Measures of Dispersion

Measures in Statistics

- Measures in Statistics
  - Measures in Central Tendency (Measures of Location)
    - Ungrouped Data
      - Mean
      - Median
      - Mode
      - Weighted Mean
    - Grouped Data
      - Mean
      - Median
      - Mode
  - Measures in Dispersion
    - Range
    - Inter Quartile Range
    - Variance
    - Standard Deviation
  - **Measures of Position**
    - **Percentiles**
    - **Quartiles**
  - **Measures of Shape**
    - **Skewness**
    - **Kurtosis**
  - **Measures of Association**
    - **Correlation Coefficient**
    - **Covariance**

Definition

- *Measures of* ***Dispersion*** are Descriptive Statistics that describe the extent to which a numerical data is likely to vary about an average value.

  - The more similar the scores are to each other, the lower the measure of dispersion will be
  - The less similar the scores are to each other, the higher the measure of dispersion will be
  - **In general, the more spread out a distribution is, the larger the measure of dispersion will be**

Measures of Dispersion

- There are three main measures of dispersion:

  1. The Range
  2. The Interquartile range
  3. Standard deviation

- Other popular measures of dispersion:
  - The semi-interquartile range (SIR)/ (Quartile Deviation)
  - Variance

The Range

- The ***Range*** is defined as the difference between the highest and lowest values of a dataset, $X_L - X_S$

Range = Highest Value – lowest Value

R = H – L

$R = X_L - X_S$

E.g.:

What is the range of the following data?:

4 8 1 6 6 2 9 3 6 9

Answer:

The largest score (H) is 9

The smallest score (L) is 1

Range is H-L = 9 - 1 = 8

When To Use the Range

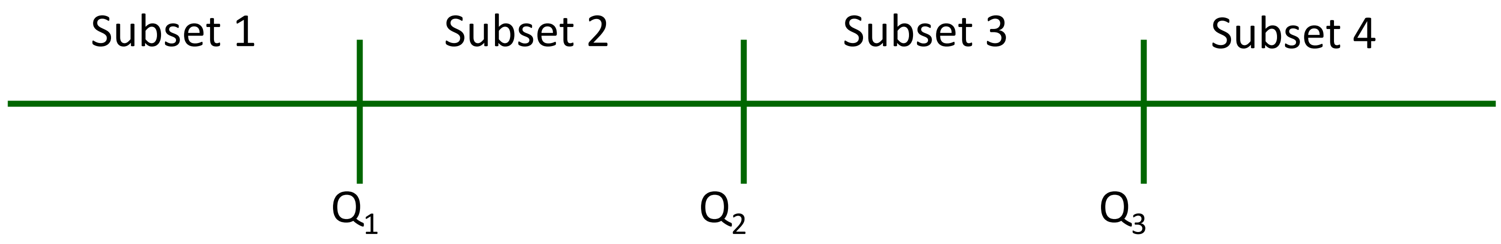
- The range is used when;
  - you have ordinal data or
  - you are presenting your results to people with little or no knowledge of statistics

- The range is rarely used in scientific work as it is fairly insensitive.
  - It depends on only two scores in the set of data, X_L and X_S
  - Two very different sets of data can have the same range:
    1 1 1 1 9 vs 1 3 5 7 9

Quartiles

Calculating Quartiles

- **A type of quantile which divides the number of data points into four parts, or quarters**
- Every data set has three quartiles, which divide it into four (04) equal parts
- The data **must be ordered** from smallest to largest to compute quartiles.
- If the horizontal line can be thought of as a data set arranged in an ordered array, three quartiles can be identified, which together produce four separate parts or subset of equal size in the data set.



First Quartile (Q_1)– Ungrouped data

- The first quartile is the value below which, at most, 25% of the observations fall, and above which the remaining 75% can be found

E.g. : Find the first quartile value.

