# Measures of dispersion

# Dispersion

- Definition: "Dispersion (also called variability, scatter or spread) is the extent to which a distribution is stretched or squeezed."
- Common examples of measures of statistical dispersion are the variance, standard deviation and interquartile range.

## Types of Measures of Dispersion

There are two main types of dispersion methods in statistics which are:

- Absolute Measure of Dispersion
- Relative Measure of Dispersion

# Common measures of Dispersion
# Absolute dispersion

- An absolute measure of dispersion contains the same unit as the original data set. Absolute dispersion method expresses the variations in terms of the average of deviations of observations like standard or means deviations. It includes range, standard deviation, quartile deviation, etc
- Range
- Inter-quartile range (IQR)
- Variance
- Standard deviation
- Mean deviation
- Standard error of the mean
- Lorenz Curve

# Relative Measure of Dispersion

- The relative measures of depression are used to compare the distribution of two or more data sets. This measure compares values without units. Common relative dispersion methods include:
- Co-efficient of Range
- Co-efficient of Variation
- Co-efficient of Standard Deviation
- Co-efficient of Quartile Deviation
- Co-efficient of Mean Deviation

The common coefficients of dispersion are:

| C.D. In Terms of | Coefficient of dispersion |
|---|---|
| Range | C.D. = $(X_{max} - X_{min}) / (X_{max} + X_{min})$ |
| Quartile Deviation | C.D. = (Q3 − Q1) / (Q3 + Q1) |
| Standard Deviation (S.D.) | C.D. = S.D. / Mean |
| Mean Deviation | C.D. = Mean deviation/Average |

# Range

- The range of a data set is the difference between the largest and smallest data values.
- Example:
- Range of the sample:53, 55, 70, 58, 64, 57, 53, 69, 57, 68, 53
- R = H – L
- Where, R= Range H = Highest value in the observation= 70
- L = Lowest value in the observation= 53
- Range, R= H-L =   70-53=    7

# RANGE

Example: Calculate the range and coefficient of range of the given dataset –

13, 20, 7, 15, 29, 35

Solution:

R = L-S

= 35 – 7 = **28**

Coefficient of range = $\frac{L-S}{L+S} \times 100$

= (35-7/35+7)x 100 = 66.66%

Example:

Find out the Range and coefficient of range of following data.

**59,46,30,23,27,40,52,35,29**

ans. 36, 0.44

# Find range of coefficient

**Calculate the coefficient of range for the following data:**

| Heights in cm. | 120–124 | 125–129 | 130–134 | 135–139 | 140–144 |
|---|---|---|---|---|---|
| No. of students | 6 | 9 | 10 | 12 | 8 |

**Answer**

First let us find the limits of heights (limits of class intervals) and their mid points.

| Height | Height limits | Midpoint of heights | No. of students |
|---|---|---|---|
| $120-124$ | $119.5-124.5$ | $122$ | $6$ |
| $125-129$ | $124.5-129.5$ | $127$ | $9$ |
| $130-134$ | $129.5-134.5$ | $132$ | $10$ |
| $135-139$ | $134.5-139.5$ | $137$ | $12$ |
| $140-144$ | $139.5-144.5$ | $142$ | $8$ |

Upper limit of the highest height, $H = 144.5$, lower limit of the lowest height, $L = 119.5$.

Coefficient of range $= \dfrac{H-L}{H+L} = \dfrac{144.5-119.5}{144.5+119.5} = \dfrac{25}{264} = 0.095.$

# Advantages & disadvantages of the Range

- **Advantages:**
- Easy to compute.
- Easy to understand.
- Scores exist in the data set.
- **Disadvantages:**
- Value depends only on two scores.
- Influenced by sample size.
- Very sensitive to outliers.
- Insensitive to the distribution of scores within the two extremes.
- For example: 1,2,2,3,4,6,7 vs. 1,1,1,1,1,1,7 both have Range, R=6

