احصاء: ولاء ياسين

*chapter 1+2 :- statistics

تعريف الاحصاء

① statistics: The science of collecting, organizing, presenting, summarizing, and analyzing data to make conclusions and better decisions.

← الإحصاء: علم يهتم في جمع، تنظيم، تمثيل، تلخيص، وتحليل البيانات للوصول الى نتائج وقرارات أفضل.

② Variable: The characteristic of interest that assumes different values.

← المتغير: الصفة أو الخاصية المراد دراستها وتأخذ قيم مختلفة.

③ Data: The values or ~~xxxx~~ measures (numbers or labels) that the variable assumes.

← البيانات: هي القيم أو القراءات سواء كانت (أرقام أو وصوف) التي يأخذها المتغير.

مثال:

-EX: suppose we want to study the gender of students stat 2371 section 2.

→ Variable: Gender ← الجنس

Data: Male, male, female, ...

ذكر أنثى

←

* Two types of variables:-

a) Qualitative: variables whose values are names or labels or categorise (not actual numbers).

المتغيرات النوعية : هي المتغيرات التي تأخذ قيم على شكل مسميات ، تصنيف ، أو فئات (ليست أرقام فعلية).

- EX: (Qualitative variables):

① Gender: Male, Female
② Blood Type: A, B, AB, O
③ clothes size: small, medium, large, ...

b) Quantitative: variables whose values are actual numbers.

المتغيرات الكمية العددية : هي المتغيرات التي تأخذ قيم على شكل أرقام فعلية

- EX: (Quantitative variables).

① family size: 1, 2, 3, ... حجم العائلة

② # of science students: 0, 1, 2, ... عدد طلاب العلوم

③ Age: 20, 25.5, 33, ...

④ Weight: 52, 65, 80, ... الوزن

* Note:

~~sam~~ sometimes, we assign numbers to qualitative variabl
still, these numbers are Just labels.
أحياناً نقوم بتحديد أرقام للمسميات لتسهيل التعامل معها ولكن تبقى مسميات.

- EX: we can assign 0 to female and 1 to ~~male~~ male
for the gender only for facilitation purpose.

- EX: student ID, Bank account number.
(الرقم الجامعي، رقم الحساب البنكي) هي $\boxed{\text{Qualitative}}$

* Remark:

Quantitative variables are either discrete or continuous.
منفصل مستمر

① Discrete variables:
منفصلة

- are quantitative whose values are finite or can be counted. (Usually take no decimals or fractions)

المتغيرات المنفصلة : متغيرات كمية يمكن عدها ، عالأغلب لا تأخذ أعشار.
نعدها لأنها محددة

② Continuous variables:

are quantitative whose values can't be counted.
(Usually take decimals or ~~the~~ fractions).

المتغيرات المتصلة أو المستمرة : متغيرات كمية لا يمكن عدها ، عادة تأخذ أعشار.

EX:

عدد أفراد العائلة

① family size: Discrete variable

② Number of prisoners: Discrete variable

عدد طلاب العلوم

③ Number of science students: Discrete variable

④ Age: continuous variable.

الوقت

⑤ Time: continuous variable.

درجة الحرارة

⑥ Temperature: continuous variable.

⑦ weight: continuous variable.

ضغط الدم

⑧ Blood pressure: continuous variable.

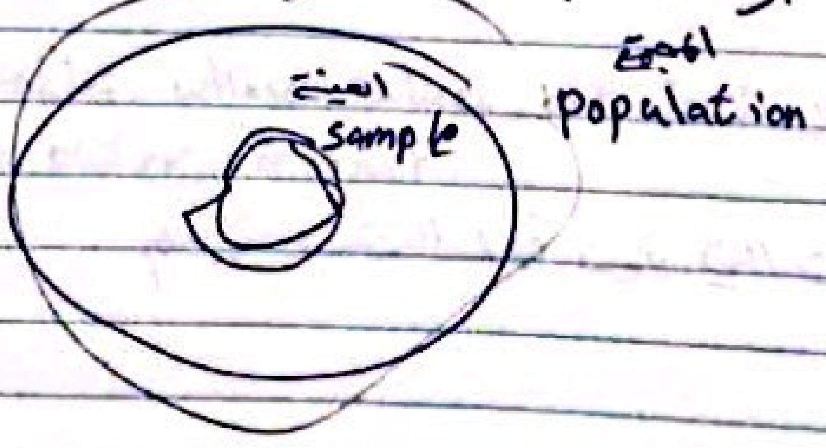
الطول

* Population: All subjects or elements of interest

المجتمع الإحصائي: جميع عناصر الدراسة

* sample: A subest or part of the population

العينة: جزء من المجتمع الإحصائي



- Note:

نستخدم العينة لتوفير الوقت والجهد

we use samples to save time, cost, effort, ...

* Parameter: Any characteristic or measure obtained from a population.

متوسط

(EX: population mean)

* Statistic: Any characteristic or measure obtained from a sample.

متوسط

(EX: sample mean)

~~1, 5, 2, 2.5~~

1, 1.5, 2, 2.5, 3, 3.5, ..., 1

متصلة

* Two types of statistics: مجالات علم الإحصاء

① Descriptive statistics: الاحصاء الوصفي

Consists of the organizing, presenting, and summarizing of data using tables, graphs and numerical measures

تنظيم البيانات وتمثيلها باستخدام الجداول والرسومات ، وأرقام عددية تعرضها.

ch4-6 سنتطرق لهذا النوع في ‹‹

② Inferential statistics: الإحصاء الاستدلالي

Concluding a characteristic or prediction of the population based on samples. It includes point and interval estimation and hypotheses testing.

بناء توقعات عن المجتمع كامل من خلال العينات ، يتضمن فترات التقريب واختبار الفرضيات . ‹‹ سنتطرق لهذا النوع في 12-11 ch .

* Levels (scales) of measurements: مستويات القياس

① Nominal level: المستوى الاسمي

Includes variables whose values are Just names or Labels with no order or rank.

يتضمن المسميات التي يكون فيها الترتيب عديم

_EX: (Nominal scale)

Gender , Religion , Eyes color , ...
الجنس الديانة لون العيون

② ordinal scale: المقياس الترتيبي

Includes variables whose values are names or labels that can be ~~ord~~ orderd or ranked.

يعني المتغيرات التي تكون قيمها الترتيب لها.

_EX: (ordinal scale)

. Letter grades: (A, B, C, ...) العلامات بالاحرف

. service ~~satisf~~ satisfication: (strongly approve, A

. Interest in football: (High, Medium, low)
الاهتمام بكرة القدم

③ Interval scale: مقياس الفترة (المسافة)

Includes variables whose values are actual numbers.
المتغيرات التي تكون قيمها ارقام فعلية.

_EX: (Interval scale)

family size , Age , weight , salary , IQ test , ...
أفراد العائلة العمر الوزن الراتب اختبار الذكاء

- Note:

- Variable
  - Qualitative
    - scales
      - Nominal
      - ordinal
  - Quantitative
    - scale
      - Interval

* Outliers (Extreme values): القيم الشاذة أو المتطرفة

are extremely high or extremely low data values, if any, compared to the rest of the data values.

القيم المختلفة عن البيانات (بعيدة أو قريبة جداً)

- Ex: consider the following data:

65, 72, 53, 67, 71, 130, 58.

→ 130 is an outlier (extreme value).

* Ch4: Frequency distributions and graphs.

- Descriptive statistics: الإحصاء الوصفي

| organizing data

tables — graphs — numerical measures

جداول — رسومات — أرقام عدديه

. Using tables:- (Qualitative variables)

① Frequency distribution: (frequency table) الجدول التكراري أو التوزيع التكراري

A table that organizes the data of a variable int
classes and frequencies.

الفئة
* Class: A qualitative category of a variable
التكرار
* Frequency: The number of data values of a var

- Ex: The following data represents the blood type
of 25 students:-

O, A, O, O, O, B, A, AB, O, O, B, B, A, AB, O, C

A, O, O, O, A, B, O, AB, AB.

a) Construct a frequency table.

class	freq
A	5
B	4
AB	4
O	12

- Notes: ① the table consists of 4 classes.

  \# of classes = # of categories (disjoint).

  ② total frequency " $\sum F$ " = n جموع التكرارات

② Relative frequency tables: جدول التكرار النسبي

Classes vs relative frequency.

→ Relative frequency "R.F" = $\frac{F}{n}$, n = total size, F = frequency

n ← حجم العينة

- Ex: for the previous example:

b) Construct a relative frequency distribution.

class	freq	relative freq
A	5	$\frac{5}{25} = 0.2$
B	4	$\frac{4}{25} = 0.16$
AB	4	$\frac{4}{25} = 0.16$
O	12	$\frac{12}{25} = 0.48$

→ $r.f = \frac{F}{n}$

$\frac{25}{25} = \boxed{1}$

- Notes: ① $\sum R.F = \underline{1}$ "always"

② The relative frequency represents a probability.

③ Percent frequency tables: الجدول المئوي

Classes vs percent relative frequency.

⟶ Percent relative frequency "P.R.F" $= \frac{F}{n} \times 100\%$

$= R.F \times 100\%$

- EX: for the previous example.

c) Construct a percent frequency distribution.

Class	freq.	relative freq.	percent relative freq. / P.F
A	5	0.2	20%
B	4	0.16	16%
AB	4	0.16	16%
O	12	0.48	48%

d) what is the number of students whose blood type O? 12

e) How many students with blood type A or B? 9

أو

تقاطع

f) How many student with blood type A and B? 0

النسبة المئوية → P.R.F

g) what is the percentage of students whose blood type is AB? 16%

= P.R.F

الاحتمال

h) what is the probability that a student has blood type B?

= R.F

0.16

- Notes: ① $\sum$ P.R.F = 100%
② P.R.F = R.F × 100%

- Exercise:

The following data represent the size (small = S, M = medium, L = Large).

M, M, S, L, L, S, S, S, S, M, L, M, M, M, L, L, L, L, L, S.

Construct a frequency, relative frequency, percent frequency tables for these data.

classes	f.	r.f	p.f
S	6	$\frac{6}{20} = 0.3$	30%
M	6	$\frac{6}{20} = 0.3$	30%
L	8	$\frac{8}{20} = 0.4$	40%

الرسوم البيانية

• Using graphs (Qualitative variables).

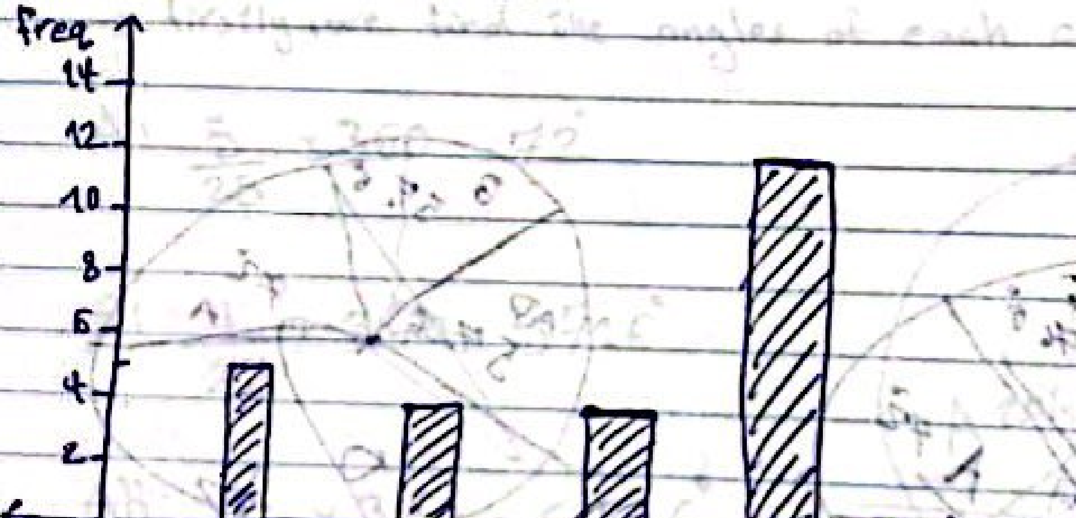
① Bar graph طريقة التمثيل بالأعمدة

→ horizontal axis: classes المحور الأفقي: الفئات

→ vertical axis: frequency, relative frequency, or percent frequency.

المحور العمودي: التكرار أو التكرار النسبي أو المئوي.

- Ex: construct a bar graph for the blood type (previous example).



* Note: the bars should be separated to emphasize the fact each class is separate. الأعمدة منفصلة

② Pie chart: طريقة القطاعات الدائرية

التكرار النسبي

we draw a circle (دائرة) and we use the P.R.F to subdivide the circle into sectors or parts (قطاعات او اجزاء)

→ the angle (الزاوية) of each sector $= \frac{f}{n} \times 360$, f: freq, n: total

_EX: Construct a pie chart for the blood example

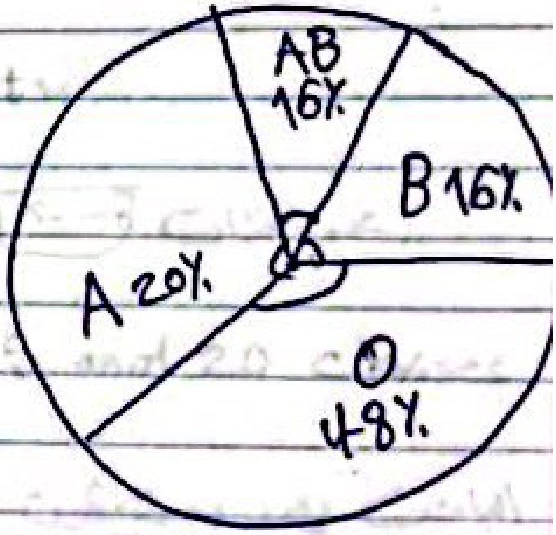
→ firstly, we find the angles of each class.

A: $\frac{5}{25} \times 360 = 72°$

B: $\frac{4}{25} \times 360 = 57.6°$

AB: $\frac{4}{25} \times 360 = 57.6°$

O: $\frac{12}{25} \times 360 = 172.8°$ or



* Note: $\sum$ angles $= 360°$

. Using tables (Quantitative variables).

