3. *Descriptive Measure*

Is a number or characteristic that is used to summarize the data set .

- Statistic : is a characteristic or measure obtained by using the data values from a sample.
- Parameter : is a characteristic or measure obtained by using all the data values from a specific population.

3.1 *Measures Central Tendency*

In statistics, a **central tendency** (or measure of central tendency) is a central or typical value for a probability distribution. It may also be called a centre or location of the distribution.

From common measures of central tendency that represent either a typical or representative score and / or a value when the data tends to the centre.

1) Mean: is the arithmetic average.
2) Median: is the midpoint or middle of data.
3) Mode: is the most frequently occurring score.

3.1.1 Mean

- **Mean of ungrouped data**: The arithmetic average (termed the mean is abbreviated as $\bar{x}$) represents the most appropriate measure of central tendency for continuous type data. It is obtained by adding all of the values (scores) and dividing this sum by the number of scores.

$$\textbf{Mean for population data}: \textbf{Mean } (\boldsymbol{\mu}) = \frac{\sum x}{N}$$

$$\textbf{Mean for sample data}: \textbf{Mean } (\bar{\boldsymbol{x}}) = \frac{\sum x}{n}$$

Where the mean can be denoted by $\bar{\boldsymbol{x}}$ (x-bar) for samples and it is denoted by $\boldsymbol{\mu}$ when we dealing with populations. $\sum x$denotes the summation of the set of values, $\boldsymbol{x}$ represents the individual raw scores and ($\boldsymbol{n}$ and $\boldsymbol{N}$) equals the number of scores.

Example: Compute the mean of the following scores that placed in ascending order:

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10$$

Solution: First, we sum up the scores of the given data as:

$$\sum \boldsymbol{x} = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 55$$

Second, we divide the summation by the total of sores or data point as:

$$\textbf{Mean } (\bar{\boldsymbol{x}}) = \frac{\sum \boldsymbol{x}}{\boldsymbol{n}} = \frac{55}{10} = 5.5$$

- **Mean of grouped data**: To calculate the mean for grouped data, first we find the midpoint of each class and then multiply the midpoints by the frequencies of the corresponding classes. The sum of these products which denoted by

($\sum frequency * midpoint$) gives an approximation for the sum of all values. To find the value of the mean, we divide this sum by the total number of observations in the data.

Calculating mean for grouped data are

$$\textbf{Mean for population data } \mu = \frac{\sum f_i * m_i}{N}$$

$$\textbf{Mean for sample data } \bar{x} = \frac{\sum f_i * m_i}{n}$$

Example: *The amount of protein (in grams) for a variety of fast-food sandwiches is reported here. Construct a frequency distribution using 6 classes:*

Class Limits	*Frequency* f_i
12 – 19	*7*
20 – 27	*17*
28 – 35	*10*
36 – 43	*4*
44 – 51	*1*
52 – 59	*1*

Solution

Class Limits	*Frequency* f_i	*Midpoint* m_i	$f_i * m_i$
12 – 19	*7*	*15.5*	*108.5*
20 – 27	*17*	*23.5*	*399.5*
28 – 35	*10*	*31.5*	*315*
36 – 43	*4*	*39.5*	*158*
44 – 51	*1*	*47.5*	*47.5*
52 – 59	*1*	*55.5*	*55.5*
Total	$\sum f_i = 40 = n$		$\sum f_i * m_i = 1084$

$$\bar{\mathbf{x}} = \frac{\sum \mathrm{f_i} * \mathrm{m_i}}{\mathrm{n}} = \frac{1084}{40} = 27.1 gram$$

Example: Compute the mean of 80 student are distributed according to their weights as the following table.

