STAT2361
Lecture Notes - First

We'am A Rostom

STUDENTS-HUB.com

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 النتيجه بتختلف حسب المطلوب
الي ما بفهم منيح بامكانه يبعتلي مسج افهمة كيف يستخدم الاله الحاسبة بكل مطلوب

برضو اي اشي مش واضح او بتشوفوا ان مش كامل وما وصلتكم فكرته بتقدروا تتواصلوا معي
.....
ملاحظة: تدققوش عالاغلاط الاملائية المهم الاخلاق🤣

وبس
شكرا عفوا

STUDENTS-HUB.com

stat 2361

chapter 1

sc. 1

Data and statistics

section 1 :

- statistics : علوم وفن جمع المعلومات، تحليلها، عرضها وتفسيرها
- Data : الرسوم والحقائق التي يتم جمعها وتفسيرها وتحليلها
- Data set : مجموعة الحقائق التي تخص دراسة معينة

- Elements : الأشياء والكيانات التي تشمل الدراسة عليها
- Variable : الصفات والخصائص التي تقوم بدراستها
- observation : المعلومات القياسية لكل عنصر

section 2 : Scales of measurments

تشير إلى طرق تحليل أو تلخيص البيانات

(1) Nominal scale : تعبر عن أسماء أو صفات بشكل عام سواء كانت البيانات رقمية أو غير رقمية

مثل : Phone number / ID's number / gender

(2) ordinal scale : الأدوار التي يتضمن أسماء أو صفات بشكل عام ولكن الترتيب حسب الأولوية الأهمية كبيرة سواء كانت رقمية أو غير رقمية

مثل : ترتيب الرتب excellent → very good → good

من الأعلى للأقل من الأولوية

STUDENTS-HUB.com

(3) Ratio scale : الجداول التي يمتلك المعرفوها فيها صفر حقيقي
يعني أوقفت صرعي الصفات الرقمية

مثل : ~~[illegible]~~ الراتب / الوقت

(4) Interval scale: جداول ذات قيمة رقمية ذات
معنى مثل درجات الحرارة

* we have 2 Types of Data:

* Quntitative data : (Intetal / Ratio) قيم عددية
* Qualitative data : (ordinal, Nominal) قيم نوعية

section 3: Data Sources

(1) Existing sources : بيانات جاهزة وموجودة مسبقاً

* internal
* external
* internet

(2) Statistical studies :

* observational : ملاحظة ومراقبة
* experiment : القيام بتجربة والاعتماد على النتائج

STUDENTS-HUB.com

* section (4,5) :

Statistic

(1) Descriptive stat. :
وصف البيانات بدلول

(2) Inferenc Stat. :
دراسة مجتمع وتناصره

* population : Collection of all elements
* sample : subjectof population
* Cunsus : survey for all elements

STUDENTS-HUB.com

Chapter 2 :

Descriptive statistic

Qualitative data :

Section 1 :

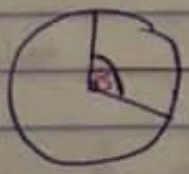
من ناحية الجداول

* Tabular :

a - Frequency distribution (Fd) = التكرار
b - Relative Frequency (RFd) = $\frac{F}{n}$ عدد العناصر
c - percentage Frequency (PFd) = $\left(\frac{F}{n}\right) * 100\ \%$

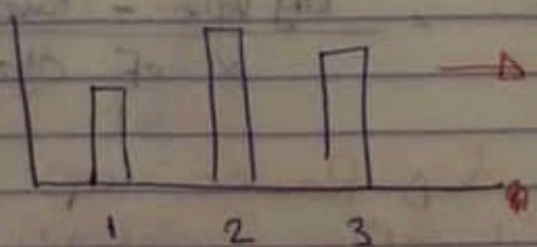
* graphical : من ناحية الرسوم البيانية

(*) piechart



قياس الزاوية
= 360 * RF

(*) bar graph



يتميز بأن الأعمدة فيه منفصلة غير متلاصقة

STUDENTS-HUB.com

* ج نحتي لون ع التكرارات متناحية تراكمية

* متناحية اكبر من: *

- Cumulative Frequency: Which are less or equal the upper class limit

- Cumulative Relevant
- " percentage

* طرق اكل:-

[1] number of classes

(2) width of " = (larg value - smallest value) / * of class

Round up

يعني بقل تقريب مثلاً اذا طلعت النتيجة

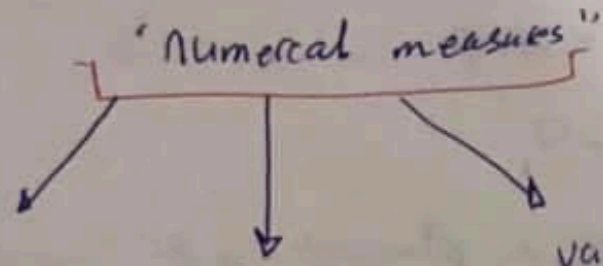
3 ← Round up ← 2.3
4 ← 3.5
3 ← 2.2
3 ← 2.7

فاهمين؟

STUDENTS-HUB.com

"Chapter 3" Sec.1

"Numerical measures"



[1] location:

- *mean
- *mode
- *median : ترتيب تصاعدي
- *percentile : ترتيب تصاعدي
- Quartiles : ترتيب تصاعدي

