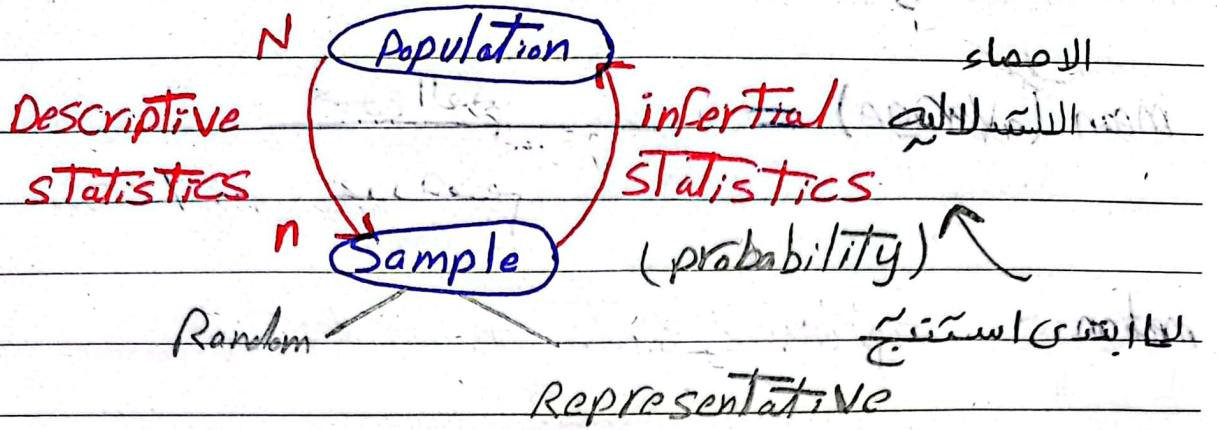
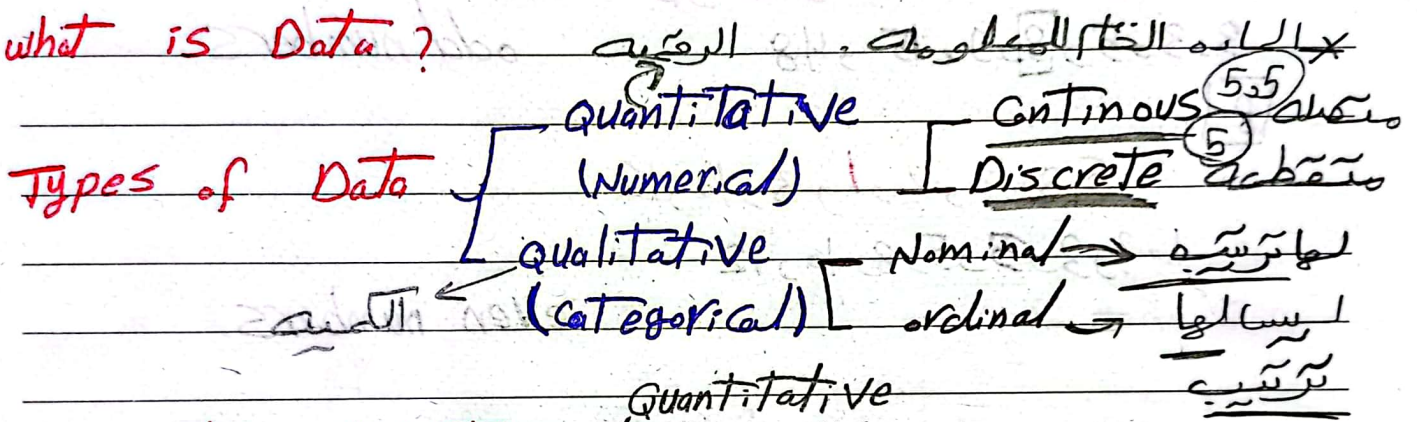


