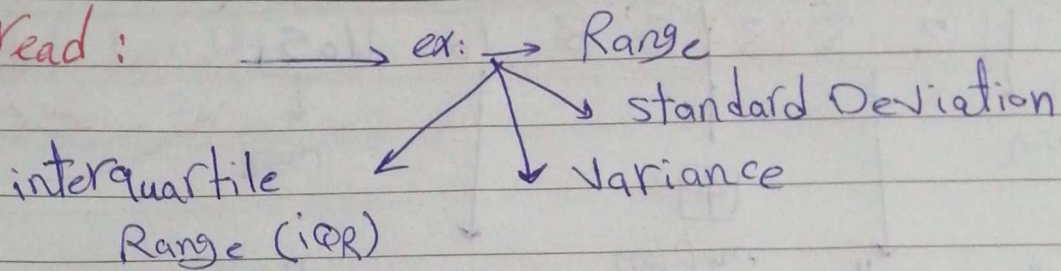


\* Spread:



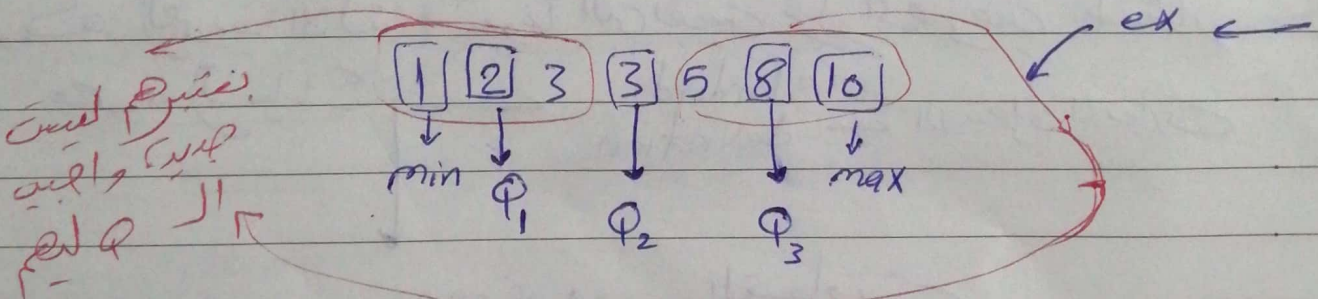
## Five Number summary

maximum

 $(Q_3)$  third Quartile $(Q_2)$  second Quartile (median) $(Q_1)$  first Quartile

minimum

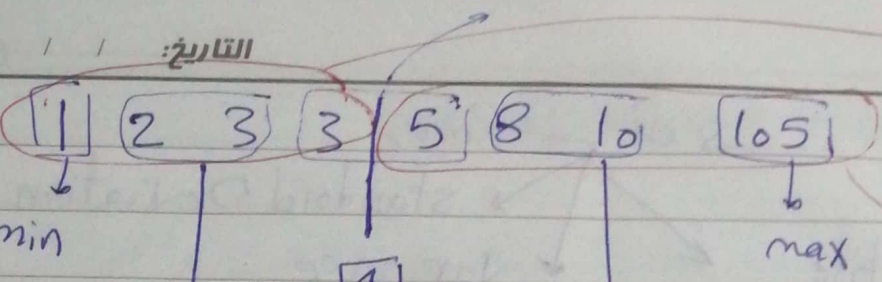
ترتيب (قيم أرتب) لنحصل على min و max

median ← وهذا هو  $(Q_2)$  و  $Q_1$  ← ثاني رقم بعد ال minو  $Q_3$  ← ثاني رقم قبل ال max

لو عدد الأرقام زوجي ←



mean و median



الموضوع: التوزيعات الاحتمالية  
المعيار: المتوسط الحسابي  
المعيار: الانحراف المعياري

mean (Q1)

2.5

median (Q2)

mean (Q3)

\* Range  $\rightarrow$  max - min ✓

\* interquartile Range (IQR)  $\rightarrow$  Q3 - Q1 ✓

Box Plot  $\rightarrow$  useful for quickly comparing the spread of two Data sets

و يمكن معرفة الأكثر تنوعاً التي تقيس مدى التفرقة في انتشار  
مجموعه البيانات  
standard deviation  $\rightarrow$  الانحراف المعياري  
variance  $\rightarrow$  التباين



\* standard deviation :

هو الفرق المتوسط بين كل قيمة من المتوسط .

