III YEAR - V SEMESTER
COURSE CODE: 7BMA5C2

CORE COURSE - X – STATISTICS - I

Unit – I

Central Tendencies – Introduction – Arithmetic Mean – Partition Values – Mode Geometric Mean and Harmonic Mean – Measures of Dispersion.

Unit – II

Moments – Skewness and Kurtosis – Curve fitting – Principle of least squares.

Unit – III

Correlation – Rank correlation Regression – Correlation Coefficient for a Bivariate Frequency Distribution.

Unit – IV

Interpolation – Finite Differences – Newton's Formula – Lagrange's Formula – Attributes – Consistency of Data – Independence and Association of Data.

Unit – V

Index Numbers – Consumer Price Index Numbers – Analysis of Time series – Time series – Components of a Time series – Measurement of Trends.

Text Book:

1. Statistics by Dr. S. Arumugam and Mr. A.ThangapandiIssac, New Gamma Publishing House, Palayamkottai, June 2015.

Unit I	Chapter 2sections 2.1 to 2.4 Chapter 3 section 3.1
Unit II	Chapter 4 sections 4.1 & 4.2 Chapter 5 section 5.1
Unit III	Chapter 6 sections 6.1 to 6.4
Unit IV	Chapter 7 sections 7.1 to 7.3 Chapter 8 sections 8.1 to 8.3
Unit V	Chapter 9 sections 9.1 & 9.2 Chapter 10 sections 10.1 to 10.3

Book for Reference:

1. Statistics Theory and Practice by R.S.N.Pillai and Bagavathi, S.Chand and Company Pvt. Ltd. New Delhi, 2007.

♣♣♣♣♣♣♣♣♣♣♣♣

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-1

PART-1

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

INTRODUCTION

Statistics means numerical description to most people.

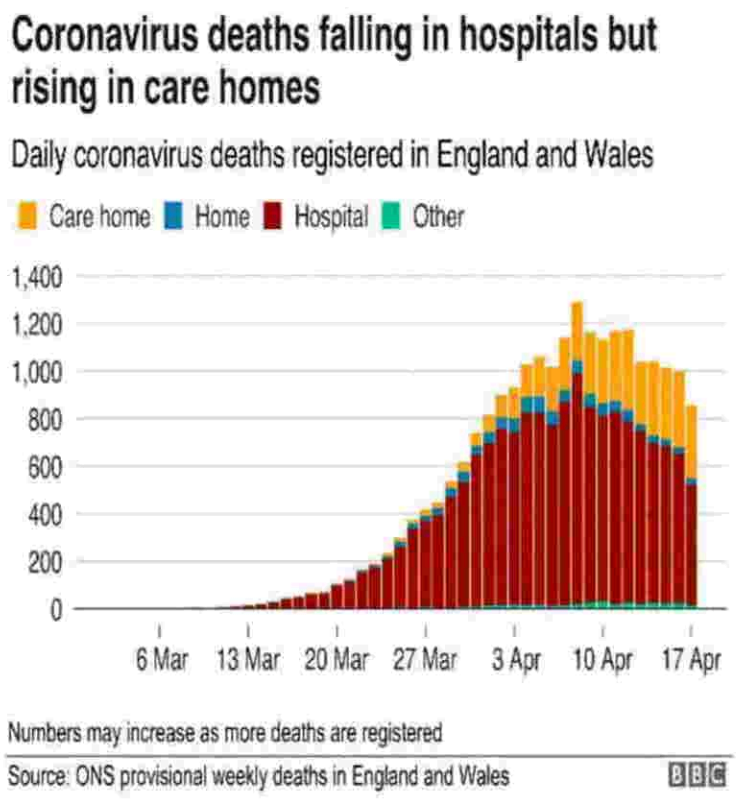
Statistics deals with collection , classification, tabulation, analysis and interpretation of numerical data.

Statistical methods have been developed for analysing data and drawing valid inferences or intelligent judgements from them.

ORIGIN OF STATISTICS

The term statistics has its origin in Latin word *Status,* Italian word *Statista* or German term statistik. All the three terms mean Political State.

statistics is not related to the administration of the state alone, but it has close relation with almost all those activities of our lives which can be expressed in quantitative terms.



DEFINITION OF STATISTICS

statistics is a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data

TYPES OF STATISTICS

DESCRIPTIVE STATISTICS

- METHODS FOR SUMMARIZING DATA
- SUMMERIES USUALLY CONSIST OF GRAPHS AND NUMERICAL SUMMARIES OF THE DATA

INFERENTIAL STATISTICS

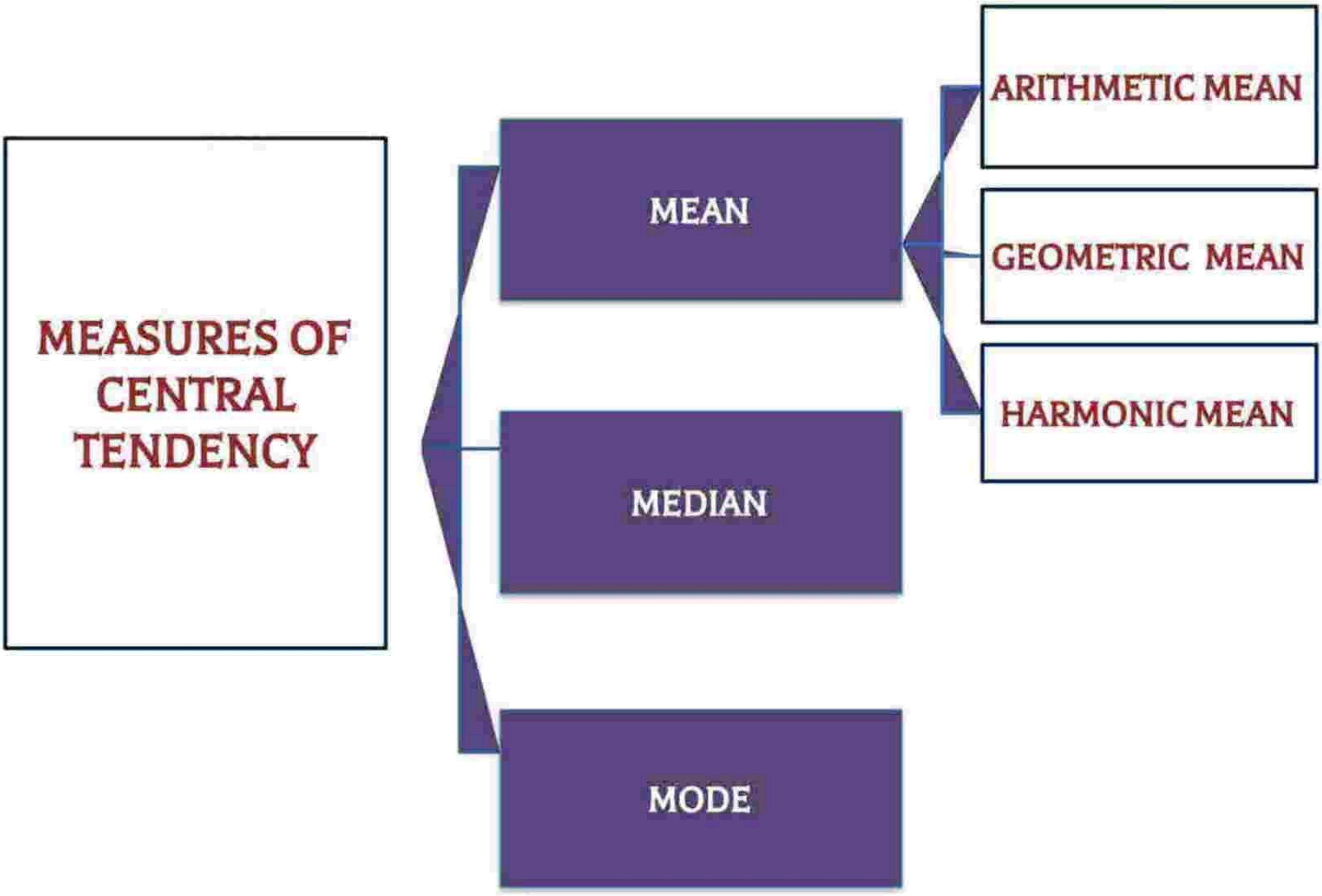
- METHOD OF MAKING DECISIONS ABOUT A POPULATIONS BASED ON SAMPLE INFORMMATION.

FOUR TYPES OF DESCRIPTIVE STATISTICS

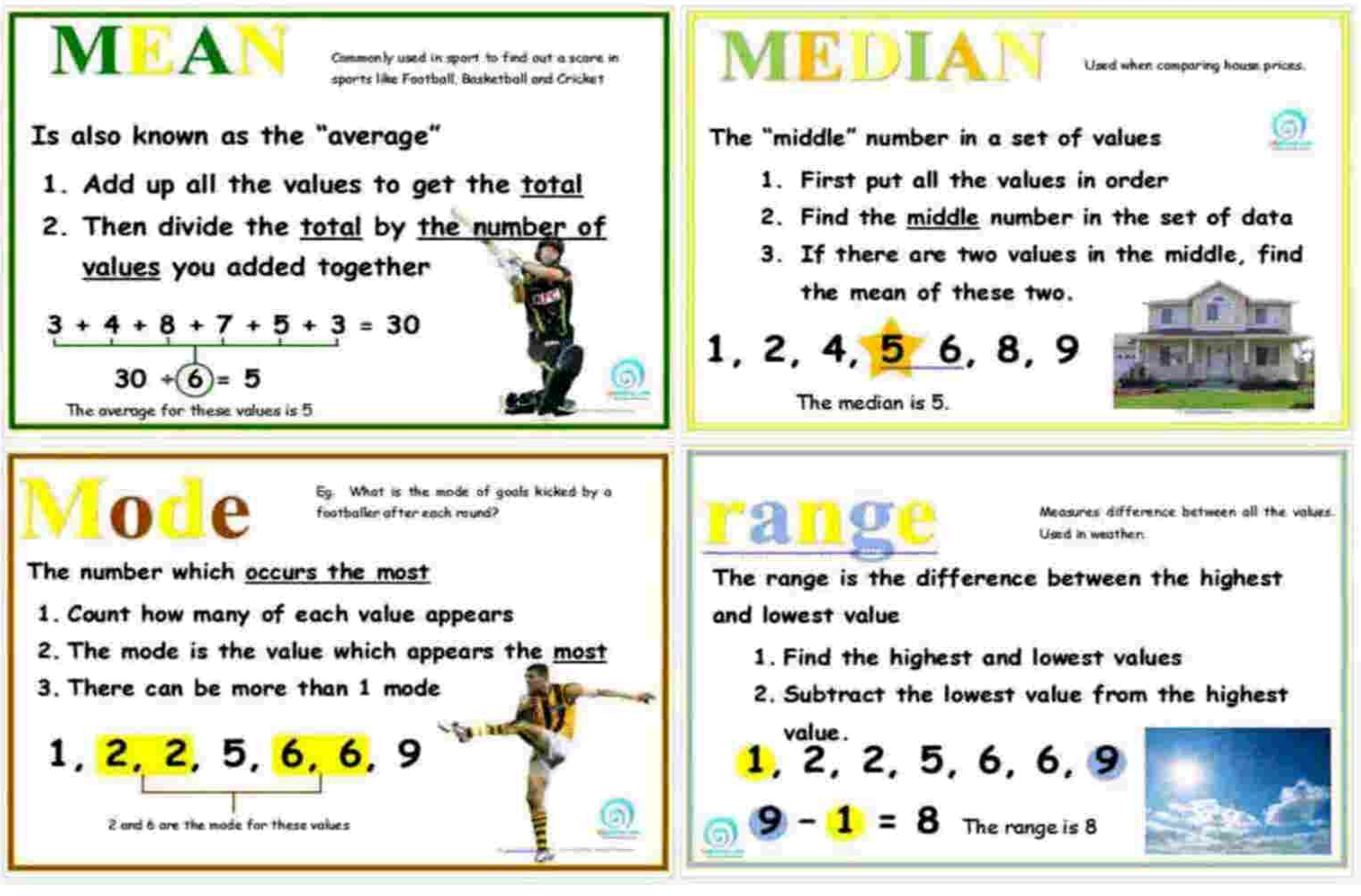
1. MEASURE OF CENTRAL TENDENCY OR MEASURE OF LOCATION.
2. MEASURE OF DISPERSION.
3. MEASURE OF SKEWNESS.
4. MEASURE OF KURTOSIS.

DEFINITION OF MEASURES OF CENTRAL TENDENCY

MEASURES OF CENTRAL TENDENCY ARE "STATISTICAL CONSTANTS WHICH ENABLE US TO COMPREHEND IN A SINGLE EFFORT THE SIGNIFICANCE OF THE WHOLE"


MEASURES OF CENTRAL TENDENCY
MEAN
MEDIAN
MODE
ARITHMETIC MEAN
GEOMETRIC MEAN
HARMONIC MEAN

Basic knowledge about central tendency



Assignment

1. Find the mean , median and mode height of the above data.

2. Collect some pictures from daily news paper also find
   3 major measures of central tendency .

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

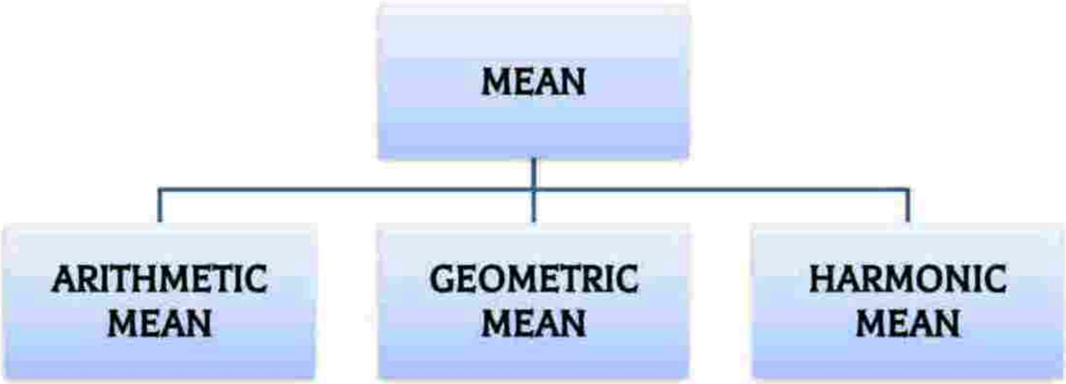
COURSE CODE: 7BMA5C2

UNIT-1

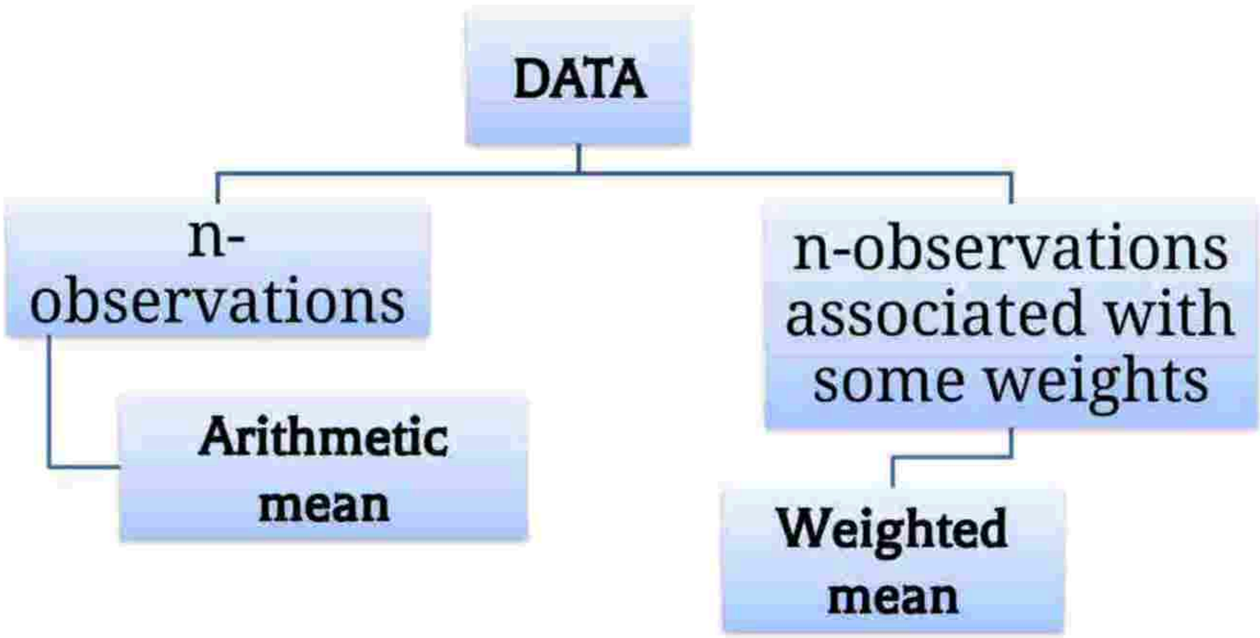
PART-2

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

content


MEAN
ARITHMETIC MEAN
GEOMETRIC MEAN
HARMONIC MEAN

classification of ARITHMETIC MEAN



Definition of arithmetic mean

Arithmetic Mean:

The Arithmetic Mean is obtained from a set of numbers by dividing the sum of those numbers by the number of observation.

If $x_1, x_2, x_3, \ldots\ldots, x_n$ are n observation, then their arithmetic mean is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots\ldots + x_n}{n}$$

$$= \sum_{i=1}^{n} \frac{x_i}{n}$$

Definition of weighted mean

Weighted Mean

Let $x_1, x_2, \ldots, x_n$ be the set of n values having weights $w_1, w_2, \ldots, w_n$ respectively, then the weighted mean is,

$$\overline{x}_w = \frac{w_1 x_1 + w_2 x_2 + \ldots + w_n x_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Problems related to arithmetic mean

1.$Consider\ the\ 10\ numbers$

$18,15,18,16,17,18,15,19,17,17$

Then $\bar{x} = \frac{18+15+18+16+17+18+15+19+17+17}{10}$

$=\frac{170}{10}=17$

2. *Consider the* 10 *numbers* 18,15,18,16,17,18,15,19,17,17

	15	**16**	**17**	**18**	**19**
	2	**1**	**3**	**3**	**1**

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}, i = 1,2, \ldots\ldots 5$$

$$= \frac{(2X15)+(1X16)+(3X17)+(3X18)+(1X19)}{2+1+3+3+1} = \frac{170}{10} = 17$$

3. *Consider the* 10 *numbers* 18,15,18,16,17,18,15,19,17,17 *are* assigned the weights 1,3,3,3,2,1,2,2,3,2 then the weighted average

$$\overline{x_w} = \frac{\sum x_i w_i}{\sum w_i}, i = 1,2, \ldots..10$$

$$= \frac{(18X1) + (15X3) + (18X3) + (16X3) + (17X2) + (18X1) + (15X2) + (19X2) + (17X3) + (17X2)}{22}$$

$$= \frac{370}{22} = 16.81$$

arithmetic mean (a. m) of a grouped frequency distribution

The Arithmetic Mean (A.M) of a grouped frequency distribution is to defined to be

$\bar{x} = \frac{\sum f_i x_i}{N}$ **Where N=$\sum f_i$ and x_i is the mid-value of the true class interval.**

Class interval			
10-20	15	2	30
20-30	25	3	75
30-40	35	2	70

$$\bar{x} = \frac{\sum f_i x_i}{N}$$

$$= \frac{175}{7}$$

= 25

Find the arithmetic mean (a. m) of a grouped frequency distribution by step-deviation method

A.M.= $\bar{x}$= A+ $\frac{\sum f_i u_i}{\sum f_i} \times h\ , where\ i = 1,2, \dots \dots . n$

A= MID VALUE OF x_i , Where $x_i = \frac{Upper\ class\ limit + Lower\ class\ Limt}{2}$

h= class size

$u_i = \frac{x_i - A}{h}$

4. For the frequency distribution given in the following table, Find the value of $\bar{x}$.

A=24.5 **h= class size =10**

Class interval			$u_i = \frac{x_i - A}{h}$	
-0.5 - 9.5	4.5	11	-2	-22
9.5 – 19.5	14.5	20	-1	-20
19.5 – 29.5	24.5	16	0	0
29.5 – 39.5	34.5	36	1	36
39.5 - 49.5	44.5	17	2	34

$$\bar{x} = A + \frac{\Sigma f_i u_i}{\Sigma f_i} \times h$$

$= 24.5 + \frac{28}{100} \times 10$

27.3

PAGE NO: 15
THEOREM NO: 2.1

The algebraic sum of the deviation of a set of n values from their A.M. arithmetic mean is zero.

In Mathematics and statistics, deviation is a measure of difference between the observed value of a variable and some other value, often that variable's mean

Proof:

Let $x_1, x_2, \ldots x_n$ be the values with frequencies $f_1, f_2, \ldots f_n$ respectively.

$$\therefore \bar{x} = \frac{\sum f_i x_i}{N}, \quad N = \sum f_i \quad \rightarrow (1)$$

The deviation of x_i from the A.M. is given by

$d_i = x_i - \bar{x}$ $(i = 1, 2, \ldots n)$

$$\therefore \sum f_i d_i = \sum f_i (x_i - \bar{x})$$

$$= \sum f_i x_i - \bar{x} \sum f_i = N\bar{x} - \bar{x}N \quad [\text{from } (1)]$$

$$= 0$$

Hence the theorem.

STATISTICS-I UNIT-I TEXT BOOK: ARUMUGAM ISAAC
7BMA5C2

Page NO: 15
Theorem NO: 2.2

The sum of the squares of the deviations of a set of n values is minimum when the deviations are taken from their mean.

Proof: Let $x_1, x_2, \dots x_n$ be the set of n values with the corresponding frequencies $f_1, f_2, \dots f_n$.

$$\therefore \bar{x} = \frac{\sum f_i x_i}{N} \quad \text{where } N = \sum f_i$$

Now, the sum of the squares of the deviations of x_i from an arbitary number A is given by $z = \sum f_i (x_i - A)^2$.

The value of A for which z is minimum is determined by the condition $\frac{dz}{dA} = 0$ and $\frac{d^2z}{dA^2} > 0$

PAGE NO: 15
THEOREM NO: 2.2

$$Z = \Sigma f_i (x_i - A)^2$$

Now, $\frac{dz}{dA} = 0 \Rightarrow -2 \Sigma f_i (x_i - A) = 0$

$$\Sigma f_i (x_i - A) = 0$$

$$\Sigma f_i x_i - \Sigma f_i A = 0$$

$$\Sigma f_i x_i = \Sigma f_i A$$

$$A = \frac{\Sigma f_i x_i}{\Sigma f_i} = \bar{x}$$

Also $\frac{d^2 z}{dA^2} = 2 \Sigma f_i = 2N > 0$

$\therefore$ Z is minimum when $A = \bar{x}$

Hence the theorm.

STATISTICS-I
7BMA5C2

UNIT-I

TEXT BOOK: ARUMUGAM ISAAC

PAGE NO: 16
THEOREM NO: 2.3

If $x_1, x_2, \ldots, x_k$ are the arithmetic means of $n_1, n_2, \ldots, n_k$ observations then the arithmetic mean of the combined set of observations is given by

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1 + n_2 + \ldots + n_k}$$

Proof

$n_1\bar{x}_1$ is the sum of all the n_1 observation in the first set.

$n_2\bar{x}_2$ is the sum of all the n_2 observation in the second set.

⋮

$n_k\bar{x}_k$ is the sum of all the n_k observation in the k^{th} set.

$\sum_{i=1}^{k} n_i\bar{x}_i$ is the sum of all the $(n_1+n_2+\ldots+n_k)$ observations in the combined set.

PAGE NO : 16

THEOREM NO : 2.3

$$\therefore \bar{x} = \frac{1}{N}\left(\sum_{i=1}^{k} n_i \bar{x}_i\right) \text{ where } N = \sum_{i=1}^{k} n_i .$$

Hence the theorem.

STATISTICS-I UNIT-I TEXT BOOK: ARMUGAM ISAAC

7BMA5C2

PAGE NO: 25

PROBLEM NO: 4

The marks scored by 60 students in an examination in statistics is given below. Form a frequency distribution with class intervals of 10 and calculate the A.M.

6 10 58 56 0 25 32 35 35 9 78 17 60 50 35 38 30
10 48 5 68 48 35 30 31 21 23 23 50 72 19 25 35 40
46 42 45 25 60 41 35 36 38 35 33 46 28 31 35 42 46
38 39 45 48 50 28 29 31 55.

RANGE = 78 - 0 = 78, A = 45, h = 10

Soln:

CLASS	MID x_i	f_i	$u_i = \frac{x_i - A}{h}$	$f_i u_i$
0 - 10	5	4	-4	-16
10 - 20	15	4	-3	-12
20 - 30	25	9	-2	-18
30 - 40	35	20	-1	-20
40 - 50	45	12	0	0
50 - 60	55	6	1	6
60 - 70	65	3	2	6
70 - 80	75	2	3	6

$\bar{x} = A + \frac{\sum f_i u_i}{N} \times h$

$= 45 + \frac{(-48)}{60} \times 10$

$= 45 - 8$

$= 37$

STATISTICS-I UNIT-I TEXT BOOK: ARMUGAM ISAAC
7BMA5C2

PAGE NO: 25
PROBLEM NO: 3

Find the mean mark of the following table.

$A = 55$, $h = 10$

MARKS	CLASS	MID x_i	C.F	f_i (No of students)	$u_i = \frac{x_i - A}{h}$	$f_i u_i$
0 and above	0 - 10	5	80	3	-5	-15
10 and above	10 - 20	15	77	5	-4	-20
20 and above	20 - 30	25	72	7	-3	-21
30 and above	30 - 40	35	65	10	-2	-20
40 and above	40 - 50	45	55	12	-1	-12
50 and above	50 - 60	A 55	43	15	0	0
60 and above	60 - 70	65	28	12	1	12
70 and above	70 - 80	75	16	6	2	12
80 and above	80 - 90	85	10	2	3	6
90 and above	90 - 100	95	8	8	4	32
100 and above	100 - 110	105	0	0	5	0

$$\bar{x} = A + \frac{\sum f_i u_i}{N} \times h$$

Where $N = \sum f_i$

$$\bar{x} = 55 + \frac{-26}{80} \times 10$$

$$\bar{x} = 51.75$$

STATISTICS-I UNIT-I TEXT BOOK: ARUMUGAM ISAAC
7BMA5C2

PAGE NO: 21
PROBLEM NO: 6

The four parts of a distribution are as follows. Find the mean of the entire distribution.

	FREQUENCY	MEAN
PART 1	50	61
PART 2	100	70
PART 3	120.	80
PART 4	30	83

$n_1 = 50 \quad \bar{x}_1 = 61$

$n_2 = 100 \quad \bar{x}_2 = 70$

$n_3 = 120 \quad \bar{x}_3 = 80$

$n_4 = 30 \quad \bar{x}_4 = 83$

Soln:

Now, $$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + n_4\bar{x}_4}{n_1 + n_2 + n_3 + n_4}$$

$$= \frac{(50 \times 61) + (100 \times 70) + (120 \times 80) + (30 \times 83)}{50 + 100 + 120 + 30}$$

$$\bar{x} = \frac{22140}{300} = 73.8$$

STATISTICS-I UNIT-I TEXT BOOK: ARUMUGAM ISAAC
7BMA5C2

PAGE NO: 26
PROBLEM NO: 6 (i)

Find the Mean of the following data

$A = 54.5$, $h = 10$

Class	Continuous class	Mid x_i	frequency f_i	$u_i = \frac{x_i - A}{h}$	$f_i u_i$
0 - 9	-0.5 - 9.5	4.5	32	-5	-160
10 - 19	9.5 - 19.5	14.5	65	-4	-260
20 - 29	19.5 - 29.5	24.5	100	-3	-300
30 - 39	29.5 - 39.5	34.5	184	-2	-368
40 - 49	39.5 - 49.5	44.5	228	-1	-228
50 - 59	49.5 - 59.5	54.5	167	0	0
60 - 69	59.5 - 69.5	64.5	98	1	98
70 - 79	69.5 - 79.5	74.5	46	2	92
80 - 89	79.5 - 89.5	84.5	20	3	60
90 - 99	89.5 - 99.5	94.5	0	4	0

$\Sigma f_i = 940$ $\quad \Sigma f_i u_i = -1066$

$$\bar{x} = A + \frac{\Sigma f_i u_i}{\Sigma f_i} \times h$$

$$= 54.5 + \frac{(-1066)}{940} \times 10$$

$$\bar{x} = 43.16$$

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-1

CHAPTER2.2

PART-5

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

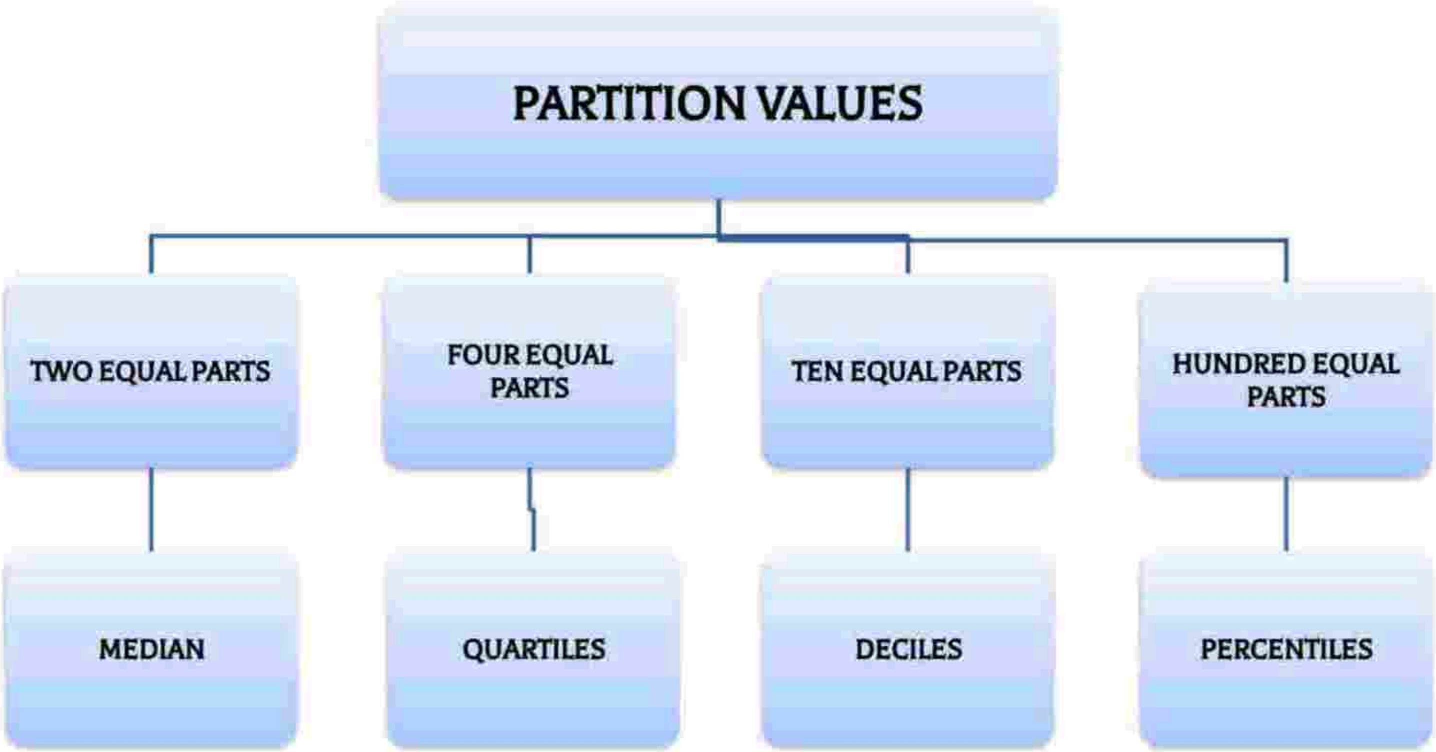
- Partition values
- median
- quartiles
- deciles
- percentiles
- problems

DEFINITION OF PARTITION VALUES

Partition values:

- The variate values dividing the total number of observations into equal number of parts is known as partition values.

- The equal parts may be two, four, ten or hundred.


PARTITION VALUES
TWO EQUAL PARTS
FOUR EQUAL PARTS
TEN EQUAL PARTS
HUNDRED EQUAL PARTS
MEDIAN
QUARTILES
DECILES
PERCENTILES

FOR EXAMPLE:

X	F	LESS THAN C.F
1	5	5
2	9	14
3	18	32
4	12	44
5	9	53
6	7	60
TOTAL	60	-

Here N=60. Hence $\frac{N}{2} = \frac{60}{2} = 30$

The value of 'x ' for which the c.f is ***just greater than 30*** is given by ***x=3.***
Therefore x=3 is the median of the frequency distribution.

DEFINITION OF MEDIAN:

FOR A GROUPED FREQUENCY DISTRIBUTION THE MEDIAN CLASS IS DEFINED TO BE THE CLASS WHERE THE LESS THAN CUMULATIVE FREQUENCY IS JUST GREATER THAN $\frac{N}{2}$.

DEFINITION OF QUARTILES:-

QUARTILES ARE VALUES THAT DIVIDE A COMPLETE GIVEN SET OF OBSERVATIONS INTO FOUR EQUAL PARTS.