① Frequency distribution:

is a tabular summary of data showing the number of items in each nonoverlapping classes.

كفئات منفصلة (غير متداخلة)

عشان نعرف شو عدد الفئات يلزمنا ٣ خطوات

3 steps to ~~dete~~ define the ~~classes~~ Classes:

يلزمنا عدد الفئات — معطى

a) Determine the number of classes. (given)

وبعدين نحدد طول الفئة

b) Determine the width of each class.

c) Determine the class limits.

→ Number of classes:- عدد الفئات $\{5 - 20\}$

Usually we use between 5 and 20 classes

الباحث يقدر يحدد بحيث يكون ملائم للبيانات ويختار المتوفر بينهما . ويرجع عدد الفئات الى الدراسة.

طول الفئة — البيانات الكبيرة — الصغيرة

$$\text{Class width} = \frac{\text{Largest data value} - \text{smallest data value}}{\#\ \text{of classes}}$$

عدد الفئات

the approximated class width can be rounded to a more convenient value.

$9.2 \xrightarrow{\text{round up}} 10$

lower limit

class width 10

70 - 79 → upper limit → class

80 - 89

90 - 99

→ Class limits: L - U - f ... class حدود الفئة

نختار النقطة البداية بحيث تحتوي وتغطي البيانات كلها عادة " بأصغر رقم أو قريب عليه.

lower limit حد سفلي , upper limit حد علوي

Ex: the following frequency table represents the ages of 40 BZU employees.

Class	f	mid point
22-30	10	26
31-39	8	35
40-48	7	44
49-57	9	53
58-66	6	62

$\frac{31+39}{2} = 35$

$\Sigma f = n = 40$

Note: ① the table consists of 5 disjoint classes: 22-30, 31-39, ...

② Each class consists of 2 limits: lower limit and upper limit

③ Each class has a mid point نقطة المنتصف / مركز الفئة

$$\text{midpoint} = \frac{\text{Lower Limit} + \text{upper Limit}}{2}$$

9 { 22 - 30
31 - 39
9 { 40 - 48

30 - 22 + 1 = 9

④ Class width is 9 = U - L + 1

30 - 22 + 1 = 9

⑤ The total frequency = n = 40

- EX: The following data represents the weights (kg) of 16 BZU students

52, 60, 79, 60, 71, 45, 64, 84, 74, 80,
59, 91, 50, 60, 76, 75.

Construct a frequency table. (use 4 classes)

→ # of classes = 4 ⟶ (given)

$$\text{class width} = \frac{\text{Largest data value} - \text{smallest data value}}{\#\text{ of classes}}$$

$$= \frac{91 - 45}{4} = 11.5 \xrightarrow{\text{round up}} 12$$

→ constructing the classes by finding the Lower limits starting from the smallest data valual.

Classes		frequency
45	- 56	3
57	- 68	5
69	- 80	6
81	- 92	2

② Relative frequency distribution: التكرار النسبي

→ $R.f = \frac{f}{n}$, f: frequency التكرار , n: toto size

③ Percent relative frequency distribution: التكرار المئوي

→ $P.r.f = \frac{f}{n} \times 100\%$

④ Cumulative frequency distribution: التكرار المتجمع الصاعد أو المتجمع الصاعد

the cumulative frequency distribution shows the number of data item with values less than or equal to the upper limit of each class.

عدد البيانات الأقل أو تساوي الحد العلوي للفئة.

- EX: Construct a relative frequency, percent frequency and cumulative frequency tables for the previous example.

class	freq	R.f	P.f	C.f
45 - 56	3	$\frac{3}{16}$ = 0.1875	0.1875 × 100% = 18.75%	3
57 - 68	5	$\frac{5}{16}$ = 0.3125	31.25%	8
69 - 80	6	$\frac{6}{16}$ = 0.375	37.5%	14
81 - 92	2	$\frac{2}{16}$ = 0.125	12.5%	16 = n

التكرار التراكمي   فقط هو

* Note: ① the cumulative frequency is only for quantitative variable.

الكمي   متغير

[illegible]

② The Last Cumulative frequency = n

_EX: ~~The~~ The following data represents the numb ~~er~~ of credit hours of 15 BZU students.

94 // 65 // 84 // 32 // 23 // 27 // 32 // 80 // 81 // 50 //
80 // 56 // 30 // 20 // 29.

f   r.f   c.f

a) Construt a frequency, relative frequency, Cumulati frequency for these data (use 4 classes)

→ # of classes = 4

→ class width = $\frac{\text{Largest data value} - \text{smallest value}}{\text{\# of classes}}$

$= \frac{94-20}{4} = 18.5 \rightsquigarrow \underline{19}$

at ~~least~~

at least على الأقل   between بين

at most على الأكثر

more than أكثر من   upper+

classes	f	$R.f = \frac{f}{n}$	$P.f = \frac{f}{n} \times 100\%$	C.f	midpoints
20_38	7	0.467	46.7%	7 +2	29 +19
39_57	2	0.133	13.3%	9 +1	48 +19
58_76	1	0.067	6.7%	10 +5	67 +19
77_95	5	0.333	33.3%	15	86

منتصفات

b) find the midpoint for each class.

النسبة المئوية

c) what is the percentage of students whose number of credt house at least 39 ?

على الأقل

ما نسبة الطلاب الذين عدد ساعاتهم

على الأقل 39 ← يعني 39 أو أكثر

$13.3\% + 6.7\% + 33.3\% = 53.3\%$

d) ~~what is the percentage of students whose number of credit hours at~~

احتمال

d) what is the probability that astudent whose number of credit hours is between 38 and 77?

أن

ما احتمال الطلاب الذين عدد ساعاتهم

المطلوبة بين 38 و 77

$0.133 + 0.067 = 0.2$

f

e) what is the number of students whose number of credit hours more than 57 ?

أكثر من

ما عدد الطلاب الذين عدد ساعاتهم

المطلوبة أكثر من 57

$1 + 5 = 6$

at most 57

على الأكثر

* Using graphs (Quantitative variables):

a) Histogram: المدرج التكراري

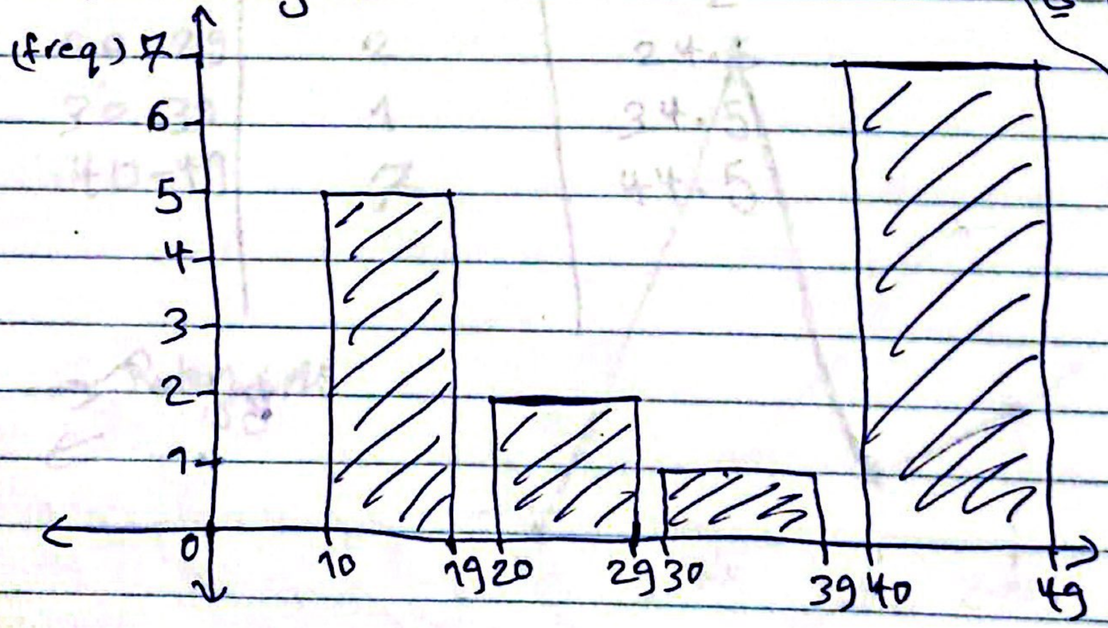
→ horizontal axis: classes المحور الأفقي: الفئات
→ vertical axis: frequency, relative frequency, or percent relative frequency.
المحور العمودي: التكرار أو التكرار النسبي أو المئوي

- EX: Construct a histogram for the following data

classes	freq
10 - 19	5
20 - 29	2
30 - 39	1
40 - 49	7

→ Histogram:

يشبه Bar graph في « Qualitative



b) frequency Polygon:

المضلع التكراري

→ horizontal axis: midpoints — المحور الأفقي: مركز الفئة

→ vertical axis: f, r.f, ~~[illegible]~~ or P.r.f

المحور العمودي: التكرار أو التكرار النسبي أو المئوي

→ Assigin the points: (midpoint, f). — نعين النقاط

connect the points by line segments. — نصل بين النقاط

close the polygon from both sides.

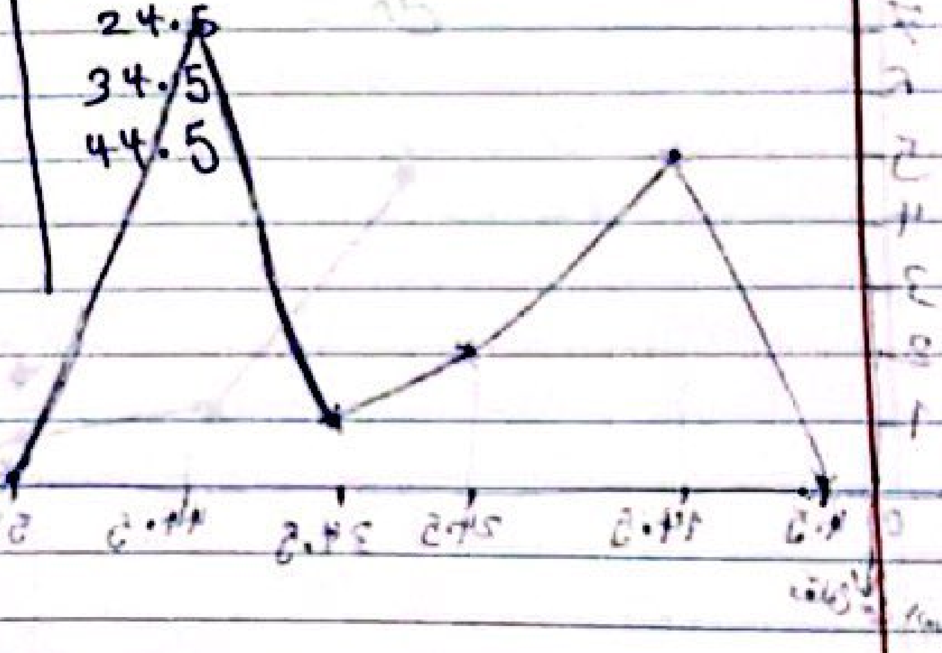
نغلق الشكل من البداية والنهاية

- EX: Construct the polygon for the following data.

classes	freq	midpoint
10-19	5	$\frac{10+19}{2} = 14.5$
20-29	2	24.5
30-39	1	34.5
40-49	7	44.5

→ Polygon:

←



c) ogive: المنحنى المتجمع الصاعد
((أو المنحنى التراكمي)).

→ horizontal axis: upper limits. المحور الأفقي : الحدود العليا

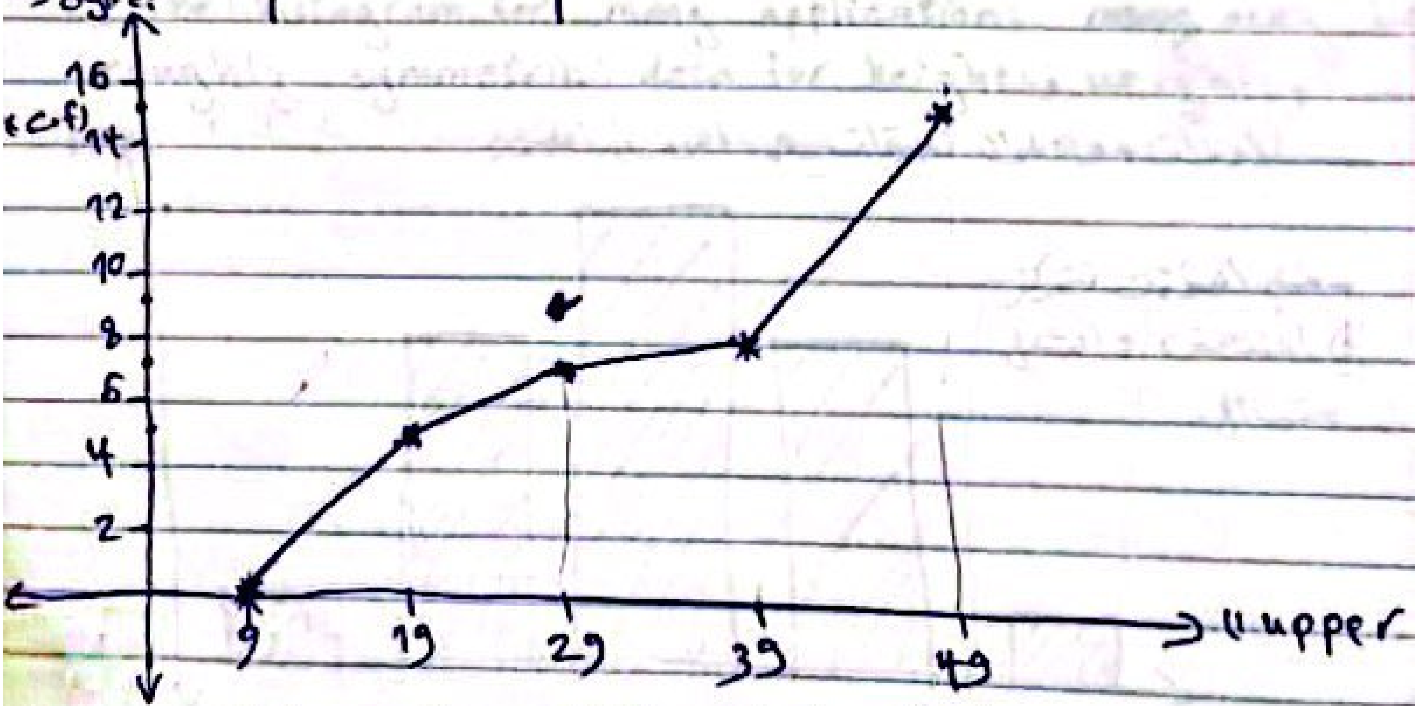
→ vertical axis: Cumulative frequency.
المحور العمودي : التكرار التراكمي

