in the official books issued after the decennial census operations held in India? Suppose it is proposed to hold quick annual sample census operations in between decennial enumerations, which items would you consider most important for that purpose, and show how would you plan those surveys?

10. Explain how the normal probability integral is used to reduce to quantitative terms on a numerical scale performances and judgements involved in various types of educational, vocational and psychological tests.

STATISTICIAN'S DIPLOMA EXAMINATION, PART I—JULY 1947

PAPER III (Practical)

Time allowed 5 hours.

Maximum Marks 100.

1. The following figures are supplied to you:

(a) Industries	Paid up Capital of Companies (in lakhs of rupees)	
	Those included in the index of profits.	At work at the end of 1939-40 coming under the various groups included in the index.
	(1)	(2)
Cotton Mills	18.15	38.99
Jute Mills	19.00	20.19
Cement	8.41	10.45
Tea	5.39	13.31
Iron & Steel	13.22	18.32
Sugar	4.03	11.64
Paper	2.15	2.19
Coal	5.55	8.72
	<u>75.00</u>	<u>124.11</u>

Total paid-up of all Companies in 1939-40

...Rs. 3,03,51 lakhs.

(b) Indices of Profits (Base 1928=100).

	Cotton Mills	Jute Mills	Tea	Coal	Sugar	Iron	Paper	All Industries
1939	154.6	13.6	96.2	139.1	170.4	239.3	151.8	72.4
1942	760.7	49.2	219.5	110.3	219.8	403.3	488.4	169.4

You are asked to write a short note (not exceeding 15 lines) on the trends of profits in Industries in India. Make comments on :

- (1) Source.
- (2) the adequacy and appropriateness of these figures to give you a picture for industrial profits in the recent years.
- (3) the representativeness of the index shown : Make such further calculations as you may think necessary.

What further information would you like to gather on the above before you feel you can give a complete and well reasoned reply? Can you suggest how these indices should be constructed so as to allow for the inclusion of new companies and exclusion of defunct ones?

2. Write down the expected cell frequencies from the observed figures on the hypothesis that the two attributes are independent. Calculate Chi-square and hence test for association.

		Musical Faculty			Total
		Good	Botwixt	Poor	
Mathe-	A	364	283	65	712
	B	312	574	162	1048
matical	C	62	87	175	324
	D	41	97	135	273
Total		779	1041	537	2357

3. Draw 150 sets of samples of size 5 from a normal population of mean = 10 and standard deviation = 5.

Show that the frequency distribution of means of the above set also follows a normal law.

(Table of normal deviates to be supplied)

4. From the following data find the correlation coefficient between blood pressure and age, and test its significance. Find the fiducial limits of the correlation coefficient on 95% probability level. Also test whether the distribution of blood pressure can be regarded as normal.

Blood Pressure in m.m.	Age in years											
	10-	15-	20-	25-	30-	35-	40-	45-	50-	55-	60-	65-70
39.			1									
45.	1			1								
51.	2	2	4	6	6	6	2	2	1			
57.	6	6	10	30	28	13	10	5	6			2
63.	28	33	121	160	191	100	75	44	34	4	6	2
69.	18	26	107	168	158	108	68	35	27	11	11	5
75.	18	20	64	71	79	53	27	18	26	3	1	2
81.	7	12	61	80	107	62	33	30	21	6	7	2
87.	10	8	14	10	19	9	9	5				2
93.	3	8	11	18	12	12	11	8	14	4	1	
99.	2	1	6	3	8	4	6	4	2			1
105.					2	2	2	1	1			
111.			1			1	1	1	1	1		
117.			1				1					
123-129								1	1			

STATISTICIAN'S DIPLOMA EXAMINATION, PART I—JULY 1947

PAPER IV (Practical)

Time allowed 5 hours.

Maximum Marks 100.

1. Arrange the following figures in a suitable diagrammatic form and make comments on the consumption of coal in various industries.

