1982-83 | B121

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications, 1982-83
Part I

Descriptive Statistics (Theory and Practical)
PERIODICAL EXAMINATION

Date : 8.9.1982. Maximum Marks : 100 Time: 2 hours

Note : Do as many questions as you can do. The paper carries 120 marks but the maximum you can score is 100 marks. Marks allotted are given at the end of each question.

1. A statistical enquiry yielded the following data of land holdings (in acres) for 60 households.

5, 7, 15, 10, 6, 15, 6, 4, 3, 2, 0, 1, 16, 11,
17, 12, 13, 15, 3, 2, 51, 2, 4, 35, 40, 60, 50, 45,
62, 65, 45, 20, 18, 15, 19, 20, 17, 10, 11, 14, 15, 25,
30, 35, 40, 45, 50, 55, 60, 65, 18, 19, 18, 29, 49, 21,
11, 19, 18, 31,

(i) Construct a grouped frequency distribution with suitable no. of classes.

(ii) Draw frequency polygon, frequency curve, cumulative frequency polygon and cumulative frequency curve and Histogram for the above data. Obtain the values of quartiles with the help of ogive. (10+20) = [30]

2.(i) What do you mean by measures of central tendency and dispersion ? Name some important measures and define them.

(ii) For the data in Q.1 (you may use grouped frequency distribution) obtain the values of mean, median, mode, first and third quartiles, variance, standard deviation and coefficient of variation. (10+20)=[30]

3.(i) Define r^{th} raw moment and central moment and express the later in terms of the former.

(ii) Obtain first four central moments for the data in Question 1. (10+10) = [20]

P.T.O.

4.(i) What do you mean by skewness and kurtosis ?

(ii) Give some measures of skewness and kurtosis and obtain their values for the data in question 1.

(10+10)=[20]

5.(i) What do you mean by sampling and non-sampling errors ? Mention major types of non-sampling errors.

(ii) Indicate major stages of a statistical investigation and the main steps required within each stage. What precautions should be taken against the non-sampling errors ?

(5+15) =[20]

1982-83 | E131

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications:82-83.
PART I

Economic Statistics(Theory and Practical)
PERIODICAL EXAMINATION

Date : 22.9.82. Maximum Marks : 100 Time : 2 hours

Note : Answer all the questions. Marks allotted to each question are given in brackets.

1. Discuss the index number problem with reference to a price index number. State the well-known index number formulae distinguishing clearly between the "aggregative" type and "average" type indices. Is it possible to interpret the Laspeyre's price index formula both as an weighted average of price relatives and as an aggregative price index ?
(5+12+ 5)

2. Discuss the problems of constructing a series of price index number over a long period of time. Do you think that the chain base method is more suitable than the fixed base method for the construction of such a series ? Give reasons for your answer.
(5+8)

3. The following table gives the sums $\Sigma p_{it}\, q_{it}$ where $t, t' = 1961, 1962, 1963, 1964$.

t \ t′	1961	1962	1963	1964
1961	356	331	378	416
1962	395	365	424	460
1963	448	423	482	529
1964	465	432	495	542

Calculate the following index numbers.

P.T.O.

Q.No.3 contd...

(i) Laspeyre's price index numbers for 1964 with 1962 as base and for 1963 with 1961 as base.

(ii) Paasche's quantity index numbers for 1962 with 1964 as base and for 1961 with 1963 as base.

(iii) Laspeyre - type chain base price index number for 1964 with 1961 as base.

(4+4+7)

1982-83	B112

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications: 1982-83.

Part: I

Probability

PERIODICAL EXAMINATIONS

Date: 15.10.82. Maximum Marks: 100 Time: 3 hours.

Note : The whole paper carries 110 marks but the maximum you can score is 100. Answer all questions.

1.(a) Show that B(n,p) distribution approaches a Poisson distribution as $n \to \infty$ if $np = c$ (a constant). [10]

(b) Define the probability generating function of a nonnegative integer valued random variable. Find the p.g.f. of the Poisson distribution with parameter λ. Prove using p.g.f. that Poisson distribution is reproductive. [3+7+6]

(c) Write down the density function of N(3,4). Find the density function of $-3X+2$ if $X \sim N(3,4)$. [3+6]

2. An urn contains 5 red, 3 blue and 2 green balls. Two balls are drawn at random from the urn with replacement. Let X (respectively, Y,Z) be the number of red (respectively, blue, green) balls obtained.

(a) What is the name of the distribution of (X,Y,Z) ? Write its (discrete) density function. [3+5]

(b) Write down the table (no proof) giving the density function of (X,Y). Obtain the marginal density functions of X,Y. What are the names of the distributions of X,Y? [8+5+3]

(c) Find the density function of X+Y. What is the name of this distribution ? Are X,Y independent ? [6+3+5]

3. Let (X,Y) be a continuous random vector with density function

$$f(x,y) = \begin{cases} c\,(y-x)^2 & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

P.T.O.

3.(a) Find the value of c. [6]

(b) Find the density function of X and the density function of X^2. [6+6]

(c) Find the conditional density function Y given $X = \frac{1}{2}$. [4]

(d) Find $P(0 < X < \frac{1}{2}, \frac{1}{2} < Y < 1)$, $P(X + Y \leq 1)$. [10]

(e) Are X, Y independent ? Why ? [5]

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications: 1982-83

PART: I

Descriptive Statistics(Theory and Practical)

PERIODICAL EXAMINATION

Date: 20.10.82. Maximum Marks: 100 Time: 2 hours.

Note: Attempt as many questions as you can. The paper carries 120 marks but the marks obtained by you are not to exceed 100. The marks allotted are specified at the end of each question.

Prove or disprove the following statements :

(i) Let μ be the mean of a Binomial distribution $B(X,n,p)$, X being the total number of successes in n Bernouli trials with p as probability of success. Then the variance of X is given by $\mu(n-\mu)/n$.

(ii) Let μ and σ denote the mean and s.d. of a Poisson distribution. Then

$$\mu = \sigma^2 = \mu_3 \quad \text{and} \quad \mu_4 = \sigma^2(1+3\sigma^2) = \mu(1+3\mu)$$

and $\beta_2 = 3+\beta_1$.

(iii) Two variates X and Y are uncorrelated if and only if they are independent.

(iv) In case of normal distribution, all odd order central moments are zero.

(v) Total regression coefficients and correlation coefficient are in G.P.

(5+5+10+5+5) = [30]

In the Higher Secondary Examinations 20% of the students fail. In the case of 400 students appearing in the above examination find the probabilities that :

(a) At least 80 students will fail.

(b) Atmost 90 students will fail.

(c) No. of failure cases will lie between 60 and 100.

(5+5+5)=[15]

P.T.O.

3. Five seeds are sown in a plot. Some of the seeds may germinate and grow into plants while some may not. Let X denote number of seeds germinating and growing into plants. The experiment is repeated in 100 plots and the following data is obtained :

X	0	1	2	3	4	5
frequency (No.of plots)	5	5	15	40	30	5

(i) Fit a binomial distribution to the above data.

(ii) Find the probable number of plots in which at least three seeds would germinate and grow into plants.

(iii) Find the mode and variance of the distribution.

(10+5+5)=[20]

4. The following data gives profits (after tax) for five factories in a particular year and the extra amount spent on due to raise in wages, bonus and social welfare of all the workers in these factories for that year.

Factory No.	1	2	3	4	5
X	25	30	10	50	70
Y	3	2	1	4	6

Where X : Profits, after tax, in Rs. lakhs.

Y : Extra amount spent on raising wages, bonus and social welfare of workers in Rs. lakhs.