# Inter-quartile Range (IQR)

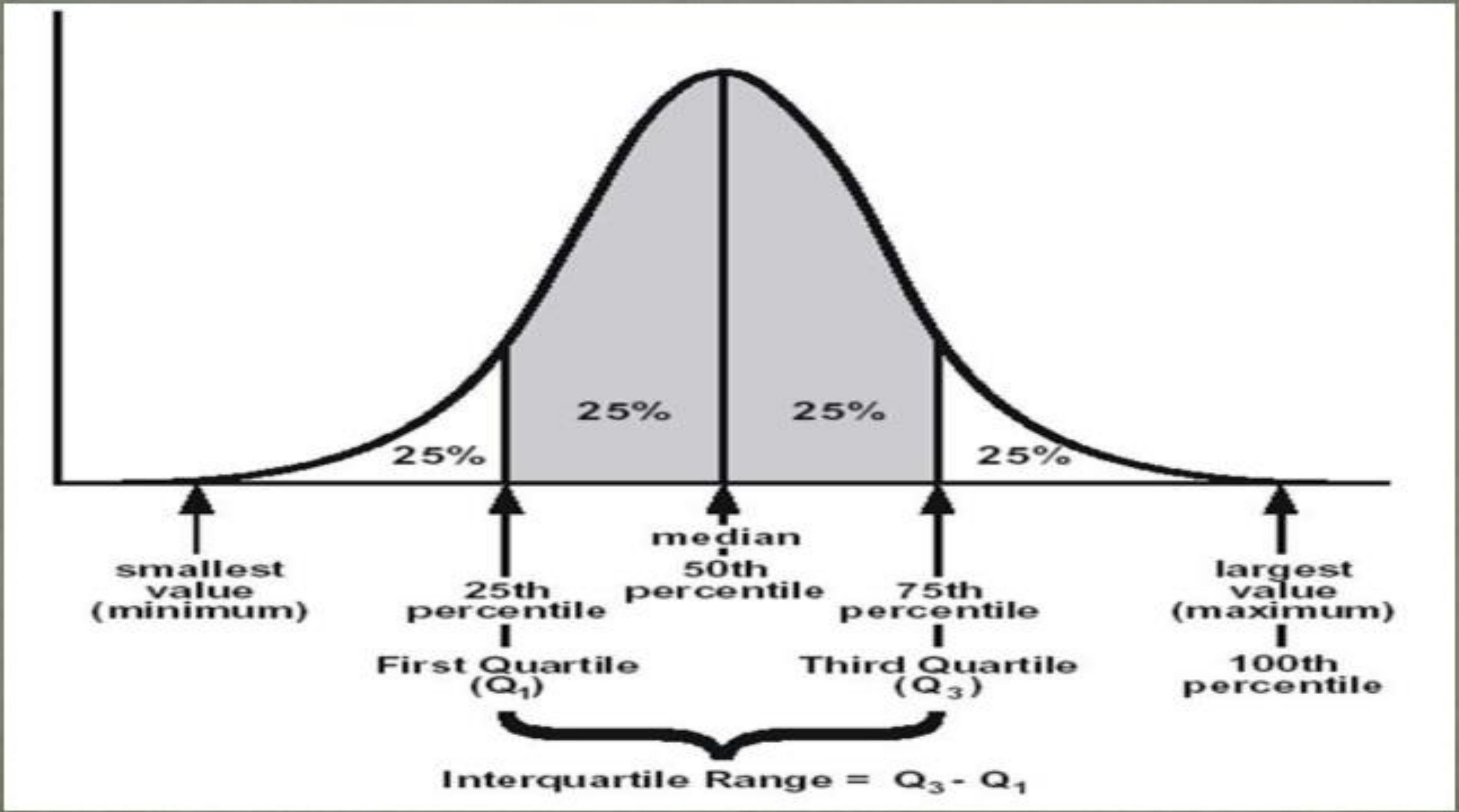
- The inter-quartile range of a data set is the difference between the third quartile(75%) and the first quartile(25%).
- It is the range for the middle 50% of the data.
- It overcomes the sensitivity to extreme data values.

**IQR= Q3 - Q1**

- Where, Q1 = First quartile Q3 = Third quartile
- Semi-inter-quartile range is also known as quartile deviation,

**QD = IQR/2 = Q3 - Q1 /2**

# Graphical Representation of IQR




www.DuloMix.com

- **Advantages of IQR**
- It is not affected by extreme values as in the case of range.
- It is useful in estimating dispersion in grouped data with open ended class.
- **Disadvantages of IQR**
- IQR as a measure of dispersion is most reliable only with symmetrical data series. Unfortunately, in social sciences most of data distributions are generally asymmetrical in nature. So, its use in social sciences is usually limited to data which are moderately skewed.