Consumption of Coal by various Industries							
	Railways	Steel works*	Bunkers	Exports	Cotton Textile	Bricks & Tiles**	Soft Coke
1939	8,457,687	2,912,927	602,000	1,688,992	1,791,000	1,221,000 (856,800)	888,982
1940	8,738,353	3,223,187	487,000	2,112,281	2,029,000	1,211,000 (868,700)	962,825
1941	9,591,582	3,395,850	480,000	1,734,580	2,435,000	1,382,000 (937,400)	957,553
1942	9,348,014	3,235,108	347,000	422,001	2,258,000	1,476,000 (1,033,200)	431,858
1943	9,784,443	2,988,800	—	346,692	—	—	354,835
1944	10,144,863	2,617,013	580,939	262,628	1,599,976	970,214 (736,289)	445,721
1945	9,173,727	2,611,488	1,140,709	122,111	2,010,820	1,243,336 (889,560)	547,232

*Including engineering works (figs. only in steel works)

**Including Potteries & Coments.

2. In order to study quantitatively how the Foreign Demand for Jute (y) was being influenced by the different factors—prices of raw jute (x_1), business conditions at the appropriate markets (x_2) and prices of related goods (x_3)—figures were collected for 10 years 1920—1938 and the values of y , x_1 , x_2 , x_3 noted. But those values instead of being given as absolute values have all been quoted in the Table below with reference to the year 1920 as base :

Develop the demand curve to show how the different factors influence separately and jointly foreign demand for raw Jute.

Can you suggest why relative values and not absolute values have been used ?

Year	y	x_1	x_2	x_3	Year	y	x_1	x_2	x_3
1920	—	—	—	—	1930	0.77	0.69	0.83	0.71
1921	0.99	0.60	0.77	0.41	1931	0.95	0.73	0.82	0.85
1922	1.24	1.12	1.26	1.02	1932	0.96	0.78	0.77	0.85
1923	1.14	0.91	1.20	1.01	1933	1.33	0.87	1.19	1.00
1924	1.08	1.11	0.93	1.13	1934	1.00	0.90	1.09	0.94
1925	0.93	1.07	1.11	1.17	1935	1.02	1.16	1.16	0.97
1926	1.09	0.85	1.06	0.80	1936	1.03	1.00	1.18	0.80
1927	1.26	0.87	0.99	1.06	1937	0.91	1.11	1.10	1.00
1928	1.01	1.04	1.04	1.02	1938	0.92	0.90	0.73	0.91
1929	0.90	0.95	1.11	0.80					

3. The following data give the commercial interest rates in U. S. A. per quarter during the year 1900—1913.

Calculate the seasonal index by assuming a straight line trend and from the trend and the seasonals, build up the normal. Finally calculate their cycle (c) and express them as percentage deviation cycle. What further calculation would you suggest to make this cycle comparable to any other cycle.

	1900	1910	1911	1912	1913
Q1.	3.8	4.7	4.1	3.9	5.3
Q2.	3.9	4.8	3.6	4.2	5.7
Q3.	4.2	5.6	4.2	5.1	6.0
Q4.	5.5	5.3	4.3	6.0	5.8

4. With a view to determine the meteorological factors which give rise to significant variation in yield per acre of cotton, following observations were recorded:

Years (1)	Yield ratio (2)	Sept. rainfall (3)	Nov. Mx. Temp. (4)
1912	107.7	67.1	101.0
1913	118.0	117.7	97.2
1914	108.2	81.1	102.5
1915	89.3	81.3	101.8
1916	102.2	153.1	97.6
1917	157.3	92.8	91.5
1918	122.2	173.0	97.6
1919	121.0	87.0	99.7
1920	108.8	143.0	89.1
1921	64.1	68.4	104.7
1922	90.7	21.6	98.1
1923	77.2	17.0	98.1
1924	75.6	103.5	105.5
1925	112.8	133.4	98.6
1926	122.0	98.8	97.1
1927	125.7	157.9	101.9
1928	119.5	167.4	96.0
1929	120.4	88.6	100.7
1930	101.4	137.9	98.5
1931	87.3	78.6	99.1

Calculate the multiple correlation coefficient of yield with the meteorological factors.