(i) Obtain correlation coefficient between X and Y and interpret the result.

(ii) Obtain regression line of y on x and that of x on y.

(iii) Obtain theoretical values of Y corresponding to various values of X in above table.

(10+5+5)= [20]

(Contd.....3).

5. A machine is used for a particular task during the day and for another task during the night. Let

X : no. of times the machine breaks down during the day.

Y : no. of times the machine breaks down during the night.

The joint probability distribution $P(X = x, Y = y)$ is found to be

Y \ X	0	1
0	0.2	0.3
1	0.08	0.12
2	0.1	0.2

(i) Obtain the marginal and conditional distributions. Are X and Y independent ?

(ii) Obtain $P(X < 1)$, $P(X > 0)$, $P(Y > 1)$, $P(1 \leq Y \leq 2)$, $P(1 \leq Y \leq 2 \mid 0 < X \leq 2)$

(iii) In how many percent cases there would be more break downs in the nights than during the days.

(10+5+5)=[20]

6. Monthly wage-distribution of 200 workers in a factory are obtained as

wages (in Rs.)	No. of workers
100 - 200	20
200 - 500	25
500 -1000	75
1000 -1500	60
1500 -2500	20

(i) Obtain mean, median, mode , first and second quartiles of the above data.

(ii) Obtain coefficient of variation for the above data.

(10+15)=[15]

1982-83 | E132

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications: 1982-83
Part I

Time Series and Official Statistics (Paper III)

PERIODICAL EXAMINATION

Date: 10.11.82. Maximum Marks: 100 Time: 2 hours.

Note : Answer all the questions. Answers should be brief and to the point. Marks allotted to each question are given in brackets.

1.(a) What is a time series ? Describe with suitable illustrations the different components that may be present in an observed time series.

(b) A time series of number of movie-goers in Calcutta is recorded daily during the month of September, 1982. Identify the systematic components that may be present in the series.

[15+10]=[25]

2.(a) Describe different trend models which can be used to forecast a time series.

(b) Interpret the meaning of the parameters for each model and mention the possible range of values for the parameters.

(c) Describe the different methods of selection (or identification) of an appropriate trend model for a given time series which contains trend and irregular components only.

(d) What trend model should be used to analyse the following time series:

Time	1	2	3	4	5	6	7	8
Y_t	5	8	13	20	29	40	53	68

Can you estimate the parameters of the identified model without doing any long calculation ?

[10+10+10+10]=[40]

P.T.O.

3. The estimated linear trend equation for the logarithm of the Index Production in the chemical and chemical product industries, 1947-65 in the U.S.A. over time is

$$\text{Log } Y_t = 1.93427 + 0.03390t$$

(a) Find out the corresponding exponential trend to this equation and hence obtain the rate of growth of production in the chemical and chemical product industries.

(b) Forecast the trend value for the year 1972, where t=9 for the year 1965.

[10+5]=15

4. On the basis of the data below, prepare (a) a long-term forecast of sales of Sigma Products, Inc., i.e. for 1968 ; and (b) a forecast for the June, July and August season of 1968. Interpret your forecasts.

SIGMA PRODUCTS, INC.
Seasonal Index

Month	Index	Month	Index
January	75.0	July	135.6
February	77.3	August	122.5
March	82.5	September	101.0
April	84.0	October	101.3
May	86.2	November	102.4
June	129.7	December	102.5

$$Y_c = 543.2 + 20.1\ t$$

Origin : July 1, 1961

t in terms of years.
Y in terms of billions of dollars.

[10+10]=20

1982-83 | E232

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications: 1982-83
Part I

Paper III

Economic Statistics (Index Number and National Income Accounting)

SEMESTRAL -I EXAMINATION

Date: 20.12.82. Maximum Marks: 100 Time: 2 hours.

Note : Answer question 5 and any three from the rest. Marks allotted to each question are given in brackets.

1. Explain the 'factor reversal' test. Examine which of the well-known index number formulae satisfy this test. Discuss the importance of this test. [22]

2. Discuss the different methods of construction of index of industrial production. Describe briefly how the official index of industrial production is constructed in India. [22]

3. Define national income and bring out clearly the distinction between (i) national income at market prices and national income at factor cost (ii) national income at current prices and national income at constant prices (iii) national income and domestic income. [22]

4. Write short notes on any <u>two</u> of the following :-
   (i) Chain base versus fixed base index numbers.
   (ii) Choice of <u>base period</u> and <u>items</u> for the construction o price index numbers.
   (iii) Real national product and its measurement.

[22]

P.T.O.

5.(a) From the following data calculate index numbers of wholesale prices of (i) machinery and transport equipment and (ii) all commodities taken together, for the last week of 1970-71 with 1961-62 as base. [9]

Group/Subgroup	group weight	weight in the group	index number for the last week of 1970-71
(1)	(2)	(3)	(4)
1. Food articles	413	-	199.8
2. Liquor and tobacco	25	-	184.9
3. Fuel, power etc.	61	-	162.7
4. Industrial raw materials	121	-	191.0
5. Chemicals	7	-	189.5
6. Machinery and transport equipment.	70	-	-
6.1. Electrical Machinery	-	241	149.2
6.2. Non-electrical machinery	-	403	160.7
6.3. Transport equipment	-	266	136.0
7. Manufactures	294	-	160.4

(b) The following data (in Rs.crores) relate to the Indian economy for the year 1965-66 :

(i) Net domestic product at factor cost	= 20786
(ii) Gross national product at market prices	= 23921
(iii) Transfer payments to individuals	= 266
(iv) Net factor income from abroad	= -165
(v) National debt interest	= 175
(vi) Indirect taxes less subsidies	= 2082
(vii) Income from domestic product accruing to government	= 345
(viii) Other current transfers from the rest of the world	= 72

Using the above information calculate (a) net national product at market prices (b) gross national product at factor cost (c) depreciation (d) private income. [25]

1982-83 | E233

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications, 1982-83
PART I

Paper III
Economic Statistics(Time Series and Official Statistics).

SEMESTRAL-I EXAMINATION

Date: 22.12.82. Maximum Marks: 100 Time : $2\frac{1}{2}$ hours.

Note : Answer all the questions. Marks are given in brackets. Answers should be brief and to the point.

GROUP A

1.(a) Distinguish between the additive and multiplicative models of time series in a set up where the trend, seasonal fluctuations and an irregular component is assumed to be present in the observed time series.

(b) A time series of monthly observations covers 50 years. The series is known to contain a trend, seasonal fluctuations and an irregular component. What types of components will be present in the series obtained by(i) a 12-month moving average (ii) a 24-month moving average, and (iii) a 4-month moving average of the series

(c) Describe briefly the alternative methods of finding out the trend component of an observed series indicating their relative advantage and disadvantages.

[8+8+9]=[25]

2.(a) A modified exponential trend curve $Y_t = K + ab^t$ may be regarded as sum of two curves, viz., $Y_t=K$ and $Y_t = a\,b^t$. Can you obtain the shape of the curve graphically or algebrically when $k > 0$, $a > 0$ and $1 > b > 0$.

(b) If the rate of growth of trend value is measured as r_t $r_t = (Y_t - Y_{t-1}) / Y_{t-1}$ where Y_t is the trend value for time t, what would be the nature of rate of growth for an increasing linear trend curve and a declining linear trend curve

P.T.O.

2.(a) The following quarterly data relate to money supply with the public in India for the period end March 1968 to end December 1971.

end of	1968	1969	1970	1971
March	54*	58	64	71
June	55	60	66	75
September	53	58	66	74
December	54	60	68	77

* unit = 000 millions of Rs.

Calculate the quaterly seasonal indices by ratio to moving average method.

