

# Statistics.

Statistics is basically the science of collecting ,analyzing and interpreting data.

Statistics can be classified to two main categories :

| Descriptive  | Inferential   |
|--|---|
| Organizing and summarizing data<br>Using numbers & graphs .          | Using sample data to make an inference or draw a conclusion of the population . |
| Data summary:<br>Bar graphs , histograms ,pie chart , etc.           | Uses probability to determine how confident we can be that the                  |
| Measures of central tendency:<br>Mean , median & mode .              | Conclusions we make are correct.  |
| Measures of variability :<br>Range , variance & standard deviation . | _____   |

|                       |  |
|-----------------------|--|
| Mean :<br>“ average “ | Sum of the Given Data/Total number of Data   |
| Median :              | is the middle number in a set of <i>data</i> .<br>If there is an even number of items of data, there will be two numbers in the middle. The median is the number that is half way between these two numbers. |
| Mode :                | the number, or item, which occurs most often in a set of <i>data</i>   |
| Renge :               | The largest number – the smallest number .   |

**Variance** : is a measure of a spread of the data .

The variance tell us how far the data is spread from the mean.

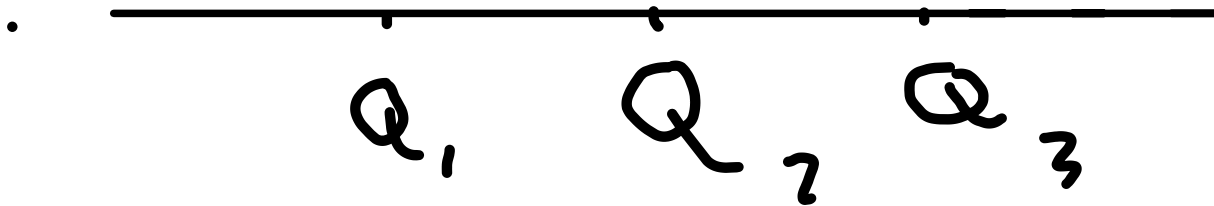
**Variance Formula** :  $s^2 = \sum (x - \bar{x})^2 / n - 1$

Where (  $\bar{x}$  ) is the mean .

standard deviation :  $SD = \sqrt{\sum (x - \bar{x})^2 / n - 1}$ .

---

**Interquartile range and any potential outliers:**



**Q1 is the median of the lower half of the data**

**Q2 is the median of the all data**

**Q3 is the median of the upper half of the data**

The interquartile range represents the middle 50% of the data .

Or IQR is the difference between the third quartile (Q3) and the first quartile (Q1).

**$IQR = Q3 - Q1$**  .

**$Outliers = [Q1 - 1.5 \cdot IQR, Q3 + 1.5 \cdot IQR]$**

if we have a set of numbers and any number of them doesn't exist in outlier range  
so that number is considered to be an outlier number

**Ex :** 11 , 31 , 21 , 19 , 8 , 54 , 35 , 26 , 23 , 13 , 29 , 17 .

Mean =  $(11+31+21+19+8+54+35+26+23+13+29+17)/12 = 22.25$ .

Median : 1- rearrange : 8,11,13,17,19,21,23,26,29,31,35,54

2- median =  $(21+23)/2 = 22$ .

Mode = no mode .

Range =  $54 - 8 = 46$  .

Variance =  $[(11-22.25)^2 + (11-22.25)^2 + (13-22.25)^2 + (17-22.25)^2 + (19-22.25)^2 +$   
 $(21-22.25)^2 + (23-22.25)^2 + (26-22.25)^2 + (29-22.25)^2 + (31-22.25)^2 + (35-22.25)^2 +$   
 $(54-22.25)^2]/11 = 150.31$ .

standard deviation =  $(\text{variance})^{0.5} = 12.26$ .

$Q1 = (17+13)/2 = 15$  ,  $Q2 = (21+23)/2 = 22$  ,  $Q3 = (29+31)/2 = 30$

$IQR = Q3 - Q1 = 15$

Outliers =  $[Q1 - 1.5 * IQR , Q3 + 1.5 IQR] = [-7.5, 52.5]$