Calculate the forecasting formula for the yield of cotton and compare the estimated yield figures with official figures.

COMPUTER'S CERTIFICATE EXAMINATION,—MAY 1947

PART IA, SECTION I

Time allowed : 2½ hours.

(No calculating machines will be provided. No credit will be given for incomplete scribbly answer).

1. Evaluate :

(i) $\frac{1-003720}{-0163} \times 3-713 \times 0-00$

(ii) $137-1 \div 2-070 - 0-31395 \div 0-10 \times 13-7$

(iii) $\frac{(1-2y)}{(1-x)(1-y)}$, when x and y take values 0, 0-2, 0-4, 0-6 and 0-8.

2. Indicate which of the following statements are likely to be wrong :

- (1) the present population of India is nearly 4000 lakhs
- (2) about 90% of the people of Madras Presidency are Muslims
- (3) the square root of 3 is 2-1437
- (4) the square of 5-43 is 29-48
- (5) in the United Provinces the number of male children born every year is nearly twice the number of female children born
- (6) wealthy families spend on food a larger proportion of their income than poor families
- (7) deaths in Calcutta are more frequent in winter than in summer
- (8) the average area of land cultivated by a family in Bengal is less than 15 bighas
- (9) the temperature in Cawnpore during Summer often rises above 105°C.
- (10) the distance from Calcutta to Bombay is more than 3000 miles
- (11) in a certain village 47 per cent of the total area is under paddy, 26 per cent is under jute, 39 per cent is waste land and the rest is lying fallows.

or

'A' is younger than 'R' by 2 yrs. 3 months, while 'C' is older than 'D' by nine months and younger than 'A' by 1½ months only. 'E' whose age is just one month more than 'F' has got the same age as 'G'. If 'G' is senior to 'B' by 6½ months and has just attained 18 years 7 months, what is the age of each, A, B, C, D, E, F ?

3. Copy out the following table correcting any mistakes you can detect.

Statistics of messages by Government Telegraphs.
(figures in thousands)

Year	State		Private		Total	
	Number	Value Rs.	Number	Value Rs.	Number	Value Rs.
1929-30	1121	1788	15144	16334	16265	18122
1930-31	1226	2129	14120	15038	16355	17167
1931-32	1197	2168	13304	14216	14591	16384
1932-33	852	1691	12391	14222	13243	15823
1933-34	803	1511	72456	14202	13239	15713
1934-35	834	1430	14080	14558	14923	15997
1935-36	844	1460	14152	94351	14096	15820
1936-37	972	163	14686	14397	16558	16030
1937-38	920	1051	12822	12790	13742	14441
1938-39	958	1560	12584	12285	13542	13845
Total	19834	10940	135767	142393	245474	169343

Item of expenditure	Amount Rs.-as.	Item of expenditure	Amount ¹ Rs.-as.
Family (5) Size of Family—7		Family (7) Size of Family—6	
Rice & Pulses	40- 2	Pulses	3- 8
Sweater	6- 4	Milk	12-13
Other Food	35- 8	Rice	40- 9
Other Clothings	12- 2	Rent	35- 7
Rent & Taxes	10- 8	Total Clothings	24-12
Others	12- 0	Other Foods	40- 2
Total	116- 8	Total	157- 3
Family (8) Size of Family—3		Family (8) Size of Family—2	
Coal etc.	8- 8	Medical	12- 0
Other Foods	42- 7	Cigarettes	14- 4
Cinema	6- 2	Tea	8- 8
Toilets	8- 4	Furniture	23- 5
Genji	2- 8	Rent	24- 6
Vegetables	16- 0	Rice	12- 0
Fish & Meat	8- 0	Other Foods	38- 4
Other Clothings	4- 6	Clothings	11- 0
Total	96-12	Total	143-11
Family (9) Size of Family—4			
Coal & Light	6- 3		
Rent	12- 8		
Total Food	65-12		
Shirt	11- 8		
Total	95-15		

5. The following are the records of output in the month of April of copying in an office with 27 computers :-

Two workers attended for 23 days each and copied a total of 21,000 items, while five copied 32,480, attending 25 days each, thirteen copied 47,560, attending 18 days each, six copied 38,212 attending 16 days each and one copied 15,200 attending 26 days.

