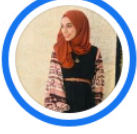



STAT2361  
Lecture Notes - First

We'am A Rostom

<  We'am A Rostom ▶ Commercial  
Zone - BZU  
34m · 

يعطيكم العافيه 🙏  
الي بعتولي بدهم تلخيص ماده الستات الفيرست  
جمعتلكم الي كنت مسجلته عالدفتر + بلص تلخيص العناوين  
والقوانين المهمة وطريقه استخدام الاله الحاسبه  
ركزوا كثير على كل عنوان وكيف منستخدم في الاله الحاسبه اغلبهم  
نفس البدايه بس اللهم النتيجة بتختلف حسب المطلوب  
الي ما بفهم منيح بامكانه يبعثلي مسج افهمه كيف يستخدم الاله  
الحاسبه بكل مطلوب

برضو اي اشي مش واضح او بتشوفوا ان مش كامل وما وصلتكم  
فكرته بتقدروا تتواصلوا معي

.....

ملاحظة: تدققوش عالغلاط الاملائية المهم الاخلاق 😊

وبس  
شكرا عفوا

stat 2361

chapter 1  
sc. 1

### Data and statistics

#### Section 1 :

- statistics : علم وفن جمع البيانات، تحليلها، عرضها وتفسيرها
- Data : الرسوم والكميات التي يتم جمعها وتفسيرها وتحليلها
- Data Set : مجموعة الكميات التي قُدمت في دراسة معينة
- Elements : العنصر والكميات التي تجعل الدراسة عليها
- Variable : الصفات والكميات التي تقوم بدراسة
- observation : الملاحظات القياسية لكل عنصر

#### Section 2 : Scales of measurements

تقسيم إلى ثلاث تدرجات أو تدرجات البيانات

(1) Nominal scale : تدرجات أسماء أو صفات في كل عام  
سواء كانت البيانات رقمية أو غير رقمية  
مثل : Phone number / ID's number / gender

(2) Ordinal scale : التدرجات التي يتغير فيها الأسماء أو الصفات  
من عام إلى عام ولكن الترتيب حسب الأولوية  
سواء رقمي أو غير رقمي  
مثل : ترتيب الترتيب good -> very good -> excellent

من الأعلى للأقل  
من الأولوية

(3) Ratio scale : اكراول التي يمتلك الصفر فيها عتبه صفر  
تظهر وتقتصر على السمات الرقمية

مثل : ~~الراتب / الوقت~~

(4) Interval scale : مداول ذات حقة رقمية ذات  
صفر من درجات الكمية

\* we have 2 Types of Data :

- \* Quantitative data : ( Interval / Ratio ) قيم كمية
- \* Qualitative data : ( ordinal , Nominal ) قيم نوعية

### Section 3: Data Sources

(1) Existing Sources : دراسة حالة خاصة موجودة

- \* internal
- \* external
- \* internet

(2) Statistical studies :

- \* observational : ملاحظة ومراقبة
- \* experiment : التجارب بحرية المتغير  
النتائج



\* section (4,5) :

## Statistic

(1) Descriptive stat. :

دراسة البيانات

(2) Inference Stat. :

دراسة الاستدلال

\* population : collection of all elements

\* sample : subject of population

\* Census : survey for all elements

Chapter 2 =

Descriptive statistic

Qualitative data

Section 1 =

من ناحية اعداد

\* Tabular :