38, 16, 24, 40, 58, 90, 30, 14, 41, 39, 61

Answer:

1. order the data set in ascending order-

   14,16, 24, 30, 38, 39, 40, 41, 58, 61, 90

2. Apply the formula- **("n" is the total number of observations in the array)**

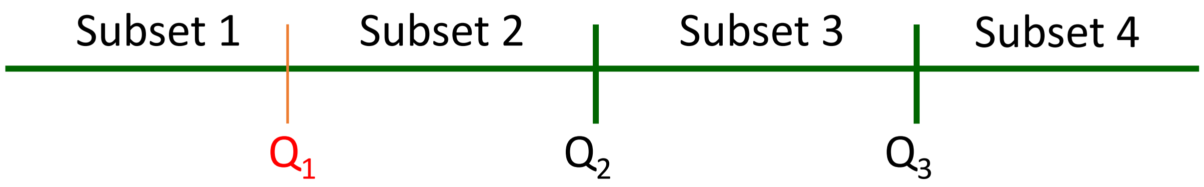
   $\mathbf{Q_1 = \frac{1}{4}(n+1)^{th}}$ **observation**

   $Q_1 = \frac{1}{4}(11+1)^{th}$ observation

   11=No. of items in the array

   $Q_1 = 3^{rd}$ observation

   Hence, $Q_1 = 24$



Second Quartile (Q_2)– Ungrouped data

- The second quartile is the value below which, at most, 50% of the observations fall, and above which the remaining 50% can be found.
- The second quartile is right in the middle. Same as the **Median**

E.g. : Find the second quartile value.
38, 16, 24, 40, 58, 90, 30, 14, 41, 39, 61

Answer:

- order the data set in ascending order-
  14,16, 24, 30, 38, 39, 40, 41, 58, 61, 90
- Apply the formula-

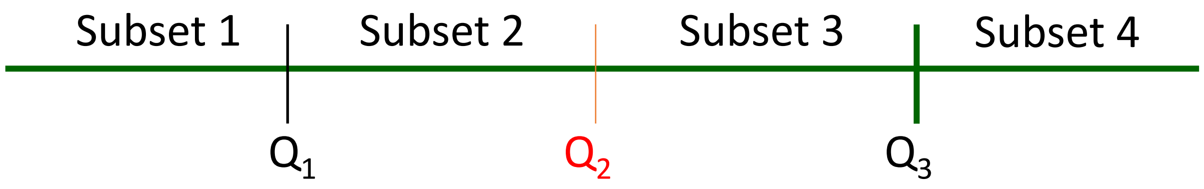
$Q_2 = \frac{2}{4}(n+1)^{th}$ observation

11=No. of items in the array

$Q_2 = \frac{2}{4}(11+1)^{th}$ observation

$Q_2 = 6^{th}$ observation

Hence, $Q_2 = 39$



Third Quartile (Q_3) – Ungrouped data

- The third quartile is the value below which, at most, 75% of the observations fall, and above which the remaining 25% can be found.

E.g. : Find the third quartile value.

38, 16, 24, 40, 58, 90, 30, 14, 41, 39, 61

Answer:

- order the data set in ascending order-

  14,16, 24, 30, 38, 39, 40, 41, 58, 61, 90

- Apply the formula-

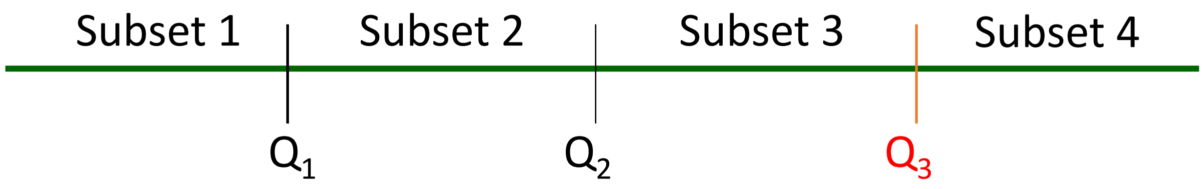
$$Q_3 = \frac{3}{4}(n+1)^{th} \text{ observation}$$