Find out (a) the average rate of output in copying from the above (i) per day per worker, (ii) per month per worker, and also (b) the average attendance per worker per month.

Or,

If 1020 three-digit figures in copying takes up a day and if again 640 of such figures can be added per day by one worker, what is the number that may be expected to be both copied and added in a day, if these are done together?

COMPUTER'S CERTIFICATE EXAMINATION,—MAY 1947

PART IA : SECTION II

Time allowed : 3 hours

(No machines allowed. No credit will be given for incomplete or scrappy answer.)

1. (a) You are given a table showing cases of sickness per 100 persons (classified according to age) in five departments of the building trade. Exhibit the material in a graphical form. Write a note on the feature of the data.

Age Groups	15-24	25-34	35-44	45-54	55-64
Trade					
Underground labourers	56.2	63.6	79.3	93.2	92.9
Cementers	46.8	48.7	50.0	73.9	80.9
Painters	37.3	43.8	44.6	50.8	76.5
Masons	36.4	35.1	38.2	43.7	45.1
Mason's Labourers	67.8	64.5	74.4	80.0	85.5

(b) By a graphical method, find the area of the triangle formed by

$$x - 2y + 12 = 0$$

$$x + y + 3 = 0$$

$$5x - y - 21 = 0$$

2. (a) If $a = 34562.73$ and $b = 28347.012$, find the value of the square root of $(a^2 - b^2)$. (With the help of logarithm tables)

(b) Find the value of the series correct to four decimal places

$$1 + \frac{1}{2.5} + \frac{1}{3.5^2} + \frac{1}{4.5^3} + \dots$$

(c) The following table gives the values of some functions for two given values of the argument. Find by simple interpolation the values of the functions corresponding to the value 1.36572 of the argument.

Argument (x)	1.365	1.366
$\psi(x)$	-0.09789	-0.09083
$\log \Gamma(x)$	0.9492017	0.9402194
e^{-x}	0.2553806	0.2551254
$\text{Sinh } x$	1.83017	1.83220

3. The following is the frequency distribution of marks in English at the school leaving certificate examination of U.P. in 1919.

Class-limits	Frequency
1-5	2
6-10	8
11-15	23
16-20	61
21-25	131
26-30	246
31-35	304
36-40	441
41-45	397
46-50	333
51-55	186
56-60	107
61-65	38
66-70	14
71-75	4
76-80	2
81-85	0

- (i) Draw a histogram.
- (ii) Calculate the mean and standard deviation of the above frequency distribution. Find approximately the number of boys getting marks above $(\text{mean}) + 2$ (standard deviation).

COMPUTER'S CERTIFICATE EXAMINATION,—MAY 1947

PART IB : SECTION I

Time allowed : 3 hours

(Machines and tables allowed. No credit will be given for incomplete or scrappy answer.)

1. The following table shows the prices of twelve items of expenditure for the years 1939 and 1946. Calculate the Cost of Living Index for 1946 with 1939 as base using the figures in column six as weights.

No.	Commodity	Units	Average Market Price in Rupees		Weights
			Price in Rupees		
			1939	1946	
1	Rice	Seer	0.14	0.40	11.67
2	Atta	"	0.13	0.34	3.24
3	Fish	"	0.57	2.10	6.51
4	Meat	"	0.60	2.06	2.29
5	Pulses	"	0.18	0.78	2.28
6	Eggs	Sooro	0.53	2.20	0.53
7	Potato	Seer	0.12	0.64	2.08
8	Other vegetables	"	0.10	0.30	4.28
9	Mustard oil	"	0.48	1.43	3.42
10	Ghee	"	1.09	5.85	4.24
11	Sugar	"	0.29	0.84	1.88
12	Coal	Md.	0.58	1.37	2.61

following figures give the weekly traffic index for the years 1944 to 1946 of an Indian Railway.