[5+5+15]=[25]

3.(a) Describe different criteria to select a good forecasting method.

(b) For a given time series data $y_1, y_2, \cdots, y_n$, $\sum_{t=1}^{10} y_t = 100$ and $n = 10$. If for the quadratic trend curve $Y_t = a+bt+ct^2$ fitted to this data $a = 2.0$ (the origin is selected so that $\sum_{t=1}^{n} t = 0$) then find out the value of C.

(c) Describe the method of estimation of parameters of a logistic curve $Y_t = \frac{k}{1+a\, e^{-bt}}$.

[9+9+7]=[25]

4. Write a note on the nature, coverage and sources of any two of the following :

i) Consumer price indices in India.

ii) Agricultural statistics available in India with particular reference to land utilisation and yield statistics.

iii) Central Statistical Organisation (C.S.O).

iv) National Sample Survey.

v) Statistical System in India.

$[12\frac{1}{2} + 12\frac{1}{2}] = [25]$

1982-83	E212

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications, 1982-83
PART I

Probability (Paper I)

SEMESTRAL-I EXAMINATION

Date: 13.12.1982. Maximum Marks: 100 Time: $2\frac{1}{2}$ hours.

Note : Answer all questions. The whole paper carries 108 marks but the maximum you can score is 100.

1. The life X of an electric bulb manufactured by a company has the exponential distribution with parameter $\lambda = \frac{1}{2}$.

   (a) Write down the density function of X and find the c.d.f. of X. [3+5]

   (b) Find the m.g.f. of X and using it find the mean and the variance of X. [5+3+4]

   (c) Suppose it costs the manufacturer Rs.5 to manufacture the bulb and he sells it for Rs.9. If the life of any bulb is less than 1, the manufacturer refunds the full sale price. If the manufacturer manufactures and sells 10,000 bulbs what is his expected profit? You may take $e^{-0.5} = 0.606$. [8]

2.(a) Write down the density function of χ^2_n. [4]

   (b) If X, Y are independent random variables such that $X \sim \chi^2_m$ and $Y \sim \chi^2_n$, find the density function of $Z = \frac{nX}{mY}$. [10]

   (c) Find the density function of $\frac{1}{1+\frac{mZ}{n}}$ using the result in 2(b). What is the name of this distribution? [6+2]

P.T.O.

3. (X,Y) has the uniform distribution on the region bounded by the triangle with vertices at $(0,0)$, $(1,0)$ and $(0,1)$.

(a) Find the marginal density of X and $E(X)$. [6+4]

(b) Find the conditional density of Y given $X=x$ and $E(Y|\ X=x)$. [4+4]

4.(a) Prove that $|\rho(X,Y)| \leq 1$ for any two random variables X and Y. [10]

(b) State clearly the uniqueness theorem and the continuity theorem for characteristic functions. [4+4]

(c) State Chebychev'sinequality. Using it solve the following problem : An unknown proportion p of the individuals in a population has a certain characteristic. A sample of size n is drawn with replacement and let Y be the proportion of individuals with the characteristic in the sample. Find the smallest n such that with probability at least 0.9, Y differs from p by less than 0.05. [4+8]

(d) State any version of central limit theorem and explain how you would use it to solve the problem in 4(c). [4+6]

1982-83 | E222

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications:'82-83
PART I

PAPER II
Descriptive Statistics(Theory)
SEMESTRAL-I EXAMINATION

Date: 15.12.1982. Maximum Marks:100 Time: 2 hours.

Note : The paper carries 120 marks. Attempt as many questions as you can but your maximum score is not to exceed 100 marks. The marks allotted are given at the end of each question.

1. Prove or disprove the following statements :

(i) For a binomial distribution $B(x, n, p)$ we have

$$\mu_{r+1} = p(1-p)[rn\ \mu_{r-1} + \frac{d\ \mu_r}{dp}]$$

where μ_r is the r^{th} central moment.

(ii) For Poisson distribution $P(x ; \lambda)$

$$\mu_{r+1} = r\lambda\ \mu_{r-1} + \lambda\ \frac{d\ \mu_r}{d\lambda}$$

(iii) For a normal distribution $N(\mu, \sigma^{-2})$

$$\mu_{2r+1} = 0$$

$$\mu_{2r} = 1.3.5.7 \ldots\ldots (2r-1)\ \sigma^{2r}$$

(iv) The sample correlation coefficient between x and y based on two observations (x_i, y_i), $i=1,2$, on each of x and y, is either +1 or -1.

(v) Regression line of y on x (based on the least squares estimates of unknown coefficients) in the sample is given by

$$y - \bar{y} = b_{yx}\ (x - \bar{x})$$

P.T.O.

1.(vi) Regression plane of y on x_1 and x_2 is given by

$$y-\bar{y} = b_{01.2}\,(x_1 - \bar{x}_1) + b_{02.1}(x_2 - \bar{x}_2)$$

where the suffixes 0,1 and 2 stand for y, x_1 and x_2 respectively and the symbols have their usual meaning.

(vii) The partial correlation coefficient $r_{01.2}$ is the geometric mean between the partial regression coefficient $b_{01.2}$ and $b_{10.2}$.

(viii) (a) $1 - R^2_{0(1,2)} = (1 - r^2_{01})(1 - r^2_{02.1})$

(b) $R_{0(1,2)} \geq r_{ij} \qquad i,j = 1,2,3, \quad i \neq j$

(c) $1 - R^2_{0(1,2)} = \dfrac{(1-\rho)(1+2\rho)}{(1+\rho)}$ if $r_{ij} = \rho$

(5+5+10+5+5+10+10+15) = [65]

2.(i) What do you mean by measures of central tendency and dispersion?

(ii) Give important measures of central tendency and dispersion.

(5+10)=[15]

3.(i) What do you mean by skewness and kurtosis?

(ii)Give some measures of skewness and kurtosis.

(5+10) =[15]

4.(i) Explain, in brief, the principle of least squares.

(ii) Obtain normal equations for fitting

$$y = a_0 + a_1 x + a_2 x^2$$

to a data (x_i, y_i) $i = 1, .., n.$

(5+10)=[15]

5. Obtain correlation coefficient between x and y whose joint distribution is given by

$$f(x,y) = 2 - x - y \qquad 0 \leq x \leq 1, \quad 0 \leq y \leq 1$$

$$= 0 \qquad \text{elsewhere.}$$

[10]

1982-83	E223

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications: 1982-83
PART I

PAPER II
Descriptive Statistics Practical
SEMESTRAL-I EXAMINATION

Date: 17.12.82. Maximum Marks : 100 Time: $2\frac{1}{2}$ hours.

Note : Attempt as many questions as you can. The paper carries 120 marks but the marks secured by you are not to be more than 100. The marks allotted are specified at the end of each question.

1. The following table gives, according to age, the frequency distribution of marks obtained by 60 students in an intelligence test.

Marks Y \ Age in years X	18	19	20	Total
10-20	4	2	2	8
20-30	5	4	6	15
30-40	6	7	10	23
40-50	4	4	6	14
Total	19	17	24	60

(i) Calculate the product moment correlation coefficient between X and Y.

(ii) Fit the regression line of Y on X and that of X on Y.

(iii) Find the theoretical values of Y for the given values of X.

(iv) Find the marks obtained by a student whose age is 22 years.

(10+4+4+2)=[20]

P.T.O.