11=No. of items in the array

$$Q_3 = \frac{3}{4}(11+1)^{th} \text{ observation}$$

9^{th} observation

Hence, $Q_3 = 58$

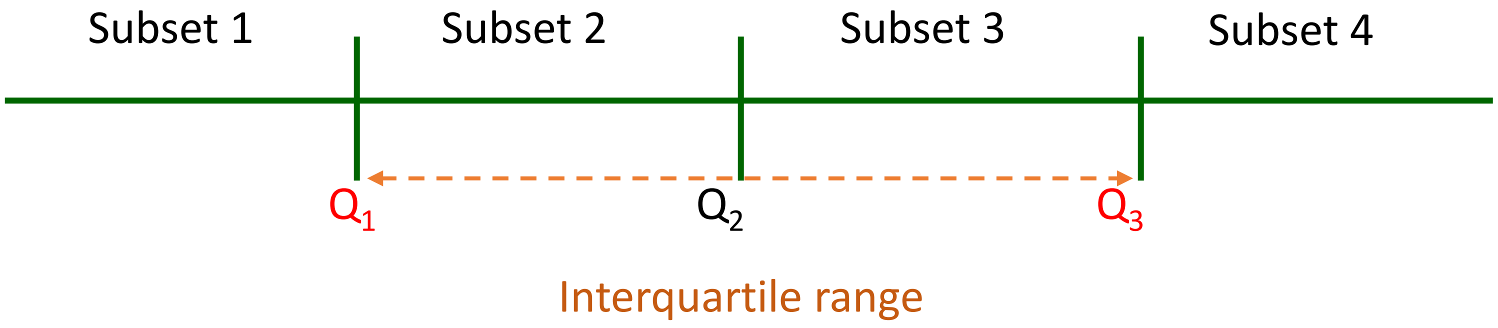


Inter-Quartile Range - Ungrouped data

- The Interquartile range is the distance between the third quartile Q_3 and the first quartile Q_1.

Inter-quartile range = Third quartile - First quartile

= $Q_3 - Q_1$



Quartile Deviation/ Semi-Interquartile Range

- Half of the Inter Quartile Range ($Q_3 – Q_2$) is know as "***semi-interquartile range***" or **"Quartile Deviation"**.
- The ***semi-interquartile range*** (or ***SIR***) is defined as the difference of the first and third quartiles divided by two

$$\textbf{Quartile deviation (SIR)} = \frac{\textbf{Third quartile – First quartile}}{2}$$

$$= \frac{Q3 - Q1}{2}$$

Example 02

- Find the **inter-Quartile range** and **Quartile Deviation** of the data set below.

55, 55, 56, 56, 56, 56, 56, 56, 56, 56, 57, 57, 57, 57, 57, 57, 57, 57, 57, 57, 57, 57, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 59, 59, 59, 59,

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
55,	55,	56,	56,	56,	56,	56,	56,	56,	56,	57,	57,	57,	57,	57,	57,	57,	57,	57,	57,	57,	57,	58,	58,	58,	58,	58,	58,	58,	58,	58,	58,	58,	58,	58,	59,	59,	59,	59,

Answer_Example 02

First Quartile (Q_1) $= \frac{1}{4}(39+1)^{th}$ observation

$= \frac{1}{4}(40)^{th}$ observation

$= 10^{th}$ observation

Hence, Q_1 $= 56$

Second Quartile (Q_2) $= \frac{1}{2}(39+1)^{th}$ observation

$= \frac{1}{2}(40)^{th}$ observation

$= 20^{th}$ observation

Hence, Q_{12} $= 57$

Third Quartile (Q_3) $= \frac{3}{4}(n+1)^{th}$ observation

$= \frac{3}{4}(40)^{th}$ observation

$= 30^{th}$ observation

Hence, Q_3 $= 58$

Inter-Quartile Range $= Q_3 - Q_1$

$=58 - 56 = 2$

Quartile Deviation $= \frac{(Q_3 - Q_1)}{2}$

$=(58 - 56)/2 = 1$

Activity

- Find the **inter-Quartile range** and **Quartile Deviation** of the data set below using frequency distribution table.