28.3	15.4	20.0	5.9	16.8	21.7	7.2	19.4	12.9	31.0	36.1
27.8	20.0	10.0	28.7	14.5	22.2	35.5	41.8	24.0	31.9	15.0
26.2	10.0	24.5	56.5	14.1	22.5	34.8	15.6	22.4	19.4	15.6
26.0	36.9	23.9	68.3	32.2	21.0	32.7	15.6	33.7	19.7	13.1
41.0	39.0	5.8	17.8	32.9	6.5	19.8	13.6	32.1	36.6	12.7
19.1	12.9	33.1	16.7	22.8	19.6	15.7	19.2	7.4	36.3	
20.5	7.3	33.1	15.7	34.1	20.0	13.1	19.5	23.8	40.4	
19.2	7.2	22.2	13.7	32.5	37.5	12.7	13.6	22.8	65.0	
19.5	35.6	21.8	13.0	31.6	37.0	19.3	13.0	24.1	15.1	
13.5	34.0	40.8	24.3	32.5	15.1	20.7	7.2	23.4		

(a) Arrange the figures in a frequency distribution form taking five units as length of class intervals.

(b) Calculate σ , β_1 , and β_2 .

(c) Prepare a cumulative frequency table and find the semi-interquartile range. Also find the ratio of semi-interquartile range to standard deviation.

3. In a plebiscite for eliciting public opinion regarding the partition of Bengal, the verdict of voters are classified under the heads yes, no, indifferent. The voters are classified by communities (Hindus, Muslims, Christians and others) and also by districts (1, 2, 28). In each district again the voters are differentiated as rural and urban.

Prepare a skeleton table to represent the result of the plebiscite in a neat form.

4. Fit a curve of the form $y = a + bx + cx^2$ to the following data.

Year		Year		Year	
(x)	(y)	(x)	(y)	(x)	(y)
1920	87	1926	606	1932	1520
1921	198	1927	1000	1933	2119
1922	247	1928	1001	1934	1948
1923	326	1929	1314	1935	2527
1924	422	1930	1179	1936	2976
1925	583	1931	1573		

COMPUTER'S CERTIFICATE EXAMINATION,—MAY 1947

PART 1B: SECTION II

Time allowed: 3 hours

(N.B. Credit will be given for neatness and precise answers.)

1. (a) The following table gives an observed frequency distribution of the heights of the students of a certain age:

Height in inches	Frequency	Height in inches	Frequency
58	2	68	258
59	1	69	231
60	7	70	118
61	10	71	99
62	20	72	38
63	40	73	15
64	74	74	2
65	142	75	1
66	220	76	0
67	230	77	1

(In the column for height the mid-point of the class-interval is indicated.)

Assuming the mean and the standard deviation, worked out from the above table, to be 67 inches and 2.4 inches respectively, calculate the expected frequencies if the distribution is taken to be normal. State whether the discrepancies between the calculated and the observed frequencies are statistically significant.

(b) Are all symmetrical distributions normal?

2. (a) A contractor purchased a very large number of a certain type of copper sheets from a manufacturer. The contract specified that the mean thickness of the sheets should be 0.022 inch. The contractors measured the thickness of a random sample of 100 sheets and found that the mean and the standard deviation in the sample were 0.020 inch and 0.003 inch respectively. Had the contractor any valid reason to complain?

(b) The heights of two groups of soldiers consisting of 100 in each group, were measured and the following results were obtained.

Group	Mean	Standard Deviation
I	67.51 inches	2.20 inches
II	62.24 "	2.25 "

If the differences in the average height indicate racial differences, do you think that the two groups of soldiers came from different races?

3. Draw bar-diagrams to represent any one of the three following enquiries (a), (b) or (c).

(Actual figures must also be quoted at the bottom of the diagram in a tabular form, mentioning also the source of the data.)

(a) Export of raw jute from India for any three consecutive years showing the following details: export to United Kingdom, export to Germany and export to the rest of the countries.

