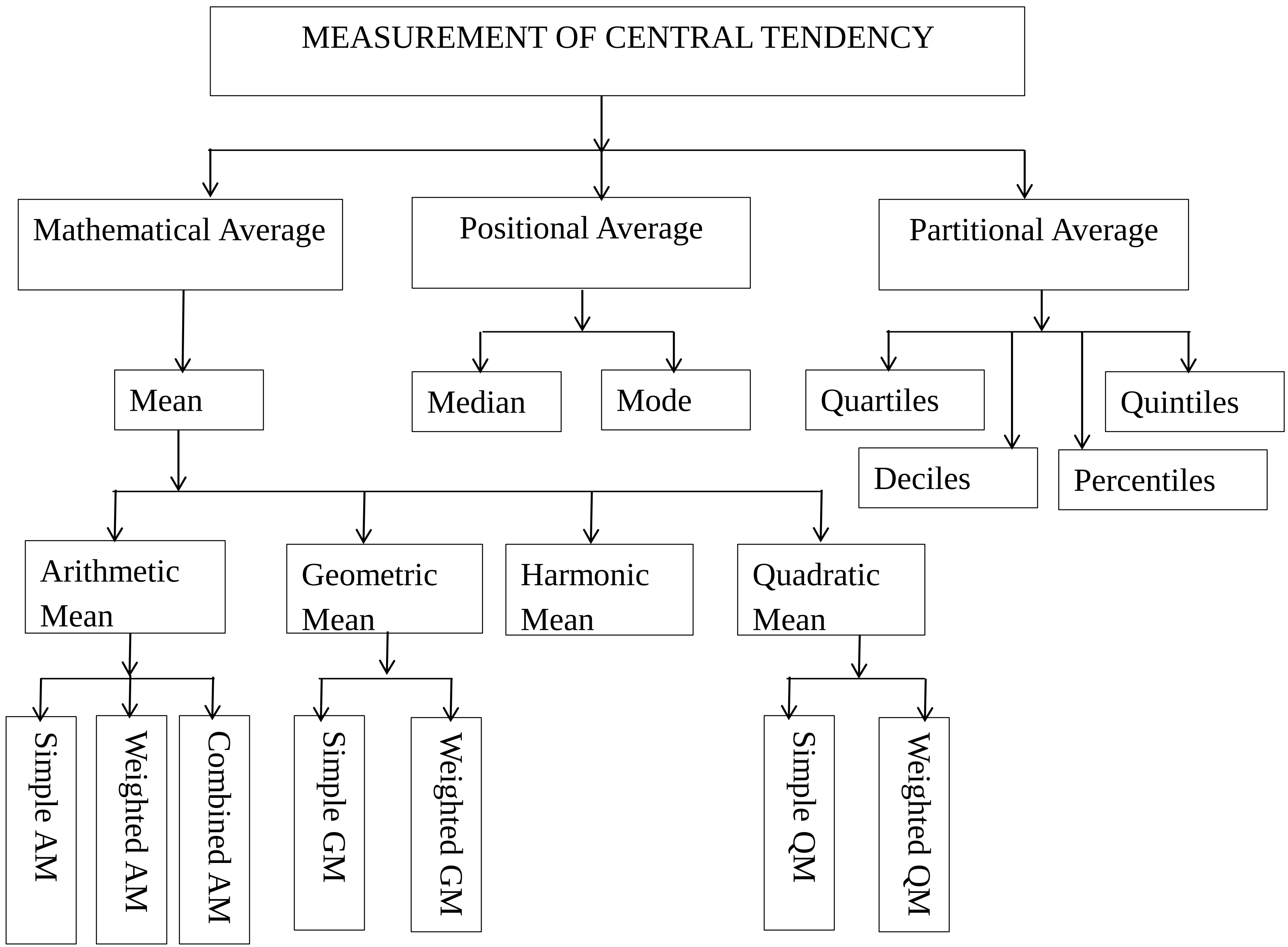
Unit 4: Measures of Central Tendencies and Variability

For
BCA sixth Semester
PU

Course of Study

- *Measure of central tendencies:* 10 hrs.
  - Mean
  - Median
  - Mode
  - Mid-hinge
- Measures of Variation
  - Range
  - Interquartile range
  - Standard deviations
  - Coefficient of variations (CV)
- Shape of five number summary
- Box and whisker plot


MEASUREMENT OF CENTRAL TENDENCY
Mathematical Average
Positional Average
Partitional Average
Mean
Median
Mode
Quartiles
Quintiles
Deciles
Percentiles
Arithmetic Mean
Geometric Mean
Harmonic Mean
Quadratic Mean
Simple AM
Weighted AM
Combined AM
Simple GM
Weighted GM
Simple QM
Weighted QM

Introduction

According to Croxton and Cowden, "*An average value is a single value within the range of the data is used to represent all of the values in the series. Since an average is somewhere in the range of the data, it is something called a measure of central value.*"

Mean (A. M.) *is defined as the ratio between the sum of observations and the number of observations.*

Calculations

1. $\overline{\mathrm{X}} = \frac{\mathrm{X_1 + X_2 + \cdots \ldots + X_n}}{\mathrm{n}}$

$$= \frac{\sum_1^n X_i}{n}$$

2. $\overline{\mathrm{X}} = \frac{\mathrm{X_1.f_1 + X_2.f_2 + \cdots \ldots + X_n.f_n}}{\mathrm{f_1 + f_2 + \cdots + f_n}}$

$$= \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{f_i . X_i})}{\mathrm{n}}$$

3. $\overline{\mathrm{X}} = \frac{\sum \mathrm{f_i.m_i}}{\mathrm{n}}$

where n = number of items, m = mid value of the respective groups.

Contd…

- Weighted arithmetic mean:
- $\overline{X}_w = \frac{W_1.X_1+W_2.X_2+\cdots+W_n.X_n}{W_1+W_2+\cdots+W_n} = \frac{\sum_{i=1}^{n} W_i X_i}{\sum_{i=1}^{n} W_i}$
- Combined mean:
- Let $\overline{X}_1$ and $\overline{X}_2$ are the arithmetic means of first and second series where n_1 and n_2 are the number of observations respectively. Then their combined mean denoted by $(\overline{\overline{X}})$ is obtained by: $\overline{\overline{X}} = \frac{n_1\ .\overline{X}_1 + n_2\ .\ \overline{X}_2}{n_1 + n_2}$
- Let $\overline{X}_1$, $\overline{X}_2$ & $\overline{X}_3$ are the arithmetic means of first, second and third series where n_1, n_2 & n_3 are the number of observations respectively. Then their combined mean denoted by $(\overline{\overline{X}})$ is obtained by: $\overline{\overline{X}} = \frac{n_1\ .\overline{X}_1 + n_2\ .\ \overline{X}_2 + n_3\ .\ \overline{X}_3}{n_1 + n_2 + n_3}$

Merits

- ***Merits***
- It is easy to understand and simple to compute.
- Its value is based on each and every item of the data with the result a change in any item would mean a change in the average itself.
- It is most commonly used in further statistical computation.
- Arrangement of data is not required while computing arithmetic mean.
- It is rigidly defined by an algebraic formula.
- It is based on all the observations.
- It is affected by the value of every item in the series.
- Of all averages, arithmetic mean is least affected by the fluctuations of sampling.
- It provides a good basis for the comparison of two or more distribution.