55, 55, 56, 56, 56, 56, 56, 56, 56, 56, 57, 57, 57, 57, 57, 57, 57, 57, 57, 57, 57, 57, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 59, 59, 59, 59,

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
55,	55,	56,	56,	56,	56,	56,	56,	56,	56,	57,	57,	57,	57,	57,	57,	57,	57,	57,	57,	57,	57,	58,	58,	58,	58,	58,	58,	58,	58,	58,	58,	58,	58,	58,	59,	59,	59,	59,

Answer_Example 02

Values (X)	Frequency (f)	Cf
55	2	2
56	8	10
57	12	22
58	13	35
59	4	39

First Quartile (Q_1) $= \frac{1}{4}(39+1)^{th}$ observation

$= \frac{1}{4}(40)^{th}$ observation

$= 10^{th}$ observation

Hence, $Q_1 = 56$

Third Quartile (Q_3) $= \frac{3}{4}(n+1)^{th}$ observation

$= \frac{3}{4}(40)^{th}$ observation

$= 30^{th}$ observation

Hence, $Q_3 = 58$

Inter-Quartile Range $= Q_3 – Q_1$

$=58 – 56 = 2$

Quartile Deviation $= \frac{(Q_3 - Q_1)}{2}$

$=(58 – 56)/2 = 1$

Activity 02

- Marks earned for Statistics by 15 students are listed below.

Marks	95	12	50	63	82	45	92
No. of students	2	1	2	4	3	2	1

Find;

a) First Quartile
b) Second Quartile
c) Third Quartile
d) Inter- Quartile range
e) Quartile deviation

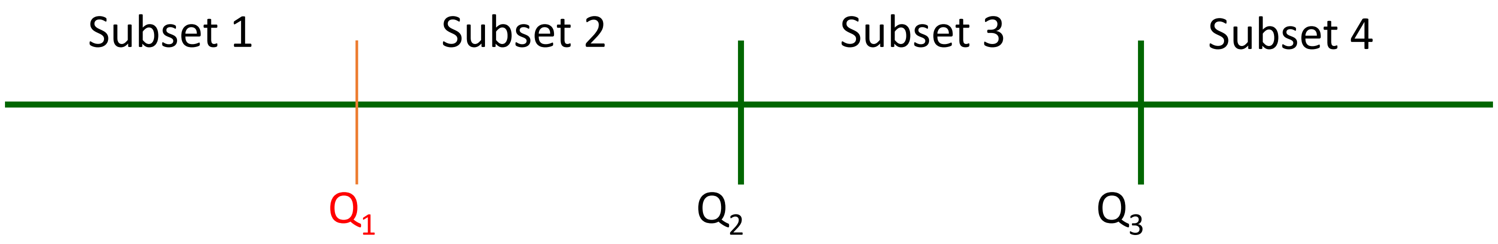
First Quartile (Q_1)– Grouped data

- The first quartile is the value below which, at most, 25% of the observations fall, and above which the remaining 75% can be found

$$\text{First Quartile } (Q_1) = L + \frac{\left(\frac{n}{4} - CF\right)}{f}(i)$$

where

L = lower boundary of the class containing Q_1,
CF = cumulative frequency preceding class containing Q_1,
f = frequency of class containing Q_1,
i = size of class containing Q_1.



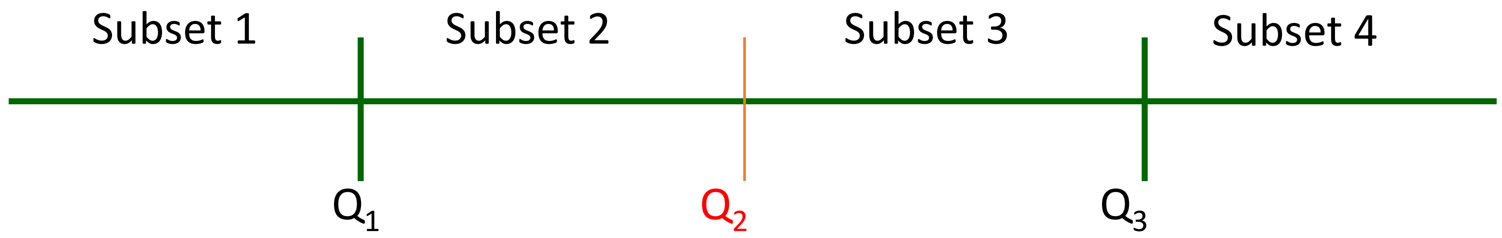
Second Quartile – Grouped data

- The second quartile is right in the middle. Same as the **median**

$$\textbf{Second Quartile } (Q_2) = L + \frac{\left(\frac{2n}{4} - CF\right)}{f}(i)$$

