# 3

# Measures of Central Tendency

## 3.1. Introduction

Usually the collected data is not suitable to draw conclusions about the mass from which it has been taken. Even though the data will be some what summarized after it is depicted using frequency distributions and presented by using graphs and diagrams, still we cannot make any inferences about the data since we have many groups. Hence, organizing a data into a frequency is not sufficient, there is a need for further condensation, particularly when we want to compare two or more distributions we may reduce the entire distribution into one number that represents the distribution we need. A single value which can be considered as a typical or representative of a set of observations and around which the observations can be considered as centered is called an average (or average value or center of location). Since such typical values tend to lie centrally within a set of observations when arranged according to magnitudes; averages are called *measures of central tendency (MCT).*

## 3.2. Objectives of MCT

- *To condense a mass of data in to one single value.* That is to get a single value which is best representative of the data (that describes the characteristics of the entire data). Measures of central tendency, by condensing masses of in to one single value enable us to get an idea of the entire data. Thus one value can represent thousands of data even more.
- *To facilitate comparison.* Statistical devices like averages, percentages and ratios used for this purpose. Measures of central tendency, by condensing masses of in to one single value, facilitates comparison. For instance, to compare two classes A and B, instead

of comparing each student result, which is practically infeasible, we can compare the average mark of the two classes.

## 3.3. Desirable Properties of Good MCT

A measure of central tendency is good or satisfactory if it possesses the following characteristics.

1. It should be calculated based on all observations.
2. It should not be affected by extreme values.
3. It should be defined rigidly which means it should have a definite value.
4. It should always exist.
5. It should be easy to understand and calculate. It should not be subject to complicated and tedious calculations, though the advent of electronic calculators and computers has made it possible.
6. It should be capable of further algebraic treatment. By algebraic treatment, we mean that the measures should be used further in the formulation of other formulae or it should be used for further statistical analysis.

## 3.4. Summation Notation

Suppose we have variable $x$ having successive values $x_1, x_2, ..., x_n$. The sum of these values can be written as $x_1 + x_2 + ... + x_n$. This can be written as using Greek letter $\sum$ as

$$x_1 + x_2 + ... + x_n = \sum_{i=1}^{n} x_i$$

By $\sum$ notation we can write

- $x_1^2 + x_2^2 + ... + x_n^2 = \sum_{i=1}^{n} x_i^2$
- $x_1y_1 + x_2y_2 + ... + x_ny_n = \sum_{i=1}^{n} x_iy_i$
- $(x_1 + x_2 + ... + x_n)^2 = (\sum_{i=1}^{n} x_i)^2$
- $\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4} = \sum_{i=1}^{4} \frac{1}{x_i}$

**Rules of Summation**

1. For two variables $x$ and $y$ we have

$$\sum_{i=1}^{n}(x_i \pm y_i) = \sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} y_i$$

2. If $k$ is constant number, we have

$$\sum_{i=1}^{n} kx_i = k\sum_{i=1} x_i$$

3. For constant number $k$, we have

$$\sum_{i=1}^{n} k = nk$$

4. $\sum_{i=1}^{n}(x_i - k)^2 = \sum_{i=1}^{n} x_i^2 - 2k\sum_{i=1}^{n} x_i + nk^2$

From now onwards we will use $\sum x_i$ in place of $\sum_{i=1}^{n} x_i$ just for simplicity.

## 3.5. Types of Measures of Central Tendency

There are many types of measures of central tendency, each possessing particular properties and each being typical in some unique way. The most frequently encountered ones are

- ▷ Mean (computed average)
  - Arithmetic mean (simple arithmetic mean, weighted arithmetic mean and combined mean)
  - Geometric mean
  - Harmonic mean
- ▷ Mode (the most frequented value)
- ▷ Positional averages
  - Median
  - Quantiles (quartiles, deciles and percentiles)

## 3.6. Mean

### 3.6.1. Arithmetic Mean (AM)

**Simple Arithmetic Mean**

1. Suppose a variable $x$ has observed values $x_1, x_2, ..., x_n$. The simple arithmetic mean denoted by $\bar{x}$ (for sample) and $\mu$ (for population) is the sum of these observations divided by the total number of observations. Symbolically,

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\mu = \frac{x_1 + x_2 + ... + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

   Simple AM is the most commonly used average.

2. Suppose the values $x_1, x_2, ..., x_n$ are accompanied by frequencies $f_1, f_2, ..., f_n$ respectively, then the simple AM is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + ... + f_nx_n}{f_1 + f_2 + ... + f_n} = \frac{\sum f_ix_i}{\sum f_i}$$

3. For data in grouped frequency distribution we use the class mark instead of each observed value and simple AM is given by

$$\bar{x} = \frac{f_1m_1 + f_2m_2 + ... + f_nm_n}{f_1 + f_2 + ... + f_n} = \frac{\sum f_im_i}{\sum f_i}$$

   where $m_i$ is the class mark of the $i^{th}$ class.