BASICALLY,

THERE ARE THREE TYPES OF QUARTILES,

If N is a total frequency,

- FIRST QUARTILE(LOWER QUARTILE)- 'Q1'- $\frac{N}{4}$
- SECOND QUARTILE(MEDIAN)- 'Q2' - $\frac{2N}{4}$
- THIRD QUARTILE(UPPER QUARTILE)-'Q3'- $\frac{3N}{4}$
- $P = \frac{1}{4}, \frac{2}{4}, \frac{3}{4}$ FOR DIFFERENT VALUES OF Q1, Q2, AND Q3 RESPECTIVELY.

DEFINITION OF DECILES:-

- Deciles are those values that divide any set of a given observation into a total of ten equal parts.
- Therefore, there are a total of nine deciles.
- These representation of these deciles are as follows – D1, D2, D3, D4, ……… D9.

- $P = \frac{1}{10}, \frac{2}{10}, \frac{3}{10} \ldots . \frac{9}{10}$ FOR DIFFERENT VALUES OF D1, D2, … D9 RESPECTIVELY.

DEFINITION OF PERCENTILES:-

- The other name for percentiles is centiles.
- A centile or a percentile basically divide any given observation into a total of 100 equal parts.
- The representation of these percentiles or centiles is given as - P1, P2,P3,P4,... P99.

- $P = \frac{1}{100}, \frac{2}{100}, \frac{3}{100}, \dots \frac{4}{100}$ FOR DIFFERENT VALUES OF P1, P2, ... P99 RESPECTIVELY.

Grouped data and ungrouped data

- **Ungrouped data is data given as individual data points.**
- **Grouped data is data given in intervals.**

Ungrouped data

42	26	32	34	57
30	58	37	50	30
53	40	30	47	49
50	40	32	31	40
52	28	23	35	25
30	36	32	26	50
55	30	58	64	52
49	33	43	46	32
61	31	30	40	60
74	37	29	43	54

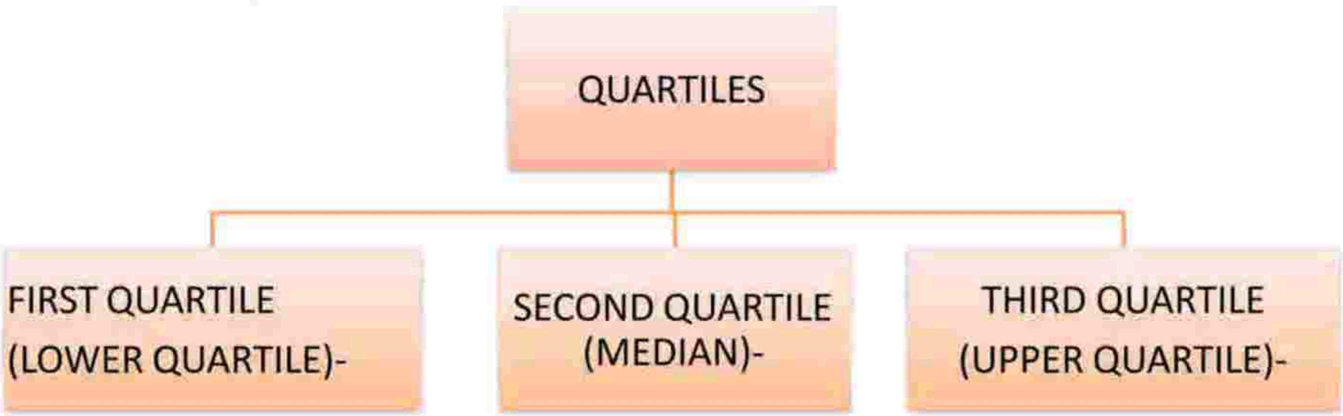
Ages of a Sample of Managers from Urban Child Care Centers in the United States

Grouped data

Frequency Distribution of Child Care Manager's Ages

Class Interval	Frequency
20-under 30	6
30-under 40	18
40-under 50	11
50-under 60	11
60-under 70	3
70-under 80	1

FORMULA TO FIND QUARTILES:-



FORMULA TO FIND FIRST QUARTILE(Q_1):-

UNGROUPED DATA

If n is total number of items then

$$Q_1 = x_i + q(x_{i+1} - x_i)$$

Where,

$i = \left[\frac{1}{4}(n+1)\right] = the\ integral\ part\ of\ \frac{1}{4}(n+1)$

$q = \frac{1}{4}(n+1) - \left[\frac{1}{4}(n+1)\right]$.

Hence q is the fractional part.

GROUPED DATA

If n is the total number of items then

$$Q_1 = l + \frac{\left(\frac{N}{4} - m\right)h}{f_k}$$

* l is the lower boundary of the median class.
* m is the C.F. Above the Median class.
* f_k is the frequency corresponding to the median class
* h is the width of the class.

FORMULA TO FIND MEDIAN(Q_2):-

UNGROUPED DATA

If 'n' is odd

$$Median = \left(\frac{n+1}{2}\right)^{th} term$$

If 'n' is even

$$Median = \frac{\left(\left(\frac{n}{2}\right)^{th\,term} + \left(\frac{n}{2}+1\right)^{th} term\right)}{2}$$

GROUPED DATA

$$\textbf{Median} = l + \frac{\left(\frac{N}{2} - m\right)h}{f_k}$$

Where

* l is the lower boundary of the median class.

* m is the C.F. Above the Median class.

* f_k is the frequency corresponding to the median class

* h is the width of the class.

FORMULA TO FIND THIRD QUARTILE(Q_3):-

UNGROUPED DATA

If n is total number of items then

$$Q_1 = x_l + q(x_{l+1} - x_l)$$

Where,

$$i = \left[\frac{3}{4}(n+1)\right]$$

$$= \textit{the integral part of} \frac{3}{4}(n+1)$$

$$q = \frac{3}{4}(n+1) - \left[\frac{3}{4}(n+1)\right].$$

GROUPED DATA

If n is the total number of items then

$$Q_3 = l + \frac{\left(\frac{3N}{4} - m\right)h}{f_k}$$

* l is the lower boundary of the median class.
* m is the C .F . Above the Median class.
* f_k is the frequency corresponding to the median class
* h is the width of the class.

FORMULA TO FIND DECILES(D_k):-

NOTE: The Median can also be denoted by D_5

UNGROUPED DATA

If n is the total number of items then

$$D_k = x_i + q(x_{i+1} - x_i)$$

Where,

$$i = \left[\frac{k}{10}(n+1)\right]$$

$$= \textit{the integral part of } \frac{k}{10}(n+1)$$

$$q = \frac{k}{10}(n+1) - \left[\frac{k}{10}(n+1)\right]$$

$$for \quad k = 1, 2, 3, 4 \ldots\ldots 9$$

GROUPED DATA

If N is the total number of items

$$D_i = l + \frac{\left(\frac{iN}{10} - m\right)h}{f_k}$$

Where ($i = 1, 2, \ldots. 9$)

* l is the lower boundary of the median class.
* m is the C.F. Above the Median class.
* f_k is the frequency corresponding to the median class
* h is the width of the class.

FORMULA TO FIND DECILES(P_k):-

- NOTE: The Median can also be denoted by P_{50}

UNGROUPED DATA

If n is the total number of items then

$$P_k = x_i + q(x_{i+1} - x_i)$$

Where,

$$i = \left[\frac{k}{100}(n+1)\right]$$

$$= \text{the integral part of } \frac{k}{100}(n+1)$$

$$q = \frac{k}{100}(n+1) - \left[\frac{k}{100}(n+1)\right]$$

$$\text{for} \quad k = 1, 2, 3, 4 \dots \dots 99$$

GROUPED DATA

If N is the total number of items

$$P_i = l + \frac{\left(\frac{iN}{100} - m\right)h}{f_k}$$

Where ($i = 1, 2, \dots. 99$)

* l is the lower boundary of the median class.
* m is the C .F. Above the Median class.
* f_k is the frequency corresponding to the median class
* h is the width of the class.

PROBLEMS:----

1. Find the median and quartile marks of 10 students in Statistics test whose marks are given as 40,90,61,68,72,43,50,84,75,33.

Solution:

$\frac{n}{2}=\frac{10}{2}=5$

$\frac{n}{2}+1=5+1=6$

$\boldsymbol{Median(Q_2)}$:

Arranging in ascending order of given data:33,40,43,50,61,68,72,75,84,90.

Here n=10 is even. Hence

$$\boldsymbol{Median(Q_2)}=\frac{\left(\left(\frac{n}{2}\right)^{th\ term}+\left(\frac{n}{2}+1\right)^{th} term\right)}{2}$$

$$\boldsymbol{Median(Q_2)}=\frac{(5^{th}term+6^{th}term)}{2}=\frac{(61+68)}{2}=-=$$

first Quartile(Q_1): Data in Ascending order : 33,**40,43**,50,61,68,72,75,84,90

$$\textbf{First Quartile } Q_1 = x_i + q(x_{i+1} - x_i)$$

Where,

$$i = \left[\frac{1}{4}(n+1)\right] = \textit{the integral part of } \frac{1}{4}(n+1)$$

$$i = \left[\frac{1}{4}(10+1)\right] = \left[\frac{11}{4}\right] = [2.75] = 2$$

$$q = \frac{1}{4}(n+1) - \left[\frac{1}{4}(n+1)\right]$$

q =2.75 - 2 = 0.75

$$\textit{Therefore}, Q_1 = x_i + q(x_{i+1} - x_i)$$

$$Q_1 = x_2 + q(x_3 - x_2)$$

$$Q_1 = 40 + 0.75(43 - 40)$$

Q_1=42.5

Third Quartile(Q_3): Data in Ascending order : 33,40,43,50,61,68,72,75,84,90

$$\textit{Third Quartile } Q_3 = x_i + q(x_{i+1} - x_i)$$

Where,

$$i = \left[\frac{3}{4}(n+1)\right] = \textit{the integral part of } \frac{3}{4}(n+1)$$

$$i = \left[\frac{3}{4}(10+1)\right] = \left[\frac{33}{4}\right] = [8.25] = 8$$

$$q = \frac{3}{4}(n+1) - \left[\frac{3}{4}(n+1)\right]$$

$$q = 8.25 - 8 = 0.25$$

$$\textit{Therefore}, Q_3 = x_i + q(x_{i+1} - x_i)$$

$$Q_3 = x_8 + q(x_9 - x_8)$$

$$Q_3 = 75 + 0.25(84 - 75)$$

$$Q_3 = 77.25$$

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-1

CHAPTER2.2

PART-6

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

PROBLEMS:----

 Text book page no:36 ----- Problem 3

1. Find the (i)Mean (ii) Median (iii)first quartile (iv)third quartile (v)9th decile (vi)19th percentile for the following frequency distribution.

Class	frequency
11-15	8
16-20	15
21-25	39
26-30	47
31-35	52
36-40	41
41-45	28
46-50	16
51-55	4
TOTAl	**250**

SOLUTIO N: A=33

h=15.5-10.5=5

Class	Mid(xi)	Freq *f*			Less than C.F
10.5-15.5	13	8	-4	-32	8
15.5-20.5	18	15	-3	-45	23
20.5-25.5	**23**	**39**	**-2**	**-78**	**62** **P19**
25.5-30.5	**28**	**47**	**-1**	**-47**	**109** **Q1**
30.5-35.5	**33**	**52**	**0**	**0**	**161** **Q2**
35.5-40.5	**38**	**41**	**1**	**41**	**202** **Q3**
40.5-45.5	**43**	**28**	**2**	**56**	**230** **D9**

(i) Mean:

$$Mean = A + \frac{\sum f_i u_i}{N} \times h$$

$$Mean = 33 + \frac{(-41)}{250} \times 5$$

$$Mean = 33 - 0.82 = 32.18$$

(ii)Median:

Here $\frac{N}{2} = \frac{250}{2} = 125$ and h=5

Hence Median class is 30.5-34.5

$Therefore\ l = 30.5; m = 109; f_k = 52$

Median$(Q_2) = \boldsymbol{l} + \frac{\left(\frac{N}{2} - m\right)h}{f_k}$

$(Q_2) = 30.5 + \frac{(125-109)}{52} \times 5 = 32.04$

(iii) First Quartile:

Here $\frac{N}{4} = \frac{250}{4} = 62.5\ and\ h = 5$

Hence the first quartile class is 25.5-30.5

Therefore $l = 25.5; m = 62; f_k = 47$

First Quartile$(Q_1) = l + \frac{\left(\frac{N}{4} - m\right)h}{f_k}$

$(Q_1) = 25.5 + \frac{(62.5-62)}{47} \times 5 = 25.55$

(iii) Third Quartile:

Here $\frac{3N}{4} = \frac{250}{4} = 187.5 \ \ and\ h = 5$

Hence the Third quartile class is 35.5-40.5

Therefore $l = 35.5; m = 161; f_k = 41$

First Quartile$(Q_1) = l + \frac{\left(\frac{3N}{4} - m\right)h}{f_k}$

(Q_1)=$35.5 + \frac{(187.5-161)}{41} \times 5 = 38.73$

(iii) 9th Decile:

Here $\frac{9N}{10} = \frac{9\times250}{10} = 225 \ \ and\ h = 5$

Hence the 9th Decile class is 40.5-45.5

Therefore $l = 40.5; m = 202; f_k$ =28

First Quartile$(D_9) = l + \frac{\left(\frac{9N}{10} - m\right)h}{f_k}$

(Q_1)=$40.5 + \frac{(225-202)}{28} \times 5 = 44.61$

(iii) 19th Percentile:

Here $\frac{19N}{100} = \frac{19\times250}{100} = 47.5 \ \ and\ h = 5$

Hence the 19th Percentile class is 20.5-25.5

Therefore $l = 20.5; m = 23; f_k = 39$

First Quartile$(P_{19}) = l + \frac{\left(\frac{19N}{100} - m\right)h}{f_k}$

(Q_1)=$20.5 + \frac{(47.5-23)}{39} \times 5 = 23.64$

Text book page no:38 ------ problem 4

2. From the following data calculate the percentage of tenants paying monthly rent (i) More than 105. (ii) between 130 and 190.

Monthly rent	No.of tenants	
60-80	18	
80-100	21	
100-120	45	← More than 105
120-140	85	} Between 130-190
140-160	88	}
160-180	75	}
180-200	18	}
TOTAL	**350**	

SOLUTION:--

(i) Number of tenants

paying more than Rs.105 is $= \left[\frac{(120-105)}{20} \times 45\right] + 85 + 88 + 75 + 18 = 300$

Required percentage $= \frac{300}{350} \times 100 = 85.7$(approximately)

(ii)Number of tenants

Paying rent between Rs.130 and Rs.190 $= \left[\frac{(140-130)}{20} \times 85\right] + 88 + 75 + \left[\frac{(190-180)}{20} \times 18\right] = 215$

Required percentage$= \frac{215}{350} \times 100 = 61.43$(approximately)

Text book page no:39 ------ problem 5

3. An in complete distribution is given below. The Median is 35 Find the missing frequencies.

Class	frequency	Less than C.F
0-10	10	10
10-20	20	30
20-30	? f1	30+f1
30-40	**40**	**70+f2**
40-50	? f2	70+f1+f2
50-60	25	95+f1+f2
60-70	15	110+f1+f2
TOTAL	**170**	

SOLUTION:-

Let the frequency corresponding to the class 20-30 be f_1 and that of class 40-50 be f_2 .

$$f_1 + f_2 + 10 + 20 + 40 + 25 + 15 = 170$$

$$f_1 + f_2 + 110 = 170$$

$$f_1 + f_2 = 170 - 110 = 60$$

Here the **Median class is 30-40**

Therefore $l = 30; m = 30 + f_1; f_k = 40$ and h=10

$$\text{Median}(Q_2) = \boldsymbol{l} + \frac{\left(\frac{N}{2} - \boldsymbol{m}\right)\boldsymbol{h}}{\boldsymbol{f_k}}$$

$$\text{Median}(Q_2) = 30 + \frac{\left(\frac{170}{2} - (30 + f_1)\right)}{40} \times 10$$

$$35 = 30 + \left[\frac{(85 - 30 - f_1)}{40}\right] \times 10$$

$$35 = 30 + \left[\frac{(55 - f_1)}{40}\right] \times 10$$

$$35 - 30 = \left[\frac{(55 - f_1)}{4}\right]$$

$$5 \times 4 = (55 - f_1)$$

$$20 = (55 - f_1)$$

$f_1 = 35$ and $f_1 + f_2 = 60$

$$35 + f_2 = 60$$

$$f_2 = 60 - 35 = 25$$

Text book page no:41------ exercise problem 3

3. An in complete distribution is given below. The Median is 35 Find the missing frequencies.

Wages in Rs.	No.of.Students
More than 100	5
More than 90	17
More than 80	37
More than 70	43
More than 60	49
More than 50	49
More than 40	51
TOTAL	

Class	No.of. workers	Less than c.f
40-50	2	2
50-60	0	2
60-70	6	8
70-80	6	14
80-90	**20**	**34 Me** ←
90-100	12	46
100-110	5	51

Median:

Here $\frac{N}{2}=\frac{51}{2}=25.5$ and h=10

Hence Median class is 80 - 90

$$Therefore$$

$l=80; m=14; f_k=20$

$$\text{Median}(Q_2)=l+\frac{\left(\frac{N}{2}-m\right)h}{f_k}$$

$$(Q_2)=80+\frac{(25.5-14)}{20}\times 10=85.75$$

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-1

CHAPTER2.3

PART-7

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

- DEFINITION OF MODE
- COMPUTING MODE FOR INDIVIDUAL DATAS
- COMPUTING MODE FOR DISCRETE AND CONTINUOUS DATAS
- COMPUTING MODE BY INSPECTION METHOD
- COMPUTING MODE BY METHOD OF GROUPING
- FORMULAE TO FIND MODE
- RELATIONSHIP BETWEEN MEAN, MEDIAN AND MODE
- PROBLEMS
- ASSIGNMENT

DEFINITION OF MODE

In a distribution the value of the variate which

occurs

most frequently and around which the other values

of

variates cluster densely is called the mode or modal

value of the distribution.

TEXT BOOK PAGE NO:51, EXERCISE Q.NO:1

- COMPUTING MODE BY INSPECTION METHOD

Example:

1. Find the mean, median and mode for the set of numbers.

(i) 6, 8, 2, 5, 9, 5, 6, 5, 2, 3.

Solution:

Here most frequently occurring data is 5.
Hence Mode is 5.

(ii)61.7, 71.8, 65.3, 70, 69.8

Solution:

In this case there is no mode because none of the numbers is repeated.

- COMPUTING MODE FOR DISCRETE AND CONTINUOUS DATAS

By inspection

- For discrete distribution, when the values of individual items are known, mode can be determined just by inspection. By inspection you can find out the value of the variate around which the items are most heavily concentrated.

By grouping

- Difficulty arises in both discrete and continuous data, when nearly equal concentrations are found in two or more neighbouring classes; i.e., there is a small . difference between the maximum frequency and the frequency preceding it or succeeding it. To locate a modal class in such situations, there is a need for Grouping and Analysis.

By inspection

Example:

2. Consider the discrete frequency distribution.

X	1	2	3	4	5	6	7	8	9	10
y	8	13	47	105	28	9	5	3	2	1

Solution:

By inspection ,

Here the maximum frequency is 105.

The value corresponding to this maximum frequency is 4.

Hence Mode is 4.

- COMPUTING MODE BY METHOD OF GROUPING

By grouping

1.

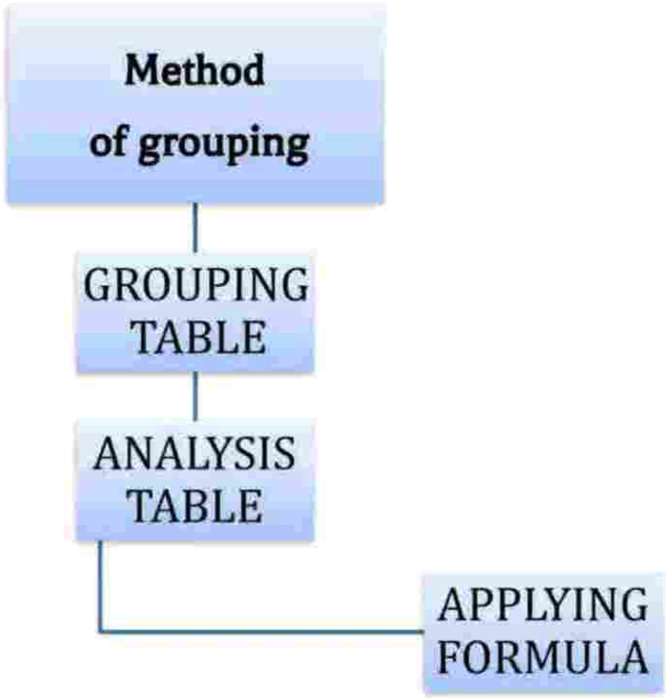
Size of wages	3	4	5	6	7	8	9	10
Persons wearing it	10	28	38	42	45	15	8	7

2.

Marks	No. Of. Students
0-9	6
10-19	29
20-29	87
30-39	181
40-49	247
50-59	263
60-69	133
70-79	43
80-89	9
90-99	2

In these both problems difficulties occurs to selecting the modal class. So, we have to choose grouping method to solve this problem.

- **COMPUTING MODE BY METHOD OF GROUPING**



Grouping Table :

A grouping table has six columns as explained below:

- **Column 1 :** It is of class frequencies written against each class.
- **Column 2 :** Frequencies are grouped in this column in two's, and totals are found. Then the highest total is marked or circled.
- **Column 3 :** Leaving first frequency from the top, the remaining frequencies are again grouped in two's and the highest total is marked.
- **Column 4 :** Starting from the top, frequencies are grouped in three's, their totals are ' obtained and the highest total is marked. '
- **Column 5 :** Leaving first frequency, they are again grouped in three's. Their totals are obtained and the highest total is marked.
- **Column 6** : . Leaving the first two frequencies from the top, remaining frequencies are grouped in three's. Their totals are calculated and the highest total is marked..

Analysis Table :

After preparing a grouping table, an analysis table is prepared. The value containing the maximum frequency are noted down for each column and are written in a table called Analysis table.

Applying formulae :

✓ In the case of grouped frequency distribution the mode is computed by the formula mode .

$$Mode = l + \frac{(f - f_1)}{2f - f_1 - f_2} \times h$$

Where,

- l is the lower boundary of the modal class(Class having maximum frequency);
- f is the maximum frequency;
- f_1 and f_2 are the frequencies of the classes preceding and following the modal class.

✓ In the case of grouped frequency distribution the mode is computed by the another formula mode .

$$Mode = l + \frac{h(f_2)}{f_1 + f_2}$$

Where,

- l is the lower boundary of the modal class(Class having maximum frequency);
- f_1 and f_2 are the frequencies of the classes preceding and following the modal class.

- **RELATIONSHIP BETWEEN MEAN, MEDIAN AND MODE**

There is an interesting empirical relationship between mean, median, mode which appears to hold for unimodal curves of moderate asymetry namely,

Mean-Mode=3(Mean-Median)

- **Mode=3Median-2Mean**

- TEXT BOOK PAGE NO:52 , EXERCISE PROBLEM NO:8

1. Determine the modal class and hence find the mode for the 100 articles which follow the following frequency distribution.

Marks	No. Of. Students
0-9	6
10-19	29
20-29	87
30-39	181
40-49	247
50-59	263
60-69	133
70-79	43
80-89	9
90-99	2

Solution:

By inspection it is difficult to say which is the modal class. Hence we determine the modal class by forming the grouping table.

GROUPING TABLE

<table>
<tr><th>Marks</th><th>CONTINUOUS CLASS</th><th>I</th><th>II</th><th>III</th><th>IV</th><th>V</th><th>VI</th></tr>
<tr><td>0-9</td><td>-0.5-9.5</td><td>6</td><td rowspan="2">35</td><td></td><td rowspan="3">122</td><td></td><td></td></tr>
<tr><td>10-19</td><td>9.5-19.5</td><td>29</td><td rowspan="2">116</td><td rowspan="3">297</td><td></td></tr>
<tr><td>20-29</td><td>19.5-29.5</td><td>87</td><td rowspan="2">268</td><td rowspan="3">515</td></tr>
<tr><td>30-39</td><td>29.5-39.5</td><td>181</td><td rowspan="2">428</td><td rowspan="3">691</td></tr>
<tr><td>40-49</td><td>39.5-49.5</td><td>247</td><td rowspan="2">510</td><td rowspan="3">643</td></tr>
<tr><td>50-59</td><td>49.5-59.5</td><td>263</td><td rowspan="2">396</td><td rowspan="3">439</td></tr>
<tr><td>60-69</td><td>59.5-69.5</td><td>133</td><td rowspan="2">176</td><td rowspan="3">185</td></tr>
<tr><td>70-79</td><td>69.5-79.5</td><td>43</td><td rowspan="2">52</td><td rowspan="3">54</td></tr>
<tr><td>80-89</td><td>79.5-89.5</td><td>9</td><td rowspan="2">11</td><td></td></tr>
<tr><td>90-99</td><td>89.5-99.5</td><td>2</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">TOTAL</td><td>1000</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

- Next we construct Analysis table

ANALYSIS TABLE

COLUMNS	-0.5-9.5	9.5-19.5	19.5-29.5	29.5-39.5	39.5-49.5	49.5-59.5	59.5-69.5	69.5-79.5	79.5-89.5	89.5-99.5
I						I				
II					I	I				
III				I	I					
IV				I	I	I				
V					I	I	I			
VI			I	I	I					
Tot			1	3	5	4	1			

CONCLUSION

:

Form the Analysis table , It is clear that the modal class is 39.5-49.5.

$$\frac{263}{1000} \times 100 = 26.3$$

$$\frac{181}{1000} \times 100 = 18.1$$

Hence l=39.5; f_1=181; f_2=263 and h=10.

$$Mode = l + \frac{h(f_2)}{f_1 + f_2}$$

$Mode(for\ the\ 1000\ articles)$

$$= 39.5 + \frac{10\times263}{181+263} = 39.5+5.92=45.42$$

Mode for the 100 articles

$$= 39.5 + \frac{10\times26.3}{18.1+26.3} = 39.5+5.92=45.42$$

- **TEXT BOOK PAGE NO:52 , EXERCISE PROBLEM NO:5**

2. Calculate the mode from the data given below

Wages in Rs.	Number of workers
Above 30	520
Above 40	470
Above 50	399
Above 60	210
Above 70	104
Above 80	45
Above 90	7
Above 100	0

Solution:

GROUPING TABLE

<table>
<tr><th>Wages in Rs.</th><th>I
Number of workers[F]</th><th>II</th><th>III</th><th>IV</th><th>V</th><th>VI</th></tr>
<tr><td>30-40</td><td>50</td><td rowspan="2">121</td><td></td><td rowspan="3">310</td><td></td><td></td></tr>
<tr><td>40-50</td><td>71</td><td rowspan="2">260</td><td rowspan="3">366</td><td></td></tr>
<tr><td>50-60</td><td>189</td><td rowspan="2">295</td><td rowspan="3">354</td></tr>
<tr><td>60-70</td><td>106</td><td rowspan="2">165</td><td rowspan="3">203</td></tr>
<tr><td>70-80</td><td>59</td><td rowspan="2">97</td><td rowspan="3">104</td></tr>
<tr><td>80-90</td><td>38</td><td rowspan="2">45</td><td rowspan="3">45</td></tr>
<tr><td>90-100</td><td>7</td><td rowspan="2">7</td><td></td></tr>
<tr><td>100-110</td><td>0</td><td></td><td></td><td></td></tr>
</table>

- Next we construct Analysis table

ANALYSIS TABLE

COLUMNS	30-40	40-50	50-60	60-70	70-80	80-90	90-100	100-110
I			I					
II			I	I				
III		I	I					
IV	I	I	I					
V		I	I	I				
VI			I	I	I			
Tot	1	3	6	3	1			

CONCLUSION

:

Form the Analysis table , It is clear that the modal class is 50-60

Hence l=50; f_1=71; f_2=106 and h=10.

$$Mode = l + \frac{h(f_2)}{f_1 + f_2}$$

$$Mode = 50 + \frac{10\times106}{71+106}=50+5.99=55.99$$

 TEXT BOOK PAGE NO:52 , EXERCISE PROBLEM NO:5

3. The Expenditure of 100 families is given below. Mode for the distribution is 24. Calculate the missing frequencies.

Expenditure	**Number of families**
0- 10	14
10-20	
20-30	27
30-40	
40-50	15

Solution:

Expenditure	Number of families	Less than C.F
0- 10	14	14
10-20		
20-30	27	
30-40		
40-50	15	

Here N= 100, we get

$$f_1 + f_2 = 100 - (27 + 14 + 15)$$
$$= 100 - 56 = 44$$

$$Mode = l + \frac{h(f - f_1)}{2f - f_1 - f_2}$$

$$Mode = 20 + \frac{10 \times (27 - f_1)}{54 - (f_1 + f_2)}$$

$$24 = 20 + \frac{10 \times (27 - f_1)}{54 - 44}$$

$$24 = 20 + \frac{10 \times (27 - f_1)}{10}$$

$24 = 20 + (27 - f_1)$

$24 - 20 = 27 - f_1$

$4 = 27 - f_1$

$\mathbf{f_1 = 23}$

$f_2 = 44 - f_1 = 44 - 23 = 21$

$\mathbf{f_2 = 21}$

- **ASSIGNMENT**

Text book page No.	Question.No.
51	3,4[exercise problems]
52	6,7[exercise problems]

Hint:

For question No 3

Use this formula Mode=3Median-2Mean

For Question No. 4

find Mean and Median . To find mode use this formula

Mode=3Median-2Mean

For question no 7

Find grouping table and Analysis table then find the highest frequency size.

Here No need to use formula.

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-1

CHAPTER2.4

PART-8

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

- DEFINITION OF GEOMETRIC MEAN
- DEFINITION OF HARMONIC MEAN
- PROBLEMS
- ASSIGNMENT

- ## DEFINITION OF GEOMETRIC MEAN

The geometric mean (G.M.) of a set of n observations $x_1, x_2, x_2, \ldots . x_n$ is the n^{th} root of their product.

Thus, Geometric mean is $G = (x_1 x_2 x_3 \ldots . x_n)^{1/n}$

Therefore,

$$logG = \frac{1}{n}(\log x_1 + \log x_2 + \cdots .. \log x_n)$$

$$= \frac{1}{n}\sum \log x_i$$

$$G = anti \ \log\left[\frac{1}{n}\sum \log x_i\right]$$

In case of grouped frequency distribution geometric mean

$$G = \left(x_1{}^{f_1} x_2{}^{f_2} \ldots \ldots \ldots \ldots . x_n{}^{f_n}\right)^{1/N} \text{ where } N=\sum f_i$$

$$G = anti \ \log\left[\frac{1}{N}\sum f_i \log x_i\right]$$

DEFINITION OF HARMONIC MEAN

- Harmonic mean of the set of n observations $x_1, x_2, x_2, \dots . x_n$ is defined to be the reciprocal of the arithmetic mean of the reciprocal of the observations.

  Thus Harmonic mean $H = \frac{1}{\frac{1}{n}\left(\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}\right)}$

- In case of a grouped frequency distribution

  Harmonic mean $H = \frac{1}{1/_N\left[\sum \frac{f_i}{x_i}\right]}$ $where\ N = \sum f_i$

- TEXT BOOK PAGE NO:57 , EXERCISE PROBLEM NO:5

1.Calculate the G.M. and H.M. of the following distribution.

Class	**2-4**	**4-6**	**6-8**	**8-12**
Frequency	20	40	30	10

Solution:

Class	Mid(Xi)	Log(Xi)	fi	fiLog(Xi)	fi/Xi
2-4	3	0.4771	20	9.542	6.6667
4-6	5	0.6989	40	27.956	8
6-8	7	0.8451	30	25.353	4.2857
8-12	10	1	10	10	1
			100	**72.851**	**19.9524**

$$G = anti\ \log\left[\frac{1}{N}\sum f_i \log x_i\right]$$

$$G = anti\ \log\left[\frac{1}{100}(72.851)\right]$$

$$G = anti\ \log[0.72851]$$

G=5.352

$$H = \frac{1}{{}^{1}/_{N}\left[\sum \frac{f_i}{x_i}\right]}$$

$$H = \frac{1}{{}^{1}/_{100}\,[19.9524]}$$

$$H = \frac{1}{0.199524}$$

H=5.012

3. Calculate A.M.,G.M., and H.M. of the following observations and Show that A.M> G.M> H.M. 32, 35, 36, 37, 39, 41, 43.

Solution:

Arithmetic Mean $= \frac{\sum x_i}{n} = \frac{32+35+36+37+39+41+43}{7} = \frac{263}{7} = 37.57$

Geometric Mean $G = anti\ \log\left[\frac{1}{n}\sum \log x_i\right]$

$$=anti\ log\left[\frac{\log 32+\log 35+\log 36+\log 37+\log 39+\log 41+\log 43}{7}\right]$$

$$= anti\ log\left[\frac{[1.505 + 1.544 + 1.556 + 1.568 + 1.591 + 1.613 + 1.633]}{7}\right]$$

$= antilog[1.57285714]$=37.399

Harmonic mean $H = \frac{1}{\frac{1}{n}\left(\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}\right)}$

Harmonic mean $H = \frac{1}{\frac{1}{7}\left(\frac{1}{32}+\frac{1}{35}+\frac{1}{36}+\frac{1}{37}+\frac{1}{39}+\frac{1}{41}+\frac{1}{43}\right)}$

H= $\frac{1}{\frac{1}{7}(0.03125+0.0285714+0.0277+0.027+0.025641+0.02439+0.023255)}$

$$= \frac{1}{0.02683}$$

H=37.27

Here , 37.57>37.399>37.27

Hence A.M>G.M>H.M

- **ASSIGNMENT**

Text book page No.	Question.No.
54	1,2
55	3
56	4,5
57	3,4

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-1
CHAPTER3.1
PART-9

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

- DEFINITION OF MESURES OF DISPERSION
- DEFINITION OF QUARTILE DEVIATION
- DEFINITION OF MEAN DEVIATION
- DEFINITION OF ROOT MEAN SQUARE DEVIATION
- THEOREMS
- PROBLEMS
- ASSIGNMENT

CHAPTER 3

MEASURES OF DISPERSIO

3.0 INTRODUCTION

Measures of central tendency give an idea of the concentration of the observations about the central part of the distributions. However these measures are inadequate to give a complete idea of the distribution.