Demerits

- ***Demerits***
- It is seriously affected by the extreme values.
- For example: the A. M. of 1, 100 and 1000 is $\overline{\mathrm{X}} = \frac{1+100+1000}{3} = \frac{1101}{3} = 367$. If we exclude the item 1, then A. M. would be $\overline{\mathrm{X}} = \frac{100+1000}{2} = \frac{1100}{2} = 550$again we omit 1000 then A. M. would be $\overline{\mathrm{X}} = \frac{1+100}{2} = \frac{101}{2} = 50.5$
- Its value cannot be determined graphically.
- A. M. cannot be computed from qualitative data like: love, safety, honesty, beauty etc.
- It fails to average the ratios and percentage properly.

Geometric mean

- Geometric mean (G. M.): is the n^{th} root of the product of 'n' observations.

1. $\mathrm{G.M.} = \sqrt[\mathrm{n}]{\mathrm{X_1.X_2 \dots X_n}} = (\mathrm{X_1.X_2 \dots X_n})^{\frac{1}{\mathrm{n}}}$

$$= \mathrm{antilog}\left(\frac{\sum \log\ \mathrm{X}}{\mathrm{n}}\right)$$

2. $GM = Antilog\left[\frac{\sum f \log X}{n}\right]$
3. $GM = Antilog\left[\frac{\sum f \log X}{n}\right]$

- Example: Find the GM of 2, 4 and 8.
- Solution $GM = \sqrt[3]{2.4.8} = \sqrt[3]{64} = \sqrt[3]{4^3} = 4$

Harmonic mean

- H. M: is defined as the reciprocal of the arithmetic mean of the reciprocal of the given non-zero observations.

Individual	Discrete	Group
$\frac{\mathrm{N}}{\left(\frac{1}{\mathrm{X}_1}+\frac{1}{\mathrm{X}_2}+\cdots+\frac{1}{\mathrm{X}_\mathrm{n}}\right)}=\frac{\mathrm{N}}{\sum\left(\frac{1}{\mathrm{X}}\right)}$	$\frac{N}{\sum f\left(\frac{1}{X}\right)}$	$\frac{N}{\sum f\left(\frac{1}{m}\right)}$

Example: A train starts from a rest and travels successively quarters of a mile at average speed of 12, 16, 24 and 48 miles per hours. Find the average speed of the train.

- Solution: Here the appropriate average is the harmonic mean of 12, 16, 24, 48; and N = 4. Now $\mathrm{H.M.}=\frac{\mathrm{N}}{\left(\frac{1}{\mathrm{X}_1}+\frac{1}{\mathrm{X}_2}+\frac{1}{X_3}+\frac{1}{\mathrm{X}_4}\right)}$
- $=\frac{4}{\left(\frac{1}{12}+\frac{1}{16}+\frac{1}{24}+\frac{1}{48}\right)}=\frac{4}{\left(\frac{4+3+2+1}{48}\right)}=\frac{4}{\left(\frac{10}{48}\right)}=\frac{4\ x\ 48}{10}=19.2\ miles/h$

Partition Values

- Partitions are the values, which divide the distribution into number of equal parts. The partition values are:
- ***Median, Quartiles, Deciles and Percentiles***.

i. Median:

- Median (M_d) = size of $\left(\frac{n+1}{2}\right)^{th}$ item (For individual and discrete)
- Median (M_d) = size of $\left(\frac{n}{2}\right)^{th}$ item. (for continuous)
- To find the actual median have to use the formula
- $M_d = L + \frac{h}{f}\left(\frac{n}{2} - c.f.\right)$
- Note: if inclusive class interval is given, we should convert it to the exclusive class intervals by correction factor using the relation
- c_f = (½){lower limit of second class – upper limit of first class}

Merits

- ***Merits***
- Median is simple to understand.
- Median is positional average. It is not affected by extreme items.
- For open-end classes, median is the most appropriate average.
- Its value can be determined graphically and by inspection.
- Median is epically useful in qualitative phenomena like honesty, inelegancy, efficiency etc.

Demerits

- ***Demerits***
- To find median, the given data should arranged in ascending or descending order but other averages do not need any arrangement.
- It is not always rigidly defined i.e. in case if number of observations is even, then there are two middle most values and in such case, median is obtained by taking the arithmetic mean of two middle most values. It cannot be determined exactly.
- Since it is a positional average, it is not based on each and every item of the distribution.
- Median is not suitable for further mathematical treatment.
- It is not as familiar as arithmetic mean.
- If the number of items is small, mean is more suitable then median.

Quartiles

- The variate values which divides the given distribution into four equal parts is called quartiles. To divide into four equal parts there are three quartiles namely: Q_1(known as lower quartile), Q_2 (known as median) and Q_3 (known as upper quartile).
- Case i: when an individual and discrete data is given:
- Size of Q_i = the value of $\frac{i\ x\ (n+1)}{4}^{th}\ item$, where ‘i’ = 1, 2 and 3.
- Case ii: when a group data is given:
- Size of Q_i = the value of $\frac{i\ x\ (n)}{4}^{th}\ item$, where ‘i’ = 1, 2 and 3.
- To get actual quartile we further have to apply the formula:
- $Q_i = L + \frac{h}{f}\left(\frac{i\ x\ n}{4} - c.f.\right)$
- Where, L = lower limit of the quartile class, h = class size, f = frequency of respective quartile class, n = total frequency
- c. f. = cumulative frequency of respective preceding the quartile class.

Deciles

- The variate values which divides the given distribution into ten equal parts is called Deciles. To divide into ten equal parts there are nine Deciles namely: $D_1, D_2, \ldots, D_9$.
- Case i: when an individual and discrete data is given:
- Size of D_i = the value of $\frac{i\,x\,(n+1)}{10}^{th}$ $item,$
- where 'i' = 1, 2,…,9.
- Case ii: when a group data is given:
- Size of D_i = the value of $\frac{i\,x\,(n)}{10}^{th}$ $item,$
- where 'i' = 1, 2,…,9.
- To get actual Deciles we further have to apply the formula: $D_i = L + \frac{h}{f}\left(\frac{i\,x\,n}{10} - c.f.\right)$
- Where, L = lower limit of the Deciles class, h = class size,
- f = frequency of respective Deciles

Percentiles

- The variate values which divides the given distribution into hundred equal parts is called Percentiles. To divide into hundred equal parts there are 99 Percentiles namely: P_1, P_2, ….,P_{99}.
- Case i: when an individual and discrete data is given:
- Size of P_i = the value of $\frac{i\ x\ (n+1)}{100}^{th}\ item,$
- where 'i' = 1, 2,….,99.
- Case ii: when a group data is given:
- Size of P_i = the value of $\frac{i\ x\ (n)}{100}^{th}\ item,$
- where 'i' = 1, 2,….,99.
- To get actual Percentiles we further have to apply the formula:
- $P_i = L + \frac{h}{f}\left(\frac{i\ x\ n}{100} - c.f.\right)$
- Where, L = lower limit of the Percentiles class, h = class size,
- f = frequency of respective Percentiles class, n = total frequency,
- c. f. = cumulative frequency of respective preceding the Percentiles class.