→ Assign the points: (upper limit, c.f). نعين النقاط
connect the points by line segents. نصل النقاط
close the ogive at the begining. نغلق فقط من البداية

- EX: Construct an ogive for the follwing data

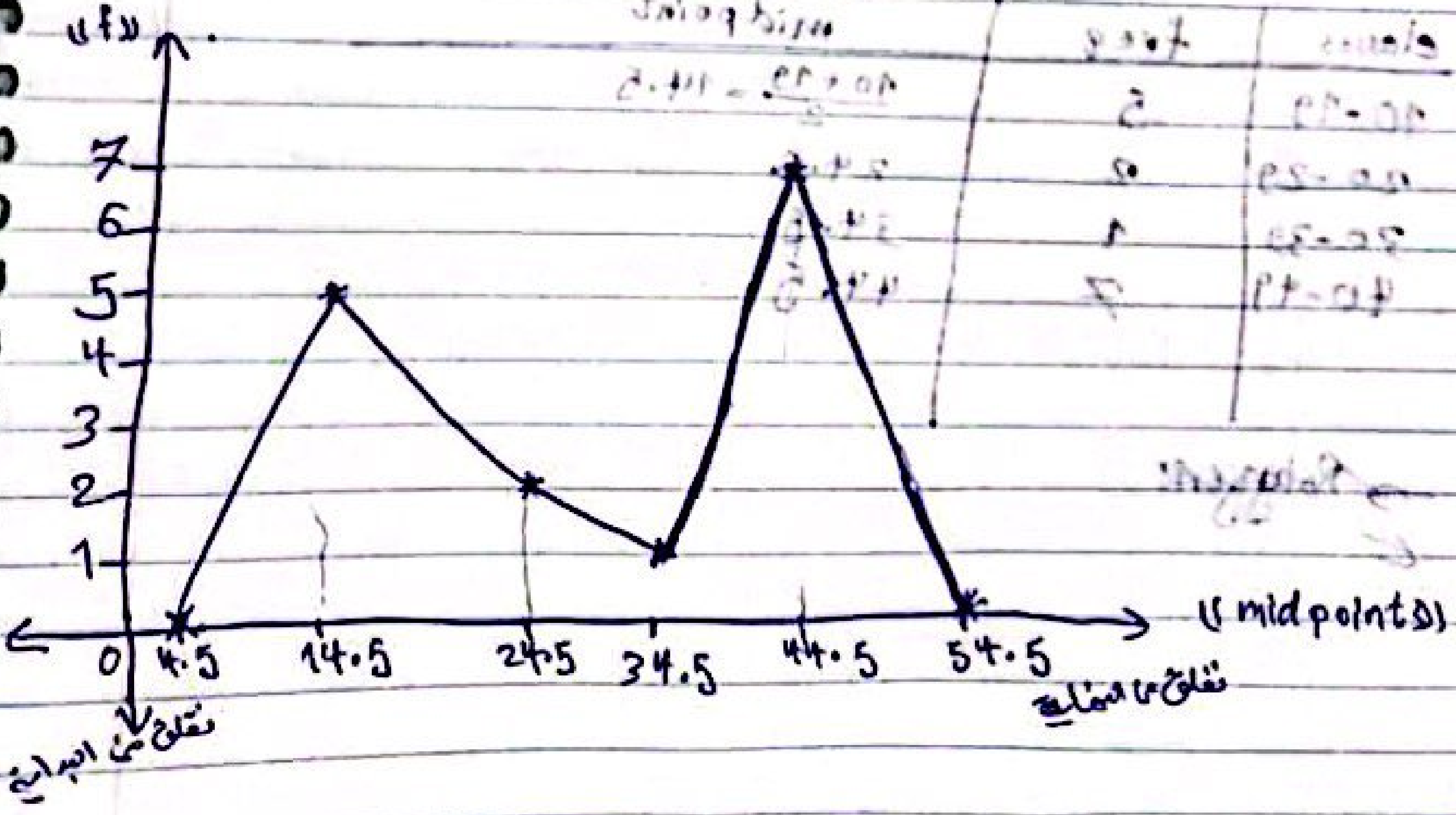
classes	f	Cumulative frequency ((c.f))
10-19	5	5
20-29	2	7
30-39	1	8
40-49	7	15

→ ogive:



- Note : An ogive is increasing. ((always)).
يزداد

→ Polygon:


(f)
7
6
5
4
3
2
1
0
4.5
14.5
24.5
34.5
44.5
54.5
(midpoints)
نغلق من البداية
نغلق من النهاية

* Distribution's shapes : أنواع شكل التوزيع

Based on the histogram, we can determine the distribution shape.

نحدد شكل التوزيع بناءً على رسمة histogram.

• Distribution shape:

describes how the data values of a variable are distributed.

يصف لنا كيفية توزيع البيانات.

- Types of distribution shapes:

① Normal distribution :- التوزيع الطبيعي

تتركز في المنتصف

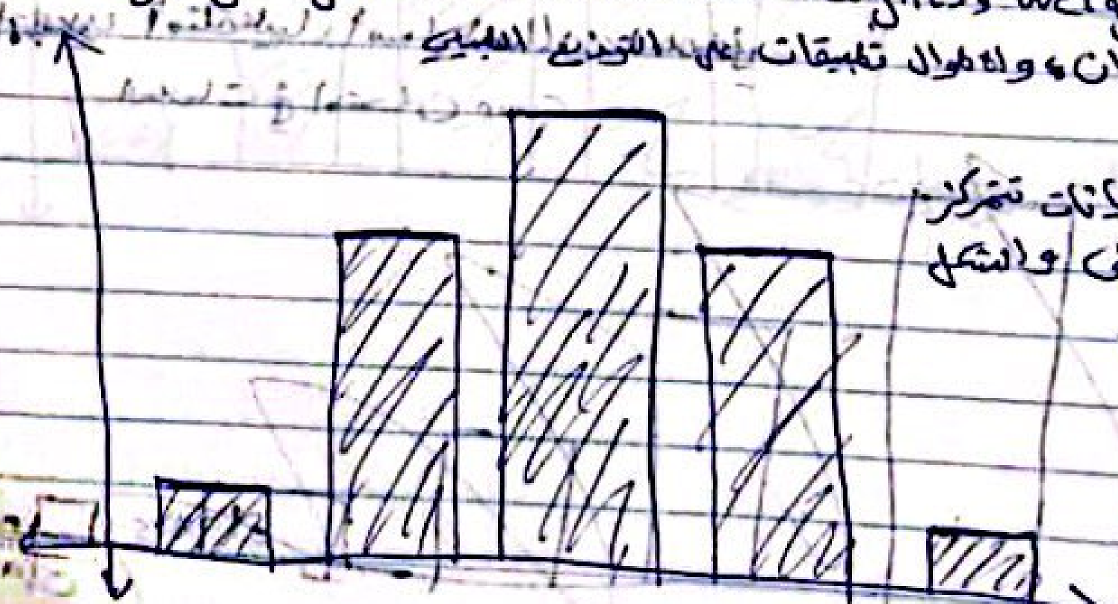
→ Data values cluster around the center "middle"

→ symmetric "that is, the left tail mirrors the shape of the right"

متماثل

→ The histogram for many applications may be roughly symmetric: data for heights, weights, ...

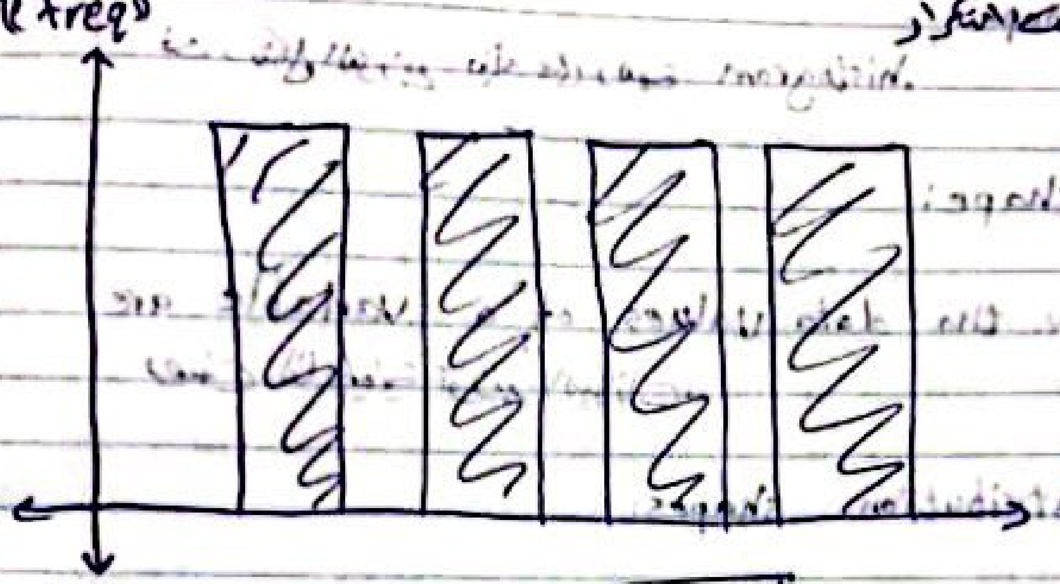
الأوزان والأطوال تطبيقات على التوزيع الطبيعي



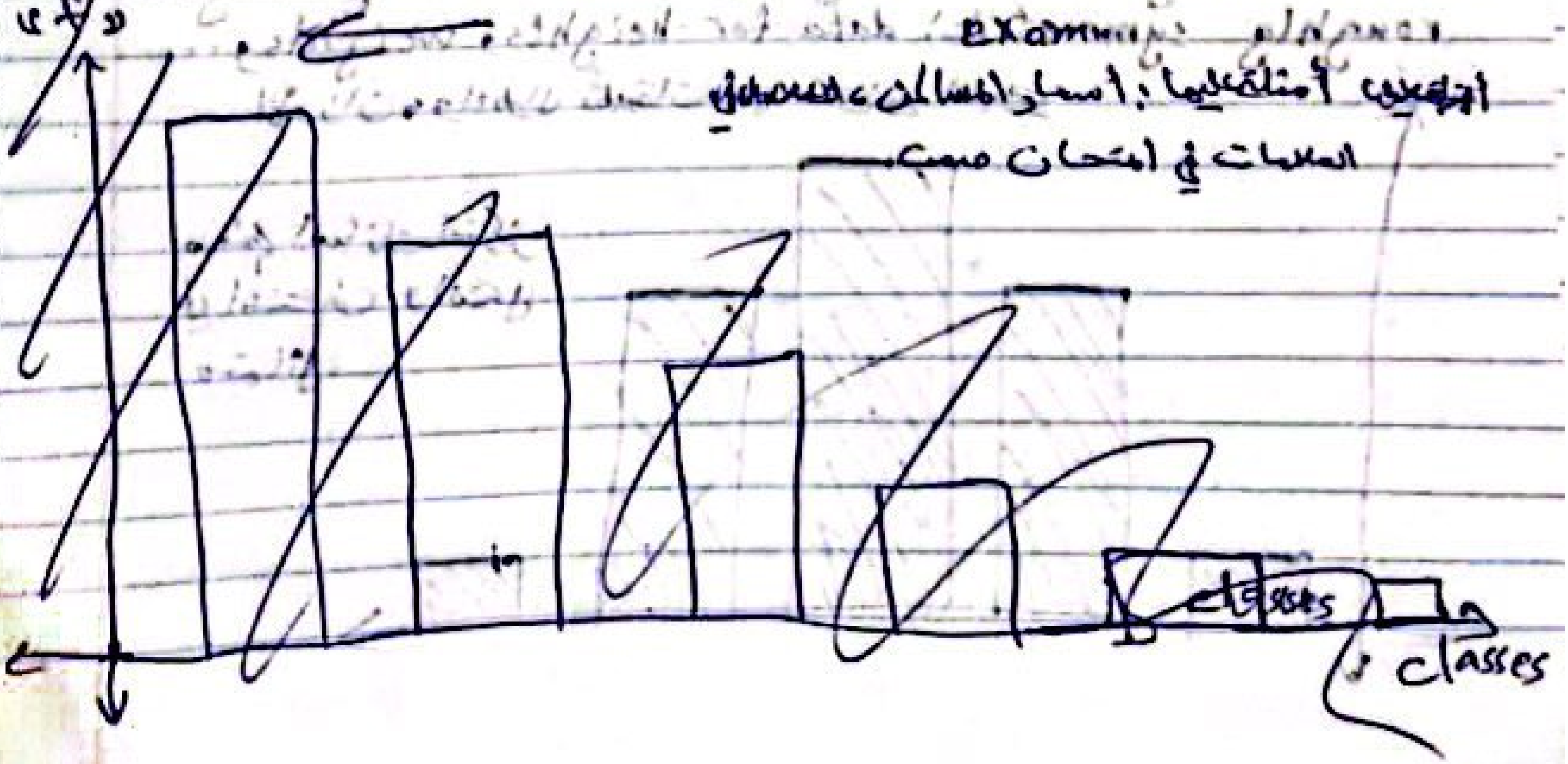
معظم البيانات تتركز في المنتصف والشكل متماثل.

