

Measure of central tendency

All the values tend to cluster in the central part of the data it is called measure of central tendency.
It also indicates the location or position of the data so it is also called measure of location or position.

Average

A single value in the centre of the whole given data it is called an average.

Desirable properties of a an ideal Average

- i) It should be rigidly defined.
- ii) It should be based on all the observations.
- iii) It should be easy to calculate and simple to understand.
- iv) It should not be unduly affected by extreme values.
- v) It should be capable of further mathematical treatment.
- vi) It should have sampling stability.

Types of Average

- i) Arithmetic mean
- ii) Geometric mean
- iii) Harmonic mean
- iv) Mode
- v) Median (Quartiles, Deciles, Percentiles)

Arithmetic mean

The Arithmetic mean or simply mean of the series $x_1, x_2, x_3, \dots, x_i, \dots, x_n$ is defined as dividing the sum of all the values by their numbers. It is denoted by \bar{x} and read as x-bar.

Or

Sum of all the values dividing by their numbers. It is called A.M and denoted by \bar{x} and read as x-bar.

Method of calculation

Un-group data

- i) Simple method/D

$$\bar{x} = \frac{\sum x}{n} \quad \text{sample mean}$$

OR

$$u = \frac{\sum X}{N} \quad \text{population mean}$$

- ii) Short-cut method/Change of origin method

$$\bar{x} = A + \frac{\sum D}{n}$$

Where

$$D = X - A$$

$A = \text{Arbitrary}$

- iii) Coding method/ Change of origin and scale method

$$\bar{x} = A + \frac{\sum u}{n} \times h$$

Where

$$u = \frac{X - A}{h}$$

$A = \text{Arbitrary}$

$h = \text{class interval}$

Group data

$$\bar{X} = \frac{\sum fx}{\sum f}$$

$$\bar{x} = A + \frac{\sum fD}{\sum f}$$

$$\bar{x} = A + \frac{\sum fu}{\sum f} \times h$$

Merits of Arithmetic mean

- i) It should be rigorously defined.
- ii) It is based on all the observations.
- iii) It is easy to calculate and simple to understand.
- iv) It provides a good basis for comparison.
- v) It is capable of further mathematical treatment.
- vi) It is based on a mathematical formula.
- vii) It is best average for less dispersion type of data.
- viii) It Can be calculated if any value is zero or -ve.
- ix) It stable in repeated sampling experiments.

Demerits of Arithmetic mean

- i) It is affected by extreme values.
- ii) Cannot be calculated from open-end table.
- iii) Not best average for rates and ratios.
- iv) Not best average for skewed type data.
- v) Cannot be calculated if any value is missing.
- vi) Cannot be calculated by graphically.
- vii) Sometimes strange answer.
- viii) Not true average.
- ix) Give weights to large values.

Example.1: Find the mean from following ungroup data by simple method.

1,5,1,0,2,5,7,4,5,3,5,4,4,7,8,9,10,9,3,5,3,6,4,8,5,6,3,4,2,3

Solution:
$$\bar{x} = \frac{\sum x}{n} = \frac{141}{30} = 4.7$$

Example.2: Find the mean from following ungroup data by short-cut and coding method.

5,10,15,20,25

Solution:

X	D=X-A	$U = \frac{X-A}{h}$	i) $\bar{x} = \frac{\sum x}{n} = \frac{75}{5} = 15$
5	-10	-2	
10	-5	-1	ii) $\bar{x} = A + \frac{\sum D}{n} = 15 + \frac{0}{5} = 15$
15	0	0	
20	5	1	iii) $\bar{x} = A + \frac{\sum U}{n} \times h = 15 + \frac{0}{5} \times 5 = 15$
25	10	2	
$\sum x = 75$	$\sum D = 0$	$\sum U = 0$	

Example.3: Find A.M from the following data by 1) Direct ii) Short-cut iii) Coding method

Classes:	10-14	15-19	20-24	25-29	30-34
Frequencies:	2	4	6	8	3

Classes	f	x	fx	D	fD	U	fu
10-14	2	12	24	-10	-20	-2	-4
15-19	4	17	68	-5	-20	-1	-4
20-24	6	22	132	0	0	0	0
25-29	8	27	216	5	40	1	8
30-34	3	32	96	10	30	2	6
Total	23		536	0	30	0	6

$$\bar{x} = \frac{\sum fx}{\sum f} = 23.30$$

$$\bar{x} = A + \frac{\sum fD}{\sum f} = 22 + \frac{30}{23} = 23.30$$

$$\bar{x} = A + \frac{\sum fU}{\sum f} \times h = 22 + \frac{6}{23} \times 5 = 23.30$$

Properties of Arithmetic mean

The Arithmetic mean has the following four properties:

i) The sum of the deviations of the observations X 's from \bar{X} is always equal to zero.

