Module 2

Page No.
Date 24 8 23

* Measure of Central Tendency *

There are 3 types of measure of central tendency :-

1) Mean (Average)

There are 3 types of mean :-
i) Arithmatic mean (AM)
ii) Geometric mean (GM)
iii) Harmonic mean (HM)

• Arithmatic mean (AM)

i) For non-frequency data :-

Suppose 'X' is a random variable having 'n' values namely $x_1, x_2, \ldots x_n$ then their A.M. is defined as

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

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

ii) For frequency data :-

Suppose 'X' is a random variable having 'n' values namely $x_1, x_2, \ldots, x_n$ with frequencies $f_1, f_2, \ldots f_n$ respectively then A.M

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{f_1 + f_2 + \cdots + f_n}$$

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

$$N = \sum_{i=1}^{n} f_i$$

sum of all frequencies

iii) Combined Arithmatic mean

Suppose there are 2 groups, there are n_1 observations in the first group with mean $\bar{x}_1$ and n_2 observation in the second group with mean $\bar{x}_2$. Then Combined mean,

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

Remark: If x & y are two random variable related as $y = a + bx$ a, b any real no. & $b \neq 0$ then their A.M. is also related as

$$\bar{y} = a + b\bar{x}$$

2) Median

3) Mode.

Q1 The frequency distribution of lot wise no. of defective items obtained for 150 lots of items in a factory is given in the following table. Calculate mean.

no. of defective in the lot (x_i)	Frequency (f_i)
1	14
2	20
3	22
4	30
5	25
6	18
7	12
8	9

Ans This is a frequency data ~~so~~.

So, $$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{f_1 + f_2 + \dots + f_n}$$

$$\bar{x} = \frac{1\times14 + 2\times20 + 3\times22 + 4\times30 + 5\times25 + 6\times18 + 7\times12 + 8\times9}{14 + 20 + 22 + 30 + 25 + 18 + 12 + 9}$$

$$\bar{x} = \frac{14+40+66+120+125+108+84+72}{150}$$

$$= \frac{\cancel{437}}{\cancel{149}} = \frac{629}{150} = 4.2$$

Q.2 Calculate mean for the following data by applying step-deviation method.

Blood pressure (in mm) (x_i)	No. of persons (f_i)
60	4
65	5
70	31
75	39
80	114
85	30
90	25
95	2

Ans By using step-deviation method

x_i	f_i	$y_i = \frac{x_i - 75}{5}$
60	4	-3
65	5	-2
70	31	-1
$\boxed{75}$	39	0
$\boxed{80}$	114	1
85	30	2
90	25	3
95	2	4
	250	

$y_i \times f_i$
-12
-10
-31
0
114
60
75
8
204

$$\bar{y} = \frac{\Sigma y_i \times f_i}{\Sigma f_i}$$

$$= \frac{204}{250}$$

and, $\because y_i = \frac{x_i - 75}{5}$

$$\bar{y} = \frac{\bar{x} - 75}{5}$$

$\Rightarrow \bar{x} - 75 = 5 \times \bar{y}$

$\Rightarrow \bar{x} = (5 \times \bar{y}) + 75$

$$= 75 + \left(5 \times \frac{204}{250}\right)$$

$$= 75 + 4.08$$

$$= 79.08.$$

Q3 Calculate mean for the following data

Height (in inches)	No. of persons (f_i)
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 – 74	8
	100

Ans This is group frequency data.

$$x_i = \frac{\text{Upper limit} + \text{lower limit}}{2}$$

So,

x_i
61
64
65
70
73

So,

$y_i = \frac{x_i - 67}{3}$	$y_i \times f_i$
−2	−10
−1	−18
0	0
1	27
2	16
	15

$$\bar{y} = \frac{\cancel{15}^{3}}{\cancel{100}_{20}}$$

$$= \frac{3}{20}$$

$$y_i = \frac{x_i - 67}{3}$$

$$\bar{y} = \frac{\bar{x} - 67}{3}$$

$$\bar{x} = (3 \cdot \bar{y}) + 67$$

$$= \left(3 \times \frac{3}{20}\right) + 67$$

$$= 67.45.$$

Q4 Find the missing frequency in the following table.

Class interval	Frequency	x_i
0-10	2	5
10-20	f	15
20-30	3	25
30-40	1	35
40-50	2	45

if the mean is 10.8.

