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

ook for Reference:

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

♣♣♣♣♣♣♣♣♣♣♣♣

Dr. V. CHITRA
Dept. of Maths

3.8.2020

STATISTICS - I

UNIT - I

The statistical constants that describe any given group of data are of four types. They are

(i) Measure of central tendency
(ii) Measure of dispersion
(iii) Measure of Skewness
(iv) Measure of kurtosis.

Measure of Central Tendencies :-

They are 5 types.

(i) Arithmetic Mean (or) Mean
(ii) Median
(iii) Mode
(iv) Geometric Mean
(v) Harmonic Mean

Arithmetic Mean

It is defined by

(1) $$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\Sigma x_i}{n}.$$

(2) Suppose frequencies are given. Then

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}, \quad i = 1, 2 \ldots n$$

3) Weighted average is $\bar{x}_w = \frac{\Sigma w_i x_i}{\Sigma w_i}$.

$i = 1, 2, \ldots n.$

Problems

1. The heights of 10 students in Om's class chosen at random are given by 164, 159, 162, 168, 168, 170, 168, 171, 154, 167. Calculate A.M.

Soln:- $n = 10$ (Given data)

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{164+159+162+\cdots\cdots+167}{10}$$

$$= 169$$

2. Calculate A.M from the following frequency table.

Weight in Kgs	50	48	46	44	42	40
No. of persons	12	14	16	13	11	09

Soln:-

x_i	f_i	$f_i x_i$
50	12	600
48	14	672
46	16	736
44	13	572
42	11	462
40	09	360
Total	75	3402

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

$$= \frac{3402}{75}$$

$$= 45.36$$

3. Calculate A.M for the following frequency distribution of the marks obtained by 50 students in a class.

Marks	No. of students
5-10	5
10-15	6
15-20	15
20-25	10
25-30	5
30-35	4
35-40	3
40-45	2

Solu:-

$h =$ interval $= 5$

choose $A = 22.5$ (origin)

New $x_i = \frac{\text{old } x_i - A}{h} = u_i$

class	Mid x_i	u_i	f_i	$u_i f_i$
5–10	7.5	−3	5	−15
10–15	12.5	−2	6	−12
15–20	17.5	−1	15	−15
20–25	22.5	0	10	0
25–30	27.5	1	5	5
30–35	32.5	2	4	8
35–40	37.5	3	3	9
40–45	42.5	4	2	8
			50	−12

$\therefore \bar{x} = A + h\bar{u}$

$= 22.5 + 5\left[\frac{\Sigma u_i f_i}{\Sigma f_i}\right]$

$= 22.5 + 5\left[\frac{-12}{50}\right]$

$= 22.5 - 1.2$

$= 21.3$

4.8.2020

4. Find the mean mark of students from the follo... table

Marks	No. of students
0 and above	80
10 and above	77
20 and above	72
30 and above	65
40 and above	55
50 and above	43
60 and above	28
70 and above	16
80 and above	10
90 and above	8
[illegible] and above	9

Solu

Marks	Mid x_i	No. of students f_i	$f_i x_i$
0-10	5	3	15
10-20	15	5	75
20-30	25	7	175
30-40	35	10	360
40-50	45	12	540
50-60	55	15	825
60-70	65	12	780
70-80	75	6	450
80-90	85	2	170
90-100	95	8	760
100	—	0	—
Take $N = \Sigma f_i$		80	4140

$$\therefore \bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{4140}{80} = 51.75$$

5. Calculate (i) Mean price (ii) weighted average price of the following food articles from the value given below.

Article of food	Quantity in kgs	price per kg
Rice	30	4.50
wheat	10	2.75
Sugar	5.5	6.25
oil	3.5	16.50
Flour	4.5	4.00
Ghee	1.5	40.00
Onion	9	3.25

Solu:-

Article of food	Price per kg x_i	Quantity w_i	$w_i x_i$
Rice	4.5	30	135.00
wheat	2.75	10	27.50
Sugar	6.25	5.5	34.38
oil	16.50	3.5	57.75
Flour	4.00	4.5	18.00
Ghee	40.00	1.5	60.00
Onion	3.25	9	29.25
	77.25	64	361.88

(i) mean price $= \frac{\Sigma x_i}{n} = \frac{77.25}{7} = 11.04$ (5)

(ii) weighted average price $= \frac{\Sigma w_i x_i}{\Sigma w_i} = \frac{361.88}{64} = 5.65$

6. The four parts of a distribution are as follows.

	Frequency	Mean
Part 1	50	61
Part 2	100	70
Part 3	120	80
Part 4	30	83

Find the mean of the entire distribution.

→ Soln:- Given $n_1 = 50$ $n_2 = 100$ $n_3 = 120$ $n_4 = 30$

$\bar{x}_1 = 61$ $\bar{x}_2 = 70$ $\bar{x}_3 = 80$ $\bar{x}_4 = 83$

Mean $\boxed{\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + n_4\bar{x}_4}{n_1+n_2+n_3+n_4}}$ ✓

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

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

7. Mean weight of 80 students in two classes A and B is 50 kgs. There are 45 students in class A. The mean weight of the students in class B is 48. Find the mean weight of the students in class A.

→ Soln:- Given $n_1 = 45$ $n_2 = 80 - 45 = 35$

$\bar{x} = 50$ $\bar{x}_2 = 48$.

To find $\bar{x}_1$ from the formula $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$

$$\Rightarrow 50 = \frac{(45\times\bar{x}_1) + (35\times 48)}{45+35}$$

$$45\bar{x}_1 = 2320$$

$$\bar{x}_1 = \frac{2320}{45} = 51.56 \text{ kgs.}$$

∴ Mean weight of the students in class A = 51.56 kgs

8. S.T (i) A.M of the first n natural numbers is $\frac{1}{2}(n+1)$

(ii) the weighted A.M of first n natural numbers whose weights are equal to the corresponding numbers is equal to $\frac{1}{3}(2n+1)$

→

(i) A.M. of first 'n' natural numbers $= \frac{\Sigma x_i}{n}$

$$= \frac{1+2+\cdots+n}{n}$$

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

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

(ii) The required weighted A.M $= \frac{\Sigma w_i x_i}{\Sigma w_i}$

$$= \frac{1^2+2^2+\cdots+n^2}{1+2+\cdots+n}$$

$$= \frac{n(n+1)(2n+1)}{6} \Big/ \frac{n(n+1)}{2}$$

$$= \frac{\cancel{n}(\cancel{n+1})(2n+1)}{\cancel{6}\,3} \times \frac{\cancel{2}}{\cancel{n}(\cancel{n+1})}$$

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

9. The mean of 20 numbers is 50. By mistake marks of two students were taken as 64 and 67 instead of 46 and 76. Find the correct mean.

→

<u>Soln</u>:-

Total marks of the students $= 20 \times 50$

$= 1000$

Total marks after correction $= 1000 - (64 - 46) - (67 - 76)$

$= 1000 - 18 + 9$

$= 1000 - 9$

$= 991$

After correction, the average $= \frac{991}{20} = 49.55$

Median

Median is the value of the variate for which the cumulative frequency is $\frac{N}{2}$, where N is the total frequency.

1. Quartiles

First Quartile (lower quartile) is $\frac{N}{4}$ and it is denoted by Q_1

Second Quartile is Median. and it is denoted by Q_2.

Third Quartile (upper quartile) is $\frac{3N}{4}$ and it is denoted by Q_3.

Also, for ungrouped data with n items,

$$Q_1 = l + \frac{\left(\frac{N}{4} - m\right)h}{f_K} \qquad Q_3 = l + \frac{\left(\frac{3N}{4} - m\right)h}{f_K}$$

where l is lower limit of class in which particular quartile lies,

f_K is frequency of this class

h is width of the class

m is the cumulative frequency of the preceeding class.

2. Decile

$$D_i = l + \frac{\left(\frac{iN}{10} - m\right)h}{f_K}, \quad i = 1, 2, \dots, 9$$

3. Percentile

$$P_i = l + \frac{\left(\frac{iN}{100} - m\right)h}{f_K}, \quad i = 1, 2, \dots, 99$$

Problems

1. Find the median and quartiles of the heights in c.m of 11 students given by 66, 65, 64, 70, 61, 60, 56, 63, 60, 67, 62.

→ Arrange the given data in ascending order

56, 60, 60, 61, 62, 63, 64, 65, 66, 67, 70

Given $n = 11$, n is odd, median is sixth item

∴ Median = 63 (sixth item).

Q_1 = size of $\frac{1}{4}(n+1)^{th}$ item = size of $\frac{1}{4}(11+1)^{th}$ item
= 3^{rd} item
= 60.

$Q_3 = \frac{3}{4}(n+1)^{th}$ item $= \frac{3}{4}(11+1)^{th}$ item $= 9^{th}$ item $= 66$.

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

→

Soln:- Arranging in ascending order

33, 40, 43, 50, 61, 68, 72, 75, 84, 90

Here $n = 10$ (even)

Median is average of two middle items 61 and 68

$\therefore$ Median $= \frac{1}{2}(61+68) = 64.5$

First Quartile

Let $i = \frac{1}{4}(n+1)$ = integral part of $\frac{1}{4}(10+1) = 2$

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

$= 0.75 = \frac{3}{4}$ $\boxed{\frac{11}{4} = 2\frac{3}{4}}$

for grouped data

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

$= x_2 + 0.75\,(x_{2+1} - x_2)$

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

$= 42.5$

Third Quartile

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

$= \frac{3}{4}(11) = \frac{33}{4} = 8\frac{1}{4}$

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

$= 0.25 = \frac{1}{4}$

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

$= x_8 + (0.25)(x_9 - x_8)$

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

$= 77.25.$

3. Find the (i) mean (ii) median (iii) Q_1 (iv) Q_3 (v) 9th decile (vi) 19th percentile for the following frequency distribution.

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

→

choose $A = 33$ $h = 5$

New $x_i = \dfrac{\text{old } x_i - A}{h} = \dfrac{\text{old } x_i - 33}{5}$

class	mid x_i	f	u_i	$f_i u_i$	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
25.5 – 30.5	28	47	−1	−47	109 ←
30.5 – 35.5 ←	33	52 ←	0	0	161
35.5 – 40.5	38	41	1	41	202 ←
40.5 – 45.5	43	28	2	56	230
45.5 – 50.5	48	16	3	48	246
50.5 – 55.5	53	4	4	16	250
		250		−41	

(i) Mean $= \bar{x} = A + h\bar{u}$ where $\bar{u} = \dfrac{\sum f_i u_i}{\sum f_i}$

$$= 33 + 5\left[\frac{-41}{250}\right] = 32.18$$

(ii) Median :-

$$\frac{N}{2} = \frac{250}{2} = 125$$

∴ Median class is 30.5 – 35.5

$l = 30.5$ $m = 109$ $f_K = 52$

$$\text{Median} = l + \frac{\left(\frac{N}{2} - m\right)h}{f_k} = 30.5 + \frac{(125-109)5}{52}$$

$$= 30.5 + \frac{80}{52} = 32.04$$

(iii) First Quartile :-

$$\frac{N}{4} = \frac{250}{4} = 62.5 \quad \text{class is } 25.5-30.5$$

∴ $l = 25.5$ $m = 62$ $f_K = 47$

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

(iv) Third Quartile :-

$$\frac{3N}{4} = \frac{3(250)}{4} = 187.5$$

3rd Quartile class is 35.5 – 40.5

$l = 35.5$ $m = 161$ $f_K = 41$

$$Q_3 = 35.5 + \left(\frac{187.5 - 161}{41}\right)5 = 38.73$$

(iv) 9th decile :-

$$\frac{9}{10}N = \frac{9}{10}(250) = 225$$

9th decile class is 40.5 – 45.5

$\therefore$ $l = 40.5$ $m = 202$ $f_k = 28$

$$\therefore D_9 = 40.5 + \left(\frac{225-202}{28}\right)5$$

$$= 44.61$$

(vi) 19th percentile :-

$$\frac{19}{100}N = \frac{19}{100}(250) = 47.5$$

$\therefore$ 19th percentile class is 20.5 – 25.5

$\therefore$ $l = 20.5$ $m = 23$ $f_k = 39$

$$\therefore P_{19} = 20.5 + \left(\frac{47.5-23}{39}\right)5$$

$$= 23.64$$

4. 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
120 – 140	85
140 – 160	88
160 – 180	75
180 – 200	18

Soln :- (i) Number of tenants paying more than 105 is

$$= \left(\frac{120-105}{20}\right) \times 45 + 85 + 88 + 75 + 18$$

$$= 34 + 266 = 300.$$

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

(ii) No. of tenants paying the rent b/w 130 and 190

$$= \left(\frac{140-130}{20}\right) \times 85 + 88 + 75 + \left(\frac{190-180}{20}\right) \times 18$$

$$= 42.5 + 88 + 75 + 9 = 215$$

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

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.

For discrete frequency distribution, mode is the value of the variate corresponding to the maximum frequency.

$$\boxed{\text{Mode} = l + \frac{hf_2}{f_1+f_2}}$$ is used for finding mode

Remark :-

$$\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})$$

(or)

$$\text{mode} = 3\,\text{Median} - 2\,\text{Mean}$$

Problems

1. The following are the heights of 10 students. Calculate the modal height 63, 65, 66, 65, 64, 65, 65, 61, 67, 68.

→.

∵ 65 occurs four times and no other item occurs 4 or more than four times, 65 c.m is the modal height.

2. Calculate the mode for the frequency distribution given below

Class:	11-15	16-20	21-25	26-30	31-35	36-40	41-45
freq:	8	15	39	47	52	41	28

class:	46-50	51-55
freq:	16	4.

→.

$\therefore l = 30.5 \quad f_1 = 47 \quad f_2 = 41 \quad h = 5$

$$\text{mode} = l + \frac{hf_2}{f_1+f_2} = 30.5 + \frac{5 \times 41}{47+41}$$

$$= 30.5 + \frac{205}{88}$$

$$= 32.83$$

class	mid x_i	f	u_i	$f_i u_i$	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
25.5-30.5	28	47	-1	-47	109
30.5-35.5	33	$\boxed{52}$	0	0	161
35.5-40.5	38	41	1	41	202
40.5-45.5	43	28	2	56	230
45.5-50.5	48	16	3	48	246
50.5-55.5	53	4	4	16	250
		250			

The maximum frequency 52 occurs in 30.5 - 35.5.