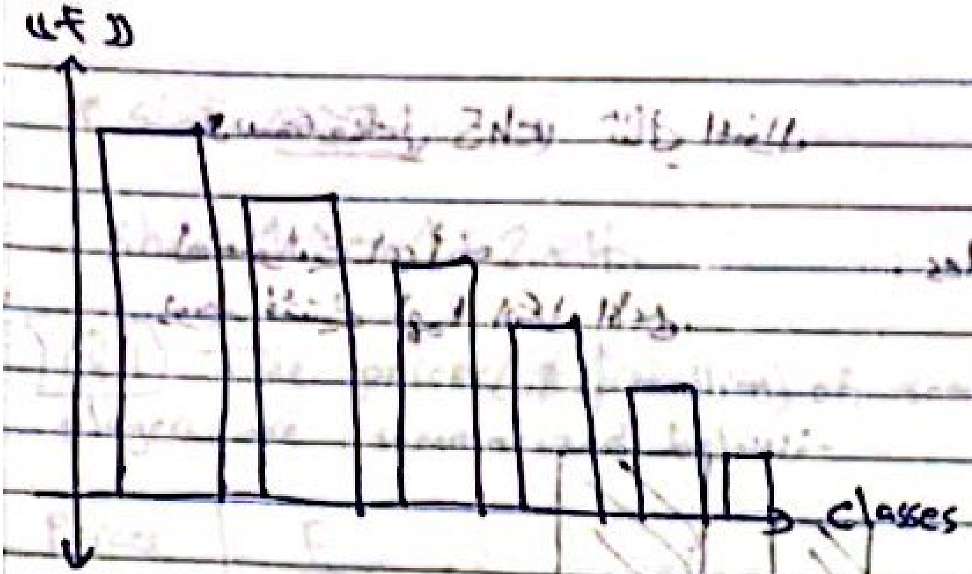
② The uniform distribution: التوزيع المنتظم

→ data values have same frequencies
→ جميع الفئات لها نفس التكرار



③ Positively skewed: التواء موجب / انحراف يميناً (right)
→ data values cluster on left. تتمركز البيانات في اليسار
→ its tail extends to the right. ذيل
→ An example of this type: ① housing prices, ② scores in a difficult exam.
أمثلة عليها: أسعار المساكن، العلامات في امتحان صعب





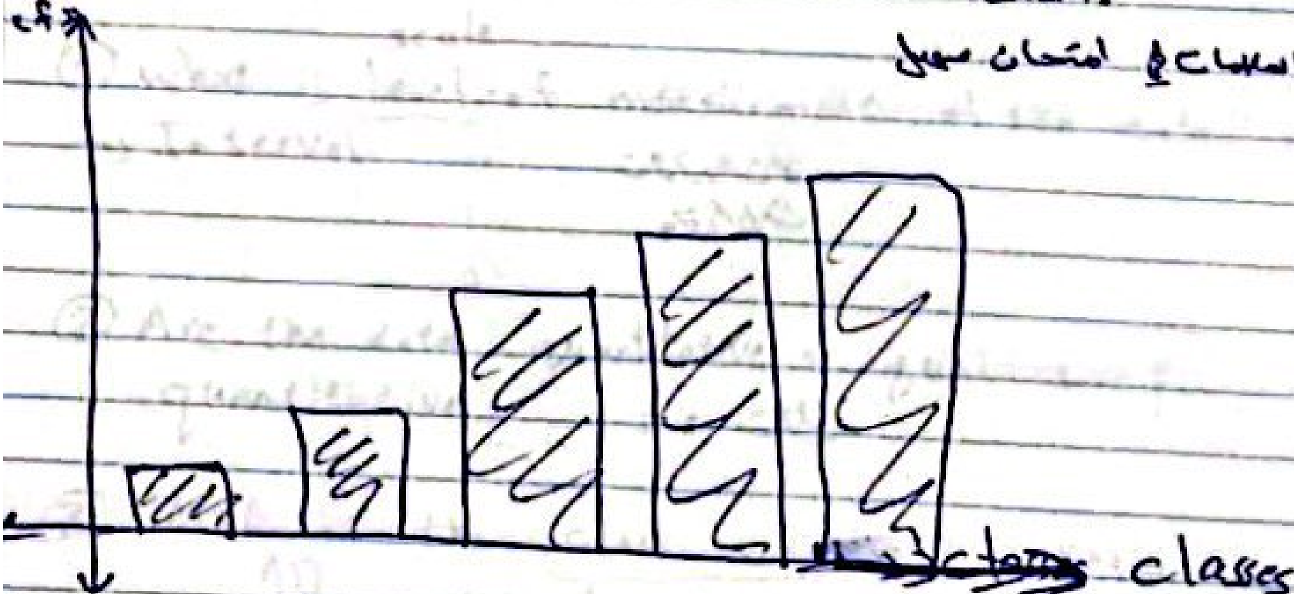
④ Negatively skewed:

انحياز يسارا

تتركز في اليمين

→ data values cluster on right.
→ its tail extends to the left.
→ An example of this type: scores in an easy exam.

العلامات في امتحان سهل

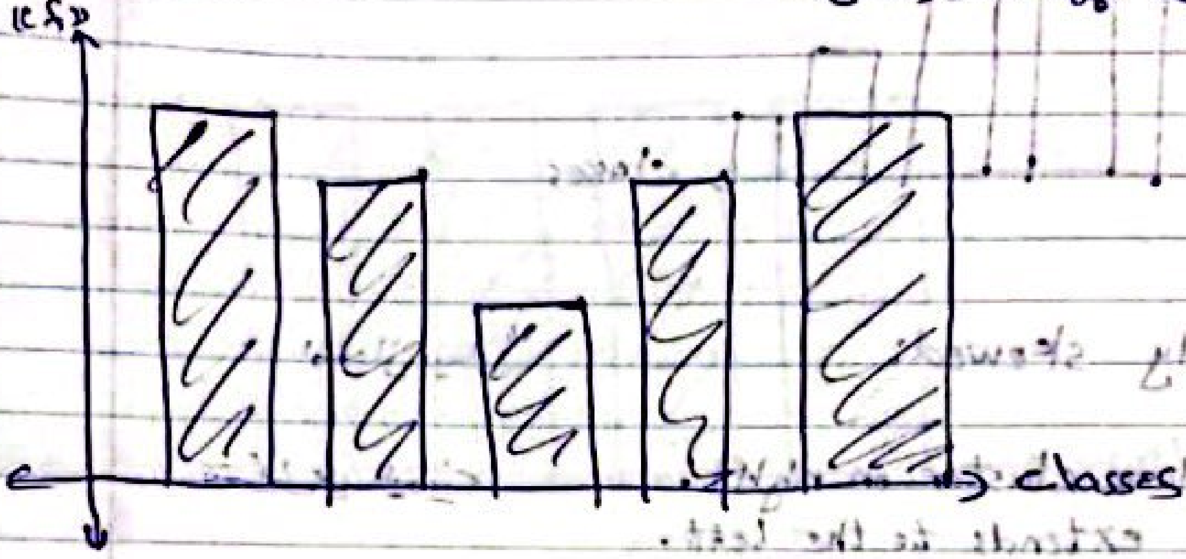


(5) Bimodal:

سنتطرق له في ch5 ثنائي المنوال

→ we have 2 modes.

له منوالين أي أنه يوجد فئتين لهما التكرار الأعلى.



* stata 2371

- Sheet 1 : ch1 + 2 + 4.

Q1 The prices ($ mallion) of some ~~...~~ footbal players are summarized below:-

Prices	f
6-15	24
16-25	26
26-35	10
36-45	8
46-55	5
56-65	7

① what is level of measkrment of the data? (scale) تصنيف
→ Interval
نوع البيانات
شكل الأرقام

② Are the data quantitative or qualitative?
quantitative.
مجالات رقمية

③ what is the class width? طول الفئة
10
أو الفرق بين أي حدين علويين أو سفليين

④ what is the midpoint (مركز الفئة) of the classes 56 - 65?

$\rightarrow$ midpoint $= \dfrac{U + L}{2}$ $\quad \dfrac{\text{الحد الأعلى} + \text{الحد الأدنى}}{2}$

$= \dfrac{56 + 65}{2} = 60.5$

⑤ what is the sample size? (حجم العينة)

sample size «n» $= \sum F$

$= 24 + 26 + 10 + 8 + 5 + 7$

$= 80$

⑥ what is the number (التكرار) of players with prices between (بين) 15 and 56?

$26 + 10 + 8 + 5 = 49$

	F
6 - 15	
16 - 25	26
26 - 35	10
36 - 45	8
46 - 55	5
56 - 65	

احتمال

(7) what is the probability that a player's price is more than 35 million?
أكثر

$R.f = \frac{f}{n} = \frac{20}{80} = 0.25$

class	f
6 - 15	
16 - 25	
26 - 35	
36 - 45	8
46 - 55	5
56 - 65	7

$8 + 5 + 7 = 20$

المطلوب probability

أي $r.f = \frac{20}{80}$

النسبة المئوية

(8) what is the percentage of players whose price are at most 45 million?
على الأكثر

$p.r.f = \frac{f}{n} \times 100\%$

$= \frac{68}{80} \times 100\%$

$= 85\%$

classes	f
6 - 15	24
16	26
26	10
36	8
46	
56	

يعني أكثر 45
أي 45 أو أقل

$24 + 26 + 10 + 8 = 68$

R.F

⑨ construct a relative frequency distribution?

P.F

⑩ construct a percent frequency distribution for these data?

C.F

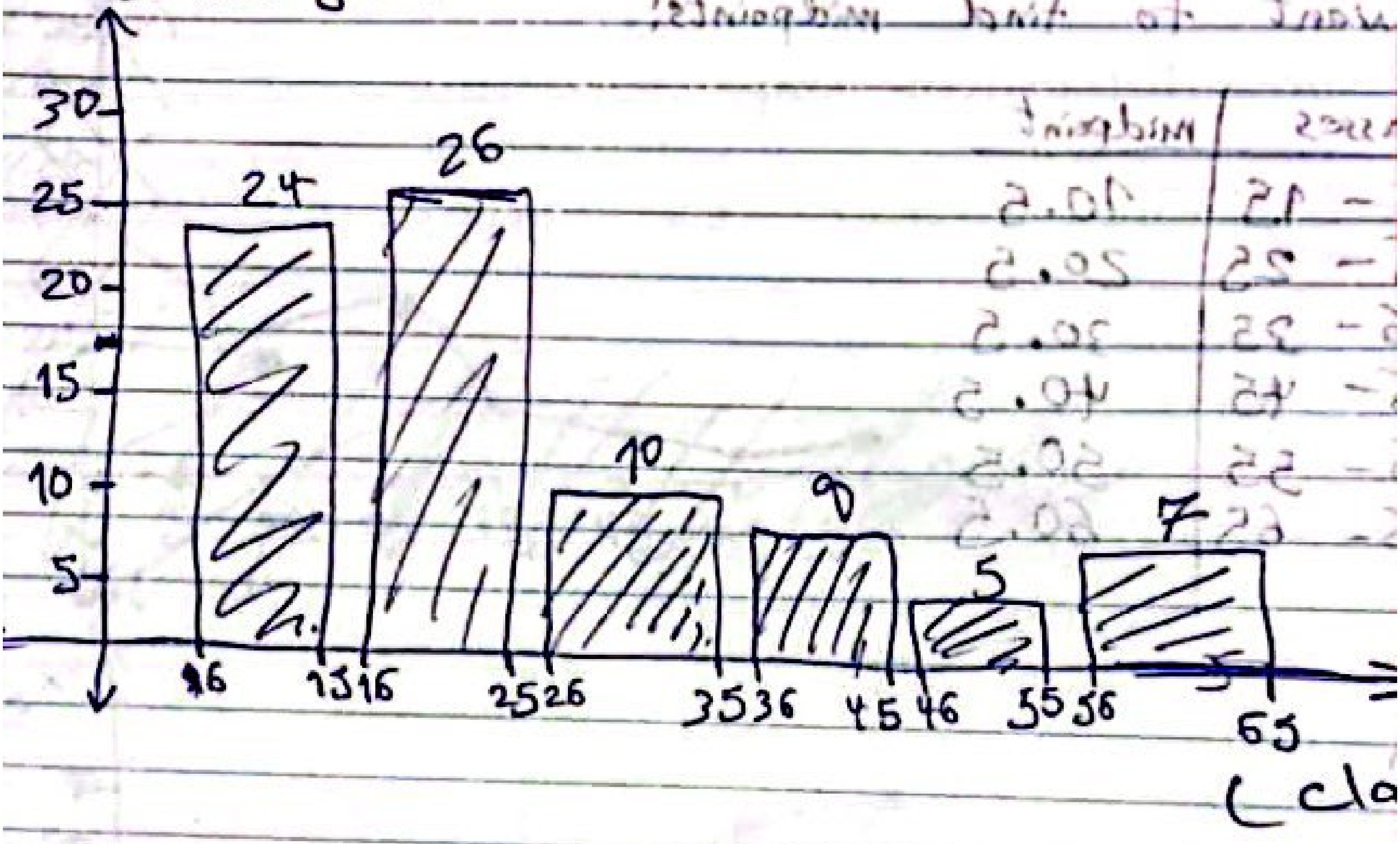
⑪ constructs a cumulative frequency distribution for these data?

ج ⑨ + ⑩ + ⑪ في جدول واحد

classes	F	$R.F = \frac{F}{n}$	$P.r.f = \frac{F}{n} \times 100\%$	C.f
6 - 15	24	$\frac{24}{80} = 0.3$	$\frac{24}{80} \times 100\% = 30\%$	24
15 - 25	26	0.325	32.5%	24 + 26 = 50
26 - 35	10	0.125	12.5%	50 + 10 = 60
36 - 45	8	0.1	10%	60 + 8 = 68
46 - 55	5	0.0625	6.5%	68 + 5 = 73
56 - 65	7	0.0875	8.75%	73 + 7 = 8[illegible]

(12) Constract a histogram, frequency polygon and ogive for these data?

