Subject: **Introduction to Statistics**
BS Mathematics Morning/Evening program spring semester 2020

Chapter # 03 Measures of Central Tendency

Measure of Central Tendency: Usually when two or more different data sets are to be compared it is necessary to condense the data, but for comparison the condensation of data set into a frequency distribution and visual presentation are not enough. It is then necessary to summarize the data set in a single value. Such a value usually somewhere in the center and represent the entire data set and hence it is called measure of central tendency or averages. Since a measure of central tendency (i.e. an average) indicates the location or the general position of the distribution on the X-axis therefore it is also known as a measure of location or position.

Types of Measure of Central Tendency

1. Arithmetic Mean
2. Geometric Mean
3. Harmonic Mean
4. Mode
5. Median

Arithmetic Mean or Simply Mean: *"A value obtained by dividing the sum of all the observations by the number of observation is called arithmetic Mean"*

$$Mean = \frac{Sum\ of\ All\ observation}{Number\ of\ observation}$$

Methods	Ungrouped data	Grouped data
Direct Method	$\bar{x} = \frac{\sum x_i}{n}$	$\bar{x} = \frac{\sum fx}{n}$; Here $n = \sum f$
Short cut Method	$\bar{x} = A + \frac{\sum D}{n}$	$\bar{x} = A + \frac{\sum fD}{n}$; Here $n = \sum f$
	Where $D = X_i - A$ and A is the provisional or assumed mean.	
Step deviation Method	$\bar{x} = A + \frac{\sum u}{n} \times h$	$\bar{x} = A + \frac{\sum fu}{n} \times h$; Here $n = \sum f$
	Where $u = \frac{X_i - A}{h}$ and h is the common width of the class intervals	

Numerical Example:

Calculate the arithmetic mean for the following the marks obtained by 9 students are given below:

Using formula of arithmetic mean for ungrouped data:

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

$$n = 9$$

$$\bar{x} = \frac{360}{9} = 40 \ marks$$

x_i
45
32
37
46
39
36
41
48
36
$\sum_{i=1}^{n} x_i = 360$

- **Numerical Example:**

✧ **Calculate the arithmetic mean for the following data given below:**

◆ Using formula of **direct method** of **arithmetic mean** for **grouped data**:

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

The weight recorded to the nearest grams of 60 apples picked out at random from a consignment are given below:

106	107	76	82	109	107	115	93	187	95	123	125
111	92	86	70	126	68	130	129	139	119	115	128
100	186	84	99	113	204	111	141	136	123	90	115
98	110	78	185	162	178	140	152	173	146	158	194
148	90	107	181	131	75	184	104	110	80	118	82

Weight (grams)	Frequency
65----84	09
85----104	10
105----124	17
125----144	10
145----164	05
165----184	04
185----204	05

Solution:

Weight (grams)	Midpoints (x_i)	Frequency (f_i)	$f_i x_i$
65----84	$(65+84)/2=74.5$	09	9×74.5=670.5
85----104	94.5	10	945.0
105----124	114.5	17	1946.5
125----144	134.5	10	1345.0
145----164	154.5	05	772.5
165----184	174.5	04	698.0
185----204	194.5	05	972.5
		$\sum_{i=1}^{n} f_i = 60$	$\sum_{i=1}^{n} f_i x_i = 7350.0$

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{7350.0}{60} = 122.5 \text{ grams}$$ **(Answer).**

◆ Using formula of **short cut method** of **arithmetic mean** for **grouped data**:

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i D_i}{\sum_{i=1}^{n} f_i}, \quad n = \sum_{i=1}^{n} f_i$$

Where $D_i = X_i - A$ and A is the provisional or assumed mean

Weight (grams)	Midpoints (x_i)	Frequency (f_i)	$D_i = X_i - A$ $A = 114.5$	$f_i\ D_i$
65----84	$(65+84)/2 = 74.5$	09	-40	-360
85----104	94.5	10	-20	-200
105----124	114.5	17	0	0
125----144	134.5	10	20	200
145----164	154.5	05	40	200
165----184	174.5	04	60	240
185----204	194.5	05	80	400
		$\sum_{i=1}^{n} f_i = 60$		$\sum_{i=1}^{n} f_i D_i$ =480

$$\overline{x} = A + \frac{\sum_{i=1}^{n} f_i D_i}{\sum_{i=1}^{n} f_i} = 114.5 + \frac{480}{60} = 122.5 \text{ grams}$$ (Answer).