**Example 1**: The heights of 7 students selected from a class are given below in centimeter. $165, 160, 172, 168, 159, 170, 173$. Calculate the simple AM of heights.

$$\bar{x} = \frac{x_1 + x_2 + ... + x_7}{7} = \frac{\sum_{i=1}^{7} x_i}{7} = \frac{1167}{7} = 166.5\, cm$$

**Example 2**: The following is the frequency distribution of marks in Stat 1011 of 46 students (out of 20). Find the mean mark of this class.

| Mark ($x_i$) | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| No of students ($f_i$) | 1 | 2 | 3 | 6 | 10 | 11 | 7 | 3 | 2 | 1 | 46 |
| $f_ix_i$ | 9 | 20 | 33 | 72 | 130 | 154 | 105 | 48 | 34 | 18 | 623 |

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + ... + f_nx_n}{f_1 + f_2 + ... + f_n} = \frac{\sum f_ix_i}{\sum f_i} = \frac{623}{46} = 13.54$$

**Example 3**: Calculate the mean amount of yield of maize from the grouped frequency distribution given below.

| Yield (in kg) | No of plots ($f_i$) | Class mark ($m_i$) | $f_im_i$ |
|---|---|---|---|
| 171-179 | 3 | 175 | 525 |
| 180-188 | 7 | 184 | 1288 |
| 189-197 | 12 | 193 | 2316 |
| 198-206 | 9 | 202 | 1818 |
| 207-215 | 4 | 211 | 844 |
| 216-224 | 4 | 220 | 880 |
| 225-233 | 1 | 229 | 229 |
| Total | 40 | | 7900 |

$$\bar{x} = \frac{f_1m_1 + f_2m_2 + ... + f_nm_n}{f_1 + f_2 + ... + f_n} = \frac{\sum f_im_i}{\sum f_i} = \frac{7900}{40} = 197.5\, kg\, per\, plot$$

**Weighted Arithmetic Mean**

It is an arithmetic mean used when all observations in data have unequal relative importance (technically termed as weight). Suppose $x_1, x_2, ..., x_n$ have weights $w_1, w_2, ..., w_n$ respectively, then weighted arithmetic mean ($\bar{x}_w$) is given by

$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + ... + w_nx_n}{w_1 + w_2 + ... + w_n} = \frac{\sum w_ix_i}{\sum w_i}$$

**Example**: Semester grade point average (GPA) of a student is a good example of weighted arithmetic mean.

| Course | Weights (Credit hours) | Grade ($x$) |
|---|---|---|
| Stat 281 | 4 | B = 3 |
| Math 261 | 4 | B = 3 |
| Math 224 | 3 | C = 2 |
| Phil 201 | 3 | B = 3 |
| Comp 201 | 3 | C = 2 |

Calculate the GPA of this student?

$$GPA = \bar{x}_w = \frac{w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5}{w_1 + w_2 + w_3 + w_4 + w_5} = \frac{\sum w_ix_i}{\sum w_i} = \frac{45}{17} = 2.64$$

**Combined Mean**

If there are $k$ different groups (having the same unit of measurement) with mean $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ and number of observations $n_1, n_2, ..., n_k$ respectively, then the mean of all the groups i.e. the combined mean is given by

$$\bar{\bar{x}} = \bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + ... + n_k\bar{x}_k}{n_1 + n_2 + ... + n_k} = \frac{\sum n_i\bar{x}_i}{\sum n_i}$$

**Example**: There are 49 students in a certain department. Among these 7 are seniors with average weight of 165 lbs, 9 are juniors with average weight of 160 lbs, 13 are sophomores with average weight of 152 lbs and 20 freshman with average weight of 150 lbs. Find the average weight of students in the department.

$$\begin{aligned}\bar{\bar{x}} &= \frac{n_s\bar{x}_s + n_j\bar{x}_j + n_{so}\bar{x}_{so} + n_f\bar{x}_f}{n_s + n_j + n_{so} + n_f} \\ &= \frac{7 \times 165 + 9 \times 130 + 3 \times 152 + 20 \times 150}{7 + 9 + 13 + 20} \\ &= 93.28\, lbs\end{aligned}$$