For example consider the two sets of observation.

I	9	11	15	14	15	16	17	18	20
II	2	5	9	15	15	15	21	25	28

We have 15 as the mean, median and mode for the two sets of observations. (verify) We observe that the individual items in the second set are scattered from the mean whereas in the first set they are closely packed. Thus we cannot form a complete idea about the distributions from these averages. Hence averages must be supported and suplemented by some other measures. One such is measure of dispersion.

3.1 MEASURES OF DISPERSION

Definition. **Dispersion** of a distribution is the amount of scatteredness of the individual values from a measure of central tendency. There are four measures of dispersion which are in common use. They are as follows (i) *Range* (ii) *Quartile* (iii) *Mean deviation* (iv) *Standard deviation*.

Range. *Range is the most simple and obvious measure of dispersion. It is the difference between the maximum and the minimum value of the variate.*

Example. For the data given in *Table 2* the maximum value is 49 and the minimum value is 1. Hence the range is 48.

Quartile deviation. (Semi inter quartile range)

The **quartile deviation** (Q.D) or **semi inter quartile range** is defined by $Q.D = \frac{1}{2}(Q_3 - Q_1)$ where Q_1 and Q_3 are the first and the third quartiles of the distribution.

Example. For the data in *Table 5*, $Q_1 = 17$ and $Q_3 = 38.5$. (refer example of quartiles in section 2.2).

Hence $Q.D = \frac{1}{2}(38.5 - 17) = 10.75$.

Mean deviation. The mean deviation of a frequency distribution from any average A is defined by $M.D = \frac{\Sigma f_i |x_i - A|}{N}$ where $N = \Sigma f_i$.

Example. For the data given in *Table 4*, $\bar{x} = 27.3$ (refer example under A.M in section 2.1) Now we find the mean deviation from the mean.

Mid x_i	f_i	$\lvert x_i - 27.3 \rvert$	$f_i \lvert x_i - 27.3 \rvert$
4.5	11	22.8	250.8
14.5	20	12.8	256.0
24.5	16	02.8	044.8
34.5	36	07.2	259.2
44.5	17	17.2	292.4
Total	100	-	1103.2

$\therefore$ M.D about mean $= \frac{1103.2}{100} = 11.032$

Standard deviation.

A common measure of dispersion which is prev... circumstances in statistics is the standard deviation.

Definition. The **standard deviation** σ of a frequency distrib... by $\sigma = \left[\frac{\Sigma f_i (x_i - \bar{x})^2}{N}\right]^{1/2}$ where $N = \Sigma f_i$ and $\bar{x}$ is the arithm... the frequency distribution.

The square of the standard deviation of a frequency distr... called the **variance** of the frequency distribution. Hence **variance** $=$...

Note. If σ_s^2 is the variance of a sample of size n the "best" estima... population variance σ_p^2 is not σ_s^2 but $\left(\frac{N}{N-1}\right)\sigma_s^2$. For this reas... authors define standard deviation by the formula $\sigma = \left[\frac{\Sigma f_i (x_i - \bar{x})^2}{N-1}\right]^{1/2}$

For large values of N the two formulae for standard deviat... practically indistinguishable. Throughout this book we use the first fo... for finding standard deviation of a frequency distribution. Both the for... for standard deviation find place in modern calculators.

Definition. The **root mean square deviation** of a frequency distributi... defined to be $s = \left[\frac{\Sigma f_i (x_i - A)^2}{N}\right]^{1/2}$ where A is any arbitrary origin... s^2 is called the **mean square deviation**.

Definition. **Coefficient of variation** of a frequency distribution is defined... be $C.V = \frac{\sigma}{\bar{x}} \times 100$.

For comparing the variability of two sets of observations of a frequency distribution we calculate the C.V for each of the set of frequency distribution. The set having smaller C.V is said to be more consistent than the other.

Example 1. Consider the numbers 1, 2, 3, 4, 5, 6, 7.

Their arithmetic mean $\bar{x} = 4$.

Now, $\Sigma (x_i - 4)^2 = 28$. (verify)

$$\therefore \sigma = \left[\frac{\Sigma (x_i - 4)^2}{7}\right]^{1/2} = \left(\frac{28}{7}\right)^{1/2} = 2.$$

Example 2. For the frequency distribution given in *Table 4*, $\bar{x} = 27.3$. Hence we have the following table.

x_i	f_i	$x_i - 27.3$	$(x_i - 27.3)^2$	$f_i (x_i - 27.3)^2$
04.5	11	− 22.8	519.84	5718.24
14.5	20	− 12.8	163.84	3276.80
24.5	16	− 2.8	7.84	125.44
34.5	36	7.2	51.84	1866.24
44.5	17	17.2	295.84	5029.28
Total	**100**	-	-	**16016**

$$\therefore \sigma^2 = \frac{1}{N}\Sigma f_i (x_i - \bar{x})^2 = \frac{16016}{100} = 160.16.$$

$$\therefore \sigma = 12.66$$

We now establish a relation between the root mean square deviation s and standard deviation σ.

Theorem 3.1 $\sigma^2 = s^2 - d^2$ where $d = \bar{x} - A$.

Proof. $s^2 = \frac{\Sigma f_i (x_i - A)^2}{N}$

$$= \frac{\Sigma f_i (x_i - \bar{x} + \bar{x} - A)^2}{N}$$

$$= \frac{1}{N}\left[\Sigma f_i (x_i - \bar{x})^2 + 2\Sigma f_i (x_i - \bar{x})(\bar{x} - A) + \Sigma f_i (\bar{x} - A)^2\right]$$

$$= \frac{\Sigma f_i (x_i - \bar{x})^2}{N} + \frac{2d}{N}\Sigma f_i (x_i - \bar{x}) + d^2.$$

$$= \sigma^2 + d^2 \quad (\text{since } \Sigma f_i (x_i - \bar{x}) = 0)$$

$$\therefore \sigma^2 = s^2 - d^2.$$

Corollary. The standard deviation is the least possible root mean square deviation.

Proof. We have $s^2 = \sigma^2 + d^2$.

$\therefore s^2$ is least when $d = 0$. Hence the least value of s^2 is σ^2.

The following theorem gives another formula for calculation of standard deviation of a frequency distribution.

Theorem 3.2 $\sigma = \left[\frac{\Sigma f_i x_i^2}{N} - \left(\frac{\Sigma f_i x_i}{N}\right)^2\right]^{1/2}$

Proof. $\sigma^2 = (1/N)\Sigma f_i (x_i - \bar{x})^2$

$$= (1/N)\left[\Sigma f_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2)\right]$$

$$= \frac{\Sigma f_i x_i^2}{N} - 2\bar{x}\left(\frac{\Sigma f_i x_i}{N}\right) + \bar{x}^2\left(\frac{\Sigma f_i}{N}\right)$$

$$= \frac{\Sigma f_i x_i^2}{N} - \bar{x}^2 = \frac{\Sigma f_i x_i^2}{N} - \left(\frac{\Sigma f_i x_i}{N}\right)^2$$

$$\therefore \sigma = \left[\frac{\Sigma f_i x_i^2}{N} - \left(\frac{\Sigma f_i x_i}{N}\right)^2\right]^{1/2}$$

Theorem 3.3 The standard deviation σ is independent of change of origin and is dependent on change of scale.

Proof. We have $\sigma_x^2 = (1/N)\Sigma (x_i - \bar{x})^2$.

Suppose we change the variable x_i to u_i where $u_i = x_i - A$, A being an arbitrary origin.

We know that $\bar{u} = \bar{x} - A$.

Now, $u_i - \bar{u} = x_i - \bar{x}$

Now, $\sigma_x^2 = (1/N)\Sigma f_i (x_i - \bar{x})^2 = (1/N)\Sigma f_i (u_i - \bar{u})^2$

$$= \sigma_u^2$$

Hence σ is independent of change of origin.

Now, suppose we change the variable x_i to v_i where $v_i = x_i/h$.

Then $\bar{v} = \bar{x}/h$

$\therefore v_i - \bar{v} = (1/h)(x_i - \bar{x})$.

Now, $\sigma_x^2 = (1/N)\Sigma f_i (x_i - \bar{x})^2 = (h^2/N)\Sigma f_i (v_i - \bar{v})^2$

$$= h^2 \sigma_v^2$$

$\therefore$ S.D is dependent on change of scale.

Note : When we effect a change in origin as well as in scale σ^2 is multiplied by the square of the scale introduced

$$\text{Hence, } \sigma_x^2 = h^2\left[\frac{\Sigma f_i u_i^2}{N} - \left(\frac{\Sigma f_i u_i}{N}\right)^2\right]$$

Theorem 3.4 (Variance of combined set). Let the mean and standard deviation of two sets containing n_1 and n_2 members be $\bar{x}_1$, $\bar{x}_2$ and σ_1, σ_2 respectively. Suppose the two sets are grouped together as one set of $(n_1 + n_2)$ members. Let $\bar{x}$ be the mean and σ be the standard deviation of this set. Then $\sigma^2 = \frac{1}{n_1 + n_2}\left[n_1\left(\sigma_1^2 + d_1^2\right) + n_2\left(\sigma_2^2 + d_2^2\right)\right]$

where $d_1 = \bar{x}_1 - \bar{x}$ and $d_2 = \bar{x}_2 - \bar{x}$.

Proof. $\sigma^2 = \frac{1}{n_1 + n_2}\left[\sum_{i=1}^{n_1+n_2} f_i (x_i - \bar{x})^2\right]$

$$= \frac{1}{n_1 + n_2}\left[\sum_{i=1}^{n_1} f_i (x_i - \bar{x})^2 + \sum_{i=n_1+1}^{n_1+n_2} (x_i - \bar{x})^2\right]$$

Now, $\sum_{i=1}^{n_1} f_i (x_i - \bar{x})^2 = \sum_{i=1}^{n_1} f_i (x_i - \bar{x}_1 + \bar{x}_1 - \bar{x})^2 = \sum_{i=1}^{n_1} f_i\left[(x_i - \bar{x}_1) + d_1\right]^2$

$$= \sum_{i=1}^{n_1} f_i (x_i - \bar{x}_1)^2 + 2d_1\sum_{i=1}^{n_1} f_i (x_i - \bar{x}_1) + d_1^2 \sum_{i=1}^{n_1} f_i$$

$$= n_1 \sigma_1^2 + n_1 d_1^2 \quad (\text{since } \Sigma f_i(x_i - \bar{x}_1) = 0)$$

Similarly $\sum_{i=n_1+1}^{n_1+n_2} f_i (x_i - \bar{x})^2 = n_2 \sigma_2^2 + n_2 d_2^2$

Hence $\sigma^2 = \frac{1}{n_1 + n_2}\left[(n_1\sigma_1^2 + n_1 d_1^2) + n_2\sigma_2^2 + n_2 d_2^2)\right]$ ...... (1)

$$= \frac{1}{n_1 + n_2}\left[n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)\right].$$

Note. The above formula for σ^2 can also be written as

$$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2} + \frac{n_1 n_2}{(n_1 + n_2)^2}(\bar{x}_1 - \bar{x}_2)^2.$$

We have $d_1 = \bar{x}_1 - \bar{x}$ and $d_2 = \bar{x}_2 - \bar{x}$

Also we know that $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$.

$$\therefore d_1 = \bar{x}_1 - \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{n_2(\bar{x}_1 - \bar{x}_2)}{n_1 + n_2}$$

Similarly $d_2 = \frac{n_1(\bar{x}_2 - \bar{x}_1)}{n_1 + n_2}$. From (1) we get,

$$\sigma^2 = \frac{1}{n_1 + n_2}\left[n_1\sigma_1^2 + n_2\sigma_2^2 + \frac{n_1 n_2^2(\bar{x}_1 - \bar{x}_2)^2}{(n_1 + n_2)^2} + \frac{n_2 n_1^2(\bar{x}_2 - \bar{x}_1)^2}{(n_1 + n_2)^2}\right]$$

$$\therefore \sigma^2 = \frac{1}{n_1 + n_2}\left[n_1\sigma_1^2 + n_2\sigma_2^2 + \frac{n_1 n_2}{n_1 + n_2}(\bar{x}_1 - \bar{x}_2)^2\right]$$

Solved Problems.

Problem 1. Find (i) mean (ii) range (iii) S.D (iv) mean deviation about mean and (v) coefficient of variation for the following marks of 10 students.

20, 22, 27, 30, 40, 48, 45, 32, 31, 35.

Solution. (i) Mean $= (1/n)\,\Sigma x_i = \frac{330}{10} = 33.$

(ii) Range = Maximum value − Minimum value

$= 48 - 20 = 28.$

Here we have $\Sigma x_i^2 = 11652$ (verify).

$$\therefore \sigma = \left[\frac{11652}{10} - \left(\frac{330}{10}\right)^2\right]^{1/2} = (76.2)^{1/2} = 8.73$$

(iv) Mean deviation about mean $= \frac{1}{10}\left[\Sigma |x_i - 33|\right]$

$$= \frac{1}{10}[13 + 11 + 6 + 3 + 7 + 15 + 12 + 1 + 2 + 2]$$

$$= 7.2.$$

(v) $C.V = \left(\frac{\sigma}{\bar{x}}\right) \times 100 = \left(\frac{8.73}{33}\right) \times 100 = 26.45.$

Problem 2. Show that the Variance of the first n natural numbe is $\frac{1}{12}(n^2 - 1)$.

Solution. $\sigma^2 = \frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2$.

We have $\Sigma x_i = 1 + 2 + \ldots\ldots + n = \frac{1}{2} n(n + 1)$ and

$$\Sigma x_i^2 = 1^2 + 2^2 + \ldots\ldots + n^2 = \frac{1}{6} n(n + 1)(2n + 1).$$

$$\therefore \sigma^2 = \frac{n(n+1)(2n+1)}{6n} - \left[\frac{n(n+1)}{2n}\right]^2$$

$$= \frac{1}{6}(n + 1)(2n + 1) - \frac{1}{4}(n + 1)^2$$

$$= \frac{1}{12}[2(n + 1)(2n + 1) - 3(n + 1)^2]$$

$$= \frac{1}{12}[(n + 1)(4n + 2 - 3n - 3)] = \frac{1}{12}[(n + 1)(n - 1)]$$

$$= \frac{1}{12}(n^2 - 1).$$

Problem 3. The following table gives the monthly wages of worker in a factory. Compute (i) standard deviation (ii) quartile deviation and (iii) coefficient of variation.

Monthly wages	No. of workers	Monthly wages	No. of workers
125 - 175	2	375 - 425	4
175 - 225	22	425 - 475	6
225 - 275	19	475 - 525	1
275 - 325	14	525 - 575	1
325 - 375	3	Total	72

Solution. Let $A = 300$; $h = 50$ and $u_i = \frac{1}{50}(x_i - 300)$. The table is

Mid x_i	f_i	u_i	$f_i u_i$	$f_i u_i^2$	c.f.
150	2	− 3	− 6	18	2
200	22	− 2	− 44	88	24
250	19	− 1	− 19	19	43
300	14	0	0	0	57
350	3	1	3	3	60
400	4	2	8	16	64
450	6	3	18	54	70
500	1	4	4	16	71
550	1	5	5	25	72
Total	72	-	− 31	239	-

(i) $\bar{x} = A + h\bar{u}$

$= 300 + 50\left(-\frac{31}{72}\right) = 300 - \frac{1550}{72}$

$= 300 - 21.53 = 278.47.$

(ii) $Q_1 = 175 + \frac{(18-2)\times 50}{22} = 175 + \frac{800}{22} = 211.36$

$Q_3 = 275 + \frac{(54-43)\times 50}{22} = 275 + \frac{550}{14} = 314.29$

$\therefore Q.D = \frac{1}{2}(Q_3 - Q_1)$

$= \frac{1}{2}(314.29 - 211.36)$

$= 51.45.$

(iii) $\sigma^2 = h^2\left[\frac{\Sigma f_i u_i^2}{N} - \left(\frac{\Sigma f_i u_i}{N}\right)^2\right]$

$= 50^2\left[\frac{239}{72} - \left(-\frac{31}{72}\right)^2\right]$

$\therefore \sigma = 88.52$ (verify)

(iv) $C.V = \frac{88.52}{278.47} \times 100$

$= 31.79.$

Problem 4 Find the arithmetic mean $\bar{x}$, standard deviation σ and percentage of cases within $\bar{x} \pm \sigma$, $\bar{x} \pm 2\sigma$ and $\bar{x} \pm 3\sigma$ in the following frequency distribution.

Marks	10	9	8	7	6	5	4	3	2	1
Frequency	1	5	11	15	12	7	3	3	0	1

Solution.

x_i	f_i	$f_i x_i$	$f_i x_i^2$
10	1	10	100
9	5	45	405
8	11	88	704
7	15	105	735
6	12	72	432
5	7	35	175
4	3	12	48
3	2	6	18
2	0	0	0
1	1	1	1
Total	**57**	**374**	**2618**

$$\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{374}{57} = 6.56$$

$$\sigma^2 = \frac{(\Sigma f_i x_i^2)}{N} - \left(\frac{\Sigma f_i x_i}{N}\right)^2$$

$$= \frac{2618}{57} - \left(\frac{374}{57}\right)^2 = \frac{2618\times 57 - 374^2}{57^2} = \frac{9350}{57^2}$$

$$\therefore \sigma = \left(\frac{1}{57}\right)\sqrt{9350} = 1.7 \text{ (approximately).}$$

Now, $\bar{x} \pm \sigma = 6.56 \pm 1.7 = 8.26,\ 4.86.$

There are 45 items [7 + 12 + 15 + 11] which lie within 4.86 and 8.26.

$\therefore$ Percentage of cases lying within the

$\bar{x} \pm \sigma = \frac{45}{57} \times 100 = 79\%$.

Now $\bar{x} \pm 2\sigma = 6.56 \pm 3.4 = 9.96$, 3.16.

There are only 53 items $[3 + 7 + 12 + 15 + 11 + 5]$ ly

within 3.16 and 9.96.

$\therefore$ Percentage of items lying within the range $\bar{x} \pm 2\sigma$

$$= \frac{53}{57} \times 100 = 93\%$$

Similarly the percentage of items lying within the r

$\bar{x} \pm 3\sigma$ is 98% (verify).

Problem 5. Mean and standard deviation of the marks of two classes of

25 and 75 are given below.

	Class A	Class B
Mean	80	85
S.D.	15	20

Calculate the combined mean and standard deviation of the marks of t

students of the two classes. Which class is performing a consistent progre

Solution. Let $\bar{x}$ and σ be the mean and standard deviation of the combin

classes.

Given $\bar{x}_1 = 80$; $\bar{x}_2 = 85$; $\sigma_1 = 15$; $\sigma_2 = 20$; $n_1 = 25$; $n_2 = 75$.

$$\therefore \bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} = \frac{25 \times 80 + 75 \times 85}{100} = \frac{8375}{100} = 83.75$$

Now, $d_1 = \bar{x}_1 - \bar{x} = 80 - 83.75 = -3.75$.

$d_2 = \bar{x}_2 - \bar{x} = 85 - 83.75 = 1.25$

We have $\sigma^2 = \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}$

$$\therefore \sigma^2 = \frac{1}{100} [25 \times 15^2 + 75 \times 20^2 + 25 (-3.75)^2 + 75 (1.25)^2]$$

$$= \frac{1}{100} [5625 + 30000 + 351.5625 + 117.1875]$$

$$= 360.9375.$$

$\therefore \sigma = 19$ (approximately).

C.V. of marks of class $A = \frac{\sigma_1}{\bar{x}_1} \times 100 = \frac{15}{80} \times 100$

$= 18.75$

C.V. of marks of class $B = \frac{\sigma_2}{\bar{x}_2} \times 100 = \frac{20}{85} \times 100$

$= 23.53$.

Since the C.V. of marks of class A is smaller than that of class B, class A is performing consistent progress.

Problem 6. Prove that for any discrete distribution standard deviation is not less than the mean deviation from mean.

Solution. Let m = mean deviation from mean.

$$\therefore m = (1/N) \left[\Sigma f_i |x_i - \bar{x}|\right].$$

We have to prove σ not less than m. (i.e.) to prove $\sigma^2 \geq m^2$.

Now, $\sigma^2 \geq m^2 \Leftrightarrow (1/N) \Sigma f_i (x_i - \bar{x})^2 \geq [(1/N) \Sigma f_i |x_i - \bar{x}|]^2$

$\Leftrightarrow (1/N) \Sigma f_i z_i^2 \geq [(1/N) \Sigma f_i z_i]^2$ where $z_i = |x_i - \bar{x}|$

$\Leftrightarrow (1/N) [\Sigma f_i z_i^2 - (\Sigma f_i z_i)^2] \Leftrightarrow \sigma_z^2 \geq 0$ which is true. Hence the result.

Problem 7. The scores of two cricketers A and B in 10 innings are given below. Find who is a better run getter and who is more consistent player.

A scores x_i	40	25	19	80	38	8	67	121	66	76
B Scores y_i	28	70	31	0	14	111	66	31	25	4

Solution. For cricketer A : $\bar{x} = \frac{540}{10} = 54$.

For cricketer B : $\bar{y} = \frac{380}{10} = 38$.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
40	− 14	196	28	− 10	100
25	− 29	841	70	32	1024
19	− 35	1225	31	− 7	49
80	26	676	0	− 38	1444
38	− 16	256	14	− 24	576
8	− 46	2116	111	73	5329
67	13	169	66	28	784
121	67	4489	31	− 7	49
66	12	144	25	− 13	169
76	22	484	4	− 34	1156
Total	-	10596	Total	-	10680

$\sigma_x = [(1/n)\,\Sigma (x_i - \bar{x})^2]^{1/2} = \left[\frac{10596}{10}\right]^{1/2} = \sqrt{1059.6} = 32.55.$

Similarly $\sigma_y = [(1/n)\,\Sigma (y_i - \bar{y})^2]^{1/2} = \left[\frac{10680}{10}\right]^{1/2}$

$= \sqrt{1068} = 32.68.$

$C.V \text{ of } A = \left(\frac{\sigma_y}{\bar{x}}\right) \times 100 = \frac{32.35}{54} \times 100 = 60.28.$

$C.V \text{ of } B = \left(\frac{\sigma_y}{\bar{y}}\right) \times 100 = \frac{32.68}{38} \times 100 = 86.$

Since $\bar{x} > \bar{y}$ cricketer A is better run getter. $C.V.$ of A < $C.V$ of B cricketer A is also a consistent player.

Problem 8. The mean and standard deviation of 200 items are found to be 60 and 30. If at the time of calculation two items are wrongly taken as 3 and 87 instead of 13 and 17, find the correct mean and standard deviation.

Solution. Here $n = 200$; $\bar{x} = 60$; $\sigma = 20$.

$\bar{x} = 60 \Rightarrow \frac{\Sigma x_i}{200} = 60.$

$\therefore \Sigma x_i = 12000$

Corrected $\Sigma x_i = 12000 - (3 + 67) + (13 + 17) = 11960.$

$\therefore$ Corrected $\bar{x} = \frac{11960}{200} = 59.8$

$\sigma^2 = \frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2$. Hence $20^2 = \frac{\Sigma x_i^2}{200} - 60^2$

$\therefore \Sigma x_i^2 = 200\,(20^2 + 60^2) = 800000.$

After correction $\Sigma x_i^2 = 800000 - (3^2 + 67^2) + (13^2 + 17^2) = 795960.$

$\therefore$ Corrected $\sigma^2 = \frac{795960}{200} - (59.8)^2 = 3979.8 - 3576.04 = 403.76$

$\therefore \sigma = 20.09.$

Problem 9. Find (i) the mean deviation from the mean (ii) variance of arithmetic progression $a, a+d, a+2d, \ldots\ldots, a+2nd$.

Solution. There are $2n+1$ terms in the A.P.

$$\therefore \bar{x} = \frac{1}{2n+1}\left[a + (a+d) + \ldots\ldots + (a+2nd)\right]$$

$$= \frac{1}{2n+1}\left[(2n+1)a + d(1+2+\ldots\ldots+2n)\right]$$

$$= \frac{1}{2n+1}\left[(2n+1)a + d\left\{\frac{2n(2n+1)}{2}\right\}\right]$$

$$= a + nd$$

(i) Mean deviation from mean $= \dfrac{1}{2n+1}\Sigma|x_i - \bar{x}|$

$$= \frac{1}{2n+1}\left[2d(1+2+\ldots\ldots+n)\right]$$

$$= \frac{n(n+1)d}{2n+1}$$

(ii) Variance $\sigma^2 = \dfrac{1}{2n+1}\Sigma(x_i - \bar{x})^2$

$$= \frac{1}{2n+1}\left[2d^2(1^2 + 2^2 + \ldots\ldots + n^2)\right]$$

$$= \frac{1}{2n+1}\,2d^2\left[\frac{n(n+1)(2n+1)}{6}\right]$$

$$= \frac{1}{3}n(n+1)d^2.$$

Exercises

1. Calculate mean, S.D and C.V of the marks obtained by 20 students in an examination.

82	85	75	81	74	58	66	72	54	84
65	59	82	62	85	52	80	86	71	75

2. Calculate the standard deviation from the following data of income of 10 employees of a firm.

100 120 140 120 180 175 185 136 200 150

3. Prepare a frequency table from the following passage taking consonants and vowels in each word as two variables x and y. Find $\bar{x}$, $\bar{y}$, σ_x and σ_y.

"Childhood shows the man as morning shows the day. All that lies between the cradle and the grave is uncertain".

4. Calculate mean deviation from (i) mean (ii) median (iii) mode for the following data.

size of item	frequency	size of item	frequency
3 - 4	3	7 - 8	85
4 - 5	7	8 - 9	32
5 - 6	22	9 - 10	8
6 - 7	60	**Total**	217

5. Calculate the mean deviation from (i) mean (ii) mode (iii) median of marks obtained by 9 students given by 7, 4, 10, 9, 15, 12, 7, 9, 7.

6. Calculate the C.V for a series for which the following results are known $N = 50$; $\Sigma d_i = -10$; $\Sigma d_i^2 = 400$ where $d_i = x_i - 75$.

7. Given $\Sigma x_i = 99$; $n = 9$; $\Sigma(x_i - 10)^2 = 79$. Find Σx_i^2. Hence find σ^2.

8. Find the standard deviation of the following heights of 100 male students.

Height in inches	60 - 62	63 - 65	66 - 68	69 - 71	72 - 74
No. of students	5	18	42	27	8

Text book page No.	Question. No.
76	1
77	2,3,4,8
78	9,14
79	18,19
80	20,22

THANK YOU

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-2

CHAPTER4.1 &4.2

PART-10

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

- introduction
- Definition of moments about mean(μ_r)
- definition of moments about any point(μ_r)
- Common moments
- Formulae for calculating moments
- Definition of skewness
- nature of skewness
- Types of Measures of skewness and nature of skewness
- Definition of kurtosis
- nature of kurtosis

introduction

Beyond the measures of central tendency and dispersion explained earlier, there are measures that further describe the characteristics of a distribution.
Some of them are discussed here.

- MOMENTS
- SKEWNESS
- KURTOSIS

➢ DEFINITION OF MOMENTS ABOUT MEAN(μ_r)

❖ I

The rth moment about mean of a distribution, denoted by μ_r, is given by

$$\mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^r, \text{ where } r = 0, 1, 2, 3, 4, \ldots\ldots\ldots$$

Thus, rth moment about mean is the mean of the rth power of deviations of observations from their arithmetic mean. In particular,

if $r = 0$, we have $\mu_0 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^0 = 1$,

if $r = 1$, we have $\mu_1 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right) = 0$,

These moments are also known as Central Moment

if $r = 2$, we have $\mu_2 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^2 = \sigma^2$,

if $r = 3$, we have $\mu_3 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^3$ and so on.

❖ II

When the Variables x_i are changed into another variable u_i where $u_i = \frac{x_i - A}{h}$
The r^{th} *moment* μ_r of the variable x_i is given by,

$$\mu_r = h^r \left[\sum \frac{f_i (u_i - \bar{u})^r}{N} \right]$$

➢ DEFINITION OF MOMENTS ABOUT ANY POINT($\mu_r^{\prime}$)

In addition to the above, we can define *raw moments* as moments about any arbitrary mean.

Let A denote an arbitrary mean, then rth moment about A is defined as

$$\mu_r' = \frac{1}{N}\sum_{i=1}^{n} f_i (X_i - A)^r \text{ , } r = 0, 1, 2, 3, \ldots\ldots\ldots$$

When $A = 0$, we get various moments about origin.

Common moments

Moments are a set of statistical parameters to measure a distribution.

Four moments are commonly used:

- 1st moment - Mean (describes central value)
- 2nd moment - Variance (describes dispersion)
- 3rd moment - Skewness (describes asymmetry)
- 4th moment - Kurtosis (describes peakedness)

- **Formulae for calculating moments**

$$1\text{st moment} = \mu_1 = \frac{\sum f(x-\bar{x})}{n}$$

$$2\text{nd moment} = \mu_2 = \frac{\sum f(x-\bar{x})^2}{n}$$

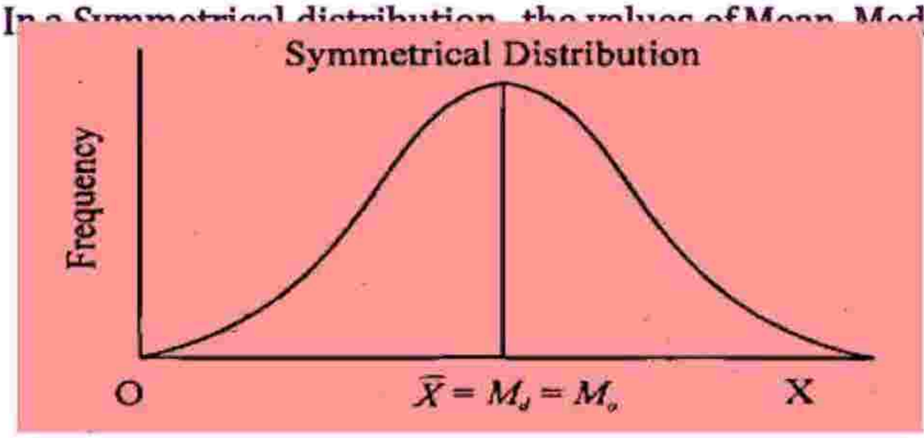
$$3\text{rd moment} = \mu_3 = \frac{\sum f(x-\bar{x})^3}{n}$$

$$4\text{th moment} = \mu_4 = \frac{\sum f(x-\bar{x})^4}{n}$$

Definition of skewness(β1)

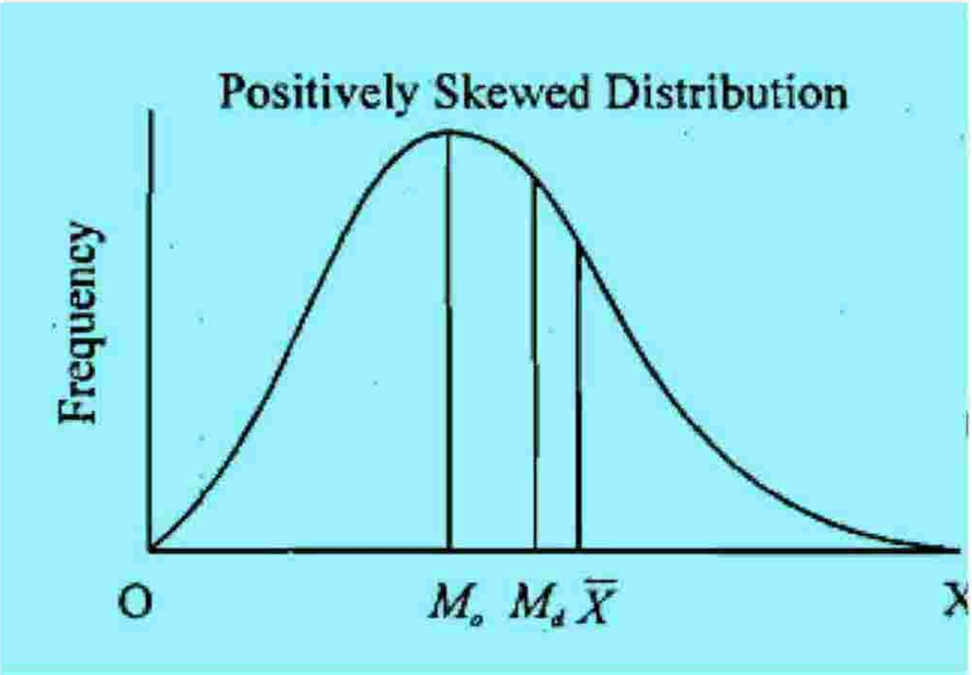
The term Skewness refers to the lack of symmetry or departure from symmetry. e.g., When a distribution not symmetrical(or asymmetrical) it is called a skewed distribution.