3. Calculate the mode for the following distribution

Marks:	0-9	10-19	20-29	30-39	40-49	50-59	60-69
No. of students:	6	29	87	181	247	263	133

marks:	70-79	80-89	90-99
No. of students:	43	9	2

→ Soln:-
We determine the modal class by forming the grouping table

Marks	I	II	III	IV	V	VI
0-9	6					
10-19	29	35		122		
20-29	87		116			
30-39	181	268			297	
40-49	247		428	$\boxed{691}$		515
50-59	263	510				
60-69	133		396		643	
70-79	43	176		185		439
80-89	9		52			
90-99	2	11			54	

From the table, the maximum freq occurs in 40 - 49. The true class limit is 39.5 - 49.5. This is the modal class

$\therefore l = 39.5$ $f_1 = 181$ (Previous) $f_2 = 263$ (Next) $h = 10$

$$\therefore \text{Mode} = l + \frac{hf_2}{f_1 + f_2}$$

$$= 39.5 + \left(\frac{10 \times 263}{181 + 263}\right)$$

$$= 45.42$$

4. Given that the mode of the following frequency distribution of 70 students is 58.75. Find the missing frequency f_1 and f_2.

Class	frequency
52-55	15
55-58	f_1
58-61	25 ←
61-64	f_2
	70

Soln Since $N = 70$, we've $f_1 + f_2 = 30$.

$\Sigma f_i = 70$ [$25 + 15 = 40$, $70 - 40 = 30$]

Mode class 58-61.

$\therefore l = 58$, $h = 3$, $f = 25$

Using the formula for finding Mode

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

$$58.75 = 58 + \frac{3(25 - f_1)}{(2 \times 25) - f_1 - f_2}$$

$$58.75 - 58 = \frac{75 - 3f_1}{50 - f_1 - f_2}$$

$$0.75 = \frac{75 - 3f_1}{50 - f_1 - f_2}$$

$$(0.75)(50 - f_1 - f_2) = 75 - 3f_1$$

$$37.5 - 0.75f_1 - 0.75f_2 = 75 - 3f_1$$
$$-0.75f_1 + 3f_1 - 0.75f_2 = 75 - 37.5$$
$$2.25f_1 - 0.75f_2 = 37.5 \rightarrow (2)$$
$$(1)\times 2.25 \quad 2.25f_1 + 2.25f_2 = 67.5$$
$$-3f_2 = -30.0$$
$$f_2 = \frac{30}{3}$$
$$\boxed{f_2 = 10}$$

Sub. $f_2 = 10$ in (1), we've $f_1 + 10 = 30$

$$f_1 = 30 - 10$$
$$\boxed{f_1 = 20}$$

H.W. The expenditure of 100 families is given below.

Expenditure	0-10	10-20	20-30	30-40	40-50
No. of families	14	—	27	—	15

Mode for the distribution is 24. Calculate the missing frequencies.

Geometric Mean (G.M)

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

$$G = (x_1 x_2 \dots x_n)^{1/n}$$
$$\log G = \frac{1}{n}(\log x_1 + \log x_2 + \dots + \log x_n)$$
$$= \frac{1}{n} \Sigma \log x_i$$
$$G = \text{Anti log}\left[\frac{\Sigma \log x_i}{n}\right]$$

For grouped freq. distribution

$$G = \text{anti log}\left[\frac{1}{N}\left(\Sigma f_i \log x_i\right)\right]$$

where $N = \Sigma f_i$

Harmonic Mean :- (H.M)

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

For grouped freq. distribution,

$$H = \frac{1}{\left(\frac{1}{N}\right)\left[\Sigma\left(\frac{f_i}{x_i}\right)\right]}, \quad N = \Sigma f_i.$$

Problems:-

1. Find the G.M and H.M of the four numbers 2, 4, 6, 27.

$\rightarrow$

$$G.M = (2\times 4\times 6\times 27)^{1/4}$$
$$= (2\times 2^2\times 2\times 3\times 3^3)^{1/4}$$
$$= (2^4\times 3^4)^{1/4}$$
$$= 2\times 3 = 6.$$

$$H.M = \frac{1}{\left(\frac{1}{4}\right)\left[\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\frac{1}{27}\right]} = \frac{4\times 108}{103}$$
$$= 4.19$$

2. Find the G.M and H.M of the following distribution.

x :	1	2	3	4	5
f :	2	4	3	2	1

$\rightarrow$

$\Sigma f_i = 2+4+3+2+1$

$N = 12$

$$G.M = \cancel{1\times 2\times 3\times 4\times 5} \; (1^2\times 2^4\times 3^3\times 4^2\times 5^1)^{\frac{1}{12}}$$
$$= (16\times 27\times 16\times 5)^{1/12}$$
$$= \text{antilog}\left(\frac{1}{12}(\log 34560)\right)$$
$$= \text{antilog}(0.3782)$$
$$= 2.384$$

$$H.M = \frac{1}{\left(\frac{1}{12}\right)\left[\frac{2}{1}+\frac{4}{2}+\frac{3}{3}+\frac{2}{4}+\frac{1}{5}\right]}$$
$$= \frac{12}{2+2+1+\frac{1}{2}+\frac{1}{5}} = 2.11$$

3. Find the G.M for the following distribution

Marks	0-10	10-20	20-30	30-40
No. of students	5	8	3	4

→

Marks	Mid x_i	f_i	$\log_{10} x_i$	$f_i \log_{10} x_i$
0-10	5	5	0.6990	3.4950
10-20	15	8	1.1761	9.4088
20-30	25	3	1.3979	4.1937
30-40	35	4	1.5441	6.1764
		20		23.2739

$$G.M = \text{antilog } \frac{1}{N}\left(\Sigma f_i \log x_i\right)$$

$$= \text{antilog}\left(\frac{1}{20}(23.2739)\right)$$

$$= \text{antilog}(1.1637)$$

$$= 14.38$$

4. Find the H.M for the following distribution

Class	0-10	10-20	20-30	30-40	40-50
freq	15	10	7	5	3

→

Class	Mid x_i	f_i	f_i/x_i
0-10	5	15	3
10-20	15	10	0.6670
20-30	25	7	0.2800
30-40	35	5	0.1430
40-50	45	3	0.0666
		40	4.1566

$$H.M = \frac{1}{\left(\frac{1}{N}\right)\left[\Sigma(f_i/x_i)\right]} = \frac{1}{\left(\frac{1}{40}\right)[4.1566]}$$

$$= \frac{40}{4.1566}$$

$$= 9.6232$$

(Q) Calculate the average speed of a train running at the rate of 20 k.m per hour during the first 100 k.m at 25 k.m.p.h during the second 100 k.m and at 30 k.m.p.h during the third 100 k.m.

→

Weighted H.M is the proper average

$$\boxed{\text{Weighted H.M} = \frac{\sum w_i}{\sum (w_i/x_i)}}$$

$$= \frac{100+100+100}{\frac{100}{20}+\frac{100}{25}+\frac{100}{30}}$$

$$= \frac{300}{5+4+3.3} = 24.4 \text{ kmph.}$$

Measures of Dispersion:-

1. Range
2. Quartile deviation
3. Mean deviation
4. standard deviation.

1. **Range:-**

It is the difference b/w the maximum and minimum value of the variate.

2. **Quartile deviation:-**

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

3. **Mean Deviation:-**

$$M.D = \frac{1}{N}\sum f_i |x_i - A|, \quad N = \sum f_i$$

4. **Standard Deviation:-**

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

$$\text{Variance} = \sigma^2$$

Root Mean square deviation is

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

A = origin

s^2 = mean square deviation

Coefficient of variation

$$C.V = \frac{\sigma}{\bar{x}} \times 100$$

Problems:

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

→.

Soln:- $n = 10$

(i) $$\text{Mean} = \frac{\sum x_i}{n} = \frac{20+22+\cdots+35}{10}$$

$$= \frac{330}{10} = 33$$

(ii) Range = Max. Value − Min. Value

$$= 48 - 20$$

$$= 28$$

(iii)

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

$$\sum x_i^2 = 20^2 + 22^2 + 27^2 + \cdots + 35^2$$

$$= 11652$$

$$\therefore \sigma = \left[\frac{11652}{10} - (33)^2\right]^{1/2}$$

$$= (1165.2 - 1089)^{1/2}$$

$$= (76.2)^{1/2}$$

$$= 8.7$$

(iv) Mean deviation about mean

$$= \frac{1}{10} \sum |x_i - 33|$$

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

$$= 7.2$$

(v) Coefft. of Variation $= \frac{\sigma}{\bar{x}} \times 100 = \frac{8.7}{33} \times 100 = 26.45$

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

Monthly wages	No. of workers
125-175	2
175-225	22
225-275	19
275-325	14
325-375	03
375-425	04
425-475	06
475-525	01
525-575	01
	72

Soln :- Let $A = 300 \quad h = 50 \quad u_i = \dfrac{\text{old } x_i - 300}{50}$

Monthly wages	Mid x_i	f_i	u_i	$f_i u_i$	$f_i u_i^2$	c.f
125-175	150	2	-3	-6	18	2 Q_1
175-225	200	22	-2	-44	88	24
225-275	250	19	-1	-19	19	43
275-325	300	14	0	0	0	57
325-375	350	3	1	3	3	60
375-425	400	4	2	8	16	64
425-475	450	6	3	18	54	70
475-525	500	1	4	4	16	71
525-575	550	1	5	5	25	72
		72		-31	239	

(i) $\bar{x} = A + h\bar{u} \quad , \bar{u} = \dfrac{\Sigma f_i u_i}{\Sigma f_i}$

$= 300 + 50\left(\dfrac{-31}{72}\right) = 300 - 21.53 = 278.47$

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

$N = \Sigma f_i = 72 \quad \dfrac{N}{4} = \dfrac{72}{4} = 18$

$\therefore$ First Quartile class is 175 – 225

$\therefore l = 175 \quad m = 2 \quad f = 22 \quad h = 50$

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

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

$\frac{3N}{4} = \frac{3}{4}(72) = 3 \times 18 = 54$

Third Quartile class is 275 – 325

$\therefore l = 275 \quad m = 43 \quad f_k = 14 \quad h = 50$

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

$= 275 + \frac{550}{14} = 314.29$

$\therefore$ Quartile deviation $= \frac{1}{2}(Q_3 - Q_1)$

$= \frac{1}{2}[314.29 - 211.36]$

$= 51.45$

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

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

$= 2500\,[3.3194 - (-0.430)^2]$

$= 2500\,(3.3194 - 0.1849)$

$= 2500\,(3.1345) = 7836.25$

$\boxed{\sigma = 88.5}$

(iv) coefficient of variation $= \frac{\sigma}{\bar{x}} \times 100$

$= \frac{88.5}{278.47} \times 100$

$= 31.79$

6. The mean and S.D. of 200 items are found to be 60 and 20 at the time of calculation two items are wrongly taken as 3 and 67 instead of 13 and 17, find the correct mean and standard deviation.

$\rightarrow$ Given $n = 200$, $\bar{x} = 60$, $\sigma = 20$

$\bar{x} = 60$

$\Rightarrow \dfrac{\sum x_i}{n} = 60$

$\dfrac{\sum x_i}{200} = 60 \Rightarrow \sum x_i = 12000$

Corrected $\sum x_i = 12000 - (3+67) + (13+17)$

$= 11960$

Corrected $\bar{x} = \dfrac{11960}{200} = 59.8$

$\sigma^2 = \dfrac{\sum x_i^2}{n} - \left(\dfrac{\sum x_i}{n}\right)^2 \Rightarrow 20^2 = \dfrac{\sum x_i^2}{200} - (60)^2$

$\Rightarrow \sum x_i^2 = 200\,(20^2 + 60^2) = 800000$

After correction

$\sum x_i^2 = 800000 - (3^2 + 67^2) + (13^2 + 17^2)$

$= 795960$

$\therefore$ Corrected $\sigma^2 = \dfrac{795960}{200} - (59.8)^2$

$= 403.76$

$\therefore \sigma = \sqrt{403.76} = 20.09$

7. Find (i) mean deviation from the mean

(ii) Variance of the arithmetic progression $a, a+d, a+2d, \ldots, a+2nd$

$\rightarrow$ There are $(2n+1)$ terms in A.P.

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

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

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

$= a + nd$

(i) Mean deviation from mean

$$= \frac{1}{2n+1} \Sigma |x_i - \bar{x}|$$

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

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

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

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

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

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

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

H.W

1. Find the S.D 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

Ans: 2.92

2. Find the S.D and Q.D for the following frequency distribution.

Marks:	30-39	40-49	50-59	60-69	70-79	80-89
No. of students:	37	50	42	21	11	3

Ans: $\sigma = 12.55$

3. The mean of 2 samples of sizes 50 and 100 respectively are 54.1 and 50.3 and S.D are 8 and 7. Obtain the mean and S.D of the sample of size 150 obtained by combining the two samples.

CHAPTER 4 UNIT - II

MOMENTS SKEWNESS AND KURTOSIS

4.0 INTRODUCTION

In previous chapters we have introduced certain measures of central tendencies and measures of dispersion with the aim of finding a "few statistical constants" that represent the entire data. In this chapter we introduce some more statistical constants known as moments.

4.1. MOMENTS

Definition. The r^{th} **moment about any point** A, denoted by μ_r' of frequency distribution (f_i/x_i) is defined by $\mu_r' = \dfrac{\Sigma f_i (x_i - A)^r}{N}$

When $A = 0$ we get $\mu_r' = \dfrac{\Sigma f_i x_i^r}{N}$ which is the r^{th} moment about the origin.

The r^{th} moment about the arithmetic mean $\bar{x}$ of a frequency distribution is given by $\mu_r = \dfrac{\Sigma f_i (x_i - \bar{x})^r}{N}$.

μ_r is also called the r^{th} **central moment.**

Note 1. The first moment about origin coincides with the A.M of the frequency distribution and μ_2 is nothing but the variance of the frequency distribution.

Note 2. $\mu_1 = \dfrac{\Sigma f_i (x_i - \bar{x})}{N} = 0.$

Note 3. $\mu_1' = \dfrac{\Sigma f_i(x_i - A)}{N} = \left[\dfrac{(\Sigma f_i x_i) - A\,\Sigma f_i}{N}\right] = \overline{x} - A.$

$\therefore\ \overline{x} = A + \mu_1'$

We now establish a relation between μ_r' and μ_r.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Note. When the variables x_i are changed into another variable u_i whe

$u_i = \dfrac{x_i - A}{h}$ the r^{th} moment μ_r of the variable x_i is given by

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

Thus the r^{th} moment of the variable x_i is h^r times the r^{th} moment of variable u_i.