**Properties of Arithmetic Mean**

- ▷ If a constant $k$ is added or subtracted from each value in a distribution, then the new mean will be

$$\bar{x}_{new} = \bar{x}_{old} \pm k$$

- ▷ If each value of a distribution is multiplied by a constant $k$, the new mean will be the original mean multiplied by $k$. That is,

$$\bar{x}_{new} = k\bar{x}_{old}$$

- ▷ Arithmetic mean can be calculated for any set of data (quantitative data), and it will be unique.We cannot calculate AM for open-ended grouped frequency distribution.
- ▷ It is highly affected by extreme values.
- ▷ It lends itself for further statistical analysis. For example, as combined mean.
- ▷ The algebraic sum of the deviations of each value from the arithmetic mean is zero. That is

$$\sum(x_i - \bar{x}) = 0$$

**Example 1**: The mean age of a group of 100 students was found to be 32.02 years. Later it was discovered that age of 57 was misread as 27. Find the correct mean.

**Solution**:

Let $\bar{x}_{cor}$ and $\bar{x}_{wr}$ are the correct and wrong means respectively. Thus, from the given problem $\bar{x}_{wr} = 32.02, n = 100, x_{wr} = 27$ and $x_{cor} = 57$.

$$\bar{x}_{wr} = \frac{(\sum x_i)_{wr}}{n}$$

$$(\sum x_i)_{wr} = \bar{x}_{wr} \times n$$

$$(\sum x_i)_{wr} = 32.02 \times 100 = 3202$$

$$(\sum x_i)_{cor} = (\sum x_i)_{wr} + x_{cor} - x_{wr}$$

$$(\sum x_i)_{cor} = 3202 + 57 - 27 = 3232$$

$$\bar{x}_{cor} = \frac{(\sum x_i)_{cor}}{n}$$

$$\bar{x}_{cor} = \frac{3232}{100} = 32.32 year$$

**Example 2**: The mean weight of 150 students in certain class is 60 kg. The mean weight of boys in the class is 70 kg and that of the girls is 55 kg. Find the number of boys and girls in the class.

**Solution**:

Let $n_b$ and $n_g$ are number of boys and girls in the class respectively. Further, suppose $\bar{\bar{x}} = 60kg$, $\bar{x}_b = 70kg$ and $\bar{x}_g = 55kg$ are the mean weight of both, boys and girls respectively.

$$n_b + n_g = 150 \tag{3.1}$$

Using combined mean formula

$$\bar{\bar{x}} = \frac{n_b\bar{x}_b + n_g\bar{x}_g}{n_b + n_g} = 60 = \frac{70n_b + 55n_g}{n_b + n_g}$$

$$n_g = 2n_b \tag{3.2}$$

Inserting equation (3.2) in equation (3.1) we obtain $n_b = 50$ and $n_g = 100$.

### 3.6.2. Geometric Mean (GM)

The geometric mean of $n$-positive numbers is the $n^{th}$ root of their product. The geometric mean of $x_1, x_2, ..., x_n$ is given by the following for raw data, ungrouped and grouped frequency respectively.

$$GM = \sqrt[n]{x_1 \times x_2 \times ... \times x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

$$GM = \sqrt[n]{x_1^{f_1} \times x_2^{f_2} \times ... \times x_n^{f_n}} = \sqrt[n]{\prod_{i=1}^{n} x_i^{f_i}}$$

$$GM = \sqrt[n]{m_1^{f_1} \times m_2^{f_2} \times ... \times m_n^{f_n}} = \sqrt[n]{\prod_{i=1}^{n} m_i^{f_i}}$$

We can also use logarithms to calculate $GM$

$$GM = \sqrt[n]{x_1 \times x_2 \times ... \times x_n} = (x_1 \times x_2 \times ... \times x_n)^{1/n}$$

$$\log GM = \frac{1}{n} \log(x_1 \times x_2 \times ... \times x_n)$$

$$\log GM = \frac{1}{n}(\log x_1 + \log x_2 + ... + \log x_n)$$

Taking antilog of both sides we get that

$$GM = anti\log\{\frac{1}{n} \log(x_1 + \log x_2 + ... + \log x_n)\} = anti\log(\frac{1}{n} \sum \log x_i)$$

If the variable values are measured as ratios, proportions or percentage and some values are larger in magnitude and others are small, then the geometric mean is a better representative of the data than the simple average. In a "geometric series", the most meaning full average is the geometric mean. The arithmetic mean is very biased toward the large numbers in the series. The main disadvantage of geometric mean is that it cannot be calculated if one or more observations are zero or negative. It is also affected by extreme values but not to the extent of $AM$.

**Examples**

1. A given epidemic was spreading at the rate of 1.5 and 2.67 in two successive days. What is its average spread rate?

   **Solution**:

$$GM = \sqrt{x_1 \times x_2} = \sqrt{1.5 \times 2.67} = \sqrt{4.005} = 2.001$$

2. The price of a commodity increased by 5% from 1989 to 1990, 8% from 1990 to 1991 and by 77% from 1991 to 1992. Find the average price increase.