Ans

$x_i \times f_i$	f_i
10	2
15f	f
75	3
35	1
90	2
210+15f	8+f

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

$$\Rightarrow 10.8 = \frac{210 + 15f}{8+f}$$

$$\Rightarrow 8.64 + 10.8f = 210 + 15f$$

$$\Rightarrow 10.8f - 15f = 210 - \cancel{8.64}\ 86.4$$

$$\Rightarrow 93f = 123.6$$

$$f = \frac{123.6}{93}$$

Q5 The A.M calculated from the following frequency distribution is 67.45 find the missing frequency.

Class-interval	Frequency	x_i
60 - 62	15	61
63 - 65	54	64
$\boxed{66 - 68}$	f	67
69 - 71	81	70
72 - 74	24	73
	$174+f$	

Ans

$y_i = \frac{x_i - 67}{3}$	$y_i \times f_i$
-2	-30
-1	-54
0	0
1	81
2	48
	45

So, $\bar{y} = \frac{\Sigma y_i f_i}{\Sigma f_i}$

$$\bar{y} = \frac{45}{175+f}$$

& $y_i = \frac{x_i - 67}{3}$

$$\bar{y} = \frac{\bar{x} - 67}{3}$$

$$\bar{x} = 67 + 3(\bar{y})$$

$$67.45 = 67 + 3\left(\frac{45}{175+f}\right)$$

$$\cancel{(67.45)(175+}$$

$$\frac{135}{174+f} = 0.45$$

$$78.3 + 0.45f = 135$$

$$0.45f = 135 - 78.3$$

$$f = \frac{56.7}{0.45}$$

$$\therefore f = 126$$

- **Geometric Mean (G.M)**

i) For non-frequency data :-

$$GM = (x_1 \times x_2 \ldots \times x_n)$$

ii) For frequency data :-

$$GM = \left(x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_n^{f_n}\right)^{1/N}$$

(N = sum of all frequency)

iii) Combined geometric mean

Suppose there are two groups with n_1 observations having mean G_1 & in the second group with n_2 observations having mean G_2. Then the combined GM is given by

$$G.M. = \left(G_1^{n_1} \times G_2^{n_2}\right)^{\frac{1}{n_1+n_2}}$$

Remark If two variables x & y are related as $y = a \cdot x$, their their GM is also related as

$$GM(y) = a \cdot GM(x)$$

- Harmonic Mean (HM)

i) For non-frequency data :-

$$HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

ii) For frequency data :-

$$HM = \frac{N}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}}$$

$$N = \sum_{i=1}^{n} f$$

Page No.
Date

iii) Combined Harmonic mean

There are two groups, In the 1st group & with n_1 obvs. & HM. H_1, & in the second group n_2 obvs. & HM. H_2

$$HM = \frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$$

Remark If x & y are two variables related as $y = ax$, then their HM. is also related as

$$HM(y) = a \cdot HM(x)$$

Q The average age of a group of 20 girls is 15 years and that of ~~it~~ a group of 25 boys is 24 years, if the two groups are taken together, what is the average age of the new group.

Ans

$$HM = \frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$$

$$= \frac{20 + 25}{\frac{20}{15} + \frac{25}{24}}$$

$$= \cancel{\frac{45}{1.3 + 1.04}} = \cancel{\frac{45}{\frac{20 \times 24}{15 \times 24}}}$$

Ans average age of both the groups

$$= \frac{x_1 b_1 + x_2 b_2}{b_1 + b_2}$$

$$= \frac{15 \times 20 + 25 \times 24}{20 + 25}$$

$$= \frac{300 + 600}{45}$$

$$= \frac{900}{45} = 20.$$

Page No
Date 26 8 22

~~2) The~~

2) **Median** :- The median of a variable is defined as the middle most value when its values are arranged in ascending or descending order of magnitude.

i) Median for ungrouped data :-

a) n = total no. of observations = odd

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

b) n = even

$$\text{Median} = \frac{1}{2}\left[\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}\right] \text{obs.}$$

ii) Median for grouped frequency data

$$\text{Median} = l + \left[\frac{\frac{N}{2} - C.F}{f}\right] \times h$$

where,

l = lower-class boundary of Median class.

N = total frequency

$C.F$ = Cumulated Frequency of the previous class of median class.

f = original frequency of the median class.

h - class width
= upper class boundary - lower class boundary.

eg1:- The scores of nine students in economics in a class test were found to be 40, 37, 41, 38, 31, 37, 44, 45, 42. Find the median score.