- In a Symmetrical distribution, the values of Mean, Median, Mode are alike.

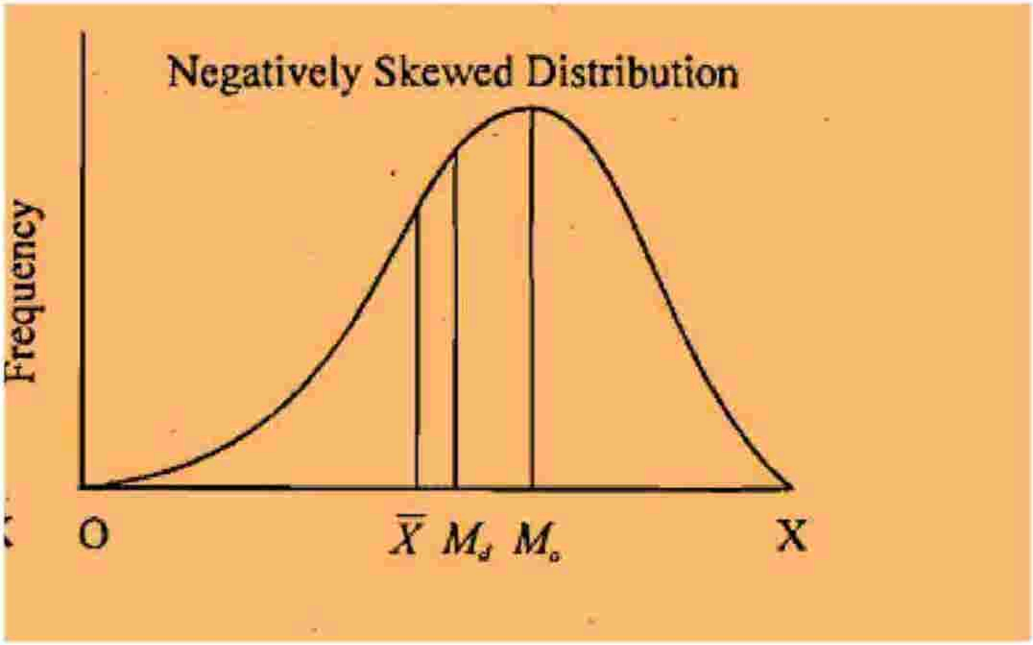


nature of skewness

- If the value of Mean>Median>Mode , Skewness is said to be Positive.



- If the value of Mean<Median<Mode , Skewness is said to be Negative.



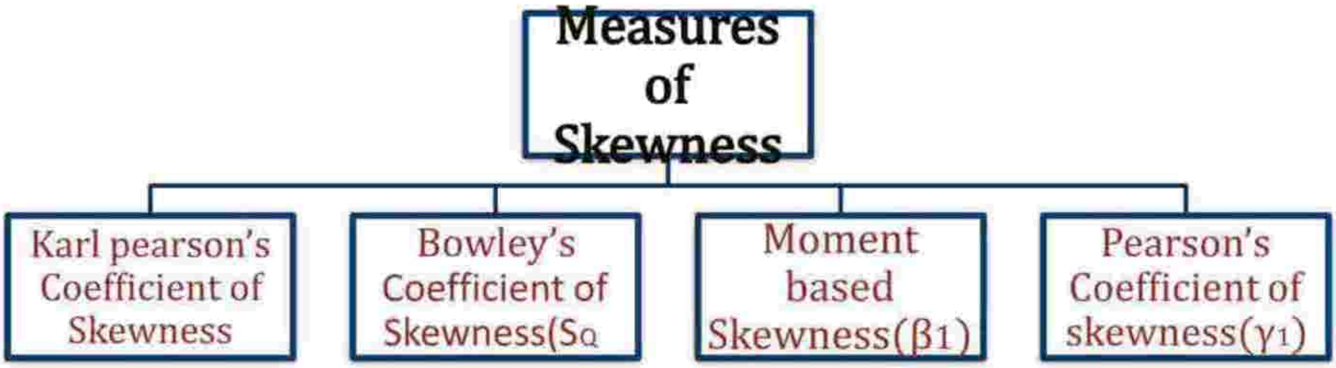
Generally,

If Mean > Mode, the skewness is positive.

If Mean < Mode, the skewness is negative.

If Mean = Mode, the skewness is zero.

Types of Measures of skewness and nature of skew of skewness



- **Karl pearson's Coefficient of Skewness**

Karl Pearson's Coefficient of Skewness = $\frac{Mean - Mode}{Standard\ Deviation}$ or $\frac{3(Mean - Median)}{Standard\ Deviation}$

- If Karl pearson's coefficient >0 then the distribution is Positively Skewed
- If Karl pearson's Coefficient<0 then the distribution is Negatively Skewed

- **Bowley's Coefficient of Skewness(S_Q)**

This measure is based on quartiles. For a symmetrical distribution, it is seen that Q_1 and Q_3 are equidistant from median. Thus $(Q_3 - M_d) - (M_d - Q_1)$ can be taken as an absolute measure of skewness.

A relative measure of skewness, known as Bowley's coefficient (S_Q), is given by

$$S_Q = \frac{Q_3 - 2M_d + Q_1}{Q_3 - Q_1}$$

Moment based Skewness(β_1)

$$\text{Moment based measure of skewness} = \beta_1 = \frac{\mu_3^{\,2}}{\mu_2^{\,3}}$$

- For symmetrical distribution $\beta_1=0$
- For Positive Skewness $\beta_1>0$
- For Negative Skewness $\beta_1<0$

- Pearson's Coefficient of skewness($\gamma 1$)

$$\text{Pearson's coefficient of skewness} = \gamma_1 = \sqrt{\beta_1}$$

Definition of kurtosis(β2)

Kurtosis refers to the degree of peakedness of a frequency curve. It tells how tall and sharp the central peak is, relative to a standard bell curve of a distribution.

Kurtosis is measured in the following ways:

Moment based Measure of kurtosis $= \beta_2 = \dfrac{\mu_4}{\mu_2^{\,2}}$

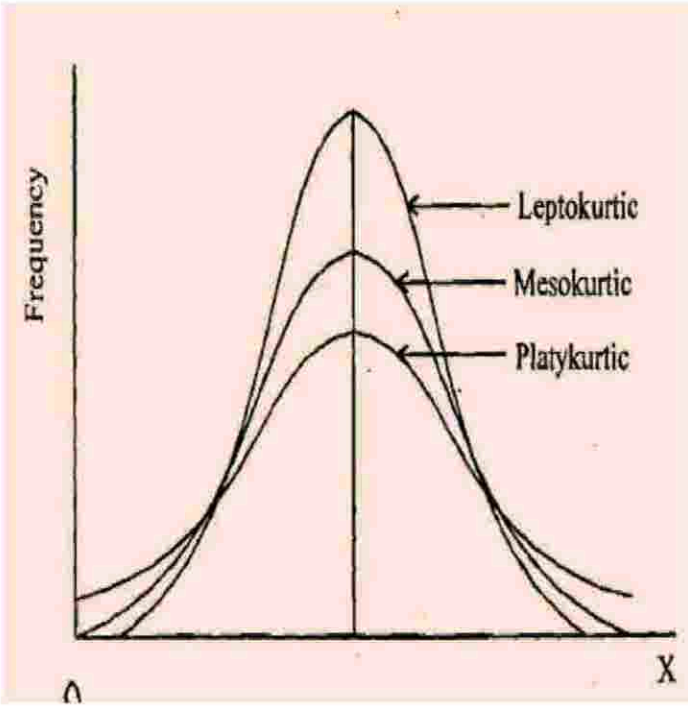
Coefficient of kurtosis $= \gamma_2 = \beta_2 - 3$

nature of kurtosis

Messokurtic:
- $\beta_2 = 3$ or $\gamma_2 = 0$
- Normal curve

Platykurtic:
- $\beta_2 < 3$ or $\gamma_2 < 0$
- Flater than Normal curve

Leptokurtic:
- $\beta_2 > 3$ or $\gamma_2 > 0$
- More Peaked than Normal curve



B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-2

CHAPTER4.1 &4.2

PART-11

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

- THEOREM NO 4.1
- THEOREM NO 4.2
- Problems
- Assignment

- **THEOREM NO 4.1**

Theorem 4.1

$$\mu_r = \mu'_r - {}^rc_1\, \mu'_{r-1}\, \mu'_1 + {}^rc_2\, \mu'_{r-1} (\mu'_1)^2 - \ldots\ldots\ldots + (-1)^{r-1} (r-1) (\mu'_1)^r.$$

Proof. $\mu_r = (1/N) \Sigma f_i (x_i - \bar{x})^r$

$$= (1/N) \Sigma f_i (x_i - A + A - \bar{x})^r$$

$$= (1/N) \Sigma f_i (x_i - A - d)^r \text{ where } d = \bar{x} - A$$

$$= (1/N) [\Sigma f_i (x_i - A)^r - {}^rc_1\, d\, \Sigma f_i (x_i - A)^{r-1} + {}^rc_2\, d^2\, \Sigma f_i (x_i - A)^{r-2}$$

$$- \ldots\ldots + {}^rc_{r-1} (-d)^{r-1} \Sigma f_i (x_i - A) + {}^rc_r (-d)^r \Sigma f_i]$$

$$= \mu'_r - {}^rc_1\, d\, \mu'_{r-1} + {}^rc_2\, d^2\, \mu'_{r-2} - \ldots\ldots + (-1)^{r-1}\, r\, d^{\,r-1} (\mu'_1) + (-1)^r d^r$$

$$= \mu'_r - {}^rc_1\, \mu'_{r-1}\, \mu'_1 + {}^rc_2\, \mu'_{r-2} (\mu'_1)^2 - \ldots\ldots + (-1)^{r-1} (r-1) (\mu'_1)^r.$$

Note. Putting $r = 2, 3, 4$ in the above theorem we have

(i) $\mu_2 = \mu'_2 - (\mu'_1)^2$

(ii) $\mu_3 = \mu'_3 - 3\mu'_2 \mu'_1 + 2(\mu'_1)^3$

(iii) $\mu_4 = \mu'_4 - 4\mu'_3 \mu'_1 + 6\mu'_2 (\mu'_1)^2 - 3(\mu'_1)^4$

Theorem 4.2. $\mu'_r = \mu_r + {}^rc_1\, \mu_{r-1}\, \mu'_1 + {}^rc_2\, \mu_{r-2} (\mu'_1)^2 + \ldots + (\mu'_1)^r.$

Proof. $\mu'_r = (1/N) \Sigma f_i (x_i - A)^r.$

$$= (1/N) \Sigma f_i (x_i - \bar{x} + \bar{x} - A)^r.$$

THEOREM NO 4.2

Theorem 4.2. $\mu_r' = \mu_r + {}^rc_1\, \mu_{r-1}\, \mu_1' + {}^rc_2\, \mu_{r-2}\, (\mu_1')^2 + \dots + (\mu_1')^r.$

Proof. $\mu_r' = (1/N)\, \Sigma f_i\, (x_i - A)^r.$

$$= (1/N)\, \Sigma f_i\, (x_i - \bar{x} + \bar{x} - A)^r.$$

$$= (1/N)\, \Sigma f_i\, (x_i - \bar{x} + d)^r \text{ where } d = \bar{x} - A = \mu_1'$$

$$= (1/N)\, \Sigma f_i\, [(x_i - \bar{x})^r + {}^rc_1\, (x_i - \bar{x})^{r-1} d + {}^rc_2\, (x_i - \bar{x})^{r-2} d^2 + \dots + d^r].$$

$$= \mu_r + {}^rc_1\, \mu_{r-1}\, \mu_1' + {}^rc_2\, \mu_{r-2}\, (\mu_1')^2 + \dots + (\mu_1')^r.$$

Note. Putting $r = 2, 3, 4$ in the above theorem and using $\mu_1 = 0$ we have

(i) $\mu_2' = \mu_2 + (\mu_1')^2$

(ii) $\mu_3' = \mu_3 + 3\,\mu_2\,\mu_1' + (\mu_1')^3$

(iii) $\mu_4' = \mu_4 + 4\,\mu_3\,\mu_1' + 6\,\mu_2\,(\mu_1')^2 + (\mu_1')^4.$

- Text book Problem no:I ,page no:86

1. Calculate the first four central moments from the following data to find β_1 and β_2 and discuss the nature of the distribution

x	0	1	2	3	4	5	6
f	5	25	17	25	19	14	5

Solution:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{300}{100} = 3$$

Let $u_i = x_i - \bar{x} = x_i - 3$

X_i	f_i	u_i	f_iu_i	$f_iu_i^2$	$f_iu_i^3$	$f_iu_i^4$
0	5	-3	-15	45	-135	405
1	15	-2	-30	60	-120	240
2	17	-1	-17	17	-17	17
3	25	0	0	0	0	0
4	19	1	19	19	19	19
5	14	2	28	56	112	224
6	5	3	15	45	135	405
Total	**100**	**-**	**0**	**242**	**-6**	**1310**

$$\mu_1 = \frac{\sum f_i(x_i - \bar{x})}{N} = 0$$

$$\mu_2 = \frac{\sum f_i(x_i - \bar{x})^2}{N} = \frac{242}{100} = 2.42$$

$$\mu_3 = \frac{\sum f_i(x_i - \bar{x})^3}{N} = \frac{-6}{100} = -0.06$$

$$\mu_4 = \frac{\sum f_i(x_i - \bar{x})^4}{N} = \frac{1310}{100} = 13.10$$

$$\beta_1 = \frac{{\mu_3}^2}{{\mu_2}^3} = \frac{(-0.006)^2}{2.42^3} = \frac{0.0036}{14.1725} = 0.0003$$

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2} = \frac{13.10}{2.42^2} = \frac{13.10}{5.8564} = 2.237$$

- ✓ *Since $\beta_1 > 0$ the distribution is positively skewed*
- ✓ *Since $\beta_2 = 2.237 < 3$ the distribution is Platykurtic*

- Text book exercise Problem no:7 ,page no:93

2. Calculate the first 4 moments of the following distribution about x=4 and hence find the moments about the mean of the distribution. Also find the values of β_1 and β_2

x	0	1	2	3	4	5	6	7	8	9	10
f	5	10	30	70	140	200	140	70	30	10	5

Solution:

Taking $u_i = x_i - 5$ we get the following table

Xi	fi	ui	fiui	fiui^2	fiui^3	fiui^4
0	5	-5	-25	125	-625	3125
1	10	-4	-40	160	-640	2560
2	30	-3	-90	270	-810	2430
3	70	-2	-140	280	-560	1120
4	140	-1	-140	140	-140	140
5 A	200	0	0	0	0	0
6	140	1	140	140	140	140
7	70	2	140	280	560	1120
8	30	3	90	270	810	2430
9	10	4	40	160	640	2560
10	5	5	25	125	625	3125
Tot	710	0	0	1950	0	18750

$$\mu_1 = \frac{\sum f_i(x_i - A)}{N} = 0$$

$$\mu_2 = \frac{\sum f_i(x_i - A)^2}{N} = \frac{1950}{710} = 2.75$$

$$\mu_3 = \frac{\sum f_i(x_i - A)^3}{N} = \frac{0}{710} = 0$$

$$\mu_4 = \frac{\sum f_i(x_i - A)^4}{N} = \frac{18750}{710} = 26.41$$

Now, μ_1=0

$\mu_2 = \mu_2 - (\mu_1)^2$ =2.75-0=2.75

$\mu_3 = \mu_3 - 3\mu_2\mu_1 + 2(\mu_1)^3$
=0-3(2.75)(0)+2(0)=0

$\mu_4 = \mu_4 - 4\mu_3 + 6\mu_2(\mu_1)^2 - 3(\mu_1)^4$

=26.41-4(0)+6(2.75)(0)-3(0)=26.41

$\beta_1 = \frac{{\mu_3}^2}{{\mu_2}^3} = \frac{(0)^2}{2.75^3} = 0$

$\beta_2 = \frac{\mu_4}{{\mu_2}^2} = \frac{26.41}{2.75^2} = \frac{26.41}{7.5625}$ =3.49

- ✓ *Since $\beta_1 = 0$ the distribution is symmetrical*
- ✓ *Since $\beta_2 = 3.49 > 3$ the distribution is Leptokurtic*

- Text book exercise Problem no:1(I) ,page no:91

3. For the following data calculate the Karl Pearson's coefficient of skewness.

Wages in Rs.	10	11	12	13	14	15
Frequency	2	4	10	8	5	1

Solution:

$$\text{Karl Pearson's Coefficient of Skewness} = \frac{Mean - Mode}{Standard\ Deviation} \text{ or } \frac{3(Mean - Median)}{Standard\ Deviation}$$

Meadian = In the Less than C.F just greater than $\frac{N}{2}$

$$\sigma = \left[\frac{\sum f_i x_i^2}{N} - \left(\frac{\sum f_i x_i}{N}\right)^2\right]^{\frac{1}{2}}$$

$$Mean = \frac{\sum f_i x_i}{N}$$

X_i	f_i	fixi	$fixi^2$	Less than C.F
10	2	20	400	2
11	4	44	484	6
12	10	120	1440	16
13	8	104	1352	24
14	5	70	980	29
15	1	15	225	30
Tot	**30**	**373**	**4881**	

$$\boldsymbol{Mean = \frac{\sum f_i x_i}{N} = \frac{373}{30} = 12.43}$$

Meadian= In the Less than C.F just greater than $\frac{N}{2}$

Median=in the lessthan C.F just greater than $\frac{30}{2}$=15

Median=12

$$\sigma = \left[\frac{\sum f_i {x_i}^2}{N} - \left(\frac{\sum f_i x_i}{N}\right)^2\right]^{\frac{1}{2}}$$

$$\sigma = \left[\frac{4881}{30} - \left(\frac{373}{30}\right)^2\right]^{\frac{1}{2}}$$

$$\sigma = \left[87.1 - (12.43)^2\right]^{\frac{1}{2}}$$

$$\sigma = [162.7 - 154.5049]^{\frac{1}{2}}$$

$$\sigma = [8.1951]^{\frac{1}{2}}$$

σ=2.863

Karl Pearson's Coefficient of Skewness

$$= \frac{3(Mean - Median)}{Standard\ Deviation}$$

$$= \frac{3(12.43 - 12)}{2.863}$$

Karl Pearson's coefficient=0.451>0

- **Text book example Problem no:3 ,page no:89**

3.The first four moments of a distribution about x=2 are 1,2.5,5.5 and 16. Calculate the four moments (i)About the mean (ii)about zero

Solution:

Here Given $\mu_1 = 1, \mu_2 = 2.5, \mu_3 = 5.5\ AND \mu_4 = 16$

(i)Moments about mean:

Now, μ_1=0

$\mu_2 = \mu_2 - (\mu_1)^2$ =2.5-1=1.5

$\mu_3 = \mu_3 - 3\mu_2\mu_1 + 2(\mu_1)^3$

= 5.5-3(2.5)(1)+2 $(1)^3$=0

$\mu_4 = \mu_4 - 4\mu_3 + 6\mu_2(\mu_1)^2 - 3(\mu_1)^4$

=16-4(5.5)+6(2.5) $(1)^2$-3 $(1)^4$=16-22+15-3=6

(ii)Moments about Zero

Note: $\bar{x} = A + \mu_1$

$\bar{x} = 2 + 1 = 3$

B. Sc., III YEAR- V SEMESTER
STATISTICS-I
COURSE CODE: 7BMA5C2

UNIT-2
CHAPTER5.1
PART-12

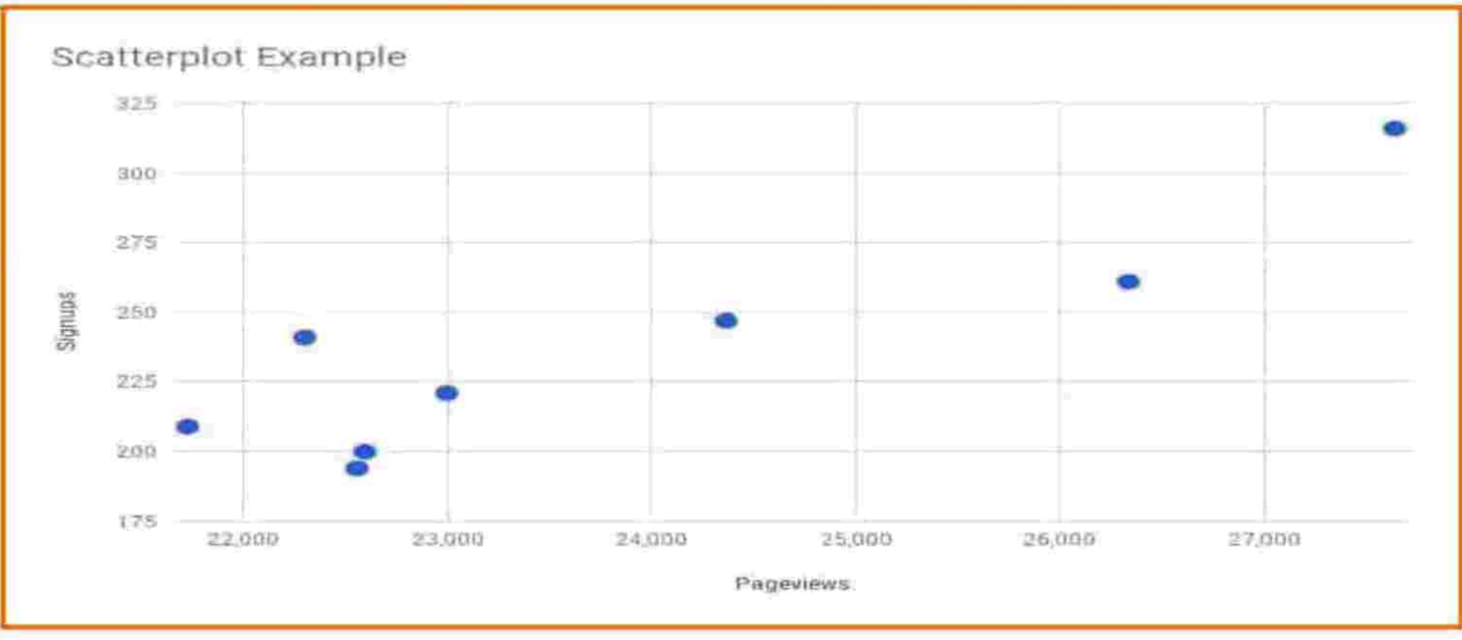
R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

- INTRODUCTION
- Definition of curve fitting
- Definition of scatter diagram
- Principle of least squares method
- Fitting a straight line
- Fitting a second degree parabola
- problems
- Assignment

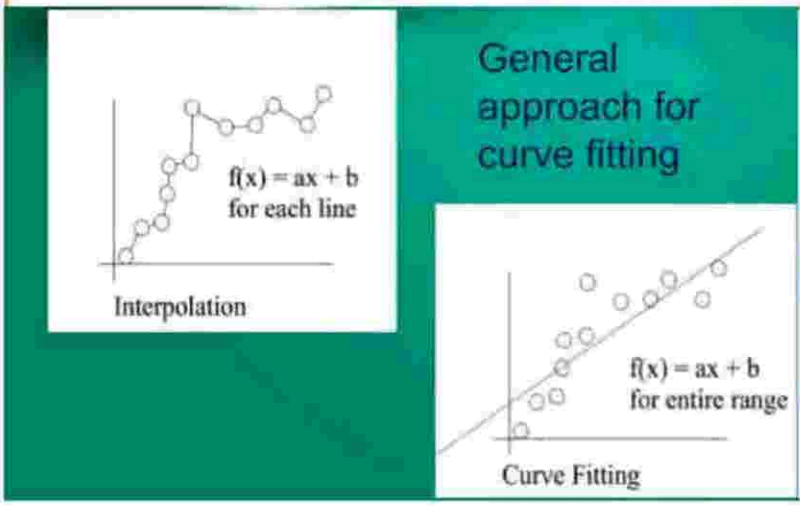
DEFINITION OF SCATTER DIAGRAM

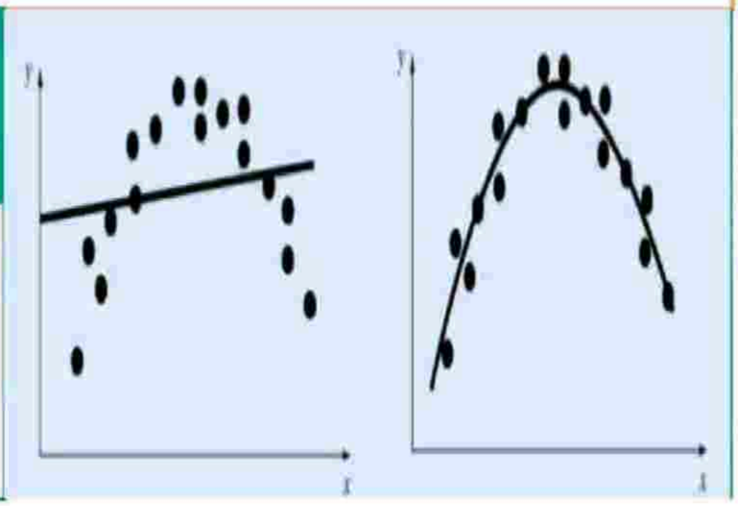
Let x_i where i=1,2,3,…n be the values of the independent variable x_i and y_i where i=1,2,3….n be the corresponding values of the dependent variables y_i. If the points (x_i, y_i) ; i=1,2,3..n are plotted on a graph paper and we obtain a Scatter diagram.



DEFINITION OF CURVE FITTING

The process of finding such a functional relationship between the variables is called curve fitting. A curve can be fitted to indicate a functional relationship between two variables x_i and y_i .





PRINCIPLE OF LEAST SQUARES

- The least squares method is a statistical procedure to find the best fit for a set of data points by minimizing the sum of the offsets or residuals of points from the plotted curve. Least squares regression is used to predict the behaviour of dependent variables.

- Let (x_i, y_i) where i=1,2….n be the observed set of values of the variables (x, y).Let y=f(x) be a functional relationship sought between the variables x and y. Then d_i=y_i-f(x_i) which is the difference between the observed value of y and the value of y determined by the functional relation is called the residuals. The principle of least squares states that the parameters involved in f(x) should be chosen in such a way that $\sum d_i^2$ is minimum.

Fitting a straight line

Fitting a straight line

Consider the fitting of the straight line $y = ax + b$ to the data (x_i, y_i), $i = 1, 2, \ldots, n$.

The residual d_i is given by $d_i = y_i - (ax_i + b)$

$\therefore \Sigma d_i^2 = \Sigma (y_i - ax_i - b)^2 = R$ (say). According to the principle of least sqaures we have to determine the parameters a and b so that R is minimum.

$$\frac{\partial R}{\partial a} = 0 \Rightarrow -2\Sigma (y_i - ax_i - b)x_i = 0$$

$$\Rightarrow \Sigma (x_i y_i - ax_i^2 - bx_i) = 0.$$

$$\therefore a\Sigma x_i^2 + b\Sigma x_i = \Sigma x_i y_i. \qquad \ldots\ldots (1)$$

$$\frac{\partial R}{\partial b} = 0 \Rightarrow -2\Sigma (y_i - ax_i - b) = 0$$

$$\therefore a\Sigma x_i + nb = \Sigma y_i. \qquad \ldots\ldots (2)$$

Equations (1) and (2) are called **normal equations** from which a and b can be found.

Fitting a second degree parabola

Fitting a second degree parabola.

Consider the fitting of the parabola $y = ax^2 + bx + c$ to the data (x_i, y_i) where $i = 1, 2, \ldots, n$.

The residual d_i is given by $d_i = y_i - (ax_i^2 + bx_i + c)$.

$\therefore \Sigma d_i^2 = \Sigma (y_i - ax_i^2 - bx_i - c)^2 = R$ (say)

By the principle of least sqaures we have to determine the parameters a, b and c so that R is minimum.

$$\frac{\partial R}{\partial a} = 0 \Rightarrow -2\Sigma(y_i - ax_i^2 - bx_i - c)x_i^2 = 0.$$

$$\Rightarrow \Sigma x_i^2 y_i - a\Sigma x_i^4 - b\Sigma x_i^3 - c\Sigma x_i^2 = 0.$$

$$\therefore a\Sigma x_i^4 + b\Sigma x_i^3 + c\Sigma x_i^2 = \Sigma x_i^2 y_i. \qquad \ldots\ldots (1)$$

$$\frac{\partial R}{\partial b} = 0 \Rightarrow -2\Sigma(y_i - ax_i^2 - bx_i - c)x_i = 0.$$

$$\Rightarrow \Sigma x_i y_i - a\Sigma x_i^3 - b\Sigma x_i^2 - c\Sigma x_i = 0.$$

$$\therefore a\Sigma x_i^3 + b\Sigma x_i^2 + c\Sigma x_i = \Sigma x_i y_i. \qquad \ldots\ldots (2)$$

$$\frac{\partial R}{\partial c} = 0 \Rightarrow -2\Sigma(y_i - ax_i^2 - bx_i - c) = 0.$$

$$\Rightarrow \Sigma y_i - a\Sigma x_i^2 - b\Sigma x_i - nc = 0.$$

$$\therefore a\Sigma x_i^2 + b\Sigma x_i + nc = \Sigma y_i. \qquad \ldots\ldots (3)$$

Equations (1), (2), and (3) are called **normal equations** from which a, b and c can be found.

Note. If the given data is not in linear form it can be brought to linear form by some suitable transformations of variables. Then using the principle of least squares the curve of best fit can be achieved.

- Text book page no:98,problem no:1

1. Fit a Straight line to the following data.

X	0	1	2	3	4
Y	2.1	3.5	5.4	7.3	8.2

Solution:

Let the Straight line to be fitted to the data be y=ax+b.
Then the parameters a and b are got from the normal equations.

$$\sum y_i = a\sum x_i + nb$$

$$\sum x_i y_i = a\sum x_i^2 + b\sum x_i$$

X_i	Y_i	X_iY_i	X_i^2
0	2.1	0	0
1	3.5	3.5	1
2	5.4	10.8	4
3	7.3	21.9	9
4	8.2	32.8	16
Total	26.5	69	30

Hence the normal equations are,

$10a+5b=26.5 \longrightarrow (1)$

$30a+10b=69 \longrightarrow (2)$

Solving (1) and (2),

$3x(1) \Rightarrow 30a+15b=79.5$

$(2) \Rightarrow 30a+10b=69$

(-) (-) (-)

$5b=10.5$

$b=2.1$

$10a+5(2.1)=26.5$

$10a+10.5=26.5$

$10a=16$

$a=1.6$

By solving (1) and (2) we get $a=1.6$ & $b=2.1$

The straight line fitted for the data is **$Y=1.6x+2.1$**

 Text book page no:100,problem no:3

3. Fit a Second degree parabola by taking x_i as the independent variable.

X	0	1	2	3	4
y	1	5	10	22	38

Solution:

Let the second degree parabola to be fitted to the data be $y = ax^2 + bx + c$. Then we have the normal equations to find a, b, c.

$$a\sum x_i^4 + b\sum x_i^3 + c\sum x_i^2 = \sum x_i^2 y_i$$

$$a\sum x_i^3 + b\sum x_i^2 + c\sum x_i = \sum x_i y_i$$

$$a\sum x_i^2 + b\sum x_i + nc = \sum y_i$$

X_i	Y_i	X_iY_i	X_i^2	$X_i^2Y_i$	X_i^3	X_i^4
0	1	0	0	0	0	0
1	5	5	1	5	1	1
2	10	20	4	40	8	16
3	22	66	9	198	27	81
4	38	152	16	608	64	256
Total	**76**	**243**	**30**	**851**	**100**	**354**

Hence the normal equations are,

$354a+100b+30c=851$? (1)

$100a+30b+10c=243$? (2)

$30a+10b+5c=76$? (3)

Solving (1) ,(2)and (3),

(1)=> $354a+100b+30c=851$

3x(2)=> $300a+90b+30c=729$

(-) (-) (-)

$54a+10b \quad =122$ □(4)

(2)=> $100a+30b+10c=243$

2x(3)=> $60a+20b+10c=152$

(-) (-) (-)

$40a+10b \quad =91$ □(5)

By solving (4) and (5)

(4)=> $54a+10b \quad =122$

(5)=> $40a+10b \quad =91$

(-) (-) (-)

$14a=31$ => **$a=2.21$**

Substitute a=2.21 in Equation (5)

40(2.21)+10b =91

88.4+10b=91

10b=91-88.4=2.6

b=0.26

30a+10b+5c=76

30(2.21)+10(0.26)+5c=76

66.3+2.6+5c=76

68.9+5c=76

5c=7.1

c=1.42

we get a=2.21,b=0.26 & c=1.42

The straight line fitted for the data is

$$y = 2.21x^2 + 0.26x + 1.42$$

- Text book page no:IOI,problem no:4

2. Fit the curve $y = bx^a$ to the following data

X	1	2	3	4	5	6
y	1200	900	600	200	110	50

Solution:

$$y = ax^b$$
$$\log y = Y \text{ and } \log x = X$$

Then the curve is transformed into Y=AX+B where A=a and B=log b.
Hence the normal equations now become

$$\sum Y = A \sum X + nB$$

$$\sum XY = A \sum X^2 + B \sum X.$$

x	y	X	Y	XY	X^2
1	1200	0	3.0792	0	0
2	900	0.3010	2.9542	0.889	0.091
3	600	0.4771	2.7782	1.325	0.228
4	200	0.6021	2.3010	1.385	0.363
5	110	0.6990	2.0414	1.427	0.489
6	50	0.7782	1.6990	1.322	0.606
Total	**-**	**2.8574**	**14.8530**	**6.348**	**1.777**

The normal equations are

2.9A+6B=14.9→(1)(approximately)

1.8A+2.9B=6.3→(2)(approximately)

(1)X1.8 =>5.22A+10.8B=26.82

(2)X2.9 =>5.22A+8.41B=18.27

(-) (-) (-)

2.39B=8.55→B=log b=3.6

Substitute B=3.6 in equation (2)

→1.8A+2.9(3.6)=6.3

1.8A+10.44=6.3

1.8A=-4.14→A= a=-2.3

A= a=-2.3

B=log b=3.6

b=antilog(3.6)=3981

Therefore the required equation to the curve is $\boldsymbol{y = 3981x^{-2.3}}$

- **Text book page no:102,problem no:5**

5. Explain the method of fitting the curve of good fit $y = ae^{bx} (a > 0)$

Solution:

$y = ae^{bx}$ →(1)

Therefore , $\log y = \log a + bx \log e$ →(2)

Let Y=log y; B=log a; A=log e

Therefore (2) become Y=Ax + B

This is linear equation in x and Y whose normal equation are

$$\sum Y_i = A \sum x_i + nB$$

$$\sum x_i Y_i = A \sum x_i^2 + B \sum x_i.$$

From the two normal equations we can get the values of A and B and consequently a and b can be obtained from a=antilog (B) and $b = \frac{A}{\log e}$. Thus the curve of best fit (1) can be obtained.