(b) Production of wheat in British India for any three consecutive years, showing the following details: production in the Punjab, production in the United Provinces, production in the rest of British India.

(c) Population in Bengal (excluding the States) for any three consecutive periods at the intervals of 10 years, showing the following details: population in the Burdwan Division, population in the Dacca Division and population of the rest of Bengal.

4. The following table gives the population of the United States at intervals of 10 years from 1820 to 1870 :—

Years	Population (in millions)	Years	Population (in millions)
1820	9.6	1850	23.2
1830	12.9	1860	31.4
1840	17.1	1870	38.6

(a) Calculate the population in 1853.

(b) Suppose the population P (in millions) is given by the equation $P = 5.3(1.3167)^t$ where $t = (x - 1800)/10$, and x represents the year (in A.D.), calculate the values for the population for the years shown in the table.

(c) Draw a graph showing the actual and the calculated values of the population for the different years.

COMPUTER'S CERTIFICATE EXAMINATION,—MAY 1947
(Supplementary)

PART 1B : SECTION 1

Time allowed : 3 hours

(Machines and tables allowed. No credit will be given for incomplete or scrappy answer.)

1. Scrutinize the following calculating sheet using some methods of check.

x	y	x^2	y^2	xy
2.5	3.5	6.25	12.25	8.75
2.5	3.0	6.25	9.00	7.50
3.0	5.0	9.05	25.00	15.00
4.0	7.0	16.00	47.00	28.00
4.5	5.5	20.39	30.25	24.05
5.0	8.0	25.01	64.00	40.00
5.5	8.0	30.25	64.00	40.00
6.0	10.0	36.00	10.00	60.00
6.0	7.0	36.00	49.00	41.25
6.5	10.5	41.79	107.25	68.25

2. Find the arithmetic mean and the standard deviation from the following data.
Weekly earnings of stenographers and secretaries,

New York City, May 1937.	
Weekly earning	Number of stenographers and secretaries.
\$ 10.00 to \$ 19.99	363
\$ 20.00 to \$ 29.99	2,177
\$ 30.00 to \$ 39.99	1,984
\$ 40.00 to \$ 49.99	579
\$ 50.00 to \$ 59.99	214
\$ 60.00 to \$ 69.99	69
\$ 70.00 to \$ 79.99	17
\$ 80.00 to \$ 89.99	5
\$ 90.00 to \$ 99.99	4
\$ 100.00 to \$ 109.99	1
	5,413

Find also the median and the mode of the above distribution and verify whether the relation $(\text{Mean} - \text{Mode}) = 3(\text{Mean} - \text{Median})$ approximately holds good or not. Calculate also the measure of skewness defined by $\frac{\text{Mean} - \text{Mode}}{\text{Standard deviation}}$.

3. The following figures give the yield of Alfalfa in tons per acre (x) and the labour spent on cultivation in hours per acre (y).

Hours (y)	Tons (x)							Total
	1.4-1.7	1.8-2.1	2.2-2.5	2.6-2.9	3.0-3.3	3.4-3.7	3.8-4.1	
18-19					2	2	3	7
16-17			1	2	9	7	1	20
14-15			1	10	23	1		35
12-13		1	5	27	12			45
10-11		3	17	15	7			42
8-9	1	6	11	5	1			24
6-7	5	3	4	1	1			14
4-5	2	1	2					5
Total	8	14	41	60	55	10	4	192

- (1) Find the regression equation of x on y , and (2) find the estimated yield per acre when the labour spent per acre is 14.5 hours.

4. The figures below give the yield of sugarcane by blocks and varieties in a randomised block experiment.

Variety	Block					
	1	2	3	4	5	6
A	248	217	227	210	218	215
B	245	217	239	210	205	219
C	238	228	205	191	224	211
D	254	223	189	180	216	209
E	249	221	226	242	245	216
F	225	212	194	211	202	215

Find by analysis of variance whether there is any significant difference between the varieties.