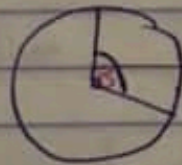
a - Frequency distribution (Fd) =

b - Relative Frequency (RF) =  $\frac{F}{n}$

c - percentage frequency (PF) =  $(\frac{F}{n}) * 100\%$

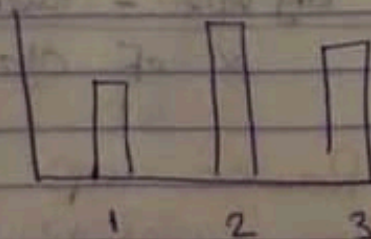
\* graphical = من ناحية الرسوم البيانية

(\*) pie chart



قياس الزاوية =  $360 * RF$

(\*) bar graph



يتميز بأنها يمكن فيه مقابلة غير متساوية

\* راجع تعریفی طور پر اسکرابان مناجیہ  
تاسیہ ←

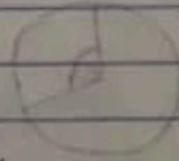
\* مناجیہ اسکرابان : Cumulative Frequency \*

- Cumulative Frequency

Which are less or equal the upper class limit

- Cumulative Relevant

- " " " " percentage



\* اسکرابان اسکرابان :-

(1) number of classes

(2) width of " =  $\frac{\text{larg value} - \text{smallest value}}{\text{* of class}}$

↳ Round up

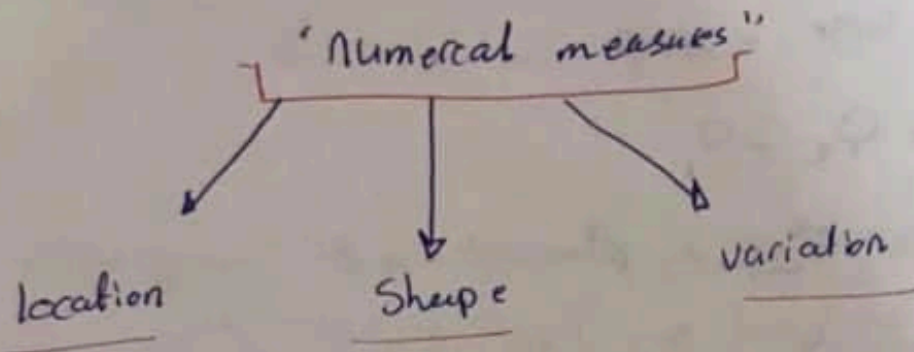
بجای بچہ تقریباً ملے گا! اس کا جواب

3	← Round up	2.3
4	←	3.5
3	←	2.2
3	←	2.4

فاسد ہے؟



"Chapter 3" SC-1



[1] ان کا ڈھونڈنا، ترتیب سے لکھنا

[1] location:

- \* mean
- \* mode
- \* median : ترتیب سے لکھنا
- \* percentile : ترتیب سے لکھنا
- \* Quartiles : ترتیب سے لکھنا

\* mean: average  $\rightarrow \bar{x}$ ,  $\mu = \frac{\text{مجموع}}{\text{تعداد}} = \frac{\sum x_i}{n}$

\* mode: الأكثر تكراراً

\* median: midpoint

\* percentile  $\rightarrow p^{\text{th}}$  :  $i = \frac{p}{100} \cdot n$

$\rightarrow \frac{x_i + x_{i+1}}{2}$

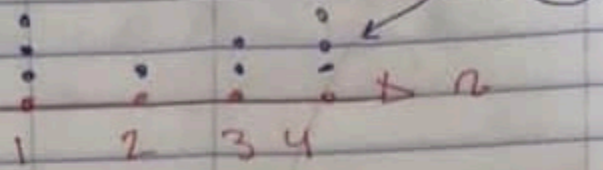
Round up

- \* Quartiles  $\rightarrow$ 
  - $Q_1 = 25\%$
  - $Q_2 = 50\% \rightarrow$  median = midpoint
  - $Q_3 = 75\%$

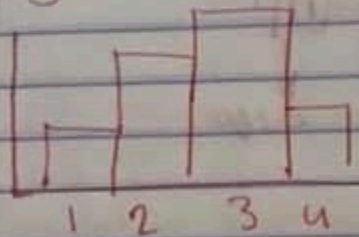


من ناحية الرسوم البيانية ~~البيانية~~ :