# MEAN DEVIATION

It is the average of the absolute values of the deviation from the mean (or median or mode).

Mean deviation or MD or $\delta = \sum |x| \div N$

where, MD= mean deviation;

x = deviation from actual mean

$| \;|$ = not considering sign (+ve or -ve )

Deviation, $x = X - \overline{X}$

- **MERITS AND DEMERITS OF MEAN DEVIATION**
- Mean deviation is easy to calculate but since mean deviation has less mathematical value , it is rarely applied for biological statistical analysis. It is also not meaningful because negative sign of deviations is ignored.

# STANDARD DEVIATION

- It may be defined as “the square root of the airthmetic mean of the squares of deviations from the airthmetic mean.”
- Sample Standard Deviation, $s=\sqrt{s^2}$
- Population Standard Deviation, $\sigma= \sqrt{\sigma^2}$

# Variance

Thus, we can take $\frac{1}{n}\sum(x_i - \bar{x})^2$ as a quantity which leads to a proper measure of dispersion. This number, i.e., mean of the squares of the deviations from mean is called the ***variance*** and is denoted by $\sigma^2$ (read as sigma square). Therefore, the variance

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

**15.5.1 *Standard Deviation*** In the calculation of variance, we find that the units of individual observations $x_i$ and the unit of their mean $\bar{x}$ are different from that of variance, since variance involves the sum of squares of $(x_i - \bar{x})$. For this reason, the proper measure of dispersion about the mean of a set of observations is expressed as positive square-root of the variance and is called *standard deviation*. Therefore, the standard deviation, usually denoted by $\sigma$, is given by

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \ldots (1)$$

# Standard Deviation

↳ Population Standard Deviation Formula :—

$$\boxed{\sigma = \sqrt{\frac{\Sigma (X-\mu)^2}{N}}}$$

X = Terms / observations / values.
$\mu$ = Population mean / $\Sigma$ = Sum of.

$\mu = \frac{\Sigma x_i}{N}$ [N = number of items in the Population]

N = number of values in Population

ex.

| Sr.No. | Students | Marks ($X_i$) |
|---|---|---|
| 1 | Tulsi | 600 |
| 2 | Visha | 470 |
| 3 | Raj | 170 |
| 4 | Tirthesh | 430 |
| 5 | Khushi | 300 |

$$\mu = \frac{\Sigma X}{N} = \frac{(600+470+170+430+300)}{5}$$

$$= 1970 / 5$$

$$= 394$$

$$\therefore \boxed{\mu = 394}$$

| Students | Mark (X) | $x-\mu$ | $(x-\mu)^2$ |
|---|---|---|---|
| Tulsi | 600 | 206 | 42436 |
| Visha | 470 | 76 | 5776 |
| Raj | 170 | -226 | 51076 |
| Tirthesh | 430 | 36 | 1296 |
| Khushi | 300 | -94 | 8836 |

$\Sigma(x-\mu)^2 = 109420$

$\mu = 394$

$x-\mu = 600-394 = 206 \qquad (206)^2$

$470-394 = 76 \qquad (76)^2$

$170-394 = -226 \qquad (-226)^2$

$430-394 = 36 \qquad (36)^2$

$300-394 = -94 \qquad (-94)^2$

$$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

$$= \sqrt{\frac{109420}{5}}$$

$$= \sqrt{21884}$$

$$\boxed{\sigma = 147.93}$$ Standard deviation for Population

↳ Sample standard deviation :-

$$\boxed{s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}}$$

$\bar{x}$ = Sample mean/Average
$x$ = Individual Values in Sample
$n$ = number of Individual Values in Sample.

eg.

| Students | Marks (x) | $x - \bar{x}$ | $(x - \bar{x})^2$ |
|---|---|---|---|
| Tulsi | 600 | 206 | 42436 |
| Visha | 470 | 76 | 5776 |
| Raj | 170 | −226 | 51076 |
| Tirthesh | 430 | 430 | 1296 |
| Khushi | 300 | 300 | 8836 |
| | $\Sigma x = 1970$ | | $\Sigma = 109420$ |

$$\bar{x} = \frac{\Sigma x}{n} = \frac{1970}{5} = 394$$

$$x - \bar{x} = 600 - 394 = 206$$

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{109420}{5-1}} = \sqrt{27355}$$

$$\therefore \boxed{s = 165.39}$$

↳ Standard Deviation of a discrete and Continuous frequency.