- **Text book page no:102,problem no:6**

6. Explain the method of fitting the curve $y = ka^{bx}$ (a, k > 0) obtaining the normal equations by the method of least squares .

Solution:

The curve can be transferred to the from of a straight line as follows .

$$\log y = \log k + b(log a)x; \ (a, k > 0)$$

Let $\log y$=Y; log k=B; blog a= A

Hence the above equation takes form Y=AX=B

By the principle of least squares the normal equations to find A and B of the above straight line are

$$\sum Y_i = A \sum x_i + nB$$

$$\sum x_i Y_i = A \sum x_i^2 + B \sum x_i.$$

After finding the values of A and B from the normal equations we can obtain the value of k, a and b and hence the curve $y = ka^{bx}$ can be fitted.

- Text book page no:IO3,problem no:7

7. Fit a curve of the form $y = ka^{bx}$ to the following data

Year(x)	1951	1952	1953	1954	1955	1956	1957
Production in tons	201	263	314	395	427	504	612

Solution:

$y = ka^{bx}$ →(1)

Therefore , $\log y = \log k + b(\log a)x$; $(a, k > 0)$ →(2)

log y=Y; log k=B; blog a= A

Therefore (2) becomes Y=AX=B →(3)

Where **X=x-1954**

x	y	X=x-1954	Y=logy	XY	X^2
1951	201	-3	2.3032	-6.9096	9
1952	263	-2	2.4200	-4.8400	4
1953	314	-1	2.4969	-2.4969	1
1954	395	0	2.5966	0	0
1955	427	1	2.6304	2.6304	1
1956	504	2	2.7024	5.4048	4
1957	612	3	2.7868	8.3604	9
Total		**0**	**17.9363**	**2.1491**	**28**

The normal equations for (3) are

$$\sum Y = A\sum X + nB$$

$$\sum XY = A\sum X^2 + B\sum X.$$

28A=2.1491

7B=17.9363

Solving the above equations we get A=0.0768 B=2.5623

b=antilog A=antilog0.0768=1.19(approximately)

a= antilog B=antilog2.5623=365.01(approximately)

There fore ,The curve of good fit is $y=(365.01)(1.19)^{x}$

$=(365.01)(1.19)^{x-1954}$

Assignment

Text book page No.	Question. No.
104	1,2
105	3,4,5,6

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-3

CHAPTER6.1

PART-13

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

DEFINITION OF BIVARIATE DATA

Bivariate means "two variables", in other words there are two types of data. Such a data$(x_i, y_i); i = 1,2, \dots \dots n$ is called a Bivariate data .

For Example:

Ice Cream Sales and **Temperature**.
Height and **Weight** of Collection of students.
Price **of commodity** and **Corresponding demand.**

So with bivariate data we are interested in **comparing** the two sets of data and finding any **relationships**.

There are two **main problems** involved in the relationship between x & y.

- The first is to find a measure of the **degree of association or correlation.**

- The second problem is to find the most suitable form of equation for determining the probable value of one variable corresponding to a given value of the other or **Regression**

DEFINITION OF CORRELATION

The word Correlation is made of **Co-** (meaning "together"), and **Relation**

Direct or Positive correlation:

Correlation is **Positive** when the values **increase** or decrease together,

ie.,The two variables deviate in the same direction.

For Example:

Height and **Weight** of Collection of students.

Income and **Expenditure** of a family

Inverse or Negative correlation:

Correlation is **Negative** when one value **decreases** as the other **increases**.

ie.,The two variables deviate in the opposite direction.

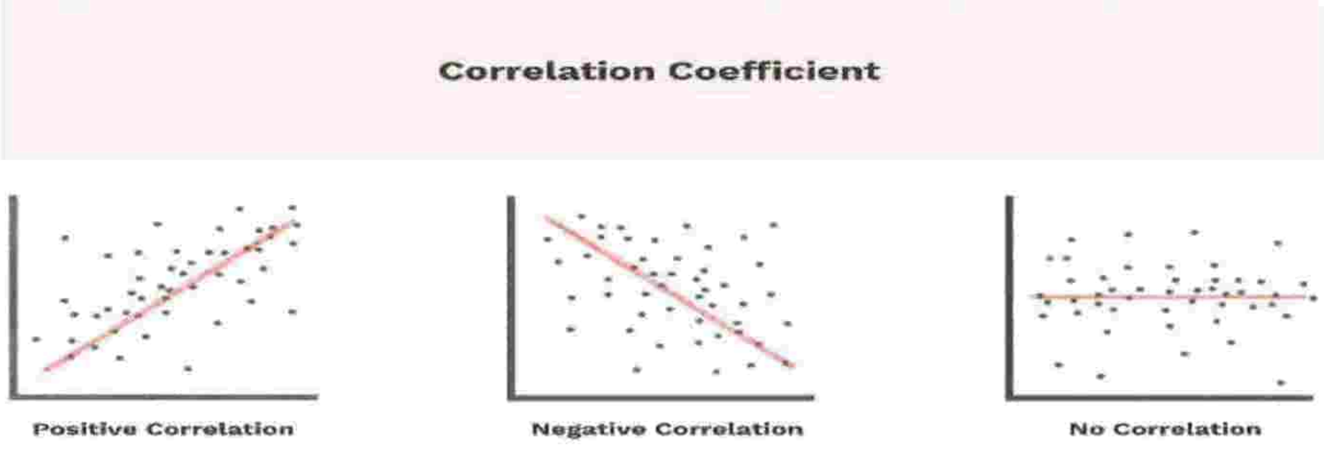
For Example:

Price and **Demand**

volume and **pressure**

Perfect correlation:

- A relationship between two variables, *x* and *y*, in which the change in value of one variable is exactly proportional to the change in value of the other.
- A perfect correlation forms a perfectly straight line.



Karl Pearson's coefficient of correlation:

Karl Pearson's coefficient of correlation between the variables x and y is defined by

$\gamma_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y}$ where $\bar{x}, \bar{y}$ are the Arithmetic Means and σ_x, σ_y

The standard deviations of the variables x and y respectively.

Covariance of x and y:

The covariance between x and y is defined by

$$cov(x, y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Hence $\gamma_{xy} = \frac{cov(x,y)}{\sigma_x\sigma_y}$

NOTE 1: If γ =1 the correlation is perfect and positive

NOTE 2: If γ = -1 the correlation is perfect and negative

NOTE 3: If γ =0 the variables are uncorrelated

NOTE: If the variables x and y are uncorrelated then Cov(x, y)=0

Theorem 6.1 $\gamma_{xy} = \dfrac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{\left[n\Sigma x_i^2 - (\Sigma x_i)^2\right]^{1/2}\left[n\Sigma y_i^2 - (\Sigma y_i)^2\right]^{1/2}}$

Proof. $\gamma_{xy} = \dfrac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n\,\sigma_x \sigma_y}$ .... (1)

Now, $\Sigma (x_i - \bar{x})(y_i - \bar{y}) = \Sigma x_i y_i - \bar{x}\Sigma y_i - \bar{y}\Sigma x_i + n\bar{x}\bar{y}$

$= \Sigma x_i y_i - \bar{x}(n\bar{y}) - \bar{y}(n\bar{x}) + n\bar{x}\bar{y}$

$= \Sigma x_i y_i - n\bar{x}\bar{y}$

$= \Sigma x_i y_i - \left(\frac{1}{n}\right)\Sigma x_i \Sigma y_i$

$= \frac{1}{n}\left[n\Sigma x_i y_i - \Sigma x_i \Sigma y_i\right]$ .... (2)

Also, $\sigma_x^2 = \frac{1}{n}\Sigma (x_i - \bar{x})^2$

$= \frac{1}{n}\left[\Sigma x_i^2 - 2\bar{x}\Sigma x_i + n(\bar{x})^2\right]$

$= \frac{1}{n}\left[\Sigma x_i^2 - 2n(\bar{x})^2 + n(\bar{x})^2\right]$

$= \frac{1}{n}\left[\Sigma x_i^2 - \left(\frac{1}{n}\right)(\Sigma x_i)^2\right]$

$= \frac{1}{n^2}\left[n\Sigma x_i^2 - (\Sigma x_i)^2\right]$

$\therefore\ \sigma_x = \frac{1}{n}\left[n\Sigma x_i^2 - (\Sigma x_i)^2\right]^{1/2}$ .... (3)

Similarly $\sigma_y = \frac{1}{n}\left[n\Sigma y_i^2 - (\Sigma y_i)^2\right]^{1/2}$ .... (4)

Substituting (2), (3) and (4) in (1) we get the required result.

the following theorem.

Theorem 6.2 The correlation coefficient is independent of the change of origin and scale.

Proof. Let $u_i = \frac{x_i - A}{h}$ and $v_i = \frac{y_i - B}{k}$ where $h, k > 0$.

$\therefore\ x_i = A + hu_i$ and $y_i = B + kv_i$.

Hence $\bar{x} = A + h\,\bar{u}$ and $\bar{y} = B + k\,\bar{v}$

$\therefore\ x_i - \bar{x} = h\,(u_i - \bar{u})$ and $y_i - \bar{y} = k\,(v_i - \bar{v})$

Also $\sigma_x = h\,\sigma_u$ and $\sigma_y = k\,\sigma_v$.

$$\therefore\ \gamma_{xy} = \frac{\Sigma\,(x_i - \bar{x})\ (y_i - \bar{y})}{n\,\sigma_x\ \sigma_y} = \frac{hk\ \Sigma\,(u_i - \bar{u})\ (v_i - \bar{v})}{n\,(h\,\sigma_u)\ (k\sigma_v)}$$

$$= \frac{\Sigma\,(u_i - \bar{u})\ (v_i - \bar{v})}{n\ \sigma_u\ \sigma_v}$$

$$= \gamma_{uv}.$$

Hence $\gamma_{xy} = \gamma_{uv}$.

Theorem 6.3 $-1 \le \gamma \le 1$.

Proof. $$\gamma_{xy} = \frac{\Sigma\,(x_i - \bar{x})\ (y_i - \bar{y})}{n\ \sigma_x\ \sigma_y}$$

$$= \frac{\left(\frac{1}{n}\right)\ \Sigma\,(x_i - \bar{x})\ (y_i - \bar{y})}{\left[\frac{1}{n}\ \Sigma\,(x_i - \bar{x})^2\right]^{1/2}\ \left[\frac{1}{n}\ \Sigma\,(y_i - \bar{y})^2\ \right]^{1/2}}$$

Let $a_i = x_i - \bar{x}$ and $b_i = y_i - \bar{y}$

$$\therefore\ \gamma_{xy}^2 = \frac{(\Sigma\ a_i\,b_i)^2}{(\Sigma\,a_i\,^2)\,(\Sigma\,b_i\,^2)}.$$

By Schwartz inequality we have $(\Sigma\, a_i b_i)^2 \le (\Sigma a_i^2)(\Sigma b_i^2)$.

Hence $\gamma_{xy}^2 \le 1$.

$\therefore |\gamma_{xy}| \le 1$.

$\therefore -1 \le \gamma \le 1$.

Note 1. If $\gamma = 1$ the correlation is perfect and positive.

Note 2. If $\gamma = -1$ the correlation is perfect and negative.

Note 3. If $\gamma = 0$ the variables are uncorrelated.

Note 4. If the variables x and y are uncorrelated then $Cor(x, y) = 0$.

The following theorem gives another formula for γ_{xy} in terms γ_x and γ_y.

Theorem 6.4 $\gamma_{xy} = \dfrac{\sigma_x^2 + \sigma_y^2 - (\sigma_{x-y})^2}{2\,\sigma_x\,\sigma_y}$

Proof $(\sigma_{x-y})^2 = \dfrac{\Sigma\,[(x_i - y_i) - (\bar{x} - \bar{y})]^2}{n}$

$$= \frac{\Sigma\,[(x_i - \bar{x}) - (y_i - \bar{y})]^2}{n}$$

$$= \frac{1}{n}\left[\Sigma\,(x_i - \bar{x})^2 - 2\Sigma\,(x_i - \bar{x})(y_i - \bar{y}) + \Sigma(y_i - \bar{y})^2\right]$$

$$= \sigma_x^2 - 2\gamma_{xy}\,\sigma_x\,\sigma_y + \sigma_y^2$$

$$\therefore \gamma_{xy} = \frac{\sigma_x^2 + \sigma_y^2 - (\sigma_{x-y})^2}{2\,\sigma_x\,\sigma_y}$$

Solved problems.

correlation coefficient between the marks of the two tests.

x	51	63	63	49	50	60	65	63	46	50
y	49	72	75	50	48	60	70	48	60	56

Solution. Choosing the origin $A = 63$ for the variable x and $B = 60$ for y and taking $u_i = x_i - A$ and $v_i = y_i - B$ we have the following table.

x_i	u_i	y_i	v_i	u_i^2	v_i^2	$u_i v_i$
51	− 12	49	− 11	144	121	132
63	0	72	12	0	144	0
63	0	75	15	0	225	0
49	− 14	50	− 10	196	100	140
50	− 13	48	− 12	169	144	156
60	− 3	60	0	9	0	0
65	2	70	10	4	100	20
63	0	48	− 12	0	144	0
46	− 17	60	0	289	0	0
50	− 13	56	− 4	169	16	52
Total	− 70	-	− 12	980	994	500

$\gamma_{xy} = \gamma_{uv}$ (by theorem **6.2**)

$$= \frac{n \Sigma u_i v_i - \Sigma u_i \Sigma v_i}{[n \Sigma u_i^2 - (\Sigma u_i)^2]^{1/2} [n \Sigma v_i^2 - (\Sigma v_i)^2]^{1/2}}$$

$$= \frac{10 \times 500 - (-70) \times (-12)}{\left[10 \times 980 - (70^2)\right]^{1/2} \left[10 \times 994 - (-12^2)\right]^{1/2}}$$

$$= \frac{4160}{70 \times 98.97} = 0.6 \qquad \text{(verify)}$$

Problem 2. If x and y are two varibales prove that the correlation coeffic[illegible] between $ax + b$ and $cy + d$ is $\gamma_{ax+b,\,cy+d} = \frac{ac}{|ac|} \gamma_{xy}$ if $a, c \neq 0$.

Proof Let $u = ax + b$ and $v = cy + d$

$\therefore \bar{u} = a\bar{x} + b$ and $\bar{v} = c\bar{y} + d$.

$$\sigma_u^{\,2} = \frac{1}{n} \Sigma (u - \bar{u})^2 = \frac{a^2}{n} \Sigma(x_i - \bar{x})^2 = a^2 \sigma_x^{\,2}$$

Similarly $\sigma_v^{\,2} = c^2 \sigma_y^{\,2}$

$$\text{Now, } \gamma_{uv} = \frac{\Sigma (u - \bar{u})(v - \bar{v})}{n\, \sigma_u \sigma_v} = \frac{\Sigma\, a(x - \bar{x})\, c(y - \bar{y})}{n\, |ac|\, \sigma_x \sigma_y}$$

$$= \frac{ac}{|ac|} \gamma_{xy}.$$

Problem 3. ...

Problem 3. A programmer while writing a program for correlation coefficient between two variables x and y from 30 pairs of observations obtained the following results $\Sigma x = 300$; $\Sigma x^2 = 3718$; $\Sigma y = 210$; $\Sigma y^2 = 2000$; $\Sigma xy = 2100$. At the time of checking it was found that he had copied down two pairs (x_i, y_i) as (18, 20) and (12, 10) instead of the correct values (10, 15) and (20, 15). Obtain the correct value of the correlation coefficient.

Solution. Corrected $\Sigma x = 300 - 18 - 12 + 10 + 20 = 300$

Corrected $\Sigma y = 210 - 20 - 10 + 15 + 15 = 210$

Corrected $\Sigma x^2 = 3718 - 18^2 - 12^2 + 10^2 + 20^2 = 3750$

Corrected $\Sigma y^2 = 2000 - 20^2 - 10^2 + 15^2 + 15^2 = 1950$

Corrected $\Sigma xy = 2100 - (18 \times 20) - (12 \times 10) + (10 \times 15) + (20 \times 15) = 2070.$

After correction the correlation coefficient is

$$\gamma_{xy} = \frac{n \Sigma xy - \Sigma x \, \Sigma y}{\left[n \Sigma x^2 - (\Sigma x)^2\right]^{1/2} \left[n \Sigma y^2 - (\Sigma y)^2\right]^{1/2}}$$

$$\therefore \gamma_{xy} = \frac{30 \times 2070 - 300 \times 210}{\left[30 \times 3750 - 300^2\right]^{1/2} \left[30 \times 1950 - 210^2\right]^{1/2}}$$

$$= \frac{62100 - 63000}{(112500 - 90000)^{1/2} \, (58500 - 44100)^{1/2}}$$

$$= \frac{-900}{(22500)^{1/2} \, (14400)^{1/2}} = -\frac{900}{150 \times 120} = -\frac{1}{20}$$

$$= -0.05$$

Problem 4. If x, y and z are uncorrelated variables each having same standard deviation obtain the coefficient of correlation between $x + y$ and $y + z$.

Solution. Given $\sigma_x = \sigma_y = \sigma_z = \sigma$ (say)

x and y are uncorrelated $\Rightarrow \Sigma (x - \bar{x})(y - \bar{y}) = 0$.

y and z are uncorrelated $\Rightarrow \Sigma (y - \bar{y})(z - \bar{z}) = 0$.

z and x are uncorrelated $\Rightarrow \Sigma (z - \bar{z})(x - \bar{x}) = 0$.

Let $u = x + y$ and $v = y + z$.

$\therefore \bar{u} = \bar{x} + \bar{y}$ and $\bar{v} = \bar{y} + \bar{z}$.

Now, $\sigma_u^2 = \frac{1}{n} \Sigma (u - \bar{u})^2 = \frac{1}{n} \Sigma [(x - \bar{x}) + (y - \bar{y})]^2$

$= \frac{1}{n} \left[\Sigma (x - \bar{x})^2 + \Sigma (y - \bar{y})^2 + 2\Sigma (x - \bar{x})(y - \bar{y}) \right]$

$= \sigma_x^2 + \sigma_y^2$ (since $\Sigma (x - \bar{x})(y - \bar{y}) = 0$

$= 2\sigma^2$.

Similarly $\sigma_v^2 = 2\sigma^2$.

Now, $\Sigma (u - \bar{u})(v - \bar{v})$

$$= \Sigma [\{(x - \bar{x}) + (y - \bar{y})\}\{(y - \bar{y}) + (z - \bar{z})\}]$$

$$= \Sigma (x - \bar{x})(y - \bar{y}) + \Sigma (y - \bar{y})^2 + \Sigma (x - \bar{x})(z - \bar{z}) + \Sigma (y - \bar{y})(z - \bar{z})$$

$$= 0 + n\sigma_y^2 + 0 + 0 = n\sigma^2.$$

$$\therefore \gamma_{uv} = \frac{\Sigma (u - \bar{u})(v - \bar{v})}{n\sigma_u \sigma_v} = \frac{n\sigma^2}{n(2\sigma^2)} = \frac{1}{2}.$$

Problem

Problem 5. Show that the variables $u = x\cos\alpha + y\sin\alpha$ and $v = y\cos\alpha - x\sin\alpha$ are uncorrelated if $\alpha = \frac{1}{2}\tan^{-1}\left(\frac{2\gamma_{xy}\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}\right)$

Solution.

$$u_i = x_i\cos\alpha + y_i\sin\alpha \text{ and } v_i = y_i\cos\alpha - x_i\sin\alpha.$$

$$\therefore \bar{u} = \bar{x}\cos\alpha + \bar{y}\sin\alpha \text{ and } \bar{v} = \bar{y}\cos\alpha - \bar{x}\sin\alpha$$

$$\therefore u_i - \bar{u} = (x_i - \bar{x})\cos\alpha + (y_i - \bar{y})\sin\alpha.$$

The variables u_i and v_i are uncorrelated if $\Sigma(u_i - \bar{u})(v_i - \bar{v}) = 0.$

$$\therefore \Sigma[(x_i - \bar{x})\cos\alpha + (y_i - \bar{y})\sin\alpha]$$

$$[(y_i - \bar{y})\cos\alpha - (x_i - \bar{x})\sin\alpha] = 0$$

$$\therefore \Sigma(x_i - \bar{x})(y_i - \bar{y})\cos^2\alpha - \Sigma(x_i - \bar{x})(y_i - \bar{y})\sin^2\alpha$$

$$- \cos\alpha\sin\alpha[\Sigma(x_i - \bar{x})^2 - \Sigma(y_i - \bar{y})^2] = 0.$$

$$\therefore n\gamma_{xy}\sigma_x\sigma_y(\cos^2\alpha - \sin^2\alpha) = n\cos\alpha\sin\alpha(\sigma_x^2 - \sigma_y^2)$$

$$\therefore \gamma_{xy}\sigma_x\sigma_y\cos 2\alpha = \frac{1}{2}\sin 2\alpha(\sigma_x^2 - \sigma_y^2)$$

$$\therefore \tan 2\alpha = \frac{2\gamma_{xy}\,\sigma_x\,\sigma_y}{\sigma_x^2 - \sigma_y^2}$$

$$\therefore \alpha = \frac{1}{2}\tan^{-1}\left(\frac{2\gamma_{xy}\,\sigma_x\,\sigma_y}{\sigma_x^2 - \sigma_y^2}\right)$$

Problem 6. Show that if X', Y' are the deviations of the random variables X and Y from their respective means then (i) $\gamma = 1 - \frac{1}{2N}\Sigma\left(\frac{X_i'}{\sigma_X} - \frac{Y_i'}{\sigma_Y}\right)^2$.

(ii) $\gamma = -1 + \frac{1}{2N}\Sigma\left(\frac{X_i'}{\sigma_X} + \frac{Y_i'}{\sigma_Y}\right)^2$. Deduce that $-1 \le \gamma \le 1$

Solution. (i) Given that $X_i' = X_i - \bar{X}$ and $Y' = Y_i - \bar{Y}$.

$$1 - \frac{1}{2N}\Sigma\left(\frac{X_i'}{\sigma_X} - \frac{Y_i'}{\sigma_Y}\right)^2 = 1 - \frac{1}{2N}\Sigma\left(\frac{X_i - \bar{X}}{\sigma_X} - \frac{Y_i - \bar{Y}}{\sigma_Y}\right)^2$$

$$= 1 - \frac{1}{2N}\left(\frac{\Sigma(X_i - \bar{X})^2}{\sigma_X^2} + \frac{\Sigma(Y_i - \bar{Y})^2}{\sigma_Y^2} - \frac{2\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sigma_X\,\sigma_Y}\right)$$

$$= 1 - \frac{1}{2N}\left[\frac{N\sigma_X^2}{\sigma_X^2} + \frac{N\sigma_Y^2}{\sigma_Y^2} - 2\gamma N\right]$$

$$= 1 - \frac{1}{2N}[N + N - 2\gamma N] = 1 - \frac{1}{2N}[2N - 2\gamma N]$$

$$= 1 - (1 - \gamma) = \gamma.$$

(ii) can similarly be proved.

Since $\Sigma\left(\frac{X_i'}{\sigma_X} + \frac{Y_i'}{\sigma_Y}\right)^2$ is always positive we have $\frac{1}{2N}\Sigma\left(\frac{X_i'}{\sigma_X} + \frac{Y_i'}{\sigma_Y}\right)^2$ is positive.

Hence $1 - \frac{1}{2N} \Sigma\left(\frac{X_i}{\sigma_X} + \frac{Y_i}{\sigma_Y}\right)^2 \leq 1$.

$\therefore$ By (i) $\gamma \leq 1$ Similarly by (ii) $-1 \leq \gamma$

Hence $-1 \leq \gamma \leq 1$.

Problem 7. Let x, y be two variables with standard deviation σ_x and σ_y respectively. If $u = x + ky$ and $v = x + \left(\frac{\sigma_x}{\sigma_y}\right) y$ and $\gamma_{uv} = 0$ (ie. u and v are uncorrelated) find the value of k.

Solution. $u = x + ky \Rightarrow \overline{u} = \overline{x} + k\overline{y}$

$$v = x + \left(\frac{\sigma_x}{\sigma_y}\right) y \Rightarrow \overline{v} = \overline{x} + \left(\frac{\sigma_x}{\sigma_y}\right) \overline{y}$$

$$u - \overline{u} = (x - \overline{x}) + k(y - \overline{y}) \text{ and } v - \overline{v} = (x - \overline{x}) + \left(\frac{\sigma_x}{\sigma_y}\right)(y - \overline{y})$$

Now, $\gamma_{uv} = 0 \Rightarrow Cov(u, v) = 0$

$$\Rightarrow \Sigma (u - \overline{u})(v - \overline{v}) = 0.$$

$$\Rightarrow \Sigma [(x - \overline{x}) + k(y - \overline{y})] \left[(x - \overline{x}) + \left(\frac{\sigma_x}{\sigma_y}\right)(y - \overline{y})\right] = 0.$$

$$\Rightarrow \Sigma (x - \overline{x})^2 + k\left(\frac{\sigma_x}{\sigma_y}\right) \Sigma (y - \overline{y})^2 + k \Sigma (x - \overline{x})(y - \overline{y}) + \left(\frac{\sigma_x}{\sigma_y}\right) \Sigma (x - \overline{x})(y - \overline{y}) = 0.$$

$$\Rightarrow n\sigma_x^2 + nk\left(\frac{\sigma_x}{\sigma_y}\right)\sigma_y^2 + n\gamma_{xy}\sigma_x\sigma_y\left(k + \frac{\sigma_x}{\sigma_y}\right) = 0.$$

$$\Rightarrow n\sigma_x [\sigma_x + k\sigma_y + \gamma_{xy}(k\sigma_y + \sigma_x)] = 0.$$

$$\Rightarrow n\sigma_x [(\sigma_x + k\sigma_y)(1 + \gamma_{xy})] = 0.$$

$$\Rightarrow \sigma_x(\sigma_x + k\sigma_y)(1 + \gamma_{xy}) = 0.$$

$$\Rightarrow \sigma_x + k\sigma_y = 0 \text{ or } \gamma_{xy} + 1 = 0 \text{ or } \sigma_x = 0$$

If $\gamma_{xy} \neq -1$ and $\sigma_x \neq 0$ we get $k = -(\sigma_x/\sigma_y)$.

Assignment

Text book page No.	Question. No.
117	1(i), 1(iii), 1(iv)
118	1(x),1(xi),4
119	7,8

THANK YOU

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-3

CHAPTER6.2

PART-14

R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

contents

- RANK CORRELATION
- THEOREM 6.5
- PROBLEMS

RANK CORERELATION

The same individuals are ranked in two ways on the basis of different characteristics or by two different persons for a single characteristic.

Let x_i and y_i be the ranks of the i^{th} individuals in the first and second ranking respectively.

The coefficient of correlation between the ranks x_i and y_i is called the Rank correlation and is denoted by ρ.

...ed by ρ.

Theorem 6.5 Rank correlation ρ is given by $\rho = 1 - \frac{6\,\Sigma(x-y)^2}{n(n^2-1)}$

Proof. Consider a collection of n individuals. Let x_i and y_i be the ranks of the i^{th} individual in the two different rankings.

$$\therefore\ \bar{x} = \frac{1}{2}(n+1) = \bar{y} \text{ and } \sigma_x^{\,2} = \frac{1}{12}(n^2-1) = \sigma_y^{\,2}.$$

Now, $\Sigma(x-y)^2 = \Sigma[(x-\bar{x}) - (y-\bar{y})]^2$ (since $\bar{x} = \bar{y}$)

$$= \Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2 - 2\Sigma(x-\bar{x})(y-\bar{y})$$

$$= n\sigma_x^{\,2} + n\sigma_y^{\,2} - 2n\rho\sigma_x\sigma_y$$

$$= 2n\sigma_x^{\,2}(1-\rho) \text{ (since } \sigma_x^{\,2} = \sigma_y^{\,2})$$

$$= \frac{1}{6}n(n^2-1)(1-\rho)$$

$$\therefore\ 1-\rho = \frac{6\,\Sigma(x-y)^2}{n(n^2-1)}$$

$$\therefore\ \rho = 1 - \frac{6\,\Sigma(x-y)^2}{n(n^2-1)}$$

$\rho = 1 - \frac{6\,\Sigma(x-y)^2}{n(n^2-1)}$ **is known as Spearman's formula.**

Note. This is known as Spear...

PROBLEM:1

Find the rank correlation coefficient between the height in c.m. and weight in kg of 6 soldiers in Indian Army.

Height	165	167	166	170	169	172
Weight	61	60	63.5	63	61.5	64

Solution:

Height X	Rank in Height x	Weight Y	Rank in Weight y	x-y	$(x-y)^2$
165	6	61	5	1	1
167	4	60	6	-2	4
166	5	63.5	2	3	9
170	2	63	3	-1	1
169	3	61.5	4	-1	1
172	1	64	1	0	0
Total					16

$N=6$

$$\rho = 1 - \frac{6\sum(x-y)^2}{n(n^2-1)}$$

$$\rho = 1 - \frac{6X16}{6(6^2-1)}$$

$=0.543$

PROBLEM:2 Three judges assign the ranks to 8 entries in a beauty contest.

Judge Mr. X	**1**	**2**	**4**	**3**	**7**	**6**	**5**	**8**
Judge Mr. Y	**3**	**2**	**1**	**5**	**4**	**7**	**6**	**8**
Judge Mr. Z	**1**	**2**	**3**	**4**	**5**	**7**	**8**	**6**

Which pair of judges has the nearest approach to common taste in beauty?

Solution: Table for the rank correlation coefficients $\rho_{xy}, \rho_{yz}, \rho_{zx}$.

x	y	z	x- y	y- z	z- x	$(x-y)^2$	$(y-z)^2$	$(z-x)^2$
1	3	1	-2	2	0	4	4	0
2	2	2	0	0	0	0	0	0
4	1	3	3	-2	-1	9	4	1
3	5	4	-2	1	1	4	1	1
7	4	5	3	-1	-2	9	1	4
6	7	7	-1	0	1	1	0	1
5	6	8	-1	-2	3	1	4	9
8	8	6	0	2	-2	0	4	4
Tot						28	18	20

$$\rho_{xy} = 1 - \frac{6\sum(x-y)^2}{n(n^2-1)}$$

$$\rho_{xy} = 1 - \frac{6X28}{8(8^2-1)}$$

$$= 1 - \frac{168}{504} = 0.67$$

$$\rho_{zx} = 1 - \frac{6\sum(z-x)^2}{n(n^2-1)}$$

$$\rho_{zx} = 1 - \frac{6X20}{8(8^2-1)}$$

$$= 1 - \frac{120}{504} = 0.76$$

$$\rho_{yz} = 1 - \frac{6\sum(y-z)^2}{n(n^2-1)}$$

$$\rho_{yz} = 1 - \frac{6X18}{8(8^2-1)}$$

$$= 1 - \frac{108}{504} = 0.79$$

Hence $\rho_{yz>}\ \rho_{xy}$ ***and*** ρ_{zx}

Therefore Judges Mr.Y and Mr. Z has the nearest approach to common taste in beauty?

Assignment

Text book page No.	Question. No.
126	1(i), 1(iv)
127	5
128	7,9

B. Sc., III YEAR- V SEMESTER

STATISTICS-I

COURSE CODE: 7BMA5C2

UNIT-3

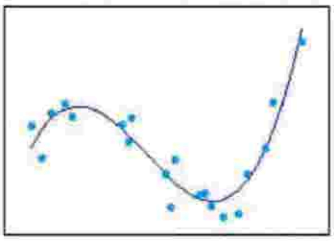
CHAPTER6.3

PART-15

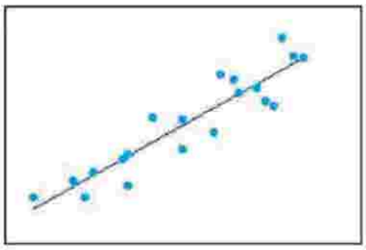
R.RAJALAKSHMI [G.L],GACW,RAMANATHAPURAM

REGRESSIO N

If there is a functional relationship between the two variables x_i and y_i the points in the scatter diagram will cluster around Some curve called the curve of regression.