[1] أول خطوة أرتب تصاعدياً

* mean: average → $\bar{x}$, μ = المجموع / العدد = $\frac{\sum x_i}{n}$

* mode: الأكثر تكراراً

* median: mid point

* percentile → pth : $i = \frac{p}{100} \cdot n$

عدد صحيح = $\frac{x_i + x_{i+1}}{2}$

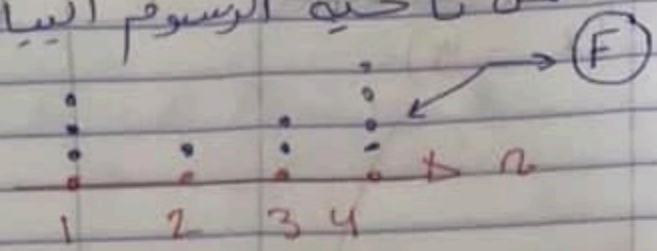
عدد عشري → Round up

* Quartiles → Q_1 = 25%
Q_2 = 50% → median = mid point
Q_3 = 75%

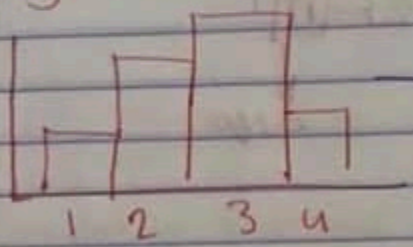
STUDENTS-HUB.com

من ناحية الرسوم البيانية :-

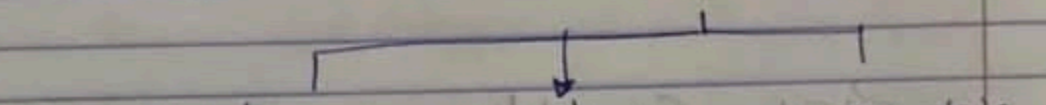
* Dot plot

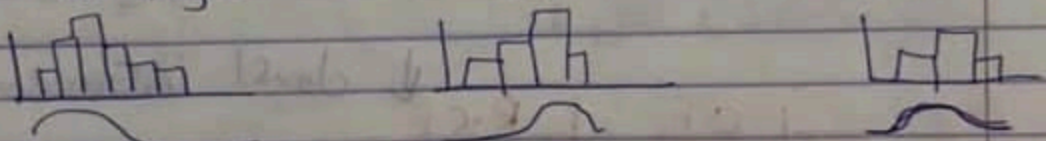


* His togram



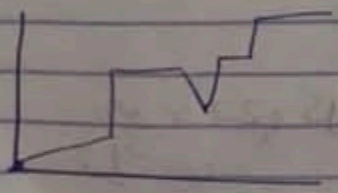
نتميز أن الأعمدة هنا متلاصقة يكون لدي ٣ أشكال





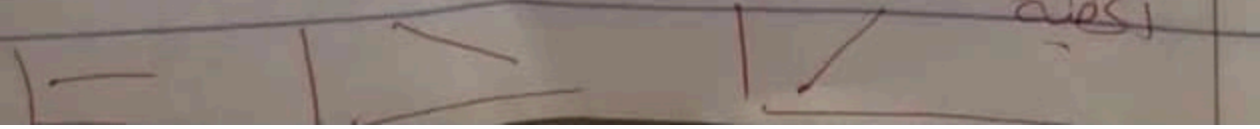
الذيل يتجه لليمين | يتجه الذيل لليسار | متناظر