(1) $$\boxed{\sigma = \sqrt{\frac{1}{N}\sum f_i (x_i - \bar{x})^2}}$$ Actual Mean Method or Direct

$N = \sum f_i$

Discrete frequency

| Marks $x_i$ | Students $f_i$ |
|---|---|
| 5 | 3 |
| 6 | 4 |
| 7 | 2 |
| 8 | 6 |
| 10 | 5 |
| 12 | 10 |

Continous frequency

| Class | Frequency |
|---|---|
| 0 – 10 | 3 |
| 10 – 20 | 4 |
| 20 – 30 | 2 |
| 30 – 40 | 6 |
| 40 – 50 | 5 |
| 50 – 60 | 10 |

Class Interval = 10

| Inclusive Class | Frequency |
|---|---|
| 0 – 9 | 1 |
| 10 – 19 | 2 |
| 20 – 29 | 3 |
| 30 – 39 | 4 |
| 40 – 49 | 5 |
| 50 – 59 | 6 |

Lower limit, Upper limit

Upper Limit ($1^{st}$) and Lower Limit of ($2^{nd}$) class are not same = Inclusive.

For class boundaries –

$$\text{Exclusive} = \frac{\text{Lower boundary of } 2^{nd} \text{ class} - \text{Upper boundary of } 1^{st} \text{ class}}{2}$$

$= \frac{10-9}{2} = \frac{1}{2} = \underline{\underline{0.5}}$

Reduce 0.5 from lower limit of all
Add 0.5 to upper limit of all class.

Exclusive class

- 0.5 — 9.5
9.5 — 19.5
19.5 — 29.5
29.5 — 39.5
39.5 — 49.5
49.5 — 59.5

(2) Short-Cut Method or ~~Actual~~ Assumed mean method :-
for discrete and Continuous frequency.

$$\boxed{\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}}$$

$d = x - A$

A = assumed mean.
n = Total number of observation.
$n = \Sigma f$

↳ Standard Deviation Formula for Grouped Data :-

$$\boxed{\sigma = \frac{1}{N}\sqrt{N\Sigma f_i x_i^2 - (\Sigma f_i x_i)^2}}$$

↳ Another Formula for Standard deviation :-

$$\boxed{\sigma = \frac{h}{N}\sqrt{N\,\Sigma f_i y_i^2 - (\Sigma f_i y_i)^2}}$$

$h$ = width of class intervals

$y_i = \frac{x_i - A}{h}$ , A = assumed mean.

(3) Step Deviation Method :-

$$\boxed{\sigma = \sqrt{\frac{\Sigma f d^2}{n} - \left(\frac{\Sigma f d}{n}\right)^2} \times i}$$

Where,

i = Common class interval

$d = \frac{x - A}{i}$ , A = assumed mean.

# Formulas for Standard Deviation

| | |
|---|---|
| Population Standard Deviation Formula | $\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}}$ |
| Sample Standard Deviation Formula | $s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$ |

## Notations for Standard Deviation

- $\sigma$ = Standard Deviation
- $x_i$ = Terms Given in the Data
- $\bar{x}$ = Mean
- n = Total number of Terms

**15.5.2 *Standard deviation of a discrete frequency distribution*** Let the given discrete frequency distribution be

$$x: \quad x_1, \quad x_2, \quad x_3, \ldots, x_n$$
$$f: \quad f_1, \quad f_2, \quad f_3, \ldots, f_n$$

In this case standard deviation $(\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2}$ ... (2)

where $N = \sum_{i=1}^{n} f_i$ .

**15.5.3** ***Standard deviation of a continuous frequency distribution*** The given continuous frequency distribution can be represented as a discrete frequency distribution by replacing each class by its mid-point. Then, the standard deviation is calculated by the technique adopted in the case of a discrete frequency distribution.

If there is a frequency distribution of $n$ classes each class defined by its mid-point $x_i$ with frequency $f_i$, the standard deviation will be obtained by the formula

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2},$$

where $\bar{x}$ is the mean of the distribution and $N = \sum_{i=1}^{n} f_i$.