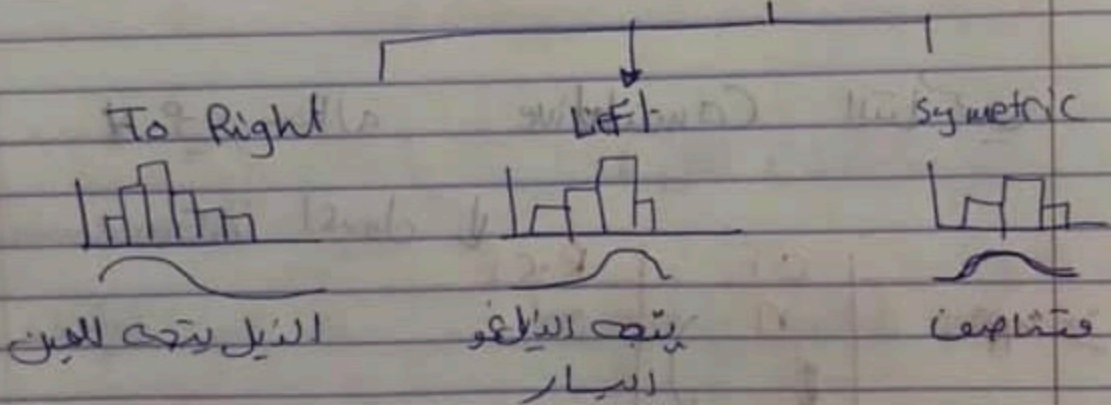
\* Dot plot



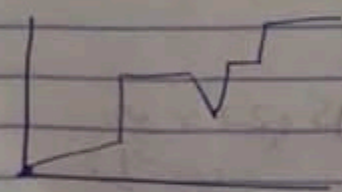
\* Histogram



تعتبر أن الأعمدة  
تأخذ مساحة  
بكونها  
3 أشكال



\* ogaive : ~~البيانية التراكمية~~  
Cumulative



• Gross-Tabulation → بيانات كمية وكمية

• Scatter diagram → ~~البيانات~~ ~~البيانية~~ ~~البيانية~~



(3) [lower  $\leftrightarrow$  upper]  
 $(x - y)$

	F	R, F
		P, F

$\Sigma F = n$

↓  
 2000

يكون صنفين  
 فترات مثل [5-10]  
 الفرق بينهم  
 classes

لا  $\rightarrow$  حالة Cumulative الترتيب

↓  
 يكون الكود

	C.F
$\geq$ upper	n
$\geq$ upper	F + n = Z
↓	F + n + Z

يكون مجموع كل نتيجة مع اي قبالها لكل الترتيب



مثال : 65, 60, 57, 50, 42  
 ملاحظة : القيمة التي هي أكبر من القيمة الوسطى

عدد موزون : 57  
 ملاحظة : كل قيمتين مع بعض والنتيجة هي 57  
 Medication

57 : medication

[4] Percentile : "the  $p^{th}$  percentile" is a value such that at least  $p\%$  of the observations are less than or equal to this value and  $(100 - p)^{th}$  of observations are greater than this value

أ) ترتيب البيانات من الأصغر إلى الأكبر  
 smallest to greatest

ب) حدد القيمة التي يكون فيها  $p\%$  من البيانات أصغر من هذا

مثال : 20% percentile  
 80%  
 20<sup>th</sup> percentile

40% percentile  
 60%  
 percentile

مثال : 75<sup>th</sup> percentile  
 76<sup>th</sup> percentile  
 75% 25%  
 60  
 75  
 25  
 ملاحظة : القيمة التي يكون فيها 75% من البيانات أصغر من هذا والباقي 25% أكبر من هذا



[2] **Mode** : the value with greatest frequency

بہاؤ سے زیادہ تکرار سے (ایک سے زیادہ)

مثلاً : 46, 54, 42, 46, 32 \*  
 ← 2 1 1 1  
 46, 54, 42, 46, 32 \*  
 ← 2 1 1 1  
 46, 54, 46, 32, 49 \*

