

## DESCRIPTIVE STATISTICS

### 3.2.3 Numerical summary of data

We will look at the following parameters to help us understand how data is summarized numerically:

1. Mode
2. Median
3. Quartiles
4. Arithmetic mean
5. Variance

#### Mode

This is the most frequent number or value. In the data set below:

2, 4, 5, 6, 7, 7, 8, 9, 10

The mode is 7.

In the set below:

2, 4, 5, 5, 6, 7, 7, 8, 9, 10

The modes are 5 and 7 i.e. a bimode and if more than two values, then we have a multimode.

#### Median

This is the middle value or observation in a ranked order of values. To get the median in the data set below:

10, 16, 12, 5, 22, 14, 19

We first rank values in increasing order

5, 10, 12, **14**, 16, 19, 22

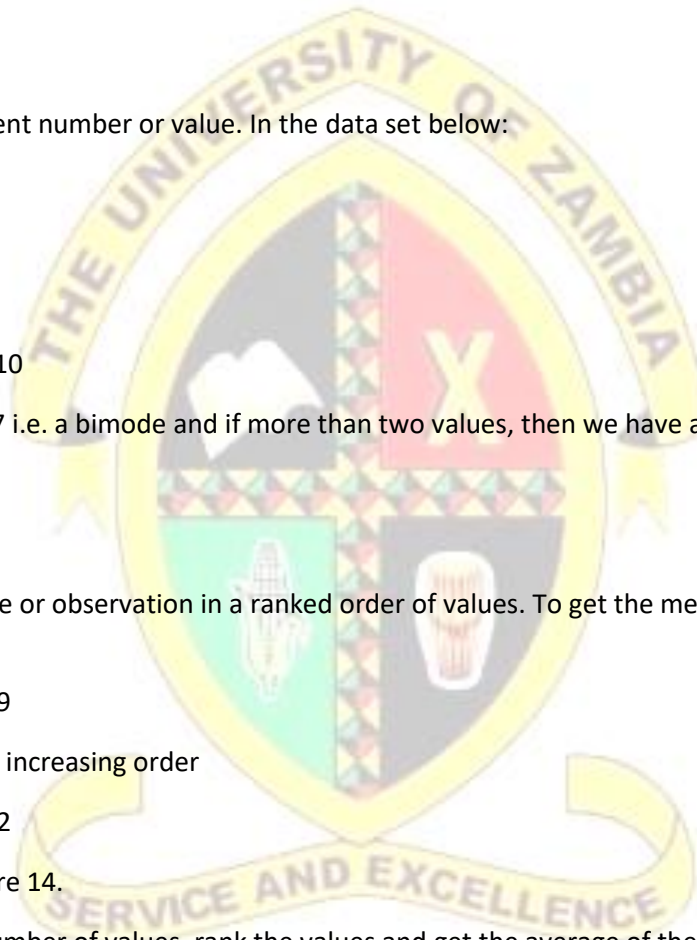
The median is therefore 14.

If you have an even number of values, rank the values and get the average of the two middle values.

If we have:

5, 10, 12, 14, 16, 19, 22, 28

The median will be 15.



## Quartiles

This term comes from the word “quarter”. While the median divides data into 2 parts, the quartile divides data into 4 parts as follows:

- $\frac{1}{4}$  of the data has values less than or equal to the lower quartile “q”
- $\frac{1}{4}$  of the data is between the lower quartile “q” and the median “m”
- $\frac{1}{4}$  of the data is between the median “m” and the upper quartile “Q”
- $\frac{1}{4}$  of the data is greater than the upper quartile “Q”

The interquartile range is the difference between the upper quartile “Q” and the lower quartile “q” i.e Q-q.

The data is therefore divided into 4 parts as follows:

Q1	Q2	Q3	Q4
q	m	Q	>Q

To calculate these parameters for an **odd-number data set** below where  $n = 7$  ( $n$  is the number of observations):

5, 22, 14, 12, 19, 16, 10

- a) Order / rank the values  
5, 10, 12, 14, 16, 19, 22
- b) Find the median position by calculating  $(n+1)/2$ . In our example this is  $(7+1)/2 = 4$ . The median is therefore the 4<sup>th</sup> observation which is **14**.
- c) Find q or Q1 by getting the total number of observations before the median and using the above formula  $(n+1)/2$ .  
**5, 10, 12, 14, 16, 19, 22** There are 3 observations before the median therefore we have  $(3+1)/2 = 2$ . Therefore q or Q1 is the 2<sup>nd</sup> observation before the median which is **10**.
- d) Find Q or Q3 by getting the total number of observations after the median and using the above formula  $(n+1)/2$ .  
5, 10, 12, **14, 16, 19, 22** There are 3 observations after the median therefore we have  $(3+1)/2 = 2$ . Therefore Q or Q3 is the 2<sup>nd</sup> observation after the median which is **19**.

To calculate these parameters for an **even-number data set** below where  $n = 10$  ( $n$  is the number of observations):

8, 8, 9, 10, 11, 11, 11, 11, 14, 15

- a) Order / rank the values  
8, 8, 9, 10, 11, 11, 11, 11, 14, 15
- b) Find the median by calculating the average of the two middle values i.e.  $(11+11)/2 = 11$

- c) Find  $q$  or  $Q_1$  by dividing the total observations into two halves and then using the formula  $(n+1)/2$  for the first half of the data to find the position of  $q$  or  $Q_1$ .  
8, 8, 9, 10, 11, 11, 11, 11, 14, 15. There are 5 observations, therefore  $(5+1)/2 = 3$ .  $Q_1$  is on the 3<sup>rd</sup> position in the first half which is **9**.
- d) Find  $Q$  or  $Q_3$  by dividing the total observations into two halves and then using the formula  $(n+1)/2$  for the second half of the data to find the position of  $Q$  or  $Q_3$ .  
8, 8, 9, 10, 11, 11, 11, 11, 14, 15. There are 5 observations, therefore  $(5+1)/2 = 3$ .  $Q_3$  is on the 3<sup>rd</sup> position in the second half which is **11**.

In the data, the interquartile range (**IQR**), is  $Q_3 - Q_1$  or  $Q - q = 11 - 9 = 2$ .

### Arithmetic Mean and Variance

This is calculated differently depending on whether the data is **grouped** or **ungrouped**. Grouped data is data organized into groups with a frequency distribution. Ungrouped data does not have frequencies attributed to the data.

I will send out videos or links to videos from youtube with meaningful examples on how to calculate means and variances.