where

L = lower boundary of the class containing Q_1,
CF = cumulative frequency preceding class containing Q_1,
f = frequency of class containing Q_1,
i = size of class containing Q_1.



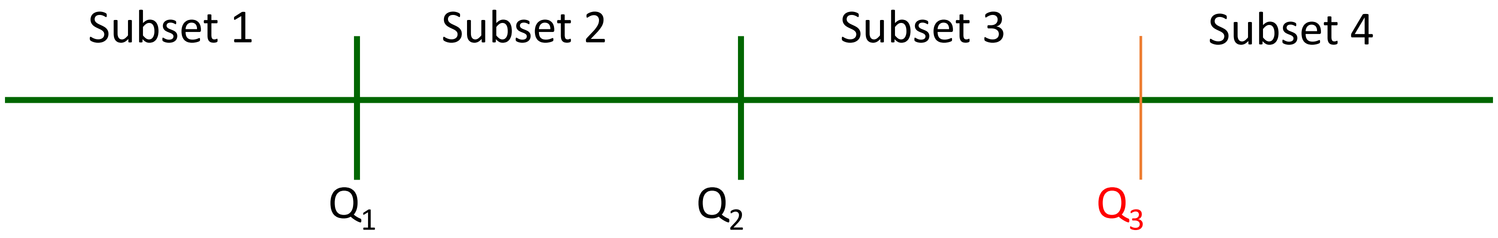
Third Quartile – Grouped data

- The third quartile is the value below which, at most, 75% of the observations fall, and above which the remaining 25% can be found

$$\textbf{Third Quartile } (Q_3) = L + \frac{\left(\frac{3n}{4} - CF\right)}{f}(i)$$

where

L = lower boundary of the class containing Q_3,
CF = cumulative frequency preceding class containing Q_3,
f = frequency of class containing Q_3,
i = size of class containing Q_3.



Activity 03

- Find the Inter quartile range and the quartile deviation.

Class limits	Frequency
0-10	2
10-20	3
20-30	5
30-40	2
40-50	6
50-60	2

Answer-Activity 02

- Find the Inter quartile range

Class limits	Frequency	CF
0-10	2	2
10-20	3	5
20-30	5	10
30-40	2	12
40-50	6	18
50-60	2	20

Q_1 = n/4 = 20/4 =5th observation

$$Q_1 = L + \left[\frac{\frac{n}{4} - CF}{f}\right] (i)$$

$$Q_1 = 10 + \left[\frac{\frac{20}{4} - 2}{3}\right] (10)$$

$$Q_1 = 10 + \left[1\right] (10)$$

$$Q_1 = 20$$

Answer-Activity 02

- Find the Inter quartile range

Class limits	Frequency	CF
0-10	2	2
10-20	3	5
20-30	5	10
30-40	2	12
40-50	6	18
50-60	2	20

Q_3 = (n/4)*3 = 20/4 =15th observation

$$Q_3 = L + \left[\frac{\frac{n*3}{4} - CF}{f}\right] \quad \text{(i)}$$

$$Q_3 = 40 + \left[\frac{15-12}{6}\right] \quad (10)$$

$$Q_3 = 40 + \left[\frac{1}{2}\right] \quad (10)$$

$$Q_3 = 45$$

Answer-Activity 02

- Find the Inter quartile range

Class limits	Frequency	CF
0-10	2	2
10-20	3	5
20-30	5	10
30-40	2	12
40-50	6	18
50-60	2	20

Inter-Quartile range $= Q_3 - Q_1$

$= 45 - 20$

$= 25$

Quartile Deviation $= \frac{(Q_3 - Q_1)}{2}$

$= \frac{(45 - 20)}{2}$

$= 12.5$

Activity 03

- Find the Inter-quartile range and quartile deviation

Class limits	Frequency
350-360	4
360-370	6
370-380	5
380-390	4
390-400	3

Activity 04

No. of complaints received to a customer care service center during 50 days are as below. Find the quartile deviation of this distribution.