# [Statistics] [for Data Science]

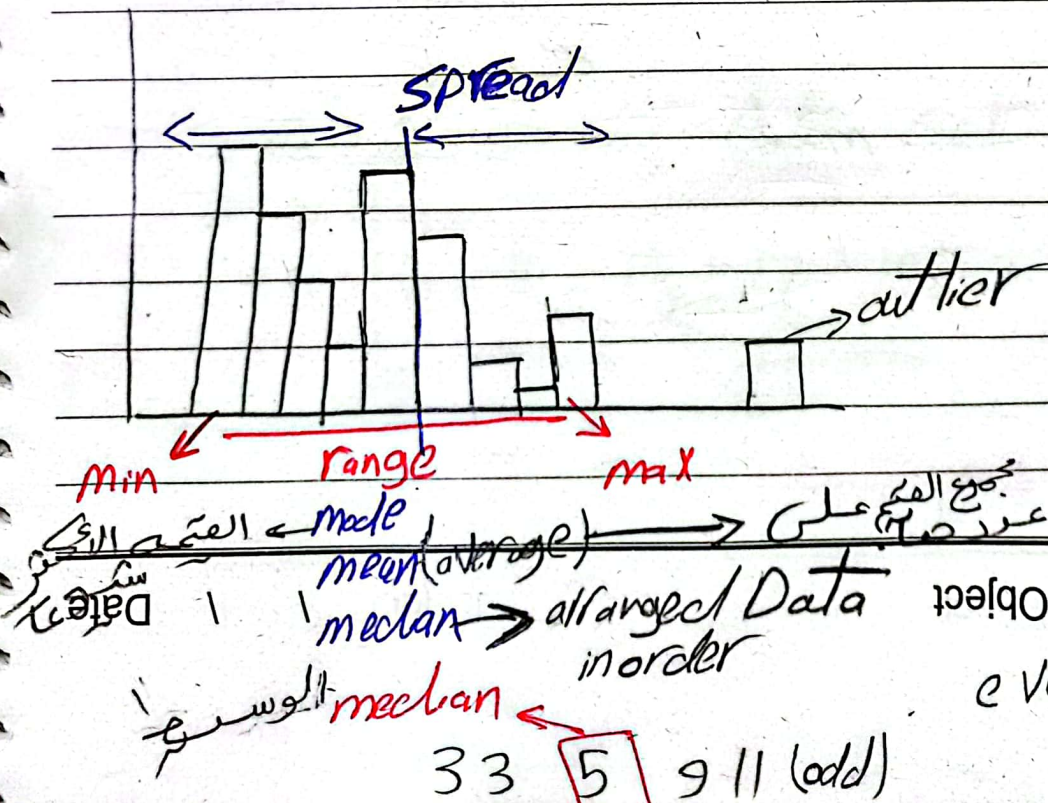


what is Data?



Descriptive statistics: Quantitative (numerical) وصف البيانات

\* لازم نفهم البيانات ونعرف كيف نحسنها ونحافظ عليها





# Data Visualization

## measures of Centers:

$$\text{Mean (average)} = \frac{\text{مجموع القيم}}{\text{عدد قيم}}$$

median arranged Data in order  
الوسط

9, 3, 5, 8, 3, 15, 48  
3, 3, 5, 8, 15, 48 odd numbers.  
8

5, 8, 3, 2, 1, 3, 10, 12  
1, 2, 3, 3, 5, 8, 10, 12 even numbers.  
median =  $\frac{3+5}{2} = 4$