- Using formula of **step deviation method** of **arithmetic mean** for **grouped data**:

$$\overline{x} = A + \frac{\sum_{i=1}^{n} f_i u_i}{\sum_{i=1}^{n} f_i} \times h\,,\ \ u_i = \frac{x_i - A}{h}$$, where h is the width of the class interval:

Weight (grams)	Midpoints (x_i)	Frequency (f_i)	$u_i = \frac{X_i - A}{h}$ $A = 114.5$, h=20	$f_i\ u_i$
65----84	$(65+84)/2 = 74.5$	09	-2	-18
85----104	94.5	10	-1	-10
105----124	114.5	17	0	0
125----144	134.5	10	1	10
145----164	154.5	05	2	10
165----184	174.5	04	3	12
185----204	194.5	05	4	20
		$\sum_{i=1}^{n} f_i = 60$		$\sum_{i=1}^{n} f_i u_i$ =24

$$\overline{x} = A + \frac{\sum_{i=1}^{n} f_i u_i}{\sum_{i=1}^{n} f_i} \times h = 114.5 + \frac{24}{60} \times 20 = 114.5 + 08 = 122.5 \text{ grams}$$ (Answer).

Properties of Arithmetic Mean: The following are the properties of arithmetic mean:

- The mean of a constant is that constant.

  Proof: By definition of arithmetic mean: $\bar{x} = \frac{\sum xi}{n}$

  If "c" is any constant, then $\bar{x} = \frac{\sum c}{n}$

  $$\Rightarrow \bar{x} = \frac{nc}{n} \qquad (\because \sum c = nc)$$
  $$\Rightarrow \bar{x} = c$$

- The sum of deviations from mean is equal to zero. i.e. $\sum(xi - \bar{x}) = 0$

  Proof: Sum of Deviation
  $$= \sum(xi - \bar{x})$$
  $$= \sum xi - \sum \bar{x}$$
  $$= \sum xi - n\bar{x} \qquad (\because \bar{x} \text{ is constant})$$
  $$= \sum xi - n\left(\frac{\sum xi}{n}\right) \qquad \left(\because \bar{x} = \frac{\sum xi}{n}\right)$$
  $$= \sum xi - \sum xi$$
  $$= 0$$

- The sum of squared deviations from the mean is smaller than the sum of squared deviations from any arbitrary value or provisional mean. i.e. $\sum(xi - \bar{x})^2 < \sum(xi - A)^2$

  Proof: Taking $\sum(xi - A)^2$
  $$= \sum(xi - A + \bar{x} - \bar{x})^2$$
  $$= \sum[(xi - \bar{x}) + (\bar{x} - A)]^2$$
  $$= \sum[(xi - \bar{x})^2 + (\bar{x} - A)^2 - 2(xi - \bar{x})(\bar{x} - A)]$$
  $$= \sum(xi - \bar{x})^2 + \sum(\bar{x} - A)^2 - 2\sum(xi - \bar{x})(\bar{x} - A)$$
  $$= \sum(xi - \bar{x})^2 + n(\bar{x} - A)^2 - 2(\bar{x} - A)\sum(xi - \bar{x})$$
  $$= \sum(xi - \bar{x})^2 + n(\bar{x} - A)^2 \qquad \{\because \sum(xi - \bar{x}) = 0\}$$

  $$\Rightarrow \sum(xi - A)^2 < \sum(xi - \bar{x})^2 \qquad \{\because n(\bar{x} - A)^2 > 0\}$$

  Note: If $A = \bar{x}$ Then $\sum(xi - A)^2 = \sum(xi - \bar{x})^2$

- The arithmetic mean is affected by the change of origin and scale i.e. when a constant is added to or subtracted from each value of a variable or if each value of a variable is multiplied or divided by a constant, then arithmetic mean is affected by these changes.