Definition. Karl Pearson's β and γ coefficients are defined as follo

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad \text{and} \quad \beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\gamma_1 = \sqrt{\beta_1} \quad \text{and} \quad \gamma_2 = \beta_2 - 3.$$

The above four coefficients depend upon the first four centr moments. They are pure numbers independent of units in which the variab x_i is expressed. Also their values are not affected by change of origin and scal These constants are used in section 4.2 in the study of skewness and kurtosi

.2 SKEWNESS AND KURTOSIS

If the values of a variable x_i are distributed symmetrically about its mean which is taken as the origin then for every positive value of $x - \bar{x}$ there corresponds a negative equal value. Hence when these values are added they retain their signs and cancel on addition.

$$\therefore \mu_3 = \frac{1}{N}\Sigma f_i (x_i - \bar{x})^3 = 0. \text{ Hence } \beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0.$$

Thus in the case of symmetrical distribution $\beta_1 = 0$. If a distribution fails to be symmetric (asymmetric) then we say that it is a skewed distribution. Thus skewness means lack of symmetry. From the above discussion we see that β_1 can be taken as a measure of skewness. We say that a frequency distribution has positive skewness if $\beta_1 > 0$ and negative skewness if $\beta_1 < 0$.

For a symmetric distribution the mean, median and mode coincide. Hence for an asymmetrical distribution the distance between the median and mean may be used as measures of skewness.

$\therefore$ Mean $-$ Mode and Mean $-$ Median may be taken as measures of skewness.

These measures were suggested by Karl Pearson.

Another measure of skewness due to Bowley is based on the fact that for a positively skewed distribution the third quartile is farther from the median than the first quartile so that $Q_3 -$ Median $>$ Median $- Q_1$. Hence $(Q_3 -$ Median$) - ($Median $- Q_1) = Q_3 + Q_1 - 2$ Median may be taken as another measure of skewness.

The above measures of skewness are the absolute measures of skewness.

To make these measures free from units of measurements, comparison with other distribution may be possible we divide them suitable measure of dispersion and obtain the following coefficien skewness.

(i) Karl Pearson's coefficient of skewness.

$$\frac{\text{Mean} - \text{Mode}}{\sigma} \text{ and } \frac{3\,(\text{Mean} - \text{Median})}{\sigma}$$ are called Pearson's coefficients of skewness.

(ii) Bowley's coefficient of skewness is given by $\frac{Q_3 + Q_1 - 2\,\text{Medi}}{Q_3 - Q_1}$

Kurtosis

Definition. Kurtosis is the *degree of peakedness* of a distribution usu taken relative to a normal distribution. Thus kurtosis enables us to have idea about the *flatness or peakedness* of a frequency curve. It is measu by the coefficient β_2.

For a normal curve $\beta_2 = 3$ or $(\gamma_2 = 0)$ **messokurtic.**

For a curve which is flater than the normal cur $\beta_2 < 3$ or $(\gamma_2 < 0)$ and such a curve is known as **platykurtic.**

For a curve which is more peaked than the normal cur $\beta_2 > 3$ or $(\gamma_2 > 0)$ and such a cure is known as **leptokurtic.**

Solved problems.

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

x	0	1	2	3	4	5	6
f	5	15	17	25	19	14	5

Solution. Here $\bar{x} = \dfrac{\Sigma f_i x_i}{\Sigma f_i} = \dfrac{300}{100} = 3.$

Choosing $u_i = x_i - \bar{x} = x_i - 3$ we have the following table.

x_i	f_i	u_i	$f_i u_i$	$f_i u_i^2$	$f_i u_i^3$	$f_i u_i^4$
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**

$$\mu_1 = (1/N)\,\Sigma f_i(x_i - \bar{x}) = 0.$$

$$\mu_2 = (1/N)\,\Sigma f_i(x_i - \bar{x})^2 = \frac{242}{100} = 2.42.$$

$$\mu_3 = (1/N)\,\Sigma f_i(x_i - \bar{x})^3 = -\frac{6}{100} = -0.06.$$

$$\mu_4 = (1/N)\,\Sigma f_i(x_i - \bar{x})^4 = \frac{1310}{100} = 13.10.$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-0.06)^2}{2.42^3} = \frac{.0036}{14.1725} = 0.0003.$$

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

Since $\beta_1 > 0$ the distribution is *positively skewed.*

Since $\beta_2 = 2.237 < 3$ the distribution is *platykurtic.*

Problem 2. Calculate the values of β_1 and β_2 for the distribution [illegible]

Table 4.

Solution. Taking $u_i = \dfrac{x_i - 24.5}{10}$ we get the following table.

x_i	f_i	u_i	$f_i u_i^2$	$f_i u_i^3$	$f_i u_i^3$	f_i [illegible]
04.5	11	− 2	− 22	44	− 88	176
14.5	20	− 1	− 20	20	− 20	20
24.5	16	0	0	0	0	0
34.5	36	1	36	36	36	36
44.5	17	2	34	68	136	272
Total	100	0	28	168	64	504

Here we have chosen $A = 24.5$ and $h = 10$.

$$\mu_1' = \frac{1}{N}\Sigma f_i (x_i - A) = \frac{1}{N}\Sigma f_i u_i \times h = \frac{28}{100} \times 10 = 2.8$$

$$\mu_2' = \frac{1}{N}\Sigma f_i u_i^2 \times h^2 = \frac{168}{100} \times 10^2 = 168.$$

$$\mu_3' = \frac{1}{N}\Sigma f_i u_i^3 \times h^3 = \frac{64}{100} \times 10^3 = 640.$$

$$\mu_4' = \frac{1}{N}\Sigma f_i u_i^4 \times h^4 = \frac{504}{100} \times 10^4 = 50400$$

Now, $\mu_1 = 0$,

$$\mu_2 = \mu_2' - (\mu_1')^2 = 168 - (2.8)^2 = 160.16.$$

$$\mu_3 = \mu_3' - 3\,\mu_2'\,\mu_1' + 2\,(\mu_1')^3 = 640 - 3 \times 168 \times 2.8 + 2\,(2.8)^3$$

$$= -727.296.$$

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

$$= 50400 - 4 \times 640 \times 2.8 + 6 \times 168 \times (2.8)^2 - 3\,(2.8)^4$$

$$= 50950.323$$

$$\text{Now, } \beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0.129 \text{ (verify)}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 1.985 \text{ (verify)}$$

Problem 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. Given $\mu_1' = 1$; $\mu_2' = 2.5$; $\mu_3' = 5.5$; $\mu_4' = 16$ where $A = 2$.

(i) *Moments about mean.*

$$\mu_1 = 0.$$

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

$$\mu_3 = \mu_3' - 3\,\mu_2'\,\mu_1' + 2\,(\mu_1')^3 = 5.5 - 3 \times 2.5 + 2 = 0.$$

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

$$= 16 - 4 \times 5.5 + 6 \times 2.5 - 3 = 6.$$

(ii) *Moments about zero.*

We have $\bar{x} = A + \mu_1'$ (refer Note 3 in 4.1)

$$= 2 + 1 = 3.$$

Now, the first moment about zero $\mu_1' = (1/N)\Sigma f_i (x_i - 0)$

$$= \bar{x} = 3.$$

Now, $\mu_2' = \mu_2 + (\mu_1')^2 = 1.5 + 3^2 = 10.5$

$\mu_3' = \mu_3 + 3\mu_2\mu_1' + (\mu_1')^3 = 0 + 3 \times 1.5 \times 3 + 3^3 = 40.5$

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

$$= 6 + (4 \times 0 \times 3) + (6 \times 1.5 \times 3^2) + 3^4 = 168.$$

Problem 4. The first three moments about the origin are given by $\mu_1' = \frac{1}{2}(n+1)$; $\mu_2' = \frac{1}{6}(n+1)(2n+1)$; $\mu_3' = \frac{1}{4}n(n+1)^2$. Examine the skewness of the distribution.

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

$$= \frac{1}{4}n(n+1)^2 - 3 \times \frac{1}{6}(n+1)(2n+1)\,\frac{1}{2}(n+1) + 2\left[\frac{1}{2}(n+1)\right]^3$$

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

$$= \frac{1}{4}(n+1)^2 \times 0 = 0. \text{ Hence } \mu_3 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = \frac{1}{6}(n+1)(2n+1) - \left[\frac{1}{2}(n+1)\right]^2$$

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

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

$\mu_2 \neq 0$ if $n \neq \pm 1$.

$\therefore$ When $n > 1$, $\beta_1 = 0$.

Hence the distribution is symmetric.

Then $d_i = y_i - \ldots$ y and the value of y determined ... residuals. The principle of least squares ... parameters involved in $f(x)$ should be chosen in such a way that Σd_i^2 is minimum.

Fitting a straight line

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

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.

(x_i, y_i) where $i = 1, 2, \ldots$

The residual d_i ...

$\therefore \Sigma d_i^2 = \Sigma (y$ ...

By the princi... parameters a, b and c so ...

$\frac{\partial R}{\partial a} = 0 \Rightarrow$ ...

$\Rightarrow$ ...

$\therefore a\Sigma x_i^4 +$ ...

$\frac{\partial R}{\partial b} = 0 \Rightarrow$ ...

$\Rightarrow$ ...

$\therefore a\Sigma x_i^3$ ...

$\frac{\partial R}{\partial c} = 0 \Rightarrow$ ...

$\Rightarrow$ ...

$\therefore a\Sigma x_i^2$ ...

Equation... which a, b and c ca...

Note. I... linear form by som... principle of least s...

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.

Curves of the form (i) $y = bx^a$ (ii) $y = ab^x$ (iii) $y = ae^{bx}$ ... of special interest which are dealt with here in solved problems.

Solved problems.

Problem 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

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

$$\Sigma x_i y_i = a\Sigma x_i^2 + b\Sigma x_i.$$

x_i	y_i	$x_i y_i$	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 10	26.5	69.0	30

Hence the normal equations are

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

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

Solving (1) and (2) we get $a = 1.6$ and $b = 2.1$.

$\therefore$ The straight line fitted for the data is $y = 1.6x + 2.1$.

Problem 2. Fit a straight line to the following data and estimate the value of y corresponting to $x = 6$.

x	0	5	10	15	20	25
y	12	15	17	22	24	30

Solution. Take $u_i = \frac{1}{5}(x_i - 15)$ and $v_i = y_i - 22$.

Let $v = au + b$ be the straight line to be fitted.

We get the following normal equations to get the parameters a and b. Then the normal equations are.

$$\Sigma v_i = a \Sigma u_i + nb.$$

$$\Sigma u_i v_i = a \Sigma u_i^2 + b \Sigma u_i.$$

x_i	y_i	u_i	v_i	$u_i v_i$	u_i^2
0	12	− 3	− 10	30	9
5	15	− 2	− 7	14	4
10	17	− 1	− 5	5	1
15	22	0	0	0	0
20	24	1	2	2	1
25	30	2	8	16	4
Total	-	− 3	− 12	67	19

∴ The normal equations are

$$-3a + 6b = -12 \quad \ldots\ldots (1)$$

$$19a - 3b = 67. \quad \ldots\ldots (2)$$

Solving for a and b we get $a = 3.49$ and $b = -0.26$.

∴ The straight line to be fitted becomes $y - 22 = 3.49\left(\frac{x - 15}{5}\right) \times$ (

$\therefore 5y - 110 = 3.49x - 52.35 - 1.30$

$\therefore 5y = 3.49x + 56.35$

$\therefore y = .698x + 11.27.$

Now for $x = 6$ the estimated value of y is $y = .698 \times 6 + 11.27 = 15$

Problem 3. Fit a second degree parabola by taking x_i as the independ 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 $y = ax^2 + bx + c$. Then we have the normal equations to find a, b, c.

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

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

$$a\Sigma x_i^2 + b\Sigma x_i + nc = \Sigma y_i.$$

x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y_i$	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 10	76	243	30	851	100	354

Now, the normal equations become

$$354a + 100b + 30c = 851 \quad \ldots\ldots (1)$$

$$100a + 30b + 10c = 243 \quad \ldots\ldots (2)$$

$$30a + 10b + 5c = 76 \quad \ldots\ldots (3)$$

Solving for a, b and c we get $a = 2.21$; $b = 0.26$ and $c = 1.42$ (verify)

$\therefore$ The second degree parabola is $y = 2.21x^2 + 0.26x + 1.42$.

Problem 4. 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. $y = bx^a$.

$\therefore \log y = a \log x + \log b$.

Let $\log y = Y$ 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

$$\Sigma Y = A \Sigma X + nB.$$

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

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**

$\therefore$ The normal equations are

$$2.9A + 6B = 14.9 \text{ approximately}$$

$$1.8A + 2.9B = 6.6 \text{ approximately}$$

$\therefore A = -2.3$ and $B = 3.6$ (verify)

$\therefore A = a = -2.3$ and $B = \log b = 3.6$

$\therefore a = -2.3$ and $b = \text{antilog } 3.6 = 3981.$

$\therefore$ The required equation to the curve is $y = 3981\, x^{-2.3}$

Problem 5. Explain the method of fitting the curve of

$y = ae^{bx}$ $(a > 0)$ ...... (1)

Solution. $y = ae^{bx}$ ...... (2)

$\therefore \log y = \log a + bx \log e$

Let $Y = \log y$; $B = \log a$; $A = b \log e$

$\therefore$ (2) between $Y = Ax + B$

This is linear equation in x and Y whose normal equation

$$\Sigma x_i Y_i = A \Sigma x_i^2 + B \Sigma x_i$$

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

From the two normal equations we can get the values of A and consequently a and b can be obtained from $a = \text{antilog}$ (B

$b = \dfrac{A}{\log e}$. Thus the curve of best fit (1) can be obtained.

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

Solution. The curve can be transferred to the form of a straight line as fo

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

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

Hence the above equation takes the form $Y = Ax + B$

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

$$\Sigma Y_i x_i = A \Sigma x_i^2 + B \Sigma x_i$$

$$\Sigma Y_i = A \Sigma x_i + nB.$$

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 $y = k\,a^{bx}$ can be fitted.

Problem 7. Fit a curve of the form $y = ab^x$ to the following data.

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

Solution. $y = ab^x$ ...... (1)

$\therefore \log y = \log a + x \log b$ ...... (2)

Let $\log y = Y$; $\log a = B$ and $\log b = A$.