   **Solution**:

   For increment, take the base line value as 100% and then add the % increase so as to get the values in successive years.

| Year | % increase | Value ($x_i$) | $\log x_i$ |
|---|---|---|---|
| 1989-1990 | 5 | 105 | 2.02 |
| 1990-1991 | 8 | 108 | 2.03 |
| 1991-1992 | 77 | 177 | 2.25 |
| Total | | | $\sum \log x_i = 6.30$ |

   Then,

$$GM = anti\log(\frac{1}{n}\sum \log x_i) = anti\log(\frac{1}{3} \times 6.30) = anti\log(2.1) = 125.89$$

   Therefore, the price increment is 25.89%.

3. A machine depreciated by 10% each in the first two years and by 40% in the third year. Find out the average rate of depreciation.

   **Solution**:

   Like the previous one, take the base line value of the machine as 100% and then deduct the % of depreciation so as to get the depreciated values in successive years.

| Year | % depreciation | Value ($x_i$) | $\log x_i$ |
|---|---|---|---|
| 1 | 10 | 90 | 1.95 |
| 2 | 10 | 90 | 1.95 |
| 3 | 40 | 60 | 1.79 |
| Total | | | $\sum \log x_i = 5.69$ |

   Then,

$$GM = anti\log(\frac{1}{n}\sum \log x_i) = anti\log(\frac{1}{3} \times 5.69) = anti\log(1.70) = 50.12$$

   Therefore, the machine depreciated by is 49.88%.

### 3.6.3. Harmonic Mean (HM)

Harmonic mean is another specialized average which is useful in averaging variables expressed as rate per unit of time such as speed, number of units produced per day. Simple harmonic

mean is the reciprocal of the arithmetic mean of the numbers.

$$HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_i}}$$

- The simple $HM$ is preferably used to calculate average speed for fixed distance, average price for fixed total cost, average time for fixed total distance.

For ungrouped frequency distribution,

$$HM = \frac{f_1 + f_2 + ... + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + ... + \frac{f_n}{x_n}} = \frac{\sum f_i}{\sum \frac{f_i}{x_i}}$$

For grouped frequency distribution,

$$HM = \frac{f_1 + f_2 + ... + f_n}{\frac{f_1}{m_1} + \frac{f_2}{m_2} + ... + \frac{f_n}{m_n}} = \frac{\sum f_i}{\sum \frac{f_i}{m_i}}$$

The weighted $HM$ of $n$ non-zero observations $x_1, x_2, ..., x_n$ having weights $w_1, w_2, ..., w_n$ respectively is given by

$$HM_w = \frac{w_1 + w_2 + ... + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + ... + \frac{w_n}{x_n}} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}}$$

- The weighted $HM$ is used to compute mean speed to cover differing distances, mean prices when the total cost is not fixed, etc.

**Examples**

1. A driver travels for 3 days at speed of 48 km/hr for about 10 hrs, 40 km/hr for 12 hrs, 32 km/hr for 15 hrs respectively. What is the average speed of the driver in 3 days?

   **Solution**:

   Using $d_i = s_i \times t_i$; $i = 1, 2, 3$ the distance covered in three days is fixed, which is $480km$. So simple $HM$ is appropriate to compute the average speed.

   $$\begin{aligned} HM &= \frac{3}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}} = \frac{3}{\sum \frac{1}{x_i}} \\ &= \frac{3}{\frac{1}{48} + \frac{1}{40} + \frac{1}{32}} = \frac{3}{0.0771} \\ &= 38.91 km/hr \end{aligned}$$

2. A driver travelled for 3 days on first days he derived for 10 hrs at speed of 48 km/hr, on the second day for 12 hrs at 45 km/hr, on third day for 15 hrs at 40 km/hr. What is the average speed?

**Solution**:

Using $d_i = s_i \times t_i$; $i = 1, 2, 3$ the distance covered in each day is not fixed, which is $480km, 540km$ and $600km$ respectively. So weighted $HM$ is appropriate to compute the average speed.

$$\begin{aligned} HM_w &= \frac{w_1 + w_2 + w_3}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \frac{w_3}{x_3}} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} \\ &= \frac{10 + 12 + 15}{\frac{10}{48} + \frac{12}{45} + \frac{15}{40}} = \frac{37}{0.892} \\ &= 41.48km/hr \end{aligned}$$

**Some Empirical Relationship among AM, GM and HM**

▷ The GM of two numbers $x_1$ and $x_2$ is equal to the GM of their AM and HM. That is,

$$GM = \sqrt{x_1 \times x_2} = \sqrt{AM \times HM}$$

▷ For $n$ positive numbers $HM \leq GM \leq AM$.