Variable	Mean
Xi	$\bar{X}$
$Xi \pm a$	$\bar{X} \pm a$
$a\,Xi$	$a\,\bar{X}$
$\frac{X_i}{a}$	$\frac{\bar{X}}{a}$

- **Geometric Mean:** "The n^{th} root of the product of "*n*" positive values is called geometric mean"

$$Geometric\ Mean = \sqrt[n]{product\ of\ "n"\ positive\ values}$$

The following are the formulae of geometric mean:

Ungrouped data	Grouped data
$G = Antilog\left(\frac{\sum logx}{n}\right)$	$G = Antilog\left(\frac{\sum f\ logx}{n}\right)$;Here $n = \sum f$

- **Numerical example of geometric Mean for both grouped and ungrouped data:**
- **Calculate the geometric mean for the following the marks obtained by 9 students are given below:**

x_i	45	32	37	46	39	36	41	48	36

- Using formula of geometric mean for ungrouped data:

$n = 9$

x_i	$\log x_i$
45	$\log 45 = 1.65321$
32	1.50515
37	1.56820
46	1.66276
39	1.59106
36	1.55630
41	1.61278
48	1.62124
36	1.55630
	$\sum_{i=1}^{n} \log x_i = 14.38700$

$$G.M = anti-\log\left(\frac{\sum_{i=1}^{n} \log x_i}{n}\right)$$

$$G.M = anti-\log\left(\frac{14.38700}{9}\right)$$

$$G.M = anti-\log(1.59856)$$

$$G.M = 39.68$$ (Answer).

- Given the following frequency distribution of weights of 60 apples, calculate the geometric mean for grouped data.

Weights (grams)	65--84	85--104	105--124	125--144	145--164	165--184	185--204
Frequency	09	10	17	10	05	04	05

$$G.M = anti-\log\left(\frac{\sum_{i=1}^{n} f_i \log x_i}{\sum_{i=1}^{n} f_i}\right)$$

Weight (grams)	Midpoints (x_i)	Frequency (f_i)	$\log x_i$	$f_i \log x_i$
65----84	$(65+84)/2 = 74.5$	09	1.8722	16.8498
85----104	94.5	10	1.9754	19.7540
105----124	114.5	17	2.0589	35.0013
125----144	134.5	10	2.1287	21.2870
145----164	154.5	05	2.1889	10.9445
165----184	174.5	04	2.2418	8.9672
185----204	194.5	05	2.2889	11.4445
		$\sum_{i=1}^{n} f_i = 60$		$\sum_{i=1}^{n} f_i \log x_i =$ 124.2483

$$G.M = anti-\log\left(\frac{\sum_{i=1}^{n} f_i \log x_i}{\sum_{i=1}^{n} f_i}\right)$$

$$= anti-\log\left(\frac{124.2483}{60}\right)$$

$$= anti-\log(2.0708)$$

$G.M = 117.7$ grams (Answer).

- **Harmonic Mean:** *"The reciprocal of the Arithmetic mean of the reciprocal of the values is called Harmonic mean"*

$$Harmonic\ Mean = reciprocal\ of\left(\frac{Sum\ of\ reciprocal\ of\ the\ values}{The\ number\ of\ values}\right)$$

The following are formulae of harmonic mean:

Ungrouped data	Grouped data
$H = \dfrac{n}{\sum\left(\frac{1}{x}\right)}$	$H = \dfrac{n}{\sum\left(\frac{f}{x}\right)}$; Here $n = \sum f$

- **Numerical example of harmonic Mean for both grouped and ungrouped data:**
- **Calculate the harmonic mean for the following the marks obtained by 9 students are given below:**

x_i	45	32	37	46	39	36	41	48	36

- Using formula of harmonic mean for ungrouped data:

$n = 9$

x_i	$1/x_i$
45	0.02222
32	0.03125
37	0.02702
46	0.02173
39	0.02564
36	0.02777
41	0.02439
48	0.02083
36	0.02777
	$\sum_{i=1}^{n} \frac{1}{x_i}$ =0.22862

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

$$H.M = \frac{9}{0.22862}$$

$$H.M = 39.36663$$ (Answer).

- Given the following frequency distribution of weights of 60 apples, calculate the harmonic mean for grouped data.

Weights (grams)	65--84	85--104	105--124	125--144	145--164	165--184	185--204
Frequency	09	10	17	10	05	04	05

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

Weight (grams)	Midpoints (x_i)	Frequency (f_i)	f_i/x_i
65----84	$(65+84)/2=74.5$	09	0.12081
85----104	94.5	10	0.10582
105----124	114.5	17	0.14847
125----144	134.5	10	0.07435
145----164	154.5	05	0.03236
165----184	174.5	04	0.02292
185----204	194.5	05	0.02571
		$\sum_{i=1}^{n} f_i = 60$	$\sum_{i=1}^{n} \frac{f_i}{x_i}$=0.53044

$$H.M = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n}\left(\frac{f_i}{x_i}\right)} = \frac{60}{0.53044} = 113.11\,\text{grams}$$ (Answer).