→ mode = 46

46, 54, 46, 32, 49 \*

→ mode = 46, 54

No mode (کوئی تکرار نہیں ہے)

مثلاً : 46, 32, 42, 46, 42, 48 \*  
 ← 2 2 1  
 46, 32, 42, 46, 42, 48 \*  
 ← 2 2 1  
 54, 60 \*

→ mode 42, 46, 45

"multi mode"

مثلاً : 3 mode (3 تکرار سے زیادہ) 42, 46, 45  
 3 سے زیادہ تکرار سے (3 سے زیادہ) mode

[3] **Median** : the value in the middle

اوسط سے زیادہ

[1] **First step** : arrange data in ascending order

(تصاعق / سیرت) سے زیادہ

example : 40, 50, 44, 60

40, 44, 50, 60 (تصاعق سے زیادہ) [1]

[2] **second step** : arrange data in ascending order

بہاؤ سے زیادہ

مثلاً : 40, 44, 50, 60 (تصاعق سے زیادہ) \*

$$\text{median} = \frac{44 + 50}{2} = 47$$

SC. 3.1 Measures of location

Type of measure location:

- 1] Mean (average)
- 2] Mode
- 3] Median
- 4] percentile
- 5] Quartiles

Sample statistic "sample mean"

\* mean: data for a sample  $x_1, x_2, x_3, \dots, x_n$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

\* population mean: data for population  $x_1, x_2, \dots, x_n$   
↳ "population parameter"

example: Consider the values and following data from

46, 59, 42, 46, 32

Compute the mean:

1] كذا هي اى نوع  $\bar{x}$  و  $M$  / نفس السؤال  
Sample mean

$$\bar{x} = \frac{46 + 42 + 46 + 32 + 59}{5} = 44$$

[2] variation

- Range = large - small
- IQR =  $Q_3 - Q_1$
- Variance =  $S_{sample}^2 \quad \sigma^2$
- Standard deviation =  $S_{sample} \quad \sigma_{population}$
- Coefficient of variation =  $\frac{S}{\bar{x}} * 100\%$

• Variance =  $\sum^{(shift)}$  (mode) (3) (=)

(mode) (2) → متغير واحد

$(x_i) (n+)$   
 $(\lambda_i) (n+)$

(Shift)  $\begin{matrix} (1) \\ (2) (=) \\ (3) \end{matrix}$  ] →  $\frac{u}{a}$   
 اعطون

$\bar{x}$	$\sigma_x$	$S_x$
1	2	3



→ Solution: *الحل*

\*  $Q_1 = 25^{\text{th}}$  percentile *إذا يوجد 25*

$\{1\} = 16, 20, 25, 26, 27, 28, 30, 34$

$\{2\} \quad i = \frac{25 * 8}{100} = 2$

$\Rightarrow 25^{\text{th}}$  percentile  $= \frac{x_2 + x_3}{2}$

$= \frac{20 + 25}{2} = \frac{45}{2} = 22.5$

$\Rightarrow Q_1 = 22.5$

\*  $Q_2 = 50^{\text{th}}$  percentile and median

16, 20, 25, 26, 27, 28, 30, 34

median  $= \frac{25 + 27}{2} = 26 = Q_2$

\*  $Q_3 = 75^{\text{th}}$  percentile

$i = \frac{75 * 8}{100} = 6$

$= \frac{x_6 + x_7}{2} = \frac{28 + 30}{2} = \frac{58}{2} = 29$

$Q_3 = 29$

→ Computing the  $p^{\text{th}}$  percentile:

1) arrange data in ascending order  
 ترتيب البيانات تصاعدياً

$$i = \frac{P \cdot n}{100}$$

if  $i$  is integer →  $x_i$  is the  $p^{\text{th}}$  percentile

$$p^{\text{th}} \text{ percentile} = \frac{x_5 + x_6}{2} = 20.5$$

if  $i$  is not an integer →  $i$  is not integer  
 ex:  $i = 1.5$

لذا نقرنها للعدد في صيغة النسبة  
 لأننا نعرفها مع

$$i = 1.5 \approx 2 \Rightarrow p^{\text{th}} \text{ percentile} = x_2$$

Example: consider a sample with data values  
 "27, 26, 20, 15, 30, 34, 28, 25"

compute the 20<sup>th</sup> percentile?