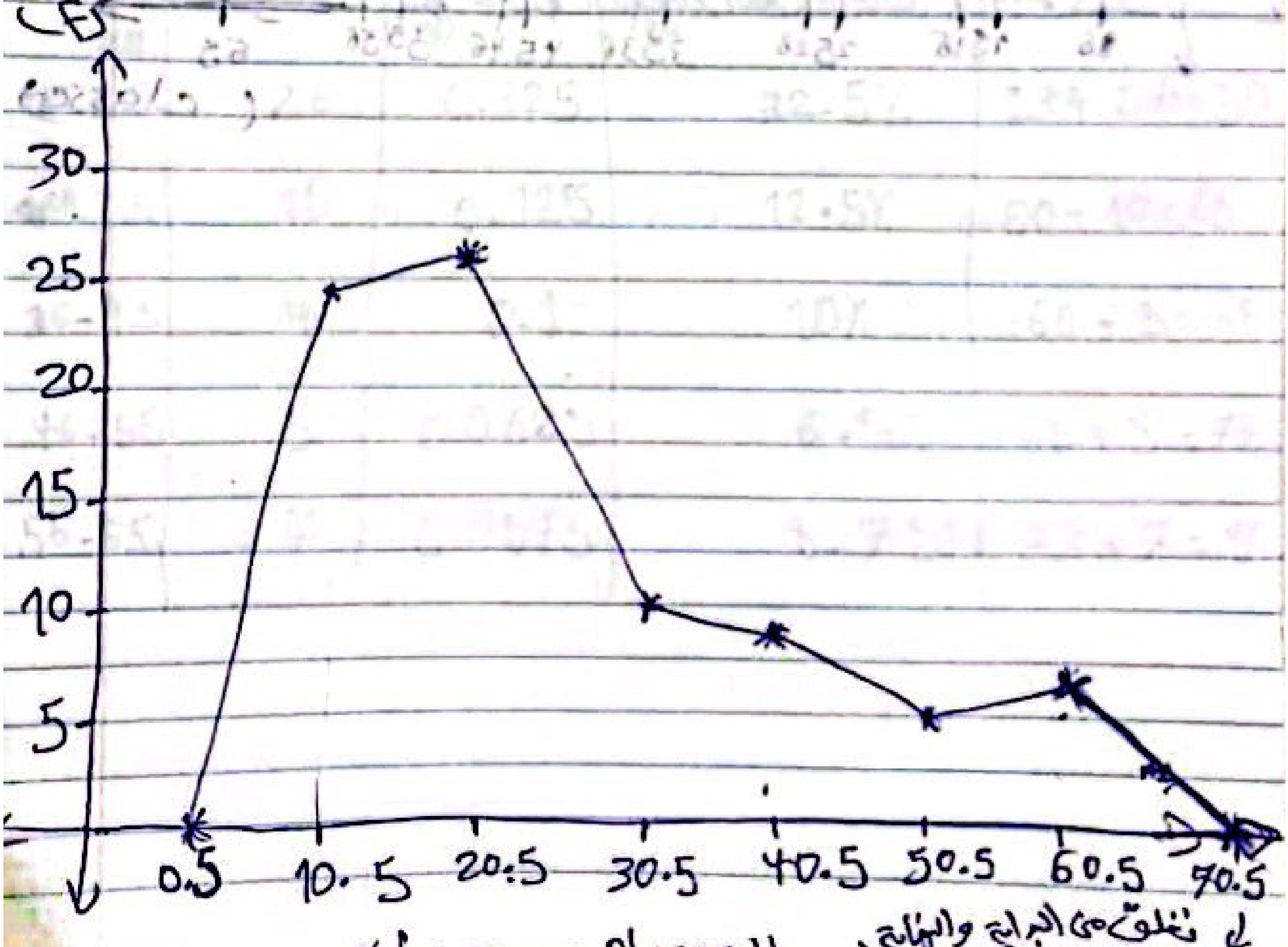
→ Histogram



→ polygon:

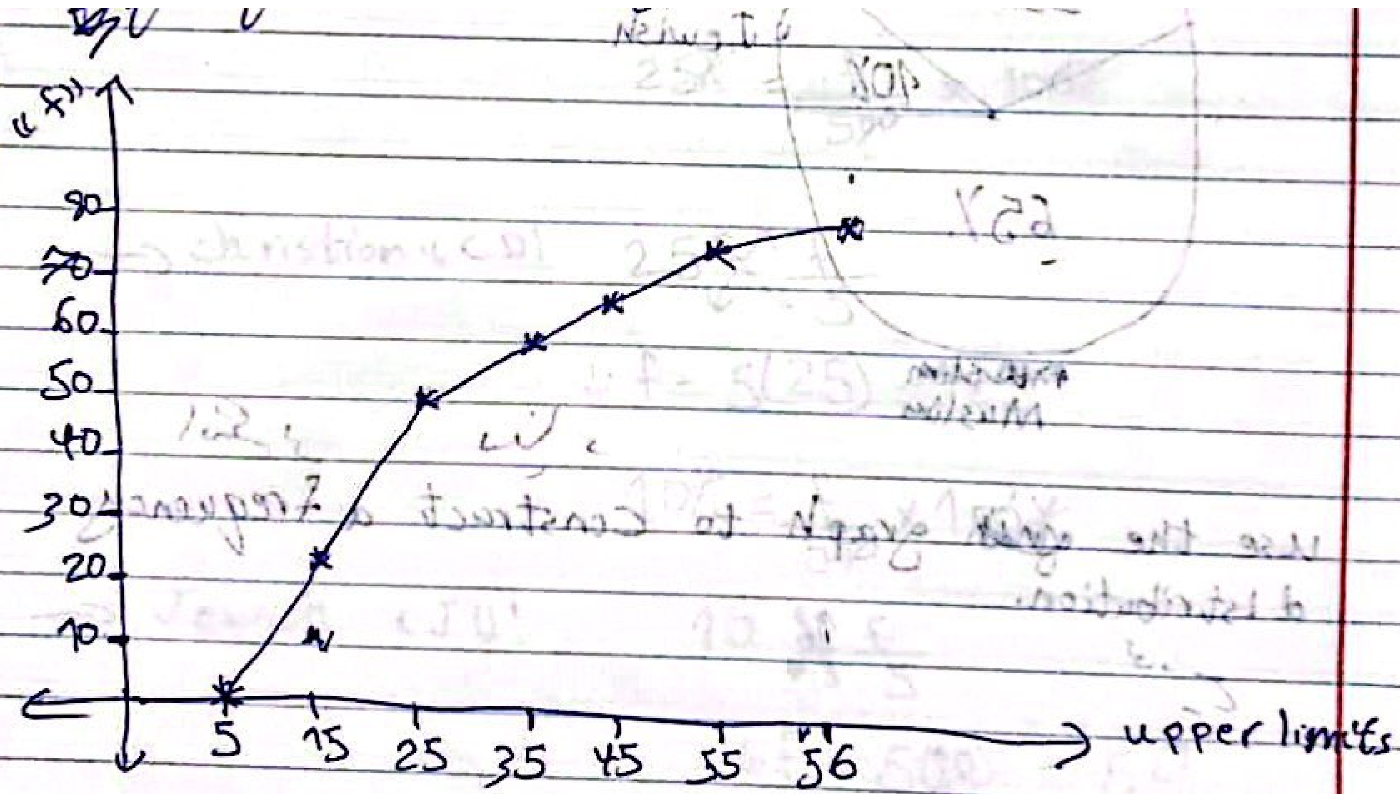
want to find midpoints:

classes	midpoint
6 – 15	10.5
16 – 25	20.5
26 – 35	30.5
36 – 45	40.5
46 – 55	50.5
56 – 65	60.5



((frequency Polygon))

نغلق من البداية والنهاية.


"f"
80
70
60
50
40
30
20
10
5
15
25
35
45
55
56
upper limits

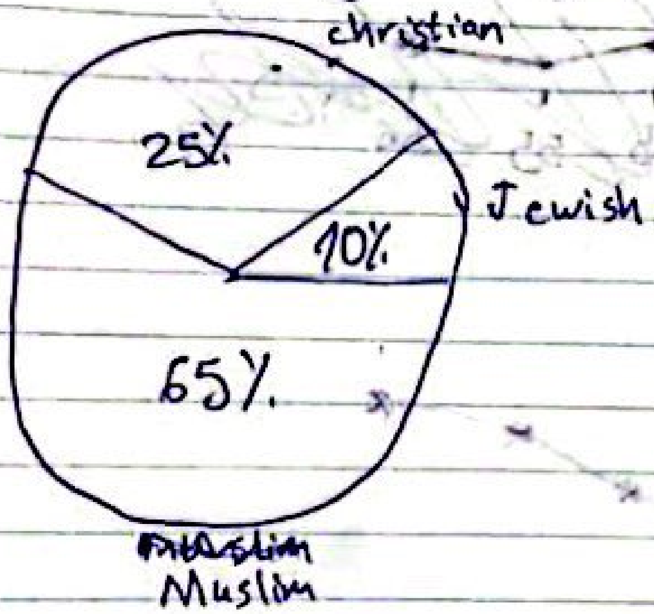
(23) what is the distribution shape for these data?

شكل التوزيع

→ Based on Histogram: Positive skewed

the distribution shape is positively skewed.

Q2 The religions of a sample of (500) residents are summarized by the graph below:-



use the graph to construct a frequency distribution.

تكرار

بناء

توزيع

* To find the frequency distribution, we want to find the frequency for each class.

→ Muslim (M): $P.r.f = \frac{f}{n} \times 100\%$

$65\% = \frac{f}{500} \times 100\%$

$65 = \frac{f}{5}$

$\therefore f = 5(65) = 325$

→ Christian (C):

$25\% = \frac{f}{500} \times 100\%$

$25 = \frac{f}{5}$

$\therefore f = 5(25) = 125$

→ Jewish (J):

$10\% = \frac{f}{500} \times 100\%$

$10 = \frac{f}{5}$

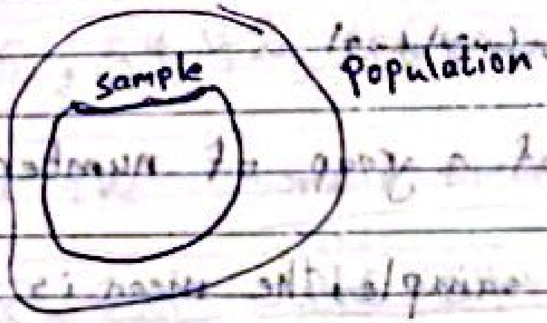
$\therefore f = 5(10) = 50$

- The frequentcy distribution:-

classes	f
M	325
J	50
C	125

اتجاه المركز مقاييس

* ch5: Measures of central tendency.



- Measures are computed for data from a sample are sample statistics. القراءات التي لها علاقة بالعينة
تسمى statistic

- Measures are computed for data from a population are parameters القراءات التي لها علاقة بالمجتمع
تسمى parameter

- A sample statistic is the point estimator of the Corresponding population parameter.

عند صعوبة دراسة المجتمع نلجأ الى عينة n وتعتبر قراءات العينة نقاط تقريبية لقراءات المجتمع.

- Measures of central tendency:

The values that determine the average of a group of data.
مقاييس النزعة المركزية: تجمع معظم المشاهدات حول قيمة متوسطة.

مقاييس المركزية

— There measures of central tendency:-

[1] Mean:

الوسط الحسابي

The arithmatic average of a group of numbers.

— If the data are for a sample, the mean is denoted by $\bar{X}$.

• the sample mean "$\bar{X}$" $= \frac{\sum X_i}{n}$; Σ = the summation sign.

إشارة الجمع

X_i: the data values.

n: the sample size

حجم العينة

— If the data are for a population, the mean is denoted by μ.

حجم المجتمع

• the population mean "μ" $= \frac{\sum X_i}{N}$, N: pop. size.

الوسط الحسابي = مجموع القيم / عددها

* Note: Most of the time, we deal with samples, so we will use $\bar{X}$.

-EX: The following data represents the class size for a sample of 5 classes:-

46, 54, 42, 46, 32.

find the sample mean

$\longrightarrow \bar{x} = \dfrac{\Sigma x_i}{n} = \dfrac{46+54+42+46+32}{5}$

$\therefore \bar{X} = 44.$

حفظها

* Properties of the mean:-

① A group of data has only on mean.

② The mean could be unrealistic.

أحياناً يكون الوسط غير واقعي.

-EX: The following data represent the family size of a sample of 8 families.

4, 7, 3, 5, 2, 4, 7, 9

$\longrightarrow \bar{X} = \dfrac{4+7+3+5+2+4+7+9}{8} = 5.125$

which isn't realistic.

لأنه يمثل عدد أفراد عائلة، فيجب أن يكون عدد صحيح

③ The mean is affected by outliers.

يتأثر الوسط الحسابي بالقيم المتطرفة.

-EX: • sample 1: 5, 7, 4, 9, 6

→ $\bar{X} = 8$! (No outliers)

• sample 2: 5, 7, 4, 9, 80

→ $\bar{X} = 21$. (with outliers).

④ The mean is used for quantitative data. [Interval scale].

يستخدم فقط للبيانات الكمية.

2) Median: الوسيط

The middle value of a group of data after being ~~ord~~ ordered. القيمة التي في المنتصف بعد ترتيب البيانات

. To find the median :- من الأصغر للأكبر

a) order the data (from smallest to largest). ترتيب البيانات

b) .If the size (n) is odd, the median is the middle value. إذا كان عدد البيانات فردي يكون الوسيط القيمة التي في المنتصف

→ the rank of the median = $\frac{n+1}{2}$

→ the median = $X_{\frac{n+1}{2}}$ ($\frac{n+1}{2}$ القيمة التي في موقعها)

- If n is even, the median is the ~~[illegible]~~ average of 2 middle values.

اذا كان حجم البيانات عدد زوجي فيكون الوسيط معدل القيمتين اللواتي في المنتصف.