## 3.7. Mode

The mode (modal value) of data set is the value that occurs *most frequently*. When two values occur with the same greatest frequency, each one is a mode and the data set is *bimodal*. When more than two values occur with the greatest frequency, each is a mode and the data set is said to be *multimodal*. When no value is repeated or values are equally repeated, we say that there is *no mode*.

**Example 1**: Find the modes of the following data sets.

- 5 5 5 3 1 5 1 4 3 5
- 1 2 2 2 3 4 5 6 6 6 7 9
- 1 2 3 6 7 8 9 10

In a frequency distribution, the mode is located in the class with highest frequency and that class is the modal class. Then the formula for mode is

$$\hat{x} = L_{\hat{x}} + \left[\frac{f_{\hat{x}} - f_{\hat{x}-1}}{(f_{\hat{x}} - f_{\hat{x}-1}) + (f_{\hat{x}} - f_{\hat{x}+1})}\right] w$$

where

$L_{\hat{x}}$ is the lower class boundary of the modal class,

$f_{\hat{x}}$ is the frequency of modal class,

$f_{\hat{x}-1}$ is the frequency of the class which precedes the modal class,

$f_{\hat{x}+1}$ is the frequency of the class which is successor of the modal class and

$w$ is the class width of the modal class.

**Example**: Use the frequency distribution of heights in the following table to find the mode of height of the 100 male students at XYZ university and interpret the result.

| Height ($in$) | Frequency ($f_i$) |
|---|---|
| 59.5-62.5 | 5 |
| 62.5-65.5 | 18 |
| 65.5-68.5 | 42 |
| 68.5-71.5 | 27 |
| 71.5-74.5 | 8 |

**Solution**:

A class having the highest frequency is considered as a modal class. Thus the $3^{rd}$ class (65.5-68.5) is the modal class.

$$\begin{aligned}
\hat{x} &= L_{\hat{x}} + \left[\frac{f_{\hat{x}} - f_{\hat{x}-1}}{(f_{\hat{x}} - f_{\hat{x}-1}) + (f_{\hat{x}} - f_{\hat{x}+1})}\right] w \\
&= 65.5 + \left[\frac{42 - 18}{(42 - 18) + (42 - 27)}\right] \times 3 \\
&= 65.5 + \left[\frac{24}{39}\right] \times 3 \\
&= 65.5 + 1.846 \\
&= 67.346
\end{aligned}$$

Mode is not affected by extreme values and can be calculated for open-ended classes. But it often does not exist and is value may not be unique. In such case mode is ill-defined.

**Properties of Mode**

1. It is simple to calculate and easy to determine.
2. It is not based on all observations.

3. The mode can be used for both qualitative (such as religious preference, gender, political affiliation, etc) and quantitative data types.

## 3.8. Median

A median is a value which divides set of data in to two equal parts such that the number of observations below it is the same as the number of observations above it. It is the middle value when the values are arranged in order of increasing (or decreasing) magnitude. To find the median, first sort the values (arrange them in order), then use one of the following procedures.

1. If the number of values is *odd*, the median is the number that is located in the exact middle of the list.

$$\tilde{x} = \left(\frac{n+1}{2}\right)^{th} value$$

**Example**: What is the median of 180, 201, 220, 191, 219, 209 and 220.

**Solution**:

First we should have to sort the data: 180, 191, 201, 209, 219, 220, 220. Since $n = 7$ is odd

$$\tilde{x} = \left(\frac{4+1}{2}\right)^{th} value = 4^{th} value = 209$$

2. If the number of values is *even*, the median is found by computing the mean of the two middle numbers.

$$\tilde{x} = \frac{\left(\frac{n}{2}\right)^{th} value + \left(\frac{n}{2}+1\right)^{th} value}{2}$$

**Example**: What is the median of 62, 63, 64, 65, 66, 66, 68 and 78.

**Solution**:

First we should have to sort the data: $62, 63, 64, \underbrace{65, 66}, 66, 68, 78$. Since $n = 8$ is even

$$\begin{aligned}\tilde{x} &= \frac{\left(\frac{n}{2}\right)^{th} value + \left(\frac{n}{2}+1\right)^{th} value}{2} \\ &= \frac{4^{th} value + 5^{th} value}{2} \\ &= \frac{65+66}{2} = 65.5\end{aligned}$$

3. For grouped frequency distributions median is given by the formula

$$\tilde{x} = L_{\tilde{x}} + \left( \frac{\frac{n}{2} - F_{\tilde{x}-1}}{f_{\tilde{x}}} \right) w$$

where

$L_{\tilde{x}}$ is the lower class boundary of the median class,
$F_{\tilde{x}-1}$ is the less than cumulative frequency just before the median class,
$w$ is the class width of the median class,
$f_{\tilde{x}}$ is the frequency of the median class and $n = \sum f_i$.