Mode

- Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster densely.
- ***Case i:*** *Determination of mode in individual and discrete data*: item(s) that is repeated maximum number of times would be referred to as mode of the given data.
- ***Case ii: Determination of mode in group data***: in group data (having exclusive class) the class corresponding to the maximum frequency is called the modal class and actual mode is defined by the following formula: $M_0 = L + \frac{f_1 - f_2}{2f_1 - f_0 - f_2}.h = L + \frac{\Delta_1}{\Delta_1 + \Delta_2}.h$
- Where,
- L = lower limit of the modal class.
- (f_1) = frequency of the modal class.
- (f_0) = frequency preceding modal class.
- (f_2) = frequency following modal class.
- (h) = class size of modal class.
- $\Delta_1 = f_1 - f_0$ & $\Delta_2 = f_1 - f_2$

Example

- Obtain the mode of the following data

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	5	7	9	6	4	1

- The class having highest frequency =“20 – 30”. So modal class = “20 – 30”
- Now L= 20, h = 10, $\Delta_1 = f_1 - f_0 = 9 - 7 = 2$
- $\Delta_2 = f_1 - f_2 = 9 - 6 = 3$
- $M_0 = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} . h = 20 + \frac{2}{2+3} . 10 = 24$

Contd…

- ***Case iii:*** *Determination of mode by grouping method*: Mode can be found by the grouping method under the following condition.
- If the maximum frequency is repeated or approximately equal concentration is found in two or more neighboring values.
- If the maximum frequency occurs either in the very beginning or at the end of the distribution.
- If there are irregularities in the distribution i. e. the frequencies of the variable increase or decrease in a haphazard way. Generally it has following two steps:
- Step i: Grouping all frequency.
- Step ii: Analysis the grouped data.

Example

- Determine the mode of the following data.

X	20	21	22	23	24	25	26	27	28	29
F	6	9	4	2	10	8	7	5	1	3

Solution: Since frequency distribution is irregular, mode is obtained by the method of grouping.

<table>
<tr><th>X</th><th>Frequency (I)</th><th>Sum of first two (II)</th><th>Sum of first two leaving first (III)</th><th>Sum of first three from first (IV)</th><th>Sum of three leaving first (V)</th><th>Sum of three's leaving from first two (VI)</th></tr>
<tr><td>20</td><td>6</td><td rowspan="2">15</td><td>x</td><td rowspan="3">19</td><td>x</td><td>x</td></tr>
<tr><td>21</td><td>9</td><td rowspan="2">13</td><td rowspan="3">15</td><td>x</td></tr>
<tr><td>22</td><td>4</td><td rowspan="2">6</td><td rowspan="3">16</td></tr>
<tr><td>23</td><td>2</td><td rowspan="2">12</td><td rowspan="3">20</td></tr>
<tr><td>24</td><td>10</td><td rowspan="2">18</td><td rowspan="3">25</td></tr>
<tr><td>25</td><td>8</td><td rowspan="2">15</td><td rowspan="3">20</td></tr>
<tr><td>26</td><td>7</td><td rowspan="2">12</td><td rowspan="3">13</td></tr>
<tr><td>27</td><td>5</td><td rowspan="2">6</td><td rowspan="3">9</td></tr>
<tr><td>28</td><td>1</td><td rowspan="2">4</td><td>x</td></tr>
<tr><td>29</td><td>3</td><td>x</td><td>x</td><td>x</td></tr>
</table>

Contd…

Analysis Table

Column No.	20	21	22	23	24	25	26	27	28	29
I					1					
II					1	1				
III						1	1			
IV				1	1	1				
V					1	1	1			
VI						1	1	1		
Total	0	0	0	1	4	5	3	1	0	0

Since highest repeated number is 25, then mode = 25.

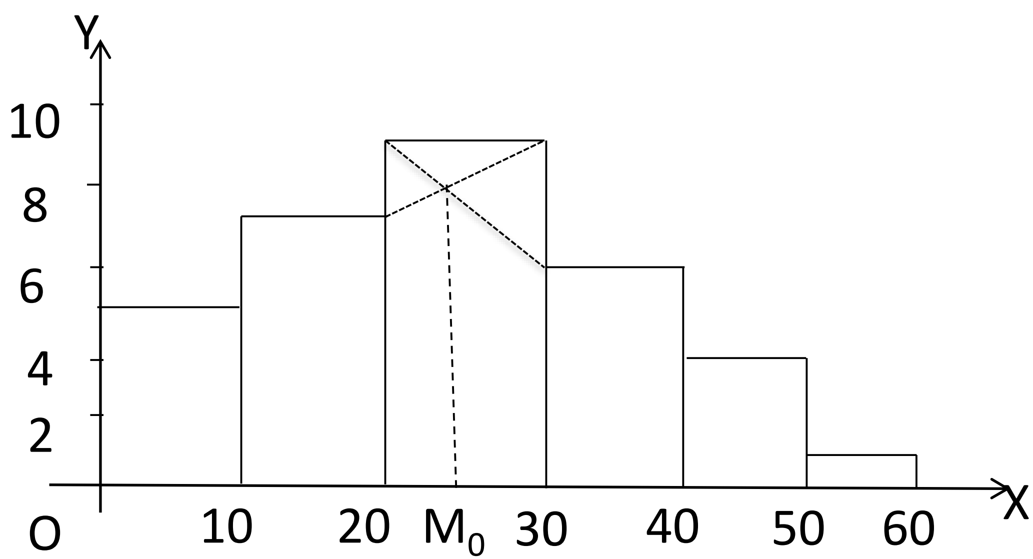
Contd…

- ***Case iv:*** *Empirical relationship between mean, median and mode*: In moderately asymmetrical(skewed) series mode is obtained by using empirical relationship defined by Prof. Karl Pearson.
- Mode = 3 Median – 2 Mean i. e.
- $\mathbf{M_0 = 3M_{\mathit{d}} - 2\,\bar{x}}$.
- Example: If the mean and median of a moderately asymmetrical series are 124 and 120. Find the mode.
- Solution: $M_0 = 3M_d - 2\,\bar{x} = 3 \times 120 - 2 \times 124$
- $= 360 - 248 = 112$.

Contd…

- ***Case v:*** *Calculation of mode by graphical method*: Mode can be calculated by using histogram when a group data is given.
- Example: Obtain the mode of the following data by using graphical method.