❖ **Median:** *"when the observation are arranged in ascending or descending order, then a value, that divides a distribution into equal parts, is called median"*

Median in case of Ungrouped Data	
In this case we first arrange the observations in **increasing** or **decreasing** order then we use the following formulae for Median:	
If "n" is odd	$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$
If "n" is even	$Median = \frac{size\ of \left\{\left(\frac{n}{2}\right)th + \left(\frac{n}{2}+1\right)th\right\} observation}{2}$

➢ **Numerical example of median for both grouped and ungrouped data:**

If "n" is odd	$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$

➢ **Calculate the median for the following the marks obtained by 9 students are given below:**

x_i	45	32	37	46	39	36	41	48	36

Arrange the data in

ascending order

32, 36, 36, 37, 39, 41, 45, 46, 48. $n = 9$ "n" is odd

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

$$Median = Size\ of\ \left(\frac{9+1}{2}\right)^{th} observation$$

$$Median = Size\ of\ 5^{th} observation$$

$$Median = 39$$ (Answer).

If "n" is even	$Median = \dfrac{size\ of\ \left\{\left(\frac{n}{2}\right)th + \left(\frac{n}{2}+1\right)th\right\} observation}{2}$

- **Calculate the median for the following the marks obtained by 10 students are given below:**

x_i	45	32	37	46	39	36	41	48	36	50

Arrange the data in ascending order

32, 36, 36, 37, 39, 41, 45, 46, 48, 50. $n = 10$ "n" is even

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

$$Median = \frac{Size\ of\ \left\{\left(\frac{10}{2}\right)^{th} + \left(\frac{10}{2}+1\right)^{th}\right\} observation}{2}$$

$$Median = \frac{Size\ of\ \left\{5^{th} + 6^{th}\right\} observation}{2}$$

$$Median = \frac{39+41}{2} = 40$$ (Answer).

- The number of values above the median balances (equals) the number of values below the median i.e. 50% of the data falls above and below the median.

Median in case of Discrete Grouped Data

In case of discrete grouped data, first we find the cumulative frequencies and then use the following formula for Median:

$$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$$

Here $n = \sum f$

- Numerical examples: The following distribution relates to the number of assistants in 50 retail establishments.

No.of assistant	0	1	2	3	4	5	6	7	8	9
f	3	4	6	7	10	6	5	5	3	1

No. of assistants	f_i	Cumulative frequency ($c.f$)
0	3	3
1	4	7
2	6	13
3	7	20
4	10	30
5	6	36
6	5	41
7	5	46
8	3	49
9	1	50
	$\sum_{i=1}^{n} f_i = 50$	

$n = \sum_{i=1}^{n} f_i = 50$ "n" is even

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

$$Median = \frac{Size\ of \left\{\left(\frac{50}{2}\right)^{th} + \left(\frac{50}{2}+1\right)^{th}\right\} observation}{2}$$

$$Median = \frac{Size\ of\ \left\{25^{th} + 26^{th}\right\}\ observation}{2}$$

$$Median = \frac{4+4}{2} = 4 \qquad \text{(Answer).}$$

Median in case of continuous Grouped Data

In continuous grouped data, when we are finding median, we first construct the class boundaries if the classes are discontinuous. Then we find cumulative frequencies and then we use the following two steps:

- First we determine the median class using n/2.
- When the median class is determined, then the following formula is used to find the value of median. i.e.

$$Median = l + \frac{h}{f}\left(\frac{n}{2} - C\right); \quad \text{Here } n = \sum f$$

Where l = lower class boundary of the median class
h = width of the median class
f = frequency of the median class
C = cumulative frequency of the class preceding the median class.

