

## STATISTICS – DAY 2

**What is Variability:** It describes how far apart data points lie from each other and from center of distribution.

- Variability is also referred as **spread or scatter or dispersion**.

### WHY VARIABILITY MATTERS

- While measure of central tendency tells you about the summary of the data.
- Variability summarizes how far (spread) they are from each other
- This is important because it tell you whether the points tend to be clustered around center or more widely spread out.
- Low variability is ideal because it means that you can better predict information about the population based on sample data.
- High Variability means that values are less consistent so it is harder to make predictions.

### Range

- It shows the spread of the data from the lowest to the highest value in distribution.
- It is one of the most easiest way of finding the measure of variability.

### **Example:**

#### **Data**

72  
110  
134  
190  
238  
287  
305  
324

#### **To calculate the range:**

Lowest Value: 72

Highest Value: 324

Range = Highest Value – Lowest Value  
324 – 72

**Range: 252**

**It does not give any information about the distribution of the values.**

**2) IQR (Inter Quartile Range):** IQR gives you the spread of the data from middle of your distribution.

Q1 => (Quartile 1) => 25% of value from your data

Q2 => (the median) => median value (mid value)

Q3 => (Quartile 3) => 75% of value from your data

Q4 => Highest Value

### **Data**

72

110

134

190

238

287

305

324

To calculate Q1:  $0.25 * 8 \Rightarrow 2$

Q1 : will be 110

Q2 => 214 (mid value)

Q3 =>  $0.75 * 8 \Rightarrow 6$

Q3 => 287

Q4 => 324

IQR = Q3 – Q1  
287 – 110  
=> **177**

### 3) Standard Deviation :

- It finds out the average of variability in your data set
- It tells you on average how far each score(points) lie from the mean
- The larger the standard deviation, the more variance (spread) the data is

#### Steps To calculate Standard Deviation:

- 1) List each score and find out the mean value
- 2) subtract the mean value from each data to get **deviation from mean**
- 3) square each of those deviations
- 4) Add up all the squared deviations
- 5) Divide the sum of squared deviation by n-1 (where n is the sample size)
- 6) find the square root of the number.

#### Data

72  
110  
134  
190  
238  
287  
305  
324

#### To calculate the stand deviation for the given data set

##### Step1:

$$72+110+134+190+238+287+305+324$$

$$\text{Mean} = 1660/8$$

$$= 207.5$$

##### Step2:

##### Deviation from Mean

72	$72 - 207.5 \Rightarrow -135.5$
110	$110 - 207.5 \Rightarrow -97.5$
134	$134 - 207.5 \Rightarrow -73.5$
190	$190 - 207.5 \Rightarrow -17.5$
238	$238 - 207.5 = 30.5$
287	$287 - 207.5 \Rightarrow 79.5$
305	$305 - 207.5 \Rightarrow 97.5$
324	$324 - 207.5 \Rightarrow 116.5$

##### Step3: square each of those deviations

**-135.5 => -18360.25**

**-97.7 => -9545.29**

**-73.5 => -5402.25**

**-17.57 => -308.70**

**30.5 => 930.25**

**79.5 => 6320.25**

**97.5 => 9506.25**

**116.5 => 13572.25**

#### **Step4:**

Add up all the squared deviations

**-3287.48**

#### **Step 5:**

**-3287.48 / 7**

**=> -469.64**

#### **Step6:**

**sqrt(-469.64)**

**21.67**