2. Following is the joint distribution of two variables X and Y :

X / Y	1	2	3	4
1	1/36	3/18	-	-
2	2/36	-	1/18	-
3	3/36	3/18	-	1/9
4	4/36	-	1/18	1/6

Find
(i) E (X+Y) , (ii) E (XY) , (iii) Var(X),
(iv) Var(Y) , (v) Cov(X,Y) and ρ_{xy}.
(5+8+6+6+5)=[30]

3. An instructor of mathematics wished to determine the relationship of marks in a final examination to those on two quizzes given during the semester, calling X_1, X_2 and X_3 the marks of a student on the first quiz, second quiz and final examination respectively. He made the following computation for a total of 120 students :

$\bar{X}_1 = 6.8$ $\bar{X}_2 = 7.0$ $\bar{X}_3 = 74$

$\sigma_1 = 1.00$ $\sigma_2 = 0.80$ $\sigma_3 = 9.0$

$r_{12} = 0.60$ $r_{13} = 0.70$ $r_{23} = 0.65$

(i) Find all the partial correlation coefficients and multiple correlation coefficients.
(ii) Fit the regression plane of X_1 on X_2 and X_3.
(25+10)=[35]

4. At 200 different points in a big city, the number of deaths were observed for a week. The inquiry yielded the following data :

Number of deaths	0	1	2	3	4
Frequency	111	63	22	3	1

Fit the Poisson distribution to the above data and calculate the theoretical frequencies.

Given m	0.4	0.5	0.6	0.7
e^{-m}	0.6703	0.6065	0.5488	0.4966

[15]

5. The following table gives the number of workers in a factory with their monthly earnings :

Range of monthly earnings (00) Rs.	Number of workers
4 - 6	74
6 - 8	376
8 -10	304
10 -12	110
12 -14	18
14 -16	0
16 -18	9

Find the average monthly earnings per worker, median, first and third quartiles and mode of the above distribution and calculate the coefficient of variation. [20]

1982-83	E241

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications 1982-83
Part II

Statistical Inference (Theory and Practical)
PERIODICAL EXAMINATION

Date: 2.2.83. Maximum Marks: 100 Time: 2 hours.

Note: Answer question No.3 and any three of the rest.
All questions carry equal marks.

1.(a) Define population, parameter of a population, statistic, sampling distribution of a statistic and standard error of a statistic.

(b) Derive the standard error of sample mean, when sample is drawn from a finite population of size N and sampling is made with replacement.

2.(a) Find the sampling distribution of the sample mean, when sample is drawn from $N(\mu, \sigma^2)$.

(b) Define student's t-statistic. Hence derive the distribution of the t-statistic.

3.(a) Distinguish between point and interval estimation of a parameter.

(b) Assume that the top speed attainable by a certain design racing car is normally distributed with unknown mean μ and s.d. 10 miles per hour. A random sample of 10 cars, built according to this design, was selected and each car was tested. The sum and sum of squares of the top speeds attained (in miles per hour) were

$$\Sigma x_i = 1652 , \quad \Sigma x_i^2 = 273,765.$$

Compute the 95% confidence interval for μ. [Given $Z_{.025} = 1.96$, where Z is a standard normal variable].

P.T.O.

4. Explain what do you mean by maximum likelihood estimator of a parameter. Find the maximum likelihood estimates of the parameters of a normal population with mean μ and variance σ^2.

Are these estimates unbiased ? If not find their unbiased estimates.

5.(a) Suppose that the time it takes a distance runner to run a mile is normally distributed with mean μ and variance σ^2. Suppose he will run n_1 mile races in May and n_2 mile races in June. Let $\bar{X}_1$ be the mean value of his times in May and $\bar{X}_2$ be that in June. Show that $T = p\bar{X}_1 + (1-p)\bar{X}_2$ is an unbiased estimator of μ for any p $(0 \leq p \leq 1)$. Find the value of p that minimizes the variance of T, assuming $\bar{X}_1$ and $\bar{X}_2$ are independent.

(b) Define sufficient statistic. Hence find the sufficient statistic for the parameter of $N(0, \theta)$, where θ is the variance of the population, on the basis of a random sample $X_1, \ldots X_n$.

1982-83	E251

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications: 1982-83
Part II

Sample Surveys (Theory and Practical)

PERIODICAL EXAMINATION

Date: 16.2.83. Maximum Marks : 100 Time: 2 hours.

Note : The paper carries 111 marks. Answer as much as you can. The maximum you can score is 100. Marks allocated to each question are given in brackets [ ].

1. Suppose that you have to conduct a sample survey of persons in your locality to find out their attitudes and preferences for T.V. programmes. Explain, briefly, the main steps you would take for conducting such a survey. [13]

2. From a class of 121 students a sample of 20 is drawn at random with replacement in which 8 students are found to be smokers. Estimate the proportion of smokers in the class and give an approximate 95% confidence interval for your estimate. [4+6]=[10]

3.(a) Explain what you understand by Linear Systematic Sampling (L.S.S.) and Circular Systematic Sampling (C.S.S.). [9]

(b) When the population size N is not divisible by the required sample size n, obtain an unbiased estimator of the population mean in the case of L.S.S. [6]

(c) Can you estimate the variance of the above estimate in (b) unbiasedly. Give reasons. [1+6]=[7]

(d) Two samples chosen independently by C.S.S. of size 4 each gave the following y-values

sample 1 : 214, 287, 268, 294
sample 2 : 200, 198, 274, 286.

(i) Estimate the population mean and obtain an estimate of its sampling error. [3+4]=[7]

P.T.O.

4.(a) Explain briefly what you understand by Split method' in the case of Probability proportional to size (PPS) sampling. [5]

(b) For estimating the total number of absentees during a week in 20 factories of an area, a probability proportional to size sample of 2 factories is drawn with replacement. Using the data given below estimate the total number of absentees in the 20 factories. Also obtain an estimate of its sampling error.

No. of workers(x) and no. of absentees (y) on 2 sample factories

Sl.No.	x	y
1	28	2
2	142	8

Total no. of workers (no. of workers taken as size measure) in the 20 factories is 1561. [11+16]=[27]

5. State whether the following statements are true or false. Substantiate your statement.

(i) For Simple Random Sampling with Replacement of size n, an unbiased estimator of the population variance σ^2 is given by $\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$, where $\bar{y}$ is the sample mean. [9]

(ii) For Lahiri's method of selection, the probability of selection of a unit is proportional to its size. [9]

(iii) In stratified sampling, an unbiased estimator of the population mean where there are k strata is given by

$\hat{\bar{Y}}_{St.} = \sum_{i=1}^{k}\hat{\bar{Y}}_i / k$, where $\hat{\bar{Y}}_i$ is the estimate of the i^{th}. Stratum (population) mean, i = 1,2, ... k. [9]

1982-83 | E252

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications, 1982-83
Part II

Sample Surveys (Theory and Practical)
PERIODICAL EXAMINATION

Date: 30.3.1983. Maximum Marks: 100 Time : 2 hours.

Note : The paper carries 120 marks. Answer as much as you can. The maximum you can score is 100.

1.(a) What are the advantages of stratified sampling ? [6]

(b) Derive the 'Neyman's optimum allocation' of a total sample of size n when Simple Random Sampling Without Replacement (SRSWOR) design is used in all strata. [12]

(c) Obtain the Variance of the estimator of the population mean when the above allocation is used. [6]

(d) How do you use this allocation in actual practice. [6]

2. A population is divided into 2 strata of sizes 68 and 32. A random sample with replacement of size 7 is selected from the first stratum and the y-values of a study variate are found to be 114,121,161,143,114,111,152 from the second stratum a sample of size 3 is selected with probability of selection proportional to the size measure of the units, where the sizes of the selected units are 64,214,159. It is also known that the total of the size measure for the 32 units in the second stratum is equal to 4216. The y-values of the study variate for the 3 selected units from the second stratum are 312, 1026, 754.