❖ Numerical example: Find the median, for the distribution of examination marks given below:

Marks	30--39	40--49	50--59	60--69	70--79	80--89	90--99
Number of students	08	87	190	304	211	85	20

$$Median = l + \frac{h}{f}\left(\frac{n}{2} - C\right) \quad \text{here} \quad n = \sum_{i=1}^{n} f_i$$

Class boundaries	Midpoints (x_i)	Frequency (f_i)	Cumulative frequency $(c.f)$
29.5---39.5	34.5	8	8
39.5---49.5	44.5	87	95
49.5---59.5	54.5	190	285
59.5---69.5	64.5	304	589
69.5---79.5	74.5	211	800
79.5---89.5	84.5	85	885
89.5---99.5	94.5	20	905
		$\sum_{i=1}^{n} f_i$ =905	

$\frac{n}{2} = \frac{905}{2}$ =452.5th student which corresponds to marks in the class 59.5---69.5

Therefore

$$Median = l + \frac{h}{f}\left(\frac{n}{2} - C\right) = 59.5 + \frac{10}{304}\left(\frac{905}{2} - 285\right) = 59.5 + \frac{10}{304}(452.5 - 285)$$

$$Median = 59.5 + \frac{1675}{304}, \quad Median = 59.5 + 5.5, \quad Median = 65 \text{ marks} \qquad \text{(Answer).}$$

- **Quartiles:** ***"when the observation are arranged in increasing order then the values, that divide the whole data into four (4) equal parts, are called quartiles"***

These values are denoted by Q_1, Q_2 and Q_3. It is to be noted that 25% of the data falls below Q_1, 50% of the data falls below Q_2 and 75% of the data falls below Q_3.

- **Deciles:** ***"when the observation are arranged in increasing order then the values, that divide the whole data into ten (10) equal parts, are called quartiles"***

These values are denoted by $D_1, D_2, \ldots, D_9$. It is to be noted that 10% of the data falls below D_1, 20% of the data falls below D_2,…, and 90% of the data falls below D_9.

- **Percentiles**: ***"when the observation are arranged in increasing order then the values, that divide the whole data into hundred (100) equal parts, are called quartiles"***

These values are denoted by $P_1, P_2, \ldots, P_{99}$. It is to be noted that 1% of the data falls below P_1, 2% of the data falls below P_2,…, and 99% of the data falls below P_{99}.

Measures	Data Type	Formulas
Quartiles j = 1, 2, 3	Ungrouped Data	$Q_j = size\ of \left(\frac{j(n+1)}{4}\right) th\ observation$
	Discrete Grouped data	$Q_j = size\ of \left(\frac{j(n+1)}{4}\right) th\ observation$; Here $n = \sum f$
Deciles j = 1, 2, . .,9	Ungrouped Data	$D_j = size\ of \left(\frac{j(n+1)}{10}\right) th\ observation$
	Discrete Grouped data	$D_j = size\ of \left(\frac{j(n+1)}{10}\right) th\ observation$; Here $n = \sum f$
Percentiles j = 1, 2, . .,99	Ungrouped Data	$P_j = size\ of \left(\frac{j(n+1)}{100}\right) th\ observation$
	Discrete Grouped data	$P_j = size\ of \left(\frac{j(n+1)}{100}\right) th\ observation$; Here $n = \sum f$

$Median = Q_2 = D_5 = P_{50}$ are same and are equal to median.

- **Calculate quartiles for ungrouped data.**
- **Calculate the quartiles for the following the marks obtained by 9 students are given below:**

x_i	45	32	37	46	39	36	41	48	36

Arranged the observation in ascending order

32, 36, 36, 37, 39, 41, 45, 46, 48.

If "*n*" is odd then we use the below formula:

$Q_j = Marks\ obtained\ by \left[\left(\frac{jn}{4}\right)+1\right]^{th} student$, generalized formula of quartiles

where *j*=1, 2, 3.

Put *j*=1, *n*=9, $Q_1 = Marks\ obtained\ by \left[\left(\frac{n}{4}\right)+1\right]^{th} student$

$$Q_1 = Marks\ obtained\ by \left[\left(\frac{9}{4}\right)+1\right]^{th} student$$

$$Q_1 = Marks\ obtained\ by\ [2.25+1]^{th} student$$

$$Q_1 = Marks\ obtained\ by\ [2+1]^{th}\ student$$

$$Q_1 = Marks\ obtained\ by\ 3^{th} student$$

$$Q_1 = 36 \quad \text{(Answer).}$$

Calculate Q_2 and Q_3 from above distribution?