$\therefore$ (2) becomes $Y = AX + B$ ...... (3) where $X = x - 1954$.

x	y	$X = x - 1954$	$Y = \log y$	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

$$\Sigma XY = A\,\Sigma X^2 + B\,\Sigma X$$

$$\Sigma Y = A\,\Sigma X + nB.$$

$28A = 2.1491$ ...... (4)

$7B = 17.9363$ ...... (5)

Solving the above equations we get $A = 0.0768 \quad B = 2.5623$

$\therefore \quad b =$ antilog $A =$ antilog $0.0768 = 1.19$ (approximately)

$a =$ antilog $B =$ antilog $2.5623 = 365.01$ (approximate

$\therefore$ The curve of good fit is $y = (365.01)\,(1.19)^X$

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

UNIT III

CORRELATION

Defn:

Consider a set of bivariate data (x_i, y_i), $i = 1, 2, \ldots n$. If there is a change in one variable corresponding to a change in the other variable, then the variables are correlated.

Karl Pearson's coefficient of correlation :-

$$r_{xy} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n \sigma_x \sigma_y}$$

where $\bar{x}, \bar{y}$ are arithmetic means

σ_x, σ_y are S.D of x and y.

Covariance :-

$$\text{Cov}(x, y) = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Then $r_{xy} = \dfrac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$.

Another form of correlation

1) $$r_{xy} = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2 \sigma_x \sigma_y}$$

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

The correlation coefficient is always lies between -1 and $+1$.

(ie) $-1 \leq r_{xy} \leq 1$.

Note

1. If $r = 1 \Rightarrow$ correlation is perfect and positive.
2. If $r = -1 \Rightarrow$ correlation is perfect and negative.
3. If $r = 0 \Rightarrow$ Variables are uncorrelated.

Ans: If variables x and y are uncorrelated then $\text{cov}(x,y) = 0$

$\Rightarrow r_{xy} = 0$

Problems

1. 10 students obtained the following percentage of marks in the college internal test (x) and in the final university examination (y). Find the 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

Soln:- Choose origin $A = 63$ for x

$B = 60$ for y

Let $u_i = x_i - A$ $\quad v_i = y_i - B$

x_i	$u_i = x_i - A$	y_i	$v_i = y_i - B$	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
	−70		−12	980	994	500

$$r = \frac{n \sum u_i v_i - \sum u_i \sum v_i}{\left[n \sum u_i^2 - (\sum u_i)^2\right]^{1/2} \left[n \sum v_i^2 - (\sum v_i)^2\right]^{1/2}}$$

Rank correlation :-

$$\rho = 1 - \frac{6 \Sigma (x-y)^2}{n(n^2-1)}.$$

This is known as Spearman's formula for rank correlation coefficient.

1. Find the rank correlation coefficient between the height in cm 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

Soln:-

Height	Rank x	weight	Rank 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
					16

Rank correlation

$$\rho = 1 - \frac{6\sum(x-y)^2}{n(n^2-1)}$$

$$= 1 - \frac{6\times 16}{6\times 35} = 1 - 0.457$$

$$= 0.543$$

2. Obtain the rank correlation coefficient for the following data.

x	5	2	8	1	4	6	3	7
y	4	5	7	3	2	8	1	6

Ans: $\frac{2}{3}$

3. From the following data of marks obtained by 10 students in physics and chemistry calculate the rank correlation coefficient.

physics (x)	35	56	50	65	44	38	44	50	15	26
Chemistry (y)	50	35	70	25	35	58	75	60	55	35

Solution:-

x	Rank x	y	Rank y	$x-y$	$(x-y)^2$
35	8	50	6	2	4
56	2	35	8	-6	36
50	3.5	70	2	1.5	2.25
65	1	25	10	-9	81
44	5.5	35	8	-2.5	6.25
38	7	58	4	3	9
44	5.5	75	1	4.5	20.25
50	3.5	60	3	0.5	0.25
15	10	55	5	5	25
26	9	35	8	1	1
					185

$$\rho = 1 - \frac{6\sum(x-y)^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 188}{10(10^2-1)}$$

$$= 1 - \frac{1128}{990} = -0.139$$

$\therefore \left[\frac{2(2^2-1)}{12} + \frac{2(2^2-1)}{12}\right] + \frac{3}{1}$

$= 3$

$\therefore 185 + 3 = 188$

$\left[\frac{1}{12} m(m^2-1)\right.$ where m is the no of times an item has repeated values.]

4. Three judges assign the ranks to 8 entries in a beauty contests.

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?

→

x	y	z	$x-y$	$(x-y)^2$	$y-z$	$(y-z)^2$	$z-x$	$(z-x)^2$
1	3	1	-2	4	2	4	0	0
2	2	2	0	0	0	0	0	0
4	1	3	3	9	-2	4	-1	1
3	5	4	-2	4	1	1	1	1
7	4	5	3	9	-1	1	-2	4
6	7	7	-1	1	0	0	1	1
5	6	8	-1	1	-2	4	3	9
8	8	6	0	0	2	4	-2	4
				28		18		20

$$\rho_{xy} = 1 - \frac{6\sum(x-y)^2}{n(n^2-1)} = 1 - \frac{6\times 28}{8(8^2-1)} = 1 - \frac{168}{504} = 0.667$$

$$\rho_{yz} = 1 - \frac{6\sum(y-z)^2}{n(n^2-1)} = 1 - \frac{6\times 18}{8(8^2-1)} = 1 - \frac{108}{504} = 0.786$$

$$\rho_{zx} = 1 - \frac{6\sum(z-x)^2}{n(n^2-1)} = 1 - \frac{6\times 20}{8(8^2-1)} = 1 - \frac{120}{504} = 0.762$$

$\rho_{yz} > \rho_{xy}$ $\rho_{yz} > \rho_{zx}$

∴ Judge Mr. Y and Mr. Z have nearest approach to common taste in beauty.

5. The coefficient of rank correlation of marks obtained by 10 students in Maths and physics was found to be 0.8. It was later discovered that the differences in ranks in two subjects obtained by one of the students was wrongly taken as 5 instead of 8. Find the correct coefficient of rank correlation.

→ W.K.T $\rho_{xy} = 1 - \dfrac{6\sum(x-y)^2}{n(n^2-1)}$.

$\rho_{xy} = 0.8$ (Given) $n = 10$.

[illegible]

$\therefore 6\sum(x-y)^2 = $ [illegible] $= 198$

$\sum(x-y)^2 = \frac{198}{6} = 33$

Corrected $\sum(x-y)^2 = 33 - 5^2 + 8^2$

$= 72$

After correction,

$$\rho_{xy} = 1 - \frac{6 \times 72}{10(10^2-1)}$$

$$= 1 - \frac{432}{990} = 0.564.$$

$\therefore$ The correct coefficient of rank correlation is 0.564

<u>Regression</u> :-

If there is a functional relationship b/w the 2 variables (x_i, y_i) in scatter diagram will cluster around some curve called the <u>curve of regression.</u>

If the curve is a <u>straight line</u> it is called a <u>line of regression</u> b/w the two variables.

<u>Regression lines</u> :-

The equation of regression line of y on x is

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

The equation of regression line of x on y is

$$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y}).$$

Result:-

1. Correlation coefficient is the geometric mean b/w the regression coefficients, (ie) $r = \pm\sqrt{b_{xy} \cdot b_{yx}}$, where

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x},$$

$b_{xy} > 1$

$b_{yx} < 1$.

2. The angle b/w the two regression lines is given by $\theta = \tan^{-1}\left[\left(\frac{r^2-1}{r}\right)\left(\frac{\sigma_x\sigma_y}{\sigma_x^2-\sigma_y^2}\right)\right]$

$\rightarrow$

Proof

W.K.T

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x}) \longrightarrow (1)$$

$$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y}) \longrightarrow (2)$$

$$(2) \Rightarrow y - \bar{y} = \frac{1}{r}\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$\therefore$ slope of the two lines (1) & (2) are $r\frac{\sigma_y}{\sigma_x}$ and $\frac{\sigma_y}{r\sigma_x}$.

Let θ be the acute angle b/w two lines of regression.

$$\tan\theta = \frac{m_1 - m_2}{1 + m_1 m_2}$$

$$\tan\theta = \frac{r\frac{\sigma_y}{\sigma_x} - \frac{\sigma_y}{r\sigma_x}}{1 + \left(r\frac{\sigma_y}{\sigma_x}\right)\left(\frac{\sigma_y}{r\sigma_x}\right)}$$

$$= \frac{r^2\sigma_y - \sigma_y}{r\sigma_x} \bigg/ \frac{r\sigma_x^2 + r\sigma_y^2}{r\sigma_x^2}$$

$$= \frac{(r^2-1)}{r}\frac{\sigma_y}{\sigma_x} \Big/ \frac{r(\sigma_x^2+\sigma_y^2)}{r\,\sigma_x^2}$$

$$= \left(\frac{r^2-1}{r}\right)\frac{\sigma_y}{\sigma_x} \times \frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}$$

$$\tan\theta = \left(\frac{r^2-1}{r}\right)\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)$$

$$\theta = \tan^{-1}\left[\left(\frac{r^2-1}{r}\right)\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)\right]$$

The obtuse angle b/w the regression lines is given by

$$\tan^{-1}\left[\left(\frac{r^2-1}{r}\right)\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)\right]$$

Note

1. If $r=0$, $\tan\theta = \infty$
$\Rightarrow \theta = \pi/2$

Thus if the two variables are uncorrelated, then the lines of regression are perpendicular to each other.

2. If $r = \pm 1$, $\tan\theta = 0$
$\Rightarrow \theta = 0$ or π

the two lines of regression are parallel. (Coincide)

Problems:

1. The following data relate to the marks of 10 students in the internal test and university examination for the maximum of 50 in each.

Internal Marks	25	28	30	32	35	36	38	39	42	45
University exam Marks	20	26	29	30	25	18	26	35	35	46

(i) obtain the two regression equations and determine

(ii) the most likely internal mark for the university mark of 25

(iii) the most likely university mark for the internal mark of 30.

→ let x = internal mark

y = university mark.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{25+28+\cdots+45}{10} = 35$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{20+26+\cdots+46}{10} = 29$$

For Regression equation

x_i	$x_i - \bar{x}$ (35)	$(x_i-\bar{x})^2$	y_i	$y_i - \bar{y}$ (29)	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
25	−10	100	20	−9	81	90
28	−7	49	26	−3	9	21
30	−5	25	29	0	0	0
32	−3	9	30	1	1	−3
35	0	0	25	−4	16	0
36	1	1	18	−11	121	−11
38	3	9	26	−3	9	−9
39	4	16	35	6	36	24
42	7	49	35	6	36	42
45	10	100	46	17	289	170
	0	358		0	598	324

$$\sigma_x^2 = \frac{\Sigma(x_i-\bar{x})^2}{n} = \frac{358}{10} = 35.8$$

$$\sigma_y^2 = \frac{\Sigma(y_i-\bar{y})^2}{n} = \frac{598}{10} = 59.8$$

$\therefore \sigma_x = \sqrt{35.8} = 5.98$

$\sigma_y = \sqrt{59.8} = 7.73$

$$r = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n \sigma_x \sigma_y}$$

$$= \frac{324}{10 \times 5.98 \times 7.73} = 0.7$$

Regression of y on x is

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$$y - 29 = (0.7)\frac{(7.73)}{(5.98)}(x - 35)$$

$$\Rightarrow y - 29 = 0.905(x - 35)$$

$$y = 0.905x - 2.675 \longrightarrow (1)$$

Regression of x on y is

$$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$$x - 35 = (0.7)\frac{(5.98)}{(7.73)}(y - 29)$$

$$\Rightarrow x = 0.542y + 19.282 \longrightarrow (2)$$

(ii) Gn $y = 25$, (2) $\Rightarrow$

$$\therefore x = 0.542(25) + 19.282$$

$$= 32.83$$

(iii) Gn $x = 30$, (1) $\Rightarrow$

$$\therefore y = 0.905(30) - 2.675$$

$$= 24.475$$

be 63.

3. The two variables x and y have the regression lines $3x+2y-26=0$ and $6x+y-31=0$.

Find (i) the mean values of x and y

(ii) the correlation coefficient b/w x and y

(iii) The variance of y if the variance of x is 25.

$\longrightarrow$.

(i) Since the two lines of regression pass through $(\bar{x}, \bar{y})$, we've

$$3\bar{x} + 2\bar{y} = 26 \rightarrow (1)$$

$$6\bar{x} + \bar{y} = 31 \rightarrow (2)$$

$$(1) \times 2 \Rightarrow 6\bar{x} + 4\bar{y} = 52$$

$$(-) \quad (-) \quad (-)$$

$$-3\bar{y} = -21$$

$$\boxed{\bar{y} = 7}$$

Sub $\bar{y} = 7$ in (1)

$3\bar{x} + 2(7) = 26$

$3\bar{x} = 26 - 14 = 12$

$\boxed{\bar{x} = 4}$

(ii) From (1), $y = -\frac{3}{2}x + 13$ $\quad b_{yx} = -3/2 = -1.5 < 1$

From (2), $x = -\frac{1}{6}y + \frac{31}{6}$, $\quad b_{xy} = -1/6 = -0.1666 < 1$

$\therefore r^2 = b_{yx} \times b_{xy}$ $\quad$ [$b_{yx} \cdot b_{xy} = -1.5 \times -0.1666 = 0.2499 < 1$, $\because r^2 < 1$ possible]

$= -\frac{3}{2} \times -\frac{1}{6}$

$r^2 = 0.2499 \quad r = 0.5$

Since the both regression coefficients are negative, $r = -\sqrt{0.2499} = -0.5 < 1$

(iii) Given $\sigma_x = 5$ $\quad \sigma_x^2 = 25$

we've $b_{yx} = r \frac{\sigma_y}{\sigma_x}$

$-\frac{3}{2} = -0.5 \left(\frac{\sigma_y}{5}\right)$

$-1.5 = -0.1\,\sigma_y$

$\sigma_y = \frac{-1.5}{-0.1}$

$\boxed{\sigma_y = 15}$

4. Out of two lines of regression given by $x + 2y - 5 = 0$ and $2x + 3y - 8 = 0$, which one is the regression line of x on y?

$\rightarrow$ Suppose $x + 2y - 5 = 0$ is eqn. of regr. line of x on y.

$2x + 3y - 8 = 0$ is eqn. of reg. line of y o

Then,

line can be written as

$3y = -2x + 8$

$y = -\frac{2}{3}x + \frac{8}{3}$

The regression coefficients

$b_{yx} = -\frac{2}{3}$ $\quad b_{xy} = -2$

Now $r^2 = b_{yx} \times b_{xy}$

$= -\frac{2}{3} \times -2$

$= \frac{4}{3} > 1$ This is impossible.

Hence our assumption is wrong.

$\therefore$ $2x + 3y - 8 = 0$ is the regression line of x on y.