■ The median class is the class which include $\left(\frac{n}{2}\right)^{th}$ *value*.

**Example**: The following table shows a frequency distribution of grades on a final examination in college algebra for 120 students. Then, obtain median and interpret the results.

| Grade | No of students |
|---|---|
| 30-39 | 1 |
| 40-49 | 3 |
| 50-59 | 11 |
| 60-69 | 21 |
| 70-79 | 43 |
| 80-89 | 32 |
| 90-99 | 9 |

**Solution**:

First we should do the following.

| Class limits | Class boundaries | Frequency | LCF |
|---|---|---|---|
| 30-39 | 29.5-39.5 | 1 | 1 |
| 40-49 | 39.5-49.5 | 3 | 4 |
| 50-59 | 49.5-59.5 | 11 | 15 |
| 60-69 | 59.5-69.5 | 21 | 37 |
| 70-79 | 69.5-79.5 | 43 | 80 |
| 80-89 | 79.5-89.5 | 32 | 112 |
| 90-99 | 89.5-99.5 | 9 | 120 |

The class which includes $\left(\frac{n}{2}\right)^{th} value = 60^{th} value$ is considered as the median class. Hence, the $5^{th}$ class is the median class.

$$\begin{aligned}\tilde{x} &= L_{\tilde{x}} + \left(\frac{\frac{n}{2} - F_{\tilde{x}-1}}{f_{\tilde{x}}}\right) w \\ &= 69.5 + \left(\frac{\frac{120}{2} - 37}{43}\right) \times 10 \\ &= 74.849\end{aligned}$$

Therefore, out of 120 students 60 of them scored less than 74.849 and 60 of them scored greater than 74.849 on college algebra examination.

**Properties of the Median**

1. It is an average of location, not the average of the values in the data set.
2. It is more affected by the number of observations than the extreme values.
3. Median can be calculated even in the case open-ended interval.

## 3.9. Quantiles

The median gives us a value which divides the data set in to two equal parts. There are also *other positional measures* that divide a given data set into more than two equal parts. These measures are collectively known as *quantiles*. Quantiles include quartiles, deciles and percentiles.

**Quartiles** are some three points that divide the array in to four parts in away each portion contains equal number of observations. The first, second and third points are called the first ($Q_1$), second ($Q_2$) and third ($Q_3$) quartiles respectively. 25% of the data fall below $Q_1$, 50% below $Q_2$ and 75% below $Q_3$ and

$$Q_1 \leq Q_2 \leq Q_3$$

**Deciles** are nine points that divide the array in to ten equal parts.The first, second, ..., ninth deciles are denoted by $D_1, D_2, ..., D_9$ respectively. 10% of the data fall below $D_1$, 20% below $D_2$, ..., 90% below $D_9$ and

$$D_1 \leq D_2 \leq \ldots \leq D_9$$

**Percentiles** are ninety nine points that divide the array in to 100 equal parts. They are denoted by $P_1, P_2, ..., P_{99}$. Always

$$P_1 \leq P_2 \leq \ldots \leq P_{99}$$

**Methods of Finding Quantiles**

1. For raw data and data in ungrouped frequency distribution. After arranging data in ascending order, we apply the following formula.

$$Q_i = \left(\frac{i(n+1)}{4}\right)^{th} value, i = 1, 2, 3$$

$$D_i = \left(\frac{i(n+1)}{10}\right)^{th} value, i = 1, 2, ..., 9$$

$$P_i = \left(\frac{i(n+1)}{100}\right)^{th} value, i = 1, ..., 99$$

**Example**: Given the data 420, 430, 435, 438, 441, 449, 490, 500, 510 and 515. Find

(a) all quartiles.

$$\begin{aligned} Q_1 &= \left(\frac{1 \times (10+1)}{4}\right)^{th} value = 2.75^{th} value \\ &= 2^{nd} value + 0.75(3^{rd} value - 2^{nd} value) \\ &= 430 + 0.75(435 - 430) \\ &= 433.75 \end{aligned}$$

$$\begin{aligned} Q_2 &= \left(\frac{2 \times (10+1)}{4}\right)^{th} value = 5.5^{th} value \\ &= 5^{th} value + 0.5(6^{th} value - 5^{th} value) \\ &= 441 + 0.5(449 - 441) \\ &= 445 \end{aligned}$$

$$\begin{aligned} Q_3 &= \left(\frac{3 \times (10+1)}{4}\right)^{th} value = 8.25^{th} value \\ &= 8^{th} value + 0.25(9^{th} value - 8^{th} value) \\ &= 500 + 0.25(510 - 500) \\ &= 502.5 \end{aligned}$$