The Weights/kg Classes	Frequency f_i	Midpoint m_i	$f_i * m_i$
31-40	1	35.5	35.5
41-50	2	45.5	91

51-60	5	55.5	277.5
61-70	15	65.5	282.5
71-80	25	75.5	1887.5
81-90	20	85.5	1710
91-100	12	95.5	1146
Total	$\sum f_i = 80 = n$		$\sum f_i * m_i = 6130 = n$

Solution: we assume that all values falling into a particular class interval are located at the midpoint of the interval. The midpoint of class interval obtained by computing the mean of upper and lower limit of the interval, the midpoint of the first class limit.

$$\overline{\boldsymbol{x}} = \frac{\sum f_i * m_i}{n} = \frac{6130}{80} = 76.6 \text{ kg}$$

<u>Some properties of mean</u>:

i) $\sum(x_i - \overline{\boldsymbol{x}}) = 0$ in ungrouped data.

ii) $\sum(x_i - \overline{\boldsymbol{x}}) * f_i = 0$ in grouped data.

iii) Uniqueness: for given set of data there is one and only one arithmetic mean.

iv) Simplicity: the arithmetic mean is easily to understand and easy to compute.

v) Since each and every value in a set data enters into the computation of the mean, it is affected by each value. Extreme values, therefor, have an influence on the mean and, in some cases, can so distort it that it becomes undesirable as a measure of central tendency.

3.1.2 <u>Median</u>

- **<u>Median of ungrouped data</u>** :The median of a set of scores represents the middle values (50th percentile) when the scores are arranged as an array in order of increasing (or decreasing) magnitude. The median is often denoted by $\tilde{\boldsymbol{x}}$ (x-tilde). The median becomes more appropriate measure of central tendency when the data are skewed that is the majority of scores tend to accumulate toward either to the high or low end of the distribution with few extreme scores at the opposite end.

To locate the median first-rank the scores and follow the following guidelines:

1) For an odd number of scores, the median in the middle score.
2) For an even number of scores, the median in the mean (arithmetic average) of two middle scores.

Example: Compute the median of the following five scores:

$$10, 30, 27, 29, 12$$

Solution:

Step 1: We arrange the scores in ascending order as:

$$10, 12, 27, 29, 30$$

Step 2: Since the total number of scores is **5** and it is an odd number, then the third number (27) becomes the median, $\tilde{\boldsymbol{x}} = 27$.

Example: Compute the median of the following scores:

$$5, 6, 8, 50, 10, 70$$

Solution:

Step 1: We arrange the scores in ascending order as:

$$5, 6, 8, 10, 50, 70$$

Step 2: Since the total number of scores is **6** and it is an even number, then we compute the average of the values of the third and fourth scores of **8** and **10** as:

$$\tilde{\boldsymbol{x}} = \frac{8 + 10}{2} = 9$$

- **<u>Median of grouped data</u>**: *If we make our data in frequency table, the first step in computing median from grouped data is to locate class interval in which is it located of median.*

The median location is: either $c_1 = \sum \frac{f_i}{2}$ *or* $\frac{n}{2}$

$$\boldsymbol{Median = L + \frac{c_1 - c_2}{c_3} * w}$$

Where L is the lower limit for median class, $\boldsymbol{n}$ *is total frequency,* $\boldsymbol{c_2}$ *is cumulative frequency previous class which contain median,* $\boldsymbol{f_i}$ *is frequency median and w is class width (interval class) upper limit−lower limit.*

Example: : If we recall the previous example , *The amount of protein (in grams) for a variety of fast-food sandwiches is reported here. Construct a frequency distribution using 6 classes:*

Class Limits	***Frequency*** $\boldsymbol{f_i}$
12 – 19	***7***
20 – 27	***17***
28 – 35	***10***
36 – 43	***4***
44 – 51	***1***
52 – 59	***1***

Solution

Class Limits	***Frequency*** $\boldsymbol{f_i}$	***Cumulative Frequency CF***
12 – 19	***7***	***7***
20 – 27	***17***	***24***

28 – 35	***10***	***34***
36 – 43	***4***	***38***
44 – 51	***1***	***39***
52 – 59	***1***	***40***
Total	$\sum f_i = 40 = n$	