* ogaive : خاص بالبيانات التراكمية Camulative

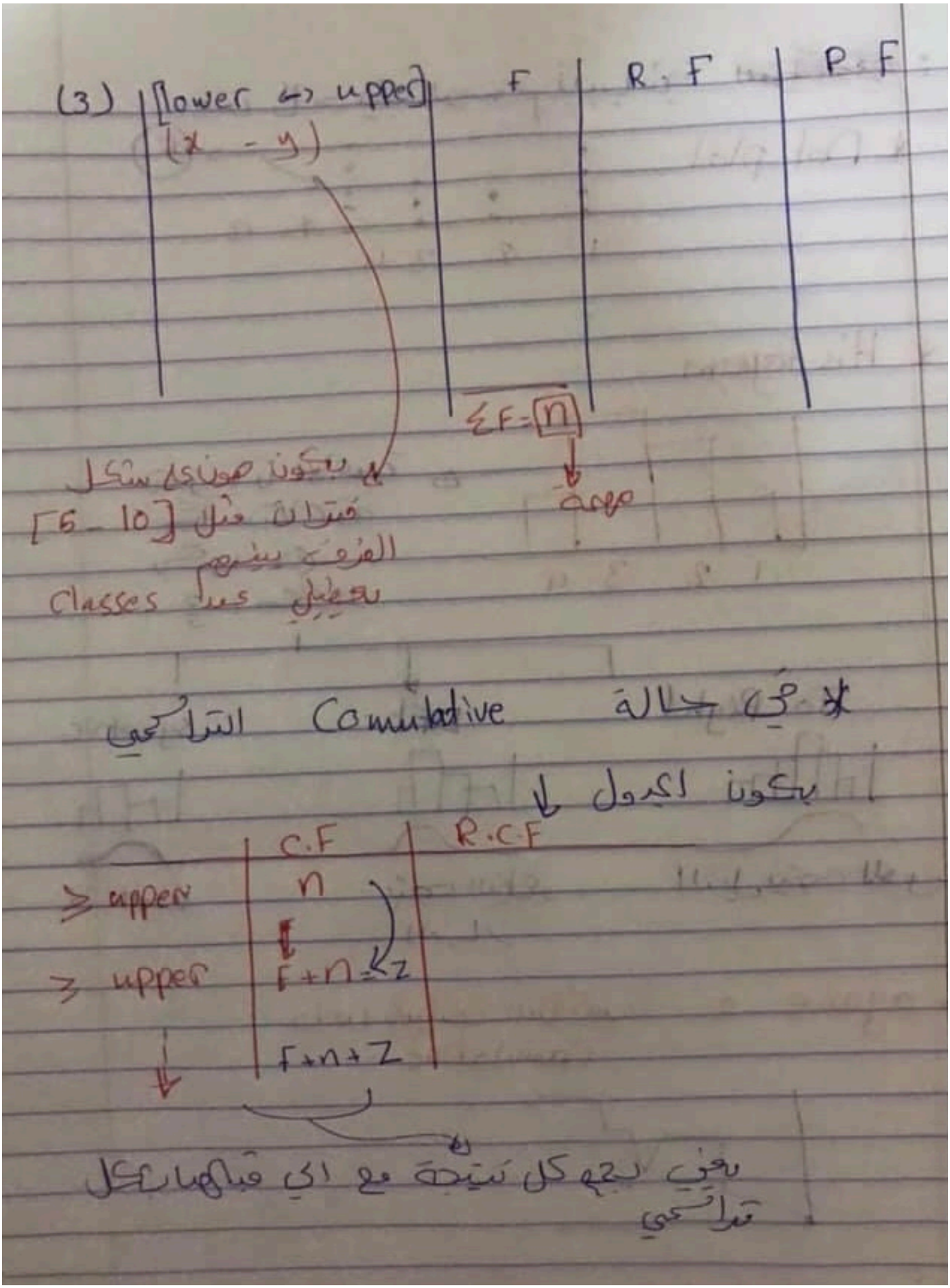


• Gross-Tabulation ← بيانات نوعية وكمية

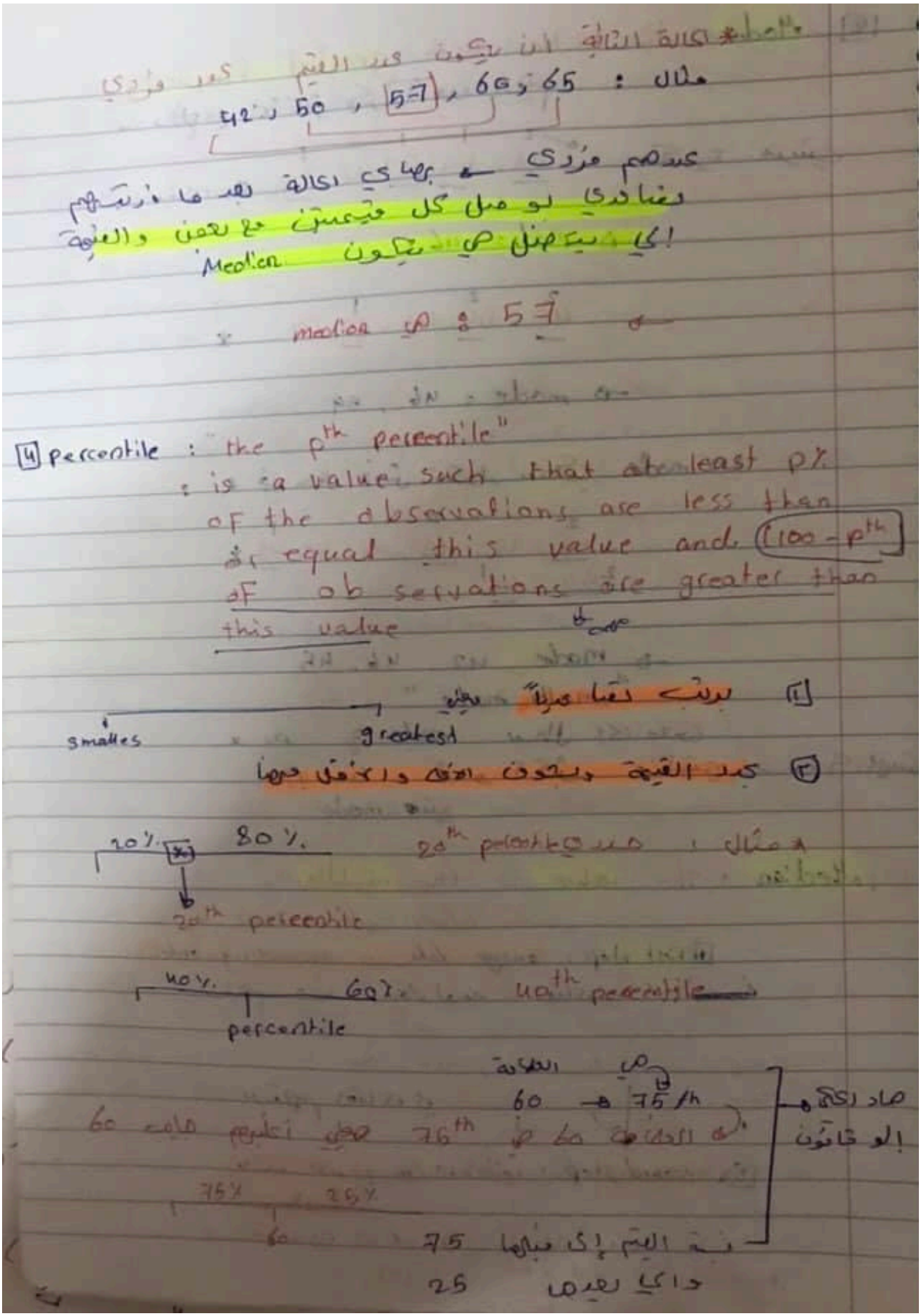
• Satter digram ← لقياس قوة العلاقة بين متغيرين



STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com

[2] Mode : the value with greatest Frequency

- يطلق على اكثر تكرارا من القيم

* مثال: 46, 54, 42, 46, 32 ← مرتبة متكررة

→ mod = 46

* 46, 54, 46, 32, 59

→ mode = 46, 54

* في حال فش تكرار بنحكي No mode

* 46, 32, 42, 46, 42, 48
54, 60

→ mode 42, 46, 45

" multi mode "

* بعد mode 3 بدل اي معنى
يعني من 3 واكثر منحكيهم بس بكتابتش mode

[3] Median : the value in the middle

- اي قيمة موجودة بالنص

[1] Fist step : arrange data in ascending order

- بترتيبهم من الاقل لاكبر تصاعدي / تنازلي

example : 40, 60, 44, 60

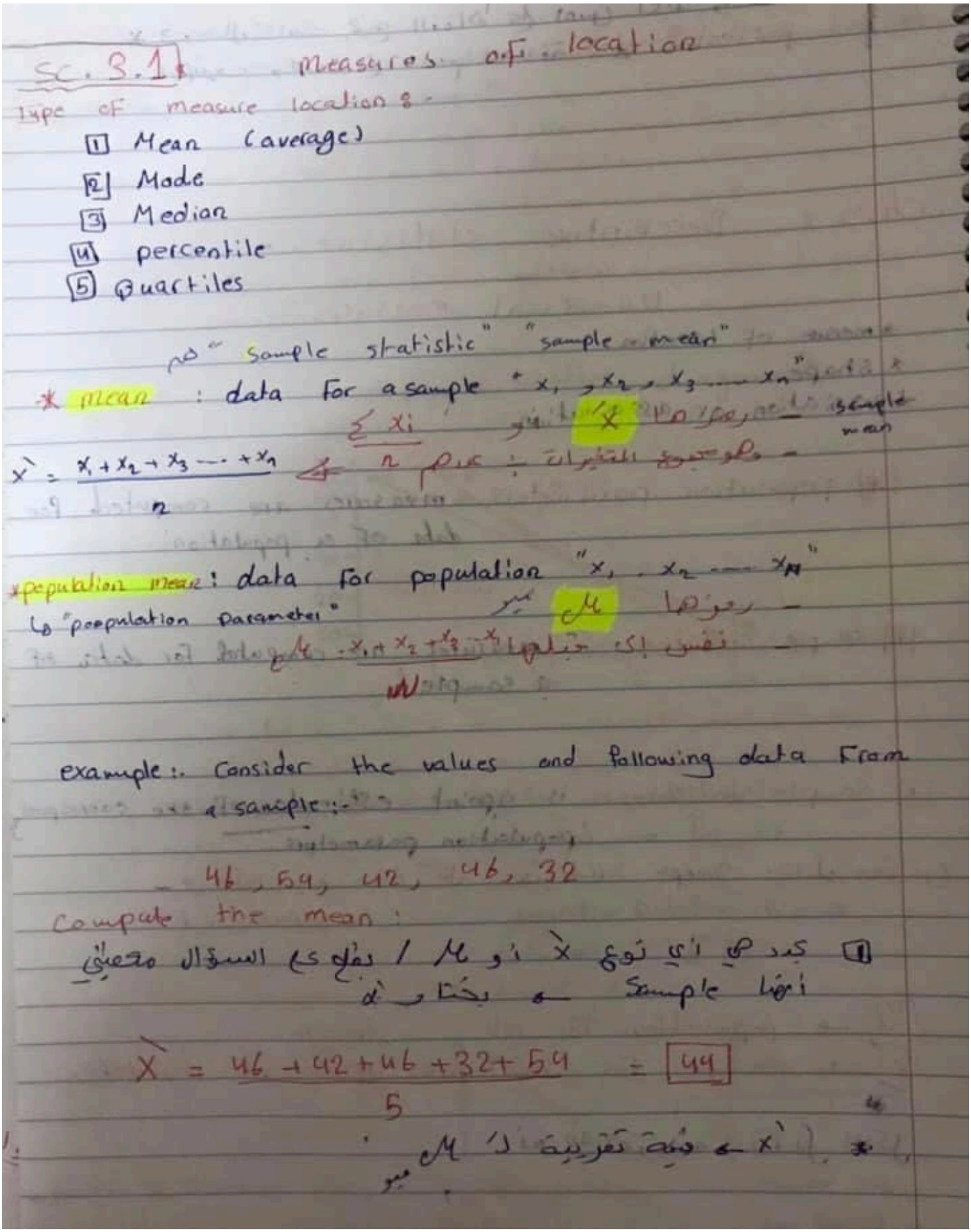
40, 44, 60, 60 [1] بترتيبهم تصاعدي

[2] second step : بأخذ المتوسط بين القيمتين

* اذا كان العدد فردي بأخذ القيمة بالنص مباشرة

median = $\frac{44 + 60}{2}$ = 47

STUDENTS-HUB.com



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 "$x_1, x_2, x_3 \dots x_n$"

