$$G.M = 145.15 \text{ Bgx}$$

The Harmonic Mean

- This kind of mean is used to determine average for speeds; It is used where different speeds were travelled for equal portions over time, space & volume.
- Hence the harmonic mean, HM is given by:

$$HM = \frac{N}{\sum_{i=1}^{n}\left(1/x_i\right)}$$

where;

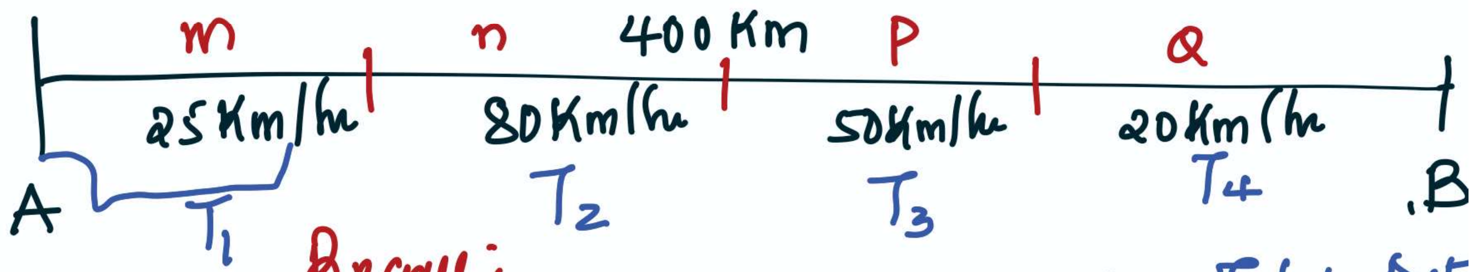
N = no of equi-distant portions

x_i = speeds [data set values]

Example:

Paul & Joan are to travel 400 km. They cover the first and 2nd parts of the journey at 25 km/hr & 80 km/hr resp; for the 3rd & 4th portions of the journey, they move at 50 km/hr and 20 km/hr respectively.

Given that their stops are equidistant, determine their average speed using the harmonic mean.



Recall:

$$H.M = N \Big/ \Sigma\left(\tfrac{1}{x_i}\right)$$

$$S = \frac{\text{Total Dist}}{\text{Total Time}}$$

$$T = D/S.$$

x_i	$1/x_i$
25	0.04
80	0.0125
50	0.02
20	0.05
Total	0.1225

$$H.M = \frac{4}{0.1225}$$

$$H.M = \underline{\underline{32.65 \text{ Km/hr}}}$$

Class 4

Monday
July 07, 2024

Harmonic mean for Grouped / Frequency Data

- Once the class mark has been obtained, we apply the formular

$$H \cdot M = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \left(f_i / x_i\right)}$$

$$\frac{f_i}{1} \cdot \frac{1}{x_i}$$

$$\frac{10}{5} = \frac{10}{1} \times \frac{1}{5}$$

where;

f_i = class frequency

x_i = class mark

Example:

The exchange rates over a uniform period were as noted.

Exchange rates (US$)	100–200	200–300	300–600	600–900	900–1200
No of Weeks	4	8	20	6	12

Determine the average exchange rate using the Harmonic mean.

Recall:

$$H \cdot M = \frac{\sum f_i}{\sum \left(f_i / x_i\right)}$$

Exchange US$	x_i	f_i	f/x
100 – 200	150	4	0.0267

200 - 300	250	8	0.032
300 - 600	450	20	0.0444
600 - 900	750	6	0.008
900 - 1200	1050	12	0.0114
Total		50	0.1225

$$H.M = \frac{50}{0.1225} = 408.16$$

$$\underline{H.M = 408.16 \text{ USD}}$$

The Mode

✓ This is the most frequent value in a given data set; For instance the mode for scores;

80, 90, 85, 45, 60, 80, 55, 70, 90, 80, 90

⇒ mode = 80 → Highest frequency.

⇒ modal frequency = 3

→ mode = 80 & 90

modal freq = 3

Mode for Grouped Data

✓ Here the modal class should be determined and then

[class with highest frequency]

We apply the formular:

$$\text{Mode} = L + \left[\frac{d_1}{d_1 + d_2}\right] \times c_w$$

where;

L = Lower class boundary of the modal class

$d_1 = f_m - f_b$

$d_2 = f_m - f_a$

$\Rightarrow$ f_m = frequency of the modal class

f_b = frequency of the class below the modal class

f_a = frequency of the class above the modal class

C_w = class width of the modal class.

Limits and boundaries

✓ A limit is the extreme value of a given range / class that is usually part of the range; Limits are used for discrete data [finite or countables]

whereas,

✓ A boundary is the extreme value of a given class that may not be part of the data set; either the lower or upper boundary is excluded in the range.