- Calculate quartiles for discrete grouped data.

$$Q_3 = Marks\ obtained\ by\ \left[\left(\frac{3n}{4}\right)+1\right]^{th}\ student$$

No. of assistants	f_i	Cumulative frequency ($c.f$)
0	3	3
1	4	7
2	6	13
3	7	20
4	10	30
5	6	36
6	5	41
7	5	46
8	3	49
9	1	50
	$\sum_{i=1}^{n} f_i = 50$	

$$Q_3 = Marks\ obtained\ by\ \left[\left(\frac{3\times 50}{4}\right)+1\right]^{th}\ student$$

$$Q_3 = Marks\ obtained\ by\ \left[\left(\frac{150}{4}\right)+1\right]^{th}\ student$$

$$Q_3 = Marks\ obtained\ by\ 38^{th} student$$

$$Q_3 = 6 \qquad \text{(Answer).}$$

- **Calculate Decile for ungrouped data.**

- **Calculate the Decile for the following the marks obtained by 9 students are given below:**

x_i	45	32	37	46	39	36	41	48	36

Arranged the observation in ascending order

32, 36, 36, 37, 39, 41, 45, 46, 48.

If "n" is odd then we use the below formula:

$$D_j = Marks\ obtained\ by\ \left[\left(\frac{jn}{10}\right)+1\right]^{th} student$$, generalized formula of quartiles

where j=1, 2, 3, 4, 5, 6, 7, 8, 9.

Put j=4, n=9, $D_4 = Marks\ obtained\ by\ \left[\left(\frac{4n}{10}\right)+1\right]^{th} student$

$$D_4 = Marks\ obtained\ by\ \left[\left(\frac{4\times 9}{10}\right)+1\right]^{th} student$$

$$D_4 = Marks\ obtained\ by\ [3.6+1]^{th} student$$

$$D_4 = Marks\ obtained\ by\ [3+1]^{th} student$$

$$D_4 = Marks\ obtained\ by\ 4^{th} student$$

$$D_4 = 37 \quad \text{(Answer).}$$

Calculate D_3, D_7 and D_9 from above distribution?

❖ Calculate decile for discrete grouped data.

$$D_6 = Marks\ obtained\ by\ \left[\left(\frac{6n}{10}\right)+1\right]^{th} student$$

No. of assistants	f_i	Cumulative frequency ($c.f$)
0	3	3
1	4	7
2	6	13
3	7	20
4	10	30
5	6	36
6	5	41
7	5	46
8	3	49
9	1	50
	$\sum_{i=1}^{n} f_i$ =50	

$$D_6 = Marks\ obtained\ by\ \left[\left(\frac{6\times 50}{10}\right)+1\right]^{th} student$$

$$D_6 = Marks\ obtained\ by\ \left[\left(\frac{300}{10}\right)+1\right]^{th}\ student$$

$$D_6 = Marks\ obtained\ by\ 31^{th} student$$

$$Q_3 = 5 \qquad \text{(Answer).}$$

- **Calculate percentile for ungrouped data.**
- **Calculate the percentile for the following the marks obtained by 9 students are given below:**

x_i	45 32 37 46 39 36 41 48 36

Arranged the observation in ascending order

32, 36, 36, 37, 39, 41, 45, 46, 48.

If "n" is odd then we use the below formula:

$P_j = Marks\ obtained\ by\ \left[\left(\frac{jn}{100}\right)+1\right]^{th}\ student$, generalized formula of quartiles

where j=1, 2, …, 99.

Put j=68, n=9, $P_{68} = Marks\ obtained\ by\ \left[\left(\frac{68n}{100}\right)+1\right]^{th}\ student$

$$P_{68} = Marks\ obtained\ by\ \left[\left(\frac{68\times 9}{100}\right)+1\right]^{th}\ student$$

$$P_{68} = Marks\ obtained\ by\ \left[6.12+1\right]^{th} student$$

$$P_{68} = Marks\ obtained\ by\ \left[6+1\right]^{th} student$$

$$P_{68} = Marks\ obtained\ by\ 7^{th} student$$

$$D_4 = 45 \quad \text{(Answer).}$$

Calculate P_9, P_{55}, P_{80} and P_{95} from above distribution?