Ans Arranging the data in ascending order

31, 37, 37, 38, 40, 41, 42, 44, 45

$$\text{median} = \frac{n+1}{2} = \frac{9+1}{2} = \frac{\cancel{10}^{5}}{\cancel{2}} = 5$$

∴ median score is 40.

eg2: The weights of 6 persons are given below, 65, 58, 56, 60, 67, 70
Find the median weight

Ans 56, 58, 60, 65, 67, 70

$$\text{median} = \frac{1}{2}\left[\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}\right]$$

Page No.
Date

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

$$= \frac{1}{2}\left[60+65\right]$$

$$= \frac{\overset{62.5}{\cancel{125}}}{\cancel{2}_1}$$

$$= \underline{\underline{62.5}}$$

eg3: Find the median for the folloving data :

Height (in inches)	No. of persons
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 – 72	8

upper class limit must be equal to next lower class limit.

Height	No. of persons	Class boundary
60 – 62	5	59.5 – 62.5
63 – 65	18	62.5 – 65.5
66 – 68	42	65.5 – 68.5
69 – 71	27	68.5 – 71.5
72 – 72	8	71.5 – 72.5
	100	

C.F (<)

5

5+18 = 23

23+42 = 65 → median class.

65+27 = 92

92+8 = 100

C.F = N → Total frequency.

N = 100

$\frac{N}{2} = 50$

So, Median $= l + \left[\frac{\frac{N}{2} - CF}{f}\right] \times h$

$= 65.5 + \left[\frac{50 - 23}{42}\right] \times 3$

$= 65.5 + \left[\frac{27}{\cancel{42}\ 14}\right] \times \cancel{3}$

$= 67.4$

eg. 4:

Class interval	Frequency	CF
140 - 150	5	5
150 - 160	10	5+10 = 15
160 - 170	20	15+20 = 35
170 - 180	9	35+9 = 44
180 - 190	6	44+6 = 50
190 - 200	2	50+2 = 52
	52	

~~N~~.

$N = 52$

$\frac{N}{2} = \frac{\cancel{52}^{26}}{\cancel{2}} = 26$

So, $\text{Median} = l + \left[\frac{\frac{N}{2} - CF}{f}\right] \times h$

$= 160 + \left[\frac{26 - 15}{20}\right] \times 10$

$= 160 + \left[\frac{11}{2\cancel{0}}\right] \times 1\cancel{0}$

$= 160 + 5.5$

$= 165.5$

Page No.
Date

3) **Mode** : The mode of a variable is that value of the variable which has the heighest frequency or frequency density accoundnding as the variable is discrete or continuous.

eg.1: Find the mode of the following data

7, 4, 3, 5, 6, 3, 3, 2, 4, 3, 4, 3, 4, 4,
3, 2, 2, 4, 3, 5, 4, 3, 4, 3, 4, 3, 1, 2, 3.

1	1
2	4
3 ←	$\boxed{12}$ → highest frequency.
4	9
5	2
6	1
7	1
	30

Mode is 3.

i) Mode for grouped frequency data :-

$$\text{Mode} = l + \left[\frac{f_0 - f_1}{2f_0 - f_{-1} - f_1}\right] \times h$$

where, l = lower class boundary of modal class.

f_0 = frequency of modal class

f_{-1} = frequency of the previous class of modal class.

f_1 = frequency of the succeeding class of the modal class.

h = class width.

eg1: Find the mode for the following data

Marks	No. of students	Class Boundary
31-40	6	30.5-40.5
41-50	14	40.5-50.5
51-60	20	50.5-60.5
61-70	7	60.5-70.5
71-80	3	70.5-80.5

modal class.

$$Mode = \overset{50.5}{\cancel{l}} + \left[\frac{20-14}{2\times 20-14-7}\right] \times 10$$

$$= 50.5 + \left[\frac{6}{40-21}\right] \times 10$$

$$= 50.5 + \left[\frac{6}{19}\right] \times 10$$

$$= 53.65$$

- **Relation b/w Mean, Median & Mode**

Mean - Mode = 3 (Mean - Median)

- **Measure of Dispersion :-**

There are four types of measure of dispersion

i) Range

ii) Mean deviation

iii) Standart deviation

iv) Quartile deviation

1) **Range** :- The Range of a variable is the simplest measure of its dispersion as it is defined as the difference between the highest and the lowest of its given set of values.

eg1: Suppose a variable takes values 3, 5, -1, 8, 4; Find Range.