5. If $x = 4y + 5$ and $y = kx + 4$ are the regression lines of x on y and y on x respectively. (i) Show that $0 \le k \le 1/4$ (ii) if $k = 1/8$ find the means of the two variables x and y and the corresponding correlation coefficient between them.

$\rightarrow$

The regression line of x on y is

$x = 4y + 5$

$\therefore$ $\boxed{b_{xy} = 4}$

The regression line of y on x is

$y = kx + 4$

$\therefore$ $\boxed{b_{yx} = k}$

Now $b_{xy}\, b_{yx} = r^2$

$\Rightarrow 4 \cdot k = r^2$

Also $r^2 \le 1$, [illegible]

[illegible]

(iii) Given $b_{xy} = 4$, $b_{yx} = \frac{1}{8}$

$\therefore r^2 = b_{xy} \cdot b_{yx}$

$= 4 \times \frac{1}{8} = \frac{1}{2} \Rightarrow r = \pm \frac{1}{\sqrt{2}}$

$\therefore r = \frac{1}{\sqrt{2}}$ ∵ both regression coefficients are positive, we take $r = +\frac{1}{\sqrt{2}}$.

For mean values

Since regression lines pass through $(\bar{x}, \bar{y})$ we've

$\bar{x} = 4\bar{y} + 5$

$\bar{y} = \frac{1}{8}\bar{x} + 4$.

Solving, $\bar{x} - 4\bar{y} = 5 \longrightarrow (1)$

$-\frac{1}{8}\bar{x} + \bar{y} = 4 \longrightarrow (2)$

$(1) \Rightarrow \bar{x} - \cancel{4\bar{y}} = 5$

$(2) \times 4 \Rightarrow -\frac{1}{2}\bar{x} + \cancel{4\bar{y}} = 16$

$\left(1 - \frac{1}{2}\right)\bar{x} = 21$

$\frac{1}{2}\bar{x} = 21$

$\boxed{\bar{x} = 42}$

Sub. $\bar{x} = 42$ in (1)

$42 - 4\bar{y} = 5$

$-4\bar{y} = 5 - 42$

$-4\bar{y} = -37$

$\bar{y} = \frac{-37}{-4}$

$\boxed{\bar{y} = 9.25}$

6. Find the correlation coefficient b/w x and y from the following table.

y \ x	5	10	15	20
4	2	4	5	4
6	5	3	6	2
8	3	8	2	3

Soln :-

$y \backslash x$	5	10	15	20	Total
4	2	4	5	4	$f_1 = 15$
6	5	3	6	2	$f_2 = 16$
8	3	8	2	3	$f_3 = 16$
Total	$g_1 = 10$	$g_2 = 15$	$g_3 = 13$	$g_4 = 9$	$N = 47$

Correlation coeff t.. b/w x and y is given by

$$r_{xy} = \frac{\sum\sum f_{ij} x_i y_j - \frac{1}{N}(\sum g_i x_i)(\sum f_j y_j)}{\sqrt{\sum g_i x_i^2 - \frac{1}{N}(\sum g_i x_i)^2}\sqrt{\sum f_j y_j^2 - \frac{1}{N}(\sum f_j y_j)^2}}$$

where $i = 1, 2, 3, 4$ (column)
$j = 1, 2, 3$ (row)

$$\sum g_i x_i = 50 + 150 + 195 + 180 = 575$$

$$\sum f_j y_j = 60 + 96 + 128 = 284$$

$$\sum g_i x_i^2 = 250 + 1500 + 2925 + 3600 = 8275$$

$$\sum f_j y_j^2 = 240 + 576 + 1024 = 1840$$

$$\sum\sum f_{ij} x_i y_j = (40 + 160 + 300 + 320) + (150 + 180 + 540 + 240) + (120 + 640 + 240 + 480)$$

$$= 3410$$

$$r_{xy} = \frac{3410 - \frac{1}{47}(575 \times 284)}{\sqrt{8275 - \frac{1}{47}(575)^2}\sqrt{1840 - \frac{1}{47}(284)^2}}$$

$$= \frac{3410 \times 47 - (575 \times 284)}{\sqrt{8275 \times 47 - (575)^2}\sqrt{1840 \times 47 - (284)^2}}$$

$$= -0.16$$

UNIT - IV - INTERPOLATION

Interpolation is the process of finding the most appropriate estimate for missing data.

Finite Differences :-

U_x is a function of the independent Variable x and if $a, a+h, a+2h, \ldots$ are a finite set of equidistant values then $U_a, U_{a+h}, U_{a+2h}, \ldots$ are the corresponding values for U_x.

The values of the independent Variable x are called arguments, the corresponding values of U_x are called entries and h is known as interval of differencing.

The operator Δ which is known as first order difference on U_x as

$$\Delta U_x = U_{x+h} - U_x$$

Note

$\Delta U_x = U_{x+h} - U_x$

$\Delta U_x = U_{x+h} - U_x$

$\Delta U_x = 0$ if U_x is constant.

$\Delta^2 U_x = \Delta U_{x+h} - \Delta U_x$

Note

$\Delta \rightarrow$ forward difference operator.

$\nabla \rightarrow$ backward difference operator.

E on U_x is defined as $EU_x = U_{x+h}$

Result :-

$$E^n U_x = U_{x+nh}$$

$E^5 U_0 = U_5$ ...... like this.

$E^3 U_4 = U_{4+3} = U_7$

Result

1. $E = 1+\Delta$

2. $E = (1-\nabla)^{-1}$

Problems :-

1. Find first and second order differences for (i) $U_x = ab^{cx}$

(ii) $U_x = \frac{x}{x^2+7x+12}$ taking interval h.

$\rightarrow$

(i) $\Delta U_x = U_{x+h} - U_x$ $\rightarrow$ first order difference

$= ab^{c(x+h)} - ab^{cx}$

$= ab^{cx} b^{ch} - ab^{cx}$

$= ab^{cx}[b^{ch} - 1]$

$\Delta^2 U_x = \Delta U_{x+h} - \Delta U_x$ $\rightarrow$ second order difference

$= ab^{c(x+h)}(b^{ch}-1) - ab^{cx}(b^{ch}-1)$

$$= (b^{ch}-1)\left[ab^{c(x+h)} - ab^{cx}\right]$$

$$= (b^{ch}-1)\left(ab^{cx}\,b^{ch} - ab^{cx}\right)$$

$$= (b^{ch}-1)\; ab^{cx}\,(b^{ch}-1)$$

$$= (b^{ch}-1)^2\; ab^{cx}$$

(ii) $U_x = \dfrac{x}{x^2+7x+12}$

Now

$$\frac{x}{x^2+7x+12} = \frac{x}{(x+4)(x+3)} = \frac{A}{x+4} + \frac{B}{x+3}$$

(using partial fraction method)

$\Rightarrow A(x+3) + B(x+4) = x$

$x = -3$, $\boxed{B = -3}$

$x = -4$, $-A = -4$

$\boxed{A = 4}$

$\therefore U_x = \dfrac{4}{x+4} - \dfrac{3}{x+3}$

$\rightarrow$ first order difference $\boxed{h = x+1}$

$$\Delta U_x = U_{x+h} - U_x$$

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

$$= \frac{4}{x+5} - \frac{3}{x+4} - \frac{4}{x+4} + \frac{3}{x+3}$$

$$\frac{4}{x+5} - \frac{7}{x+4} + \frac{3}{x+3}$$

Next second order difference

$$\Delta^2 U_x = \Delta U_{x+h} - \Delta U_x$$

$$= \left[\frac{4}{(x+1)+5} - \frac{7}{(x+1)+4} + \frac{3}{(x+1)+3}\right] - \left[\frac{4}{x+5} - \frac{7}{x+4} + \frac{3}{x+3}\right]$$

$$= \frac{4}{x+6} - \frac{7}{x+5} + \frac{3}{x+4} - \frac{4}{x+5} + \frac{7}{x+4} - \frac{3}{x+3}$$

$$= \frac{4}{x+6} - \frac{11}{x+5} + \frac{10}{x+4} - \frac{3}{x+3}$$

2. Evaluate $\dfrac{\Delta^2 x^3}{E x^2}$ taking $h=1$.

$\rightarrow$

$$\Delta x^3 = (x+1)^3 - x^3$$
$$= x^3 + 3x^2 + 3x + 1 - x^3$$
$$= 3x^2 + 3x + 1$$

$$\Delta^2 x^3 = \Delta(\Delta x^3)$$
$$= \Delta(3x^2 + 3x + 1)$$
$$= 3\Delta x^2 + 3\Delta x + \Delta(1)$$
$$= 3\left[(x+1)^2 - x^2\right] + 3\left[(x+1) - x\right] + 0$$
$$= 3(x^2 + 2x + 1 - x^2) + 3(1)$$
$$= 3(2x+1) + 3$$
$$= 6x + 3 + 3 = 6x + 6$$
$$= 6(x+1)$$

Now $E x^2 = (x+1)^2$

$$\frac{\Delta^2 x^3}{E x^2} = \frac{6(x+1)}{(x+1)^2} = \frac{6}{x+1}$$

3. If $U_0 = 1$, $U_1 = 5$, $U_2 = 8$, $U_3 = 3$, $U_4 = 7$, $U_5 = 0$ find $\Delta^5 U_0$

$\rightarrow$

Soln :-

$$\Delta^5 U_0 = (E-1)^5 U_0$$
$$= \left[E^5 - 5C_1 E^4 + 5C_2 E^3 - 5C_3 E^2 + 5C_4 E - 5C_5\right] U_0$$
$$= \left[E^5 - 5E^4 + \frac{5\times 4}{1\times 2} E^3 - \frac{5\times 4\times 3}{1\times 2\times 3} E^2 + \frac{5\times 4\times 3\times 2}{1\times 2\times 3\times 4} E - 1\right] U_0$$
$$= \left[E^5 - 5E^4 + 10E^3 - 10E^2 + 5E - 1\right] U_0$$
$$= U_5 - 5U_4 + 10U_3 - 10U_2 + 5U_1 - U_0$$
$$= 0 - 5(7) + 10(3) - 10(8) + 5(5) - 1$$
$$= -61$$

$\Rightarrow a = 14.25 \qquad b = 29.5$

Given that $U_1 + U_2 + U_3 = 25$, $U_4 = 29$, $U_5 + U_6 = 113$. find the polynomial U_x and hence find U_{10}.

$\rightarrow$

let $U_x = ax^2 + bx + c$. [$\because$ 3 values are given U_x is a polyn of degree 2].

$U_1 = a + b + c$

$U_2 = a(2^2) + b(2) + c$
$= 4a + 2b + c$

$U_3 = a(3^2) + b(3) + c$
$= 9a + 3b + c$

Given, $U_1 + U_2 + U_3 = 25$

$(a+b+c) + (4a+2b+c) + (9a+3b+c) = 25$

$14a + 6b + 3c = 25 \longrightarrow (1)$

$U_4 = a(4^2) + b(4) + c = 29$

$\Rightarrow 16a + 4b + c = 29 \longrightarrow (2)$

$U_5 = 25a + 5b + c$

$U_6 = 36a + 6b + c$

$U_5 + U_6 = 113$ (Given)

$(25a + 5b + c) + (36a + 6b + c) = 113$

$61a + 11b + 2c = 113 \longrightarrow (3)$

Solving (1), (2) and (3), we've

$a = 2,\ b = -1,\ c = 1$ [in calculator eqn. mode unknowns = 3]

$\therefore U_x = 2x^2 - x + 1$

Put $x = 10$

$U_{10} = 2(10^2) - 10 + 1$
$= 200 - 10 + 1 = 191$

Newton's formula:-

Newton's forward interpolation formula

1. Newton's

$$U_{a+rh} = U_a + \frac{r}{1!}\Delta U_a + \frac{r(r-1)}{2!}\Delta^2 U_{a} + \dots\dots + \frac{r(r-1)\dots\dots(r-(n-1))}{n!}\Delta^n U_a$$

where $r = \frac{x - x_a}{h}$. This formula is applied to equal intervals.

2. Newton's **Backward** interpolation formula

(Equal intervals)

$$U_{a+nh+rh} = U_{a+nh} + \frac{r}{1!}\nabla U_{a+nh} + \frac{r(r+1)}{2!}\nabla^2 U_{a+nh} + \dots\dots + \frac{r(r+1)\dots\dots(r+(n-1))}{n!}\nabla^n U_{a+nh}$$

Where $r = \frac{x - x_{a+nh}}{h}$

Problems:-

1. Using Newton's formula find U_x for the 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

$\Rightarrow$ Form the difference table

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				

$a = 0 \quad h = 1$

$$U_x = U_a + \frac{(x-a)}{1!}\Delta U_a + \frac{(x-a)(x-a-h)}{2!}\Delta^2 U_a$$

$$+ \frac{(x-a)(x-a-h)(x-a-2h)}{3!}\Delta^3 U_a + \dots\dots$$

$$= 1 + \frac{(x-0)}{1!}(10) + \frac{(x-0)(x-1)}{2!}(0)$$

$$+ \frac{(x-0)(x-1)(x-2)}{3!}(-3) + \frac{(x-0)(x-1)(x-2)(x-3)\cdot 0}{4!}$$

$$= 1 + 10x + \frac{x(x-1)}{2}(0) + \frac{x(x-1)(x-2)}{\cancel{6}\,2}(-\cancel{3})$$

$$= \frac{1}{2}\left(2 + 20x - x^3 + 3x^2 - 2x\right)$$

$$= \frac{1}{2}\left(-x^3 + 3x^2 + 18x + 2\right) \longrightarrow (1)$$

Put $x = 1.5$ in (1)

(i) $U_{1.5} = \frac{1}{2}\left[-(1.5)^3 + 3(1.5)^2 + 18(1.5) + 2\right]$

$= 16.188$

Put $x = 9$ in (1)

(ii) $U_9 = \frac{1}{2}\left[-(9)^3 + 3(9)^2 + 18(9) + 2\right]$

$= -161$

2. If $U_{75} = 246$, $U_{80} = 202$, $U_{85} = 118$, $U_{90} = 40$ find U_{79} -

x	75	80	85	90
U_x	246	202	118	40

$h = 5$

Equal interva

Soln

x	U_x	ΔU_x	$\Delta^2 U_x$	$\Delta^3 U_x$
75	246			
		-44		
80	202		-40	
		-84		46
85	118		6	
		-78		
90	40			

To find U_{79}:

Since 79 is nearer to beginning of the table, we use Newton's forward difference formula.

$$U_x = U_a + \frac{r}{1!}\Delta U_a + \frac{r(r-1)}{2!}\Delta^2 U_a + \frac{r(r-1)(r-2)}{3!}\Delta^3 U_a$$

where $r = \frac{x-a}{h}$

$r = \frac{79-75}{5}$

$= \frac{4}{5}$

$= 0.8$

Put $x = 79$

$$U_{79} = 246 + \frac{0.8}{1!}(-44) + \frac{(0.8)(0.8-1)}{2!}(-40) + \frac{(0.8)(0.8-1)(0.8-2)}{3!}(46)$$

$= 246 - 35.2 + 3.2 + 1.472$

$= 215.472$

3. The following data gives the melting point of a alloy of lead and zinc, θ is the temperature in degrees centigrades and x is the temperature of lead.

x	40	50	60	70	80	90
θ	184	204	226	250	276	304

Find θ when (i) $x = 42$ (ii) $x = 38$

→

x	θ	$\Delta\theta$	$\Delta^2\theta$	$\Delta^3\theta$	$\Delta^4\theta$	$\Delta^5\theta$
40	184					
		20				
50	204		2			
		22		0		
60	226		2		0	
		24		0		0
70	250		2		0	
		26		0		
80	276		2			
		28				
90	304					

(i) To find θ when $x = 42$.