20% Let's use the usual method

1) ترتيب البيانات 15, 20, 26, 26, 27, 28, 30, 34

$$i = \frac{P}{100} \times n = \frac{20}{100} \times 8 = 1.6$$

$i \approx 2$  Round up.

→ The 20<sup>th</sup> percentile =  $x_2 = 20$

(\* Chapter 3

Weighted Data

المطلوب بالسؤال  
 \*  $x_i$  → العينة :  
 $w_i$  → الوزن

(Shift) (mode) (3) (=)

- (mode) (2) (=)

( $x_i$ ) و ( $w_i$ ) (mt)

}

- (Shift) (2) (1) (2) (3) (=)

- group data :

مجموعة | نقطة الوسط  
مجموعة |  $\frac{T}{2}$

- (Shift) (mode) (2) (=)

= (mode) (2) (=)

- (mid) و (F) (mt)

- (Shift) (2) (1) (2) (3) (=)



## • Detecting outliers

- unusual large or small values
- extreme value.

• Determined by the empirical and Z-score Rules

\* Any value less than -3 "Z-score"

\* Any value more than 3 "Z-score"

## \* Exploratory Data analysis

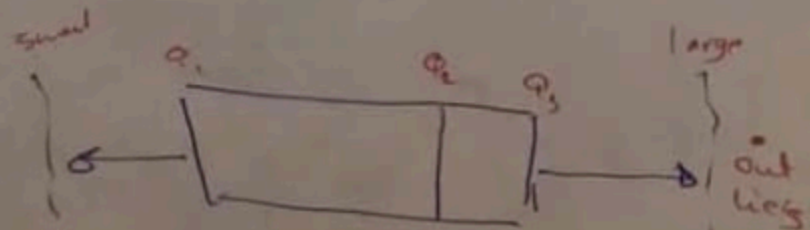
- Five-number summary  $\rightarrow$  small,  $Q_1$ ,  $Q_2$ ,  $Q_3$ , large  
 $Q_2$  is median

### • Box-plot

①  $Q_1$  to  $Q_3$  is range

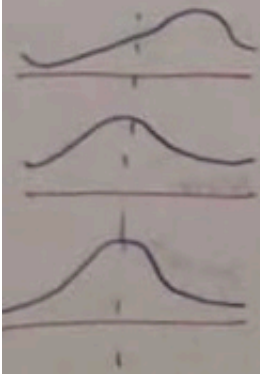
② limitation  $\rightarrow$  upper  $[Q_3 + 1.5(IQR)]$   
lowest  $[Q_1 - 1.5(IQR)]$

③ out lies for values



{3} "distribution shape"

• Histogram → • Skewness → Relationship

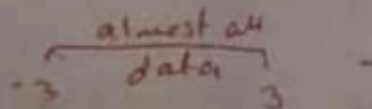
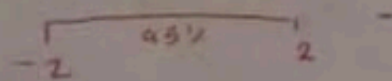
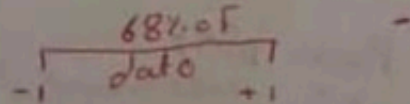


- to left → Mean < median
- to Right → mean > median
- Symtric → mean = median

• Relative location → • Z-score =  $\frac{x_i - \bar{x}}{s}$

• Empirical Rule  
تستخدم في حالة وحدة

- توزيع البيانات على شكل bell-shaped



• استخدام Z-score كما يلي

توضیح :-

\*  $r_{xy}$  between -1 and 1

- 1] values near 1 indicate a strong positive relationship
- 2] values near -1 indicate a strong negative relationship
- 3] values near 0 indicate a weak [positive or negative] relation
- 4]  $r_{xy} = 1$  " " " perfect positive relation
- 5]  $r_{xy} = -1$  " " " negative " "