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

* population mean : data for population "$x_1, x_2 \dots x_N$"

"population parameter"

$$\mu = \frac{x_1 + x_2 + x_3 \dots + x_N}{N}$$

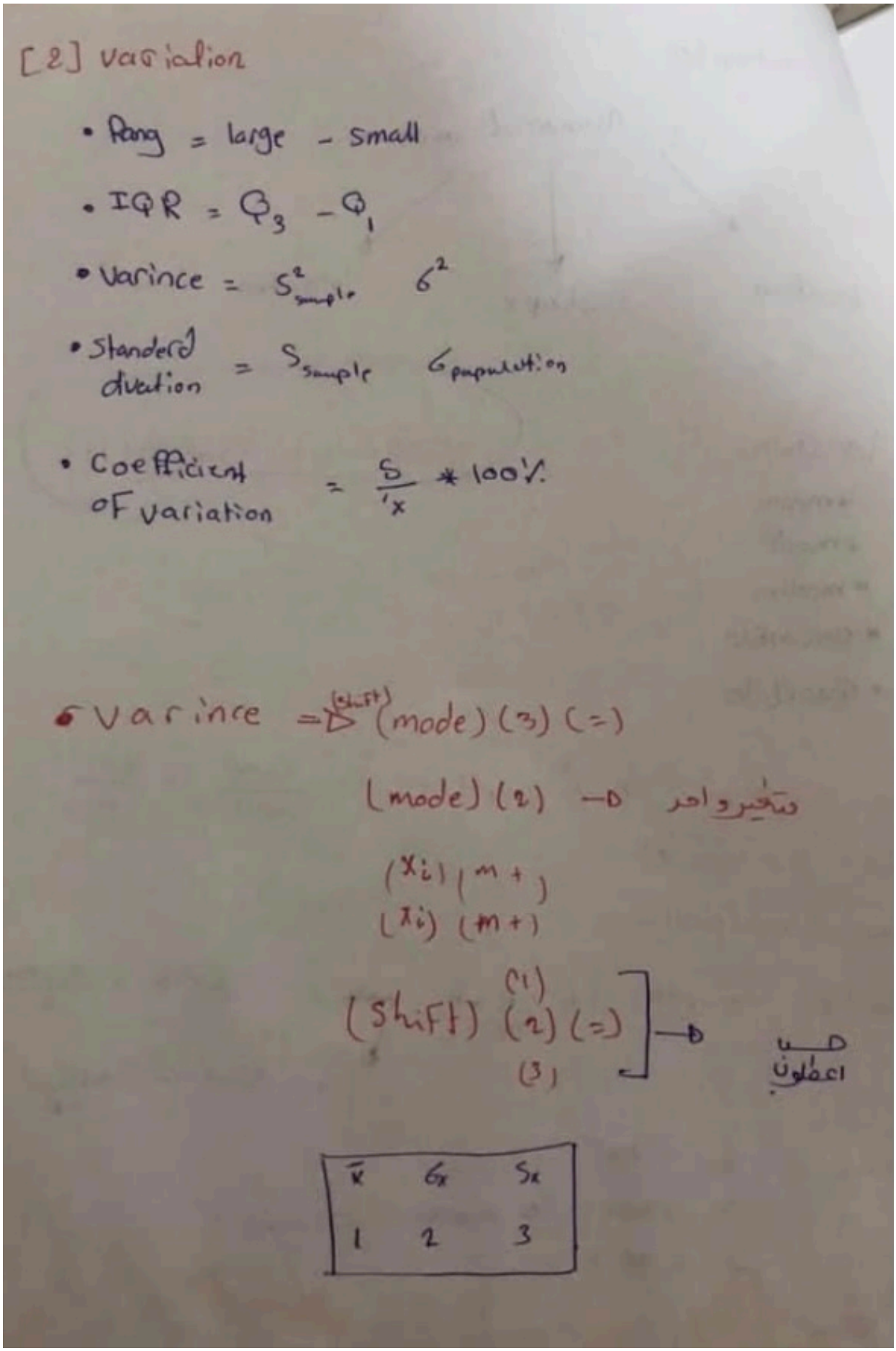
example : Consider the values and following data from a sample :-

46, 54, 42, 46, 32

Compute the mean :

$$\bar{X} = \frac{46 + 42 + 46 + 32 + 54}{5} = \boxed{44}$$

STUDENTS-HUB.com



[2] variation

- Rang = large - small
- IQR = $Q_3 - Q_1$
- varince = S^2_{sample} σ^2
- Standerd diviation = S_{sample} $\sigma_{population}$
- Coefficient of variation = $\frac{S}{\bar{x}} * 100\%$

- varince = (shift) (mode) (3) (=)

(mode) (2) ← متغير واحد

(x_i) $(m+)$

(x_i) $(m+)$

(shift) (1) (2) (3) (=) ← صيغ اعطوني

$\bar{x}$	σ_x	S_x
1	2	3

STUDENTS-HUB.com

→ Solution: * ثم اسبق اني أرتب تصاعدي قبل ما أحل

* $Q_1 = 25^{th}$ percentile & إذا بدو ما حسب

[1] = 15, 20, 25, 25, 27, 28, 30, 34

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

→ 25^{th} percentile $= \frac{X_2 + X_3}{2}$

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

→ $Q_1 = 22.5$

* Q_2 - 50^{th} percentile and median

15 ; 20, 25, 25, 27, 28, 30, 34

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

* $Q_3 = 75^{th}$ percentile

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

$= \frac{X_6 + X_7}{2} = \frac{28+30}{2} = \frac{58}{2} = 29$

$Q_3 = 29$

STUDENTS-HUB.com

→ Computing the p^{th} percentile:

1. arrange data in ascending order

2. $i = \frac{p}{100} \cdot n$

if i is integer → ex: $i = 5$

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

if i is not an integer

ex: $i = 1.5$

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

Example: consider a sample with data values

"27, 25, 20, 15, 30, 34, 28, 25"

Compute the 20th percentile?