## Standard Deviation Formula Based on Discrete Frequency Distribution

For discrete frequency distribution of the type:

**x: $x_1, x_2, x_3, \ldots x_n$** and

**f: $f_1, f_2, f_3, \ldots f_n$**

The formula for standard deviation becomes:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}$$

Here, N is given as:

**$N = {}^{n}\sum_{i=1} f_i$**

## Standard Deviation Formula for Grouped Data

There is another standard deviation formula which is derived from the variance. This formula is given as:

$$\sigma = \frac{1}{N}\sqrt{\sum_{i=i}^{n} f_i x_i^2 - \left(\sum_{i=1}^{n} f_i x_i\right)^2}$$

Another formula for standard deviation :

$$\sigma_x = \frac{h}{N}\sqrt{N\sum_{i=1}^{n} f_i y_i^2 - \left(\sum_{i=1}^{n} f_i y_i\right)^2} \qquad \ldots (5)$$

where $h$ is the width of class intervals and $y_i = \dfrac{x_i - A}{h}$ and A is the assumed mean.

# MERITS AND DEMERITS OF STD. DEVIATION

- Std. Dev. summarizes the deviation of a large distribution from mean in one figure used as a unit of variation.
- It indicates whether the variation of difference of a individual from the mean is real or by chance.
- Std. Dev. helps in finding the suitable size of sample for valid conclusions.
- It helps in calculating the Standard error.
- DEMERITS-
- It gives weightage to only extreme values. The process of squaring deviations and then taking square root involves lengthy calculations.

## CALCULATION OF STANDARD DEVIATION- DISCERETE SERIES OR GROUPED DATA

Three Methods

a) Actual Mean Method or **Direct Method**

b) Assumed Mean Method or **Short-cut Method**

c) Step Deviation Method

**Example :** During a survey, 6 students were asked how many hours per day they study on an average? Their answers were as follows: 2, 6, 5, 3, 2, 3. Evaluate the sample standard deviation.

**Solution:**

Find the mean of the data:

$$\frac{(2+6+5+3+2+3)}{6} = 3.5$$

**Step 2:** Construct the table:

| $x_1$ | $x_1 - \bar{x}$ | $(x_1 - \bar{x})^2$ |
|---|---|---|
| 2 | -1.5 | 2.25 |
| 6 | 2.5 | 6.25 |
| 5 | 1.5 | 2.25 |
| 3 | -0.5 | 0.25 |
| 2 | -1.5 | 2.25 |
| 3 | -0.5 | 0.25 |
| **= 13.5** | | |

**Step 3:** Now, use the Standard Deviation formula

Sample Standard Deviation = $s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$

$=\sqrt{(13.5/[6-1])}$

$=\sqrt{[2.7]}$

$=1.643$

**Example :** Find the mean respiration rate per minute and its standard deviation when in 4 cases the rate was found to be : 16, 13, 17 and 22.

Standard deviation = $\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum d^2}{n}}$

- **Solution:**

Here **Mean =** $\bar{x} = \frac{\sum x}{n} = \frac{16+13+17+22}{4} = \frac{68}{4} = 17$

| $(x)$ | $d = x - \bar{x}$ | $d^2 = (x - \bar{x})^2$ |
|---|---|---|
| 16 | -1 | 1 |
| 13 | -4 | 16 |
| 17 | 0 | 0 |
| 22 | 5 | 25 |
| $\sum x = 68$ | | $\sum d^2 = 42$ |

Standard deviation = $\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum d^2}{n}} = \sqrt{\frac{42}{4}} = 3.2$

**Example 11** Find the standard deviation for the following data :

| $x_i$ | 3 | 8 | 13 | 18 | 23 |
|---|---|---|---|---|---|
| $f_i$ | 7 | 10 | 15 | 10 | 6 |

$$\sigma = \frac{1}{N}\sqrt{N \Sigma f_i x_i^2 - (\Sigma f_i x_i)^2}$$

| $x_i$ | $f_i$ | $f_i x_i$ | $(f_i x_i) x_i$ <br> $f_i x_i^2$ |
|---|---|---|---|
| 3 | 7 | 21 | 63 |
| 8 | 10 | 80 | 640 |
| 13 | 15 | 195 | 2535 |
| 18 | 10 | 180 | 3240 |
| 23 | 06 | 138 | 3174 |
| | $\Sigma f = 48$ | $\Sigma f_i x_i = 614$ | $\Sigma f_i x_i^2 = 9652$ |