Ans Highest value = 8
lowest value = -1

$R = 8 - (-1)$
$= 9$

Page No.
Date

2) Mean deviation :-

i) mean deviation about Mean

$$M.D_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$$ (for non-frequency data)

ii) mean deviation about Median

$$M.D_{Me} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \text{Median}|$$

3) Standard deviation :-

$$S.d = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2}$$ (for non-frequency data)

$$S.d = \sqrt{\frac{1}{N}\sum_{i=1}^{n} x_i^2 f_i - (\bar{x})^2}$$ (for frequency data)

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} x_i f_i$$

$$N = \sum_{i=1}^{n} f_i$$

- mean deviation about mean

$$M.D_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}| f_i$$

$$N = \sum_{i=1}^{n} f_i$$

- mean deviation about median

$$M.D_{Me} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \text{median}| f_i$$

$$N = \sum_{i=1}^{n} f_i$$

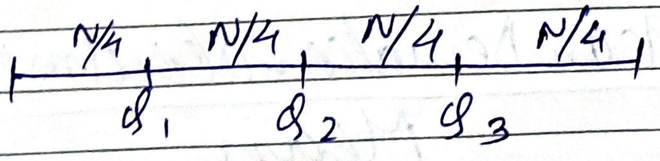
Remark 1: Standard deviation can never be negative

2: Sd^2 is called variance.

3: If two variable x & y are related as
$y = a + bx$

$$Sd(y) = |b| \cdot Sd(x)$$

4) <u>Quartile deviation</u> :-



$$Q_1 = \frac{N}{4}$$

$$Q_2 = \frac{N}{4} + \frac{N}{4} = \frac{N}{2}$$

Remark 4: If x & y are two variables related as $y = ax$ then their sd. is related as $sd(y) = a^2 . sd(x)$

$$Q_3 = \frac{N}{4} + \frac{N}{4} + \frac{N}{4} = \frac{3N}{4}$$

$$Q_1 = l + \left[\frac{\frac{N}{4} - CF}{f}\right] \times h$$

$$Q_2 = l + \left[\frac{\frac{N}{2} - CF}{f}\right] \times h$$

Q_2 or second quartile is also known as median

$$Q_3 = l + \left[\frac{\frac{3N}{4} - CF}{f}\right] \times h$$

$$\boxed{Q.D = \frac{Q_3 - Q_1}{2}}$$

also known as Semi-interquartile range.

Remark1: Coefficient of mean deviation about mean =

$$\left(\frac{\text{Mean Deviation about mean}}{\text{Mean}} \times 100\right)\%$$

Page No.
Date

Remark 2: Coefficient of mean deviation about median =

$$\left(\frac{\text{Mean deviation about median}}{\text{median}} \times 100 \right) \%$$

Remark 3: Coefficient of quartile deviation ~~ab~~ =

$$\left(\frac{\text{quartile deviation}}{\text{median}} \times 100 \right) \%$$

Remark 4: Coefficient of variation (CV) =

$$CV = \left(\frac{S.d}{\text{mean}} \times 100 \right) \%$$

31/8/23

- Problems from Measure of Dispersion :-

Q1 The first of two subgroups has 100 items with mean 15 and S.D = 3. If the whole group has 250 items with mean 15.6 and S.D = $\sqrt{13.44}$, find the S.D of second sub-group.

Ans
$n_1 = 100$
$\bar{x}_1 = 15$
$s_1 = 3$

$n_1 + n_2 = 250$
$n_2 = 150$

Combined mean

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

$$15.6 = \frac{(100\times 15)+(150\times \bar{x}_2)}{250}$$

$$\cancel{(15.6\times 250) - 1500 - 150 = \bar{x}_2}$$

$$\cancel{15.6 = \frac{1500 + 150\times\bar{x}_2}{250}}$$

$$\cancel{15.6 = \frac{1650 + \bar{x}_2}{250}}$$

$\cancel{15.}$

$$\bar{x}_2 = 16$$

$$(s.d)^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2} + \frac{n_1 n_2 (\bar{x}_1 - \bar{x}_2)^2}{(n_1+n_2)^2}$$

$$13.4 = \frac{100\times 9 + 150\times s_2^2}{250} + \frac{(100\times 150)\times 1}{(250)^2}$$

$$s_2^2 = 16$$

$$s = 4$$

cannot be negative

Q2 Calculate sd. for the following table.