No. of Complaints	No. of days
26-50	4
51-75	5
76-100	7
101-125	11
126-150	9
151-175	8
176-200	6

Answer Activity 04

No. of Complaints	No. of days	cf
26-50	4	4
51-75	5	9
76-100	7	16
101-125	11	27
126-150	9	36
151-175	8	44
176-200	6	50

Q_1 class = 76-100 , Hence, Q_1 = 88

Q_2 class = 101-125 , Hence, Q_2 = 120.95

Q_3 class = 151-175 , Hence, Q_3 = 155.19

Usage of Quartiles

- Quartiles often are used **in sales and survey data to divide populations into groups**

- The quartile deviation helps **to examine the spread of a distribution about a measure of its central tendency**, usually the mean or the average. Hence, it is in use to give you an idea about the range within which the central 50% of your sample data lies.

Variance and Standard Deviation

Variance– Ungroup Data

- **Variance;**
  - is a statistical **measurement of the spread between numbers in a data set and it is the average of the squared differences from the mean.**
  - measures how far each number in the set is from the mean and thus from every other number in the set.

- **Variance** of the POPULATION is denoted by "σ^2" (sigma squared)
- **Variance** of a SAMPLE is denoted by "s^2"

Sigma (uppercase **Σ**, lowercase **σ**)

Standard Deviation (SD) σ – Ungroup Data

- **Standard Deviation;**
  - is a statistical **measure of how dispersed the data is in relation to the mean and it is the square root of the variance.**
  - A low standard deviation indicates that the values tend to be close to the mean while a high standard deviation indicates that the values are spread out over a wider range.
  - abbreviated **SD**
  - represented in mathematical texts and equations by the **lower case Greek letter sigma σ**
  - can be computed by deriving the positive square root of Variance
  - **Standard Deviation** of the POPULATION is denoted by "σ"
  - **Standard Deviation** of a SAMPLE is denoted by "s"

Population Variance and Standard deviation – Ungrouped data

- The population variance is the mean of the squared deviations of the observations from their population mean.

$$\sigma^2 \text{ (Variance)} = \frac{\sum (Xi - \mu)^2}{n} \quad \text{OR} \quad \sigma^2 = \frac{\sum x^2}{n} - \mu^2$$

- The population standard deviation

$$\sigma \text{ (SD)} = \sqrt{\sigma^2} = \sqrt{\frac{\sum (X_i - \mu)^2}{n}}$$

Sample Variance and Standard deviation – Ungrouped data

- The sample variance is

$$s^2 \text{ (Variance)} = \frac{\sum(Xi - \bar{x})^2}{n - 1}$$

- The sample standard deviation is

$$s \text{ (SD)} = \sqrt{s^2} = \sqrt{\frac{\sum(Xi - \bar{x})^2}{n - 1}}$$

Variance and Standard Deviation – Ungroup Data

Method 01

- *Variance* is defined as the average of the square deviations:

$$\textbf{Variance } (\sigma^2) = \frac{\sum(x-\mu)^2}{n}$$

Where;

μ = the mean (population)
x = stands for each data value in turn
n = the number of data values

$$\textbf{Standard Deviation } (\sigma) = \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{\sigma^2}$$

Variance and Standard Deviation – Ungroup Data

Method 02

- *Variance* is defined as the average of the square deviations:

$$\text{Variance } (\sigma^2) = \frac{\sum x^2}{n} - \mu^2$$

Where;

μ = the mean (population)

x = stands for each data value in turn

n = the number of data values

$$\text{Standard Deviation } (\sigma) = \sqrt{= \frac{\sum x^2}{n} - \mu^2} = \sqrt{\sigma^2}$$