→ the rank of the median $= \frac{n}{2}, \frac{n}{2}+1$

→ the median $= \dfrac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$

(القيمة التي موقعها $\frac{n}{2}$ + القيمة التي موقعها $\frac{n}{2}+1$) مقسوماً على 2

- EX: Given the sample: 46, 54, 42, 46, 32.
find the median.

$n = 5$ ((odd)).

→ 32, 42, 46, 46, 54.

the median is 46
the rank of the median = 3

بإمكاننا بعد ترتيب البيانات إيجاد الوسيط من خلال إيجاد ~~[illegible]~~ الرتبة $\frac{n+1}{2} = \frac{6}{2} = 3$

* الوسيط = القيمة الثالثة = 46

- EX: Given the sample: 13, 10, 15, 10, 10, 9, 12, 14.
find the median.

$n = 8$ (even)

→ 9, 10, 10, $\boxed{10, 12}$, 13, 14, 15

∴ median $= \frac{10+12}{2} = 11$

* properties of the median:

① A group of data has only one median

② The median could be unrealistic.

- EX: The following data represent the family size of a sample of 8 families.

4, 7, 3, 5, 2, 4, 7, 9.

→ 2, 3, 4, $\boxed{4, 5}$, 7, 7, 9

median $= \frac{4+5}{2} = 4.5$ which isn't realistic.

③ The median isn't affected by outliers.

- EX: a. Given the sample 1, 4, 2, 1, 5, 4.

→ 1, 2, 4, 4, 5.

median = 4

b. Given the sample: 4, 2, 1, 95, 4

→ 1, 2, 4, 4, 95

→ median = 4

Notes that 95 is an exteme value and the median isn't affected by it.

④ The median is ~~Used~~ used for interval and ordordinal data but not for ~~nominal~~ nominal.

تستطيع ايجاد median لأي بيانات يمكننا ترتيبها.

(interval or ordinal) رتبي

⑤ The median is the 50% percentile, that is 50% of the data value are less than or equal the median, and 50% of the data values are greater than or equal to the median.

الوسيط هو المئين الخمسين يعني أن نصف البيانات أقل منه أو تساوي والنصف الآخر أكبر منه أو ~~تساويه~~ تساويه. يساويه

-EX: If the ages of a sample of 36 ~~employeses~~ employees has a median of 45 years, than 18 employees are less than or equal to the median «45» and 18 are greater than or equal to 45 «median»

18 ← 36 employees to 45 «median»

18 ← 45 median → 18

3 ~~Mode~~ Mode:

The values occurs with greatest frequency (the highest ~~freq~~ frequency data.

الأكثر البيانات تكراراً او شيوعاً

- EX: ① Given the sample: 4, 2, 2, 1, 3, 10. find the mode.

→ mode = 2

② Given the sample: 4, 2, 2, 1, 4, 10

→ Mode = 4, 2

③ Given the sample: 4, 2, 2, 1, 4, 4, 10

→ Mode = 4.

④ Given the sample: 4, 2, 2, 1, 4, 1, 10, 10

→ No mode

جميع البيانات لها نفس التكرار.

⑤ Given the sample: 4, 2, 3, 1, 10 :المطلوب ايجاد E

→ No mode

- Note that:

we could have no mode, 1 mode, 2 mode, 3 mode, ...

- If the sample has 1 mode: unimodal data.
- If the sample has 2 modes: bimodal. ثنائي المنوال
- If the sample has 3 modes: trimodal. ثلاثي المنوال
- If the sample has ~~4 modes~~ more than 3 modes: ~~multi~~ multimodal. متعدد المنوال

- Ex: The following data represents the blood type for 10 student:

AB, O, O, A, B, O, A, A, A, B.

find the mode:

→ Mode: A. اكثر البيانات تكرارا

---

① Qualitative → (nominal + ordinal) نوعية

→ mode

② Quantitative كمية

interval

* Properties of the mode:-

① A group of data could have one mode or more or even no mode.

② The mode is always realistic.

③ The mode isn't affected by outliers.

- EX: ⓐ sample 1: 4, 3, 2, 5, 1, 7, 5.
→ mode : 5

ⓑ sample 2 : 4, 3, 2, 5, 1, 100, 5
→ mode = 5

④ The mode is used for both quantitative and qualitative data (Nominal, ordinal, Interval).

* How to determine the best measure of central tendency for a sample?

① If the data values of a variable is nominal or ordinal (qualitative), then the mode is the best.

② If the data values of a variable is interval;
→ If the sample has more than 1 mode, then the mode is the best.

→ If the sample has 1 mode or has no mode, check the outliers:

ⓐ If there are outliers, than the median is the ~~best~~ best.
ⓑ If there are no outliers, than the mean is the best.

← إذا كانت البيانات نوعية (مسميات حروف)، يكون المنوال هو الأفضل

← إذا كانت البيانات رقمية: في حال وجود أكثر من منوال، تتجنب المنوال
وفي حال وجود منوال واحد أو عدم وجود منوال، ننظر إلى القيم
المتطرفة، إذا كان في متطرفة يكون الوسيط هو الأفضل، وإذا لم يكن في
متطرفة فيكون الوسط الحسابي هو الأفضل.

- Ex: which measure of central tendency is the best?

① sample: F, M, M, M, F, M, F, F, F, F, F, M, F.

→ the variable is qualitative, so the mode is the best.

② sample: 1, 2, 2, 4, 5, 2, 4, 1, 1

→ Mode = 1, 2 « more than 1 Mode »
∴ the mode is the best.

③ sample: 1, 2, 2, 4, 5, 4, 7, 2

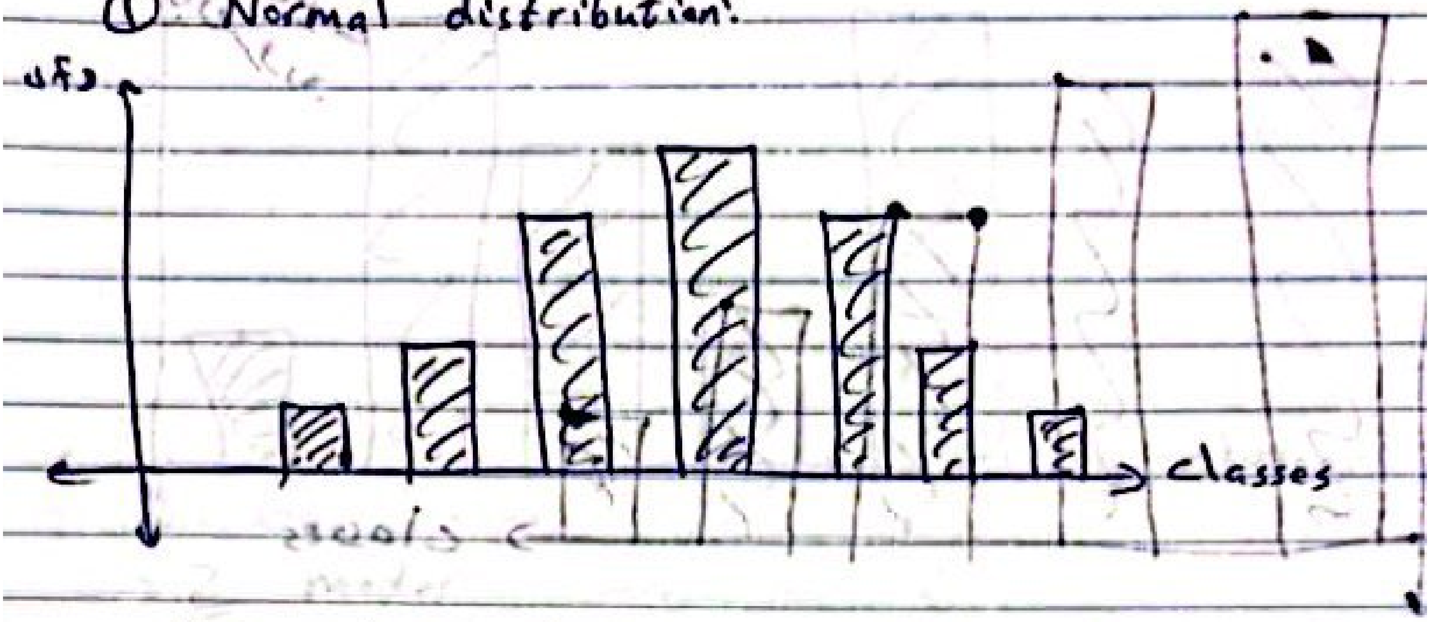
→ Mode = 2 « one mode »

No extreme values (outliers), ∴ the mean is the best.

⑤ sample: 1, 2, 2, 4, 5, 4, 70, 2
→ Mode = 2 « one mode »
70 is an extreme value, ∴ the midian is the best.

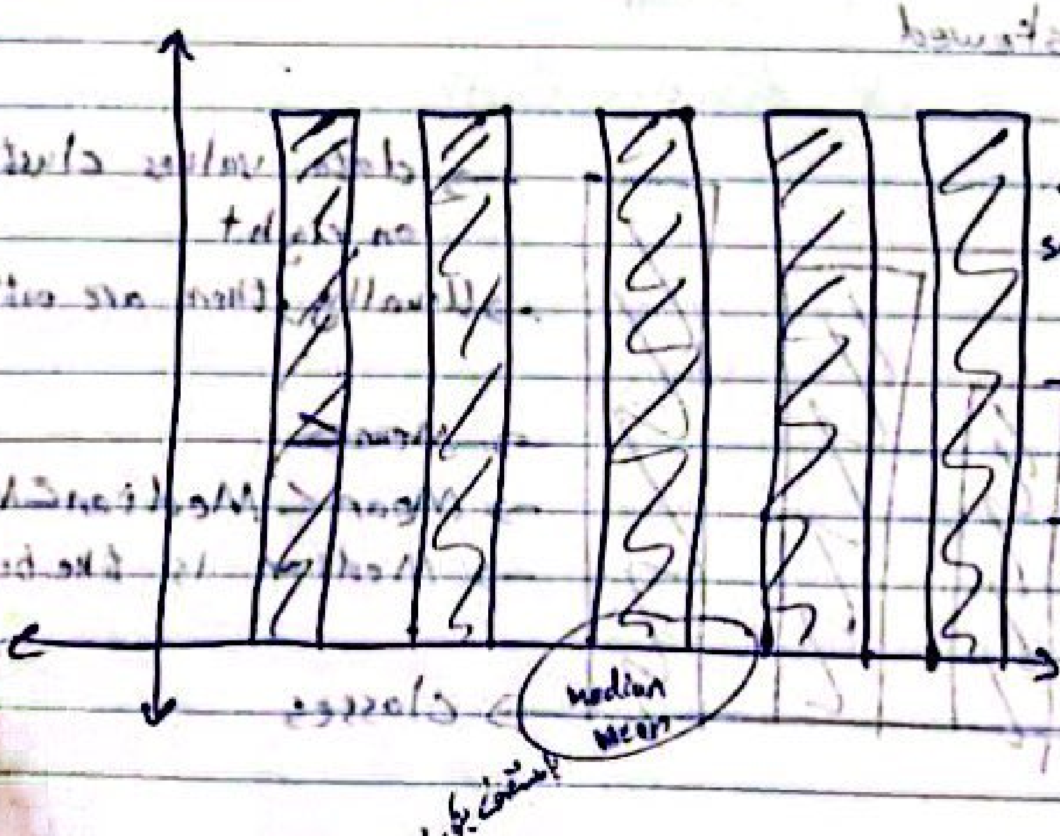
* Distribution shapes vs. measures of central tende

① Normal distribution:



→ data values cluster around the center.
→ the graph is symmetric
→ Mean = Median = Mode