— perfect: توجہ جمع التعداد بتیجی کا

(19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2)

\* (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2) (19.2.2)



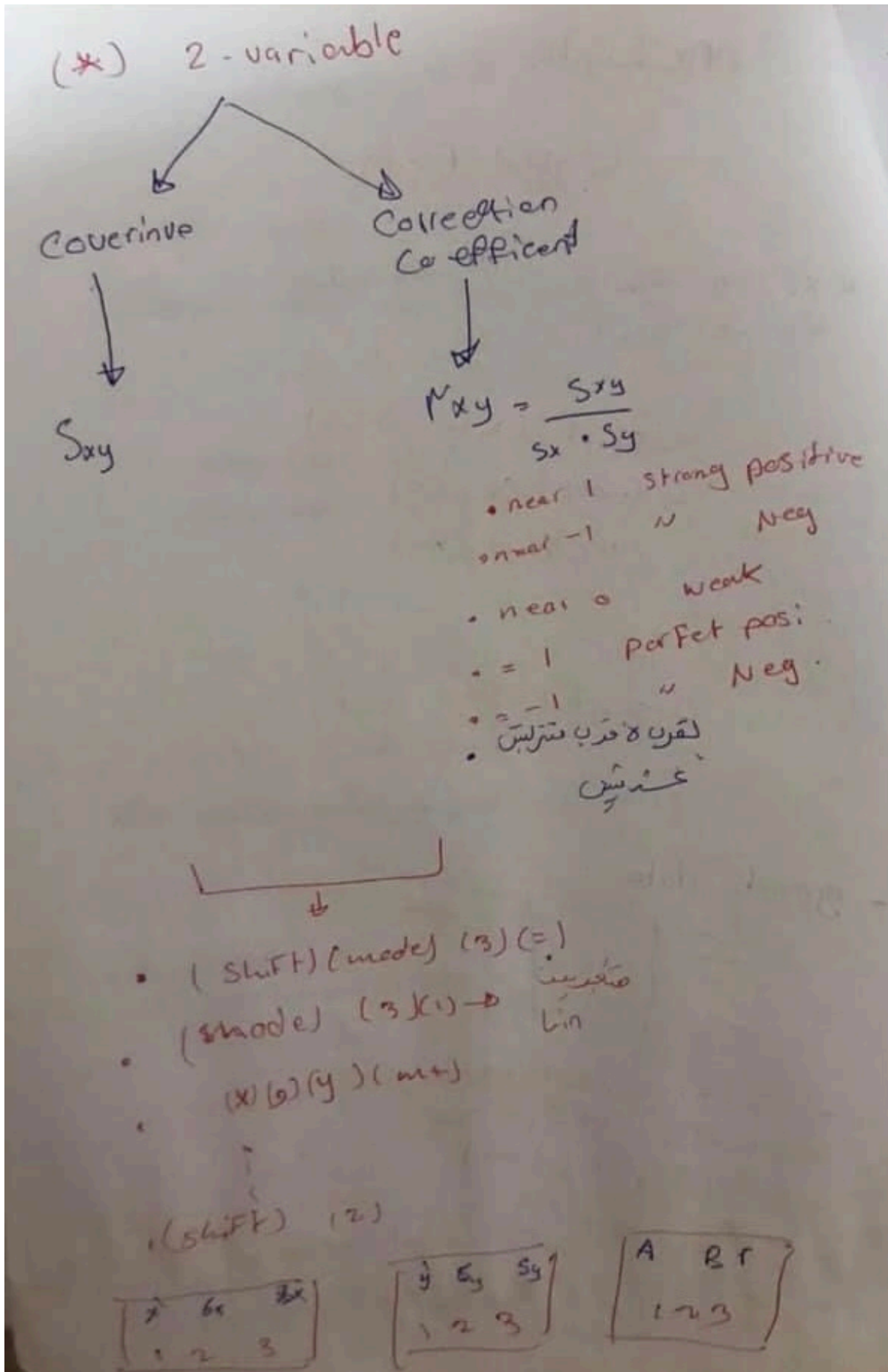
Example: (1) Skewed  
Frequency

$\Rightarrow$  Data are "1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 5, 5"  
 $\Rightarrow$  Data are skewed to Right  
 $\Rightarrow$  Positive skewed