(b) the $1^{st}$ and $7^{th}$ deciles.

$$\begin{aligned} D_1 &= \left(\frac{1\times(10+1)}{10}\right)^{th} value = 1.1^{th} value \\ &= 1^{st}\, value + 0.1(2^{nd} value - 1^{st} value) \\ &= 420 + 0.1(430-420) \\ &= 421 \end{aligned}$$

$$\begin{aligned} D_7 &= \left(\frac{7\times(10+1)}{10}\right)^{th} value = 7.7^{th} value \\ &= 7^{th} value + 0.7(8^{th} value - 7^{th} value) \\ &= 490 + 0.7(500-490) \\ &= 497 \end{aligned}$$

(c) the $40^{th}$ and $75^{th}$ percentiles.

$$\begin{aligned} P_{40} &= \left(\frac{40\times(10+1)}{100}\right)^{th} value = 4.4^{th} value \\ &= 4^{th} value + 0.4(5^{th} value - 4^{th} value) \\ &= 438 + 0.4(441-438) \\ &= 439.2 \end{aligned}$$

$$\begin{aligned} P_{75} &= \left(\frac{75\times(10+1)}{100}\right)^{th} value = 8.25^{th} value \\ &= 8^{th} value + 0.25(9^{th} value - 8^{th} value) \\ &= 500 + 0.25(510-500) \\ &= 502.5 \end{aligned}$$

2. For data in grouped frequency distribution.

$$Q_i = L_{q_i} + \frac{(\frac{in}{4} - F_{q_{i-1}})}{f_{q_i}} w$$

$$D_i = L_{d_i} + \frac{(\frac{in}{10} - F_{d_{i-1}})}{f_{d_i}} w$$

$$P_i = L_{p_i} + \frac{(\frac{in}{100} - F_{p_{i-1}})}{f_{p_i}} w$$

where

$L_{q_i}, L_{d_i}, L_{p_i}$ are the lower class boundaries of the classes containing the concerned quantile points,

$F_{q_{i-1}}, F_{d_{i-1}}, F_{p_{i-1}}$ are the LCF of the class which precedes the class containing the concerned quantile points,

$f_{q_i}, f_{d_i}, f_{p_i}$ are frequencies of classes containing the concerned quantile points and

$w$ is the class width of a class containing the concerned quantile point.

**Note**

- $Q_i$ is found in the class containing the $\left(\frac{in}{4}\right)^{th}$ observation.
- $D_i$ is found in the class containing the $\left(\frac{in}{10}\right)^{th}$ observation.
- $P_i$ is found in the class containing the $\left(\frac{in}{100}\right)^{th}$ observation.

**Example**: Calculate all quartiles, the $5^{th}$ and $8^{th}$ deciles, and the $30^{th}$ and $80^{th}$ percentiles for the students score data and interpret the results.

| Class boundaries | Frequency ($f_i$) | LCF |
|---|---|---|
| 10.5-14.5 | 4 | 4 |
| 14.5-18.5 | 7 | 11 |
| 18.5-22.5 | 8 | 19 |
| 22.5-26.5 | 10 | 29 |
| 26.5-30.5 | 12 | 41 |
| 30.5-34.5 | 7 | 48 |
| 34.5-38.5 | 8 | 56 |

**Solution**:

$Q_1$ is found in the $3^{rd}$ class (18.5-22.5) because this class include the $\left(\frac{1\times 56}{4}\right)^{th} = 14^{th} value$

$$\begin{aligned} Q_1 &= L_{q_1} + \frac{(\frac{1\times 56}{4} - F_{q_0})}{f_{q_1}} \times 4 \\ &= 18.5 + \frac{(\frac{1\times 56}{4} - 11)}{8} \times 4 \\ &= 18.5 + 1.5 = 20 \end{aligned}$$

$Q_2$ is found in the $4^{th}$ class (22.5-26.5) because this class include the $\left(\frac{2\times 56}{4}\right)^{th} = 28^{th} value$

$$\begin{aligned} Q_2 &= L_{q_2} + \frac{(\frac{2\times 56}{4} - F_{q_1})}{f_{q_2}} \times 4 \\ &= 22.5 + \frac{(\frac{2\times 56}{4} - 19)}{10} \times 4 \\ &= 22.5 + 3.6 = 26.1 \end{aligned}$$