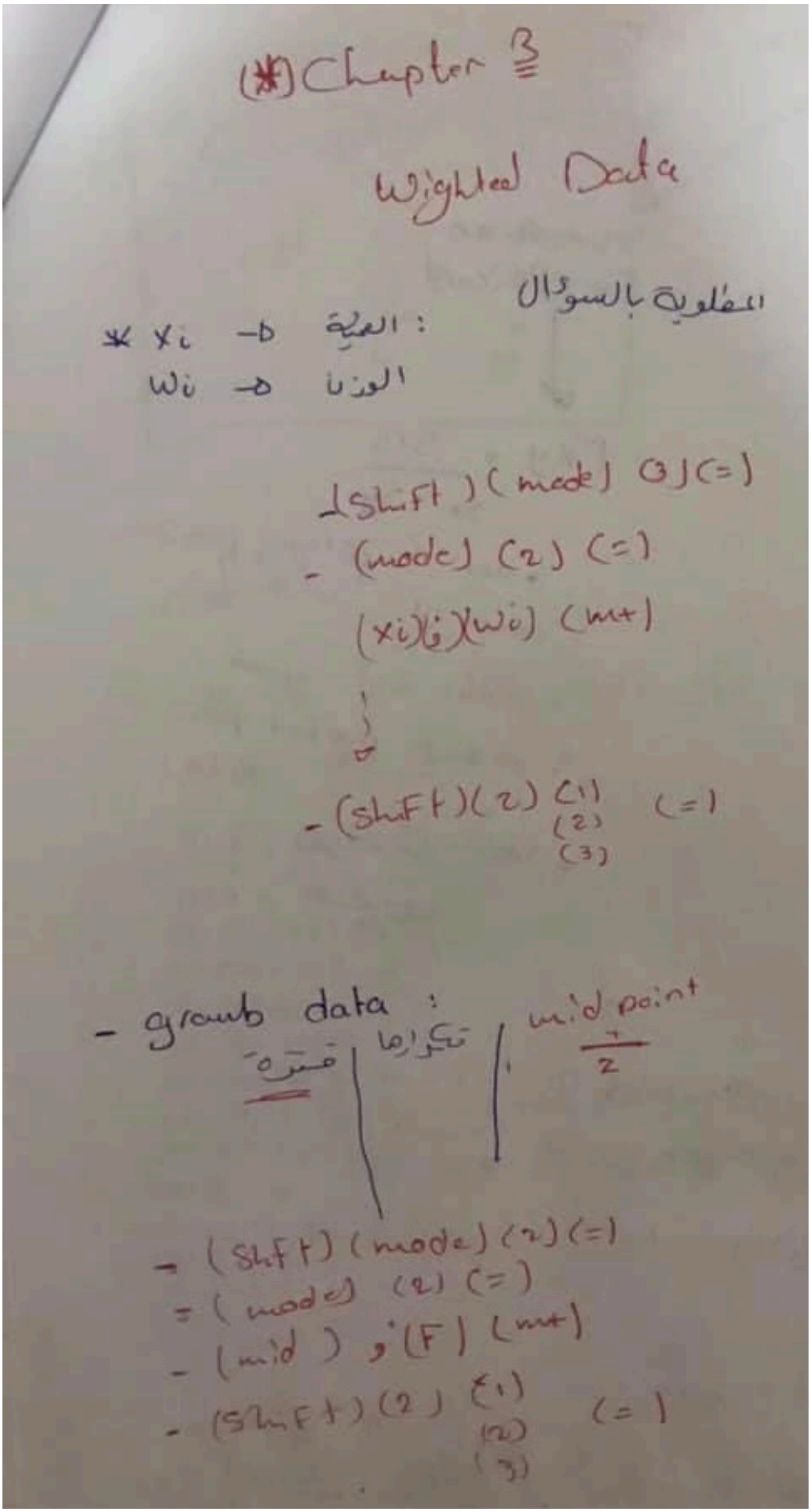
1. 15, 20, 25, 25, 27, 28, 30, 34

2. $i = \frac{p}{100} * n = \frac{20}{100} * 8 = 1.6$

$i \simeq 2$ Round up.

$\therefore$ The 20th percentile $= x_2 = 20$

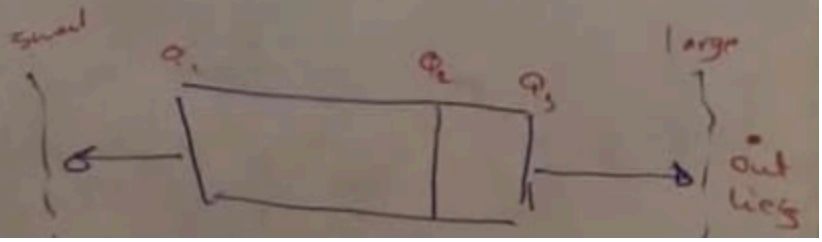
STUDENTS-HUB.com



STUDENTS-HUB.com

- Detecting outliers
  - unusual large or small values
  - extreme value.
- Detemin by the empirical and Z-score Rules
  - *Any value less than -3 "Z-score"
  - *Any value more than 3 "Z-score"

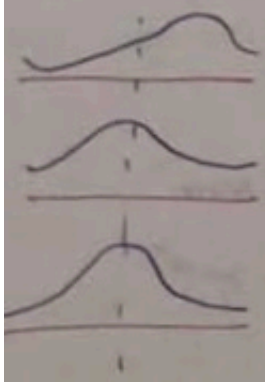
* Exploratory Data analysis
- Five-number-summary → small, Q_1, Q_2 (median), Q_3, larg
- Box-plot
  - Q_1 Q_2 Q_3
  - limition → upper $[Q_3 + 1.5(IQR)]$ lowest $[Q_1 - 1.5(IQR)]$
  - outliers



STUDENTS-HUB.com

[3] "distribution shape"

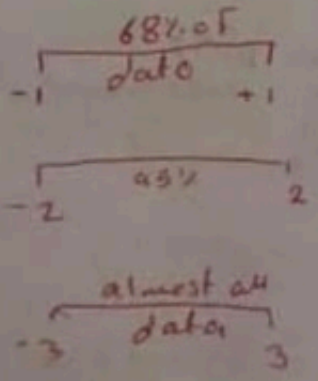
• Histogram → • Sekwenses → Rationship



- to left → mean < mediane
- to Right → mean > mediane
- Symtric → mean = mediane