② uniform توزيع منتظم



→ data valu
same frequencie

لا يوجد منوال
→ No mode

→ mean and me
near center.
المتوسط والوسيط
في المنتصف تقريبا؟

(3) positively skewed

يوجد outliers

→ data valuse cluster on left

→ Usually, there are outliers. لأنه يوجد قيم متطرفة

→ Mean > Median > Mode. الوسط أكبر من الوسيط أكبر من المنوال

(4) Negative skewed

→ data values cluster on right

→ Usually, there are outlie

→ ~~Mean~~

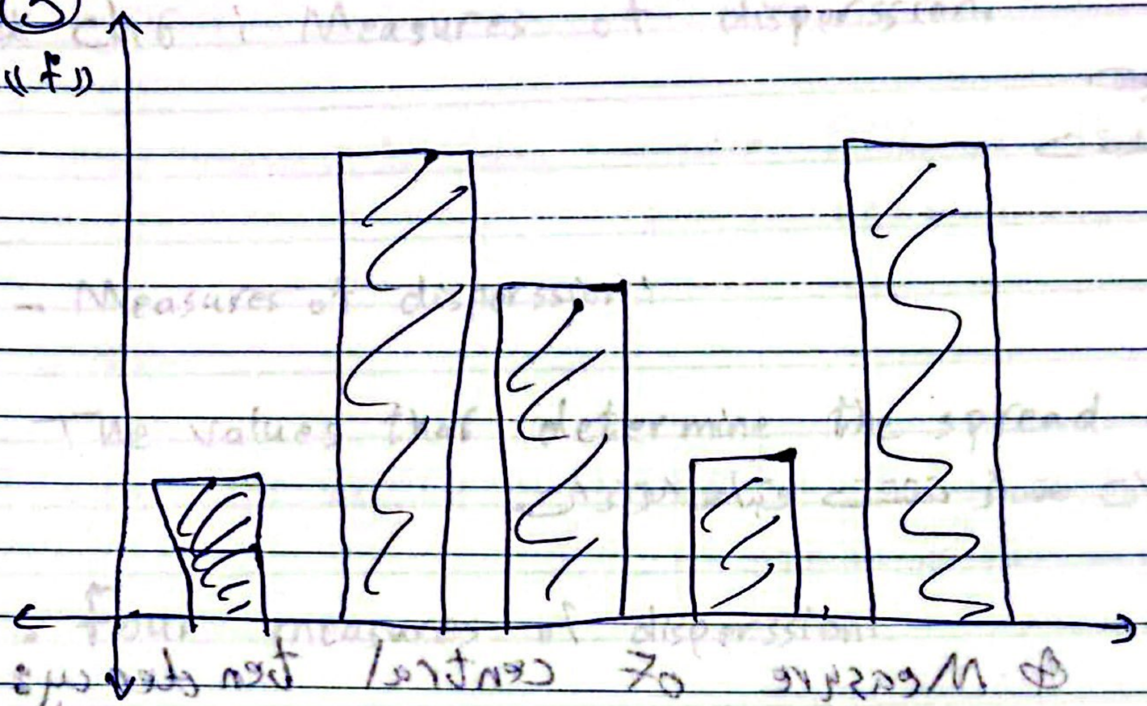
→ Mean < Median < Mo

→ Median is the best

classes

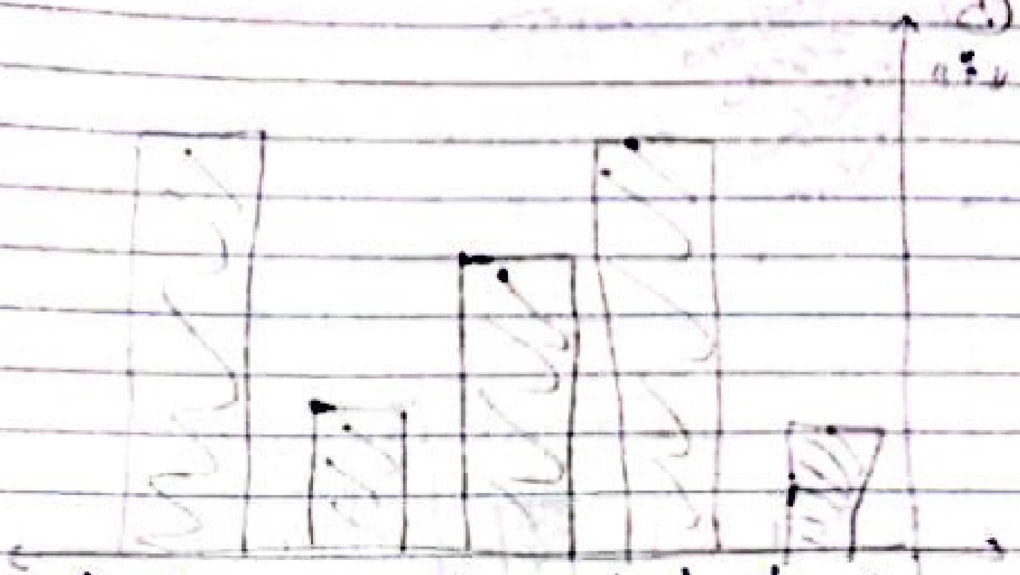
Bimodal

(5)



→ 2 modes

→ Mode is the best



* Measure of central tendency:

مقاييس الاتجاه المركزي

1) Mean الوسط الحسابي ← يتأثر بالقيم المتطرفة

2) median الوسيط ← لا يتأثر بالقيم المتطرفة

3) mode المنوال ← لا يتأثر بالقيم المتطرفة

مقاييس التشتت

* Measure of dispersion

1) Range ← يتأثر بالقيم المتطرفة

2) Interquartile range ← لا يتأثر بالقيم المتطرفة

3) variance ← يتأثر بالقيم المتطرفة

4) standard deviation ← يتأثر بالقيم المتطرفة

80

* ch6 : Measures of disperssion.

مقاييس التشتت

- Measures of disperssion:

The values that determine the spread of the data values.

تقيس هذه القراءات مقدار تشتت وتباعد القراءات

. four measures of disperssion:

① Range: المدى

The difference ~~between~~ between the largest data value and the smallest data value.

الفرق بين أعلى قيمة وأصغر قيمة

$\rightarrow$ Range = $\underset{}{\overset{\text{أعلى قيمة}}{Max}}$ - $\overset{\text{أصغر قيمة}}{Min}$

- Ex: Given the following grades 85, 73, 80, 62, 96 find the range.

$\rightarrow$ Range = 96 - 62 = 34

- Note:

The rang is a affected by outliers.

يتأثر المدى بالقيم المتطرفة.

② Intequartile range (IQR): نصف المدى الربيعي

The range for the middle 50% of the data.
المجال الذي تنتشر فيه 50% من البيانات

- Note:

The interquartile range (IQR) <u>isn't affected</u> by outliers.
الـ IQR لا يتأثر بالقيم المتطرفة

← طريقة حساب IQR غير مطلوبة.

③ Variance: التباين

The average of the squared deviations about the mean. « measure whether the data cluster around the mean ».

متوسط مربعات الانحراف عن الوسط الحسابي

→ Population variance « σ^2 » $= \dfrac{\sum (x_i - \mu)^2}{N}$; x_i: data valuse

μ: pop mean

N: pop. size

→ sample variance « S^2 » $= \dfrac{\sum (x_i - \bar{x})^2}{n-1}$; x_i: data valuse

$\bar{x}$: sample mean

n: sample size

- Ex: Given the sample: 46, 54, 42, 46, 32.
find the variance.

$\rightarrow \bar{X} = \frac{\sum X_i}{n} = \frac{46+54+42+46+32}{5} = 44$

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
46	2	4
54	10	100
42	-2	4
46	2	4
32	-12	144

$\therefore S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} = \frac{4+100+4+4+144}{5-1} = 64$

- Note: بيتأثر بالقيم المتطرفة
The variance is affected by outliers.

④ standard deviation: الانحراف المعياري

The square root of the average squared deviation of the data from the mean. « the positive square root of the variance ».

الجذر التربيعي لمتوسط مربعات انحراف القيم عن الوسط الحسابي

→ sample standard ~~dev~~ deviation ((s)) = $\sqrt{\frac{\sum(\bar{x}-x_i)^2}{n-1}}$

→ Populations ~~stand~~ standard deviation ((σ)) = $\sqrt{\frac{\sum(x_i-\mu)^2}{N}}$

لما تزيد القيم المتطرفة 100

- Note:

The standard deviation is ~~a~~ affected by outliers.

- EX: find the standard deviation for the prevsous example.

→ $s = \sqrt{s^2} = \sqrt{64} = 8$

- Not:

$\sum(x_i-\bar{x})=0$ always

مجموع انحرافات القيم عن
الوسط الحسابي = صفر دائماً.

5 2

* Using calculator:

① To find the mean.

نعّف الثاني sample او pop.

$\boxed{\text{Mode}}$ $\boxed{2}$

$\boxed{x_1}$ $\boxed{M+}$

$\boxed{x_2}$ $\boxed{M+}$

⋮

$\boxed{x_n}$ $\boxed{M+}$ $\boxed{\text{ON}}$

$\boxed{\text{shift}}$ $\boxed{2}$ $\boxed{1}$ $\boxed{=}$

② To find the sample standard devion.

$\boxed{\text{shift}}$ $\boxed{2}$ $\boxed{3}$ $\boxed{=}$

③ To find the population standard deviat

$\boxed{\text{shift}}$ $\boxed{2}$ $\boxed{2}$ $\boxed{=}$

④ To find the variance.

Just square S or σ.

10.76 ب   5 2

بوحدانه مرتب لو منحه تقريب

- EX: Given the data   2 5

25, 30, 47, 80, 56, 62, 74, 80

ⓐ find the mean

$\boxed{Mode}$ $\boxed{2}$ 25 $\boxed{M+}$ 30 $\boxed{M+}$ ...... 80 $\boxed{M+}$

$\boxed{shift}$ $\boxed{2}$ $\boxed{1}$ $\boxed{=}$

$\therefore \bar{X} = 56.75$

ⓑ find the sample standard deviation.

$\boxed{shift}$ $\boxed{2}$ $\boxed{3}$ $\boxed{=}$

$\therefore S = 21.47$

ⓒ find the population standard deviation

$\boxed{shift}$ $\boxed{2}$ $\boxed{2}$ $\boxed{=}$

$\therefore \sigma = 20.08$

ⓓ find the sample variance

21.47 $\boxed{X^2}$ $\boxed{=}$

$\therefore S^2 = 460.96$

78

*ch7: Probability. الاحتمال

- Probability is the basis of inferential statistic Predictions estimations, and hypotheses testing are all based on probability.

الاحتمالات هي الأساس في الاحصاء الاستدلالي

- Probability: The chance of an event occuring.

- Probability Experiment: A chance process that produces results called outcomes. (مخرجات)

التجربة الاحتمالية

- sample space: The set of all possible outcomes of an experiment, denoted by Ω.

الفضاء العيني : جميع المخرجات

*Remark:
$n(\Omega)$: number of total out comes.

عدد عناصر الفضاء العيني العيني

- Ex:

① Toss a coin.

$\rightarrow \Omega = \{Head, Tail\}$

$\therefore n(\Omega) = 2$

زهر

② Rolling a die. رمي حجر نرد

$\rightarrow \Omega = \{1, 2, 3, 4, 5, 6\}$

$n(\Omega) = 6$

③ Determine a child gender. تحديد جنس الطفل

$\rightarrow \Omega = \{\text{Male}, \text{Famale}\}$

$n(\Omega) = 2$

④ Answer a T/F question. اجابة نعم او لا

$\rightarrow \Omega = \{T, F\}$

$n(\Omega) = 2$

- Event: A subset of the sample space, denoted by A, B, C, ...

الحدث: جزء من الفضاء العيني

- EX: In rolling a die رمي حجر النرد

① let $\boxed{A}$ be the event of odd numbers.

الاعداد الفردية

$\rightarrow A = \{1, 3, 5\}$

② let B be the event of numbers greater than 4.

الاعداد الاكبر من 4

$B = \{5, 6\}$.

③ let $\boxed{C}$ be the event of numbers less than or equal to 4.

العدد الأقل من أو تساوي 4

→ $C = \{1, 2, 3, 4\}$.

④ let $\boxed{D}$ be the event of even numbers less than 3.

عدد زوجي أقل من 3

→ $D = \{2\}$.

الأكبر من 6

⑤ let $\boxed{E}$ be the event of numbers greater than 6.

→ $E = \{\}$ or $\emptyset$ « Empty set or null set

المجموعة الخالية

⑥ let $\boxed{F}$ be the event of getting an even number or greater than 4.

الأعداد الزوجية أو الأكبر من 4

→ $F = \{2, 4, 6, 5, 6\}$.

⑦ let $\boxed{G}$ be the event of numbers less that 10.

الأعداد الأقل من 10

→ $G = \{1, 2, 3, 4, 5, 6\} = \Omega$.

* Probability of an event A «P(A)»

احتمال الحادثة الحادث.

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{\text{عدد عناصر الحادث}}{\text{عدد عناصر } \Omega}$$

Ex: for the previous example, find the probability of each event:- $\Omega = \{1,2,3,4,5,6\} \rightarrow n(\Omega) = 6$

① $P(A)$

$\rightarrow P(A) = \frac{n(A)}{n(\Omega)}$, $A = \{1,3,5\} \rightarrow n(A) = 3$

$= \frac{3}{6} = 0.5$

② $P(B)$, $B = \{5,6\} \rightarrow n(B) = 2$

$\rightarrow P(B) = \frac{n(B)}{n(\Omega)}$

$= \frac{2}{6} = 0.333$

③ $P(C)$, $C = \{1,2,3,4\} \rightarrow n(C) = 4$

$\rightarrow P(C) = \frac{n(C)}{n(\Omega)}$

$= \frac{4}{6} = 0.667$

④ $P(D)$

$\rightarrow P(D) = \frac{n(D)}{n(\Omega)}$, $D = \{2\} \rightarrow n(D) = 1$

$= \frac{1}{6} = 0.167$

⑤ $P(E)$, $E = \{\ \} \rightarrow n(E) = 0$

$\longrightarrow P(E) = \dfrac{n(E)}{n(\Omega)}$

$= \dfrac{0}{6} = 0$

⑥ $P(F)$, $F = \{2, 4, 6, 5\} \rightarrow P(F) = 4$

$\rightarrow P(F) = \dfrac{n(F)}{n(\Omega)}$

$\Rightarrow = \dfrac{4}{6} = 0.667$

⑦ ~~P(G)~~ $P(G)$, $G = \{1, 2, 3, 4, 5, 6\} \rightarrow n(G) = 6$

$\rightarrow$ ~~P(G)~~ $P(G) = \dfrac{n(G)}{n(\Omega)}$

$= \dfrac{6}{6} = 1$

* Remark:- ملاحظة

① $0 \leq P(A) \leq 1$, for any event A. قيمة أي احتمال محصورة من 0 الى 1

② $P(\emptyset) = 0$ «always»

③ $P(\Omega) = 1$ «always»

-EX: Toss a coin

$\rightarrow \Omega = \{H, T\}$

① $P(H) = \frac{1}{2} = 0.5$

② $P(T) = \frac{1}{2} = 0.5$

-EX: A box contains 4 red balls, 5 black balls, and a green ball. suppose you want to draw a ball.

① what is the probability of drawing a red ball?

$\rightarrow P(\text{red}) = \frac{n(\text{red})}{n(\text{red})} = \frac{4}{10} = 0.4$

1 green
5 black
4 red

② what is the probability of drawing a green ball?

$\rightarrow P(\text{green}) = \frac{1}{10} = 0.1$

③ what is the probability of drawing a black ball?

$\rightarrow P(\text{black}) = \frac{5}{10} = 0.5$

* Multiple step experiment:
An experiment with more than one step.
التجربة متعددة المراحل:

- Ex:

① Roll a die twice.
② Toss a coin 3 times.
③ Roll a die than toss a coin.

- Note:

The total number of the outcomes of a multiple-step experiment is the product of the number of outcomes of the steps.
عدد عناصر Ω = حاصل ضرب عدد المخرجات في كل محاولة

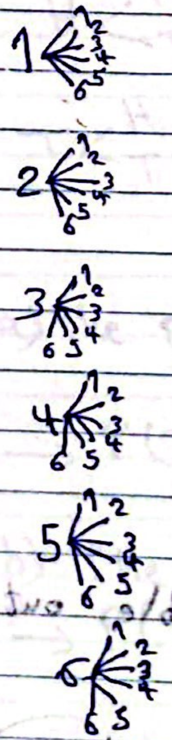
- Ex: Toss a coin twice.