$\bar{x} = \frac{\sum x}{n} = \frac{60}{20} = 3.0$   
 $\bar{y} = \frac{2+2}{2} = 2$

$\Rightarrow$  it provides that  $\text{mean} > \text{median}$  when it is positive skewed.

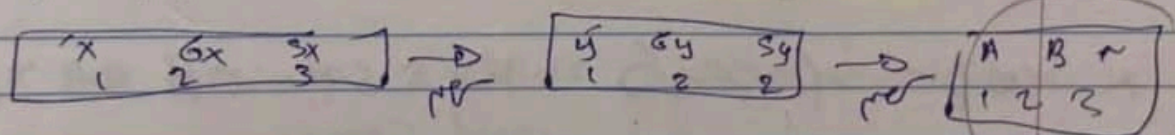
[3] When the skewness is Zero,  
 \* Data are symmetric  
 \*  $\text{mean} = \text{median}$



طريقة التحويل من  $ba$  الى  $ba/ba$

- 1] (Shift) (mode) (3)
- 2] (mode) (3) (1)
- 3] Data  $ba$

4] (Shift) (2)



(A)  $\rightarrow$   $b_0$  (= )  
 (B)  $\rightarrow$   $b_1$  (= )

\* ملاحظة بعد (Shift) (2) بعد التحويل الى  $ba$  بين  $b_1$  و  $b_0$



Ch. 12

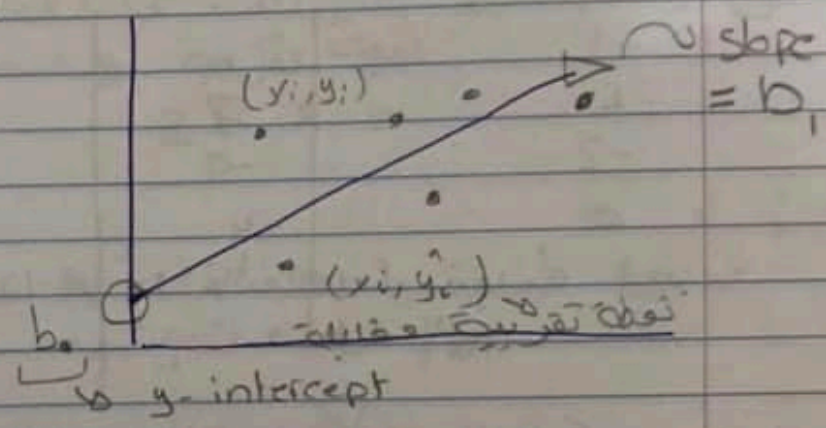
~~Simple Linear Regression~~

Simple Linear Regression  
 $y = a + b_1x$

\* estimated linear regression equation:  
 $\hat{y} = b_0 + b_1x$

example to explain:

x	y
1	3
2	7
3	5
4	11
5	14



### Sc 12.2 Least squares method.

\* least squares method : is a procedure to use data from sample to find the estimated regression equation.

$$\Rightarrow b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\Rightarrow b_0 = \bar{y} - b_1 \bar{x}$$

\* Covariance  $\Rightarrow r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$

يوجد من الآلة

نظروا علاقة جيدة يوجد فيها  $S_{xy}$   
تفاضلي إياها بين الجدول الكبيرة :

$$S_{xy} = (r_{xy}) (S_x) (S_y)$$