Marks	No. of students	x_i
31-40	6	35.5
41-50	14	45.5
51-60	20	55.5
61-70	7	65.5
71-80	3	75.5
	50	

y_i	$y_i \times f_i$	$y_i^2 \times f_i$
-2	-12	24
-1	-14	14
0	0	0
1	7	7
2	6	12
	-13	57

$$\bar{y} = \frac{\Sigma y_i f_i}{\Sigma f_i}$$

$$= \frac{-13}{50}$$

$$sd(y) = \sqrt{\frac{1}{N}\Sigma y_i^2 f_i - (\bar{y})^2}$$

$$= \sqrt{\frac{57}{50} - \left(\frac{-13}{50}\right)^2}$$

$$= \sqrt{1.14 - 0.0676}$$

$$= \sqrt{1.07} = 1.03$$

$$y_i = \frac{x_i - 55.5}{10}$$

$$x_i = 55.5 + 10 y_i$$

$$sd(x) = |10| \cdot sd(y)$$
$$= 10 \times 1.03$$
$$= 10.3$$

Q3 Calculate 1st quartile and 3rd quartile and hence quartile deviation for the following frequency distribution.

height (in cm)	No. of persons	Class boundary
141–145	7	140.5–145.5
146–150	9	145.5–150.5
151–155	15	150.5–155.5
156–160	23	155.5–160.5
161–165	21	160.5–165.5
166–170	10	165.5–170.5
171–175	5	170.5–175.5
	90	

C.F (<)

7

7+9 = 16

16 + 15 = 31 $\longrightarrow Q_1$

31 + 23 = 54

54 + 21 = 75 $\longrightarrow Q_3$

75 + 10 = 85

85 + 5 = 90

Page No.
Date

$$\frac{N}{4} = \frac{90}{4} = 22.5$$

$$Q_1 = l + \left(\frac{\frac{N}{4} - C.F}{f}\right) \times h$$

$$= 150.5 + \left(\frac{22.5 - 16}{15}\right) \times 5$$

$$= 152.666$$

$$= \cancel{15.2}\ 152.67$$

$$\frac{3N}{4} = \frac{3 \times 90}{4} = \frac{270}{4} = 67.5$$

$$Q_3 = l + \left(\frac{\frac{3N}{4} - cf}{f}\right) \times h$$

$$= 160.5 + \left(\frac{67.5 - 54}{21}\right) \times 5$$

$$= 163.71$$

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

$$= \frac{163.71 - 152.67}{2}$$

$$= 5.52$$

~~Moments :~~

Q4 If A.M $= 10$, CV $= 50\%$. find sd.

Ans $CV = \frac{sd}{mean} \times 100$

$50 = \frac{sd}{10} \times 100$

$sd = 5$.

Moments :-

1) Raw moments : ~~It is defined as~~ r^{th} order is defined as m_r'

$m_r' = \frac{1}{n} \sum_{i=1}^{n} x_i^r$ (For non-frequency)

$m_r' = \frac{1}{N} \sum_{i=1}^{n} x_i^r \cdot f_i$ (Frequency)

If we put $r = 1, 2, 3$, we will get 1^{st} order, 2^{nd} order, 3^{rd} order raw moments respectively

2) ## Central moments :-

r^{th} order central moment is defined as m_r

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r \text{ (for non-frequency)}$$

$$m_r = \frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^r \cdot f_i \text{ (frequency)}$$

Remark: 1^{st} order central moment i.e. m_1 is always 0, for any distribution.

- Relation b/w raw moments and central moments :-

1) $m_1 = 0$
2) $m_2 = m_2' - (m_1')^2$
3) $m_3 = m_3' - 3m_2'm_1' + 2(m_1')^3$

- Skewness :-

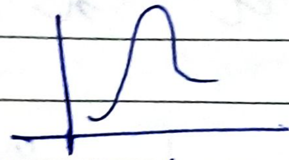
In relation to a frequency distribution the term skewness ~~reps~~ refers to its departure from symmetry