$= (\Sigma f_i x_i)^2$

$= (614)^2$

$= 376996$

$$\sigma = \frac{1}{48}\sqrt{48(9652) - 376996}$$

$$= \frac{1}{48}\sqrt{86300}$$

$$= \frac{293}{48}$$

$$\therefore \boxed{\sigma = 6.12}$$

www.DuloMix.com

**Solution :** In the given data,

$f_i x_i = 21, 80, 195, 180, 138$

$f_i x_i^2 = 63, 640, 2535, 3240, 3174$

Now, standard deviation is given by,

$$\sigma = \frac{1}{N}\sqrt{N\sum f_i x_i^2 - \left(\sum f_i x_i\right)^2}$$

Here, $N = \sum f_i = 48$

$\sum f_i x_i = 614$

$\sum f_i x_i^2 = 9652$

So, putting these values,

$$\sigma = \frac{1}{48}\sqrt{48(9652) - 614^2}$$

$$\Rightarrow \sigma = \frac{1}{48}\sqrt{463296 - 376998}$$

$$\Rightarrow \sigma = \frac{1}{48}\sqrt{86300} = \frac{293.76}{48}$$

$$\Rightarrow \sigma = 6.12$$

**Example 9** Find the variance and standard deviation for the following data:

| $x_i$ | 4 | 8 | 11 | 17 | 20 | 24 | 32 |
|---|---|---|---|---|---|---|---|
| $f_i$ | 3 | 5 | 9 | 5 | 4 | 3 | 1 |

Table 15.8

| $x_i$ | $f_i$ | $f_i x_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $f_i (x_i - \bar{x})^2$ |
|---|---|---|---|---|---|
| 4 | 3 | 12 | –10 | 100 | 300 |
| 8 | 5 | 40 | –6 | 36 | 180 |
| 11 | 9 | 99 | –3 | 9 | 81 |
| 17 | 5 | 85 | 3 | 9 | 45 |
| 20 | 4 | 80 | 6 | 36 | 144 |
| 24 | 3 | 72 | 10 | 100 | 300 |
| 32 | 1 | 32 | 18 | 324 | 324 |
| | 30 | 420 | | | 1374 |

$$N = 30, \sum_{i=1}^{7} f_i x_i = 420, \sum_{i=1}^{7} f_i (x_i - \bar{x})^2 = 1374$$

Therefore $\bar{x} = \dfrac{\sum_{i=1}^{7} f_i x_i}{N} = \dfrac{1}{30} \times 420 = 14$

Hence variance $(\sigma^2) = \dfrac{1}{N} \sum_{i=1}^{7} f_i (x_i - \bar{x})^2$

$$= \frac{1}{30} \times 1374 = 45.8$$

and Standard deviation $(\sigma) = \sqrt{45.8} = 6.77$

**Example 10** Calculate the mean, variance and standard deviation for the following distribution :

| Class | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
|---|---|---|---|---|---|---|---|
| Frequency | 3 | 7 | 12 | 15 | 8 | 3 | 2 |

Page No.: 
Date: 

| Class | Freq. $f_i$ | Mid Point $x_i$ | $f_ix_i$ | $f_ix_i^2$ |
|---|---|---|---|---|
| 30-40 | 3 | 35 | 905 | 3675 |
| 40-50 | 7 | 45 | 315 | 14175 |
| 50-60 | 12 | 55 | 660 | 36300 |
| 60-70 | 15 | 65 | 975 | 63375 |
| 70-80 | 8 | 75 | 600 | 45000 |
| 80-90 | 3 | 85 | 255 | 21675 |
| 90-100 | 2 | 95 | 190 | 18050 |
| | 50 | | 3100 | 202250 |

$(\Sigma f_ix_i)^2 = (3100)^2 = 9100$

$$\sigma = \frac{1}{N}\sqrt{N\Sigma f_ix_i^2 - (\Sigma f_ix_i)^2}$$

$$= \frac{1}{50}\sqrt{50(202250) - 9090000}$$

$$= \frac{1}{50}\sqrt{502500}$$

$$= 14.177$$

$$\boxed{\sigma = 14.18}$$

www.DuloMix.com

| Class | Frequency ($f_i$) | Mid-point ($x_i$) | $f_i x_i$ | $(x_i - \bar{x})^2$ | $f_i(x_i - \bar{x})^2$ |
|---|---|---|---|---|---|
| 30-40 | 3 | 35 | 105 | 729 | 2187 |
| 40-50 | 7 | 45 | 315 | 289 | 2023 |
| 50-60 | 12 | 55 | 660 | 49 | 588 |
| 60-70 | 15 | 65 | 975 | 9 | 135 |
| 70-80 | 8 | 75 | 600 | 169 | 1352 |
| 80-90 | 3 | 85 | 255 | 529 | 1587 |
| 90-100 | 2 | 95 | 190 | 1089 | 2178 |
| | 50 | | 3100 | | 10050 |