- Calculate percentile for discrete grouped data.

$$P_{49} = Marks\ obtained\ by\ \left[\left(\frac{49n}{100}\right)+1\right]^{th}\ student$$

No. of assistants	f_i	Cumulative frequency ($c.f$)
0	3	3
1	4	7
2	6	13
3	7	20
4	10	30
5	6	36
6	5	41
7	5	46
8	3	49
9	1	50
	$\sum_{i=1}^{n} f_i = 50$	

$$P_{49} = Marks\ obtained\ by\ \left[\left(\frac{49 \times 50}{100}\right) + 1\right]^{th} student$$

$$P_{49} = Marks\ obtained\ by\ \left[\left(\frac{2450}{100}\right) + 1\right]^{th} student$$

$$P_{49} = Marks\ obtained\ by\ 24^{th} student$$

$$Q_3 = 4 \qquad \text{(Answer).}$$

Continuous Grouped Data	
In continuous grouped data, we use the following two steps:	
Quartiles	• First we determine the j^{th} quartile class using jn/4. • When the j^{th} quartile class is determined, then the following formula is used to find the value of j^{th} quartile i.e. $$Q_j = l + \frac{h}{f}\left(\frac{jn}{4} - C\right); \quad \text{Here } n = \sum f$$ l = lower class boundary of the j^{th} quartile class h = width of the j^{th} quartile class f = frequency of the j^{th} quartile class C = cumulative frequency of the class preceding the j^{th} quartile class.
Deciles	• First we determine the j^{th} decile class using jn/10. • When the j^{th} decile class is determined, then the following formula is used to find the value of the j^{th} decile. i.e. $$D_j = l + \frac{h}{f}\left(\frac{jn}{10} - C\right); \quad \text{Here } n = \sum f$$ l = lower class boundary of the j^{th} decile class h = width of the j^{th} decile class f = frequency of the j^{th} decile class C = cumulative frequency of the class preceding the j^{th} decile class.
Percentiles	• First we determine j^{th} percentile class using jn/100. • When the j^{th} percentile class is determined, then the following formula is used to find the value of the j^{th} percentile. i.e. $$P_j = l + \frac{h}{f}\left(\frac{jn}{100} - C\right); \quad \text{Here } n = \sum f$$ l = lower class boundary of the j^{th} percentile class h = width of the j^{th} percentile class f = frequency of the j^{th} percentile class C = cumulative frequency of the class preceding the j^{th} percentile class.

❖ Numerical example of Quartile, Decile and Percentile for continuous grouped data:

Class boundaries	Midpoints (x_i)	Frequency (f_i)	Cumulative frequency $(c.f)$
29.5---39.5	34.5	8	8
39.5---49.5	44.5	87	95
49.5---59.5	54.5	190	285
59.5---69.5	64.5	304	589
69.5---79.5	74.5	211	800
79.5---89.5	84.5	85	885
89.5---99.5	94.5	20	905
		$\sum_{i=1}^{n} f_i$ =905	

Calculate the 3rd quartile

$$Q_2 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right)$$

$$\frac{3n}{4} = \frac{3\times 905}{4} = 678.75$$

$$Q_2 = 69.5 + \frac{10}{211}(678.75 - 589)$$

$$Q_2 = 69.5 + \frac{897.5}{211}$$

$$Q_2 = 73.7535 \quad \text{(Answer)}$$

Calculate the 4th and 8th decile

$$D_4 = l + \frac{h}{f}\left(\frac{4n}{10} - C\right)$$

$$\frac{4n}{10} = \frac{4\times 905}{10} = 362$$

$$D_4 = 59.5 + \frac{10}{304}(362 - 285)$$

$$D_4 = 59.5 + \frac{770}{304}$$

$$D_4 = 62.0328 \quad \text{(Answer)}$$

$$D_8 = l + \frac{h}{f}\left(\frac{8n}{10} - C\right)$$

$$\frac{8n}{10} = \frac{8\times 905}{10} = 724$$

$$D_8 = 69.5 + \frac{10}{211}(724 - 589)$$

$$D_8 = 69.5 + \frac{1350}{211}$$

$$D_8 = 75.8981 \quad \text{(Answer)}$$

Calculate the 18th and 76th Percentile

$$P_{18} = l + \frac{h}{f}\left(\frac{18n}{100} - C\right)$$

$$\frac{18 \times 905}{100} = 162.9$$

$$P_{18} = 49.5 + \frac{10}{190}(162.5 - 95)$$

$$P_{18} = 49.5 + \frac{675}{190} = 53.05 \quad \text{(Answer).}$$

$$P_{76} = l + \frac{h}{f}\left(\frac{76n}{100} - C\right)$$

$$\frac{76 \times 905}{100} = 687.8$$

$$P_{76} = 69.5 + \frac{10}{211}(687.8 - 589)$$

$$P_{76} = 69.5 + \frac{988}{211} = 74.182 \quad \text{(Answer).}$$

Calculate Q_2, D_5, D_8, P_{25}, P_{50} and P_{88}?