- To translate a limit to a boundary we use the correction factor ± 0.5; where 0.5 is deducted from the lower limit & added to the upper.
- For limits, in order to have the class width, 1 is added to the difference whereas for the boundaries we take the difference as is.

For instance:

Limits $\xrightarrow{\pm 0.5}$ Boundaries

Limits	Boundaries
10 – 19	9.5 – 19.5
20 – 29	19.5 – 29.5
30 – 39	29.5 – 39.5

Example:

- The weights of participants at a workshop were as noted;

Weights (Kg) 10–19	20–29	30–39	40–49	50–59	60–69	70–79
No of Participants	4	8	14	12	3	6

Determine the modal weight.

Recall:

$$\text{Mode} = L + \left[\frac{d_1}{d_1 + d_2}\right] \times C_W$$

Modal class = 40 - 49 ✓✓

$\Rightarrow$ Boundaries ± 0.5 : $\boxed{39.5 - 49.5}$

$d_1 = f_m - f_b$

$= 14 - 8 = 6$

$d_2 = f_m - f_a = 14 - 12 = 2$

$$\text{Mode} = 39.5 + \left(\frac{6}{6+2}\right) \times 10$$

$$= 39.5 + (6/8 \times 10) = 47 \checkmark\checkmark$$

Mode = $\underline{\underline{47 \text{ Kg}}}$.

The Median

✓ This refers to the middle value in a given data set; it is obtained by re-arranging the data set in an array format [either ascending or descending order] first.

✓ The median should leave an equal number of data values on either side.

1, 3, 5, 7, 9

Case I: Median for ungrouped data

(a) where N is ODD: N = no of data values.

✓ Here the median shall be the middle value;
For instance, the median for the scores;

90, 88, 95, 65, 75, 55, 70 ;

⇒ Re-arranging we have:

55	65	70	75	88	90	95
1	2	3		3	2	1

→ Median = 75.

(b) Where N is even:

✓ Here, the median shall be the average of the two middle values. For instance, the median sales for:

120, 140, 160, 90, 75, 80, 120, 98;

Re-arranging:

1	2	3	4	4	3	2	1
75	80	90	98	120	120	140	160
1	2	3	4	4	3	2	1

→ median = $\frac{98+120}{2}$

Median = $\underline{\underline{109}}$

Side-walking

(1) For ODD data, the position of the median is given by:

$$x = \left(\frac{N+1}{2}\right)\text{th}$$

For instance where; $N = 7$

$$x = \frac{7+1}{2} = 8/2 = 4^{th}$$

② For even data, the position of the median is given by:

$$x = N/2^{th}$$

For instance, $N = 8$:

$$x = N/2 = 8/2 = 4^{th}.$$

Case II: For Grouped Data

In this case, a cumulative frequency distribution must be generated and then we apply the formular:

$$Median = L + \left(\frac{x - F_b}{f_m}\right) \times C_w$$

where:

L = lower class boundary of the median class

x = position of the median value.

ie it gives the median class;

$$x = N/2^{th}$$

F_b = cumulative frequency of the class below the median class

f_m = frequency of the median class

C_W = class width of the median class

Example:

The sales of xxx Inc were as summarised i

Sales units	20-29	30-39	40-49	50-59	60-69
No of clients	4	8	12	15	11

Determine the median sales.

Recall: $49.5 + 25 -$

$$\text{Median} = L + \left(\frac{x - F_b}{f_m}\right) \times C_W$$

Sales	f	C-F
10-19	0	
20-29	4	4
30-39	8	12
40-49	12	24
$\boxed{50-59}$	$\boxed{15}$	$\boxed{39}$
60-69	11	50

F_b

$x = 50\%$ of N

$\frac{50}{100} \times N = \frac{1}{2} \times N$

$x = N/2 = 50/2$

$x = 25^{th}$

Median class = 50-59

Boundaries = 49.5 - 59.5

| 50 |

$$\text{Median} = 49.5 + \left(\frac{25-24}{15}\right) \times 10$$

$$= 49.5 + \left(\frac{1}{15} \times 10\right) = 50.17$$

Median = 50.17 units

Cumulative frequency refers to the total no. of items that is below or above a given class limit / boundary.