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	5	7	9	6	4	1

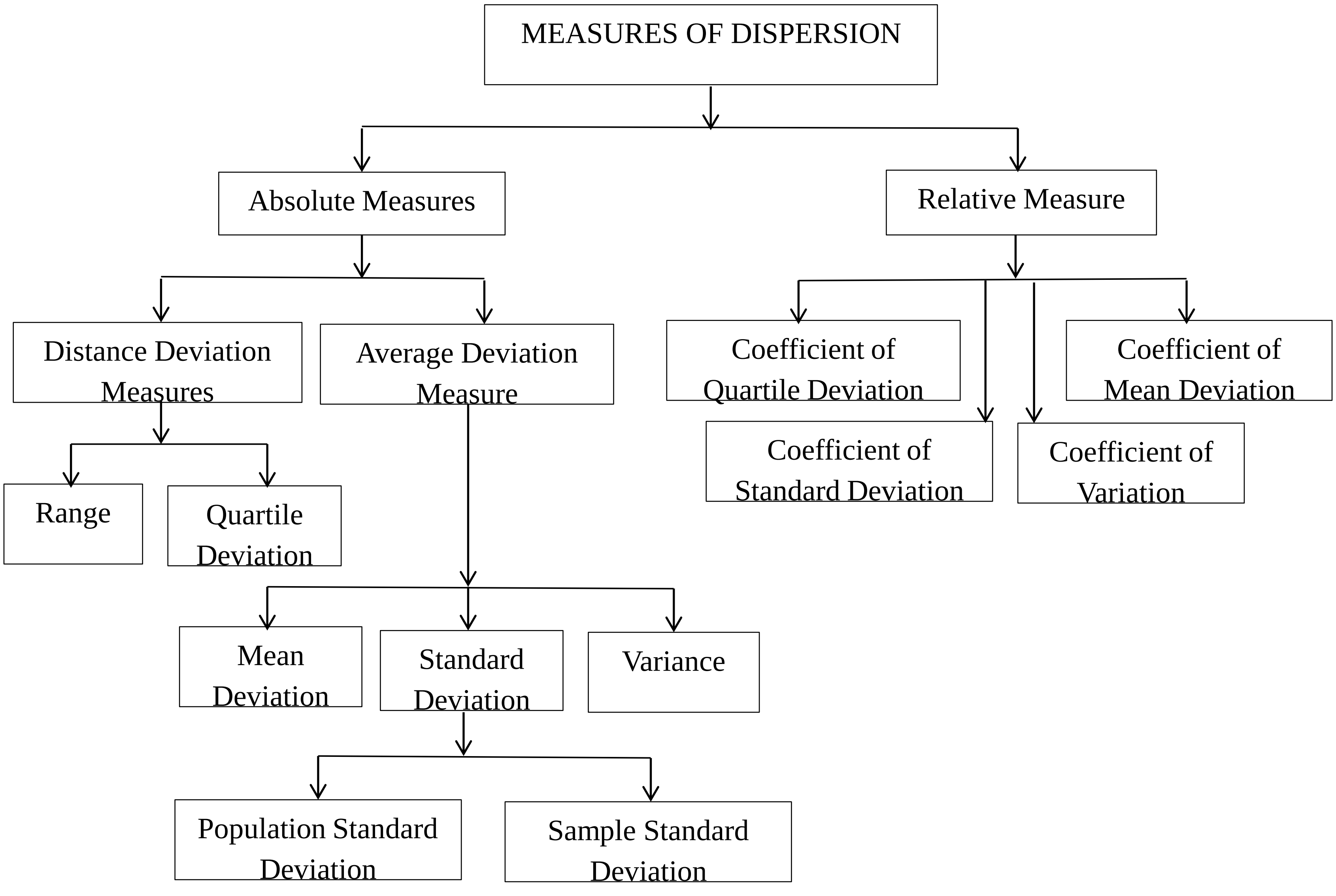


Mode = 24

Mid-hinge

Choice of an Average

- Selection of an average depends upon the nature of the data.
- ***A. M. is suggested when***:
- The average of quantitative data is calculated.
- When the distribution is not very skewed.
- The distribution does not have open and end classes.
- The distribution does not have very large and very small items.
- ***Median is recommended when***:
- The average of qualitative data is calculated.
- The distributions have open ended class.
- The distributions have very highly skewed.
- ***Mode is recommended for***
- Most repeated values.
- Most favourable items.
- Most fashionable items.
- Most common items.
- Consumer’s preferences.



Source: Rastogi, V. R. (2015): *Bio-statistics* (3rd ed). New Delhi: MEDTEC

Introduction

Series X	40	40	40	40	40	$\sum X = 200$	$\overline{X} = 40$
Series Y	38	39	40	41	42	$\sum X = 200$	$\overline{X} = 40$
Series Z	10	35	45	70	40	$\sum X = 200$	$\overline{X} = 40$

- Scatteredness of items from the central value of data is called measure of Variability
- *Methods of measuring Dispersion*:
- Range
- Quartile Deviation
- Mean Deviation
- Standard Deviation

Contd…

- *Absolute Measure of Dispersion*: expressed in terms of original units of the data.
- *Relative Measure of Dispersion*: it is independent of units of the data. It is free of any unit.

Note: In order to compare the variability of two or more than two sets of data expressed in different units, relative measure of dispersion would be used.

- Range(R) = Highest or largest item – Lowest or smallest item
- = $X_{(Largest)} - X_{(Smallest)}$.
- Coefficient of range is the ratio between difference of the extreme values and the sum of the extreme values.
- Coefficient of Range = $\frac{X_{(largest)} - X_{(smallest)}}{X_{(largest)} + X_{(smallest)}} = \frac{X_L - X_S}{X_L + X_S}$
- ***Quartile Deviation(Q. D.)***: is the half of the difference between the upper quartile (third quartile) and the lower quartile (first quartile) is also known as semi-inter quartile range. Thus, $Q.D. = \frac{Q_3 - Q_1}{2}$
- coefficient of $\mathrm{Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Contd…

- *Mean Deviation*: is the arithmetic mean of the absolute deviations of the various items from average value taken from mean or median or mode
- M. D. (from mean) = $\frac{\sum|X-\bar{X}|}{N}$.
- *Individual Data*: M. D. (from mean) = $\frac{\sum|X-\bar{X}|}{N}$
- *Discrete Data*: M. D. (from mean) = $\frac{\sum f|X-\bar{X}|}{N}$
- *Continuous Data*: M. D. (from mean) = $\frac{\sum f|m-\bar{X}|}{N}$, m = mid value.
- ***Note***: *In the above formula if we replace $\bar{X}$ by M_d or M_0 then we get the formula of mean deviation from median or mean deviation from mode*.
- Coefficient of mean deviation from mean = $\frac{\text{M.D.from mean}}{\text{mean}(\bar{X})}$
- Coefficient of mean deviation from median = $\frac{\text{M.D.from median}}{\text{median }(M_d)}$
- Coefficient of mean deviation from mode = $\frac{\text{M.D.from mode}}{\text{mode}(M_0)}$

Contd…

- *Standard Deviation*: is the positive square root of the arithmetic mean of the squares of the deviations of the given observations from their arithmetic mean.
- $S.D.(\sigma) = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}\ or\ \sqrt{\frac{\sum f(X-\bar{X})^2}{n}}\ or\ \sqrt{\frac{\sum f(m-\bar{X})^2}{n}}$
- $\mathrm{S.D.}(\sigma) =$
  $\sqrt{\frac{\sum \mathrm{X}^2}{\mathrm{n}} - \left(\frac{\sum \mathrm{X}}{\mathrm{n}}\right)^2}\ \mathrm{or}\ \sqrt{\frac{\sum \mathrm{fX}^2}{\mathrm{n}} - \left(\frac{\sum \mathrm{fX}}{\mathrm{n}}\right)^2}\ \mathrm{or}\ \sqrt{\frac{\sum \mathrm{fm}^2}{\mathrm{n}} - \left(\frac{\sum \mathrm{fm}}{\mathrm{n}}\right)^2}$
- $\mathrm{S.D.}(\sigma) =$
  $\sqrt{\frac{\sum \mathrm{d}^2}{\mathrm{n}} - \left(\frac{\sum \mathrm{d}}{\mathrm{n}}\right)^2}\ \mathrm{or}\ \sqrt{\frac{\sum \mathrm{fd}^2}{\mathrm{n}} - \left(\frac{\sum \mathrm{fd}}{\mathrm{n}}\right)^2}\ \mathrm{or}\ \sqrt{\frac{\sum \mathrm{fd}'^2}{\mathrm{n}} - \left(\frac{\sum \mathrm{fd}'}{\mathrm{n}}\right)^2}$

Contd…

- ***Variance***: Square of the standard deviation.
- Thus variance = σ^2.
- ***Coefficient of Standard Deviation***: $= \frac{\sigma}{\overline{X}}$
- ***Coefficient of Variation***: C.V.= $\frac{\sigma}{\overline{X}} x100$
- Population SD(σ) = $\sqrt{\frac{\sum(X-\overline{X})^2}{N}}$
- Note: The mean is more representative in that distribution which has least standard deviation value. The standard deviation is extremely useful in judging the representativeness of the mean. The greater the standard deviation, the greater the dispersion or variability and the greater will be the magnitude of the variations of the values from the mean.