• Diferent measures of skewness:-

1) Skewness is measured by ~~g~~ g_1

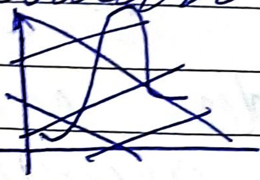
$$g_1 = \frac{m_3}{(s.d)^3}$$

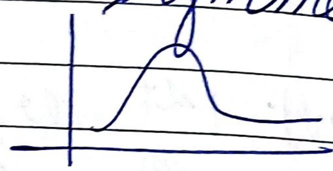
If $g_1 = 0$



$\Rightarrow$ the distribution is symmetric

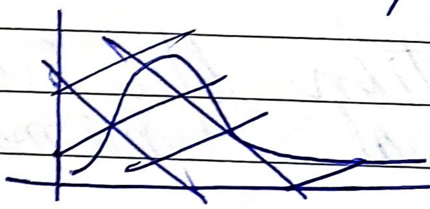
$g_1 > 0$

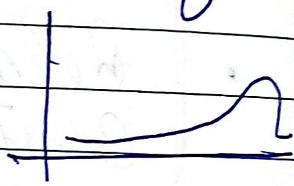




$\Rightarrow$ the distribution is positively skewed.

$g_1 < 0$





$\Rightarrow$ the distribution is negatively skewed.

~~2) Coefficient~~

2) Pearson's coefficient of skewness:-

$$S_k = \frac{\text{mean} - \text{mode}}{sd}$$

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

3) Bowley's coefficient of skewness :-

$$S_K = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

• Kurtosis :-

By Kurtosis of a frequency distribution we mean its degree of Peakedness or Steepness

A Measure of Kurtosis is given by

$$g_2 = \frac{m_4}{(s.d)^4} - 3 \qquad s.d \neq 0$$

Remark: It is to be noted that the quantities g_1^2 and $(g_2 + 3)$ are sometimes used as measure of skewness & Kurtosis respectively. These are known as the b_1 & b_2 coefficients.

$$b_1 = g_1^2 = \frac{m_3^2}{(s.d)^6}$$

$$b_2 = g_3 + 3 = \frac{m_4}{(s.d)^4}$$

1. Sheppard's correction
2. Box plot
3. Outlair detection.

} left

Q The following data were obtained by observing the no. of cars passing a point on a road during a certain hour of the day the observations being made for 30 days.

Ans

No. of Cars (x_i)	5	10	11	12	15	Total
No. of days (f_i)	3	8	5	10	4	30
$x_i \times f_i$	15	80	55	120	60	330

Find the average no. of car passing per day, its ~~sd~~, MD about mean & CV.

Ans.

$x_i^2 \times f_i =$	75	800	605	1440	900

Total = 3820

$$\overline{x} = \frac{\Sigma x_i f_i}{\Sigma f_i}$$

$= \frac{\cancel{330}^{11}}{\cancel{30}}$

$\bar{x} = 11$

→ avg. cars passing per day

$sd = \sqrt{\frac{1}{N} \Sigma x_i^2 f_i - (\bar{x})^2}$

$= \sqrt{\frac{3820}{30} - 11^2} = \sqrt{6.333}$

$= 2.51$

						Total
$\lvert x_i - \bar{x} \rvert =$	6	1	0	1	4	
$\lvert x_i - \bar{x} \rvert . f_i =$	18	8	0	10	16	52

$MD_{\bar{x}} = \frac{1}{N} \Sigma \lvert x_i - \bar{x} \rvert f_i$

$= \frac{1}{30} \times 52$

$= 1.73$

$C.V = \frac{sd}{mean} \times 100$

$= \frac{2.51}{11} \times 100 = 22.81\%$

Home Work

Q1 In a distubution the difference b/w Q_3 & Q_1 is 2.03 and their sum is 72.67 and the median is 36.18. Calculate the coefficient of skewness

Ans
$$S_k = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

here, $Q_3 - Q_1 = 2.03$
& $Q_3 + Q_1 = 72.67$
median $M = 36.18 = Q_2$

So,
$$S_k = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$= \frac{72.67 - 2 \times 36.18}{2.03}$$

$$= \frac{72.67 - 72.36}{2.03}$$

$$= 0.152$$

Page No.
Date

Q2 Find the first, second & third central moment for the data.

Expenditure	No. of families
3-6	28
6-9	292
9-12	389
12-15	212
15-18	59
18-21	18
21-24	2
	1000

Q3 An Analysis of weekly wages paid to workers in two farms, belonging to the same industry gives the following result.