$Q_3$ is found in the $6^{th}$ class (30.5-34.5) because this class include the $\left(\frac{3\times 56}{4}\right)^{th} = 42^{th} value$

$$\begin{aligned} Q_3 &= L_{q_3} + \frac{(\frac{3\times 56}{4} - F_{q_2})}{f_{q_3}} \times 4 \\ &= 30.5 + \frac{(\frac{3\times 56}{4} - 41)}{7} \times 4 \\ &= 30.5 + 0.57 = 31.07 \end{aligned}$$

$D_5$ is found in the $4^{th}$ class (22.5-26.5) because this class include the $\left(\frac{5\times 56}{10}\right)^{th} = 28^{th} value$

$$\begin{aligned} D_5 &= L_{d_5} + \frac{(\frac{5\times 56}{10} - F_{d_4})}{f_{d_5}} \times 4 \\ &= 22.5 + \frac{(\frac{2\times 56}{4} - 19)}{10} \times 4 \\ &= 22.5 + 3.6 = 26.1 \end{aligned}$$

$D_8$ is found in the $6^{th}$ class (30.5-34.5) because this class include the $\left(\frac{8\times 56}{10}\right)^{th} = 44.8^{th} value$

$$\begin{aligned} D_8 &= L_{d_8} + \frac{(\frac{8\times 56}{10} - F_{d_7})}{f_{d_8}} \times 4 \\ &= 30.5 + \frac{(\frac{8\times 56}{10} - 41)}{7} \times 4 \\ &= 30.5 + 2.17 = 32.67 \end{aligned}$$

$P_{30}$ is found in the $3^{rd}$ class (18.5-22.5) because this class include the $\left(\frac{30\times 56}{100}\right)^{th} = 16.8^{th} value$

$$\begin{aligned} P_{30} &= L_{p_{30}} + \frac{(\frac{30\times 56}{100} - F_{p_{29}})}{f_{p_{30}}} \times 4 \\ &= 18.5 + \frac{(\frac{30\times 56}{100} - 11)}{19} \times 4 \\ &= 18.5 + 1.22 = 19.72 \end{aligned}$$

$P_{90}$ is found in the $7^{th}$ class (34.5-38.5) because this class include the $\left(\frac{90\times 56}{100}\right)^{th} = 50.4^{th} value$

$$\begin{aligned} P_{90} &= L_{p_{90}} + \frac{(\frac{90\times 56}{100} - F_{p_{89}})}{f_{p_{90}}} \times 4 \\ &= 34.5 + \frac{(\frac{90\times 56}{100} - 48)}{8} \times 4 \\ &= 34.5 + 1.2 = 35.7 \end{aligned}$$

## 3.10. Exercises

1. Define and compare the characteristics of the mean, the median and the mode.

2. Your statistics instructor tells you on the first day of class that there will be five tests during the term. From the scores on theses tests for each student he will compute a measures of central tendency that will serve as the student's final course grade. Before taking the first test you must choose whether you want your final grade to be the mean or the median of the five test scores. Which would you choose? Why? Justify your answer.

3. A student's final grades in mathematics, physics, chemistry and sport are, respectively, 82, 86, 90, and 70. If the respective credits received for these courses are 3, 5, 3, and 2, determine an appropriate average grade.

4. A large department store collects data on sales made by each of its sales people. The number of sales made on a given day by each of 20 sales people is shown below.

   9 6 12 10 13 15 16 14 14 16 17 16 24 21 22 18 19 18 20 17

   Then, find $Q_3$, $D_8$, $P_{80}$ and $P_{90}$ and interpret all results.

5. In a certain investigation, 460 persons were involved in the study, and based on an enquiry on their age, it was known that 75% of them were 22 or more. The following frequency distribution shows the age composition of the persons under study.

| Mid age in years | 13 | 18 | 23 | 28 | 33 | 38 | 43 | 48 |
|---|---|---|---|---|---|---|---|---|
| Number of persons | 24 | $f_1$ | 90 | 122 | $f_2$ | 56 | 20 | 33 |

   (a) Find the median and modal life of condensers and interpret them.

   (b) Find the values of all quartiles.

   (c) Compute the $5^{th}$ decile, $25^{th}$ percentile, $50^{th}$ percentile and the $75^{th}$ percentile and interpret the results.

6. Given the following frequency distribution,

| Mid price of a commodity | 15 | 25 | 35 | 45 | 55 |
|---|---|---|---|---|---|
| Number of items sold | 27 | A | 28 | B | 19 |

(a) If 75% of the items were sold in birr 45 or less and most items were sold in birr 34, find the missing frequencies.

(b) If 25% of the items were sold in greater than or equal to birr 45 and most items were sold in birr 34, find the missing frequencies.