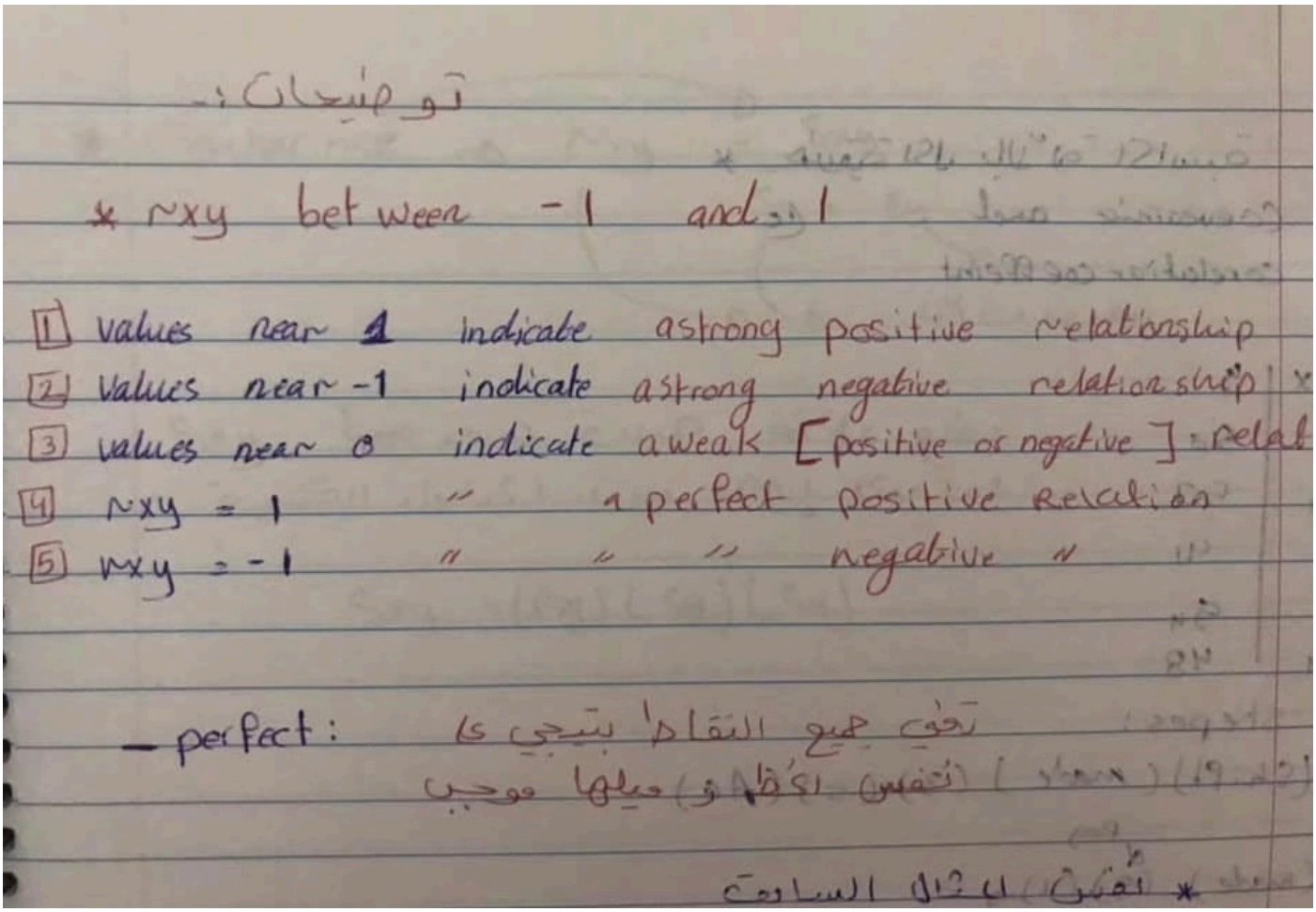
• Relative location → • z-score = $\frac{x_i - \bar{x}}{s}$

• Empirical Rule
- تستخدم في حالة وحدة
- توزيع البيانات على شكل bell-shaped



• يستخدم z-score كما بيّن أعلاها

STUDENTS-HUB.com



تو هنيجان :-

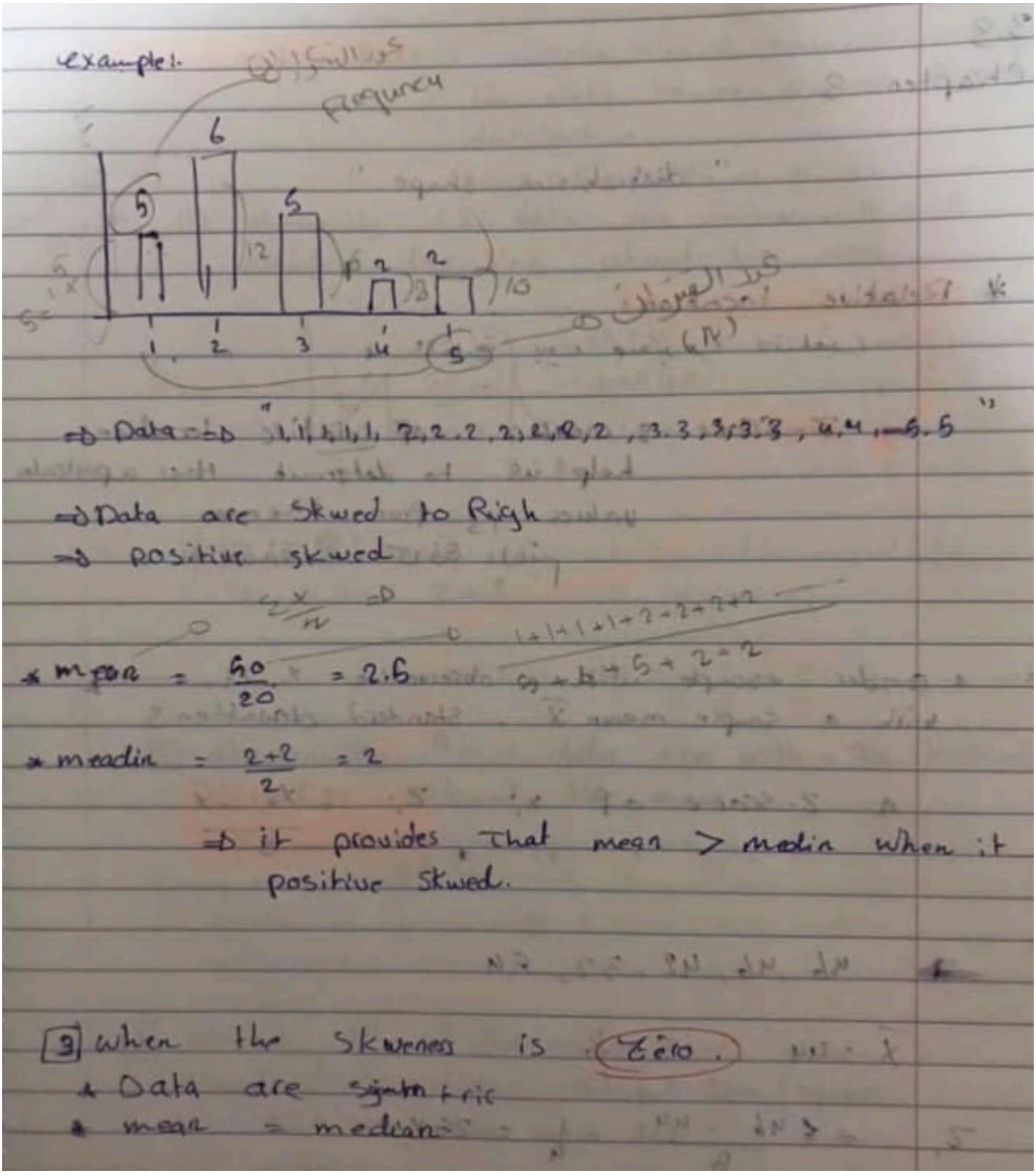
* rxy between -1 and 1

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] relat
4. rxy = 1 " a perfect positive Relation
5. rxy = -1 " " " negative " "