(a) Obtain an unbiased estimate of the population total of y-values.

(b) Obtain an estimate of the sampling error of your estimate in (a) above. [11+19]=[30]

P.T.O.

3. When you have auxiliary information on a variable taking values $X_1, X_2, \ldots, X_N$ on the units $U_1, U_2, \ldots, U_N$ of a population correlated with the study variable y taking values $Y_1, Y_2, \ldots, Y_N$ on these units,

(a) Explain how you would build up a 'ratio estimator' $\hat{Y}_{RATIO}$ for the population total $Y = \sum_1^N Y_i$, using the x - values.

(b) Obtain expression for the Bias and Mean Squared Error of $\hat{Y}_{RATIO}$.

(c) Obtain conditions under which $\hat{Y}_{RATIO}$ is better than the estimator $\hat{Y}$ which does not use any auxiliary information.

[6+14+10]=[30].

4.(a) When do you use 'Cluster sampling'?

(b) From a district it is required to select a sample of villages for which the sampling frame is not available. Explain how you would proceed to obtain a sample and suggest an unbiased estimator for the population total.

(c) Briefly, explain the possible sources of non-sampling errors in the above problem and give your suggestions to control these.

[4+8+10+8]= [30]

1982-83 | E261

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications : 1982-83,
Part II

Design of Experiments(Theory and Practical)
PERIODICAL EXAMINATION

Date: 13.4.1983. Maximum Marks: 100 Time: 2 hours.

Note: Attempt as many questions as you can. Answers to Questions 1-3 should be written on the question paper itself. Answer to Question 4 should be written in the answer book.

Rough work, if any, for Questions 1-3 may be done in the answer book.

1. Fill in the blanks in the following :

(a) Among the three basic principles of design the principles of _____ and _____ are essential and the principle of _____ is desirable.

(b) The completely randomised design(CRD) does not involve the basic principle of _____.

(c) A factorial experiment ensures greater _____ and greater _____ than a single factor experiment.

(d) In two way classified data with t and r classes and n observations per cell, the error df. is _____.

(e) Another name of least significant difference is _____ difference. $8 \times 2\frac{1}{2} = 20$.

2. Indicate whether each of the following statements is true(T) or false(F) :

(a) In analysis of variance the error SS is obtained by subtraction and hence it can be negative. T / F

(b) The RBD is loss flexible than the CRD. T / F

P.T.O.

2.(c) In RBD randomisation is done afresh for each block. T/F

(d) In RBD, "block SS + treatment SS " may sometimes seceed the total SS. T/F

$4 \times 2\frac{1}{2} = 10.$

3. In each of the following indicate the correct alternative:

(a) The technique of paired comparison is based on

(A) χ^2 test, (B) t test, (C) F test, (D) none of these.

(b) In RBD plots within the same block should be

(A) as heterogeneous as possible,
(B) as homogeneous as possible,
(C) as large as possible,
(D) none of the above.

(c) Among the following which one is not a technique of achieving local control ?

(A) confounding, (B) analysis of covariance,
(C) taking more experimental units.

(d) In the analysis of two-way classified data with m observations per cell, the interaction component is not included if m equals

(A) 1, (B) 2, (C) 3, (D) 4.

(e) The analysis of CRD is the same as the analysis of

(A) one way classified data,

(B) 2^2 factorial design,

(C) two-way classified data with one observation per cell,

(D) none of the above.

(f) The missing plot technique was suggested by

(A) Snedecor, (B) Yates, (C) Fisher, (D) Wilks.

(contd...3)

3. **The following is an incomplete ANOVA table of an RBD with 4 blocks and 3 treatments. Complete the table (No calculations or explanations are to be shown).**

Source	d.f.	SS	MS	F
Blocks		6		
Treatments				3
Error			.5	
Total				

25

4. Three treatments were compared in a completely randomised design and the following data on yield figures (in suitable units) were obtained

Treatment 1	Treatment 2	Treatment 3
11	8	18
13	10	20
12	9	21
15	7	
	5	

Examine whether the treatments can be considered equivalent. If not, examine whether treatments 2 and 3 differ significantly.

40.

1982-83 | E262

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications, 1982-83.
Part II.

Industrial Statistics (Theory and Practical)
PERIODICAL EXAMINATION

Date: 25.4.1983. Maximum Marks: 100 Time: 2 hours.

Note: Answer any four questions. All questions carry equal marks. Answers should be to the point and brief.

1. How would you define 'Quality'? when a quality characteristic is said be attribute ? Give 4 examples of quality characteristics. What quality characteristic in industry follows Poisson distribution. What type of control charts would you recommend for control of such characteristic ? Work out the limits for such a chart.

[2+2+4+2+5+10]

2. The dimensions of a particular characteristic of a R.E.Core at milling are specified as 7.4 $\pm$ .05 unit. The following table gives the average and range of 24 subgroups taken from normal production. The subgroup size is 5. Draw a suitable control chart to analyse the data. Examine the state of control with particular reference to process capability and given specification.

Sub-group	$\bar{X}$	R	Sub-group	$\bar{X}$	R
1	7.380	0.03	13	7.392	0.10
2	7.372	0.02	14	7.400	0.04
3	7.351	0.05	15	7.410	0.08
4	7.340	0.08	16	7.350	0.02
5	7.385	0.06	17	7.362	0.04
6	7.400	0.07	18	7.370	0.03
7	7.450	0.07	19	7.385	0.04
8	7.410	0.06	20	7.385	0.05
9	7.420	0.05	21	7.375	0.05
10	7.415	0.10	22	7.390	0.05
11	7.395	0.07	23	7.401	0.02
12	7.430	0.08	24	7.370	0.08

[25]

P.T.O.

3. In order to control a variable characteristic an $\bar{X}$-R chart was installed. How would you interpret the following situations in the control chart.

   (a) R chart exhibits control but most of the points are beyond limits on both sides of $\bar{X}$-chart, only a few points lie within limits.

   (b) R chart in control, but points beyond the control limits on $\bar{X}$-chart were found only on one side.

   (c) R chart out of control, and on $\bar{X}$-chart several points were found to lie beyond control limits on both sides though many points lie within the limits.

   (d) R chart exhibits lack of control and many points are beyond the upper control limit in the $\bar{X}$-chart. [25]

4. The following data gives the daily inspection result of low tension insulators manufactured in a ceramic plant. Observing high rejection the causes were investigated. Most of the rejections were due to a fine hair line crack. To control it the firing schedule was modified. The result of this change, if any, should be reflected from subgroup No.12 onwards. Analyse the data with a suitable control chart and examine if the change in firing schedule has any effect on rate of rejection. What process standard you would recommend for future control ?

Subgroup	No inspected	No. of defectives
1	500	95
2	500	101
3	500	100
4	500	85
5	500	90
6	500	120
7	500	75
8	500	94
9	500	110
10	500	81
11	500	90
12	500	90
13	500	70
14	500	85
15	500	70
16	500	65
17	500	68
18	500	73
19	500	55
20	500	75

[25]

(contd...3)

5. Write short notes on any two :

(a) Specification limits vs. Control limits

(b) Use of frequency distribution in Quality Control Work

(c) Modified Control limits.

[25]

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications : 1982-83
Part I

Economic Statistics(Theory and Practical)
SUPPLEMENTARY SEMESTRAL-I EXAMINATION

Date: 28.4.83. Maximum Marks: 100 Time: 2 hours

GROUP A Max.Marks: 50.

Note : Answer all the Questions.