Firm	No. of workers	Average weekly wage	sd of wages
A	476	Rs. 34.5	Rs. 5
B	524	Rs. 28.5	Rs. 4.5

In which farm A or B is there greater variability in individual wages? Find the combined sd. if ~~the~~ two farms are taken together

Ans 2

Expenditure	No. of families (f_i)	x_i	$x_i f_i$
3 - 6	28	4.5	126
6 - 9	292	7.5	2190
9 - 12	389	10.5	4084.5
12 - 15	212	13.5	2862
15 - 18	59	16.5	973.5
18 - 21	18	19.5	351
21 - 24	2	22.5	45
	1000	94.5	10632

$$m_r = \frac{1}{N} \sum_{i=1}^{n} (x_i - \bar{x})^r . f_i$$

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{10632}{1000}$$

$$= 10.632$$

$(x_i - \bar{x})$	$(x_i - \bar{x})^2 f_i$	$(x_i - \bar{x})^3 f_i$
-6.1	~~170.8~~ 1041.88	-6355.46
-3.1	~~905.2~~ 2806.12	-8698.97
-0.1	~~38.9~~ 3.89	-0.38
2.9	~~614.8~~ 1782.92	5170.46
5.9	~~348.1~~ 2053.79	12117.36
8.9	~~160.2~~ 1425.78	12689.44
11.9	~~23.8~~ 283.22	3370.31
	9397.6	18292.76

$m_1 = 0$

$$m_2 = \frac{1}{N} \sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot f_i$$

$$= \frac{1}{1000} \cdot 9397.6$$

$$= 9.3976$$

$$m_3 = \frac{1}{N} \sum_{i=1}^{n} (x_i - \bar{x})^4 \cdot f_i$$

$$= \frac{1}{1000} \cdot 18292.76$$

$$= 18.29276$$

Ans 3 sd (firm A) = 5

sd (firm B) = 4.5

$$CV \text{ (firm A)} = \frac{sd}{mean} \times 100$$

$$= \frac{5}{34.5} \times 100$$

$$= 14.49$$

$$CV \text{ (firm B)} = \frac{4.5}{28.5} \times 100$$

$$= 15.78$$

$CV(B) > CV(A)$

∴ firm B has greater variability in individual wages than A.

$n_1 = 476$ $\quad$ $n_2 = 524$

$\bar{x}_1 = 34.5$ $\quad$ $\bar{x}_2 = 28.5$

$\sigma_1 = 5$ $\quad$ $\sigma_2 = 4.5$

Combined group.

$N = 1000$

$\sigma = ?$

So, $N\sigma^2 = (n_1\sigma_1^2 + n_2\sigma_2^2) + (n_1 d_1^2 + n_2 d_2^2)$

also, $N\bar{x} = n_1\bar{x}_1 + n_2\bar{x}_2$ —— ①

$1000\bar{x} = 476 \times 34.5 + 524 \times 28.5$

$1000\bar{x} = 16422 + 14934$

$\bar{x} = 31.356$

$\bar{x} = 31.35$

$d_1 = 34.5 - 31.35$

$= 3.15$

$d_2 = 28.5 - 31.35$

$= -2.65$

So, putting in ②

$$1000\sigma^2 = (476 \times 5^2 + 524 \times 4.5^2) + (476 \times 3.15^2 + 524 \times (-2.85)^2)$$

$$1000\sigma^2 = 22511 + 8977.99$$

$$\sigma^2 = \frac{31488.99}{1000}$$

$$\sigma^2 = 31.48$$

$$\sigma = 5.61$$

$\therefore$ The combined s.d. is 5.61.

Page No.
Date 2 8 23

• Sheppard's Correction :-

In calculating moments for a grouped frequency distribution using formula of raw moments and central moments, all the values included in a class are taken to be equal to the class mark of that class. By this we introduce some error, known as 'error due to grouping'. So some corrections are necessary for removing this grouping error.

W.H. Sheppard has developed a method for adjustment of moments in case of grouped frequency distribution having classes of equal width.

⇒ **Sheppard's Correction for raw moments**

$$m_1' \text{(corrected)} = m_1'$$

$$m_2' \text{(corrected)} = m_2' - \frac{c^2}{12}$$

$$m_3' \text{(corrected)} = m_3' - \frac{c^2}{4} m_1'$$

⇒ **Sheppard's correction for central moments**

$$m_2 \text{(corrected)} = m_2 - \frac{c^2}{12}$$

$$m_3 \text{(corrected)} = m_3$$

where, C = is the common class width.