Use of C. V.	
More CV implies	**Low CV implies**
high variability high dispersed more heterogeneous low consistency low uniformity	low variability low dispersed more homogeneous high consistency high uniformity

Combined Standard Deviation

- Let $\overline{\mathrm{X}}_1$ and $\overline{\mathrm{X}}_2$ are the means of first and second series with n_1 and n_2 number of observations and σ_1 and σ_2 their respective standard deviations. Then the combined standard deviation is denoted by
- σ_{12} is given by: $\sigma_{12} = \sqrt{\dfrac{n_1\left(\sigma_1{}^2 + d_1{}^2\right) + n_2\left(\sigma_2{}^2 + d_2{}^2\right)}{(n_1 + n_2)}}$
- Let $\overline{\mathrm{X}}_1$, $\overline{\mathrm{X}}_2$ and $\overline{\mathrm{X}}_3$ are the means of first, second and third series with n_1, n_2 and n_3 number of observations and σ_1, σ_2 and σ_3 their respective standard deviations. Then the combined standard deviation is denoted by σ_{123} is given by:
- $\sigma_{123} = \sqrt{\dfrac{\mathrm{n}_1\left(\sigma_1{}^2 + \mathrm{d}_1{}^2\right) + \mathrm{n}_2\left(\sigma_2{}^2 + \mathrm{d}_2{}^2\right) + \mathrm{n}_3\left(\sigma_3{}^2 + \mathrm{d}_3{}^2\right)}{(\mathrm{n}_1 + \mathrm{n}_2 + \mathrm{n}_3)}}$
- Where $\mathrm{d}_1 = \overline{\mathrm{X}}_1 - \overline{\overline{\mathrm{X}}}, \mathrm{d}_2 = \overline{\mathrm{X}}_2 - \overline{\overline{\mathrm{X}}}, \mathrm{d}_3 = \overline{\mathrm{X}}_3 - \overline{\overline{\mathrm{X}}}, \overline{\overline{\mathrm{X}}} = \dfrac{\mathrm{n}_1.\overline{\mathrm{X}}_1 + \mathrm{n}_2.\overline{\mathrm{X}}_2}{(\mathrm{n}_1 + \mathrm{n}_2)}$
- $\mathrm{or}\ \overline{\overline{\mathrm{X}}} = \dfrac{\mathrm{n}_1.\overline{\mathrm{X}}_1 + \mathrm{n}_2.\overline{\mathrm{X}}_2 + \mathrm{n}_3.\overline{\mathrm{X}}_3}{(\mathrm{n}_1 + \mathrm{n}_2 + \mathrm{n}_3)}$

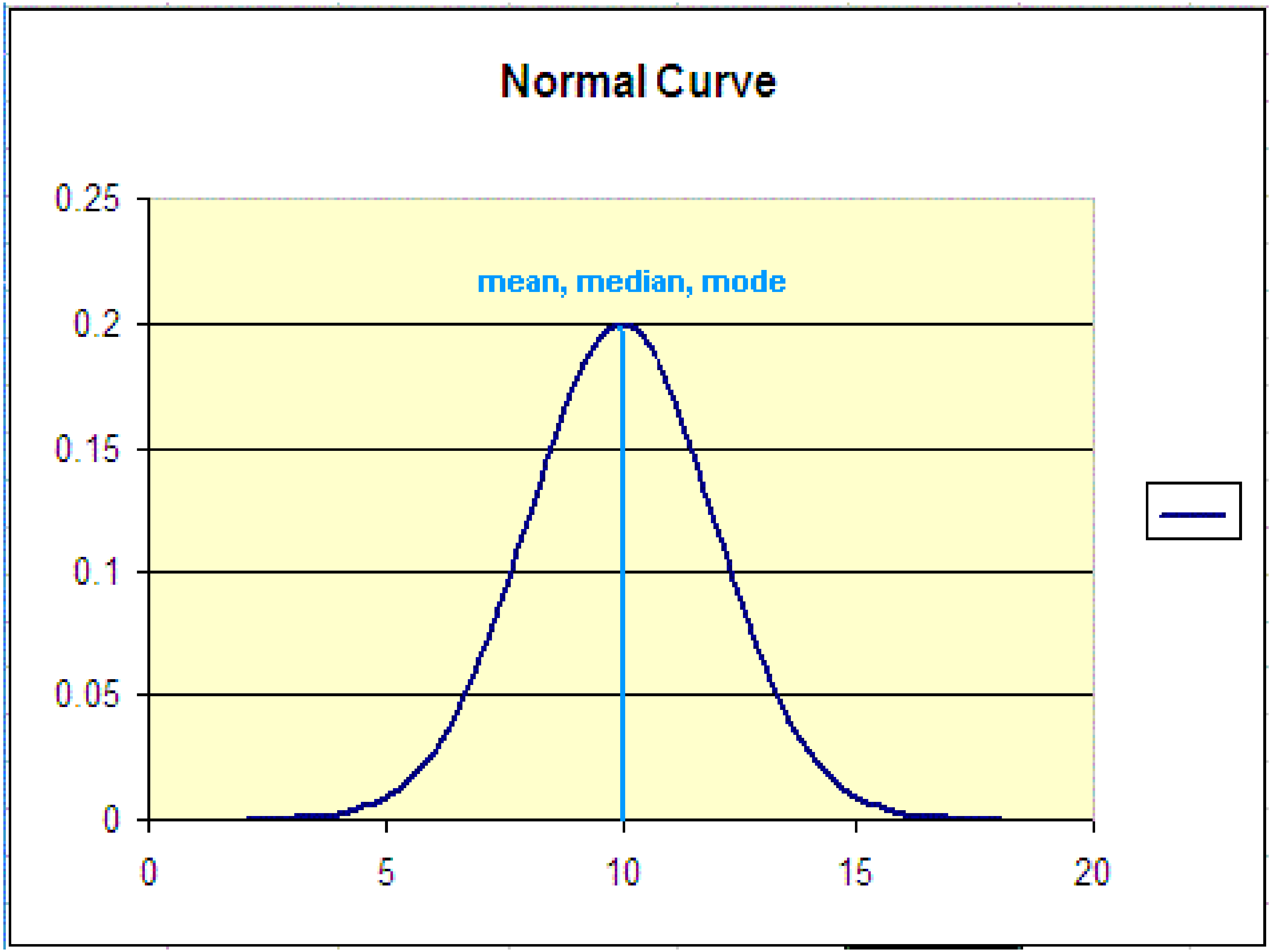
Relationship between mean, median and mode

- The shape of the data describes the manner of the distribution of the data. Data are either symmetrical or asymmetrical.
- A distribution is said to be symmetrical if, mean = median = mode.
- In any distribution if mean, median and mode are not equal then it is said to be asymmetrical (or skewed).
- A distribution is said to be positive skewed if, Mean $>$ median $>$ mode.
- A distribution is said to be negative skewed if, Mean $<$ median $<$ mode.