Main Objects of Average

- The main object (purpose) of the average is to give a bird's eye view (summary) of the statistical data. The average removes all the unnecessary details of the data and gives a concise (to the point or short) picture of the huge data under investigation.
- Average is also of great use for the purpose of comparison (i.e. the comparison of two or more groups in which the units of the variables are same) and for the further analysis of the data.
- Averages are very useful for computing various other statistical measures such as dispersion, skewness, kurtosis etc.

❖ **Requisites (desirable qualities) of a Good Average:** An average will be considered as good if:

- It is mathematically defined.
- It utilizes all the values given in the data.

- It is not much affected by the extreme values.
- It can be calculated in almost all cases.
- It can be used in further statistical analysis of the data.
- It should avoid to give misleading results.

❖ **Uses of Averages in Different Situations**

- **A.M** is an appropriate average for **all the situations** where there are no extreme values in the data.
- **G.M** is an appropriate average for calculating **average percent increase** in sales, population, production, etc. It is one of the best averages for the construction of **index numbers**.
- **H.M** is an appropriate average for calculating the **average rate of increase** of profits of a firm or finding average speed of a journey or the average price at which articles are sold.
- **Mode** is an appropriate average in case of **qualitative data** e.g. the opinion of an average person; he is probably referring to the most frequently expressed opinion which is the modal opinion.
- **Median** is an appropriate average in a **highly skewed distribution** e.g. in the distribution of wages, incomes etc.

Mode in case of Ungrouped Data: *"A value that occurs most frequently in a data is called mode"*

OR

"if two or more values occur the same number of times but most frequently than the other values, the there is more than one whole"

"If two or more values occur the same number of times but most frequently than the other values, then there is more than one mode"

- The data having one mode is called uni-modal distribution.
- The data having two modes is called bi-modal distribution.
- The data having more than two modes is called multi-modal distribution.

Mode in case of Discrete Grouped Data: *"A value which has the largest frequency in a set of data is called mode"*

Mode in case of Continuous Grouped Data: In case of continuous grouped data, mode would lie in the class that carries the highest frequency. This class is called the modal class. The formula used to compute the value of mode, is given below:

❖ Numerical examples of Mode for ungrouped and grouped data

- Calculate Mode for ungrouped data

x_i : 2, 3, 8, 4, 6, 3, 2, 5, 3.

Mode = 3 (Answer).

- Calculate Mode in discrete grouped data

No. of assistants	f_i
0	3
1	4
2	6
3	7
4	10
5	6
6	5
7	5
8	3
9	1
	$\sum_{i=1}^{n} f_i$ =50

Mode = 4

- Mode in case of Continuous grouped data:

Class boundaries	Midpoints (x_i)	Frequency (f_i)	Cumulative frequency $(c.f)$
29.5---39.5	34.5	8	8
39.5---49.5	44.5	87	95
49.5---59.5	54.5	190	285
59.5---69.5	64.5	304	589
69.5---79.5	74.5	211	800
79.5---89.5	84.5	85	885
89.5---99.5	94.5	20	905
		$\sum_{i=1}^{n} f_i$ =905	

$$Mode = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h \quad = 59.5 + \frac{304 - 190}{(304 - 190) + (304 - 211)} \times 10$$

$$Mode = 59.5 + \frac{114}{114 + 93} \times 10 \quad \Rightarrow Mode = 59.5 + 5.05072$$

Mode = 4 (Answer)

Q: What is a measure of location? What is the purpose served by it? What are its desirable qualities?

Measure of location: A central value that represents the whole data is called an average. Since average is a value usually somewhere in the center and represents the entire data set therefore it is called measure of central tendency. Measure of central tendency indicates the location or the general position of the data on the X-axis therefore it is also known as a measure of location or position

Purpose:

- It removes all the unnecessary details of the data and gives a concise picture of the huge data.
- It is used for the purpose of comparison.
- It is very useful in computing other statistical measures such as dispersion, skewness and kurtosis etc.

Desirable qualities of a good average: An average will be considered as good if:

- It is mathematically defined.
- It utilizes all the observations given in a data.
- It is not much affected by the extreme values.
- It is capable of further algebraic treatment.
- It is not affected by fluctuations of sampling.