1. Distinguish between fixed base and chain base index numbers and discuss their advantages and disadvantages. Under what condition will these index numbers give identical results ? [15]

2. The table below gives the wholesale prices and quantities produced of a number of foodgrains in India.

Commodity	1951		1952		1953	
	p	q	p	q	p	q
Rice	16.87	20964	17.50	22537	17.50	27769
Jowar	10.09	5981	11.93	7243	12.08	7954
Bajra	10.07	2309	13.33	3142	13.33	4475
Maize	21.75	2043	15.67	2825	14.70	2991
Ragi	9.45	1291	9.30	1316	15.96	1846
Wheat	18.60	6085	23.67	7382	21.93	7890
Barley	20.66	2330	17.29	2882	13.80	2905
Gram	24.09	3334	19.01	4142	19.60	4756

p : price in Rs. per 40 kgs.

q : quantity produced in thousand tons.

Calculate Laspeyre's, Paasche's, Edgeworth - Marshall and Fisher's Ideal Price Index Numbers for each of two years 1952 and 1953, taking 1951 as the base year.

[35]

P.T.O.

GROUP B Max.Marks: 50

Note : Answer all the questions.

1. The following time series is the all India Consumer Price Index Number for Agricultural Labourers (Base 1960-61=100) for food.

All India Consumer Price Index Number (Base 1960-61 = 100)

Month \ year	1973-74	1974-75	1975-76
July	270	354	353
August	280	366	349
September	280	383	350
October	286	382	346
November	281	379	349
December	288	385	350
January	300	383	338
February	304	388	328
March	309	380	312
April	316	367	314
May	339	366	309
June	346	367	309

(a) What systematic components of a time series are present in this series ?

(b) Calculate the monthly seasonal index by ratio to moving average method.

(c) Explain how would you forecast for the months of the year 1976-77. (Do not perform any calculation. Just explain the method).

(d) Describe different criteria to select a good forecasting method. [25]

2.(a) For a given time series data, it is given that $\sum_{t=1}^{10} y_t = 100$. If for the quadratic trend curve fitted to this data, a = 2 then find the value of c. The origin is so selected that $\sum_{t=1}^{10} t = 0$.

(b) Describe different trend models which can be used to forecast a time series.

(c) Write a note on the nature, coverage and sources of any one of the following :

(i) Consumer price index number.

(ii) Statistical System in India.

(iii) Industrial Statistics. [25]

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications : 1982-83
Part I

Supplementary Semestral-I Examination
PROBABILITY

Date: 28.4.83. Maximum Marks: 100 Time: 2 hours.

Note: Answer all questions. The whole paper carries 108 marks but the maximum you can score is 100.

1. X,Y are continuous random variables with joint density function

$$f(x,y) = \begin{cases} c(x+y) & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ & \text{otherwise.} \end{cases}$$

(a) Find the value of c. [6]

(b) Find the marginal density function of X. [6]

(c) Find E(X), V(X). [6+6]

(d) Find the conditional density function of Y given $X = \frac{1}{2}$. [5]

(e) Find cov (X,Y). [8]

2. (a) State and prove Chebychev's inequality. [12]

(b) State and prove the law of large numbers. [10]

3. (a) Write down the density function of the χ^2_n distribution. [5]

(b) Find using the density function, the m.g.f. of χ^2_n. Using the m.g.f., find the mean and variance. [8+6+6]

(c) Let X,Y be independent random variables with $X \sim \chi^2_m$ and $Y \sim \chi^2_n$. Write down the joint density function of X,Y and obtain the joint density function of U,V where $U = \frac{X}{X+Y}$ and V = X+Y. From the joint density function of U,V find the density function of V. What is the name of the distribution of V ? [4+12+6+2]

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications : 1982-83
Part I

Descriptive Statistics (Theory and Practical)
SUPPLEMENTARY SEMESTRAL-I EXAMINATION

Date: 28.4.83. Maximum Marks : 100 Time: 2 hours.

Note: Attempt as many questions as you can. The paper carries 120 marks but the maximum you can score is 100 marks. Marks allotted are given at the end of each question.

1.(a) X is a random variable having Poisson distribution with parameter λ. Given that

$$P(X=1) = P(X=2),$$

find (i) $P(X \leq 3)$, (ii) $P(X > 5)$.

(b) In an examination 25% of the candidates score less than 40, while 10% score above 70. Assuming the distribution of scores to be normal, estimate the usual parameters μ and σ from this information. Hence estimate the expected percentage of candidates scoring 60 or more.

[12+12=24]

2. Let X,Y be independent random variables having the same probability distribution

$$P(X=i) = P(Y=i) = \frac{1}{3}, \quad i = 1,2,3.$$

Define Z = Max (X,Y)

(a) Obtain the joint probability distribution of X and Z.
(b) Obtain E(X), E(Z).
(c) Obtain Var(X), Var(Z).
(d) Obtain the correlation coefficient between X and Z.
(e) Are X, Z independent ?

[8+4+8+10=2 = 32]

3. On the basis of 100 triplets of observations on three characters X_1, X_2, X_3, the following have been computed

$$\bar{X}_1 = 7.0, \quad \bar{X}_2 = 9.0, \quad \bar{X}_3 = 14.0,$$

$$\sigma_1 = 2.3, \quad \sigma_2 = 2.9, \quad \sigma_3 = 5.8,$$

$$r_{12} = r_{13} = r_{23} = .5$$

P.T.O.

Q.No.3 (contd..)

(a) Obtain the multiple linear regression equation of X_1 on X_2 and X_3.

(b) Find the multiple correlation coefficient between X_1 and X_2, X_3.

(c) Find the partial correlation coefficient between X_1 and X_2 eliminating the effect of X_3.

[12+8+8=28]

4. With bivariate data on two characters X and Y, the regression lines are $2x-3y+6=0$ and $5x-y+8=0$

(a) Identify which one is the regression line of Y on X and which one is that of X on Y.

(b) Obtain the correlation coefficient between X and Y.

(c) Find $\bar{X}$, $\bar{Y}$, the means of X,Y respectively.

[12+6+4=22]

5. Write short notes on any <u>two</u> of the following :

(a) Pie-chart.
(b) Scatter Diagram,
(c) Coefficient of Variation,
(d) Measures of Kurtosis.

[7+7 = 14]

1982-83 | E421

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications : 1982-83
Part II

Statistical Inference (Theory)
SEMESTRAL-II EXAMINATION

Date: 16.5.83. Maximum Marks: 100 Time: 2 hours.

Note: Answer any four questions. All questions carry equal marks.

1.(a) Define Chi-square statistic and derive its distribution.

(b) Let $x_1, .. x_n$ be n i.i.d. random variables distributed as $N(0,1)$. If $\bar{x} = \sum_1^n x_i/n$ and $S^2 = \sum_1^n (x_i - \bar{x})^2$, then prove that

$$P[\bar{x} < 0, S^2 \geq 2] = \frac{1}{2} P[\chi^2(n-1) \geq 2].$$

(c) Let t be distributed as student's t-distribution with n d.f. what is the distribution of t^2 (statement only).

2.(a) Define unbiased estimator of a parameter. Let X be a random variable with mean μ and variance σ^2. Also let $X_1, .. X_n$ be a simple random sample of X.
Show that $C \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2$ is an unbiased estimator of σ^2 for an appropriate choice of C. Hence find C.

(b) An urn contains white and black balls in unknown proportions, the total number of balls being 3. Three balls are taken at random of which 2 are found to be white and 1 black. Find the maximim likelihood estimate of the number of white balls.

(c) Let p.d.f. of a random variable X is

$$f(x) = \frac{\lambda(\lambda x)^{\gamma-1} e^{-\lambda x}}{\Gamma\gamma}, \quad x > 0$$
$$= 0, \quad \text{otherwise}.$$

where γ is known .