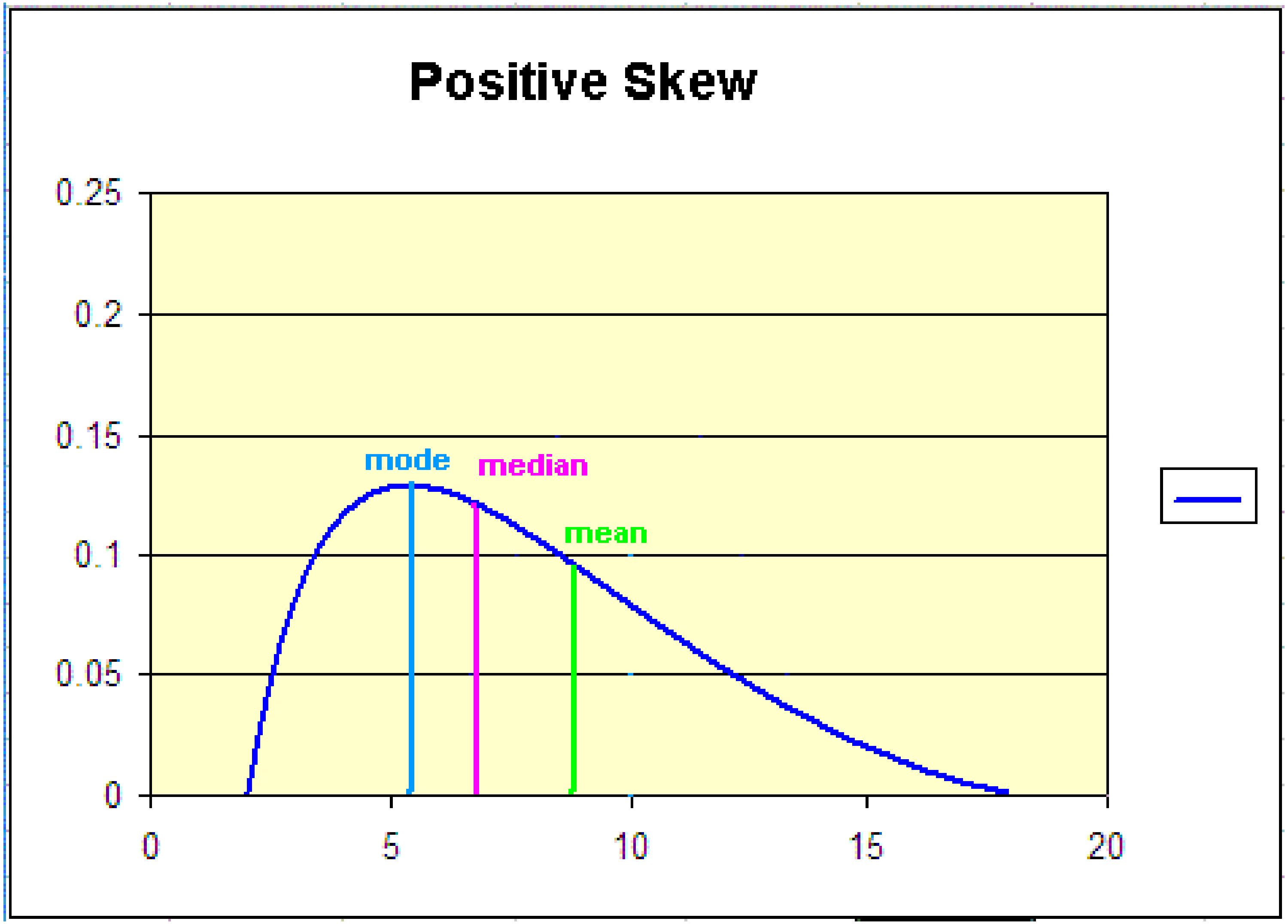
Skewness

- *No Skewness (or Symmetrical)*: A distribution is said to be symmetrical if: ***mean = median = mode*** i. e. mean, median and mode coincides. The graph of symmetrical distribution is shown in the figure (i) below. A distribution is said to be skewed if it is not symmetrical. In this case: ***mean*** $\neq$ ***median*** $\neq$ ***mode***.
- *Positively Skewed*: A distribution said to be positive skewed if it satisfies the condition: ***mean*** $>$ ***median*** $>$***mode***. The graph of positive Skewness is shown in the figure (ii) below.
- *Negatively Skewed*: A distribution said to be negative skewed if it satisfies the condition: ***mean*** $<$ ***median*** $<$ ***mode***. The graph of negative Skewness is shown in the figure (iii) below.