∵ 42 is nearer to the beginning of the difference table, we use forward difference formula.

$$U_x = U_0 + \frac{r}{1!}\Delta U_0 + \frac{r(r-1)}{2!}\Delta^2 U_0 + \cdots$$

$$r = \frac{x - x_0}{h} = \frac{42-40}{10} = \frac{2}{10} = \frac{1}{5} = 0.2$$

$$U_{42} = 184 + \frac{(0.2)}{1!}(20) + \frac{(0.2)(0.2-1)}{2!}(2) + \cdots$$

$$= 187.84$$

(ii) To find θ when $x = 38$

∵ 38 is nearer to beginning of the difference table, we use Newton's forward difference formula.

$$r = \frac{38-40}{10} = -\frac{2}{10} = -0.2$$

$$U_{38} = 184 + \frac{(-0.2)}{1!}(20) + \frac{(-0.2)(-0.2-1)}{2!}(2) + \cdots$$

$$= 180.24$$

4. The following table gives the census population of a town for the years 1931 - 1971. Estimate the population (i) for the year 1965 by using appropriate interpolation formula.

Year	1931	1941	1951	1961	1971
Population in Lakhs	36	66	81	93	101

→

Year x	population U_x	∇U_x	$\nabla^2 U_x$	$\nabla^3 U_x$	$\nabla^4 U_x$
1931	36				
		30			
1941	66		−15		
		15		12	
1951	81		−3		−13
		12		−1	
1961	93		−4		
		8			
1971	101				

To find U_{1965}

Since 1965 is nearer to end of the table, we use Newton's Backward difference formula,

$$U_{a+4h+rh} = U_{a+4h} + \frac{r}{1!}\nabla U_{a+4h} + \frac{r(r+1)}{2!}\nabla^2 U_{a+4h}$$

$$+ \frac{r(r+1)(r+2)}{3!}\nabla^3 U_{a+4h}$$

$$+ \frac{r(r+1)(r+2)(r+3)}{4!}\nabla^4 U_{a+4h}$$

$$\text{Where } r = \frac{x - x_{a+4h}}{h}$$

$$= \frac{1965 - 1971}{10}$$

$$= \frac{-6}{10} = -0.6$$

$$U_{1965} = 101 + \frac{(-0.6)}{1!}(8) + \frac{(-0.6)(-0.6+1)}{2!}(-4)$$

$$+ \frac{(-0.6)(-0.6+1)(-0.6+2)}{3!}(-1)$$

$$+ \frac{(-0.6)(-0.6+1)(-0.6+2)(-0.6+3)}{4!}(-13)$$

$$= 101 - 4.8 + 0.48 + 0.056 + 0.4368$$

$$= 97.1728$$

Hw

1. If $\log_{10} 5 = 0.6990$, $\log_{10} 10 = 1$, $\log_{10} 15 = 1.161$ and $\log_{10} 20 = 1.3010$, find $\log_{10} 12$.

Hint :-

x :	5	10	15	20
$U_x = \log_{10} x$:	0.6990	1	1.161	1.3010

$h = 5$

5. From the following data estimate the number of persons whose daily wage is between Rs.4[illegible]

Daily wages	0-20	20-40	40-60	60-80	80-100
No. of Persons	120	145	200	250	150

Soln:-

Wages less than x : 20 40 60 80 100

No. of persons (c.f) : 120 265 465 715 865

x	U_x	ΔU_x	$\Delta^2 U_x$	$\Delta^3 U_x$	$\Delta^4 U_x$
20	120				
		145			
40	265		55		
		200		−5	
60	465		50		−145
		250		−150	
80	715		−100		
		150			
100	865				

Number of persons whose earnings is b/w Rs. 40–50 is got from finding $U_{50} - U_{40}$.

W.K.T $U_{40} = 265$.

∴ Find U_{50} only.

Since 50 is nearer to beginning of the table, we use Newton's forward difference formula,

$$r = \frac{x - x_0}{h} = \frac{50-20}{20} = 3/2 = 1.5$$

$$U_{50} = 120 + \frac{1.5}{1!}(145) + \frac{(1.5)(1.5-1)}{2!} \times (55) + \frac{(1.5)(1.5-1)(1.5-2)}{3!}(-5) + \frac{(1.5)(1.5-1)(1.5-2)(1.5-3)}{4!}(-145)$$

$$= 120 + 217.5 + 20.625 + 0.3125 - 3.3984$$

$$= 355 \text{ (approximately)}$$

∴ Number of persons whose earnings is between Rs. 40–50 $= U_{50} - U_{40}$

$$= 355 - 265$$

$$= 90.$$

Lagrange's Formula :- (For both equal and unequal intervals)

$$\phi(x) = \frac{(x-x_1)(x-x_2)\cdots(x-x_n)}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_n)} \cdot y_0 + \frac{(x-x_0)(x-x_2)\cdots(x-x_n)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)} \cdot y_1 + \cdots + \frac{(x-x_0)(x-x_1)\cdots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\cdots(x_n-x_{n-1})} \cdot y_n$$

Problems:-

1. Find U_5 given that $U_1 = 4$, $U_2 = 7$, $U_4 = 13$ and $U_7 = 30$.

[Here x is n... at unequal interval]

x	1	2	4	7
U_x	4	7	13	30

Here $x_0 = 1$, $x_1 = 2$, $x_2 = 4$, $x_3 = 7$

$y_0 = U_{x_0} = 4$, $U_{x_1} = 7 = y_1$, $U_{x_2} = 13 = y_2$, $U_{x_3} = 30 = y_3$

By Lagrange's formula,

$$U_x = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)} \times y_0 + \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)} \times y_1 + \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)} \times y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)} \times y_3$$

Here $x = 5$

$$U_5 = \frac{(5-2)(5-4)(5-7)}{(1-2)(1-4)(1-7)}(4) + \frac{(5-1)(5-4)(5-7)}{(2-1)(2-4)(2-7)}(7) + \frac{(5-1)(5-2)(5-7)}{(4-1)(4-2)(4-7)}(13) + \frac{(5-1)(5-2)(5-4)}{(7-1)(7-2)(7-4)}(30)$$

$$= \frac{(3)(1)(-2)}{(-1)(-3)(-6)}(4) + \frac{(4)(1)(-2)}{(1)(-2)(-5)}(7) + \frac{(4)(3)(-2)}{(3)(2)(-3)}(13) + \frac{(4)(3)(1)}{(6)(5)(3)}(30)$$

$$= \frac{4}{3} - \frac{28}{5} + \frac{52}{3} + 4 = 17.06$$

$$\boxed{U_5 = 17.06}$$

m.Rs.

Ans : 61.5

Theory of Attributes :-

The qualitative characteristics of a population are called attributes. It cannot be measured by numeric quantities.

The positive class denote the presence of the attribute and it is denoted by A, B, C, D...

The negative class denote the absence of the attribute and it is denoted by $\alpha, \beta, \gamma, \delta, \dots$

2×2 Contingency table

Attribute	B	β	Total
A	(AB)	(Aβ)	(A)
α	(αB)	($\alpha\beta$)	(α)
Total	(B)	(β)	

N denotes the total number in the population.

The following table gives the class frequencies of all orders and the total number of all class frequencies upto 3 attributes.

Order	Attributes	Class frequencies of all orders	No. in each order	Total No.
0		N		1
0	A	N	1	3
1		$(A), (\alpha)$	2	
0	A, B	N	1	9
1		$(A), (B), (\alpha), (\beta)$	4	
2		$(AB), (A\beta), (\alpha B), (\alpha\beta)$	4	
0	A, B, C	N	1	27
1		$(A), (B), (\alpha), (\beta), (C), (\gamma)$	6	
2		$(AB), (A\beta), (\alpha B), (\alpha\beta), (AC), (A\gamma), (\alpha C), (\alpha\gamma), (BC), (B\gamma), (\beta C), (\beta\gamma)$	13	
3		$(ABC), (AB\gamma), (A\beta C), (A\beta\gamma), (\alpha BC), (\alpha B\gamma), (\alpha\beta C), (\alpha\beta\gamma)$	8	

The classes of highest order are called the ultimate classes and their frequencies are called the ultimate class frequencies.

<u>Example</u>

1. $(AB) = (ABC) + (AB\gamma)$

$$\rightarrow (AB\gamma) = AB\gamma \cdot N$$
$$= AB(1 - C) \cdot N$$
$$= AB \cdot N - ABC \cdot N$$
$$= (AB) - (ABC)$$
$$\Rightarrow (AB) = (ABC) + (AB\gamma).$$

1. **Note** For two attributes A, ...

$N = (A) + (\alpha) = (B) + (\beta)$.

For three attributes A, B, C, $N = (A) + (\ldots$

$N = (AB) + (A\beta) + (\alpha B) + (\alpha\beta)$

$\therefore N = (ABC) + (AB\gamma) + (A\beta C) + (A\beta\gamma) + (\alpha BC)$
$\quad + (\alpha B\gamma) + (\alpha\beta C) + (\alpha\beta\gamma)$

2. **Note**

$\therefore \alpha = 1 - A \qquad \beta = 1 - B \qquad \gamma = 1 - C$

Problems

1. Given $(A) = 30$, $(B) = 25$, $(\alpha) = 30$, $(\alpha\beta) = 20$.
Find (i) N (ii) (β) (iii) (AB) (iv) $(A\beta)$ (v) (αB).

$\rightarrow$

(i) $N = (A) + (\alpha)$
$= 30 + 30 = 60$

(ii) $(\beta) = (A\beta) + (\alpha\beta)$
Also $(\beta) = N - (B)$
$= 60 - 25$
$= 35$

(iii) $(AB) = AB \cdot N$
$= (1-\alpha)(1-\beta) \cdot N$
$= N - (\alpha) - (\beta) + (\alpha\beta)$
$= 60 - 30 - 35 + 20 = 15$

(iv) $(A\beta) = A\beta \cdot N$
$= A(1-B) \cdot N = A \cdot N - AB \cdot N$
$= (A) - (AB)$
$= 30 - 15 = 15$

(v) $(\alpha B) = \alpha B \cdot N = (1-A) B \cdot N$
$= B \cdot N - AB \cdot N = (B) - (AB)$
$= 25 - 15 = 10$.

2. Given the following ultimate class frequencies of 2 attributes A and B, find the frequencies of positive and negative class frequencies and the total number of observation.

$(AB) = 975 \quad (\alpha B) = 100 \quad (A\beta) = 25$
$(\alpha\beta) = 950$

→ **Soln:-**

Positive class frequencies are (A) & (B)

$(A) = (AB) + (A\beta) = 975 + 25 = 1000$

$(B) = (AB) + (\alpha B) = 975 + 100 = 1075$

Negative class frequencies are $(\alpha) + (\beta)$

$(\alpha) = (\beta\alpha) + (\alpha B) = 950 + 100 = 1050$

$(\beta) = (A\beta) + (\alpha\beta) = 25 + 950 = 975$

$N = (A) + (\alpha) = 1000 + 1050 = 2050$

$= (B) + (\beta) = 1075 + 975 = 2050$

3. Given the following positive class frequencies. Find the remaining class frequencies

$N = 20, \quad (A) = 9, \quad (B) = 12, \quad (C) = 8$
$(AB) = 6, \quad (BC) = 4, \quad (CA) = 4, \quad (ABC) = 3$

⟹ There are 3 attributes A, B, C.

The total number of class frequencies is $3^3 = 27$

we are given only 8 class frequencies

To find the remaining 19 class frequencies

Order 1

$(\alpha) = N - (A) = 20 - 9 = 11$

$(\beta) = N - (B) = 20 - 12 = 8$

$(\gamma) = N - (C) = 20 - 8 = 12$

Order 2

$(A\beta) = A(1-B)\cdot N = (A) - (AB) = 9-6 = 3$

$(\alpha B) = (1-A)B\cdot N = (B) - (AB) = 12-6 = 6$

$(A\gamma) = A(1-C)\cdot N = (A) - (AC) = 9-4 = 5$

$(\alpha C) = (1-A)C\cdot N = (C) - (AC) = 8-4 = 4$

$(B\gamma) = B(1-C)\cdot N = (B) - (BC) = 12-4 = 8$

$(\beta C) = (1-B)C\cdot N = (C) - (BC) = 8-4 = 4$

$(\alpha\beta) = (1-A)(1-B)\cdot N$

$= N - (A) - (B) + (AB)$

$= 20 - 9 - 12 + 6 = 5$

$(\beta\gamma) = (1-B)(1-C)\cdot N$

$= N - (B) - (C) + (BC)$

$= 20 - 12 - 8 + 4 = 4$

$(\alpha\gamma) = (1-A)(1-C)\cdot N$

$= N - (A) - (C) + (AC)$

$= 20 - 9 - 8 + 4 = 7$

Order 3

$(AB\gamma) = AB(1-C)\cdot N = (AB) - (ABC) = 6-3 = 3$

$(A\beta C) = A(1-B)C\cdot N = (AC) - (ABC) = 4-3 = 1$

$(A\beta\gamma) = A(1-B)(1-C)\cdot N$

$= (A) - (AC) - (AB) + (ABC)$

$= 9 - 4 - 6 + 3 = 2$

$(\alpha BC) = (1-A)BC\cdot N$

$= (BC) - (ABC) = 4 - 3 = 1$

$(\alpha B\gamma) = (1-A)B(1-C)\cdot N$

$= (B) - (BC) - (AB) + (ABC)$

$= 12 - 4 - 6 + 3 = 5$

$(\alpha\beta C) = (1-A)(1-B)C\cdot N$

$= (C) - (AC) - (BC) + (ABC)$

$= 8 - 4 - 4 + 3 = 3$

$(\alpha\beta\gamma) = (1-A)(1-B)(1-C)\cdot N$

$= N - (A) - (B) - (C) + (AB) + (BC) + (CA) - (ABC)$

$= 20 - 9 - 12 - 8 + 6 + 4 + 4 - 3 = 2$

atleast two semesters is 83.