Find the maximum likelihood estimate of λ.

p.t.o.

3.(a) Let X be a random variable with mean μ and variance σ^2. Also let $X_1, .. X_n$ be a random sample of X. Prove that the sample mean $\bar{X}$ is the best linear unbiased estimate of μ.

(b) State and prove Rao-Cramer inequality. Show that the variance of the sample variance when the population is normal with known mean μ and unknown variance σ^2, attains Rao-Cramer lower bound. Hence obtain the lower bound.

4.(a) What do you mean by a Statistical test ? Define type I error and type II error and explain their role in testing of hypothesis.

(b) State and prove Neyman-Pearson's theorem for obtaining most powerful test for a simple null hypothesis against a simple alternative.

Let $X_1, .. X_n$ be independent random sample from a population having p.d.f.

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad , \quad 0 < x < \infty$$
$$0 < \theta < \infty$$
$$= 0 \quad , \text{ otherwise.}$$

Derive a most powerful test for testing $H_0[\theta = 2]$ against $H[\theta = 4]$.

Hence explain how will you compute the size and power of this test.

5.(a) Consider a bivariate distribution of two random variables. Let the conditional distribution of y given x is normal with mean $\eta_x = \alpha+\beta x$ and variance σ^2 on the basis of a random sample of size n, show that the least squares estimates of α and β are the same as their maximum likelihood estimates.

(b) What is Pearsonian Chi-Square ? For a 2 x 2 contingency table as below

Row	Columns 1	2
1	a	b
2	c	d

Show that the Pearsonian Chi-square can be written as

$$\chi^2 = \frac{(ad-bc)^2 (a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)} .$$

(contd..3)

6. What is one-way classified analysis of variance model ? State the underlying assumptions of the model and compute the expectation of the relevant mean squares. Hence write down the ANOVA table.

1982-83 E422

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications, 1982-83
Part II

Statistical Inference (Practical)

SEMESTRAL-II EXAMINATION

Date: 18.5.83. Maximum Marks: 100 Time: 2 hours.

Note : This paper carries 105 marks. Maximum one can score is 100.

1. The following gives the scores in two subjects, Statistical Inference (Theory) and Statistical Inference (Practical) of ten students selected at random.

% score in Statistical Inference (Theory)	67	53	57	65	48	50	67	59	75	42
% score in Statistical Inference (Practical)	65	40	47	58	40	55	49	61	62	37

Under suitable normality assumptions test :

(a) If there is any dependence between the scores in the two subjects.

(b) Whether it is justified to claim that the mean score in Statistical Inference (Theory) is 10 marks more than the mean score in Statistical Inference (Practical).

[20+15=35]

2. In a certain town, the average number of intercaste marriages per week was known to be 1 from previous records. However in two consecutive weeks 2 and 3 intercaste marriages took place respectively. Would you conclude from this data that the average number of intercaste marriages per week has increased in recent times ?

[12]

3. The following data relate to the level of literacy and the level of income of a section of people from rural Bengal. Examine whether there is any association between the level of income and the level of literacy.

Level of Literacy / Level of income	Illiterate	Semi-literate	Literate
Low income	45	26	12
Medium income	22	30	15
High income	8	12	18

[22]

p.t.o.

4.(a) The following table gives the additional hours of sleep gained by 10 patients in an experiment to test the efficiency of a certain soporific drug. Do the data give evidence that the drug produces at least one hour's additional sleep ?

Additional hours of sleep gained

Patient	1	2	3	4	5	6	7	8	9	10
Hours gained	1.7	-.1	.6	2.2	1.1	4.0	1.8	4.7	2.8	3.0

(b) Using the data in 4(a),obtain a 95%, two sided confidence interval for the variance of the additional hours of sleep gained. [16+10]

5. A random sample of 50 tubelights manufactured by a firm contained 6 defectives. Find an upper bound ''p_c'' for the proportion ''p'' of defective tubelights such that Prob. $(p \leq p_o) = .90$. [10]

1982-83 | E521

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications, 82-83
PART II

Sample Surveys (Theory)

SEMESTRAL-II EXAMINATION

Date: 20.5.83. Maximum Marks: 100 Time: 2 hours.

Note : The paper carries 112 marks. Answer as much as you can. The maximum you can score is 100.

1. Explain briefly the main sources of non-sampling errors in a sample survey. With reference to any particular survey of your choice indicate how you would assess and control the non-sampling errors in that survey. [14+8+8=30]

2. Consider a finite population of size N with the corresponding values of the study variate ($\mathcal{Y}$) and auxiliary variate ($\mathcal{X}$) as Y_i and X_i on the ith unit, respectively. Let $\hat{R}_1 = \bar{y}/\bar{x}$ be an estimator for $R = \frac{Y}{X} = \sum_{i=1}^{N} Y_i \Big/ \sum_{1}^{N} X_i$ based on a simple random sample of size n selected with replacement where $\bar{y}$ = sample mean of selected y-values and $\bar{x}$ = sample mean of selected x values.

   (a) Is $\hat{R}_1$ unbiased for R ? If not, obtain B_1 = Bias ($\hat{R}_1$). [6]

   (b) Consider an alternative estimator $\hat{R}_2 = \frac{1}{n} \sum_{i=1}^{n} y_i/x_i$. Is $\hat{R}_2$ unbiased for R ? If not, obtain B_2 = Bias($\hat{R}_2$). [6]

   (c) How are B_2 and B_1 related ? [3]

   (d) Consider another estimator $\hat{R}' = \frac{n\hat{R}_1 - \hat{R}_n}{n-1}$. Is this unbiased for R ? If not, obtain B = Bias ($\hat{R}'$). What conclusions do you draw from the above results. [6+1+4]=[11]

p.t.o.

3.(a) Give an example of a sample survey where you use a two-stage sampling design. [8]

(b) For a particular choice of Simple Random Sampling Without Replacement (SRSWOR) design at both stages, write down an unbiased estimator of the population total of your study variate. [6]

(c) If in the first stage you also have extra information on an auxiliary variate related to the study variate how do you use this information (i) in choosing a design and (ii) in estimation ? [6+6] = [12]

4.(a) How is 'Systematic Sampling' different from a 'simple random sampling design ? [5]

(b) For a 'linear systematic sampling design', how do you estimate the population total unbiasedly ? [6]

(c) If only one systematic sample is available from a population, suggest one method of estimating the sampling error. Is your estimator unbiased ? [9+2]=[11]

(d) If you have two independent systematic samples, then explain how you would estimate the sampling error and assess the non-sampling error ? [8]

1982-83	E522

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications, 1982-83
Part II

Sample Surveys (Practical)
SEMESTRAL-II EXAMINATION

Date: 23.5.83. Maximum Marks : 100 Time: 2 hours.

Note : Solve as much as you can. The paper carries 125 marks, however, the maximum you can score is 100.

1. A population of 150 villages has been divided into four strata having 35, 25, 50 and 40 villages respectively on the basis of type of available auxiliary information. We have a sample of 4 villages selected with SRSWR from the first stratum, a PPSWR sample of 3 villages from the second stratum, two Linear Systematic samples of 4 villages each with replacement from the third stratum and an SRSWOR sample of 5 villages from the fourth stratum. For each selected village, the total area under wheat (y) is observed. The observed values and the other relevant information are given below.

Stratum	1	2		3		4	
				Sample 1	Sample 2		
Sample village	(y)	(x)	(y)	(y)	(y)	(x)	(y)
1	75	729	247	427	395	421	101
2	5	870	359	326	412	327	89
3	78	569	223	481	503	215	65
4	45			445	348	503	169
5						289	51

where (x) : total area under cultivation.