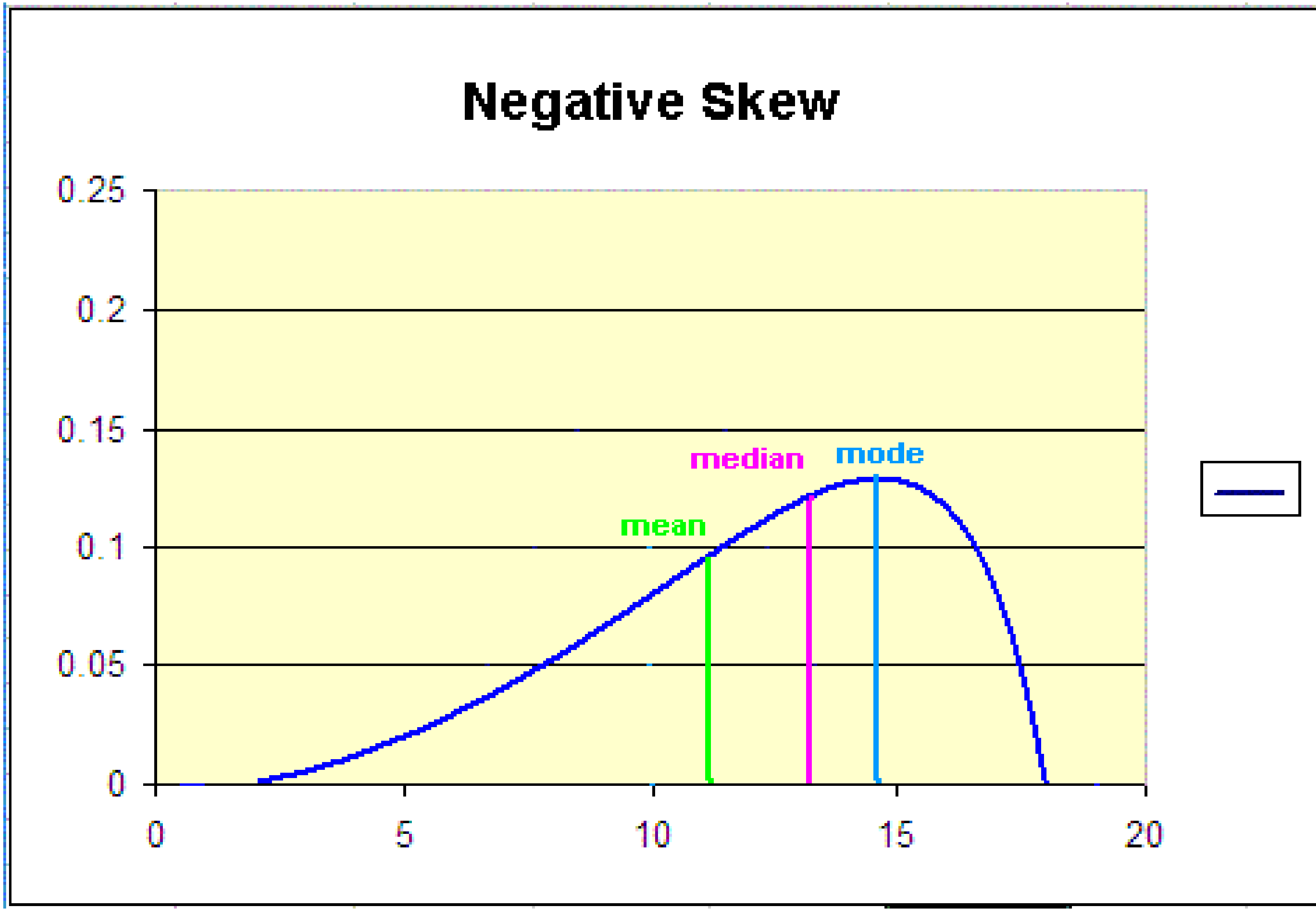
Normal Curve



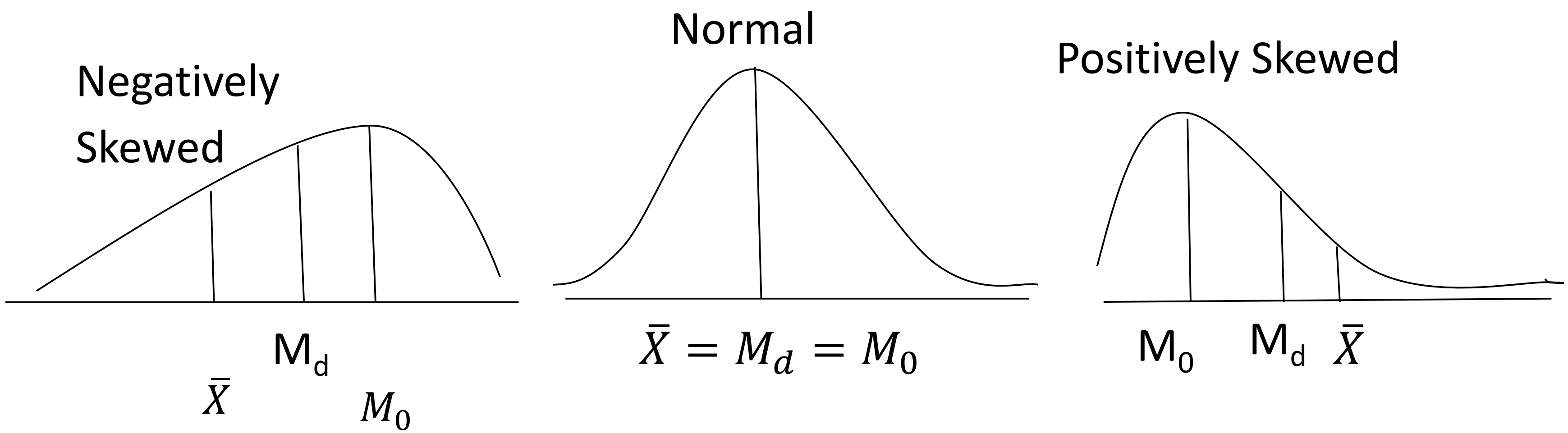
Positive Skew Curve



Negative Skew Curve



Shape



Example: Describe the shape of the following data: 55, 65, 80, 60, 50, 60, 65, 45 and 60.

Solution: Now arranging the given data in ascending order: 45, 50, 55, 60, 60, 60, 65, 65, 80.

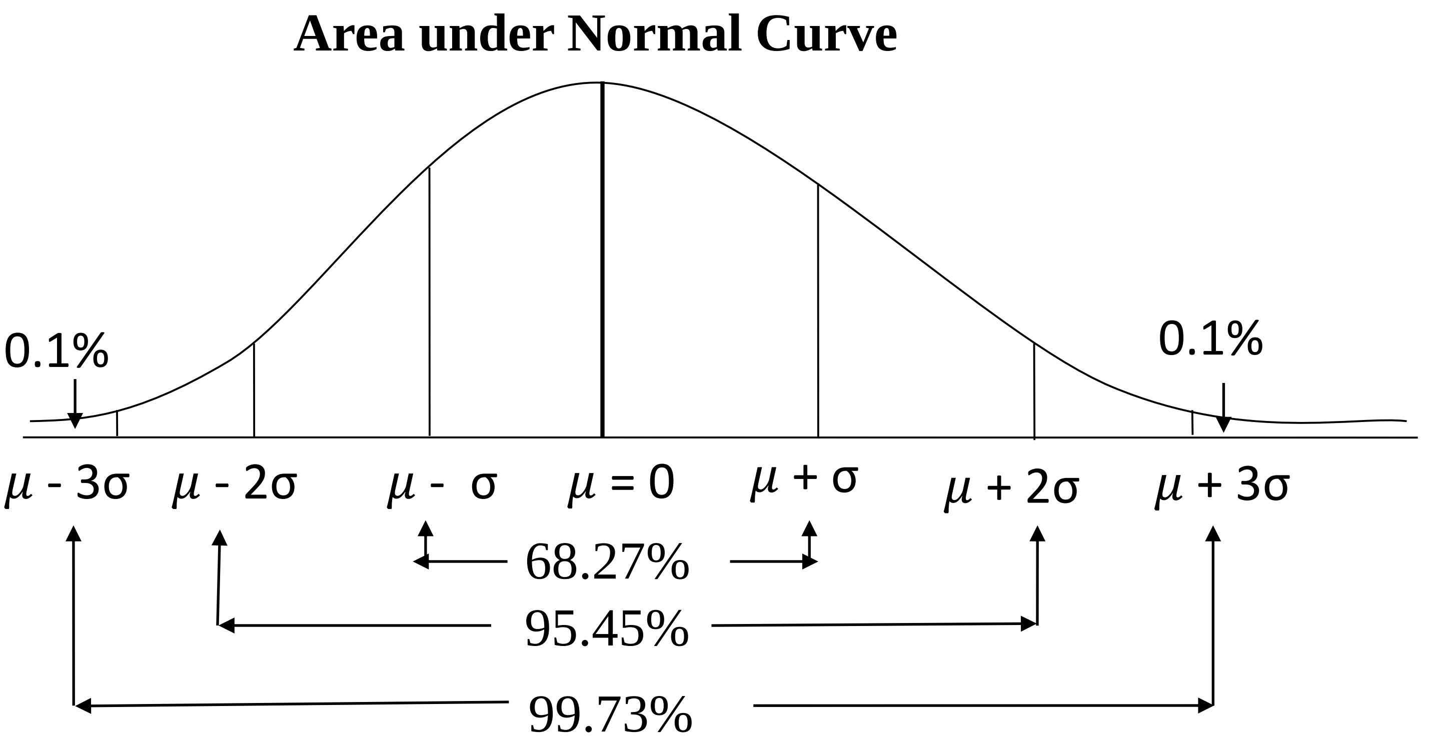
Now $\bar{X} = \frac{45 + 50 + 55 + \ 60 + 60 + 60 + 65 + 65 + 80}{9} = \frac{540}{9} = 60$

Median = the value of $\left(\frac{n+1}{2}\right)^{th}$ $item = \frac{9+1}{2} = 5^{th} item = 60$

And Mode = 60 [Since 60 repeats maximum times

Now, mean = median = mode = 60 then above distribution is normal distribution.