- perfect: تحقق جميع النقاط بتيجي ع

* تحقق ا ؟ال السابق

STUDENTS-HUB.com



⇒ Data ⇒ 1,1,1,1,1, 2,2,2,2,2,2,2, 3,3,3,3,3, 4,4, 5,5

⇒ Data are skwed to Righ

⇒ positive skwed

$\Sigma x / N \Rightarrow$

$1+1+1+1+2+2+2+2+2+2+2+3+3+3+3+3+4+4+5+5$

* mean $= \frac{50}{20} = 2.5$

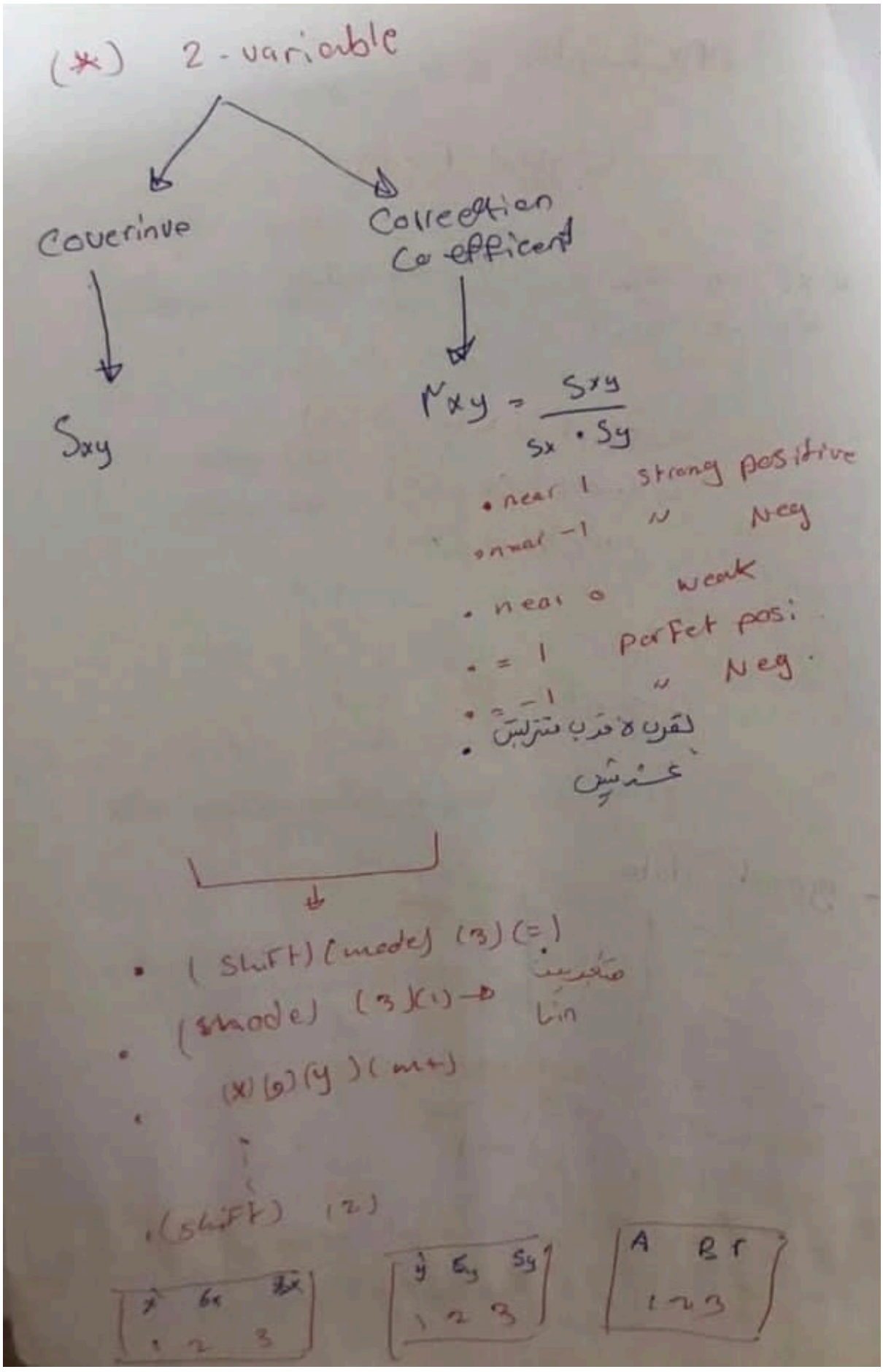
* median $= \frac{2+2}{2} = 2$

⇒ it provides that mean > median when it positive skwed.

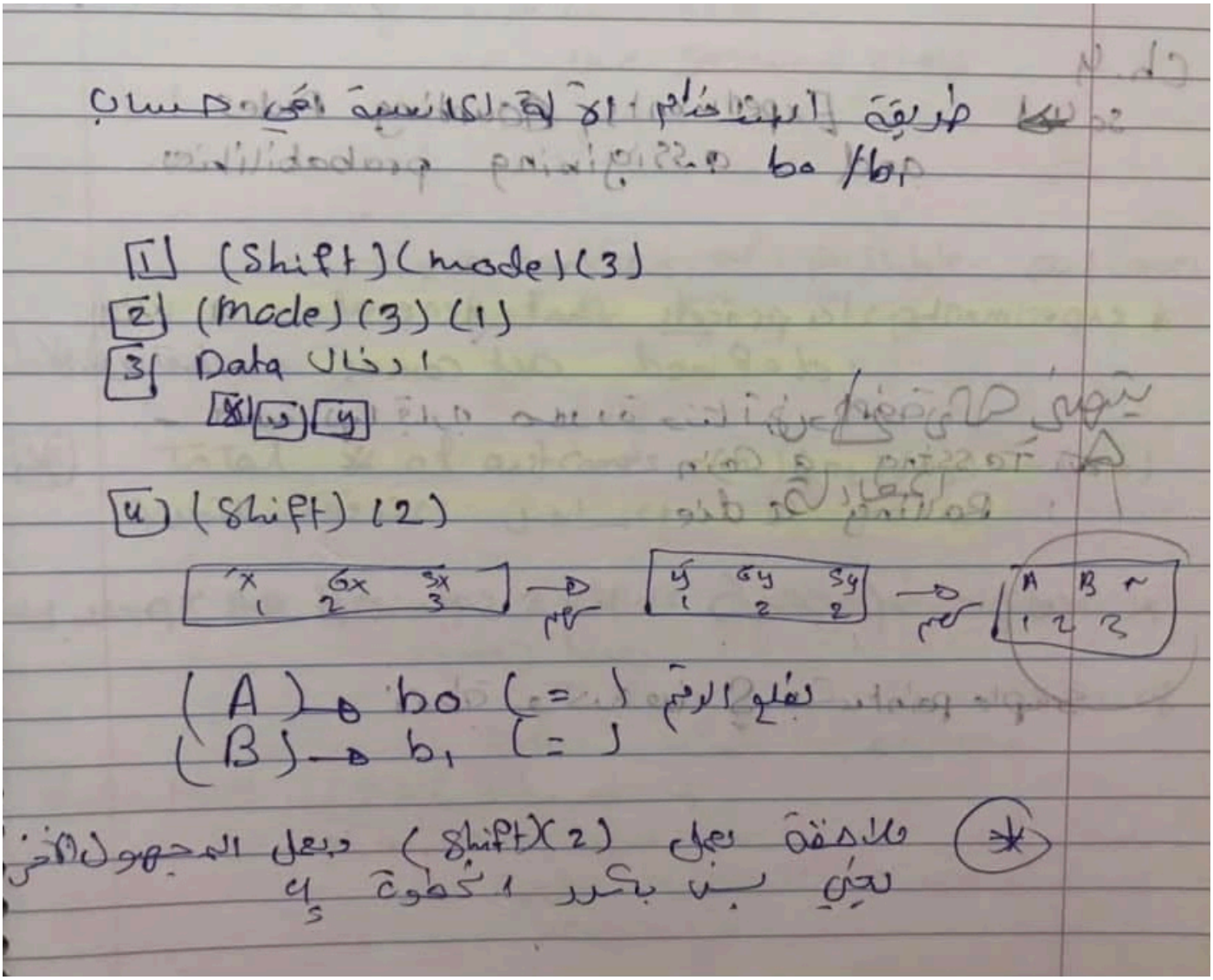
[3] when the skweness is Zero.