Or

Sum of deviation from mean always equal to zero.

i.e.
$$\sum_{i=1}^n (X - \bar{X}) = 0$$

ii) The sum of squared deviations of the observations X 's from \bar{X} is less than or equal to the sum of squared deviations X 's from 'a'.

Or

Sum of square deviation from mean is minimum or least.

i.e.
$$\sum_{i=1}^n (X - a)^2 \geq \sum_{i=1}^n (X - \bar{X})^2$$

This property is called minimal property of Arithmetic mean

iii) **(Combined mean)** If $\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_i, \dots, \bar{X}_k$ are means of K subgroups data having frequencies $n_1, n_2, n_3, \dots, n_i, \dots, n_k$ where $i=1,2,3,\dots,k$ and $\sum_{i=1}^k n = n$ then combined mean \bar{X}_c is

given as
$$\bar{X}_c = \frac{\sum n_i \bar{X}_i}{\sum n_i}$$

iv) If $Y_i = a + bX_i$ Then $\bar{Y} = a + b\bar{X}$

Q. The mean weights of four groups of students consisting of 15,20,10 and 18 students were 162,148,153 and 140 pounds respectively. Find mean weight of all the students.

Solution: where $n_1 = 15$ $n_2 = 20$ $n_3 = 10$ $n_4 = 18$
 $\bar{x}_1 = 162$ $\bar{x}_2 = 148$ $\bar{x}_3 = 153$ $\bar{x}_4 = 140$

$$\bar{x}_c = \frac{\sum_{i=1}^k n_i \bar{x}_i}{\sum_{i=1}^k n_i} = \frac{n_1 \bar{x}_1 + \dots + n_4 \bar{x}_4}{n_1 + \dots + n_4} = \frac{9440}{63} = 149.84$$

Weighted mean

Sometimes, some items are not of equal important. To make them equal important, we assign some numerical values to such items. Such numerical values are called weights. The Arithmetic mean calculated from such type of data is called weighted Arithmetic mean. It is denoted by \bar{X}_w .

$$\bar{X}_w = \frac{W_1 X_1 + W_2 X_2 + W_3 X_3 + \dots + W_n X_n}{W_1 + W_2 + W_3 + \dots + W_n} = \frac{\sum W_i X_i}{\sum W_i} \quad i=1,2,3,\dots,n$$

Q.10: Salman obtained the following marks in a certain examination. Find the weighted mean if weights 4,3,3,2 and 2 respectively are allotted to the subjects.

English	Urdu	Math	Stat	Physics
73	82	80	57	62

Solution:

$$\begin{aligned} \bar{X}_w &= \frac{W_1 X_1 + W_2 X_2 + W_3 X_3 + \dots + W_n X_n}{W_1 + W_2 + W_3 + \dots + W_n} \\ &= \frac{\sum W_i X_i}{\sum W_i} = \frac{1016}{14} = 72.57 \end{aligned}$$

Geometric mean

The Geometric mean of the series $x_1, x_2, x_3, \dots, x_i, \dots, x_n$ is defined as the n th root of the product of the +ve values

i.e.
$$G.M = \sqrt[n]{(X_1 * X_2 * X_3, \dots, X_i, \dots, X_n)}$$

$$G.M = (X_1 * X_2 * X_3, \dots, X_i, \dots, X_n)^{\frac{1}{n}} \text{ or } G.M = \text{Antilog} \frac{\sum \log x}{n}$$

Method of calculation

Un-group data

Group data

$$G.M = (X_1 * X_2 * X_3, \dots, X_i, \dots, X_n)^{\frac{1}{n}}$$

$$G.M = \text{Antilog} \frac{\sum f \log x}{\sum f}$$

$$G.M = \text{Antilog} \frac{\sum \log x}{n}$$

$$G.M = (X_1^{f_1} * X_2^{f_2} * X_3^{f_3}, \dots, X_i^{f_i}, \dots, X_n^{f_n})^{\frac{1}{\sum f}}$$

Merits of Geometric mean

- i) It should be rigorously defined.
- ii) It is based on all the observations.
- iii) It is based on a mathematical formula.
- iv) Stable average.
- v) Not affected by extreme values.
- vi) True average.
- vii) Used in advance statistics.
- viii) Best average for rates, ratios and percentage.