Example

6, 7, 10, 11, 11, 13, 16, 18, 25

1. **Find the sample variance and sample Standard Deviation of given data set**
2. **Find the population variance and population SD of given data set.**

Answer for Q1

6, 7, 10, 11, 11, 13, 16, 18, 25

- Frist find mean $\bar{x} = \frac{\sum x}{n} = \frac{117}{9} = 13$

x		$x - \bar{x}$	$(x - \bar{x})^2$
6	Mean =13	-7	49
7		-6	36
10		-3	9
11		-2	4
11		-2	4
13		0	0
16		3	9
18		5	25
25		12	144

Variance (s^2) = $\frac{\sum(x-\bar{x})^2}{n} = \frac{(49+36+9+4+4+0+9+25+144)}{9}$

$= \frac{280}{8}$

= 35

Standard Deviation (s) $= \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{s^2}$

$= \sqrt{35}$

= 5.9

Answer for Q2 - method 01 $(\sigma^2) = \frac{\sum(x-\mu)^2}{n}$

6, 7, 10, 11, 11, 13, 16, 18, 25

- Frist find mean $\mu = \frac{\sum x}{n} = \frac{117}{9} = 13$

x		$x - \bar{\mu}$	$(x-\mu)^2$
6	Mean =13	-7	49
7		-6	36
10		-3	9
11		-2	4
11		-2	4
13		0	0
16		3	9
18		5	25
25		12	144

Variance $(\sigma^2) = \frac{\sum(x-\mu)^2}{n} = \frac{(49+36+9+4+4+0+9+25+144)}{9}$

$= \frac{280}{9}$

= 31.11

Standard Deviation (σ) $= \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{\sigma^2}$

$= \sqrt{31.11}$

= 5.57

Answer for Q2- Method 02 [$\sigma^2 = \frac{\sum x^2}{n} - \mu^2$]

- Frist find mean $\mu = \frac{\sum x}{n} = \frac{117}{9} = 13$

x	x^2
6	36
7	49
10	100
11	121
11	121
13	169
16	256
18	324
25	625
117	**1801**

Total

Variance (σ^2) $= \frac{\sum x^2}{n} - \mu^2 = \frac{1801}{8} - 13^2$

=200.11 - 169

= 31.11

Standard Deviation (σ) $= \sqrt{\frac{\sum x^2}{n} - \mu^2} = \sqrt{\sigma^2}$

$=\sqrt{31.11}$

= 5.58

Activity 05

Marks obtained by students in class have been given. Find the standard deviation of marks

65, 70, 62, 90, 92, 50, 48, 32, 60, 71

σ^2 = 304.2

σ = 17.44

Method 01:

$$\sigma^2 = \frac{\sum(x-\mu)^2}{n}$$

Method 02:

$$\sigma^2 = \frac{\sum x^2}{n} - \mu^2$$

Variance and Standard Deviation – Grouped Data

Method 01

- *Variance* is defined as the average of the square deviations:

$$\textbf{Variance } (\sigma^2) = \frac{\sum f(x-\mu)^2}{n}$$

Where;

μ = the mean

x = stands for each data value in turn

n = the number of data values

$$\textbf{Standard Deviation } (\sigma) = \sqrt{\frac{\sum f(x-\mu)^2}{n}} = \sqrt{\sigma^2}$$

Variance and Standard Deviation – Grouped Data

Method 02

- *Variance* is defined as the average of the square deviations:

$$\text{Variance } (\sigma^2) = \frac{\sum fx^2}{n} - \mu^2$$

Where;

μ = the mean

x = stands for each data value in turn

n = the number of data values

$$\text{Standard Deviation } (\sigma) = \sqrt{= \frac{\sum fx^2}{n} - \mu^2} = \sqrt{\sigma^2}$$

Example

- Find an estimate of the variance and standard deviation of the following data for the marks obtained in a test by 88 students

Marks	Frequency (f)
$0 \leq x < 10$	6
$10 \leq x < 20$	16
$20 \leq x < 30$	24
$30 \leq x < 40$	25
$40 \leq x < 50$	17
Total	88