COMPUTER'S CERTIFICATE EXAMINATION,—MAY 1947

(Supplementary)

PART 1B : SECTION II

Time allowed : 3 hours

(Machines and tables allowed. No credit will be given for incomplete or scrappy answer.)

1. The following table gives the frequency distribution of heights of 1000 students.

Height in cms.	No. of students	Height in cms.	No. of students
155-157	4	179-181	125
158-160	8	182-184	92
161-163	26	185-187	60
164-166	53	188-190	22
167-169	89	191-193	4
170-172	146	194-196	1
173-175	188	197-199	1
176-178	181		

Mean height is 175.3 cms. and standard deviation is 6.58. Fit a normal curve to the above frequency distribution and test the goodness of fit.

2. (a) Two groups of rats (89 in each group) were weighed and found to have the following mean weights and standard deviations:

	Group I	Group II
Mean weight (grams)	214.9	167.3
S.D. of weight (grams)	52.0	20.5

Is the difference in weight between the two groups significant?

(b) An analysis of the wage payments before and after the settlement of a wage dispute shows that the average pay has increased from Rs. 25 a month to Rs. 29.5 a

month, and that the standard deviation has changed from Rs. 3 to Rs. 4-5. The pay is not only higher after the settlement but also more uniform. Do you agree ?

3. Draw bar diagrams to represent the following facts:—"India exported in 1936-37, 648,495 tons of cotton, in 1937-38, 413,123 tons and in 1938-39, 405,067 tons. Of the total quantity exported in 1936-37, 111,147 tons were sent to U.K., 433,223 tons to Japan and the rest to other countries. In 1937-38, 70,554 tons were exported to U.K., 242,695 tons to Japan and the rest to other countries. In 1938-39, 73,397 tons were exported to U.K., 216,248 tons to Japan and the rest to other countries".

Arrange the above figures in a table.

Name the publications from which these data could be collected.

4. The production of artificial silk from 1923 to 1930 in a certain country is given below in millions of pounds.

Year	Production
1923	95
1924	127
1925	171
1926	233
1927	323
1928	452
1929	654
1930	918

Calculate the expected production figures taking as the graduating formula

$$y = 60.05 (1.417)^x$$

where y = production ; x = number of years that have elapsed since 1922.

Show graphically the discrepancy between the observed and expected figures. Do you consider the agreement between the observed and expected figures satisfactory ?

COMPUTER'S CERTIFICATE EXAMINATION,—MAY 1947

PART IC: SECTION I

Time allowed : 3 hours

(Machines and tables allowed. No credit will be given for incomplete or scrappy answer.)

1. (a) Money invested in National Savings Certificate increases by 50% within 12 years. Assuming uniform compound interest, find the rate of interest per annum. Assuming the same rate of compound interest to hold good, within how many years the amount should have doubled ?

(b) In the Binomial Expansion $\left(3x^2 + \frac{5}{8x}\right)^8$ find the value of the term independent of x .

2. Find the three partial correlations and the three multiple correlations from the data :

Serial No.	x	y	z	Serial No.	x	y	z
1	105	3.0	10.0	13	114	6.8	42.5
2	163	6.9	47.0	14	86	6.1	20.0
3	130	6.3	35.3	15	133	6.6	54.0
4	118	4.2	31.5	16	108	5.4	28.0
5	98	4.6	15.0	17	75	4.7	11.7
6	108	8.1	48.5	18	129	5.7	26.5
7	99	4.8	18.5	19	171	6.6	5.9
8	161	7.1	59.8	20	104	3.7	11.3
9	84	5.6	27.8	21	135	6.5	37.5
10	82	5.5	19.0	22	165	6.4	33.0
11	114	4.8	20.0	23	92	4.0	22.2
12	94	5.8	20.7	24	139	4.4	40.7
				25	86	7.5	47.0

3. In a certain experiment on manurial treatment with Nitrogen (n), Phosphorus (p) and Potassium (k) the following layout was arranged :

p	n	np	k	nk	(1)	npk	pk
18.8	12.2	18.3	15.2	11.4	11.5	19.4	18.9
n	nk	pk	npk	p	k	np	(1)
12.9	7.3	17.4	17.2	19.7	12.0	10.0	15.6
nk	np	n	p	(1)	npk	pk	k
10.7	17.5	10.4	18.0	9.8	16.6	17.5	14.3

In the above table (1) denotes no treatment, n, p, k denote only one kind of manure, nk, pk, denote two kinds of manures, and npk all manures. Draw up an Analysis of Variance Table to test the significance of the main effects and interactions of the manures.