Thus Mean $\bar{x} = \frac{1}{N}\sum_{i=1}^{7} f_i x_i = \frac{3100}{50} = 62$

$$\text{Variance } (\sigma^2) = \frac{1}{N}\sum_{i=1}^{7} f_i (x_i - \bar{x})^2$$

$$= \frac{1}{50} \times 10050 = 201$$

and Standard deviation $(\sigma) = \sqrt{201} = 14.18$

**Examples 12** Calculate mean, variance and standard deviation for the following distribution.

| Classes | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
|---|---|---|---|---|---|---|---|
| Frequency | 3 | 7 | 12 | 15 | 8 | 3 | 2 |

**Solution** Let the assumed mean A = 65. Here $h = 10$

We obtain the following Table 15.11 from the given data :

**Table 15.11**

| Class | Frequency $f_i$ | Mid-point $x_i$ | $y_i = \frac{x_i - 65}{10}$ | $y_i^2$ | $f_i y_i$ | $f_i y_i^2$ |
|---|---|---|---|---|---|---|
| 30-40 | 3 | 35 | – 3 | 9 | – 9 | 27 |
| 40-50 | 7 | 45 | – 2 | 4 | – 14 | 28 |
| 50-60 | 12 | 55 | – 1 | 1 | – 12 | 12 |
| 60-70 | 15 | 65 | 0 | 0 | 0 | 0 |
| 70-80 | 8 | 75 | 1 | 1 | 8 | 8 |
| 80-90 | 3 | 85 | 2 | 4 | 6 | 12 |
| 90-100 | 2 | 95 | 3 | 9 | 6 | 18 |
| | N=50 | | | | – 15 | 105 |

Therefore $$\bar{x} = A + \frac{\sum f_i y_i}{50} \times h = 65 - \frac{15}{50} \times 10 = 62$$

Variance $$\sigma^2 = \frac{h^2}{N^2}\left[N\Sigma f_i y_i^2 - (\Sigma f_i y_i)^2\right]$$

$$= \frac{(10)^2}{(50)^2}\left[50 \times 105 - (-15)^2\right]$$

$$= \frac{1}{25}[5250 - 225] = 201$$

and standard deviation $(\sigma) = \sqrt{201} = 14.18$

www.DuloMix.com

- Example.

Find the variance and standard deviation of the following data by using assume mean method.

6, 8, 10, 12, 14, 16, 18, 20, 22, 24

- Solution:

| $x_i$ | $d_i = \frac{x_i - 14}{2}$ | Deviations from mean $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ |
|---|---|---|---|
| 6 | –4 | –9 | 81 |
| 8 | –3 | –7 | 49 |
| 10 | –2 | –5 | 25 |
| 12 | –1 | –3 | 9 |
| 14 | 0 | –1 | 1 |
| 16 | 1 | 1 | 1 |
| 18 | 2 | 3 | 9 |
| 20 | 3 | 5 | 25 |
| 22 | 4 | 7 | 49 |
| 24 | 5 | 9 | 81 |
| | 5 | | 330 |

Therefore Mean $\bar{x}$ = assumed mean + $\frac{\sum_{i=1}^{n} d_i}{n} \times h = 14 + \frac{5}{10} \times 2 = 15$

and Variance $(\sigma^2) = \frac{1}{n}\sum_{i=1}^{10}(x_i - \bar{x})^2 = \frac{1}{10} \times 330 = 33$

Thus Standard deviation $(\sigma) = \sqrt{33} = 5.74$

9. Find the mean, variance and standard deviation using short-cut method

| Height in cms | 70-75 | 75-80 | 80-85 | 85-90 | 90-95 | 95-100 | 100-105 | 105-110 | 110-115 |
|---|---|---|---|---|---|---|---|---|---|
| No. of children | 3 | 4 | 7 | 7 | 15 | 9 | 6 | 6 | 3 |

# “ Thank You ”

www.DuloMix.com