mode Common Value

mode

القيمة الأكثر شيوعاً  
و أكثر قيم تكرار  
\* ممكنة ما يكثر فيه



## Random Variables

Category لازم يعبر عنه

كله منه Data

any observation

$X = X_1, X_2, X_3, X_4, \dots, X_n$

$$X = \sum_{i=1}^n X_i = X_1 + X_2 + X_3 + X_4 + \dots + X_n$$

$$\text{Sample } \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

mean

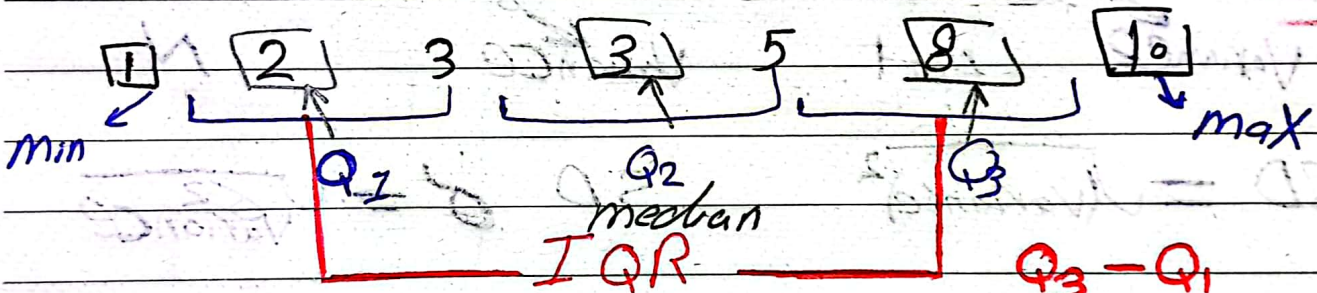
population

$$\mu = \frac{\sum_{i=1}^N X_i}{N}$$

$N \rightarrow$  عدد

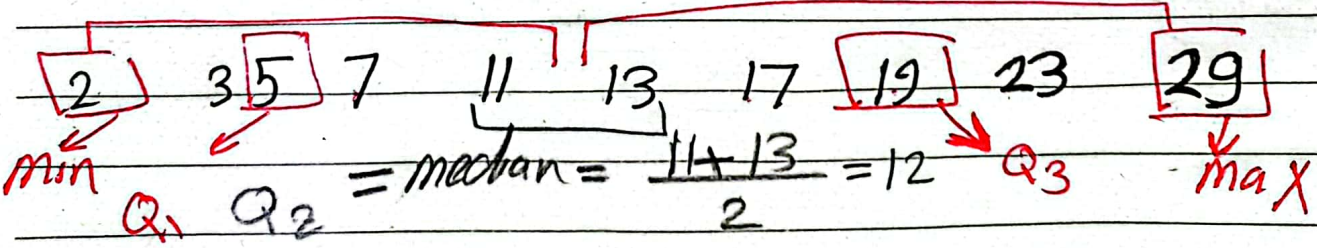
## 5 Numbers Summer

"Measures of Spread"



Quartile  $\rightarrow$  median of each quarter  
 $IQR = Q_3 - Q_1$

$Q_2 =$  median in any Data set



$$IQR = Q_3 - Q_1 = 19 - 5 = 14$$

interquartile Range

Date

Object



# Standard deviation

الانحراف المعياري

الوسط  
بند النقطة عين (mean)

10 14 10 6

$$\bar{X} = \frac{40}{4} = (10) \text{ mean}$$

$$X_i - \bar{X} \begin{cases} 10 - 10 = 0 \\ 14 - 10 = 4 \\ 10 - 10 = 0 \\ 6 - 10 = -4 \end{cases}$$

$$SD = \sqrt{\sum_{X_i=0}^n (X_i - \bar{X})^2}$$

$$= 0 \quad \times$$

Sample Variance

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Variance

population

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

Variance

$$SD = \sqrt{\text{Variance}}$$

$$SD \sigma = \sqrt{\text{Variance}}$$

G-Variance (نسبة التباين بالنسبة للتباين)

$$G\text{-relation} = \frac{SD}{Co\text{-Variance}}$$

مقياس مدى تغلف  
متغير أو أكثر  
بعض

G Salty

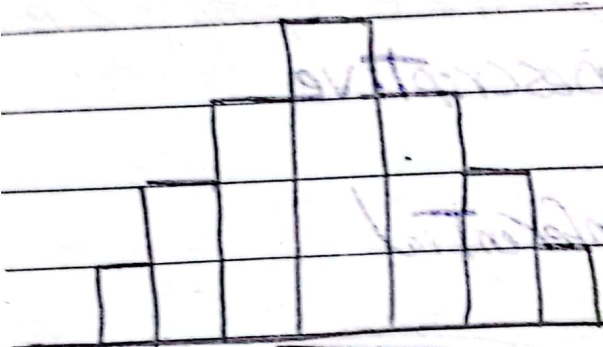
Date

Object



IR Shapes:

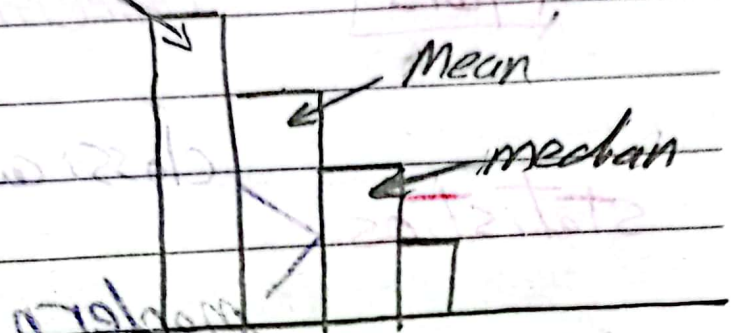
Symmetrical



mean = median

= mode

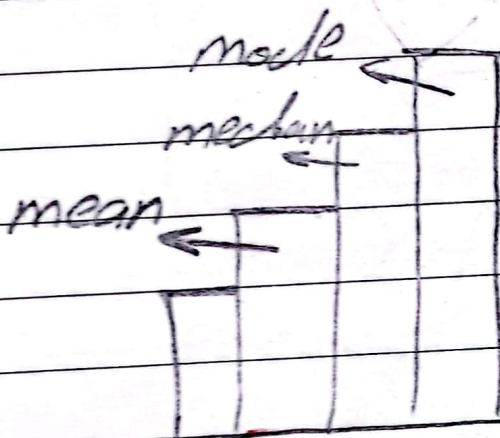
mode "Right Skewed"



mean > median



left skewed



outlier ⇒

بعيد ه عنه قيمه الانا اللي عندي

بشقي عنده عنده القيمه اللي عندي

\*



statistics → تحمل الخطأ

"Analysis = question"

Tools

Collecting (visualize) Analyze  
Descriptive


statistics   
 classical   
 modern   
 inferential

Estimation ← انما يعرفش او صلا للتخمين بالتقدير  
بعمل تقدير لاداءات الاشياء اشتغالها  
بجاء لا اقله ستة  
الخطأ فيه مفسد

Population   
 Sample   
 relative Random   
 Population   
 لا يمكن ان ندرس الـ Population  
لانها عدد كبير جداً  
ياخذ عينات تمثل العدد الكامل  
كله لانهم متجانسون وواحد ومثاله  
لا Data كامله  
يستخرج من Sample كل حاجه

[Descriptive statistics]

Frequency Distribution   
 التكرارات

Histogram   
    
 بحالينه   
 التكرارات

## measures of centers:-

- Mean
- Median
- Mode

**Data:** Distinct pieces of information

\* Data represent in Video / Audio / Spreadsheet / Images / Text

Data used to understand every facet of our lives.

## Types of Data

Quantitative

Number of Dogs

Qualitative (Categorical)

Breed of Dogs

Quantitative Data: - numeric values that allow mathematical operations

Qualitative Data (Categorical): - a group or set of items

## Types of Qualitative Data

Nominal

not order

الترتيب

الترتيب

ordinal

ordered

الترتيب

Ranked

Object



# ~~Types~~ Types of Quantitative Data

Continuous data

Discrete data

age  
decimal number and  
integer

number of Dogs  
not include decimal  
only integer

any numeric values  
Sometimes negative  
numbers.

Zip Code Categorical

لا تلو ضيفت في مولوجيتا في بيت موهيبي في  
useful information

1 3 5 7 [8] 10 15 15 22

1 2 3 5 [7 8] 10 15 15 22

Ashley	2
Steve	1
Jeff	<u>2</u>
Kyle	3
Lisa	8