If the curve is a straight line it is called a line of regression between the to variables



LINE OF REGRESSION

If we fit a straight line by the principle of least squares to the points of the scatter diagram in such a way that the sum of the squares of the distance parallel to the y-axis from the points to the line is minimized we obtain a line of best fit for the data and it is called the **regression line of y on x.**

Similarly we can define the **regression line of x on y.**

Theorem 6.6 The equation of the regression line of y on x is given by

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

Proof. Let $y = ax + b$ be the line of regression of y on x.

According to the principle of least squares the constants a and b are to be determined in such a way that $S = \Sigma\,[y_i - (ax_i + b)]^2$ is minimum.

$$\frac{\partial S}{\partial a} = 0 \Rightarrow -2\Sigma(y_i - ax_i - b)x_i = 0$$

$$\Rightarrow \Sigma x_i y_i = a\Sigma x_i^2 + b\Sigma x_i \quad \text{..........(1)}$$

$$\frac{\partial S}{\partial b} = 0 \Rightarrow -2\Sigma(y_i - ax_i - b) = 0$$

$$\Rightarrow \Sigma y_i = a \Sigma x_i + nb \quad \text{.........(2)}$$

Equations (1) and (2) are called normal equations.

From (2) we obtain $\bar{y} = a\bar{x} + b$ ....(3)

The line of regression passes through the point $(\bar{x}, \bar{y})$.

Now, shifting the origin to this point $(\bar{x}, \bar{y})$ by means of the transformation $X_i = x_i - \bar{x}$ and $Y_i = y_i - \bar{y}$ we obtain $\Sigma X_i = 0 = \Sigma Y_i$ and the equation of the line of regression becomes $Y = aX$ ....(4)

Corresponding to this line $Y = aX$, the constant a can be determined from the normal equation $a \Sigma X_i^2 = \Sigma X_i Y_i$

$$\therefore a = \frac{\Sigma X_i Y_i}{\Sigma X_i^2} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{\gamma \sigma_x \sigma_y}{\sigma_x^2} = \gamma \frac{\sigma_y}{\sigma_x}$$

$\therefore$ The required regression line (4) becomes $Y = \left(\gamma \frac{\sigma_y}{\sigma_x}\right) X$

$$y - \bar{y} = \gamma \frac{\sigma_y}{\sigma_x}(x - \bar{x}).$$

Theorem ...

Theorem 6.7 The equation of regression line of x on y is given by

$$x - \bar{x} = \gamma \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

Proof. Proof is similar to that of theorem 6.6

Note. $(\bar{x}, \bar{y})$ is the point of intersection of the two regression lines.

Definition. The slope of the regression line of y on x is called the **regression coefficient of y on x** and it is denoted by b_{yx}. Hence $b_{yx} = \gamma \frac{\sigma_y}{\sigma_x}$

The regression coefficient of x on y is given by $b_{xy} = \gamma \frac{\sigma_x}{\sigma_y}$

We now give some properties of the regression coefficients.

Theorem 6.8 Correlation coefficient is the geometric mean between the regression coefficients. (i.e) $\gamma = \pm \sqrt{b_{xy}\, b_{yx}}$

Proof. We have $b_{yx} = \gamma \dfrac{\sigma_y}{\sigma_x}$ and $b_{xy} = \gamma \dfrac{\sigma_x}{\sigma_y}$

$$\therefore\ b_{yx}\, b_{xy} = \gamma^2$$

$$\therefore\ \gamma = \pm \sqrt{b_{xy}\, b_{yx}}$$

Note. The sign of the correlation coefficient is the same as that of regression coefficients.

Theorem 6.9 If one of the regression coefficients is greater than unity the other is less than unity

Proof. We have $b_{xy}\, b_{yx} = \gamma^2 \leq 1$ so that $b_{xy}\, b_{yx} \leq 1$

Hence $b_{xy} > 1 \Rightarrow b_{yx} < 1$

Hence the theorem

Theorem 6.10 Arithmetic mean of the regression coefficients is greater than or equal to the correlation coefficient

Proof. Let b_{xy} and b_{yx} be the regression coefficients.

We have to prove $\frac{1}{2}(b_{xy} + b_{yx}) \geq \gamma$.

Now, $\frac{1}{2}(b_{xy} + b_{yx}) \geq \gamma \Leftrightarrow b_{yx} + b_{xy} \geq 2\gamma$

$$\Leftrightarrow \gamma \frac{\sigma_y}{\sigma_x} + \gamma \frac{\sigma_x}{\sigma_y} \geq 2\gamma$$

$$\Leftrightarrow \frac{\sigma_y}{\sigma_x} + \frac{\sigma_x}{\sigma_y} \geq 2$$

$$\Leftrightarrow \sigma_x{}^2 + \sigma_y{}^2 - 2\sigma_x \sigma_y \geq 0.$$

$$\Leftrightarrow (\sigma_x - \sigma_y)^2 \geq 0.$$

This is always true. Hence the theorem.

em 6.11 Regr

...the theorem.

Theorem 6.11 Regression coefficients are independent of the change of origin but dependent on change of scale.

Proof. Let $u_i = \frac{x_i - A}{h}$ and $v_i = \frac{y_i - B}{k}$

Let $x_i = A + hu_i$ and $y_i = B + kv_i$

We know that $\sigma_x = h\sigma_u$; $\sigma_y = k\sigma_v$ and $\gamma_{xy} = \gamma_{uv}$

Now, $b_{yx} = \gamma_{xy} \frac{\sigma_y}{\sigma_x} = \gamma_{uv} \left(\frac{k\,\sigma_v}{h\,\sigma_u}\right) = \frac{k}{h}\, b_{uv}$ ——(1)

Similarly $b_{xy} = (h/k)\, b_{uv}$ ——(2)

From (1) and (2) we observe that b_{yx} and b_{xy} depend upon the scales h and k but not on the origins A and B

Hence the theorem.

Theorem 6.12 The angle between the two regression lines is given by

$$\theta = \tan^{-1}\left[\left(\frac{\gamma^2 - 1}{\gamma}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)\right]$$

Proof. The equations of lines of regression of y on x and x on y respectively

are $y - \bar{y} = \gamma \frac{\sigma_y}{\sigma_x}(x - \bar{x})$ ....(1)

$x - \bar{x} = \gamma \frac{\sigma_x}{\sigma_y}(y - \bar{y})$ ....(2)

(2) can also be written as $y - \bar{y} = \frac{1}{\gamma}\frac{\sigma_y}{\sigma_x}(x - \bar{x})$ ....(3)

$\therefore$ Slopes of the two lines (1) and (2) are $\gamma \frac{\sigma_y}{\sigma_x}$ and $\frac{\sigma_y}{\gamma \sigma_x}$.

Let θ be the acute angle between the two lines of regression.

$$\therefore \tan\theta = \frac{\frac{\gamma \sigma_y}{\sigma_x} - \frac{\sigma_y}{\gamma \sigma_x}}{1 + \left(\gamma \frac{\sigma_y}{\sigma_x}\right)\left(\frac{\sigma_y}{\gamma \sigma_x}\right)}$$

$$= \frac{\gamma^2 - 1}{\gamma}\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$$

$$= \frac{1 - \gamma^2}{\gamma}\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right) \text{ (since } \gamma^2 \le 1 \text{ and } \theta \text{ is acute)}$$

$$\therefore \theta = \tan^{-1}\left[\left(\frac{1 - \gamma^2}{\gamma}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)\right]$$

Note 1 The obtuse angle between the regression lines is given by

$$\tan^{-1}\left[\left(\frac{\gamma^2-1}{\gamma}\right)\left(\frac{\sigma_x\,\sigma_y}{\sigma_x^2+\sigma_y^2}\right)\right]$$

Note 2. If $\gamma = 0$ then $\tan\theta = \infty$ Hence $\theta = \pi/2$. Thus if the *two varibles are uncorrelated then the lines of regression are prependicular to each other.*

Note 3. If $\gamma = \pm 1$ then $\tan\theta = 0$.

Hence $\theta = 0$ or π.

$\therefore$ The two lines of regression are parallel.

Further the two lines have the common point $(\bar{x}, \bar{y})$ and hence the must be coincident.

Therefore if there is a perfect correlation (positive or negative between the two variables then the two lines of regression coincide.

THANK YOU

UNIT- IV

Interpolation

Defn:

Interpolation is the process of finding the Most appropriate estimate for Missing data.

Interpolation technique is used in Various disciplines like economics, business, population studies, price determination etc. It is used to fill in the gaps in the statistical data for the sake of continuity of information.

There are two Methods in interpolation.

(i) Graphical Method

(ii) Algebraic Method.

i) Graphic Method:

It is a Simple Method in Which We just plot the available data on a graph sheet and read off the Value for the missing period from the graph itself.

ii) Algebraic Method:

There are Several Method used for interpolation.

(a) Finite difference.

(b) Gregory - Newton's formula

(c) Lagrange's formula.

Finite Difference

Defn:

We define an operator Δ which is known as the first order difference on U_x as $\Delta U_x = U_{x+h} - U_x$, where $x = a, a+h, a+2h, \ldots$

In particular,

(i) $\Delta U_a = U_{a+h} - U_a$

(ii) $\Delta U_x = 0$ if U_x is constant

Note:

(i) Δ is linear

ie) $\Delta (aU_x + bV_x) = a\,\Delta U_x + b\,\Delta V_x$

(ii) Δ satisfies the law of indices for multiplication

ie) $\Delta^m \Delta^n U_x = \Delta^{m+n} U_x$

The difference ΔU_x, $\Delta^2 U_x$ are called forward difference of U_x.

We define an operator ∇ which is known as first order backward difference on U_x as $\nabla U_x = U_x - U_{x-h}$ where, $x = a, a+h, a+2h$.

Note:

(i) $\nabla U_{x+h} = \Delta U_x$

(ii) $\nabla^n U_{a+nh} = \Delta^n U_a$

The operator E:

The operator E on U_x is defined as $EU_x = U_{x+h}$

The higher order operator of E can similarly be defined.

Generally, $E^n U_x = U_{x+nh}$

If $h=1$, the $E^n U_x = U_{x+n}$

For eg:

$E^5 U_0 = U_{0+5} = U_5$

$E^3 U_4 = U_{4+3} = U_7$

$E^4 U_{-1} = U_{4-1} = U_3$

Theorem: 7.1 (i) $E = 1+\Delta$ (ii) $E = (1-\nabla)^{-1}$

Let U_x be a function of x.

(i) $\Delta U_x = U_{x+h} - U_x$

$= EU_x - U_x$

$= (E-1)U_x$

$\therefore \Delta = E-1$ Hence $E = 1+\Delta$

(ii) $\nabla U_x = U_x - U_{x-h}$

$\therefore U_{x-h} = U_x - \nabla U_x$

$= (1-\nabla)U_x$

$E^{-1} U_x = (1-\nabla) U_x$ (since $EU_{x-h} = U_x \Rightarrow U_{x-h} = E^{-1} U_x$)

$\therefore E^{-1} = 1-\nabla$. Hence $E = (1-\nabla)^{-1}$

Note 1:

It is very easy to verify that the operator E also satisfies the basic law of algebra such a linearity and law of indices for multiplication.

Lemma:

The two operators Δ and E are commutative under composition of operation

(ie) $\Delta \circ E = E \circ \Delta$

Proof:

$$(\Delta \cdot E)(U_x) = \Delta\left(E(U_x)\right) = \Delta(U_{x+h}) = U_{x+2h} - U_{x+h}$$

$$= EU_{x+h} - EU_x$$

$$= E(U_{x+h} - U_x)$$

$$= E(\Delta(U_x))$$

$$= (E \cdot \Delta)(U_x)$$

Hence $\Delta \cdot E = E \cdot \Delta$

Problems:

1) Find first and second order difference for (i) $U_x = ab^{cx}$

(ii) $U_x = \dfrac{x}{x^2+7x+12}$ taking interval of differencing as h.

(i) $\Delta U_x = U_{x+h} - U_x = ab^{c(x+h)} - ab^{cx}$

$$= ab^{cx} b^{ch} - ab^{cx}$$

$$= ab^{cx}(b^{ch} - 1)$$

$$\Delta^2 U_x = (b^{ch} - 1)\Delta(ab^{cx}) = (b^{ch}-1)^2 ab^{cx}$$

(ii) $U_x = \dfrac{x}{x^2+7x+12} = \dfrac{x}{x+4} - \dfrac{3}{x+3}$ (by partial fraction)

$$\therefore \Delta U_x = \left[\frac{4}{(x+1)+4} - \frac{3}{(x+1)+3}\right] - \left[\frac{4}{x+4} - \frac{3}{x+3}\right]$$

$$= \frac{4}{x+5} - \frac{3}{x+4} - \frac{4}{x+4} + \frac{3}{x+3}$$

$$= \frac{4}{x+5} - \frac{7}{x+4} + \frac{3}{x+3}$$

Similarly, $\Delta^2 U_x = \frac{4}{x+6} - \frac{11x}{x+5} + \frac{10}{x+4} - \frac{3}{x+3}$ (Verify)

Problem: 3

Find $\Delta^n \sin x$ taking $h=1$

$$\Delta \sin x = \sin(x+1) - \sin x = 2\cos\left(x+\frac{1}{2}\right)\sin\left(\frac{1}{2}\right)$$

$$= 2\sin\left(\frac{1}{2}\right)\sin\left(x+\frac{1}{2}+\frac{\pi}{2}\right)$$

Now, $\Delta^2 \sin x = \Delta\left[2\sin\left(\frac{1}{2}\right)\sin\left(x+\frac{1}{2}+\frac{\pi}{2}\right)\right]$

$$= 2\sin\left(\frac{1}{2}\right)\left[\sin\left(x+\frac{1}{2}+\frac{\pi}{2}+1\right) - \sin\left(x+\frac{1}{2}+\frac{\pi}{2}\right)\right]$$

$$= 2\sin\left(\frac{1}{2}\right)\left[2\cos\left(x+1+\frac{\pi}{2}\right)\sin\left(\frac{1}{2}\right)\right]$$

$$= \left[2\sin\left(\frac{1}{2}\right)\right]^2 \sin\left[x+2\left(\frac{1}{2}+\frac{\pi}{2}\right)\right]$$

Preceeding like this we get, $\Delta^n \sin x = \left[2\sin\left(\frac{1}{2}\right)\right]^n \sin\left[x+n\left(\frac{1}{2}+\frac{\pi}{2}\right)\right]$

Problem: 4

P.T $\Delta(\log U_x) = \log\left(1+\frac{\Delta U_x}{U_x}\right)$

$$\Delta \log U_x = \log U_{x+h} - \log U_x = \log\left(\frac{U_{x+h}}{U_x}\right) = \log\left(\frac{EU_x}{U_x}\right)$$

$$= \log\left[\frac{(1+\Delta)U_x}{U_x}\right]$$

$$= \log\left(\frac{U_x+\Delta U_x}{U_x}\right)$$

$$= \log\left(1+\frac{\Delta U_x}{U_x}\right)$$

Pbm: 5

Evaluate $\frac{\Delta^2 x^3}{E x^2}$ taking $h=1$

$\Delta x^3 = (x+1)^3 - x^3 = 3x^2 + 3x + 1$

$\Delta^2 x^3 = \Delta(\Delta x^3) = \Delta(3x^2 + 3x + 1)$

$= 3\Delta x^2 + 3\Delta x + \Delta(1)$

$= 3\left[(x+1)^2 - x^2\right] + 3\left[(x+1) - x\right] + 0$

$= 6(x+1)$

Now, $Ex^2 = (x+1)^2$

$\therefore \frac{\Delta^2 x^3}{Ex^2} = \frac{6(x+1)}{(x+1)^2} = \frac{6}{x+1}$

Pbm: 7

If $U_0 = 1$, $U_1 = 5$, $U_2 = 8$, $U_3 = 3$, $U_4 = 7$, $U_5 = 0$ Find $\Delta^5 U_0$.

Consider $\Delta^5 U_0 = (E-1)^5 U_0$

$= (E^5 - 5E^4 + 10E^3 - 10E^2 + 5E - 1) U_0$

$= E^5 U_0 - 5E^4 U_0 + 10E^3 U_0 - 10E^2 U_0 + 5EU_0 - U_0$

$= U_5 - 5U_4 + 10U_3 - 10U_2 + 5U_1 - U_0$

$= 0 - (5 \times 7) + (10 \times 3) - (10 \times 8) + (5 \times 5) - 1$

$= -35 + 30 - 80 + 25 - 1$

$= -61$

Alter'

x	U_x	ΔU_x	$\Delta^2 U_x$	$\Delta^3 U_x$	$\Delta^4 U_x$	$\Delta^5 U_x$
0	1					
		4				
1	5		-1			
		3		-7		
2	8		-8		24	
		-5		17		-61
3	3		9		-37	
		4		-20		
4	7		-11			
		-7				
5	0					

Here $\Delta^5 U_0 = -61$

pbm:8

Estimate the Missing term in the following table.

x	0	1	2	3	4
U_x	1	3	9	-	81

Explain Why the resulting Value differ from 3^3.

Let the Missing term in U_x be a

The difference table is given below

x	U_x	ΔU_x	$\Delta^2 U_x$	$\Delta^3 U_x$	$\Delta^4 U_x$
0	1				
		2			
1	3		4	a-19	
		6			124-4a
2	9		a-15	105-3a	
		a-9			
3	a		90-20		
		81-a			
4	81				

Since 4 value of U_4 are given it is a polynomial of degree 3.

Hence by fundamental theorem of finite difference $\Delta^4 U_x = 0$ for all x.

In particular $\Delta^4 U_0 = 0$. Hence $124 - 4a = 0$

$\therefore \boxed{a = 31}$

Aliter:

consider $\Delta^4 U_0 = 0$

$\therefore (E-1)^4 U_0 = 0$

$\therefore (E^4 - 4E^3 + 6E^2 - 4E + 1) U_0 = 0$

$\therefore U_4 - 4U_3 + 6U_2 - 4U_1 + U_0 = 0$

$\therefore 81 - 4a + 6\times 9 - 4\times 3 + 1 = 0$

$\therefore 124 - 4a = 0$

$\therefore \boxed{a = 31}$

pbm: 9

Given an estimate of the population in 1971 from the following table.

Year	1941	1951	1961	1971	1981	1991
Population in Lakhs	363	391	421	?	467	501

Let the population in 1971 be a.

$u_0 = 363$; $u_1 = 391$; $u_2 = 421$; $u_3 = a$; $u_4 = 467$ and $u_5 = 501$

$\Delta^5 u_x = 0$

$\therefore (E-1)^5 u_0 = 0$

$\therefore (E^5 - 5E^4 + 10E^3 - 10E^2 + 5E - 1) u_0 = 0$

$\therefore u_5 - 5u_4 + 10u_3 - 10u_2 + 5u_1 - u_0 = 0$

$\therefore 501 - 5(467) + 10a - 10(421) + 5(391) - 363 = 0$

$\therefore 501 - 2335 + 10a - 4210 + 1955 - 363 = 0$

$\therefore 10a - 4452 = 0$

$\therefore \boxed{a = 445.2 \text{ lakhs}}$.

Hence the estimated population in 1971 is 445.2 lakhs.

pbm: 10

Find the Missing figure in the following table.

x	0	5	10	15	20	25
u_x	7	11	?	18	?	32

Here 2 value are Missing

Let the Missing value be a and b.

$u_0 = 7, \; u_1 = 11, \; u_2 = a, \; u_3 = 18, \; u_4 = b, \; u_5 = 32$

$\Delta^4 u_x = 0$

In particular $\Delta^4 u_0 = 0$ and $\Delta^4 u_1 = 0$

$\Delta^4 u_0 = 0$

$\therefore (E-1)^4 u_0 = 0$

$\therefore (E^4 - 4E^3 + 6E^2 - 4E + 1) u_0 = 0$

$\therefore u_4 - 4u_3 + 6u_2 - 4u_1 + u_0 = 0$

$b - 72 + 6a - 44 + 7 = 0$

i.e) $6a + b = 109$ ——— (1)

Taking $\Delta^4 U_1 = 0$

$\therefore (E-1)^4 U_1 = 0$

$\therefore (E^4 - 4E^3 + 6E^2 - 4E + 1) U_1 = 0$

$U_5 - 4U_4 + 6U_3 - 4U_2 + U_1 = 0$

$32 - 4b + 108 - 4a + 11 = 0$

i.e) $4a + 4b = 151$ ——— (2)

Solving (1) and (2)

We get $a = 14.25$ and $b = 23.5$

Hence the Missing Value are 14.25 and 23.5

Pbm:11

Given that $U_0 + U_8 = 80$; $U_1 + U_7 = 10$; $U_2 + U_6 = 5$;

$U_3 + U_5 = 10$ find U_4

Since 4 values are given $\Delta^n U_x = 0$ for all $n \geq 4$ and for all x

In particular $\Delta^8 U_0 = 0$.

$(E-1)^8 U_0 = 0$

$U_8 - 8U_7 + 28U_6 - 56U_5 + 70U_4 - 56U_3 + 28U_2 - 8U_1 + U_0 = 0$

$(U_0 + U_8) - 8(U_1 + U_7) + 28(U_2 + U_6) - 56(U_3 + U_5) + 70U_4 = 0$

$\therefore 80 - 80 + 140 - 560 + 70U_4 = 0$

$\therefore 70U_4 = 420$

$\therefore \boxed{U_4 = 6}$

pbm: 12

Given that $U_1 + U_2 + U_3 = 25$, $U_4 = 29$, $U_5 + U_6 = 113$

Find polynomial U_x and hence find U_{10}.

Since three values are given U_x is a polynomial of degree 2.

Let $U_x = ax^2 + bx + c$

We note: $U_1 = a + b + c$; $U_2 = 4a + 2b + c$; $U_3 = 9a + 3b + c$

Given $U_1 + U_2 + U_3 = 25$

$\therefore 14a + 6b + 3c = 25$ ——— (1)

Now, $U_4 = 24$, $\Rightarrow 16a + 4b + c = 29$ ——— (2)

$U_5 + U_6 = 113 \Rightarrow 61a + 11b + 2c = 113$ ——— (3)

Solving (1), (2) and (3) we get

$a = 2, \; b = -1, \; c = 1$

$\therefore U_x = 2x^2 - x + 1$

Now, $U_{10} = 100a + 10b + c = 200 - 10 + 1$

$= 191$

Theorem 7.3 (Newton-Gregory Interpolating formula for equal intervals)

Let $U_a, U_{a+h}, \ldots, U_{a+nh}$ be the value of the function U_x at the point $a, a+h, a+2h, \ldots a+nh$, which are of equal interval of difference.

Then $U_x = U_a + (x-a)\dfrac{\Delta U_a}{1!\,h} + (x-a)(x-a-h)\dfrac{\Delta^2 U_a}{2!\,h^2} + \cdots$

$+ (x-a)(x-a-h)\ldots(x-a-\overline{n-1}h)\dfrac{\Delta^n U_a}{n!\,h^n}$

Pbm:1

If $U_{75} = 246$; $U_{80} = 202$; $U_{85} = 118$ and $U_{90} = 40$ find U_{79}.

Here, $a = 75$, $h = 5$. We have to find $U_{a+rh} = U_{79}$

$\therefore a + rh = 79$. Hence $75 + 5r = 79$

$\therefore r = 4/5 = 0.8$

By Newton-Gregory formula for equal intervals.

$$U_{a+rh} = U_a + \frac{r}{1!}\Delta U_a + \frac{r(r-1)}{2!}\Delta^2 U_0 + \cdots$$

$$\therefore U_{79} = 246 + \frac{0.8(-44)}{1} + \frac{0.8(0.8-1)}{1.2}(-40) + \frac{0.8(0.8-1)(0.8-2)}{1.2.3}(4$$

$= 246 - 35.2 + 3.2 + 1.472$

$= 215.472$

x	U_x	ΔU_x	$\Delta^2 U_x$	$\Delta^3 U_x$
75	246			
		-44		
80	202		-40	
				46
85	118	-84	6	
		-78		
90	40			

pbm:2

By using Gregory-Newton formula find U_x for following data. Hence estimate (i) $U_{1.5}$ (ii) U_9

U_0	U_1	U_2	U_3	U_4
1	11	21	28	29

x	u_x	Δu_x	$\Delta^2 u_x$	$\Delta^3 u_x$	$\Delta^4 u_x$
0	1				
		10			
1	11		0		
		10		-3	
2	21		-3		0
		7		-3	
3	28		-6		
		1			
4	29				

$$u_x = u_a + (x-a)\frac{\Delta u_a}{1!} + (x-a)(x-a-h)\frac{\Delta^2 u_0 h^2}{2!} + \dots$$

Here $a=0$ and $h=1$

$$\therefore u_x = 1 + (x-0)\times\frac{10}{1!} + x(x-1)\times\frac{0}{2!} + x(x-1)(x-2)\times\frac{(-3)}{3!}$$

$$= 1 + 10x - \frac{x(x-1)(x-2)}{2}$$

$$= \frac{1}{2}(2 + 20x - x^3 + 3x^2 - 2x)$$

$$\therefore u_x = \frac{1}{2}(-x^3 + 3x^2 + 18x + 2)$$

(i) $\therefore u_{1.5} = \frac{1}{2}[-(1.5)^3 + 3(1.5)^2 + 18(1.5) + 2]$

$$= \frac{1}{2}[-3.375 + 6.75 + 27 + 2]$$

$$= 16.188$$

(ii) $u_9 = \frac{1}{2}[-9^3 + 3\times 9^2 + 18\times 9 + 2]$

$$= \frac{1}{2}[-729 + 243 + 162 + 2]$$

$$= -161$$

Pbm: 3.

Population was recorded as follows in Village

Year	1941	1951	1961	1971	1981	1991
Population	2500	2800	3200	3700	4350	5225

Estimate the population for the year 1945

Year x	population U_x	ΔU_x	$\Delta^2 U_x$	$\Delta^3 U_x$	$\Delta^4 U_x$	$\Delta^5 U_x$
1941	2500					
		300				
1951	2800		100			
		400		0		
1961	3200		100		50	
		500		50		-25
1971	3700		150		25	
		650		75		
1981	4350		225			
		875				
1991	5225					

We have to find U_{1945}

$a = 1941$ and $h = 10$

$\therefore U_{a+rh} = U_{1945}$

Hence $1941 + 10r = 1945$

$r = 0.4$

By Newton Gregory formula we get.

$U_{1945} = 2500 + 0.4 \times \frac{300}{1} + 0.4(0.4-1) \times \frac{100}{2!} + 0.4(0.4-1)(0.4-2) +$

$(0.4)(0.4-1)(0.4-2)(0.4-3) \times \frac{50}{4!} + 0.4(0.4-1)(0.4-2)(0.4-3)(0.4-4)$

$\frac{-25}{}$

$= 2500 + 120 - 12 - 2.08 + 0.75$

$= 2606.67$

$= 2607$ appl..

Lagrange's formula:

Let $U_{a_1}, U_{a_2} \ldots U_{a_n}$ be the Value of U_x at $a_1, a_2 \ldots a_n$ (not necessarily at equal intervals) then an interpolating polynomial $\phi(x)$ for U_x is given by.

$$\phi(x) = \frac{(x-a_2)(x-a_3)\ldots(x-a_n)}{(a_1-a_2)(a_1-a_3)\ldots(a_1-a_n)} \times U_{a_1} + \frac{(x-a_1)(x-a_3)(x-a_4)\ldots(x-a_n)}{(a_2-a_1)(a_2-a_3)\ldots(a_2-a_n)} U_{a_2}$$

$$+ \ldots + \frac{(x-a_1)(x-a_2)\ldots(x-a_{n-1})}{(a_n-a_1)(a_n-a_2)\ldots(a_n-a_{n-1})} \times U_{a_n}$$

Pbm: 1

Find U_5 given that $U_1 = 4$; $U_2 = 7$; $U_4 = 13$ and $U_7 = 30$

The argument 1, 2, 4, 7 are not at equal interval.

Lagrange's formula to find U_5

Take $a_1 = 1$; $a_2 = 2$; $a_3 = 4$; $a_4 = 7$; and $x = 5$

$$U_5 = \left[\frac{(5-2)(5-4)(5-7)}{(1-2)(1-4)(1-7)}\right] \times 4 + \left[\frac{(5-1)(5-4)(5-7)}{(2-1)(2-4)(2-7)}\right] \times 7 +$$

$$\left[\frac{(5-1)(5-2)(5-7)}{(4-1)(4-2)(4-7)}\right] \times 13 + \left[\frac{(5-1)(5-2)(5-4)}{(7-1)(7-2)(7-4)}\right] \times 30$$

$$= \left[\frac{3 \times 1 \times (-2)}{(-1)(-3)(-6)}\right] \times 4 + \left[\frac{4 \times 1 \times (-2)}{1(-2)(-5)}\right] \times 7 + \left[\frac{4 \times 3(-2)}{3 \times 2(-3)}\right] \times 13 + \left[\frac{4 \times 3 \times 1}{6 \times 5 \times 3}\right] 30$$

$$= \frac{4}{3} - \frac{28}{5} + \frac{52}{3} + 4$$

$= 17.06$

Pbm: 2

Find form of the function U_x for the following data. Find U_3

x	0	1	2	5
U_x	2	3	12	147

Here $a_1 = 0;\ a_2 = 1;\ a_3 = 2;\ a_4 = 5$

$\therefore U_{a1} = 2;\ U_{a2} = 1;\ U_{a3} = 12;\ U_{a4} = 142$

Applying Lagrange's formula.

$$U_x = \left[\frac{(x-1)(x-2)(x-5)}{(0-1)(0-2)(0-5)}\right] \times 2 + \left[\frac{(x-0)(x-2)(x-5)}{(1-0)(1-2)(1-5)}\right] \times 1 + \left[\frac{(x-0)(x-1)(x-5)}{(2-0)(2-1)(2-5)}\right] \times 12 + \left[\frac{(x-0)(x-1)(x-2)}{(5-0)(5-1)(5-2)}\right] \times 147$$

$$= -\frac{(x^3-8x^2+17x-10)}{5} + \frac{3(x^3-7x^2+10x)}{4} - 2(x^3-6x^2+5x) + \frac{x^3-3x^2+2x}{60} \times 147$$

$$= \frac{1}{60}\left[x^3(-12+45-120+147) + x^2(96-315+720-441) + x(-204+450-600+294) + 120\right]$$

$$= \frac{1}{60}\left[60x^3 + 60x^2 - 60x + 120\right]$$

$$\therefore U_x = x^3 + x^2 - x + 2$$

$$\therefore U_3 = 3^3 + 3^2 - 3 + 2$$

$$= 35$$

Theory of Attributes

Defn:-

Suppose the population is divided into two classes according to the presence (or) absence of single attribute. The positive class denote the presence of attribute. The +ve class denote the presence of attribute and negative class denote the absence of the capital letter Such as A, B, C, D are used to denote +ve class and the Corresponding lower case Greek letter Such as $\alpha, \beta, \gamma, \delta, \ldots$ are used denote Negative classes.

Attributes	B	β
A	(AB)	$(A\beta)$
α	(αB)	$(\alpha\beta)$

Hence, N denote the total number in the population.

$$\therefore N = (A) + (\alpha) = (B) + (\beta)$$

from the table.

$$(A) = (AB) + (A\beta)$$
$$(B) = (AB) + (\alpha B)$$
$$(\alpha) = (\alpha B) + (\alpha\beta)$$
$$(\beta) = (A\beta) + (\alpha\beta)$$

Note:

$$\boxed{N = (A) + (\alpha)}$$
$$N = A.N + \alpha.N$$
$$N = (A+\alpha).N$$
$$1 = A+\alpha$$
$$\boxed{A+\alpha = 1}$$

11'y

$B+\beta=1$

$C+\gamma=1$

1) P.T $(AB) = (ABC)+(AB\gamma)$

$$(AB\gamma) = AB\gamma.N$$
$$= AB(1-c).N$$
$$= AB - ABc.N$$
$$= AB.N - ABC.N$$
$$(AB\gamma) = (AB)-(ABC)$$
$$-(AB) = -(ABC)-(AB\gamma)$$
$$(AB) = (ABC)+(AB\gamma)$$

2) Given $(A)=30$; $(B)=25$; $(\alpha)=30$; $(\alpha\beta)=20$ Find (i) N (ii) (β) (iii) (AB) (iv) $(A\beta)$ (v) (αB)

i) $N = (A)+(\alpha) = 30+30 = 60$

ii) $(\beta) = \beta.N = (1-\beta).N$
$$= N - B.N$$
$$= N-(B)$$
$$= 60-25$$
$$= 35$$

iii) $(AB) = AB.N$
$$= (1-\alpha)(1-\beta).N$$
$$= 1-\beta-\alpha+\alpha\beta.N$$
$$= N-(\beta)-(\alpha)+(\alpha\beta)$$
$$= 60-35-30+20$$
$$= 15$$

(iv) $(A\beta) = A\beta.N$
$$= A(1-B).N$$

$= A \cdot AB \cdot N$

$= (A) - (AB)$

$= 30 - 15$

$= 15$

v) $(\alpha\beta) = \alpha\beta \cdot N$

$= (1-A)(1-\beta) N$

$= 1 - \beta - A + A\beta$

$= N - (B) - (A) + (A\beta)$

$= 60 - 35 - 30 + 15$

$= 10$

3) given the following altimate the class frequency of two attitude A and B.

Find the frequencies of positive and Negative class and the total number of observation.

$(AB) = 975$, $(\alpha\beta) = 100$, $(A\beta) = 25$, $(\alpha\beta) = 950$

Positive.