$c_1 = \sum \frac{f_i}{2} = \frac{40}{2} = 20$

$\therefore 20$ is included in the cumulative frequency of **24**, the class median related to **24** is ***20 – 27***, $L = 20$, $c_2 = 7$ is the cumulative frequency before the class which contains the median,

$$\boldsymbol{Median = L + \frac{c_1 - c_2}{c_3} * w = 20 + \frac{20 - 7}{17} * 8 = 26.1}$$

Example: 30 children whose got receiving treatment for haemolytic anaemia which show in the following table:

haemolytic anaemia Class	No. of children Frequency (f_i)	Midpoint (m)	Cumulative frequency (c. f)
6.5-7.5	1	7	1
7.5-8.5	5	8	6
8.5-9.5	11	9	17
9.5-10.5	9	10	26
10.5-11.5	3	11	29
11.5-12.5	1	12	30
Total	$\sum f_i = 30 = n$		

Solution: The median location is

$$\sum \frac{f_i}{2} \quad \text{or} \quad \frac{n}{2} = \frac{30}{2} = 15$$

$\therefore 15$ is included in the cumulative frequency of **17**, the class median related to **17** is **8.5-9.5** , $L = 8.5$, $c_2 = 6$ is the cumulative frequency before the class which contains the median,

$$\boldsymbol{w} = 9.5 - 8.5 = 1$$

The frequency of median class $(c_3) = 11$

$$\boldsymbol{Median = L + \frac{c_1 - c_2}{c_3} * w}$$

$$\text{Median} = 8.5 + \left[\left(\frac{30}{2} - 6\right)/11\right] * 1$$

$$= 8.5 + [(9)/11] * 1 = 8.5 + 0.81 = 9.31$$

<u>Some properties of Median</u>::

i) Uniqueness: As is true with the mean, there is only one median for a given set of data.

ii) Simplicity: The median is easy to calculate.

iii) It is not as drastically affected by the extreme value as the mean.

3.1.3 <u>Mode</u>

- **<u>Mode of ungrouped data</u>** : The mode (denoted my M) represents the most frequently occurring score. When two scores occur with the same greatest frequently, each one equals the mode the data set is considered to be bimodal. When more than two occur with the same greatest frequency the data set is said to be multimodal. To determine the mode, it must be locate the most frequently appearing number.

Example: Compute the median of the following scores:

$$10, 29, 26, 28, 15, 10, 25, 27, 10, 29$$

Solution:

We see the 10 is more frequently number (3 repetition) then the mode $M = 10$.

- **<u>Mode of grouped data</u>**: *When we need to calculate the mode in case of grouped frequency distribution, we will first identify the mode class, the class that has the highest frequency. Then, we will use the formula given below to calculate the mode*

$$\boldsymbol{Mode = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * w}$$

Where L is the lower limit of the mode class, Δ_1 *is frequency of class preceding the mode class,* Δ_2 *is* frequency of class succeeding the mode class *and w is class width.*

Example: : If we recall the previous example , *The amount of protein (in grams) for a variety of fast-food sandwiches is reported here. Construct a frequency distribution using 6 classes:*

Class Limits	***Frequency*** $\boldsymbol{f_i}$
12 – 19	***7***

20 – 27	***17***
28 – 35	***10***
36 – 43	***4***
44 – 51	***1***
52 – 59	***1***

Solution

$$\boldsymbol{Mode = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * w = 20 + \frac{10}{10 + 7} * 8 = 24.7}$$

<u>Exercises:</u>

For each of the following data sets compute **(a) mean, (b) median and (c) mode**:

1. 13 HIV positive patients whose are treated with highly active antiretroviral therapy (HAART) for at least 6 months. The CD4 T cell counts at baseline for the 13 subjects are listed below:

   $$230, 205, 313, 207, 227, 245, 173, 58, 103, 181, 105, 301, 169$$

2. The following frequency distribution table gives the age distraction of drivers whose were at fault in auto accidents that occurred during the 1-week period I a city.

Age (years)	Frequency (f)
Less than 20	2
20-25	12
25-30	18
30-35	14
35-40	15
45-50	16
50 and over	35