Consistency of Data :-

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.

1. Find whether the following data are consistent.

$N = 600$, $(A) = 300$, $(B) = 400$, $(AB) = 50$.

$\rightarrow$

We calculate the ultimate class frequencies $(\alpha\beta)$, (αB) and $(A\beta)$.

$$(\alpha\beta) = \alpha\beta \cdot N = (1-A)(1-B) \cdot N = N - (A) - (B) + (AB)$$
$$= 600 - 300 - 400 + 50$$
$$= -50.$$

Since $(\alpha\beta) < 0$, the data are inconsistent.

2. Show that there is some error in the following data. 50% of people are wealthy and healthy, 35% of people are wealthy but not healthy, 20% are healthy but not wealthy.

$\Longrightarrow$

Sol:- A → [illegible]

B → [illegible]

$N = 100$ $(AB) = 50$ $(A\beta) = 20$ $(\alpha B) = 35$

For consistency,

we find $(\alpha\beta)$.

$(\alpha\beta) = \alpha\beta . N = ((1-A)(1-B)).N$

$= N - (A) - (B) + (AB)$.

But $(A) = (AB) + (A\beta) = 50 + 20 = 70$

$(B) = (AB) + (\alpha B) = 50 + 35 = 85$

$\therefore (\alpha\beta) = 100 - 85 - 70 + 50 = -5$

< 0.

Hence there is error in the data.

Consistency Table

Attributes	Condition of consistency	Equivalent +ve class conditions	No. of conditions
A	$(A) \geq 0$	$(A) \geq 0$	2
	$(\alpha) \geq 0$	$(A) \leq N$	
A, B	$(AB) \geq 0$	$(AB) \geq 0$	2^2
	$(A\beta) \geq 0$	$(AB) \leq (A)$	
	$(\alpha B) \geq 0$	$(AB) \leq (B)$	
	$(\alpha\beta) \geq 0$	$(AB) \geq (A) + (B) - N$	
A, B, C	$(ABC) \geq 0$	(i) $(ABC) \geq 0$	2^3
	$(AB\gamma) \geq 0$	(ii) $(ABC) \leq (AB)$	
	$(A\beta C) \geq 0$	(iii) $(ABC) \leq (AC)$	
	$(\alpha BC) \geq 0$	(iv) $(ABC) \leq (BC)$	
	$(A\beta\gamma) \geq 0$	(v) $(ABC) \geq (AB) + (AC) - (A)$	
	$(\alpha B\gamma) \geq 0$	(vi) $(ABC) \geq (AB) + (BC) - (B)$	
	$(\alpha\beta C) \geq 0$	(vii) $(ABC) \geq (AC) + (BC) - (C)$	
	$(\alpha\beta\gamma) \geq 0$	(viii) $(ABC) \leq (AB) + (BC) + (AC) - (A) - (B) - (C) + N$	

Note.

(ix). $(AB) + (BC) + (AC) \geq (A) + (B) + (C) - N$

(x). $(AC) + (BC) - (AB) \leq (C)$

(xi). $(AB) + (BC) - (AC) \leq (B)$

(xii). $(AB) + (AC) - (BC) \leq (A)$.

3. Find the limits of (BC) for the following available data. $N = 125$ $(A) = 48$ $(B) = 62$ $(C) = 45$, $(A\beta) = 7$ and $(A\gamma) = 18$.

→

First of all we find (AB) and (AC).

$(AB) = (A) - (A\beta) = 48 - 7 = 41$

$(AC) = (A) - (A\gamma) = 48 - 18 = 30$

By (ix) condition,

$(AB) + (BC) + (AC) \geq (A) + (B) + (C) - N$

$\Rightarrow 41 + (BC) + 30 \geq 48 + 62 + 45 - 125$

$\therefore (BC) \geq -41 \longrightarrow$ ①

Also using (xii), $(AB) + (AC) - (BC) \leq (A)$

$\Rightarrow (BC) \geq (AB) + (AC) - (A)$

$= 41 + 30 - 48 = 23$

$\therefore (BC) \geq 23 \longrightarrow$ ②

using (xi), $(AB) + (BC) - (AC) \leq (B)$.

$\Rightarrow (BC) \leq (B) + (AC) - (AB)$

$= 62 + 30 - 41 = 51$

$\therefore (BC) \leq 51 \longrightarrow$ ③

using (x), $(AC) + (BC) - (AB) \leq (C)$.

$\Rightarrow (BC) \leq (C) + (AB) - (AC)$

$= 45 + 41 - 30 = 56$.

$\therefore (BC) \leq 56 \longrightarrow$ ④

From ①, ②, ③ + ④, $23 \leq (BC) \leq 56$.

Independence and Association of D.A. :-

1) A and B are independent iff

$$(AB) = \frac{(A)(B)}{N}$$

$$(A\beta) = \frac{(A)(\beta)}{N}$$

$$(\alpha\beta) = \frac{(\alpha)(\beta)}{N}$$

$$(\alpha B) = \frac{(\alpha)(B)}{N}.$$

2) A and B are independent if

$$(AB)(\alpha\beta) - (A\beta)(\alpha B) = 0$$

Association

If $(AB) = \frac{(A)(B)}{N}$, we say that A and B are associated.

If $(AB) > \frac{(A)(B)}{N}$, we say that A and B are positively associated.

If $(AB) < \frac{(A)(B)}{N}$, we say that A and B are negatively associated.

Coefficient of Association :-

Yule's coefficient of association

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

Coefficient of colligation

$$Y = \frac{1 - \sqrt{\frac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}}{1 + \sqrt{\frac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}}$$

Note $\delta = \frac{1}{N}\left[N(AB) - (A)(B)\right]$.

$= (AB) - \frac{(A)(B)}{N}$.

A and B are independent if $Q = Y = 0$.

and $\delta = 0$.

1. Check whether the attributes A and B are independent given that

1) $(A) = 30$, $(B) = 60$, $(AB) = 12$, $N = 150$.

2) $(AB) = 256$, $(\alpha B) = 768$, $(A\beta) = 48$, $(\alpha\beta) = 144$

→

1) Since the given class frequencies are of first order, the condition for independence is $(AB) = \frac{(A)(B)}{N}$.

$$\frac{(A)(B)}{N} = \frac{30 \times 60}{150} = 12 = (AB).$$

∴ A and B are independent.

2) $(A) = (AB) + (A\beta) = 256 + 48 = 304$

$(B) = (AB) + (\alpha B) = 256 + 768 = 1024$

$(\alpha) = (\alpha B) + (\alpha\beta) = 768 + 144 = 912$

$(\beta) = (\alpha\beta) + (A\beta) = 144 + 48 = 192$.

$N = (A) + (\alpha) = 304 + 912 = 1216$.

$= (B) + (\beta) = 1024 + 192 = 1216$.

Now $\frac{(A)(B)}{N} = \frac{304 \times 1024}{1216} = 256 = (AB)$.

∴ A and B are independent.

Aliter method:

$(AB)(\alpha\beta) - (A\beta)(\alpha B) = (256 \times 144) - (768 \times 48)$

$= 0$.

∴ A and B are independent.

2. In a class test in which 135 candidates were examined into proficiency in physics and chemistry, it was discovered that 75 students failed in physics, 90 failed in chemistry and 50 failed in both. Find the magnitude of association and state if there is any association between failing in physics and chemistry.

$\longrightarrow$

A → fail in physics

B → fail in chemistry

$(A) = 75 \quad (B) = 90 \quad (AB) = 50 \quad N = 135$

Magnitude of association is measured by

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

Now

$(\alpha) = N - (A) = 135 - 75 = 60$

$(\beta) = N - (B) = 135 - 90 = 45$

$(\alpha B) = (B) - (AB) = 90 - 50 = 40$

$(A\beta) = (A) - (AB) = 75 - 50 = 25$

$(\alpha\beta) = (\alpha) - (\alpha B) = 60 - 40 = 20$

$$\therefore Q = \frac{(50 \times 20) - (25 \times 40)}{(50 \times 20) + (25 \times 40)} = 0$$

$\therefore$ A and B are independent. Hence failure in Physics and chemistry are completely independent of each other.

3. S.T whether A and B are independent or positively or negatively associated in the following data.

(i) $N = 930 \quad (A) = 300 \quad (B) = 400 \quad (AB) = 2$[illegible]

(ii) $(AB) = 327$, $(A\beta) = 545$, $(\alpha B) = 741$, $(\alpha\beta) = 235$

(iii) $(AB) = 66$, $(A\beta) = 88$, $(\alpha B) = 102$, $(\alpha\beta) = 136$

→

(i) $\dfrac{(A)(B)}{N} = \dfrac{300 \times 400}{930} = 129.03$

Now $\delta = (AB) - \dfrac{(A)(B)}{N} = 230 - 129.03 = 100.97$

∴ $\delta > 0$, A and B are positively associated.

(ii) (coeff. of association) $Q = \dfrac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$

$= \dfrac{(327 \times 235) - (545 \times 741)}{(327 \times 235) + (545 \times 741)}$

$= -0.6803$

< 0

∴ $Q < 0$ Hence A and B are negatively ~~indepen~~ associated. (or) ($\delta < 0$, negatively associated).

(iii) $Q = \dfrac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$

$= \dfrac{(66 \times 136) - (88 \times 102)}{(66 \times 136) + (88 \times 102)} = 0$

∴ A and B are independent.

4. Investigate from the following data between inoculation against small pox and prevention from attack?

	Attacked	Not attacked	Total
Inoculated	25	220	245
Not inoculated	90	160	250
Total	115	380	495

→

A → inoculated

B → attacked

$\therefore (AB) = 25$, $(A\beta) = 220$, $(\alpha B) = 90$

$(\alpha\beta) = 160$

$$\therefore Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) - (A\beta)(\alpha B)}$$

$$= \frac{(25 \times 160) - (220 \times 90)}{(25 \times 160) + (220 \times 90)}$$

$$= -0.6638$$

$\therefore$ A & B have negative association. Thus inoculation against small pox can be taken as the preventive measure.

UNIT-V - INDEX NUMBERS

An index number is a widely used statistical device for comparing the level of a certain phenomenon with the level of same phenomenon at some standard period.

For example, To compare the price of a food article at a particular period with the price of the same article at a previous period of time.

They are classified into two types

(i) Simple index Number

(ii) weighted index number

Two standard methods of computations are

(i) Aggregate method

(ii) Average of price relative method.

Aggregate method

$$P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

where Σp_1 is total of the current year

Σp_0 is the total of the base year.

Average of price relative method:-

Index number of for the current year is $P_{01} = \frac{p_1}{p_0} \times 100$ where $\frac{p_1}{p_0}$ is called price relatives.

1) Arithmetic mean index number

$$P_{01} = \frac{\sum\left(\frac{p_1}{p_0}\right) \times 100}{n}$$

2) Geometric mean index number

$$P_{01} = \left[\Pi\left(\frac{p_1}{p_0}\right)\right]^{1/n} \times 100$$

where Π denotes the product.

$$\log P_{01} = \frac{\sum\left[\log\left(\frac{p_1}{p_0}\right) \times 100\right]}{n}$$

Problems

1. From the following data of the whole sale price of rice for the 5 years construct the index numbers using (i) 1987 as the base (ii) 1990 as the base.

Years:	1987	1988	1989	1990	1991	1992
Price of rice per kg	5.00	6.00	6.50	7.00	7.50	8.00

$\Rightarrow$

(i) 1987 as the base year.

Years	price of rice per kg	Index Numbers (1987 base)
1987	5.00	$100 = \frac{5}{5} \times 100$
1988	6.00	$\frac{6}{5} \times 100 = 120$
1989	6.50	$\frac{6.5}{5} \times 100 = 130$
1990	7.00	$\frac{7}{5} \times 100 = 140$
1991	7.50	$\frac{7.5}{5} \times 100 = 150$
1992	8.00	$\frac{8}{5} \times 100 = 160$

1) 1990 as base year.

Years	price of rice per kg	Index Numbers (1990 Base)
1987	5	$\frac{5}{7} \times 100 = 71.4$
1988	6	$\frac{6}{7} \times 100 = 85.7$
1989	6.5	$\frac{6.5}{7} \times 100 = 92.7$
1990	7	$\frac{7}{7} \times 100 = 100$
1991	7.5	$\frac{7.5}{7} \times 100 = 107.1$
1992	8	$\frac{8}{7} \times 100 = 114.3$

2. Construct the whole sale price index number for 1991 and 1992 from the data given below using 1990 as the base year.

Commodity	Whole Sale price in Rupees per quintal 1990	1991	1992
Rice	700	750	825
Wheat	540	575	600
Ragi	300	325	310
Cholam	250	280	295
Flour	320	330	335
Ravai	325	350	360

→ Taking 1990 as base year.