Class 5

Monday
July 15, 2024

Quantiles:

✓ These divide a data set into equal percentage intervals; they include:

- Quartiles
- Deciles
- Percentiles

(a) Quartiles:

✓ These use equal intervals of 25% to analyse a data set; For instance

(i) The lower Quartile: Q_1

⇒ This covers 25% of the data set; its position is given by:

$$x = 25\% \text{ of } N \rightarrow \Sigma f$$

N = total data set values

(ii) The second Quartile: Q_2

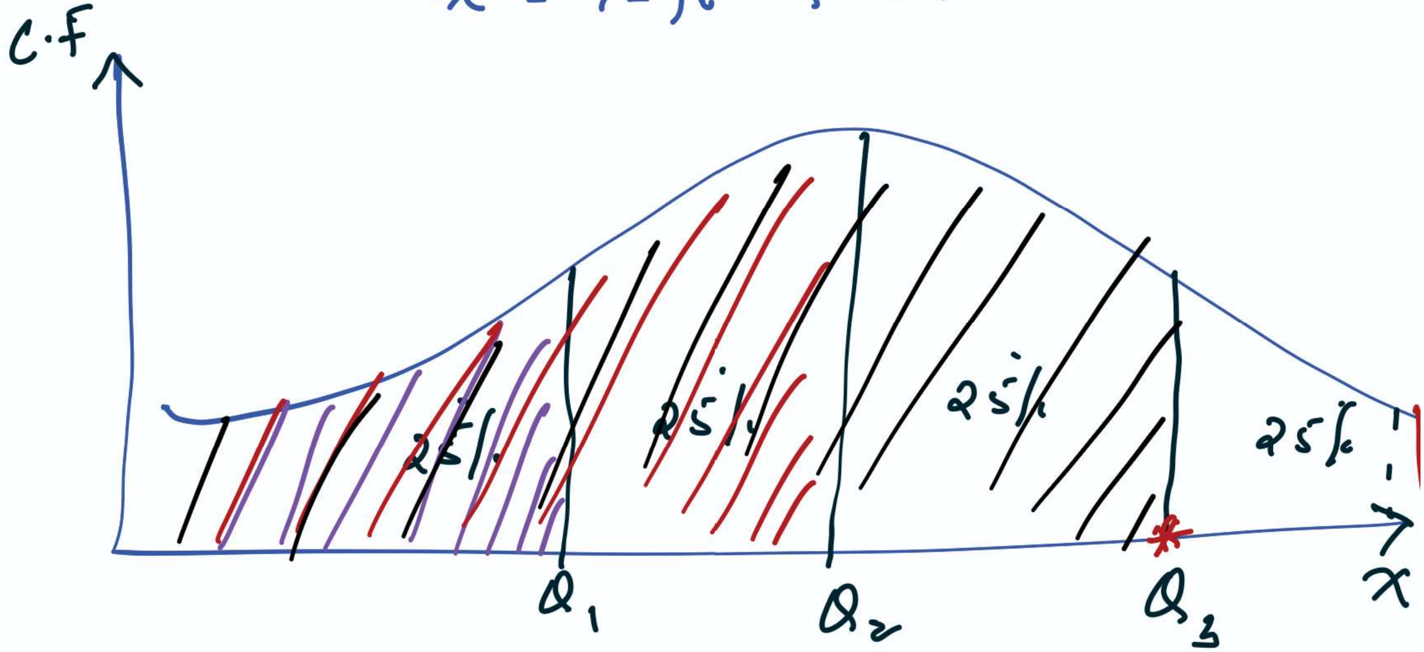
⇒ This covers 50% of the data set; its position is given by:

$$x = 50\% \text{ of } N \text{ [median]}$$

(iii) The upper Quartile: Q_3

⇒ This represents 75% of the data set; it has a position of:

$$x = 75\% \text{ of } N$$



Deciles:

✓ These use equal percentages in intervals of 10;

For instance; • D_1 would cover 10% of the data set; its position shall be given by:

$x = 10\%$ of N

• D_2 covers 20% of the data set; its position is given by:

$x = 20\%$ of N

⋮

D_9 would cover 90% of the data set with a position

$x = 90\%$ of N.

Percentiles:

✓ These use equal intervals of 1% to analyse a data set; For instance;

• P_1: covers 1% of the data set with a position of: $x = 1\% \times N$.

• P_2: its position would be:

$x = 2\%$ of N

⋮

P_{99}: this will cover 99% of the data set: with a position of $x = 99\%$ of N

Generally, a Quantile is given by:

$$Q_i = L_{Q_i} + \left[\frac{x_{Q_i} - F_b}{f_{Q_i}}\right] \times Cw_{Q_i}$$

where:

L_{Q_i} = lower class boundary of the Quantile class

x_{Q_i} = Position of the Quantile in Question i.e it gives the Quantile class

F_b = Cumulative frequency of the class below the Quantile class

f_{Q_i} = frequency of the Quantile class

Cw_{Q_i} = class width of the Quantile in Question.