→ The total number of outcomes "$n(\Omega)$" $= (2)(2) = 4$

$\Omega = [(H,H), (H,T), (T,H), (T,T)]$.

- Ex: Roll a die twice.

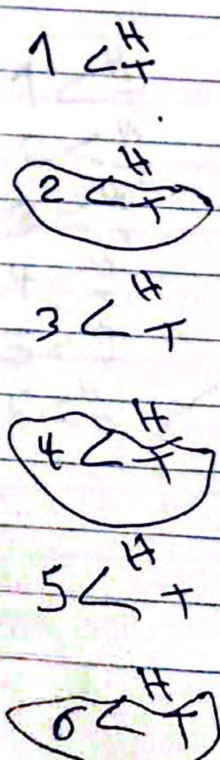
$\rightarrow n(\Omega) = (6)(6) = 36$



$\therefore \Omega = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)$
$(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)$
$(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)$
$(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)$
$(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)$
$(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$

- Ex: Roll a ~~die~~ die then toss a coin

$\rightarrow n(\Omega) = (6)(2) = 12$



$\therefore \Omega = \{(1,H), (1,T), (2,H), (2,T)$
$(3,H), (3,T), (4,H), (4,T)$
$(5,H), (5,T), (6,H), (6,T)\}$

$6 \times 6 \times 2 \times 2 \times 2 \times 2$

* Tree diagram:

A graphical tool used to find all possible outcomes of a a multiple-step experiment.

- Ex: In rolling a die twice, than tossing a coin 3 times. find the total number of outcomes

$$\rightarrow n(\Omega) = (6)(6)(2)(2)(2)$$
$$= 288$$

- Ex: In tossing a coin twice, find the following:-

$H <^{H}_{T}$

$T <^{H}_{T}$

$\Omega = \{(HH)(HT)(TH)(TT)\}$

a) The probability of getting no heads?

$\rightarrow P(TT) = \frac{1}{4} = 0.25$

b) The probability of getting one head?

$\rightarrow P((HT)(TH)) = \frac{2}{4} = 0.5$

c) The probability of getting at ~~most~~ least one head

$P((HT)(TH)(HH)) = \frac{3}{4} = 0.75$

- Ex: In rolling a die then tossing a coin:-

$1 <^{H}_{T}$

$2 <^{H}_{T}$

$3 <^{H}_{T}$

$4 <^{H}_{T}$

$5 <^{H}_{T}$

$6 <^{H}_{T}$

$\Omega = \{(1H)(1T)(2H)(2T)(3H)(3T)(4H)(4T)(5H)(5T)(6H)(6T)\}$.

H

i) The probability of getting a head.

$\rightarrow P(\{(1H)(2H)(3H)(4H)(5H)(6H)\}) = \frac{6}{12} = 0.5$

ii) The probability of getting an odd number and a tail.

$\rightarrow P(\{(1T)(3T)(5T)\}) = \frac{3}{12} = 0.25$

* Probability distributions and ~~rand~~ random variables

• Random Variable: A variable x that we can choose in a probability experiment.

المتغير العشوائي: متغير يأخذ قيمة من التجربة الاحتمالية بناء على تعريفه

• Probability distribution: A table that shows the probability of the values of a random variable X in some experiment.

جدول التوزيع الاحتمالي: جدول يوضح قيم المتغير العشوائي مع احتمال كل قيمة

- EX: In Tossing a coin twice. let X be a random variable defined by: X = # of heads.

H < H, T

T < H, T

$\Omega = \{(HH)(HT)(TH)(TT)\}$

2 1 1 0

a) Find the values of X.

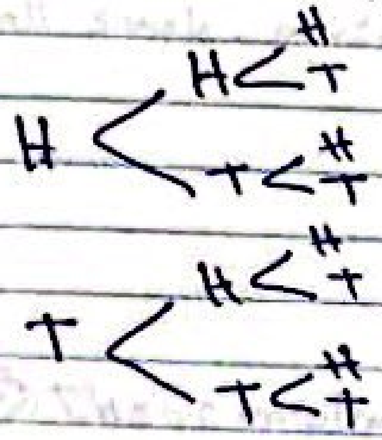
X = 0, 1, 2

b) constrakitnon probability distribution of X.

X	0	1	2
P(X)	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

-EX: In tossing a coin 3 times. let X be the number of tails.

a) find the sample space.



$\Omega = \{(HHH)\ (HHT)\ (HT$
$(HTT)\ (THH)\ (TH$
$(TTH)\ (TTT)\}$

b) find the values of X

→ X = 0, 1, 2, 3

c) Construct a probability distribution.

X	0	1	2	3
P(X)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

- Note: Let X be a random variable, then
  1. $0 \le P(X) \le 1$
  2. $\sum P(X) = 1$

- EX: Is this probability distribution valid? صحيح

هل يشكل جدول توزيع احتمالي

①

X	0	1	2	3
$P(X)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{-1}{5}$	$\frac{3}{5}$

→ No, since $P(2) = \frac{-1}{5}$ ✗

②

X	-1	0	3
$P(X)$	$\frac{1}{4}$	$\frac{2}{4}$ or $\frac{1}{2} = \frac{2}{4}$	$\frac{1}{4}$

→ Yes, since $\sum P(X) = 1$ and $0 \le P(X) \le 1$

③

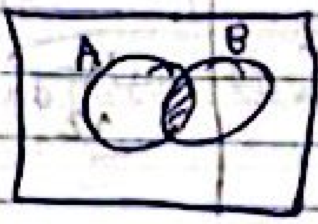
X	1	-2	-1
$P(X)$	$\frac{3}{8}$	$\frac{5}{8}$	$\frac{1}{4}$

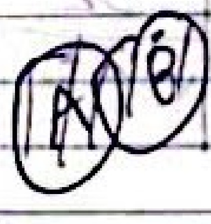
→ $\underline{\underline{\text{No}}}$, since $\frac{3}{8} + \frac{5}{8} + \frac{1}{4} \ne 1$

أنا كتبتها بالمحاضرة

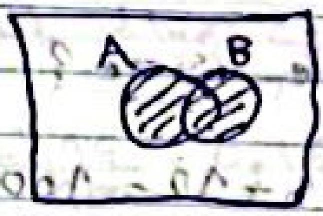
* some Basic relationships:-

1) The intersection of A and B: $A \cap B$
the event containing the sample points belongin both A and B.
التقاطع

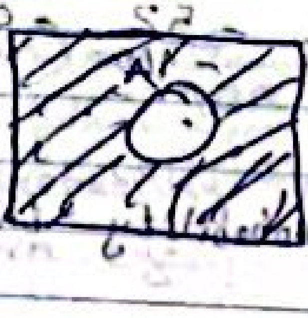


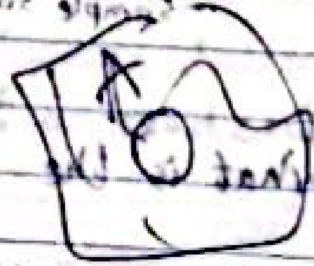


2) Union of two event A and B: $A \cup B$
all sample points belonging to A or B or both.
أو
الاتحاد



3) The complement of A: $\bar{A}$
all sample points that are not in A.





* Cross table:
is a tabular summary of data for 2 variables.
جدول نظم البيانات المأخوذة لمتغيرين.

-EX! The cross table below summarizes the gender and eyes color of a sample of students.

Gender \ Eyes color	brown	blue	Green
famale	32	16	4
male	28	8	12

we will use the table to answer the following questions!

① what is the sample size?

$32 + 16 + 4 + 28 + 8 + 12 = 100$

② what is the probability that a student is female

$P(F) = \frac{n(F)}{\text{sample size}} = \frac{52}{100} = 0.52$

③ what is the probability that a student is male?

$P(M) = \frac{n(M)}{\text{sample size}} = \frac{48}{100} = 0.48$

④ what is the probability that a student has brown eyes?

$P(\text{brown}) = \frac{60}{100} = 0.6$

(5) what is the probability that a student has blue eyes?

$$P(blue) = \frac{24}{100} = 0.24$$

(6) what is the probability that a student has green eyes?

$$P(green) = \frac{16}{100} = 0.16$$

(7) what is the probability that a student is female and has blue eyes?

$$\rightarrow P(F \cap blue) = \frac{16}{100} = 0.16$$

(8) what is the probability that a student is male and has green eyes?

$$\rightarrow P(M \cap green) = \frac{12}{100} = 0.12$$

(9) what is the probability that a student has both blue and brown eyes?

$$\rightarrow P(blue \cap brown) = \frac{0}{100} = 0$$

* The probability of union of events: ($\cup$: or) (7)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Ex: for the previous example. find the following:-

(10) what is the probability that a student is male or has blue eyes?

$$\rightarrow P(\text{male} \cup \text{blue}) = P(\text{male}) + P(\text{blue}) - P(\text{male} \cap \text{blue}).$$

$$= \frac{48}{100} + \frac{24}{100} - \frac{8}{100} = 0.64$$

(11) what is the probability that a student is female or has brown eyes?

~~$P(\text{female} \cup \text{brown}) = P(\text{female}) + P(\text{brown}) - P(\text{female}$~~

$$\rightarrow P(\text{female} \cup \text{brown}) = P(\text{female}) + P(\text{brown}) - P(\text{femal} \cap \text{brown})$$

$$= \frac{52}{100} + \frac{60}{100} - \frac{32}{100} = 0.8$$

⑫ what is the probability that a student has either blue or brown ~~ey~~ eyes?

$\longrightarrow P(\text{blue} \cup \text{brown}) = P(\text{blue}) + P(\text{brown}) - P(\text{blue} \cap \text{bro}$

$$= \frac{24}{100} + \frac{60}{100} - 0 = 0.84$$

بعد علمنا بأن

• The conditional ~~probaba~~ probability: (I: given that)

~~$\rightarrow P(A|B) = P(AA$~~

الاحتمال المشروط:
احتمال حصول حدث بشرط
حصول حدث آخر.

$$\longrightarrow P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\longrightarrow P(B|A) = \frac{P(A \cap B)}{P(A)}$$

—Ex: for the previous example. find the followings:

⑬ what is the probability that a student is female given that she has green eyes?

$$\longrightarrow P(F \mid \text{green}) = \frac{P(F \cap \text{green})}{P(\text{green})} = \frac{4/100}{16/100} =$$

(14) What is the probability that a student has blue eyes if he is a male?

$\rightarrow P(\text{blue} \mid \text{male}) = \dfrac{8/100}{48/100} = 0.17$

* Two events A, B are called <u>Mutually exclusive</u> they cannot occure at the same time.

الحوادث المتباعدة أو المنفصلة
لا تحدث بنفس الوقت يعني أن التقاطع = 0
احتمال

$\rightarrow P(A \cap B) = 0$
and $P(A \cup B) = P(A) + P(B)$.

- Ex: (1) The male and female.
  (2) The brown and green eyes.
  (3) true/false question
  (4) odd/even number.

* Two events A, B are called <u>independent</u> if the occurence of one does not affect the other.

الحوادث المستقلة احدوث الأول
لا يؤثر على نسبة حدوث الثاني

$\rightarrow P(A \mid B) = P(A)$
$P(B \mid A) = P(B)$
and $P(A \cap B) = P(A) \cdot P(B)$

- Ex: (1) In tossing a coin, the first and second tosses are independent.

(14) what is the probability that a student has blue eyes if he is a male?

$\rightarrow P(\text{blue} \mid \text{male}) = \frac{8/100}{48/100} = 0.17$

* Two events A, B are called Mutually exclusive they cannot occure at the same time.

الحوادث المتباعدة أو المنفصلة
لا تحدث بنفس الوقت يعني أن التقاطع = 0
الاحتمال

$\rightarrow P(A \cap B) = 0$
and $P(A \cup B) = P(A) + P(B)$.

- Ex: ① The male and female.
② The brown and green eyes.
③ true/false question
④ odd/even number.

* Two events A, B are called independent if the occurence of one does not affect the other

الحوادث المستقلة حدوث الأول
لا يؤثر على نسبة حدوث الثاني

$\rightarrow P(A \mid B) = P(A)$
$P(B \mid A) = P(B)$
and $P(A \cap B) = P(A) \cdot P(B)$

- Ex: ① In tossing a coin, the first and second tosses are independent.

EX: If $P(A) = 0.4$, $P(B) = 0.6$ and $P(A \cap B) = 0.24$

i) find $P(A \cup B)$.

$\rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= 0.4 + 0.6 - 0.24$

$= 0.76$

ii) find $P(B|A)$.

$$\rightarrow P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.24}{0.4} = 0.6$$

iii) Are A and B mutually exclusive? why?

$\rightarrow$ No, since $P(A \cap B) \neq 0$

iv) Are A and B independent? why?

$\rightarrow$ Yes, because $P(B|A) = P(B)$.

امتحان

- EX1 The following is a cross tabulation summ
of a sample of BZU's student.

Gender	1st year	2nd year	3rd year	4th year	Total
Male	90	70	45	25	230
famale	100	80	55	25	270
Total	200	150	100	50	500

a) what is the probability that a student is a 4th
$\rightarrow P(4th\ year) = \frac{50}{500} = 0.1$

b) what is the probability that ~~a studen~~ a studen is 3rd year student given that she is a female?

$$\rightarrow P(3rd \mid F) = \frac{P(3rd \cap F)}{P(F)}$$

$$= \frac{55/500}{270/500} = 0.2$$

c) what is the probability that a student is eith a 1st year or a 2nd year?

$$\rightarrow P(1st \cup 2nd) = P(1st) + P(2nd) - P(1st \cap 2$$

$$= \frac{200}{500} + \frac{150}{500} - 0$$

$$= \frac{350}{500} = 0.7$$