4. Correct all the mistakes in the following tables :

Jagadlal Cost of Living Index, 1942 and 1945 with 1941 as base

1942 Cost of Living Index						
Blocks	Investigators					mean ± S. D.
	I	II	III	IV	V	
1	120	122	122	124	124	122 ± 0.75
2	122	120	121	124	118	121 ± 1.00
3	121	123	121	121	121	121 ± 0.40
4	119	122	120	123	121	121 ± 0.71
5	121	124	124	124	124	123 ± 0.80
Mean ± S. D.	121 ± 2.00	121 ± 2.04	121 ± 1.96	123 ± 2.17	122 ± 2.32	122 ± 0.39

1945 Cost of Living Index						
1	270	280	275	271	270	276 ± 2.0
2	263	265	284	260	269	269 ± 3.0
3	264	274	278	272	269	272 ± 2.4
4	263	274	271	269	263	275 ± 6.2
5	260	279	266	274	284	279 ± 6.2
Mean ± S. D.	264 ± 2.00	273 ± 4.51	279 ± 4.52	276 ± 4.87	272 ± 5.21	274 ± 1.7

COMPUTER'S CERTIFICATE EXAMINATION,—MAY 1947

PART IC : Section II

Time allowed : 4 hours.

(Machines and tables allowed. No credit will be given for an incomplete or scrappy answer.)

1. By taking logarithms, and solving by the use of determinants, the resulting linear equations, determine the values of x , y , z which satisfy the equations

$$2^x 3^y 4^z = 290$$

$$3^x 4^y 5^z = 2105$$

$$4^x 5^y 7^z = 11225$$

2. Supply the missing value in the following two-way table

		X				
		2	3	4	5	6
Y	4	54.00	50.07	47.94	46.56	45.63
	6	32.76	29.34	?	26.26	25.41
	8	25.95	22.77	21.03	19.89	19.11

3. Analyse the results of the following Latin Square experiment and present your results in the form of a neat analysis of variance table.

E 633	B 627	F 652	A 390	C 504	D 416	3122
B 480	C 475	D 415	E 488	F 571	A 282	2720
A 384	E 481	C 483	B 422	D 334	F 646	2750
F 620	D 448	E 505	C 439	A 323	B 384	2719
D 462	A 432	B 411	F 617	E 594	C 466	2672
C 500	F 505	A 259	D 366	B 320	E 420	2376
3078	2868	2725	2722	2652	2614	16659

Can all the treatments be regarded equally effective ?

Is the treatment F better than A ?

4. The following figures show the net earnings (in units of 1 million dollars) of Public Utility concerns over a period of 6 years :

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1926	66.9	61.6	60.7	59.5	51.9	55.7	49.2	49.9	56.9	60.9	65.8	73.0
' 27	74.4	66.9	65.4	64.0	61.2	59.2	53.0	53.6	61.0	65.3	70.2	78.9
' 28	70.0	74.3	72.8	68.0	67.7	67.5	62.3	61.8	68.2	73.7	81.4	91.0
' 29	92.0	80.0	85.0	83.0	82.5	70.0	71.0	73.0	80.0	83.0	92.0	100.0
' 30	92.0	90.0	88.0	89.5	86.0	83.0	70.6	71.5	80.8	84.1	88.3	89.0

Calculate the index of seasonal variation in the net earnings.

5. Indicate the chief sources of statistics available for either of the following :

(1) Agriculture

(2) Industrial Production & Industrial Labour Conditions,

and indicate in broadest outlines the nature of information available.