Demerits of Geometric mean

- i) Cannot be calculated from open-end table.
- ii) Not best average for skewed type data.
- iii) Cannot be calculated if any value is missing.
- iv) Cannot be calculated by graphically.
- v) Sometimes strange answer.
- vi) Difficult to calculate.
- vii) Cannot be calculated if any value is zero.
- viii) Cannot be calculated if any value is -ve.

Q: Find Geometric Mean by basic definition and logarithms methods

X: 32 35 36 37 39 41 43

Solution:

$$\log X: 1.5051 \quad 1.5441 \quad 1.5563 \quad 1.5682 \quad 1.5911 \quad 1.6127 \quad 1.6335 \quad \sum \log X = 11.011$$

$$G.M = \text{Antilog} \frac{\sum \log x}{n} = \text{Antilog} \frac{11.011}{7} = \text{Antilog}(1.573) = 37.41$$

Q: The reciprocal of 8 values of x are given below. Calculate G.M

0.0500, 0.0454, 0.0400, 0.0333, 0.0285, 0.0232, 0.0213, 0.0200

Solution:

(1/x)	X	Logx
0.05	20	1.30103
0.0454	22.02643	1.342944
0.04	25	1.39794
0.0333	30.03003	1.477556
0.0285	35.08772	1.545155
0.0232	43.10345	1.634512
0.0213	46.94836	1.67162
0.02	50	1.69897
Total		12.06973

$$G.M = \text{Antilog} \frac{\sum \log x}{n}$$

$$G.M = \text{Antilog} \frac{12.06973}{7} = \text{Antilog}(1.5081)$$

$$G.M = 32.22$$

Q: Calculate Geometric mean for the following data.

classes	f	X	Logx	f(logx)
10-19	15	14.5	1.161368	17.42052
20-29	35	24.5	1.389166	48.62081
30-39	45	34.5	1.537819	69.20186
40-49	25	44.5	1.64836	41.209

$$G.M = \text{Antilog} \frac{\sum f \log x}{\sum f}$$

50-59	12	54.5	1.736397	20.83676
Total	132			197.289

$$G.M = \text{Antilog} \frac{197.289}{132} = \text{Antilog}(1.4946) = 31.23$$

Harmonic mean

The Harmonic mean of the series $x_1, x_2, x_3, \dots, x_i, \dots, x_n$ is defined as the reciprocal of the average and reciprocals of the values.

i.e.

$$H.M = \text{reciprocal of } \frac{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}}{n}$$

$$= \text{reciprocal of } \frac{\sum \left(\frac{1}{x_i}\right)}{n}$$

$$= \frac{n}{\sum \left(\frac{1}{x_i}\right)} \quad \text{For ungroup data}$$

$$H.M = \frac{\sum f}{\sum f \left(\frac{1}{x_i}\right)} \quad n = \sum f \quad \text{Or} \quad H.M = \frac{n}{\sum f \left(\frac{1}{x_i}\right)} \quad \text{For group data}$$

Q: Show that for the following data $H.M < G.M < A.M$ or $A.M > G.M > H.M$

classes	f	X	fx	Logx	Flogx	1/x	f(1/x)
9.3-9.7	2	9.5	19	0.9777	1.9554	0.1053	0.2105
9.8-10.2	5	10	50	1	5	0.1	0.5
10.3-10.7	12	10.5	126	1.0212	12.254	0.0952	1.1429
10.8-11.2	17	11	187	1.0414	17.704	0.0909	1.5455
Total	36		382		36.913		3.3988

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{382}{36} = 10.611$$

$$H.M = \frac{\sum f}{\sum f \left(\frac{1}{x_i}\right)} = \frac{36}{3.3988} = 10.59$$

$$G.M = \text{Antilog} \frac{\sum f \log x}{\sum f} = \text{Antilog} \left(\frac{36.913}{36} \right) = \text{Antilog}(1.025) = 10.60$$

$$10.61 > 10.60 > 10.59$$

Hence prove that $H.M < G.M < A.M$ or $A.M > G.M > H.M$

Merits of Harmonic mean

- i) It should be rigorously defined.
- ii) It is based on all the observations.
- iii) Best average for rates, ratios and speed..
- iv) Based on a mathematical formula.
- v) It is not much affected by sampling variability.