$(A) = (AB) + (A\beta)$

$= 975 + 25 = 1000$

$(B) = (AB) + (\alpha B)$

$= 975 + 100$

$= 1075$

Negative

$(\alpha) = (\alpha B) + (\alpha\beta)$

$(\alpha) = 100 + 950 = 1050$

$(B) = (A\beta) + (\alpha\beta)$

$= 25 + 950 = 975$

Total frequency.

$$N = (A) + (\alpha)$$
$$= 1000 + 1050$$
$$= 2050$$

4) In a class test in which 135 candidates where examin, for proficiency in English (A) and Math (B) it was discovered that 75 student failed in English ($\alpha = 75$), 90 student failed in Math (β) and 50 (N) failed in both. How Many candidates.

(i) have passed in Math (B)

(ii) have passed in English, failed in math ($A\beta$)

(iii) have passed in both (AB)

Let A and B denote pass in English & math respectively and α, β denote fail in English and Math respectively.

Here, $N = 135$, $(\alpha) = 75$, $(\beta) = 90$, $(\alpha\beta) = 50$

(i) have passed in Math (B)

$$(B) = N - (\beta)$$
$$= 135 - 90$$
$$(B) = 45$$

(ii) $(\beta) = (A\beta) + (\alpha\beta)$

$$(A\beta) = (\beta) - (\alpha\beta)$$
$$= 90 - 50$$
$$= 40$$

iii) $AB \cdot N = (1-\alpha)(1-\beta) \cdot N$

$(AB) = 1-\beta-\alpha+\alpha\beta \cdot N$

$= N - (\beta) - (\alpha) + \alpha\beta$

$= 135 - 90 - 75 + 50$

$(AB) = 20$

5) Given that $(A) = (\alpha) = (B) = (B) = N/2$ S.T (i) $(AB) = (\alpha\beta)$

(ii) $(A\beta) = (\alpha B)$

$(AB) = AB \cdot N$

$= (1-\alpha)(1-\beta) \cdot N$

$= 1-\beta-\alpha+\alpha\beta \cdot N$

$= (N) - (\beta) - (\alpha) + (\alpha\beta)$

$= N - \frac{N}{2} - \frac{N}{2} + (\alpha\beta)$

$(AB) = (\alpha\beta)$

(ii) $(A\beta) = (\alpha\beta)$

$= (1-\alpha)(1-B) \cdot N$

$= 1 - B - \alpha + \alpha B \cdot N$

$= (N) - (B) - (\alpha) + (\alpha B)$

$= N - \frac{N}{2} - \frac{N}{2} + (\alpha B)$

$(AB) = (\alpha B)$

6) Of 500 (N) men in a locality exposed to cholera 172 in all were attached 178 Were inoculated and of these 128 Were attack (A = 172 B = 178)

find the number of person

(i) not inoculated not attacked (β α)

(ii) inoculated not attacked. (B)

(iii) not inoculated attacked. $(\alpha\beta)$

denoting the attribute A as attack and the attribute B as inoculated.

denoting the Attribute α as not attacked and attribute β as not inoculated.

(i) Here $N = 500$, $(A) = 172$, $(B) = 178$, $(AB) = 125$

$$= \alpha\beta . N$$

$$\alpha\beta = (1-A)(1-B) . N$$

$$= 1 - B - A + AB . N$$

$$= N - (B) - (A) + (AB)$$

$$= 500 - 178 - 172 + 128$$

$$\alpha\beta = 278$$

(ii) $(\alpha B) = \alpha B . N$

$$= (1-A)B . N$$

$$= B - AB . N$$

$$= (B) - (AB)$$

$$= 178 - 128$$

$$= 50$$

$$(A\beta) = A\beta . N$$

$$= A(1-\beta) . N$$

$$= A - AB . N$$

$$= (A) - (AB)$$

$$= 172 - 128$$

$$= 44$$

Consistency of Data

Defn: A set of class frequencies is said to be consistent if none of them is negative. Otherwise, the given set of class frequencies is said to be inconsistent.

pbm:1

Find whether the following data are consistent.

$N = 600,\ (A) = 300,\ (B) = 400\ \ (AB) = 50$

$$(\alpha\beta) = \alpha\beta \cdot N$$
$$= (1-A)(1-B)\cdot N$$
$$= 1 - B - A + AB \cdot N$$
$$= N - (B) - (A) + (AB)$$
$$= 600 - 400 - 300 + 50$$
$$\alpha\beta = -150 < 0 \qquad \therefore \alpha\beta < 0$$

$\therefore$ The given frequencies are inconsistent.

2) Of 2000 people consulted 1854 speak Tamil; 1507 speak Hindi, 572 speak English, 676 speak Tamil & Hindi. 286 speak Tamil & English, 270 speak Hindi & English, 114 speak Tamil, Hindi & English. S.T the information as it stands is incorrect.

Let A, B, C denote the attribute of speaking Tamil, Hindi, English respectively.

$N = 2000,\ (A) = 1854,\ (B) = 1507,\ (C) = 572,\ (AB) = 676,$

$(BC) = 270,\ (AC) = 286,\ (ABC) = 114$

$$(\alpha\beta\gamma) = \alpha\beta\gamma \cdot N$$
$$= (1-A)(1-B)(1-C)\cdot N$$
$$= (1 - B - A + AB)(1-C)\cdot N$$

$= 1 - C - B + BC - A + AC + AB - ABC \cdot N$

$= N - (C) - (B) + (BC) - (A) + (AC) + (AB) - (ABC)$

$= 2000 - 572 - 1507 + 270 - 1854 + 286 + 676 - 114$

$(\alpha\beta\gamma) = -815 < 0$

$\therefore \alpha\beta\gamma < 0$

∴ The given data is inconsistent.

5m

1) If $\frac{(A)}{N} = x$; $\frac{(B)}{N} = 2x$, $\frac{(C)}{N} = 3x$ and $\frac{(AB)}{N} = \frac{(BC)}{N} = \frac{(AC)}{N} = y$

P.T Neither x nor y can exceed $1/4$.

W.K.T x and y are obviously positive integers.

∴ The condition of consistency

$(AB) \le (A) \Rightarrow \frac{(AB)}{N} \le \frac{(A)}{N}$

$\boxed{y \le x}$ — (1)

Similarly, $y \le 2x$; $y \le 3x$

Now, $(AB) \ge (A) + (B) - N$

$\frac{(AB)}{N} \ge \frac{(A)}{N} + \frac{(B)}{N} - 1$

$y \ge x + 2x - 1$

$\boxed{y \ge 3x - 1}$

Similarly $y \ge 5x - 1$ — (2)

$y \ge 4x - 1$

From (1) and (2)

$5x - 1 \leq y \leq x$

$x \leq 1/4$

Taking $y \leq x \leq 1/4$

$\therefore$ Neither x nor y can exceed $1/4$

Independent and Association of data:

Defn: Two attribute A and B are said to be independent if there is Same proportion of A's amongest B's as amongest β's. otherwise, the proportion of B's amongst A's is the Same as amongst the α's.

Note:

(1) The given two attributes are independent if

Where $\delta = (AB) - \frac{(A)(B)}{N}$

(2) A and B are Independent if

$(AB)(\alpha\beta) - (A\beta)(\alpha B) = 0$

Association:

(i) If $(AB) \neq \frac{(A)(B)}{N}$ We Say that A and B are associated.

(ii) If $\delta > 0$ (or) $(AB) > \frac{(A)(B)}{N}$ We say that A and B are Positively associated.

(iii) If $\delta < 0$ (or) $(AB) < \frac{(A)(B)}{N}$, We Say that A and B are Negatively associated.

Coefficient of Association

(i) Yule's Coefficient of association.

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

(ii) Coefficient of Colligation

$$\gamma = \frac{\left[1 - \sqrt{\frac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}\right]}{\left[1 + \sqrt{\frac{(AB)(\alpha\beta)}{(AB)(\alpha\beta)}}\right]}$$

1) Check whether the attributes A and B are independent given that (i) $(A) = 30$, $(B) = 60$, $(AB) = 12$; $N = 150$.

(ii) $(AB) = 256$, $(\alpha B) = 768$, $(A\beta) = 48$, $(\alpha\beta) = 144$

Given that $(A) = 30$

$(B) = 60$

$(AB) = 12$

$(N) = 150$

$\therefore$ The given attributes are independent if

$$\delta = (AB) - \frac{(A)(B)}{N}$$

$$\delta = 12 - \left(\frac{30 \times 60}{150}\right)$$

$$\delta = 0$$

$\therefore$ given attributes are independent.

(iii) $(AB)(\alpha\beta) - (A\beta)(\alpha B) = (12)(144) - (48)(768)$

$= 0$

$\therefore$ given attributes are independent.

2) In a class test in which 135 candidates where examined for proficiency in physics and chemistry it was discovered that 75 students failed in physics. 90 student failed in chemistry and 50 failed both. Find the magnitude of association and state if there is any association between failing in physics and chemistry.

Let A and B denoting failed in physics and chemistry respectively.

$(N) = 135, \; (A) = 75, \; (B) = 90, \; (AB) = 50$

Magnitude of Association is given by.

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$(A) = (AB) + (A\beta)$

$\Rightarrow 75 = 50 + A\beta$

$\boxed{(A\beta) = 25}$

$(B) = (AB) + (\alpha B)$

$90 = 50 + (\alpha B)$

$(\alpha B) = 40$

$N = (B) + (\beta)$

$135 = 90 + (\beta)$

$(\beta) = 45$

$(\beta) = (A\beta) + (\alpha\beta)$

$45 = 25 + \alpha\beta$

$\boxed{\alpha\beta = 20}$

$$Q = \frac{(50)(20) - (25)(40)}{(50)(20) + (25)(40)}$$

$$= \frac{1000 - 1000}{1000 + 1000}$$

$Q = 0$

$\therefore$ The Attribute A and B are Independent.

Hence failure in physics and chemistry are completely Independent each other.

3) S.T attribute A and B are independent are not and Positively or Negatively associated in following cases.

(i) $N = 930$, $(A) = 300$, $(B) = 400$, $(AB) = 230$

$$\delta = (AB) - \frac{(A)(B)}{N}$$

$$= 230 - \frac{(300)(400)}{930}$$

$$= 230 - \frac{1200}{930}$$

$\delta = 100.97$.

Hence $\delta > 0$. The attribute A and B are +ve associated.

ii) $(AB) = 327$, $(A\beta) = 545$, $(\alpha B) = 741$, $(\alpha\beta) = 235$

The Coefficient of association Q is given by.

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(327)(235) - (545)(741)}{(327)(235) + (545)(741)}$$

$$= \frac{76845 - 403845}{76845 + 403845}$$

$$= -\frac{327000}{480690}$$

$$Q = -0.68 < 0$$

∴ The attribute are Negatively associated.

(iii) $(A) = 470$, $(AB) = 300$, $(\alpha) = 530$, $(\alpha B) = 150$

$$\delta = (AB) - \frac{(A)(B)}{N}$$

Here, $N = (A) + (\alpha)$

$= 470 + 530$

$N = 1000$

$(B) = (AB) + (\alpha B)$

$= 300 + 150$

$(B) = 450$

$$\delta = (AB) - \frac{(A)(B)}{N}$$

$$= (300) - \frac{(470)(450)}{1000}$$

$$= 300 - 211.5 = 88.5 > 0$$

Here $\delta > 0$. The attribute A and B are positively associated.

iv) $(AB) = 66$, $(A\beta) = 88$, $(\alpha B) = 102$, $(\alpha\beta) = 136$

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(66)(136) - (88)(102)}{(66)(136) + (88)(102)}$$

$$Q = 0$$

∴ The attribute are independent.

4) calculate the coefficient of association between intelligence of father and son from the following data.

(i) Intelligence father with intelligence son 200.

(ii) Intelligence father with dull sons 50.

(iii) dull father with intelligence son 110.

(iv) dull father with dull sons 600

Command on the result.

Let A and B denoting the intelligence of father and Son respectively

$(AB) = 200$, $(A\beta) = 50$, $(\alpha B) = 110$, $(\alpha\beta) = 600$

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(200)(600) - (50)(110)}{(200)(600) + (50)(110)}$$

$$= \frac{120,000 - 5500}{120000 + 5500}$$

$$= \frac{114,500}{125,500}$$

$$Q = 0.91 > 0$$

Hence, Q > 0. It means that intelligence father are likely to have intelligence sons.

5) Investigate from the following data between inoculation against small pox and prevention from attack.

~~Pass in P~~	(B) Attacked	(β) not attacked	total
(A) Inoculated	(A 25	220	245
(α) not Inoculated	90	160	250
total	115	380	495

Let A and B denoting inoculated and attack respectively.

Here $(AB) = 25$, $(A\beta) = 220$, $(\alpha B) = 90$, $(\alpha\beta) = 160$

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) - (A\beta)(\alpha B)}$$

$$= \frac{(25)(160) - (220)(90)}{(25)(160) - (220)(90)}$$

$$= \frac{4000 - 19800}{4000 + 19800}$$

$$= \frac{-15,800}{23,800}$$

$$Q = -0.66$$

$$\therefore Q < 0$$

The A and B are Negatively associated.

∴ Inoculation and attack from small pox are Negatively associated.

6) From the following data compare the association b/w Mark in physics and chemistry in MKU and MSU

University	MSU	MKU
Total.no.of. Candidate	200 N	1600
Pass in physics	80 (A)	320
Pass in chemistry	40 (B)	90
Pass in physics and chemistry	20 (AB)	30

Let A and B denoting pass in physics and chemistry for the following data. for MKU and MSU.

MSU	MKU
$N = 200$	$N = 1600$
$(A) = 80$	$(A) = 320$
$(B) = 40$	$(B) = 90$
$(AB) = 20$	$(AB) = 30$

MSU	MKU
$(A) = (AB) + (A\beta)$	$(A) = (AB) + (A\beta)$
$80 = 20 + (A\beta)$	$320 = 30 + (A\beta)$
$(A\beta) = 60$	$(A\beta) = 290$
$(B) = (AB) + (\alpha B)$	$(B) = (AB) + (\alpha B)$
$40 = 20 + (\alpha B)$	$90 = 30 + (\alpha B)$
$(\alpha B) = 20$	$(\alpha B) = 60$
$N = (A) + (\alpha)$	$N = (A) + (\alpha)$
$200 = 80 + (\alpha)$.	$1600 = 320 + (\alpha)$
$(\alpha) = 120$	$\alpha = 1280$
$(\alpha) = (\alpha B) + (\alpha\beta)$	$(\alpha) = (\alpha B) + (\alpha\beta)$
$120 = 20 + (\alpha B)$	$1280 = 60 + \alpha\beta$
$(\alpha\beta) = 100$	$(\alpha\beta) = 1220$

The Coefficient of association.

MSU	MKU
$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$	$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$
$= \frac{(20)(100) - (60)(20)}{(20)(100) + (60)(20)}$	$= \frac{(30)(1220) - (290)(60)}{(30)(1220) + (290)(60)}$
$= \frac{2000 - 1200}{2000 + 1200}$	$= \frac{36600 - 17400}{36600 + 17400}$
$= \frac{800}{3200}$	$= \frac{19200}{54000}$
$= 0.25$	$Q = 0.356$

Hence Q of MKU > Q of MSU.

Thus the association between the Knowledge in Physics and Chemistry. is greater than on MKU with MSU.

UNIT- V

Index Numbers!-

An index number is widely used satisfied device for comparing the level of a certain phenomenon with the level of the same phenomenon at some standard period.

For example!

We may whish to compare price of food article at a particular period with the price of the same article previous period of time.

Types!

Fixed base Method

In an index number in the base year used for comparison is kept constant through out then it is called fixed base Method..

Chain base Method

If for every year the previous year is used as a base for comparison than the Method is called chain base Method.

Types of index Number:

(i) Unweighted (or) Simplex index Numbers.

(ii) Weighted index Number.

(iii) Aggregate Method.

(iv) Average of price relatives Method.

iii) Aggregate Method:

$$P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

If p_0 denotes the price of the base year. P_1 denote the price of current year.

pbm:1 from the following data Construct Simple Aggregate index number for 1992.

Commodity	price in 1991	Price in 1992
Rice	7	8
Wheat	3.5	3.75
oil	40	45
Gas	78	35
floor	4.5	5.25

Commodity	P_0	P_1
Rice	7	8
Wheat	3.5	3.75
oil	40	45
Gas	78	35
floor	4.5	5.25
	$\Sigma p_0 = 133$	$\Sigma p_1 = 147$

$$P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

$$= \frac{147}{133} \times 100$$

$$P_{01} = 110.5$$

(IV) Average of price relatives Method (SIMPLEX INDEX METHOD)

Index number for current year is

$$P_{01} = \frac{p_1}{p_0} \times 100$$

(i) A.M Index Number

$$P_{01} = \frac{\Sigma (p_1/p_0) \times 100}{n}$$

(ii) G.M Index Number.

$$\log p_{01} = \frac{\Sigma (\log p_1/p_0 \times 100)}{n}$$

1) find the Index Number of price relatives being 1991 as base year using

(i) A.M (ii) G.M

Commodities	price in 1991	Price in 1992
Rice	7	8
Wheat	3.5	3.75
oil	40	45
Gas	78	85
flour	4.5	5.25

Commodities	(1991) p_0	(1992) p_1	$\frac{p_1}{p_0} \times 100$	$\log p_1/p_0 \times 100$
Rice	7	8	$8/7 \times 100 = 114.29$	2.058
Wheat	3.5	3.75	$\frac{3.75}{3.5} \times 100 = 107.14$	2.029
oil	40	45	$\frac{45}{40} \times 100 = 112.5$	2.051
Gas	78	85	$\frac{85}{78} \times 100 = 108.97$	2.037
Flour	4.5	5.25	$\frac{5.25}{4.5} \times 100 = 116.27$	2.067
			559.57	10.242

Here, $n = 5$

(i) A.M

$$P_{01} = \frac{\Sigma(p_1/p_0)\times 100}{n}$$

$$= \frac{559.61}{5}$$

$$P_{01} = 111.922$$

(ii) G.M

$$\log P_{01} = \frac{\Sigma(\log(p_1/p_0\times 100))}{n}$$

$$= 10.243/5$$

$$\log P_{01} = 2.0486$$

$$P_{01} = \text{antilog}(2.0486)$$

$$= 111.84$$

Hw

1) From the following data Construct the Simple Aggregate Index Number 1992

Commodity	Price in 1991 (p_0)	Price in 1992 (p_1)
Rice	50	70
Wheat	40	60
Oil	80	90
Gas	110	120
Flour	20	20
	300	360

$$P_{01} = \frac{\Sigma p_1}{\Sigma p_0}\times 100$$

$$= \frac{360}{300}\times 100$$

$$= 1.2\times 100$$

$$= 120$$

2) Find Index Number of price relatives taking 1990 as base year using (i) A.M (ii) G.M

Commodities	1990	1991
Rice	158	272
Corn	168	326
Cumbu	157	309
Ragi	155	304

Commodities	p_0 (1991)	p_1 (1991)	$\frac{p_1}{p_0} \times 100$	$\log p_1/p_0 \times 100$
Rice	158	272	$\frac{272}{158} \times 100 = 172.15$	2.235
Corn	168	326	$\frac{326}{168} \times 100 = 194.04$	2.287
Cumbu	157	309	$\frac{309}{157} \times 100 = 196.81$	2.294
Raagi	155	304	$\frac{304}{155} \times 100 = 196.12$	2.292
			759.12	9.108

Here, n = 4

(i) A.M

$$P_{01} = \frac{\Sigma (P_1/p_0) \times 100}{n}$$

$$= \frac{759.12}{4}$$

$$= 189.78$$

(ii) G.M

$$\log P_{01} = \frac{\Sigma (\log P_1/p_0 \times 100)}{n}$$

$$\log P_{01} = \frac{9.108}{4} = 2.277$$

$$P_{01} = \text{antilog} (2.277)$$

$$= 189.234$$

1) From the following data whole sale price of rice for five year construct index number taking (i) 1987 as base year (ii) 1990 as base year.

Years	1987	1988	1989	1990	1991	1992
Price of rice	5	6	6.50	7	7.50	8

(i) Construction of I.N taking 1987 as base

Years	Price of Rice	Index no 1987 as base
1987	(5)	5/5 × 100 = 100
1988	6	6/5 × 100 = 120
1989	6.50	6.50/5 × 100 = 130
1990	7	7/5 × 100 = 140
1991	7.50	7.50/5 × 100 = 150
1992	8	8/5 × 100 = 160.
		800

(ii) Construction of Index no. taking 1990 as base year.

Year	Price of rice	Index no (1990) base year
1987	5	5/7 × 100 = 71.4
1988	6	6/7 × 100 = 85.4
1989	6.50	6.50/7 × 100 = 92.8
1990	(7)	7/7 × 100 = 100
1991	7.50	7.50/7 × 100 = 107.8
1992	8	8/7 × 100 = 114.3
		571.3

2) For the data given below calculate Index Number takeing (i) 1984 as base year (ii) 1991 as base year.

Year	1984	1985	1986	1987	1988	1989	1990	1991	1992
Price of heat per kg	4	5	6	7	8	10	9	10	11

ii) Construction of I.N taking 1984 base year.

Years	Price of Rice	Index no (1984)
1984	4	4/4×100 = 100
1985	5	5/4×100 = 125
1986	6	6/4×100 = 150
1987	7	7/4×100 = 175
1988	8	8/4×100 = 200
1989	10	10/4×100 = 250
1990	9	9/4×100 = 225
1991	10	10/4×100 = 250
1992	11	11/4×100 275

Total index no for 1984 as base years 1750

(ii)

Years	Price of Rice	Index no (1991)
1984	4	4/10×100 = 40
1985	5	5/10×100 = 50
1986	6	6/10×100 = 60
1987	7	7/10×100 = 70
1988	8	8/10×100 = 80
1989	10	10/10×100 = 100
1990	9	9/10×100 = 90
1991	10	10/10×100 = 100
1992	11	11/10×100 = 110

Total index no for 1991 as base year 700

1) construct the Whole Sale price index Number for 1991 and 1992 from the following data using 1990 as base year.

Commodity	Rice	Wheat	Ragi	corn	flour	Rava
1990	700	540	300	250	320	325
1991	750	575	325	280	330	350
1992	825	600	310	295	335	360

Soln:

Commodity	(1990) P_0	(1991) P_1	(1992) P_2	IN 1990 as base Year	(1991) as base Year
Rice	700	750	825	$\frac{750}{700} \times 100 = 107.1$	$\frac{825}{700} \times 100 = 117.9$
Wheat	540	575	600	$\frac{575}{540} \times 100 = 106.5$	$\frac{600}{540} \times 100 = 111.1$
Ragi	300	325	310	$\frac{325}{300} \times 100 = 108.3$	$\frac{310}{300} \times 100 = 103.3$
Corn	250	280	295	$\frac{280}{250} \times 100 = 112$	$\frac{295}{250} \times 100 = 118$
flour	320	330	335	$\frac{330}{320} \times 100 = 103.1$	$\frac{335}{320} \times 100 = 104.7$
Rava	325	350	360	$\frac{350}{325} \times 100 = 107.7$	$\frac{360}{325} \times 100 = 110.8$
				644.7	665.8

The A.M of the index no in 1991 the A.M

$$P_{01} = \frac{\Sigma (P^1/P_0) \times 100}{n}$$

Here $n = 6$

$$= \frac{644.7}{6}$$ [∴ Index no 107.45]

$= 107.45$

In 1992 I.N

$$P_{01} = \frac{665.8}{6}$$

$$= 110.9$$

Index No is 110.9

1) From the following Average price of three group of Commodities are given find (i) Fixed based index no (ii) chain based index no. With 1988 as year.

Commodity	1988	1989	1990	1991	1992
A	2	3	4	5	6
B	8	10	12	15	18
C	4	5	8	10	12

Soln:

Fixed Based Index Number.

Com	1988 (p_0)	1989 (p_1)	1990 (p_1)	1991 (p_1)	1992 (p_1)
A	$2/2 \times 100 = 100$	$3/2 \times 100 = 150$	$4/2 \times 100 = 200$	$5/2 \times 100 = 250$	$6/2 \times 100 = 800$
B	$8/8 \times 100 = 100$	$\frac{10}{8} \times 100 = 125$	$12/8 \times 100 = 150$	$15/8 \times 100 = 187.5$	$18/8 \times 100 = 225$
C	$4/4 \times 100 = 100$	$5/4 \times 100 = 125$	$8/4 \times 100 = 200$	$10/4 \times 100 = 250$	$12/4 \times 100 = 300$
Total	300	400	550	687.5	825
I.N A.M	$300/3 = 100$	$400/3 = 133.3$	$550/3 = 183.3$	$\frac{687.5}{3} = 229.2$	$\frac{825}{3} = 275$

Chain Based Index Number.

Commodity	1988	1989	1990	1991	1992
A	2/2×100 =100	3/2×100 =150	4/3×100 =133.3	5/4×100 =125	6/5×100 =120
B	8/8×100 =100	10/8×100 =125	12/10×100 =120	15/12×100 =125	18/15×100 =120
C	4/4×100 =100	5/4×100 =125	8/5×100 =160	10/8×100 =125	12/10×100 =120
Total	300	400	413.3	375	360
I.N (A.M)	300/3 =100	400/3 =133.5	413.3/3 =137.5	375/3 =125	360/3 =120

2) from the following data to find i) fixed Based I.N ii) Chain Based I.N

com	1988	1989	1990	1991	1992
I	2	3	5	7	6
II	8	10	2	4	18
III	4	3	7	2	12

i) Fixed Based I.N

Com	1988	1989	1990	1991	1992
I	2/2×100 =100	3/2×100 =150	5/2×100 =250	7/2×100 =350	6/2×100 =300
II	8/8×100 =100	10/8×100 =125	12/8×100 =150	4/8×100 =50	18/8×100 =225
III	4/4×100 =100	3/4×100 =75	7/4×100 =175	9/4×100 =225	12/4×100 =300
Total	300	850	575	625	825
I.N (A.M)	100	350/3 =116.7	575/3 =191.7	625/3 =208.3	825/3 =275

ii) Chain Based I.N

Com	1988	1989	1990	1991	1992
I	$2/2 \times 100 = 100$	$3/2 \times 100 = 150$	$5/3 \times 100 = 166.7$	$7/5 \times 100 = 140$	$6/7 \times 100 = 85.7$
II	$8/8 \times 100 = 100$	$10/8 \times 100 = 125$	$12/10 \times 100 = 120$	$4/12 \times 100 = 33.3$	$18/4 \times 100 = 450$
III	$4/4 \times 100 = 100$	$3/4 \times 100 = 75$	$7/3 \times 100 = 233.3$	$9/7 \times 100 = 128.6$	$12/4 \times 100 = 133.3$
Total	300	350	520	301.9	669
I N (A.M)	$300/3 = 100$	$350/3 = 116.7$	$520/3 = 173.3$	$301.9/3 = 100.6$	$669/3 = 233$

Weighted Index Numbers

(i) Weighted Aggregative Method:

(a) Laspeyre's Index Number.

According to Laspeyre's Method the prices of the Commodities in the base Year as well as the current Year Know and there are Weighted by the Quantities used in the base year, Hence the Laspeyre's Index Number is defined by

$$L_{I01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

Where p_0, p_1 denote the price of base year and current Year respectively.

q_0, q_1 denote Quantities consumed in the base Year and current Year respectively.

b) paasche's Index Number.

According to paasche's Method current Year Quantities are taken as Weights and hence paasches Index Number is defined by

$$P_{I01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

(c) Marshell - Edgeworth Index Number.

According to this Method, the Weight is the Sum of the Quantities of the base ~~Year~~ period and current Period. hence the Marshell. Edgeworth index number. is defined by

$$M_{I01} = \frac{\Sigma p_1 (q_0+q_1)}{\Sigma p_0 (q_0+q_1)} \times 100$$

(d) Bowley's Index Number:

The A.M of Laspeyre's and paasche's Index Number is defined to be Bowley's Index Number.

$$B_{I01} = \frac{L_{I01} + P_{I01}}{2}$$

(e) Fisher's Index Number.

The G.M of Laspeyre's and paasche's Index Number is defined

$$I_{01} = \sqrt{L_{I01} \times P_{I01}}$$

(f) Kelley's Index Number:

According to kelley Weight May be taken as the Quantities of period which is not neccessarly the base Year as current Year.

The Average Quantity of two or More Year taken as Weight.

2)

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity.
A	2	8	4	6
B	5	10	6	5
C	4	14	5	10
D	2	19	2	12

Com	p_0	q_0	p_1	q_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
A	2	8	4	6	16	12	32	24
B	5	10	6	5	50	25	60	30
C	4	14	5	10	56	40	70	50
D	2	19	2	12	38	20	38	26
					160	103	200	130

i) Laspeyre's I.N

$$L_{I01} = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$$

$$= \frac{200}{160} \times 100$$

$$= 125$$

ii) Paasche's I.N

$$P_{I01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times 100$$

$$= \frac{130}{103} \times 100$$

$$= 126.21$$

iii) Marshell Edge worth

$$M_{I01} = \frac{\Sigma p_1(q_0+q_1)}{\Sigma p_0(q_0+q_1)} \times 100 = \frac{\Sigma p_1q_0 + \Sigma p_1q_1}{\Sigma p_0q_0 + \Sigma p_0q_1} \times 100$$

$$= \frac{200+130}{160+103} \times 100$$

$$= \frac{3300}{263} \times 100 = 125.48$$

iv) Bowley's I.N

$$B_{I01} = \frac{L_{I01} + P_{I01}}{2}$$

$$= \frac{125 + 126.21}{2}$$

$$= \frac{251.21}{2}$$

$$= 125.60$$

v) Fisher's I.N

$$I_{01} = \sqrt{L_{I01} \times P_{I01}}$$

$$= \sqrt{125 \times 126.21}$$

$$= \sqrt{15776.25}$$

$$= 125.60$$

3) calculate the different type of index number.

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	6	50	10	56
B	2	100	2	120
C	4	60	6	60
D	10	30	12	24
E	8	40	12	26

com	p_0	q_0	p_1	q_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
A	6	50	10	56	300	336	500	560
B	2	100	2	120	200	240	200	240
C	4	60	6	60	240	240	360	360
D	10	30	12	24	300	240	360	288
E	8	40	12	26	320	208	480	312
					1360	1264	1900	1760

i) Laspeyre's I.N is

$$L_{I01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

$$= \frac{1900}{139} \times 100$$

$$= 139.70$$

ii) Paasche's I.N

$$P_{I01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

$$= \frac{892.5}{727.5} \times 100$$

$$= 122.68$$

iii) Fisher's I.N

$$I_{01} = \sqrt{L_{I01} \times P_{I01}}$$

$$= \sqrt{122.5 \times 122.68}$$

$$= \sqrt{15028.3}$$

$$= 122.59$$

5) Compute Fisher's Index Number take 1980 base year

	Tomato		Brinjal		Onion	
	P	Q	P	Q	P	Q
1980	4	50	3	10	2	5
1990	10	40	8	5	4	4

Com	P_0	q_0	P_1	q_1	P_0q_0	P_0q_1	P_1q_0	P_1q_1
Tomato	4	50	10	40	200	~~4~~ 160	500	400
Brinjal	3	10	8	5	30	15	80	64 ~~400~~
Onion	2	5	4	4	10	8	20	16
					240	172	600	480

i) Laspeyre's I.N

$$L_{I01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

$$= \frac{600}{240} \times 100$$

$$= 250$$

ii) paasche's I.N

$$P_{I01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

$$= \frac{480}{192} \times 100$$

$$= 250$$

iii) Marshall - Edge worth I.N

$$M_{I01} = \frac{\Sigma p_1 (q_0+q_1)}{\Sigma p_0 (q_0+q_1)} \times 100$$

$$= \frac{600+480}{240+192}$$

$$= \frac{1080}{432} \times 100$$

$$= 250$$

iv) Bowley's I.N

$$B_{I01} = \frac{L_{I01} + P_{I01}}{2}$$

$$= \frac{250+250}{2} = \frac{500}{2}$$

$$= 250$$

v) Fishers I.N

$$I_{01} = \sqrt{L_{I01} \times P_{I01}}$$

$$= \sqrt{250 \times 250}$$

$$= 250$$

4) Calculate Fishers Index Number for the following data

5)

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	6	50	12	56
B	2	100	2	120
C	4	60	6	60
D	10	30	12	24

com	P_0	q_0	P_1	q_1	P_0q_0	P_0q_1	P_1q_0	P_1q_1
A	6	50	12	56	300	336	600	672
B	2	100	2	120	200	240	200	240
C	4	60	6	60	240	240	360	360
D	10	30	12	24	300	240	360	288
					1040	1056	1520	1560

$$LI_{01} = \frac{\Sigma P_1q_0}{\Sigma P_0q_0} \times 100$$

$$= \frac{1520}{1040} \times 100$$

$$= 146.15$$

$$PI_{01} = \frac{\Sigma P_1q_1}{\Sigma P_0q_1} \times 100$$

$$= \frac{1560}{1056} \times 100$$

$$= 147.73$$

Fisher's Index Number

$$I_{01} = \sqrt{LI_{01} \times PI_{01}}$$

$$= \sqrt{146.15 \times 147.73}$$

$$= \sqrt{21590.74}$$

$$= 146.94$$

3) Find Laspeyre's Index Number and paasche's Index Number for the following data.

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	1	10	1.5	8
B	5	12	6	10
C	8	5	10	2

com	p_0	q_0	p_1	q_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
A	1	10	1.5	8	10	8	15	12
B	5	12	6	10	60	50	72	60
C	8	5	10	2	40	16	50	20
					110	74	137	92

Laspeyre's Index Number

$$LI_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

$$= \frac{137}{110} \times 100$$

$$= \frac{13700}{110}$$

$$LI_{01} = 124.55$$

paasche's Index Number

$$PI_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

$$= \frac{92}{74} \times 100$$

$$= \frac{9200}{74}$$

$$= 124.32$$

1) Find the Missing price in the following data if the ratio b/w Laspreye's and paasche's Index Number is 25:24

Com	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	1	15	2	15
B	2	15	x	30

Com	Base Year		Current Year		p_0q_0	p_0q_1	p_1q_0	p_1q_1
	p_0	q_0	p_1	q_1				
A	1	15	2	15	15	15	30	30
B	2	15	x	30	30	60	$15x$	$30x$
					45	75	$30+15x$	$30+30x$

$$L_{I01} = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100 \qquad P_{I01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times 100$$

$$= \frac{30+15x}{45} \times 100 \qquad = \frac{30+30x}{75} \times 100$$

$$L_{I01} : P_{I01} = 25 : 24$$

$$\left(\frac{30+15x}{45} \times 100\right) : \left(\frac{30+30x}{75} \times 100\right) = 25 : 24$$

$$24\left(\frac{30+15x}{45}\right) = 25\left(\frac{30+30x}{75}\right)$$

$$\frac{720+360x}{45} = \frac{750+750x}{75}$$

$$16+8x = 10+10x$$

$$16+8x-10-10x = 0$$

$$6-2x = 0$$

$$x = 3$$

$\therefore$ The Missing price $x = 3$

2) Find the Missing price in following data of the ratio between Laspreye's and paasche's Index number 28 : 27.

Com	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	1	10	2	5
B	1	5	x	2

Com	Base Year		Current Year		p_0q_0	p_0q_1	p_1q_0	p_1q_1
	p_0	q_0	p_1	q_1				
A	1	10	2	5	10	5	20	10
B	1	5	x	2	5	2	$5x$	$2x$
					15	7	$20+5x$	$10+2x$

$$L_{I01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 \qquad P_{I01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

$$= \frac{20+5x}{15} \times 100 \qquad = \frac{10+2x}{7} \times 100$$

$$L_{I01} : P_{I01} = 28:27$$

$$\left(\frac{20+5x}{15}\right) : \left(\frac{10+2x}{7}\right) = 28:27$$

$$27\left(\frac{20+5x}{15}\right) = 28\left(\frac{10+2x}{7}\right)$$

$$\frac{540+135x}{15} = \frac{280+56x}{7}$$

$$36+9x = 40+8x$$

$$36-40+9x-8x=0$$

$$-4+x=0$$

$$x=4$$

The Missing price is 4.

Formula to find the cost of living index Number.

i) Aggregate expenditive Method:

cost of living index Number.

$$I_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 \quad \text{(Laspreye's Method)}$$

ii) Family Budget Method.

cost of living I.N $I_{01} = \frac{\Sigma PV}{V}$

Where $P = \frac{P_1}{P_0} \times 100$

V = Value of Weight $P_0 q_0$

$I_{01} = \frac{\Sigma PW}{W}$, Where, $P = \frac{P_1}{P_0} \times 100$

and W is Weight.

1) Find the cost of living index Number for 1992 on the base of 1991 on the basis from the following data. using (i) Family Budget Method (ii) Aggregate Expenditure Method.

Com	Price		Quantity 1992
	1991	1992	
Rice	7	7.5	6
Wheat	6	6.75	3.5
flour	5	5	0.5
oil	30	32	3
Sugar	8	8.5	1

i) Family Budget Method.

Com	1991 (P_0)	1992 (P_1)	q_0	$P = \frac{P_1}{P_0} \times 100$ (P)	$P_0 q_0$ (V)	PV
Rice	7	7.5	6	$\frac{7.5}{7} \times 100 = 107.1$	42	4498.2
Wheat	6	6.75	3.5	$\frac{6.75}{6} \times 100 = 112.5$	21	2362.5
flour	5	5	0.5	$5/5 \times 100 = 100$	2.5	250
oil	30	32	3	$32/30 \times 100 = 106.7$	9	960.3
Sugar	8	8.5	1	$\frac{8.5}{8} \times 100 = 106.3$	8	850.4
					163.5	17564.

cost of living index Number

$$= \frac{\Sigma Pv}{V}$$

$$= \frac{17564.1}{163.5} = 107.4$$

ii) Aggregative expenditure Method.

Com	P_0	P_1	q_0	P_1q_0	P_0q_0
Rice	7	7.5	6	45	42
Wheat	6	6.75	3.5	23.6	21
floor	5	5	0.5	2.5	2.5
oil	30	32	3	96	90
Sugar	8	8.5	1	8.5	8
				175.6	163.5

$$I_{01} = \frac{\Sigma P_1 P_0}{\Sigma p_0 q_0} \times 100$$

$$= \frac{175.6}{168.5} \times 100$$

$$= 107.4$$

2) An enquiry into the budget of the Middle class families in a city in India gave the following data.

Com	food	rent	clothing	fuel	mise
Weights	35	15	20	10	20
Price in 1991	1500	300	450	70	500
Price in 1992	1650	325	500	90	550

What changes in cost of living index of 1992 as Compared with that of 1991 ar Seen.

Com	1991 price (p_0)	Price in 1992 (p_1)	Weight (W)	$P = \frac{P_1}{P_0} \times 100$ (P)	$p_0 q_0$ (PW)
Food	1500	1650	35	$\frac{165000}{1500} = 110$	3850
Rent	300	325	15	$\frac{32500}{300} = 108.3$	1624.5
Clothing	450	500	20	$\frac{50000}{450} = 111.11$	2222.2
Fuel	70	90	10	$\frac{9000}{700} = 128.57$	1285.7
misc	500	550	20	$\frac{55000}{500} = 110$	2200
			100		11182.4

Cost of living index Number

$$I_{01} = \frac{\Sigma PW}{W}$$

$$= \frac{11182.4}{100}$$

$$= 111.82$$

Find the cost of living index for the following data in a Middle class family.

Items	Price		Weight
	1991	1992	
Food	700	850	40
clothing	300	280	15
Rent	200	225	7
Fuel	70	82	5
Medicine	100	135	9
Education	500	550	12
Entertainment	100	90	10
Misc	475	425	23

Items	Price in 1991 (P_0)	Price in 1992 (P_1)	W	$P = \frac{P_1}{P_0} \times 100$	$P_0 q_0$ PW
Food	700	850	40	$\frac{85000}{700} = 121.42$	48568
clothing	300	280	15	$\frac{28000}{300} = 93.33$	1399.95
Rent	200	225	7	$\frac{22500}{200} = 112.5$	787.5
Fuel	70	82	5	$\frac{8200}{70} = 117.4$	585.7
Medicine	100	135	9	$\frac{13500}{100} = 13.5$	~~585.7~~ 1215
Education	500	550	12	$\frac{55000}{500} = 110$	1320
Entertainment	100	90	10	$\frac{9000}{100} = 90$	900
Misc	475	425	23	$\frac{42500}{475} = 59.47$	2057.81
			121		13122.76

Cost of living index Number

$$I_{01} = \frac{\Sigma PW}{W}$$

$$= \frac{13122.76}{121}$$

$$I_{01} = 108.45$$

Calulate Index Number of price for 1992 on the base of 1990 from the data given below.

Commodities	Weights	Price per unit 1990 (P_0)	Price per unit 1992 (P_1)
A	40	80	86
B	25	60	55
C	5	345	50
D	20	35	40
E	10	25	20

Com	Price 1990 (p_0)	Price 1992 (p_1)	W	$P = \frac{P_1}{P_0} \times 100$	P_{090} (W)
A	80	85	40	$\frac{8500}{80} = 106.25$	4250
B	60	55	25	$\frac{5500}{60} = 91.67$	2291.75
C	345	50	5	$\frac{5000}{345} = 14.49$	72.45
D	35	40	20	$\frac{4000}{35} = 114.29$	2285.8
E	25	20	10	$\frac{2000}{25} = 80$	800
			100		9700

Cost of living index Number

$$I_{01} = \frac{\Sigma PW}{W}$$

$$= \frac{9700}{100}$$

$$= 97$$

Formula:

$$\text{Fixed base index of current Year} = \frac{\left(\text{chain base index of current year}\right) \times \left(\text{Fixed base index of previous Year}\right)}{100}$$

Hence chain base index Number of the

$$\text{current Year} = \frac{\left(\text{Fixed base index of current Year}\right) \times 100}{\left(\text{Fixed base index number of Previous Year}\right)}$$

From the fixed base Index Number given below prepare a Chain based index Number.

Year	1975	1976	1977	1978	1979	1980
Fixed based I.N	90	105	102	98	120	125

$$\text{Chain based Index Number for current Year} = \frac{\left(\text{Fixed based I.N of current Year}\right) \times 100}{\text{Fixed based I.N for Previous Year}}$$

chain based I.N

Year	Fixed based I.N	Chain based I.N
1975	90	$90/90 \times 100 = 100$
1976	105	$105/90 \times 100 = 116.7$
1977	102	$102/105 \times 100 = 97.19$
1978	98	$\frac{98}{102} \times 100 = 96.08$
1979	120	$\frac{120}{98} \times 100 = 122.45$
1980	125	$\frac{125}{120} \times 100 = 104.17$

2) From the Chain based I.N to prepare fixed based I.N

Year	1985	1986	1987	1988	1989	1990	1991
Chain base I.N	105	108	110	107	115	120	125

$$\text{Fixed Based I.N of the current Year} = \frac{\left(\text{Chain Based I.N of current Year}\right) \times \left(\text{Fixed based I.N of Previous Year}\right)}{100}$$

Fixed Based I.N

Year	Chain Based I.N of Current Year	Fixed Based I.N
1985	105	105
1986	108	$\frac{108 \times 105}{100} = 113.4$
1987	110	$\frac{110 \times 113.4}{100} = 124.7$
1988	107	$\frac{107 \times 124.7}{100} = 133.4$
1989	115	$\frac{115 \times 133.4}{100} = 153.4$
1990	120	$\frac{120 \times 153.4}{100} = 184.1$
1991	125	$\frac{125 \times 184.1}{100} = 230.1$

1) Convert the following chain based I.N into fixed based I.N

Years	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Chain Index	100	112.8	86.4	102.6	120.5	105.3	103.3	109.8	88.4	75.8

$$\text{Fixed base index Number} = \frac{\left(\text{Chain base index of current year}\right) \times \left(\text{Fixed base index of Previous Year}\right)}{100}$$

Years	Chain Based I.N of current year	Fixed Based I.N
1983	100	100
1984	112.8	$\frac{112.8 \times 100}{100} = 112.8$
1985	86.4	$\frac{86.4 \times 112.8}{100} = 97.5$
1986	102.6	$\frac{102.6 \times 97.5}{100} = 100.04$
1987	120.5	$\frac{120.5 \times 100.04}{100} = 120.5$
1988	105.3	$\frac{105.3 \times 120.5}{100} = 126.9$
1989	103.3	$\frac{103.3 \times 126.9}{100} = 131.09$
1990	109.8	$\frac{109.8 \times 131.09}{100} = 143.9$
1991	88.4	$\frac{143.9 \times 88.4}{100} = 127.2$
1992	75.8	96.4

2) Given the following chain base index number construct the fixed base index Number.

Years	1986	1987	1988	1989	1990
Chain Index	80	110	120	90	140

$$\text{Fixed base Index Number} = \frac{\left(\text{chain base index Current Year}\right) \times \left(\text{Fixed base index Previous year}\right)}{100}$$

Years	Chain Based I.N of current Year	Fixed Based I.N
1986	80	80
1987	110	$\frac{110\times80}{100}=88$
1988	120	$\frac{120\times88}{100}=105.6$
1989	90	$\frac{90\times105.6}{100}=95.04$
1990	140	133.05

3) From the F.B.I Number given below prepare chain base Index Number.

a)

Year	1987	1988	1989	1990	1991	1992
F.B.I (1987 as base)	94	98	102	95	98	100

b)

Year	1983	1984	1985	1986	1987	1988	1989
F.B.I (1983 as base)	100	112.8	97.4	100	120.5	126.9	131.07

$$C.B.I.N = \frac{\left(\text{Fixed Base I.N of Current Year}\right) \times 100}{\text{Fixed Base I.N for the Previous Year.}}$$

a) Chain Based I.N

Years	Fixed Base I.N	Chain Based I.N
1987	94	$\frac{94 \times 100}{94} = 100$
1988	98	$\frac{98 \times 100}{94} = 104.8$
1989	102	$\frac{10200}{98} = 104.08$
1990	95	$\frac{9500}{102} = 93.1$
1991	98	103.2
1992	100	102.04

b) Chain Based I.N

Years	Fixed Index	Chain based I.N
1983	100	$\frac{100 \times 100}{100} = 100$
1984	112.8	$\frac{112.8 \times 100}{100} = 112.8$
1985	97.4	$\frac{97.4 \times 100}{112.8} = 86.3$
1986	100	$\frac{100 \times 100}{97.4} = 102.7$
1987	120.5	$\frac{12050}{100} = 120.5$
1988	126.9	$\frac{126.5 \times 100}{120.5} = 104.97$
1989	131.08	$\frac{131.08 \times 100}{126.5} = 103.6$

Analysis of time Series

Time Series:

Time Series is a Series of Value of a Variable over a period of time arranged chronologically.

Example:

The time Series of retail prices of rice is the result of Combined influences of rainfall, availability of fertilisers, good yield, transport facilities, Consumer's demand and so on.

Components of a time Series:

The Various forces affecting the Value of a phenomenon in a time Series May be broadly classified into the following 3 categories generally Known as the Component of time Series.

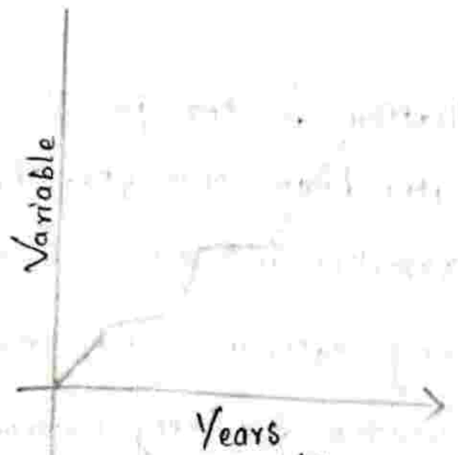
1. Long time trend (or) Secular trend.
2. Short term fluctuations (or) periodic Movements
3. Irregular fluctuations.

1) Long time trend:

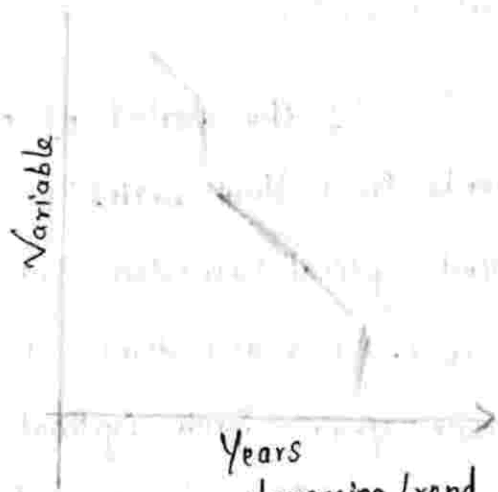
The general tendency of a time Series is to increase or decrease or stagnate over a period of several Years (long period). Such a long ran tendency of a time Series to increase or decrease over a period of time is Known as Secular trend (or) Simply trend.

Though the term 'long' is a relative item it depends upon the nature of the Series under consideration.

The long term trend does not Mean that the series should continuously move in one direction only. It is possible that different tendencies of increase and decrease persist together. A graphical representation indicating a long term increase or decrease or stability is given in the following figures.



Stability



decreasing trend

2) Short term fluctuations

In Most of the time Series a number of forces repeat themselves periodically over a period of time prevending the value of the Series of move in a particular direction. The Variations caused by such forces are called short term fluctuations.

(a) Seasonal ~~fluctua~~ Variation

(b) cyclical Variation

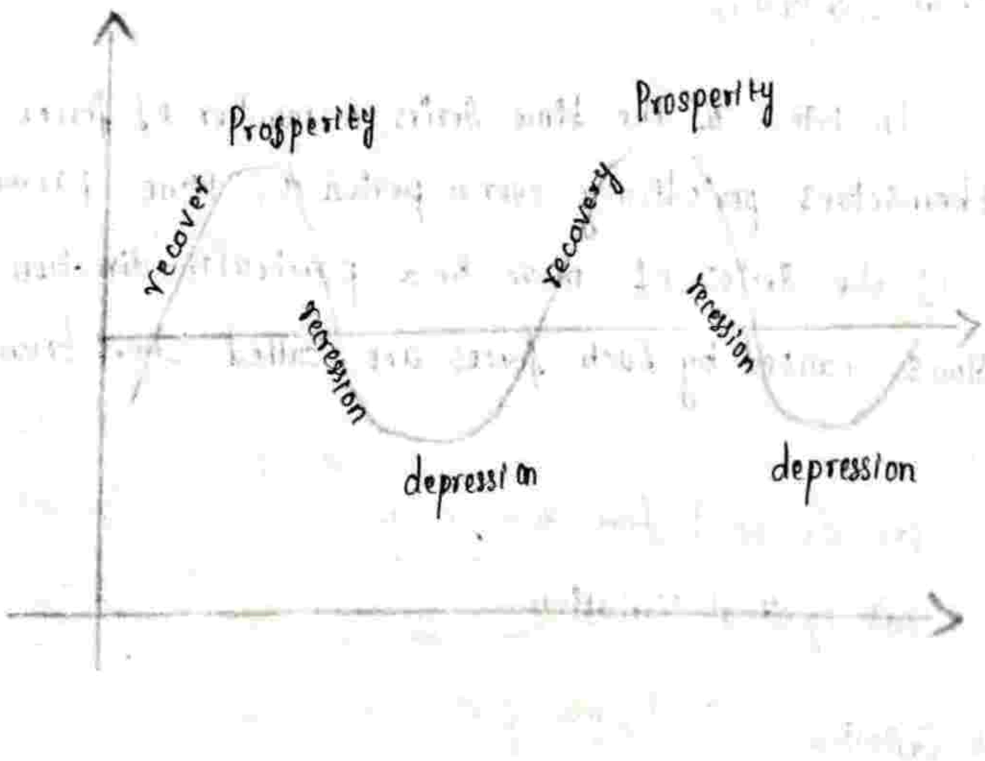
a) Seasonal Variation:

Seasonal Variation are rhythemic Short term fluctuation Which May occur due to changes in climatic factor, customery practices etc. Generally Seasonal Variations are Considered as Short term fluctuation that occur Within a Year. These influctuation May be regular as well as irregular Within a period of one Year.

For example, We find the prices of agricultural Commodities go down in the harvesting Season and slowly rise thereafter. Sales of umbrellas, rain coats, increase in the rainy season.

b) Cyclical Variation:

If the period of oscillation for the periodic Movements in a time Series is greater than one year then it is called cyclical variation. These Variation in a time Series are due to ups and down. recurring after a period greater than one year. Such cyclinal Variation are Very common in Many economic time Series and business Series.



The period between two Successive peaks thoughts is Known as the period of the cycle. In cyclical Variation generally the period of a cycle in three to eleven years.

Irregular fluctuations:

Other than the fluctuation discussed above there is one More random factor affecting the time Series. The fluctuation Which are purely random and due to unforseen and unpredictable forces are called irregular fluctuations.

Generally the Irregular fluctuation are caused by the unusual occurance of event like floods, Storms, famine, epidemics, wars, Strikes and lockout etc.

Measurement of Trends:

* A graphical representation of time Series exhibits the ~~graphical~~ general upward and downward tendencies.

* The following are the four Method Which are generally used for the Study of Measurement of trend in a time Series.

i) Graphic (Free hand curve fitting) Method
ii) Method of curve fitting by principal of least Square
iii) Method of Semi averages
iv) Method of Moving averages.

Measurement for Seasonal Variations:

There is a Simple Method for Measuring the Seasonal variation Which involves Simple averages.

Simple averages Method:

Step1: All the data are arranged by Years and Month (or) quartors.

Step2: Compute the Simple averages (A.M) $\overline{x_i}$ for ith Month.

Step3: obtain the over all averages $\overline{x}$ of those averages $\overline{x_i}$ and $\overline{x} = \dfrac{\overline{x_1} + \dots + \overline{x_{12}}}{12}$

Step4:

Seasonal indices for different Month s are Calulated by expressing Monthly averages as the Percentage of the overall average $\overline{x}$

Thus Seasonal index for ith Month $= \dfrac{\overline{x_i}}{\overline{x}} \times 100$

1) use the Method of least Square are fit a St.line trend to the following data given from 82 to 92. Hence estimate the trend Value for 1993.

Year	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Production in Quintals	45	46	44	47	42	41	39	42	45	40	48

Let the line of best fit be

$$Y = ax + b$$

Take $X = x - 1987$ (Mid value) and

$Y = y - 41$ (Mid value)

Then the line of best fit become

$$Y = aX + b$$

$$\Sigma Y_i = a\Sigma X_i + nb \quad \text{—— (1)}$$

$$\Sigma X_i Y_i = a\Sigma X_i^2 + b\Sigma X_i \quad \text{—— (2)}$$

$n = 11$

x	y	X_i (x - 1987)	Y_i (y-41)	X_i^2	X_iY_i
1982	45	-5	4	25	-20
1983	46	-4	5	16	-20
1984	44	-3	3	9	-9
1985	47	-2	6	4	-12
1986	42	-1	1	1	-1
1987	41	0	0	0	0
1988	39	1	-2	1	-2
1989	42	2	1	4	2
1990	45	3	4	9	12
1991	40	4	-1	16	-4
1992	48	5	7	25	35
			28	110	-19

From (1) $\Rightarrow$ $28 = 0 + 11b$

$11b = 28$

$b = 2.54$

From ② ⇒ $-19 = a(110) + 0$

$$a(110) = -19$$

$$a = -19/110$$

$$a = -0.17$$

Substitute a & b value for $y = ax+b$

$$\boxed{y = -0.17x + 2.54} \quad — ③$$

Substitute y & x value for ③ ⇒

$$y - 41 = (-0.17)(x - 1987) + 2.54$$

$$y - 41 = -0.17x + 337.79 + 2.54$$

$$y = -0.17x + 337.79 + 2.54 + 41$$

$$\boxed{y = -0.17x + 381.33} \quad — ④$$

Calculate the trend value:

Let $x = 1982$ in ④ ⇒

$$y = -0.17(1982) + 381.33$$

$$\boxed{y = 44.39}$$

Let $x = 1983$ in ④ ⇒

$$y = -0.17(1983) + 381.33$$

$$\boxed{y = 44.22}$$

Let $x = 1984$ in ④ ⇒

$$y = -0.17(1984) + 381.33$$

$$\boxed{y = ~~47.88~~ \; 44.05}$$

Let $x = 1985$ in ④ ⇒

$$y = -0.17(1985) + 381.33$$

$$\boxed{y = 43.88}$$

Let $x = 1986$ in ④ ⇒

$y = -0.17(1986) + 381.33$

$\boxed{y = 43.71}$

Let $x = 1987$ in ④

$y = -0.17(1987) + 381.33$

$\boxed{y = 43.54}$

Let $x = 1988$ in ④

$y = -0.17(1988) + 381.33$

$\boxed{y = 43.37}$

Let $x = 1989$ in ④

$y = -0.17(1988) + 381.33$

$\boxed{y = 43.2}$

Let $x = 1990$ in ④

$y = -0.17(1990) + 381.33$

$\boxed{y = 43.03}$

Let $x = 1991$ in ④

$y = -0.17(1991) + 381.33$

$\boxed{y = 42.86}$

Let $x = 1992$ in ④

$y = -0.17(1992) + 381.33$

$\boxed{y = 42.68}$

Let $y = 1993$ in ④

$y = -0.17(1993) + 381.33$

$\boxed{y = 42.52}$

∴ Thus the trend Value are 44.39, 44.22, 44.05, 43.88, 43.71, 43.54, 43.87, 43.2, 43.03, 42.86, 42.62

∴ Hence estimate the trend Value for 1993 is 42.52

Formula:

Seasonal index for i^{th} term $= \frac{\bar{x_i}}{\bar{x}} \times 100$

Q) Compute the Seasonal Indices for the following data by Simple average Method.

Season	1990	1991	1992	1993	1994
Summer	68	70	68	65	60
Monsoon	60	58	63	56	55
Autumn	61	56	68	56	55
Winter	63	60	67	55	18

Seasonal Indices for i^{th} term $= \frac{\bar{x_i}}{\bar{x}} \times 100$

Year	Summer	Monsoon	Autumn	Winter
1990	68	60	61	63
1991	70	58	56	60
1992	68	63	68	67
1993	65	56	56	55
1994	60	55	55	58
Total	331	292	296	303

Average $\bar{x}$ $\frac{331}{5} = 66.2$ $292/5 = 58.4$ $\frac{296}{5} = 59.2$ $303/5 = 60.6$ $\bar{x} = \frac{244.4}{4}$

$\frac{\bar{x_i}}{\bar{x}} \times 100$ $\frac{66.2}{61.1} \times 100 = 108.3$ $\frac{58.4}{61.1} \times 100 = 95.6$ $\frac{59.2}{61.1} \times 100 = 96.9$ $\frac{60.6}{61.1} \times 100 = 99.2$ $\bar{x} = 61.1$

1) Compute the Seasonal index for the following data assuming that there is no need to adjust the data for the trend.

Quarter	1989	1990	1991	1992	1993	1994
I	3.5	3.5	3.5	4.0	4.1	4.2
II	3.9	4.1	3.8	4.6	4.4	4.6
III	3.4	3.7	3.7	3.8	4.2	4.3
IV	3.6	4.8	4.0	4.5	4.5	4.7

Season Indices for pth term = $\frac{\bar{x_i}}{\bar{x}} \times 100$

Year	I	II	III	IV	
1989	3.5	3.9	3.4	3.6	
1990	3.5	4.1	3.7	4.8	
1991	3.5	3.9	3.7	4.0	
1992	4.0	4.6	3.8	4.5	
1993	4.1	4.4	4.2	4.5	
1994	4.2	4.6	4.3	4.7	
Total	22.8	25.5	23.1	26.1	
Average $\bar{x_i}$	$\frac{22.8}{6}$ = 3.8	$\frac{25.5}{6}$ = 4.25	$\frac{23.1}{6}$ = 3.85	$\frac{26.1}{6}$ = 4.35	$\bar{x} = \frac{16.25}{4}$ = 4.06
$\frac{\bar{x_i}}{\bar{x}} \times 100$	$\frac{3.8}{4.06} \times 100$ = 93.59	$\frac{4.25}{4.06} \times 100$ = 104.67	$\frac{3.85}{4.06} \times 100$ = 94.58	$\frac{4.35}{4.06} \times 100$ = 107.14	

2) Calculate (i) Three Yearly Moving average (ii) short time fluctuation for the data given problem.

Year	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
In Quintals	45	46	44	47	42	41	39	42	45	40	48

I Ye	II Production in Quintals	III 3 Yearly Moving total	IV 3yrly Moving average	Short term fluctuation
1982	45			
1983	46	135	45	1
1984	44	137	45.7	-1.7
1985	47	133	44.3	2.7
1986	42	130	43.3	-1.3
1987	41	~~122~~	~~40.7~~	~~0.3~~
1988	39	122	40.7	-1.7
1989	42	126	42	0
1990	45	127	42.3	2.7
1991	40	133	44.3	-4.3
1992	48			

Trend the Value for the given time Series are given in Column IV

Short term fluctuation are given in the last column.

F–2476

Sub. Code
7BMA5C2

B.Sc. DEGREE EXAMINATION, NOVEMBER 2019

Fifth Semester

Mathematics

STATISTICS — I

(CBCS 2017 onwards)

Time : 3 Hours　　　　Maximum : 75 Marks

Part A　　　　$(10 \times 2 = 20)$

Answer all questions.

1. Write the formula for standard deviation.

   திட்ட விலக்கம் சூத்திரம் எழுதுக.

2. What is co-efficient of variation

   மாறுபாட்டின் கெழு என்றால் என்ன?

3. Write μ_3 in terms of μ_i'.

   μ_3ஐ μ_i' எழுது.

4. Define Co-efficient of skewness.

   கோட்டக் கெழுவை வரையறு.

5. Between what values the correlation lie?

   எந்த மதிப்புகளுக்குள் மாறுபாட்டின் கெழு இருக்கும்?

6. What is a regression?

   தொடர்பு என்றால் என்ன?

7. Obtain Newtons forward difference formula.

நியூட்டனின் முன் வித்தியாச சூத்திரத்தை எழுது.

8. Obtain Lagranges interpolation formula.

லக்ராஞ்சியின் இடைச்செருகல் சூத்திரத்தை எழுது.

9. Give two uses of index numbers.

குறியீட்டெண்களின் இரண்டு பயன்களை எழுதுக.

10. Give the components of time series.

நேர தொடரின் கூறுகளை எழுது.

Part B

(5 × 5 = 25)

Answer all questions, choosing either (a) or (b).

11. (a) Find Harmonic mean.

x:	10	20	30	40	50
f:	20	30	50	15	5

இசைச் சராசரி காண்க.

x:	10	20	30	40	50
f:	20	30	50	15	5

Or

(b) Find the co-efficient of quartile deviation.

x: 20, 28, 40, 12, 30, 15, 50

கால்மான விலக்க கெழுவை காண்க.

x: 20, 28, 40, 12, 30, 15, 50

12. (a) Find μ_1 and μ_2.

x:	80	100	120	140	160
f:	8	11	18	9	4

μ_1 மற்றும் μ_2 காண்க.

x:	80	100	120	140	160
f:	8	11	18	9	4

Or

(b) Explain Curve fitting by least squares.

குறைந்த வர்க்கத்தில் வளைகோட்டை பொருத்துதலை விளக்குக.

13. (a) Find the two regression co-efficients.

x:	48	35	17	23	47
y:	45	20	40	25	45

இரு தொடர்பு கெழுக்கள் காண்க.

x:	48	35	17	23	47
y:	45	20	40	25	45

Or

(b) Explain how do you find correlation co-efficient for a bivariate distribution.

ஈருறுப்பு பரவலுக்கு மாறுபாட்டின் கெழு எவ்வாறு காண்பாய் என விளக்குக.

14. (a) Prove the following :

(i) $\Delta = E - 1$

(ii) $\mu = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right)$

(iii) $\mu^2 = (\delta^2 + 4)/4$

பின்வருவனவற்றை நிறுவுக.

(i) $\Delta = E - 1$

(ii) $\mu = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right)$

(iii) $\mu^2 = (\delta^2 + 4)/4$

Or

(b) Find y(2).

x :	0	1	3	4
$y(x)$:	–12	0	12	24

$y(2)$ காண்க.

x :	0	1	3	4
$y(x)$:	–12	0	12	24

15. (a) Find consumer price index number.

Commodity	Price (1980)	Price (1981)	Weightage %
A	400	550	35
B	250	300	25
C	500	600	15
D	200	350	20
E	150	225	5

நுகர்வோர் விலை குறியீட்டு எண்ணை காண்க.

பொருள்	விலை (1980)	விலை (1981)	(எடை விகிதம்) %
A	400	550	35
B	250	300	25
C	500	600	15
D	200	350	20
E	150	225	5

Or

(b) Explain measurement of trends.

போக்குகளை அளவிடுதல் பற்றி விளக்குக.

Part C (3 × 10 = 30)

Answer any **three** questions.

16. Find which Batsman A or B is consistent.

A :	32	28	47	63	71	39	10	60	96	14
B :	19	31	48	53	67	90	10	62	40	80

எந்த மட்டையாளர் A அல்லது B பொருத்தமுடையவர் என்று காண்க.

A :	32	28	47	63	71	39	10	60	96	14
B :	19	31	48	53	67	90	10	62	40	80

17. Find Karl Pearson co-efficient of skewness :

x :	70-80	60-70	50-60	40-50	30-40	20-30
f :	11	22	30	35	21	11

கார்ல் பியர்சன் கோட்டக் கெழுவை காண்க.

x :	70-80	60-70	50-60	40-50	30-40	20-30
f :	11	22	30	35	21	11

18. Find the rank correlation co-efficient.

Mark 1 :	20	22	28	23	30	30	23	24
Mark 2 :	28	24	24	25	26	27	32	30

தர வரிசையின் கெழுவை காண்க.

மதிப்பெண் 1 :	20	22	28	23	30	30	23	24
மதிப்பெண் 2 :	28	24	24	25	26	27	32	30

19. Estimate the value of y when $x = 22$.

x :	20	25	30	35	40
y :	73	198	573	1198	1450

$x = 22$ எனும்போது y ன் மதிப்பைக் காண்க.

x :	20	25	30	35	40
y :	73	198	573	1198	1450

20. Construct Fishers Ideal Index no.

Items	p_0	q_0	p_1	q_1
A	10	40	12	45
B	11	50	11	52
C	14	30	17	30
D	8	28	10	29
E	2	15	13	20

ஃபிஷரின் மாதிரி குறியீட்டெண்ணை அமைக்க.

பொருட்கள்	p_0	q_0	p_1	q_1
A	10	40	12	45
B	11	50	11	52
C	14	30	17	30
D	8	28	10	29
E	2	15	13	20

———————