Example:

The weights of participants at a game were as recorded;

Weight (Kg)	20 - 30	30-40	40-50	50-60	60-70	70-80
No of Participants	3	4	12	8	16	17

Determine (a) the lower Quartile weight

(b) the 6^{th} Decile weight

(c) the 90^{th} percentile weight.

(a) Recall:

$$Q_1 = L_{Q_1} + \left[\frac{x_{Q_1} - F_b}{f_{Q_1}}\right] \times C_{wQ_1}$$

40 − 49

−0.5   +0.5

39.5   49.5

10 − 2

Weight	f	Cum. Freq (F)
20 − 30	3	3
30 − 40	4	7
40 − 50	12	19
50 − 60	8	27
60 − 70	16	43
70 − 80	17	60
	60	

Lower Quartile class

x_{Q_1} = 25% of N

$= \frac{25}{100} \times 60$

$x = 15^{th}$ *

Q_1 class = 40 − 50

$$Q_1 = 40 + \left(\frac{15 - 7}{12}\right) \times 10$$

$$= 40 + \left(\frac{8}{12} \times 10\right) = 46.67$$

$$Q_1 = \underline{\underline{46.67}}$$

(b) 6^{th} Decile: D_6.

$$x = 60\% \text{ of } N$$

$$= \frac{60}{100} \times 60 = 36^{th}$$

$$D_6 \text{ class} = 60 - 70 \; *$$

$$D_6 = L + \left[\frac{x - F_b}{f_{D_6}}\right] \times C_w$$

$$= 60 + \left[\frac{36 - 27}{16}\right] \times 10$$

$$= 60 + \left(\frac{9}{16} \times 10\right) = 65.63$$

$$D_6 = \underline{\underline{65.63 \text{ Kg}}} . *$$

(c) $\underline{90^{th} \text{ Percentile}}$: P_{90}

$$x = 90\% \text{ of } N$$

$$x = 90/100 \times 60$$

$$x = 54^{th}$$

P_{90} class $= 70 - 80$

$$P_{90} = L + \left[\frac{x - F_b}{f_{P_{90}}}\right] \times CW$$

$$= 70 + \left[\frac{54 - 43}{17}\right] \times 10$$

$$= 70 + \left(\frac{11}{17} \times 10\right) = 76.47$$

$$P_{90} = \underline{76.47 \text{ Kg}}.$$

✓ moving away!
✓ scatter
✓ spread
✓ distribute
✓ Quitting
✓ separate
✓ depart

Measures of Dispersion

✓ Whenever we use a measure of location to estimate the data set, an error is created; this error is the measure of dispersion.

✓ Hence, a measure of dispersion refers to the degree of Variability of the data set values from a given measure of location.

✓ Measures of dispersion include

- Range

- Variance & standard Deviation
- Skewness
- Coefficient of Variation
- Mean Deviation
- Quartile Deviation

(a) The Range:

✓ This refers to the difference btn the highest value and the lowest value. For instance, the range for the scores; 88, 90, 42, 92, 40, 92, 62, 40; is

$$\text{Range} = 92 - 40 = 52$$

✓ For the grouped data the range shall be the difference btn the upper class boundary / limit of the highest class & the lower class boundary / limit of the lowest class

(b) The Mean Deviation:

✓ This refers to the average of the absolute difference btn the data set and its mean;

$$\text{ie} \quad M.D = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{N}$$

where;

x_i = data set values

$\bar{x}$ = the average of x_i

ie $\bar{x} = \frac{\sum x}{N}$

N = data size

-10

$+10$

$| \quad |$ = Absolute symbol; this will translate the -ves to magnitude ie "+ve"

Example:

The costs of production in W-Inc were:

Costs (US$): 650, 750, 600, 800, 1200

Determine the mean deviation in the costs

Recall:

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

x_i	$\bar{x}$	$(x_i - \bar{x})$ Deviation	$\lvert x_i - \bar{x} \rvert$
$\boxed{650}$	800	-150 ✓✓	150

750	800	-50	50
600	800	-200	200
800	800	0	0
$\boxed{1200}$	800	400	400
4000	Total		800

$$\bar{x} = \frac{\sum x}{N} = \frac{4000}{5}$$

$$\bar{x} = \underline{\underline{800}}$$

$$M.D = \frac{800}{5} = 160$$

$$MD = 160 \text{ USD}$$

Mean Deviation for Grouped / Frequency Data

✓ once the class mark has been obtained, we apply the formular:

$$MD = \frac{\sum_{i=1}^{n} \left[f_i \cdot |x_i - \bar{x}| \right]}{\sum_{i=1}^{n} f_i}$$

where: