Lecture 01-02 (Basic definition)

Definition: Statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting, and analyzing data as well as helps to draw valid conclusions and make reasonable decisions on the basis of such analysis.

Data: The facts and statistics collected together for reference or analysis. **OR,** Data is a collection of facts, such as numbers, words, measurements, observations or just descriptions of things.

Types of Data:

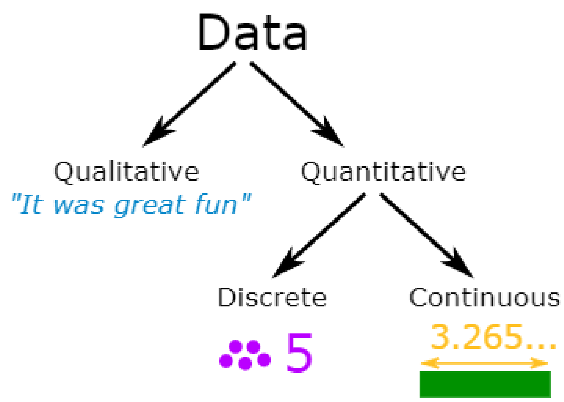


Fig. 1 Types of data.

Example: What do we know about the Dog Fig. 2?

Qualitative:

- He is brown and black
- He has long hair
- He has lots of energy

Quantitative:

- Discrete:
    - He has 4 legs
    - He has 2 eyes
- Continuous:
    - He weighs 25.5 kg
    - He is 565 mm tall

Fig. 2 A dog is jumping.

Collecting: Data can be collected in many ways.

1. Direct observation (It is the simplest way).
2. Survey questions and etc. (Your TER form)

Example: Counting Cars
You want to find how many cars pass by a certain point on a road in a 10-minute interval.
So: stand near that road, and count the cars that pass by in 10 minutes.
You might want to count many 10-minute intervals at different times during the day, and on different days too!

Fig. 3 Cars passing through a certain points of road.

Population: Population means an aggregate of elements possessing certain characteristics of interest in any particular investigation or enquiry.

Sample: Sample is a representative part of the population.

Variable: The measurements of elements of a population having certain characteristics may vary from elements to elements either in magnitude or in quantity. These measurable characteristics are called variable. If variable can take only one value, it is called constant.

Raw data: Raw data are collected data that have not been organized numerically. An example is the set of heights of 100 male students obtained from an alphabetical listing of university records.

Array: An array is an arrangement of raw numerical data in ascending or descending order of magnitude. The difference between the largest and smallest numbers is called the range of the data. For example, if the largest height of 100 male students is 74 inches (in) and the smallest height is 60 in, the range is 74 60 ¼ 14 in.

Frequency distribution: When summarizing large masses of raw data, it is often useful to distribute the data into classes, or categories, and to determine the number of individuals belonging to each class, called the class frequency. A tabular arrangement of data by classes together with the corresponding class frequencies is called a frequency distribution, or frequency table. Table 1 is a frequency distribution of heights (recorded to the nearest inch) of 100 students at Northern University Bangladesh.

Table 1 Heights of 100 male students at Northern University Bangladesh.

Height (in)	Number of students
60-62	5

63-65	18
66-68	42
69-71	27
72-74	8
Total	100

Problem 2 Construct a frequency distribution table for the following raw data (height of 20 students): 50, 54, 56, 60, 55, 45, 48, 51, 53, 57, 49, 50, 52, 58, 57, 59, 57, 58, 55, and 54.

Graph: A graph is a pictorial presentation of the relationship between variables. Many types of graphs are employed in statistics, depending on the nature of the data involved and the purpose for which the graph is intended. Among these are bar graphs, pie graphs, pictographs, etc.

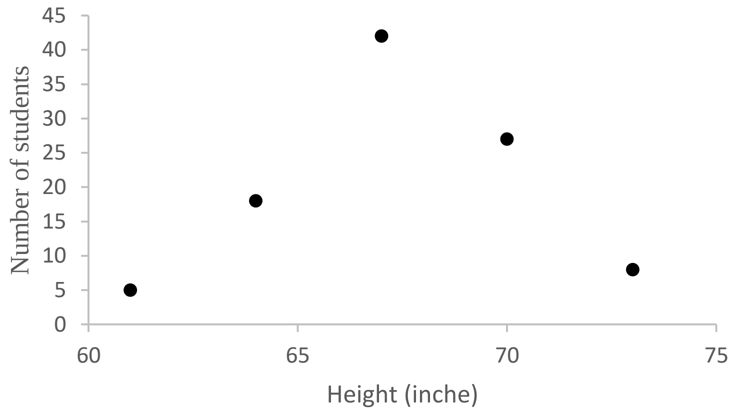
Graphical representation of frequency distribution:

a. Dot frequency diagram
b. Histogram
c. Frequency polygon
d. Cumulative frequency polygon
e. Cumulative frequency curve

Dot frequency diagram: The frequency distribution of 100 students is given in Table 1. We have to plot the dot frequency diagram. In this respect, firstly it is required to find the middle class. Then, we have to plot frequencies relative to the respective middle classes.

Height (in)	Middle class (x_i)	Number of students (f_i)
60-62	61	5
63-65	64	18
66-68	67	42
69-71	70	27
72-74	73	8
Total 100		

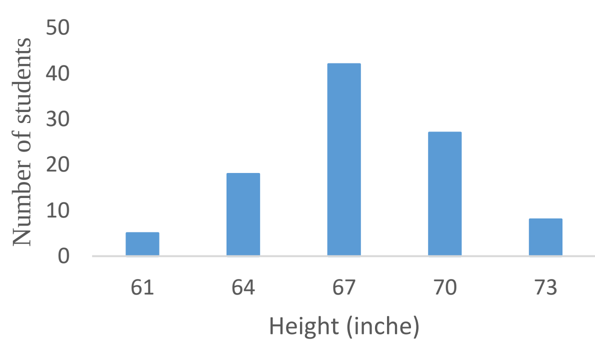
Figure 1 Dot frequency diagram for given data 1



Bar Diagram: The frequency distribution of 100 students is given in Table 1. We have to plot the Bar diagram. In this respect, firstly it is required to find the middle class. Then, we have to plot frequencies relative to the respective class.

Height (in)	Middle class (x_i)	Number of students (f_i)
60-62	61	5
63-65	64	18
66-68	67	42
69-71	70	27
72-74	73	8
	Total	100

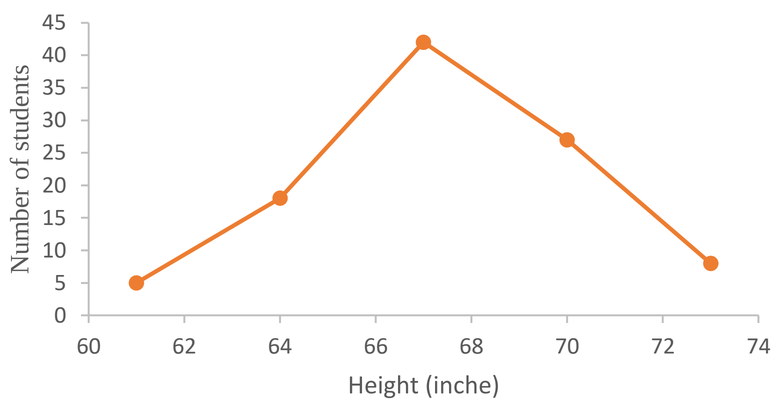
Figure 2 Bar diagram for the given data



Frequency polygon: The frequency distribution of 100 students is given in Table 1. We have to draw frequency polygon. In this respect, firstly it is required to find the middle class. Then, we have to plot frequencies relative to the respective class.

Height (in)	Middle class (x_i)	Number of students (f_i)
60-62	61	5
63-65	64	18
66-68	67	42
69-71	70	27
72-74	73	8
	Total	100

Figure 3 Frequency polygon for given data



Cumulative frequency polygon: (Do yourself)

Cumulative frequency curve: (Do yourself)

Class intervals and class limits:

A symbol defining a class, such as 60–62 in Table 1, is called a class interval. The end numbers, 60 and 62, are called class limits; the smaller number (60) is the lower class limit, and the larger number (62) is the upper class limit. The terms class and class interval are often used interchangeably, although the class interval is actually a symbol for the class. A class interval that, at least theoretically, has either no upper class limit or no lower class limit indicated is called an open class interval. For example, referring to age groups of individuals, the class interval ''65 years and over'' is an open class interval.

The size, or width, of a class interval:

The size, or width, of a class interval is the difference between the lower and upper class boundaries and is also referred to as the class width, class size, or class length. If all class intervals of a frequency distribution have equal widths, this common width is denoted by c. In such case c is equal to the difference between two successive lower class limits or two successive upper class limits. For the data of Table 1, for example, the class interval is C = 62:5-59:5 = 65:5 - 62:5 = 3.

Assignment:

1. Write down the uses of statistics in different sectors in your own words.
2. Why will we study statistics in textile sector?
3. Produce a frequency distribution table of GPA of last semester final examination for the students of this section.

Home work:

1. Write down some examples of population, sample, and variable in your own words.

Lecture 03 (Measures of Location)

Average or measures of central tendency: An average is a value that is typical, or representative, of a set of data. Since such typical values tend to lie centrally within a set of data arranged according to magnitude, averages are also called measures of central tendency.

Several types of averages can be defined, the most common being the arithmetic mean, the median, the mode, the geometric mean, and the harmonic mean. Each has advantages and disadvantages, depending on the data and the intended purpose.

Arithmetic mean: The arithmetic mean of a set of N numbers $x_1, x_2, x_3, \dots x_N$ is denoted by $\overline{X}$ and defined by,

$$\overline{X} = \frac{x_1+x_2+x_3+\dots+x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}\sum x.$$

For group data/frequency distribution, arithmetic mean can defined as $AM = \frac{1}{N}\sum_{i=1}^{N} f_i x_i$.

Properties of arithmetic mean:

a. The algebraic sum of the deviations of a set of numbers from their arithmetic mean is zero.

Geometric mean: The geometric mean of a set of N positive numbers $x_1, x_2, x_3, \ldots x_N$ is denoted by G and defined by, $G = \sqrt[N]{x_1 x_2 x_3 \ldots x_N}$.

For group data, $G = 10^{\frac{1}{N}\sum f_i \log(x_i)}$.

Harmonic mean: The harmonic mean of a set of N nonzero numbers $x_1, x_2, x_3, \ldots x_N$ is denoted by H and defined by, $H = \frac{1}{\frac{1}{N}\left(\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_N}\right)} = \frac{1}{\frac{1}{N}\sum \frac{1}{x_i}}$.

For group data, $H = \frac{1}{\frac{1}{N}\sum f_i \frac{1}{x_i}}$.

Quadratic mean (Root mean square): The root mean square (RMS), or quadratic mean, of a set of numbers $x_1, x_2, x_3, \ldots x_N$ is sometimes denoted by $\sqrt{\overline{X}^2}$ and is defined by,

$$RMS = \sqrt{\overline{X}^2} = \sqrt{\frac{1}{N}\sum x_i^2}.$$

$$RMS = \sqrt{\overline{X}^2} = \sqrt{\frac{1}{N}\sum f_i x_i^2}. \text{ (For group data)}$$

Problem: Find the arithmetic mean, geometric mean and harmonic mean of data 8, 5, 6, 3, 10, 12, 13, 9, 7, 6, 5, 3, 8.

Solution: $AM = \frac{8+5+6+3+10+12+13+9+7+6+5+3+8}{13} = 7.308$

$GM = \sqrt[13]{8 \times 5 \times 6 \times 3 \times 10 \times 12 \times 13 \times 9 \times 7 \times 6 \times 5 \times 3 \times 8}$ =6.6623 (do yourself)

$HM = \frac{1}{\frac{1}{13}\left(\frac{1}{8}+\frac{1}{5}+\frac{1}{6}+\frac{1}{3}+\frac{1}{10}+\frac{1}{12}+\frac{1}{13}+\frac{1}{9}+\frac{1}{7}+\frac{1}{6}+\frac{1}{5}+\frac{1}{3}+\frac{1}{8}\right)}$ =6.0068 (do yourself)

Problem: Daily wages range of 35 labors in a certain factory are given bellow. Find the arithmetic mean by **both direct** and **short-cut method**.

Class	11-13	13-15	15-17	17-19	19-21	21-23	23-25
Number of labors	3	4	5	10	6	4	3

Solution:

Class of interval	Middle class (x_i)	Number of labors (f_i)	$f_i x_i$	$u_i = \frac{x_i - A}{h}$	$f_i u_i$
11-13	12	3	36	-3	-9
13-15	14	4	56	-2	-8
15-17	16	5	80	-1	-5
17-19	18	10	180	0	0
19-21	20	6	120	1	6
21-23	22	4	88	2	8
23-25	24	3	72	3	9
	Total	35	632		1

For direct method

$$AM = \frac{1}{N}\sum f_i x_i = \frac{632}{35} = 18.06.$$

For short-cut method

$$AM = A + \left(\frac{1}{N}\sum f_i u_i\right) \times h = 18 + \frac{1}{35} \times 2 = 18 + \frac{2}{35} = 18.06.$$

Problem: Daily wages range of 35 labors in a certain factory are given bellow. Find the Geometric mean and Harmonic mean.

Class	11-13	13-15	15-17	17-19	19-21	21-23	23-25
Number of labors	3	4	5	10	6	4	3

Solution:

Geometric mean (GM):

Class of interval	Middle class (x_i)	Number of labors (f_i)	$log(x_i)$	$f_i\, log(x_i)$
11-13	12	3	1.0792	3.2376
13-15	14	4	1.1461	4.5844
15-17	16	5	1.2041	6.0205
17-19	18	10	1.2553	12.5530
19-21	20	6	1.3010	7.8060
21-23	22	4	1.3424	5.3696
23-25	24	3	1.3802	4.1406
	Total	N =35	$\sum f_i\, log(x_i)$ =43.7117	

It is known to us, $GM = 10^{\frac{1}{N}\sum f_i\, log(x_i)} = 10^{\frac{1}{35}(43.7117)} = 10^{1.2489}$

or, $GM = 10^{1.2489} = 17.74$ (approximately)

Harmonic mean (HM):

Class of interval	Middle class (x_i)	Number of labors (f_i)	$\frac{1}{x_i}$	$f_i \frac{1}{x_i}$
11-13	12	3	0.0833	0.2499
13-15	14	4	0.0714	0.2856
15-17	16	5	0.0625	0.3125
17-19	18	10	0.0556	0.5560
19-21	20	6	0.0500	0.3000
21-23	22	4	0.455	0.2176
23-25	24	3	0.0417	0.1251
	Total	N =35	$\sum f_i \frac{1}{x_i}$ =2.0467	

It is known to us harmonic mean, $HM = \frac{1}{\frac{1}{N}\sum f_i \frac{1}{x_i}} = \frac{1}{\frac{1}{35}(2.0467)} = \frac{35}{2.0467} = 17.10$ (approximately)

Problem: Find the relation between arithmetic mean, geometric mean and harmonic mean.

Problem: For two positive data, show that $AM \times HM = GM^2$. (HW)

Problem: Daily wages range of 35 labors in a certain factory are given bellow. Find the quadratic mean or root mean square (RMS)

Class	11-13	13-15	15-17	17-19	19-21	21-23	23-25
Number of labors	3	4	5	10	6	4	3

Lecture 04 (Median, Mode, Quartile, Percentile)

Median: The median of a set of numbers arranged in order of magnitude (i.e., in an array) is either the middle value or the arithmetic mean of the two middle values.

Examples:

For ungroup data:

If total number of data (N) is odd then, $\left(\frac{N+1}{2}\right)th$ data will be median.

If total number of data (N) is even then, the average of $\left(\frac{N}{2}\right)th$ and $\left(\frac{N}{2}+1\right)th$ data will be median.

For group data:

$$median = L + \frac{\frac{N}{2} - F}{f} \times h,$$

where, L is the lower limit of **median class**, F **cumulative frequency** of pre **median class**, f frequency of **median class**, h length of **median class**, N total number of observations.

Mode: The mode of a set of numbers is that value which occurs with the greatest frequency; that is, it is the most common value. The mode may not exist, and even if it does exist it may not be unique. A distribution having only one mode is called unimodal.

Examples:

a. The set 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, and 18 has mode 9.
b. The set 3, 5, 8, 10, 12, 15, and 16 has no mode.
c. The set 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, and 9 has two modes, 4 and 7, and is called bimodal.

For group data

$$\text{mode} = L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) h,$$

where, L is the lower limit of modal class, Δ_1 difference in frequencies of modal and pre-modal class, Δ_2 difference in frequencies of modal and post modal class, h length of modal class.

Problem: Daily wages range of 35 labors in a certain factory is given bellow. Find the median and mode from the given frequency distribution.

Class	11-13	13-15	15-17	17-19	19-21	21-23	23-25

Number of labors	3	4	5	10	6	4	3

Solution:

Class of interval	Number of labors (f_i)	Cumulative frequency (F)
11-13	3	3
13-15	4	7
15-17	5	12
17-19	10	22
19-21	6	28
21-23	4	32
23-25	3	35
	Total=35	

It is known to us, $median = L + \frac{\frac{N}{2}-F}{f} \times h.$ (1)

Now, we have to identify the median class. Here the half of the total frequency is $\frac{35}{2} = 17.5$ which lies in in range of 22 (cumulative frequency) and 22 lies in the class 17-19. Hence the class 17-19 is the median class. Thus $L = 17$, $N = 35$, $F = 12$, $f = 10$ and $h = 19 - 17 = 2$.

Now, from Eq. (1), we have

$$median = L + \frac{\frac{N}{2} - F}{f} \times h = 17 + \frac{\frac{35}{2} - 12}{10} \times 2 = 17 + \frac{17.5 - 12}{10} \times 2$$
$$= 17 + 0.55 \times 2 = 18.1$$

Again, mode $= L + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right) h.$ (2)

Here, class 17-19 is the modal class.

Thus, $L = 17$, $\Delta_1 = 10 - 5 = 5$, $\Delta_2 = 10 - 6 = 4$, and $h = 19 - 17 = 2$.

Now, from Eq. (2), we have

mode $= L + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right) h = 17 + \left(\frac{5}{5+4}\right) \times 2 =$ (do yourself)

Quartile, deciles and percentiles: If a set of data is arranged in order of magnitude, the middle value (or arithmetic mean of the two middle values) that divides the set into two equal parts is the median. By extending this idea, we can think of those values which divide the set into four equal parts. These values, denoted by Q_1, Q_2, and Q_3, are called the first, second, and third quartiles, respectively, the value *Q2* being equal to the median.

Similarly, the values that divide the data into 10 equal parts are called deciles and are denoted by $D_1, D_2, \ldots, D_9$, while the values dividing the data into 100 equal parts are called percentiles and are denoted by $P_1, P_2, \ldots, P_{99}$.

For a grouped frequency distribution, the quartiles are given by,

$$Q_i = L_i + \frac{\frac{N}{4}i - F_i}{f_i} \times h; i = 1,2,3.$$

For a grouped frequency distribution, the deciles are given by,

$$D_i = L_i + \frac{\frac{N}{10}i - F_i}{f_i} \times h; i = 1,2,3\ldots,9.$$

For a grouped frequency distribution, the percentile is given by,

$$P_i = L_i + \frac{\frac{N}{100}i - F_i}{f_i} \times h; i = 1,2,3\ldots,99.$$

Problem 03: Frequency distribution of the weekly wages of 65 employees at the P&R Company is given bellow

Wages ($)	Number of employees
250-260	8
260-270	10
270-280	16
280-290	14
290-300	10
300-310	5
310-320	2
Total=	65

Find $Q_1, Q_2, Q_3, D_1, \ldots\ D_9, P_{10}, P_{53}, P_{76}$, and P_{95}.

Solution: Please see the following table

class	f_i	F_i
250-260	8	8
260-270	10	18
270-280	16	34
280-290	14	48
290-300	10	58
300-310	5	63
310-320	2	65
Total=	65	

We know, quartiles are given by,

$$Q_i = L_i + \frac{\frac{N}{4}i - F_i}{f_i} \times h; i = 1,2,3.$$

For first quartile (i=1) $\frac{N}{4}i = \frac{65}{4} \times 1$=16.25 which lies in class 260-270. Thus, $L_1 = 260, F_1 = 8, f_1 = 10$, and $h = 10$. Therefore,

$$Q_1 = L_1 + \frac{\frac{N}{4} \times 1 - F_1}{f_1} \times h = 260 + \frac{16.25 - 8}{10} \times 10 = 260 + 8.25 = 268.25.$$

For second quartile (i=2) $\frac{N}{4}i = \frac{65}{4} \times 2$=32.50 which lies in class 270-280. Thus, $L_2 = 270, F_2 = 18, f_2 = 16$, and $h = 10$. Therefore,

$$Q_2 = L_2 + \frac{\frac{N}{4} \times 2 - F_2}{f_2} \times h = 270 + \frac{32.50 - 18}{16} \times 10 = 270 + \frac{14.50 \times 10}{16} =?$$

For third quartile (i=3) $\frac{N}{4}i = \frac{65}{4} \times 3$=48.75 which lies in class 290-300. Thus, $L_3 = 290, F_3 = 48, f_3 = 10$, and $h = 10$. Therefore,

$$Q_3 = L_3 + \frac{\frac{N}{4} \times 3 - F_3}{f_3} \times h = 290 + \frac{48.75 - 48}{10} \times 10 =?$$

Problem for practice:

1. Write down the merits and demerits of arithmetic mean, geometric mean and harmonic mean. (Assignment)
2. Write down the merits and demerits of median and mode. (Assignment)
3. For any two positive data show that $(GM)^2 = AM \times HM$.

4. Prove that the quadratic mean (QM) of any two positive unequal numbers *a* and *b* is greater than their geometric mean.
5. For *n* positive observations, show that *arithmetic mean≥geometric mean≥harmonic mean.*
6. Form a frequency distribution table from the following data and hence draw a different frequency diagrams (scattered dot plot, histogram, polygon).

   3, 4, 4, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 9, 10, 10, 10, 10, 10, 10, 12, 13, 15, 15, 16, 17, 17, 18, 19, 21, 23, 23, 25, 26, 27, 28, 29, 31, 33, 35, 34, 34, 35, 36, 37, 37, 38, 39, 42, 45.
7. Find the arithmetic mean, median and mode for the following frequency distribution:

Class	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45
frequency	3	4	6	7	12	9	6	3

8. Find the arithmetic mean (in both direct and shortcut method), geometric mean, harmonic mean, median and mode for the following frequency distribution.

Class	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
frequency	2	5	7	13	21	16	8	3

9. Suppose you have some data including positive, negative and zero. Which mean do you like to choose for measuring data location and why?

Lecture 05 (Measures of Dispersion)

Dispersion, or variation: The degree to which numerical data tend to spread about an average value is called the dispersion, or variation, of the data. Various measures of this dispersion (or variation) are available, the most common being the range, mean deviation, semi-interquartile range, 10–90 percentile range, and standard deviation.

Range: The range of a set of numbers is the difference between the largest and smallest numbers in the set.

The mean deviation: The mean deviation, or average deviation, of a set of N numbers $x_1, x_2, x_3, \ldots x_N$ is abbreviated by MD and is defined by

$$MD = \frac{1}{N}\sum |x_i - \overline{x}|.$$

For group data

$$MD = \frac{1}{N}\sum f_i |x_i - \overline{x}|.$$

The semi-interquartile range: The semi-interquartile range, or quartile deviation, of a set of data is denoted by Q and is defined by

$$Q = \frac{Q_3 - Q_1}{2},$$

where Q_1 and Q_3 are the first and third quartiles for the data. The interquartile range $(Q_3 - Q_1)$ is sometimes used, but the semi-interquartile range is more common as a measure of dispersion.

The 10–90 percentile range: The 10–90 percentile range of a set of data is defined by

$$\text{10-90 percentile range} = P_{90} - P_{10},$$

where P_{10} and P_{90} are the 10^{th} and 90^{th} percentiles for the data. The semi-10–90 percentile range, $\frac{1}{2}(P_{90} - P_{10})$, can also be used but is not commonly employed.

Problem 02: Find the semi-interquartile range, or quartile deviation, and 10–90 percentile range from the following frequency distribution.

Wages ($)	Number of employees
250-260	8
260-270	10
270-280	16
280-290	14
290-300	10
300-310	5
310-320	2
Total=	65

Solution: Please see the following table

class	f_i	F_i
250-260	8	8
260-270	10	18
270-280	16	34
280-290	14	48
290-300	10	58
300-310	5	63

310-320	2	65
Total=	65	

We know, quartiles are given by,

$$Q_i = L_i + \frac{\frac{N}{4}i - F_i}{f_i} \times h; i = 1, 2, 3.$$

For first quartile (i=1) $\frac{N}{4}i = \frac{65}{4} \times$ 1=16.25 which lies in class 260-270. Thus, $L_1 = 260, F_1 = 8, f_1 = 10$, and $h = 10$. Therefore,

$$Q_1 = L_1 + \frac{\frac{N}{4} \times 1 - F_1}{f_1} \times h = 260 + \frac{16.25 - 8}{10} \times 10 = 260 + 8.25 = 268.25.$$

For third quartile (i=3) $\frac{N}{4}i = \frac{65}{4} \times$ 3=48.75 which lies in class 290-300. Thus, $L_1 = 290, F_1 = 48, f_1 = 10$, and $h = 10$. Therefore,

$$Q_3 = L_3 + \frac{\frac{N}{4} \times 3 - F_3}{f_3} \times h = 290 + \frac{48.75 - 48}{10} \times 10 = 290.75$$

Now, the semi-interquartile range $Q = \frac{1}{2}(Q_3 - Q_1) = \frac{1}{2}(290.75 - 268.25) = \frac{1}{2}.22.50 = 11.25$

10-90 percentile range= $P_{90} - P_{10}$ =?

The semi 10-90 percentile range= $\frac{1}{2}(P_{90} - P_{10})$ =?

The standard deviation: The standard deviation of a set of *N* numbers $x_1, x_2, x_3, \ldots x_N$ is denoted by σ and is defined by

$$\sigma = \sqrt{\frac{1}{N}\sum(x_i - \overline{x})^2}.$$

Thus σ is the root mean square (RMS) of the deviations from the mean, or, as it is sometimes called, the root-mean-square deviation.

For the group data,

$$\sigma = \sqrt{\frac{1}{N}\sum f_i(x_i - \overline{x})^2}.$$

Short method for computing standard deviation

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

If sample data are given then standard deviation is dented by s and defined by

$$s = \sqrt{\frac{1}{N-1}\sum (x_i - \overline{x})^2}$$

Problem: Find the standard deviation (SD) and variance from the following frequency distribution.

Class	11-13	13-15	15-17	17-19	19-21	21-23	23-25
Number of labors	3	4	5	10	6	4	3

Solution:

Class	x_i	f_i	$f_i x_i$	${x_i}^2$	$f_i {x_i}^2$

11-13	12	3	36	144	432
13-15	14	4	56	196	784
15-17	16	5	80	256	1280
17-19	18	10	180	324	3240
19-21	20	6	120	400	2400
21-23	22	4	88	484	1936
23-25	24	3	72	576	1728
	Total	35	632		11800

We know,

$$\text{Standard deviation, } SD = \sqrt{\frac{1}{N}\sum f_i x_i^2 - \left(\frac{1}{N}\sum f_i x_i\right)^2} = \sqrt{\frac{1}{35} \times 11800 - \left(\frac{1}{35} \times 632\right)^2}$$

$$= \sqrt{337.143 - 326.06} = \sqrt{11.083} = 3.329 \text{ (Approximately)}$$

$$\text{Variance, } \sigma^2 = \frac{1}{N}\sum f_i x_i^2 - \left(\frac{1}{N}\sum f_i x_i\right)^2 = 11.083 \text{ (approximately)}$$

The variance: The variance of a set of data is defined as the square of the standard deviation and is thus given by σ^2. When it is necessary to distinguish the standard deviation of a population from the standard deviation of a sample drawn from this population, we often use the symbol s for the latter and σ (lowercase Greek sigma) for the former. Thus s^2 and σ^2 would represent the sample variance and population variance, respectively.

Coefficient of variation (CV): If the absolute dispersion is the standard deviation s and if the average is the mean $\bar{x}$, then the relative dispersion is called the coefficient of variation, or coefficient of dispersion; it is denoted by V and is given by

$$\text{Coefficient of variation}, CV(\%) = \frac{\text{SD}}{\bar{\text{x}}} \times 100\%,$$

and is generally expressed as a percentage. Note that the coefficient of variation is independent of the units used. For this reason, it is useful in comparing distributions where the units may be different. A disadvantage of the coefficient of variation is that it fails to be useful when $\bar{x}$ is close to zero.

Problem Show that for two observations, standard deviation is the half of the range.

Solution: Let *a* and *b* be two data.

The mean of the data $\bar{x} = \frac{a+b}{2}$.

We know standard deviation, $SD = \sqrt{\left(\frac{1}{N}\sum(x_i - \bar{x})^2\right)}$.

or, $SD = \sqrt{\left(\frac{1}{N}\sum(x_i - \bar{x})^2\right)}$

or, $SD = \sqrt{\left(\frac{1}{2}\left\{\left(a - \frac{a+b}{2}\right)^2 + \left(b - \frac{a+b}{2}\right)^2\right\}\right)}$

or, $SD = \sqrt{\left(\frac{1}{2}\left\{\left(\frac{2a-a-b}{2}\right)^2 + \left(\frac{2b-a-b}{2}\right)^2\right\}\right)}$

or, $SD = \sqrt{\left(\frac{1}{2}\left\{\left(\frac{a-b}{2}\right)^2 + \left(\frac{-a+b}{2}\right)^2\right\}\right)}$

or, $SD = \sqrt{\left(\frac{1}{2}\left\{\left(\frac{a-b}{2}\right)^2 + \left(\frac{a-b}{2}\right)^2\right\}\right)}$

or, $SD = \sqrt{\left(\frac{1}{2}.2.\left(\frac{a-b}{2}\right)^2\right)}$

or, $SD = \sqrt{\left(\left(\frac{a-b}{2}\right)^2\right)}$

or, $SD = \sqrt{\left(\left|\frac{a-b}{2}\right|^2\right)}$

or, $SD = \left|\frac{a-b}{2}\right|$

or, $SD = \frac{1}{2}|a-b|$

or, $SD = \frac{1}{2} \times range$. (proved)