Further X_2 = Total of x-values for the second stratum =18691

and X_4 = Total of x-values for the fourth stratum =14852.

p.t.o.

Q.No.1 contd...

USE RATIO ESTIMATOR I.. STRATUM FOUR.

(a) Estimate the total area under wheat for each of the strata separately. [31]

(b) Suggest an estimator for the average area under wheat for the entire population. Is your estimator unbiased ? If not what can you say about the bias ? [28]

(c) How efficient is the estimator that you have used in (b) ? [56]

(d) From the available information, in the above set up, how would you allocate a total sample of size 30 to different strata using
(i) proportional allocation
(ii)optimal allocation. [10]

Roll No.

1982-83	E621

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications, 1982-83

Design of Experiments(Theory and Practical)

SEMESTRAL-II EXAMINATION

Date: 25.5.83. Maximum Marks : 100 Time: 2 hours.

Note : Attempt as many questions as you can. Write answers to questions 1 and 2 on the question paper itself. Answers to questions 3 and 4 and rough work, if any, for questions 1 and 2 should be written on the answer book.

1. In each of the following identify the correct alternative :

(i) In a 2^2 factorial design interactions AB and BA are
(A) identical, (B) different, (C) sometimes identical sometimes different, (D) none of the above.

(ii) In a 2^2 factorial design b stands for the level combination in which
(A) A is at higher and B is at lower level,
(B) A is at lower and B is at higher level,
(C) A is at higher and B is at higher level,
(D) None of the above holds.

(iii) In a 2^2 factorial design, [A], the factorial effect total corresponding to main effect A is defined as
(A) [A] = [ab] - [b] + [a] - [1],
(B) [A] = [ab] + [b] - [a] - [1],
(C) [A] = [ab] - [b] - [a] + [1],
(D) None of the above.
(notations are as usual)

(iv) The computational method for obtaining factorial effect totals in a 2^m experiment is due to
(A) Fisher, (B) Yates , (C) Snedecor, (D) Hotelling.

p.t.o.

1.(v) If a 2^3 design is conducted in three replicates then error df. equals

(A) 7 , (B) 14, (C) 2, (D) None of these.

(vi) In a 2^3 design, given (in usual notations) let

[1]=5, [a] =7, [b]=3, [ab]=11, [c]=2, [ac]=8,

[bc]=16, [abc]=4, the value of [AC] the factorial effect total for interaction AC is given by

(A) -6, (B) 8, (C) -7, (D) None of these.

(vii) In (vi) above if the experiment is conducted in three replicates then SS due to main effect A is given by

(A) 2/3 , (B) 16/3, (C) 2, (D) None of these.

(viii) Given that there exist 5 mutually orthogonal latin squares of order m which one of the following is guaranteed ?

(A) $m \leq 5$, (B) $m \geq 5$, (C) $m = 5$, (D) None of these.

(ix) In the statement 'a latin square design is a (an) _____ _____ - way layout'', the blanks can be filled respectively by

(A) complete, three, (B) incomplete, two (C) incomplete three, (D) complete, two.

(x) In (m x m) latin square design the error df. equals

(A) $(m-1)^2$, (B) (m-1)(m-2), (C) (m-2)(m-3), (D) None of these.

(xi) By a complete set of mutually orthogonal latin squares (MOLS) of order m we mean

(A) m MOLS of order m ,
(B) m-1 MOLS of order m ,
(C) m+1 MOLS of order m,
(D) None of these.

(xii) Compared with a randomised block design, a latin square design is

(A) more efficient and more flexible,
(B) less efficient but more flexible,
(C) more efficient but less flexible,
(D) less efficient and less flexible.

3 x 12 =36

(contd...3)

2.(a) Examine whether the following two latin squares are mutually orthogonal.

0	1	2	3
1	0	3	2
2	3	0	1
3	2	1	0

0	1	2	3
2	3	0	1
3	2	1	0
0	1	2	3

Ans: Yes/ No.

(b) Examine whether the following two latin squares are mutually orthogonal :

0	1	2	3
1	2	3	0
2	3	0	1
3	0	1	2

0	1	2	3
2	3	0	1
3	0	1	2
1	2	3	0

Ans: Yes/No.

4 x 2=8

3. Consider the linear model :

$$Y_i = \alpha + i\beta + i^2\gamma + e_i , \quad i = 1,2, \ldots, n$$

where $\alpha, \beta,$ are unknown parameters and e_i's represent error components assumed to be independently identically distributed with zero means and constant variance.

Apply the technique of analysis of variance to derive a test for the null hypothesis $H_0 : \beta = \gamma = 0$.

30

4. The following data relate to a 2^3 factorial design conducted in two blocks. Figures within parentheses stand for yields (in suitable units). Analyse the data and test the significance of the main effects and interactions.

Block I	(1)	a	b	ab	c	ac	bc	abc
	(7)	(9)	(5)	(16)	(12)	(11)	(18)	(24)
Block II	(1)	a	b	ab	c	ac	bc	abc
	(3)	(6)	(4)	(18)	(11)	(15)	(13)	(21)

40.

1982-83	E62

INDIAN STATISTICAL INSTITUTE

One Year Evening Course in Statistical Methods and Applications, 1982-83
Part II

Industrial Statistics(Theory and Practical)
SEMESTRAL-II EXAMINATION

Date: 27.5.83. Maximum Marks : 100 Time: 2 hours.

Note : Answer any four questions. Marks allotted to part are shown in [ ]. Answers should be to the point and brief.

1. The specification for thickness of a component is 10cm $\pm$ .25cm. Analysis of past data shows that the process capability is just adequate to meet the specification. For the control of thickness on $\bar{X}$-R chart is set up with 3σ control limit with each subgroup containing a sample of 5 consecutive pieces. It is assumed that the process standard deviation remains under control.

   (a) Find the probability of detection by the $\bar{X}$-chart if the process mean (i) increases to 10.50 cm and (ii) decreases to 9.25 cm.

   (b) What will be the probabilities of detection for case (i) and case (ii) if 2σ control limits were used.

   (c) In the light of your results which control scheme would you adopt ? Give reasons for your preference.

[10+10+5]

2.(a) What is meant by work sampling ?

   (b) How would you adopt this technique to study the efficiency in a textile mill ? Also indicate the type of analysis to be made.

   (c) A weaving shed consists of 500 looms. 4 snap rounds at random timings were taken every day and the weekly summary (6 days) showed the weaving utilisation to be 75%. Obtain a 95% confidence interval of the given estimate.

   (d) The over all loss in utilisation for mechanical repair is 5.0%. How would you identify the looms which were more prone to frequent break downs. [4+7+7+7]

p.t.o.

3.(a) Why would you prefer sampling inspection to 100% inspection ?

(b) Explain with reasons if a defect free sample will ensure a defect free lot.

(c) What are your arguments against sampling plans where sample size bears a constant ratio to the lot size ?

[7+9+9]

4. Define the following terms :

(a) (i) AQL, (ii) LTPD, (iii) AOQ and (iv) AOQL

(b) Given AQL, TPD, the consumer's risk and the producer's risk how would you select a single sampling plan ?

[16+9]

5.(a) What is meant by OC of a sampling plan ? Write down the expression for OC of a double sampling plan.

(b) Give your assessment of the Double sampling scheme in comparison to the single sampling scheme.

(c) Draw the OC curve of the following sampling plan :

$n = 50, \quad c = 2.$

[6+5+14]

6. Write in details about any standard sampling plan indicating its scope and special features.

[25]