Answer - Method 01----Using $\left[\boldsymbol{\sigma^2} = \frac{\sum \boldsymbol{f(x-\mu)^2}}{\boldsymbol{n}} \right]$

- Frist find mean $\bar{x} = \frac{\sum fx}{n} = \frac{2510}{88}$ =28.52

Marks	Frequency (f)	Mid Point (x)	fx	Mean =29 (approximately)	$x-\mu$	$(x-\mu)^2$	$f(x-\mu)^2$
0 ≤ x < 10	6	5	30		-24	576	3456
10 ≤ x < 20	16	15	240		-14	196	3136
20 ≤ x < 30	24	25	600		-4	16	384
30 ≤ x < 40	25	35	875		6	36	900
40 ≤ x < 50	17	45	765		16	256	4352
Total	88		2510				12228

- **Variance** σ^2 $= \frac{\sum f(x-\mu)^2}{n}$

$= \frac{12228}{88}$

= 138.95

- **Standard Deviation** (σ) $= \sqrt{\frac{\sum f(x-\mu)^2}{n}} = \sqrt{\text{Variance}}$

$= \sqrt{138.95}$

= 11.78

Answer - Method 02----Using [$\sigma^2 = \frac{\sum fx^2}{n} - \mu^2$]

- Frist find mean $\mu = \frac{\sum fx}{n} = \frac{2510}{88}$ =28.52

Marks	Frequency (f)	Mid Point (x)	fx	$\sum fx^2$
0 ≤ x < 10	6	5	30	150
10 ≤ x < 20	16	15	240	3600
20 ≤ x < 30	24	25	600	15000
30 ≤ x < 40	25	35	875	30625
40 ≤ x < 50	17	45	765	34425
Total	88		2510	83800

- **Variance** σ^2 $= \frac{\sum fx^2}{n} - \mu^2$

$= \frac{83800}{88} - 28.52^2$ **=952.27 - 813.39**

= 138.88

- **Standard Deviation (** σ **)** $= \sqrt{\frac{\sum(x-\mu)^2}{n}}$ $= \sqrt{\text{Variance}}$

$= \sqrt{138.88}$

= 11.78

Activity 06

ABC company's sales volume in 100 days is listed as below. Find the Standard Deviation.

Sales (Rs. 000)	**No. of days**
10 – 20	5
20 – 30	10
30 – 40	20
40 – 50	30
50- 60	20
60 – 70	10
70 – 80	5
Total	100

Summary_Quartiles:

Ungrouped data:	**Grouped data:**
• First Quartile (Q_1) = $\frac{1}{4}(n+1)^{th}$ observation • Second Quartile (Q_2) = $\frac{1}{2}(n+1)^{th}$ observation • Third Quartile (Q_3) = $\frac{3}{4}(n+1)^{th}$ observation	• First Quartile (Q_1) = L + $\left[\frac{(\frac{n}{4} - CF)}{f}\right]$(i) • Second Quartile (Q2) = L + $\left[\frac{(\frac{2n}{4} - CF)}{f}\right]$(i) • Third Quartile (Q3) = L + $\left[\frac{(\frac{3n}{4} - CF)}{f}\right]$(i)

Interquartile Range = $Q_3 - Q_1$

Semi-Interquartile Range (SIR)/ Quartile Deviation = $\frac{(Q_3 - Q_1)}{2}$

Summary_Variance & Standard Deviation:

Ungrouped data:

- Variance:

$$\sigma^2 = \frac{\sum(x-\mu)^2}{n} \quad \textbf{OR} \quad \sigma^2 = \frac{\sum x^2}{n} - \mu^2$$

- Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x-\mu)^2}{n}}$$

Grouped data:

- Variance:

$$\sigma^2 = \frac{\sum f(x-\mu)^2}{n} \quad \textbf{OR} \quad \sigma^2 = \frac{\sum fx^2}{n} - \mu^2$$

- Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{= \frac{\sum fx^2}{n} - \mu^2}$$