Commodity	1990 P_0	1991 P_1	1992 P_2	Relatives for 1991	Relatives for 1992
Rice	700	750	825	$\frac{750}{700} \times 100 = 107.1$	$\frac{825}{700} \times 100 = 117.[illegible]$
Wheat	540	575	600	$\frac{575}{540} \times 100 = 106.5$	$\frac{600}{540} \times 100 = 111.1$
Ragi	300	325	310	$\frac{325}{300} \times 100 = 108.3$	$\frac{310}{300} \times 100 = 103.3$
Cholam	250	280	295	$\frac{280}{250} \times 100 = 112$	$\frac{295}{250} \times 100 = 118$
Flour	320	330	335	$\frac{330}{320} \times 100 = 103.1$	$\frac{325}{320} \times 100 = 101.5$
Ravai	325	350	360	$\frac{350}{325} \times 100 = 107.7$	$\frac{360}{325} \times 100 = 110.8$
				644.7	

$\div$ by 6

Index No. for 1991 as base year 1990 is 107.5

Index No. for 1992 as base year 1990 is 110.5

3. From the following average prices of the 3 groups of commodities given in rupees per unit find (i) fixed base index number (ii) chain base index number with 1988 as the base year and

Commodity	1988	1989	1990	1991	1992
A	2	3	4	5	6
B	8	10	12	15	18
C	4	5	8	10	12

→

(i) <u>fixed base index number</u> :-

Commodity	1988	1989	1990	1991	1992
A	100	$\frac{3}{2}\times100=150$	$\frac{4}{2}\times100=200$	$\frac{5}{2}\times100=250$	$\frac{6}{2}\times100=300$
B	100	$\frac{10}{8}\times100=125$	$\frac{12}{8}\times100=150$	$\frac{15}{8}\times100=188$	$\frac{18}{8}\times100=225$
C	100	$\frac{5}{4}\times100=125$	$\frac{8}{4}\times100=200$	$\frac{10}{4}\times100=200$	$\frac{12}{4}\times100=300$
Total	300	400	550	688	825
Index No. (A.M)	100	133.3	183.3	229.3	275

(ii) <u>chain base index number</u> :-

Comm	1988	1989	1990	1991	1992
A	$\frac{2}{2}\times100=100$	$\frac{3}{2}\times100=150$	$\frac{4}{3}\times100=133.3$	$\frac{5}{4}\times100=125$	$\frac{6}{5}\times100=120$
B	$\frac{8}{8}\times100=100$	$\frac{10}{8}\times100=125$	$\frac{12}{10}\times100=120$	$\frac{15}{12}\times100=125$	$\frac{18}{15}\times100=120$
C	$\frac{4}{4}\times100=100$	$\frac{5}{4}\times100=125$	$\frac{8}{5}\times100=160$	$\frac{10}{8}\times100=125$	$\frac{12}{10}\times100=120$
Total	300	400	413.3	375	360
Index Number (A.M)	100	133.3	137.3	125	120

<u>Weighted index Numbers</u> :-

(i) <u>Weighted aggregative method</u> :-

1) Laspeyre's index Number

$$L\,I_{01} = \frac{\sum P_1 q_0}{\sum P_0 q_0} \times 100$$

27

2) Paasche's Index Number

$$PI_{01} = \frac{\sum p_1 q_0}{\sum p_0 q_1} \times 100$$

3) Marshall - Edgeworth's Index Number

$$MI_{01} = \left(\frac{\sum p_1 q_0 + \sum p_1 q_1}{\sum p_0 q_0 + \sum p_0 q_1} \right) \times 100$$

4) Bowley's Index Number

$$BI_{01} = \frac{LI_{01} + PI_{01}}{2}$$

$$= \frac{1}{2} \left[\frac{\sum p_1 q_0}{\sum p_0 q_0} + \frac{\sum p_1 q_1}{\sum p_0 q_1} \right] \times 100$$

5) Fisher's Index Number

$$FI_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$$

$$= \sqrt{LI_{01} \times PI_{01}}$$

6) Kelley's Index Number

$$KI_{01} = \frac{\sum p_1 q}{\sum p_0 q} \times 100$$ where

q is the average quantity of two or more years.

1. Calculate (i) Laspeyre's (ii) Paasche's

Weighted Average of Price [illegible]

$P_{01} = \frac{\sum PV}{\sum V}$ where P is the price relative. V is the value weight $p_0 q_0$.

Ideal Index Number :-

An index number is said to be ideal index number if it is satisfying the following three tests

(i) time reversal test.

(ii) factor reversal test.

(iii) commodity reversal test.

Time reversal test :-

$$I_{01} \times I_{10} = I$$

Factor reversal test :-

$$I_{pq} \times I_{qp} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

Where I_{pq} is the price index of current year relative to base year

I_{qp} is the quantity index of the current year relative to base year.

Commodity Reversal test

This test is satisfied by all index number.

Note

1. Fisher's index number is an ideal index number.

2. Laspeyre's index number does not satisfy the time reversal test.

3. Laspeyre's & Paasche's index number do not satisfy factor reversal tests.

1. Construct with the help of the data given below Fisher index numbers and show that it satisfies both the factor reversal and time reversal test.

Commodity	A	B	C	D
Base year price in Rs.	5	6	4	3
Base year quantity in Quintals	50	40	120	30
Current year price	7	8	5	4
Current year quantity	60	50	140	35

→

Soln :-

Commodity	Base year p_0	q_0	Current year p_1	q_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
A	5	50	7	60	250	300	350	420
B	6	40	8	50	240	300	320	400
C	4	120	5	110	480	440	600	550
D	3	30	4	35	90	105	120	140
					1060	1145	1390	1510

Fisher's index number

$$I_{01} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}} \times 100$$

$$= \sqrt{\frac{1390}{106} \times \frac{1510}{1145}} \times 100$$

Time Reversal test :-

$$I_{01} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}} = \sqrt{\frac{1390}{1060} \times \frac{1510}{1145}}$$

$$I_{10} = \sqrt{\frac{\sum p_0q_1}{\sum p_1q_1} \times \frac{\sum p_0q_0}{\sum p_1q_0}} = \sqrt{\frac{1145}{1510} \times \frac{1060}{1390}}$$

$\therefore I_{01} \times I_{10} = 1$

Factor Reversal test :-

$$I_{pq} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}} = \sqrt{\frac{1390}{106} \times \frac{1510}{1145}}$$

Interchanging $p \to q$

$$I_{qp} = \sqrt{\frac{\sum q_1p_0}{\sum q_0p_0} \times \frac{\sum q_1p_1}{\sum q_0p_1}} = \sqrt{\frac{1145}{106} \times \frac{1510}{1390}}$$

$$I_{pq} \times I_{qp} = \frac{\sum p_1q_1}{\sum p_0q_0} = \frac{1510}{1060}$$

Consumer price Index Numbers
(Cost of living index numbers)

1) Aggregate expenditure method :-

$$I_{01} = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times 100$$

ii) Family budget method :-

$$I_{01} = \frac{\Sigma PV}{\Sigma V} \;;\; P = \frac{P_1}{P_0} \times 100$$

V = Value weight B.

$$= \frac{\Sigma PW}{\Sigma W} \quad \text{''} \quad \text{''}$$

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.

Commodity	Price in Rs. 1991	Price in Rs. 1992	Quantity in Quintals in 1991
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 :-

Commodities	P_0	P_1	q_0	P_0q_0 V	P $\frac{P_1}{P_0} \times 100$	PV
Rice	7	7.5	6	42	107.1	4498.2
Wheat	6	6.75	3.5	21	112.5	2362.5
flour	5	5	0.5	2.5	100	250
Oil	30	32	3	90	106.7	9603
Sugar	8	8.5	1	8	106.3	850.4
Total				163.5		17564.1

cost of living index $= \frac{\Sigma pV}{\Sigma V} = \frac{17564.1}{161.5}$

$= 107.4$

(ii) Aggregate expenditure method :-

Commodities	p_0	p_1	q_0	p_0q_0	p_1q_0
Rice	7	7.5	6	42	45
Wheat	6	6.75	3.5	21	23.6
Flour	5	5	0.5	2.5	2.5
Oil	30	32	3	90	96
Sugar	8	8.5	1	8	8.5
				163.5	175.6

cost of living index $= \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$

$= \frac{175.6}{163.5} \times 100$

$= 107.4$

d. An enquiry into the budgets of the middle class ~~frequencies~~ families in a city in India gave the following information.

	Food	Rent	clothing	Fuel	Misc
weights	35%	15%	20%	10%	20%
Prices 1991	1500	3000	450	70	500
Prices 1992	1650	3250	500	90	550

What changes in cost of living index of 1992 as compared with that of 1991 are seen?

→ Base year is chooses as 1991 (=100)

Items	prices 1991	prices 1992	Index No. 1992 $P = \frac{P_1}{P_0} \times 100$	W	PW
food	1500	1650	$\frac{1650}{1500} \times 100 = 110$	35	3850
Rent	300	325	108.3	15	1624.5
clothing	450	500	111.1	20	2222
Fuel	70	90	128.6	10	1286
Misc.	500	550	110	20	2200
				100	11182.5

$$\text{Cost of living index} = \frac{\Sigma PW}{\Sigma W}$$

$$= \frac{11182.5}{100}$$

$$= 111.8$$

The price in 1992 compared with the prices in 1991 has risen to 11.8%.

3. Find the cost of living index for the following data in a middle class family.

Items	Price 1991	Price 1992	weight
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

→

Taking 1991 as base year.

Items	Price 1991	Price 1992	Index No. for 1992 P	W	PW
Fuel	700	850	$\frac{850}{700}\times100 = 121.4$	40	4856
Clothing	300	280	$\frac{280}{300}\times100 = 93.3$	15	1399.5
Rent	200	225	$\frac{225}{200}\times100 = 112.5$	7	787.5
Fuel	70	82	117.1	5	585.5
Medicine	100	135	135	9	1215
Education	500	550	110	12	1320
Entertainment	100	90	90	10	900
Misc.	475	425	89.5	23	2058.5
Total				= 121	13122

$\therefore$ Cost of living index No. $= \frac{\Sigma PW}{\Sigma W} = \frac{13122}{121}$

$= 108.4$

Analysis of Time Series :-

Components of Time Series :

1. Secular ~~or~~ trend
2. Periodic movements
3. Irregular fluctuations

Time series is a series of values of a variable over a period of time arranged chronologically.

Measurement of trends :-

1. Graphic method.
2. Method of curve fitting by the principles of least squares.
3. Method of semi average
4. Method of Moving average.

1. Use the method of least squares and fit a straight line trend to the following f... given from 1982 to 1992. Hence estimate the trend value for 1993.

Year	: 82	83	84	85	86	87	88	89	90		
Production	: 45	46	44	47	42	41	39	42	45		

→ Let the line of best fit be $y = ax+b$

Take $X = x-1987$ $\quad Y = y-42$

Then the line of best fit become

$$Y = aX+b.$$

The normal equations are

$$\Sigma XY = a\Sigma X^2 + b\Sigma X$$

$$\Sigma Y = a\Sigma X + nb \qquad \text{where } n=11$$

x	$X = x-1987$	y	$Y = y-42$	XY	X^2
1982	-5	45	3	-15	25
83	-4	46	4	-16	16
84	-3	44	2	-6	9
85	-2	47	3	-10	4
86	-1	42	0	0	1
87	0	41	-1	0	0
88	1	39	-3	-3	1
89	2	42	0	0	4
90	3	45	3	9	9
91	4	40	-2	-8	16
92	5	48	6	30	25
	0		17	-19	110

Normal equs are

$$-19 = a(110) + b(0)$$

$$17 = a(0) + a\,b\ldots$$

$$\Rightarrow 110a = -19$$

$$a = -19/110 = -0.17$$

∴ 17 = Σb

⇒ 17 = [illegible] b 11 (b) [∵ h = given data [illegible]]

$b = \frac{17}{11} = 1.55$

∴ The line of best fit is $y = -0.17x + 1.55$

(ie) $y - 42 = -0.17(x - 1987) + 1.55$

$y = -0.17x + 1987 \times 0.17 + 1.55 + 42$

$\boxed{y = -0.17x + 381.34}$ is the straight line trend.

Trend values :-

When $x = 1982$, $y = 44.4$

" $x = 1983$, $y = 44.23$

" $x = 1984$, $y = 44.06$

" $x = 1985$, $y = 43.89$

" $x = 1986$, $y = 43.72$

" $x = 1987$, $y = 43.55$

" $x = 1988$, $y = 43.38$

" $x = 1989$, $y = 43.21$

" $x = 1990$, $y = 43.04$

" $x = 1991$, $y = 42.87$

" $x = 1992$, $y = 42.7$

2. Calculate the seasonal Variation indices from the following data.

Month	Monthly sales in lakhs I	II	III	IV	Total	$\bar{x}$	Seasonal Variation
Jan	10	11	11.5	13.5	46	11.5	$\frac{11.5}{12} \times 100 = 95.8$
Feb	8.5	8.5	9	10	36	9	$\frac{9}{12} \times 100 = 75$
Mar	10.5	12	11	12.5	46	11.5	$\frac{11.5}{12} \times 100 = 95.8$
Apr	12	14	16	18	60	15	$\frac{15}{12} \times 100 = 125$
May	10	9	12	15	46	11.5	$\frac{11.5}{12} \times 100 = 95.8$

	I	II	III	IV	Total	$\bar{x}_i$	Seasonal Variation
					46	11.5	$\frac{11.5}{12}\times 100 = 96$
June	10.5	10.5	11	14	56	14	$\frac{14}{12}\times 100 = 116$
July	12	14	13	17	44	11	$\frac{11}{12}\times 100 = 91.7$
Aug	9	8	11	16	48	12	$\frac{12}{12}\times 100 = 100$
Sep	11	11	12.5	13.5	44	11	$\frac{11}{12}\times 100 = 91.7$
Oct	10	9.5	10.5	13	48	12	$\frac{12}{12}\times 100 = 100$
Nov	11	12.5	10.5	14	56	14	$\frac{14}{12}\times 100 = 116.7$
Dec	12	13	15	16			
Total					144		
Average					12		

3. Compute the trend values by the method of 4-yearly moving average for the data given in problem

Year	1982	83	84	85	86	87	88	89	90	91	92
Production	45	46	44	47	42	41	39	42	45	40	48

I Year	II production	III 4-yearly moving total	IV 4-yearly moving average	V 2 period moving total	VI Trend values (V/2)
1982	45				
83	46				
		182	45.50		
				90.25	
84	44				
		179	44.75		45.13
85	47				
		174	43.50	88.25	44.13
86	42				
		169	42.25		
				85.75	42.88
87	41				
		164	41.00		
88	39				

					41.63
		167	41.75	83.75	
	42			82.75	41.38
89		166	41.50		
	45			83.25	41.63
90		175	43.75		
				85.65	42.93
	40				
91					
	48				
92					

4. Calculate (i) 3-yearly moving average (ii) Short term fluctuations for the data given in above problem 3.

I Year	II Production	III 3-yearly moving total	IV 3 yearly moving average	V Short term fluctuations II − IV
1982	45	–	–	–
83	46	135	45	1
84	44	137	45.7	−1.7
85	47	133	44.3	2.7
86	42	130	43.3	−1.3
87	41	122	40.7	0.3
88	39	122	40.7	−1.7
89	42	126	42	0
90	45	127	42.3	2.7
91	40	133	44.3	−4.3
92	48	–	–	–

Trend values for the given time series are given in column IV.

Short term fluctuations are given in column V.

5. Compute the seasonal indices for the following data by simple average method.

Seasons	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	58

→ Soln

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	~~662~~ 331	292	296	303
Average	66.2	58.4	59.2	60.6
Seasonal Index	$\frac{66.2}{61.1}\times 100 = 108.31$	$\frac{58.4}{61.1}\times 100 = 95.6$	$\frac{59.2}{61.1}\times 100 = 96.9$	$\frac{60.6}{61.1}\times 100 = 99.2$