* Data are symmetric
* mean = median

STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com

Ch. 12

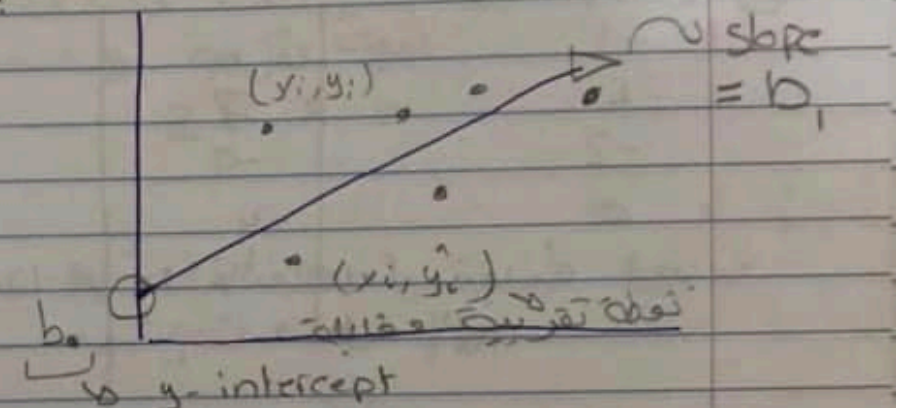
Simple linear Regression

* estimated linear regression equation:

$$\hat{y} = b_0 + b_1 x$$

example to explain:

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



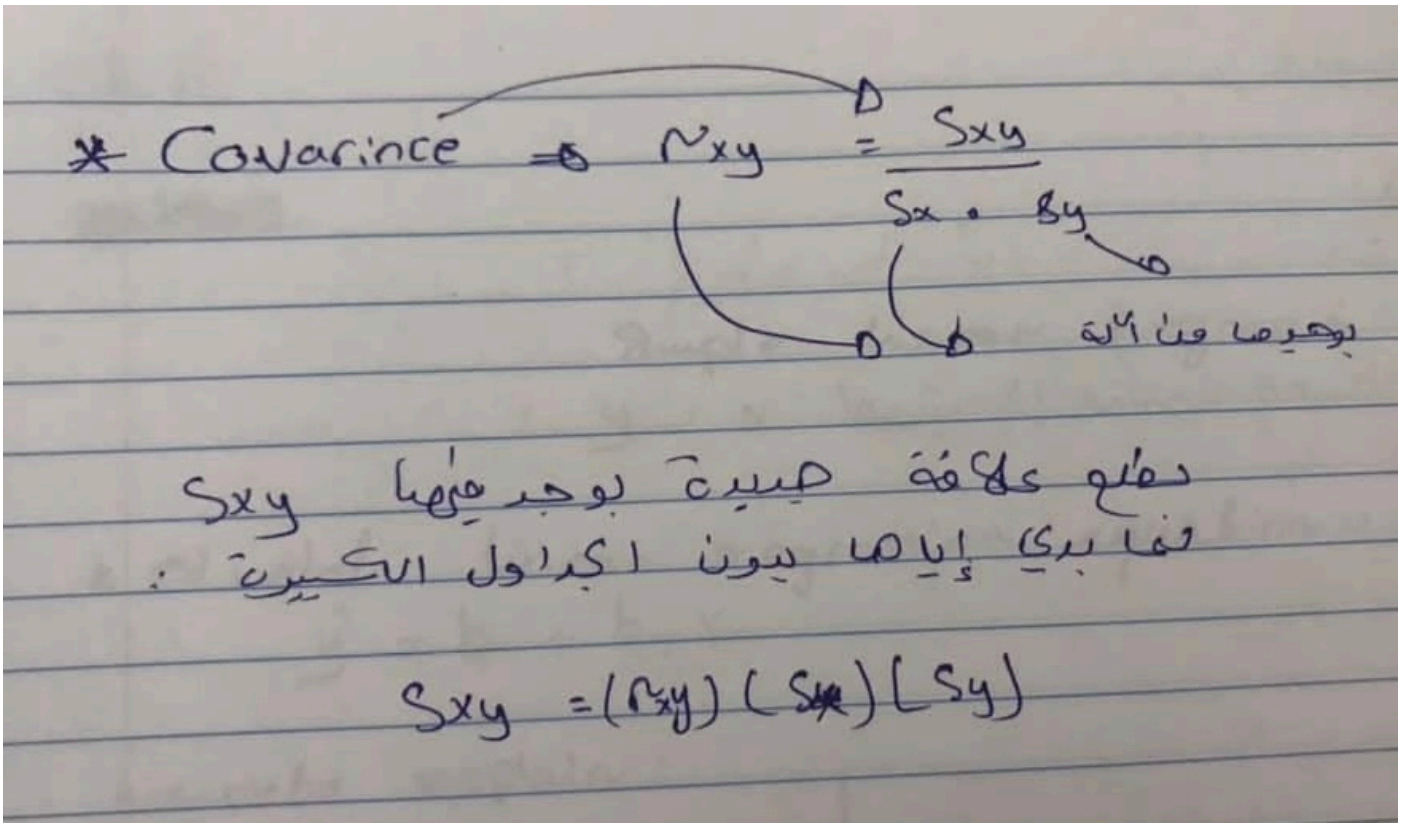
Sc 12.2 Least squares method.

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

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

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

STUDENTS-HUB.com



STUDENTS-HUB.com