Ex:  $\begin{matrix} x_1 & x_2 & x_3 & x_4 \\ 6 & 10 & 14 & 10 \end{matrix}$

$$① \text{ mean } (\bar{x}) = \frac{\sum_{i=1}^n x_i}{n} = \frac{6+10+14+10}{4} = \boxed{10}$$

② Distance of each one from this mean:  $(x_i - \bar{x})$

$$6 - 10 = -4$$

$$10 - 10 = 0$$

$$14 - 10 = 4$$

$$10 - 10 = 0$$

$$③ \text{ the average of these distances} = \frac{0+4+0-4}{4} = \boxed{0}$$

← هذا يدل على أنه كل القيم متساوية أو أنه لا يوجد spread

لذلك نحتاج لعدد كل القيم موجبة

عدد

$$(x_i - \bar{x})^2$$

$$(10 - 10)^2 \longrightarrow 0$$

$$(14 - 10)^2 \longrightarrow 16$$

$$(10 - 10)^2 \longrightarrow 0$$

$$(6 - 10)^2 \longrightarrow 16$$

Variance  $\rightarrow$  Average Squared Difference of each value from the mean

$$\rightarrow = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{0 + 16 + 0 + 16}{4} = \boxed{8}$$

Variance

\* standard deviation =  $\sqrt{\text{Variance}} = \sqrt{8} = 2.83$  ✓

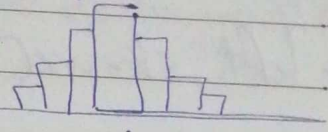
$\rightarrow$  on Average, How far each Point in our Dataset is from the mean.



## \* shape \*

### Normal distribution

the Bell curve



this shape

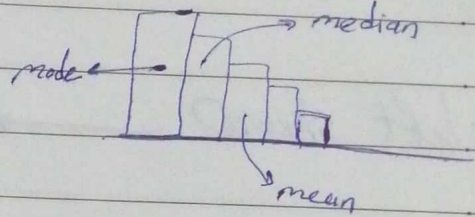
mean = median = mode

Right skewed

\* Distribution → left skewed

symmetric

mode أقرب لا



يمكن أن تكون البيانات غير متوزعة

## \* outliers \* القيم الشاذة

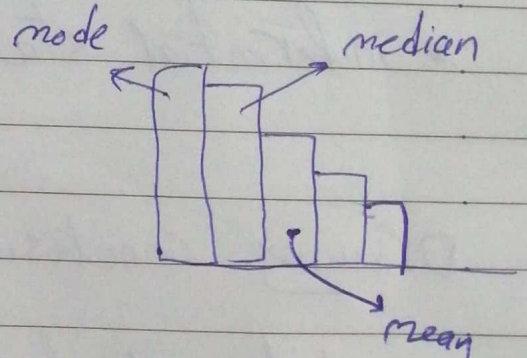
نقاط بيانات تقع بعيداً عن باقي البيانات

## \* shape \*

→ ~~left~~ skewed  
Right

median closer to mode

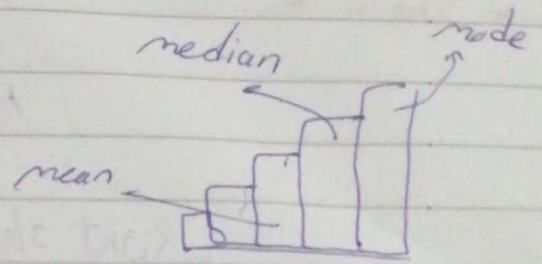
mean > median





\* left skewed \*

→ median closer to mode  
→ mean < median



\* left skewed → GPA ) age of death ) price changes

\* Right skewed → Bloodstream ) Distribution of wealth  
human athletic abilities.

الخطوات الأولى  
Descriptive statistics ← الإحصاء الوصفي

الخطوات الثانية  
\* inferential statistics \*

Drawing conclusions about A Population Based  
on data collected from a sample of individuals  
from that Population.

استنتاجات حول مجتمع كبرى على عينة