Demerits of Harmonic mean

- i) It is affected by extreme values.
- ii) Cannot be calculated from open-end table.
- iii) Not best average for skewed type data.
- iv) Cannot be calculated if any value is missing.
- v) Cannot be calculated by graphically.
- vi) Sometimes strange answer.
- vii) Difficult to calculate.
- viii) Cannot be calculated if any value is zero.

Median

The Median of the series $x_1, x_2, x_3, \dots, x_i, \dots, x_n$ is defined as the middle value of the arrange series if the number of values is odd and the average of the two middle values if the number of values is even. The Median is denoted by \tilde{X}

Median = \tilde{X} = The value of $\frac{n+1}{2}$ th item

For ungroup data and discrete frequency distribution.

$$\tilde{X} = l + \frac{h}{f} \left(\frac{n}{2} - c \right)$$

Where

h = class interval, l = lower class boundary of median group. , $n = \sum f$

f = frequency corresponding to median group and c = cumulative frequency corresponding to the group preceding the median group.

Merits

- i) Best average if some values are missing.
- ii) Can be calculated from open-end.
- iii) Can be calculated from graphically.
- iii) Can be calculated if any value is -ve or zero.
- iv) Easy to calculate and understand.
- v) Calculated every type of data.
- vi) Best average of skewed type of data.

Demerits

- i) Not best average for ratios and rates.
- ii) Not based on all the values.
- iii) Ignore large or small values.
- iv) Effected by extreme values.
- v) Arranging the data is difficult

Mode

The Median of the series $x_1, x_2, x_3, \dots, x_i, \dots, x_n$ is defined as the most repeated value in the data. It is denoted by \hat{X} . A series has more than one mode or no mode in the data.

For ungroup data and discrete frequency distribution

$$\hat{X} = l + \frac{f_m - f_1}{f_m - f_1 + f_m - f_2} \times h$$

Where h = class interval of the modal group, L = lower class boundary of modal class.

f_m = highest frequency in the data f_1 = frequency associated with the class proceeding the modal class.

f_2 = frequency associated with the class following the modal class.

Merits

- i) Speedily calculated.
- ii) Best average if some values are missing.
- iii) Can be calculated from open-end or graphically.
- iv) Can be calculated if any value is -ve or zero.
- v) Easy to analysis or interpret.
- vi) Easy to calculate.
- vii) Easy to understand

Demerits

- i) Not best average for ratios and rates.
- ii) Not best average for skewed type of data.
- iii) Not based on all the values.
- iv) Ignore large or small values.

Quartiles

The three values which divide the array data into four equal parts are called quartiles.

$Q_1 = \text{The value of } \left(\frac{n+1}{4}\right)\text{th item}$ It is called lower quartile

$Q_2 = \text{The value of } 2\left(\frac{n+1}{4}\right)\text{th item}$ It is equal to median

$Q_3 = \text{The value of } 3\left(\frac{n+1}{4}\right)\text{th item}$ it is called upper quartile

for ungroup data and discrete frequency distribution

$$Q_1 = l + \frac{h}{f} \left(\frac{n}{4} - c \right) \quad \text{Lower quartile}$$

Where h = class interval L = lower class boundary of quartile group group.

f = frequency corresponding to quartile group and $n = \sum f$

c = cumulative frequency corresponding to the group preceding the quartile group.

$$Q_3 = l + \frac{h}{f} \left(\frac{3n}{4} - c \right) \quad \text{Upper quartile}$$

Deciles

The nine values which divide the array data into four equal parts are called quartiles.

$D_1 = \text{The value of } \left(\frac{n+1}{10}\right)\text{th item}$

$D_2 = \text{The value of } 2\left(\frac{n+1}{10}\right)\text{th item}$ up to so on

$D_9 = \text{The value of } 9\left(\frac{n+1}{10}\right)\text{th item}$

For ungroup data and discrete frequency distribution

$$D_1 = l + \frac{h}{f} \left(\frac{n}{10} - c \right)$$

$$D_2 = l + \frac{h}{f} \left(\frac{2n}{10} - c \right) \quad \text{Up to so on}$$

$$D_9 = l + \frac{h}{f} \left(\frac{9n}{10} - c \right)$$

Where h = class interval, L = lower class boundary of deciles group.

f = frequency corresponding to deciles group and $n = \sum f$

c = cumulative frequency corresponding to the group preceding the deciles group.

Percentiles

The ninety nine values which divide the array data into equal 100(hundred) parts are called quartiles.