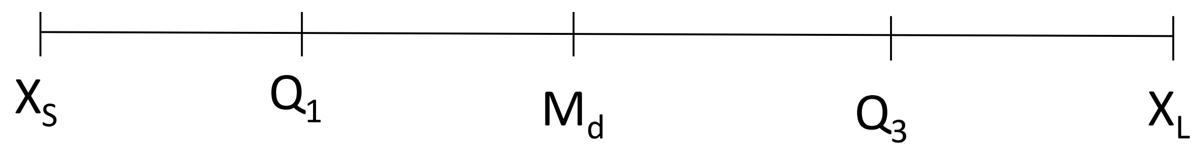
Relationship



- Range = 6 σ $\Rightarrow \sigma$ = (1/6) Range
- Q.D. = (2/3) σ $\Rightarrow \sigma$ = (3/2) Q.D.
- M.D. = (4/5) σ $\Rightarrow \sigma$ = (5/4) M.D.
- Q.D. = (5/6) M.D. $\Rightarrow$ 6 Q.D. = 5 M.D
- $\Rightarrow$ 4 S.D. = 5 M.D. = 6 Q.D.

Five Number Summaries

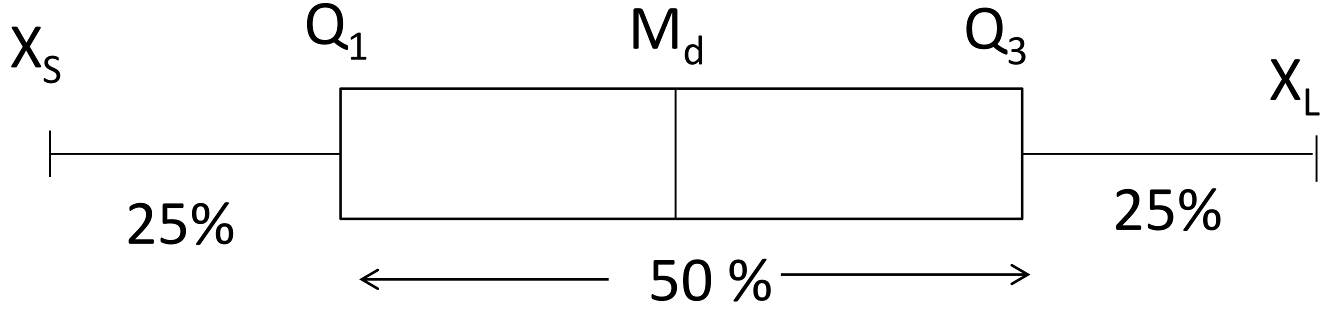
- Five number summaries provide five descriptive measures of the given data. It consists of X_S, Q_1, M_d, Q_3 and X_L. These values are measured in a scale line as:



Left skewed	Normal (no skewed)	Right skewed
$M_d - X_S > X_L - M_d$ $M_d - Q_1 > Q_3 - M_d$ $Q_1 - X_S > X_L - Q_3$ $M_d > \overline{X}$	$M_d - X_S = X_L - M_d$ $M_d - Q_1 = Q_3 - M_d$ $Q_1 - X_S = X_L - Q_3$ $M_d = \overline{X}$	$M_d - X_S < X_L - M_d$ $M_d - Q_1 < Q_3 - M_d$ $Q_1 - X_S < X_L - Q_3$ $M_d < \overline{X}$

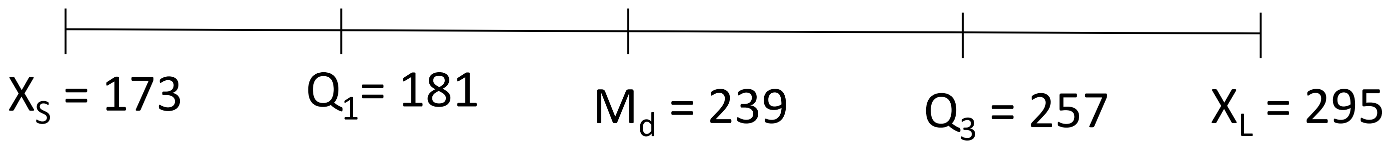
Box and Whisker plots

- A box and whisker plot is a graphical presentation of the data that displays a five number summary.

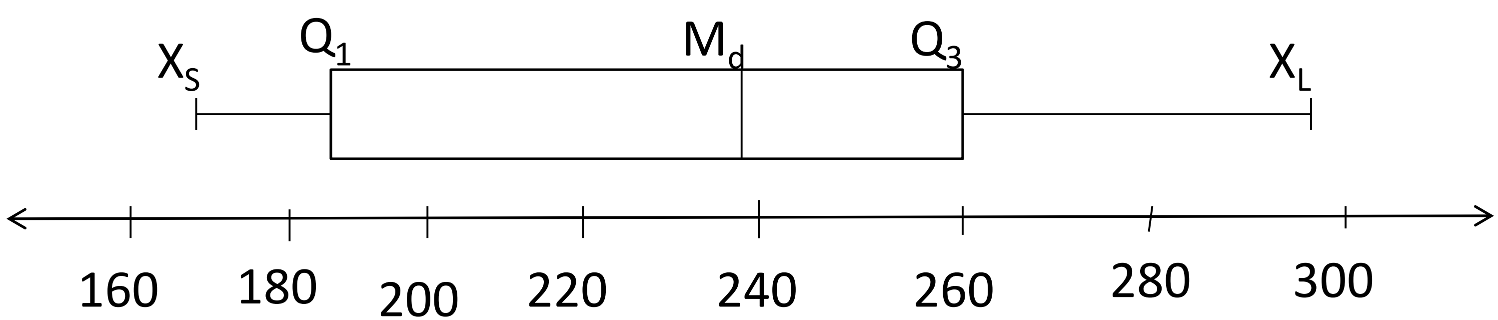


- Example: Construct box and whisker plot of the data set given below: 173, 206, 179, 257, 198, 251, 239, 246, 295, 181 and 261.
- Solution: Arranging the given data in ascending order of magnitude: 173, 179, 181, 198, 206, 239, 246, 251, 257, 261 & 295.
- Here, n = 11, X_S = 173, X_L = 295.
- Size of Q_1 = the value of $\left(\frac{n+1}{4}\right)^{th}$ $item$ = $\frac{11+1}{4} = \frac{12}{4}$
- $= 3^{rd}$item = 181
- Size of M_d = the value of $\left(\frac{n+1}{2}\right)^{th}$ $item$ = $\frac{11+1}{2} = \frac{12}{2}$
- $= 6^{th}$item = 239
- Size of Q_3 = the value of $\left(\frac{3(n+1)}{4}\right)^{th}$ $item$ = $\frac{3(11+1)}{4} = \frac{36}{4}$
- $= 9^{th}$item = 257

Now the five number summaries can be shown as:



- Box and whisker plot of the above summaries as:



- Example: The residences of Kathmandu valley city are busy in their day to day work. But the pollution of dust particles and heavy traffic and its smoke presence in an environment, health conditions of residences of this city getting deteriorate day by day. Therefore, Kathmandu Health club conducts massive campaign about the importance of physical fitness. Club has organized the training program to provide physical fitness to the people. After one month training a test is conducted. Following data represents the time, in minutes, taken in a fitness trail to complete a certain task for 45 different participants.

57	23	35	18	21	26	51	47	29
21	46	50	43	29	23	39	41	19
36	28	31	42	52	29	18	28	46
33	28	29	30	40	43	55	57	46
18	16	26	28	20	36	38	40	27

i. Compute mean, median, first quartile and third quartile.
ii. Compute range, interquartile range, standard deviation and C. V.
iii. Construct box and whisker plot and write about the shape of the distribution.

Thank You