$p_1 = \text{The value of } \left(\frac{n+1}{100}\right)\text{th item}$

$p_2 = \text{The value of } 2\left(\frac{n+1}{100}\right)\text{th item}$ Up to so on

$p_{99} = \text{The value of } 99\left(\frac{n+1}{100}\right)\text{th item}$

For ungroup data and discrete frequency distribution

$$p_1 = l + \frac{h}{f} \left(\frac{n}{100} - c \right)$$

$$p_2 = l + \frac{h}{f} \left(\frac{2n}{100} - c \right) \quad \text{Up to so on}$$

$$p_{99} = l + \frac{h}{f} \left(\frac{99n}{100} - c \right) \quad \text{For group data.}$$

Where

h= class interval

L= lower class boundary of percentile group

f= frequency corresponding to percentile group and $n = \sum f$

c= cumulative frequency corresponding to the group preceding the percentile group.

Q: Following are the marks obtained by 20 students, Calculate, Mode, Median, Q₁, Q₃, D₆, P₄₇

38,60,41,40,29,40,51,56,40,56,39,40,54,40,53,54,37,53,45,50.

Solution: First we arrange the data in ascending order

Arrange

29,37,38,39,40,40,40,40,40,41,45,50,51,53,53,54,54,56,56,60

Mode=Most repeated value= 40 Ans

Median=The value of $(\frac{n+1}{2})$ th item = (21/2)th=10.5th value

$$=10^{\text{th}} \text{ value} + 0.5(11^{\text{th}} \text{ value} - 10^{\text{th}} \text{ value}) = 41 + 0.5(45 - 41) = 41 + 0.5(4) = 41 + 2 = 43 \text{ Ans}$$

Q₁=The value of $(\frac{n+1}{4})$ th item = (21/4)th=5.25th value

$$=5^{\text{th}} \text{ value} + 0.25(6^{\text{th}} \text{ value} - 5^{\text{th}} \text{ value}) = 40 + 0.25(40 - 40) = 40 + 0.25(0) = 40 + 0 = 40 \text{ Ans}$$

Q₃=The value of $3(\frac{n+1}{4})$ th item = 3 (21/4)th=3(5.25)th value=15.75th value

$$=15^{\text{th}} \text{ value} + 0.75(16^{\text{th}} \text{ value} - 15^{\text{th}} \text{ value}) = 53 + 0.75(54 - 53) = 53 + 0.75(1) = 53 + 0.75 = 53.75 \text{ Ans}$$

D₆=The value of $6(\frac{n+1}{10})$ th item = 6 (21/10)th=3(2.1)th value=12.6th value

$$=12^{\text{th}} \text{ value} + 0.6(13^{\text{th}} \text{ value} - 12^{\text{th}} \text{ value}) = 50 + 0.6(51 - 50) = 50 + 0.6(1) = 50 + 0.6 = 50.6 \text{ Ans}$$

P₄₇=The value of $47(\frac{n+1}{100})$ th item = 47 (21/100)th=47(0.21)th value=9.87th value

$$=9^{\text{th}} \text{ value} + 0.87(10^{\text{th}} \text{ value} - 9^{\text{th}} \text{ value}) = 41 + 0.87(45 - 41) = 41 + 0.87(4) = 41 + 3.48 = 44.48 \text{ Ans}$$

Q: Calculate Mode and 39th percentiles from the following group data.

Solution:

classes	F	c.b	c.f	$\hat{X} = l + \frac{f_m - f_1}{f_m - f_1 + f_m - f_2} \times h$
10-19	15	9.5-19.5	15	$f_m=45, f_1=35, f_2=25, l=29.5, h=10$
20-29	35	19.5-29.5	50	
30-39	45	29.5-39.5	95	
40-49	25	39.5-49.5	120	
50-59	12	49.5-59.5	132	
Total	$\sum f$ 132=n			

$$\hat{X} = 29.5 + \frac{45 - 25}{45 - 25 + 45 - 35} \times 10 = 29.5 + \frac{10}{20} \times 10 = 29.5 + 5 = 34.5$$

$$p_{39} = l + \frac{h}{f} \left(\frac{39n}{100} - c \right) \quad \frac{39n}{100} \text{th value} = \frac{39 \times 132}{100} \text{th value} = 51.48 \text{th value}$$

Where h=10, c=50, f=45, l=29.5

$$p_{39} = l + \frac{h}{f} \left(\frac{39n}{100} - c \right) = 29.5 + \frac{10}{45} (51.48 - 50) = 29.5 + 0.222(1.48) = 29.5 + .3286 = 29.83$$