• Chapter 2: Descriptive statistics:

- Sec 2.1: Summarizing Qualitative Data: تمثيل البيانات النوعية

*Descriptive statistics: organizing data by tables, graphs or numerical measures. الإحصاء الوصفي

→ Qualitative data:

① By tables:

I Frequency distribution: جدول التوزيع التكراري

Is a tabular summary of data showing the number of items (frequency) التكرار in each of nonoverlapping classes (categories). الفئات

- Ex: The following data represents the blood type of 10 students.

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

Construct a frequency distribution.

Class	freq. «f»	tally
A	3	///
B	1	/
AB	2	//
O	4	////

* عند رسم الإشارات «tally» نرسم خط لتمثيل كل تكرار كل 5 نضعهم في مجموعة أي لو كان التكرار 7 ##// .

II Relative frequency and percent frequency distribution

→ Relative frequency of a class = $\frac{\text{freq. of the class}}{\text{total size}}$
التكرار النسبي

$$r.f = \frac{f}{n} \quad ; f: \text{frequency}, n: \text{total size}.$$

→ Percent frequency of a class = $\frac{f}{n} \times 100\%$
التكرار المئوي

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

• Note:

① $\sum f = n$.

② $\sum r.f = 1$.

③ $\sum p.f = 100\%$.

④ relative frequency ≡ proportion ≡ probability.

⑤ percent frequency ≡ percentage.

-Ex: Construct the relative frequency and percent frequency for the blood type example.

→

Class	f	r.f	p.f
A	3	$\frac{3}{10}=0.3$	$0.3\times100=30\%$
B	1	$\frac{1}{10}=0.1$	10%
AB	2	0.2	20%
O	4	0.4	40%

② By graphs:

I Bar graph: التمثيل بالأعمدة

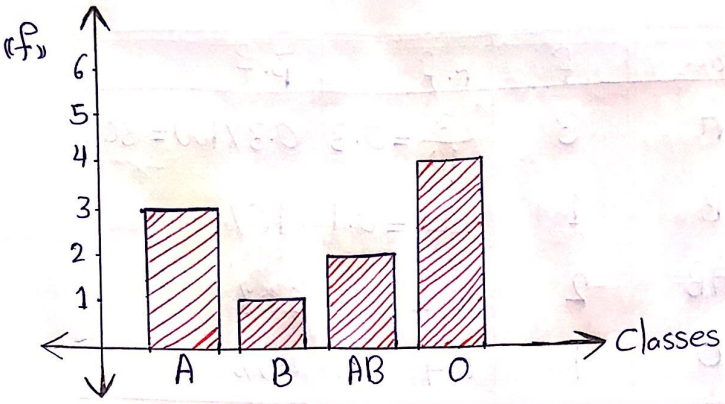
is a graphical device for depicting qualitative data summarized in a frequency, relative frequency or percent frequency.

→ horizontal axis: labels (classes) = categories
vertical axis: frequency, relative freq., percent freq.

→ the bars should be separated to emphasize the fact that each class is separate. يوجد فراغات بين الأعمدة

-Ex: Construct a bar graph for the blood type example. (with f)

→



II Pie chart: التمثيل بالقطاعات الدائرية

is a graphical device for presenting relative frequency and percent frequency for qualitative data.

→ We draw a circle, then we use the relative frequency to subdevide the circle into sectors or قطاعات parts.

→ the angle of each sector $= \frac{f}{n} \times 360 = r.f \times 360^\circ$

; f: frequency.
n: total size.

- Ex: Construct a pie chart for the blood type example. (with p.f).

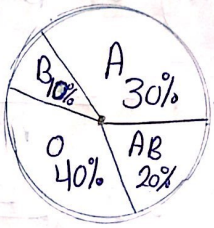
→ firstly, we find the angles of each sector

A: $0.3 \times 360 = 108°$

B: $0.1 \times 360 = 36°$

AB: $0.2 \times 360 = 72°$

O: $0.4 \times 360 = 144°$



-Note: $\sum angles = 360°$

-Ex: A partial relative frequency distribution is given:

Class	r.f
A	0.22
B	0.18
C	0.4
D	

a) What is the relative frequency of class D?

$\longrightarrow \sum f = 1$

$0.22 + 0.18 + 0.4 + x = 1$

$0.8 + x = 1$

$\therefore \boxed{x = 0.2}$

b) If the total sample size is 200. What is the frequency of D?

$$\longrightarrow \text{r.f} = \frac{f}{n}$$

$$0.2 = \frac{f}{200}$$

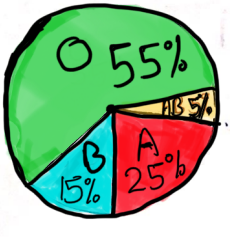
$$\therefore f = 200(0.2) = 40$$

note: $f = n(\text{r.f})$

c) Show the frequency distribution.

Class	r.f	f
A	0.22	44
B	0.18	36
C	0.4	80
D	0.2	40

Ex: The pie chart below represents the blood type of 40 students.



Find the number of students with blood type O.

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

$$55\% = \frac{f}{40} \times 100\%$$

$$0.55 = \frac{f}{40}$$

$$\therefore f = 40(0.55) = 22$$

-Sec 2.2: Summarizing Quantitative data

① By tables:

I Frequency distribution: جدول التكرار النسبي

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

* 3 steps to define the classes.

1. Determine the number of nonoverlappin classes. عدد الفئات
2. Determine the width of each class. طول الفئة
3. Determine the class limits. حدود الفئة

→ Number of classes:

Usually we use between 5 and 20 classes.

الباحث من يُقرر عدد الفئات بحيث يكون ملائم للبيانات ويغطي التنوع بينها ويرجع عدد الفئات على الدراسة.

→ Class width = $\frac{\text{largest data value} - \text{smallest data value}}{\text{number of classes}}$

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

9.2 might be rounded to → 10

→ Class limits:

lower limit – upper limit

نختار الرقم بحيث يغطي أغلب البيانات بدءاً أعادةً بأصغر رقم أو قريب منه.

- Ex: The following data represent the waiting times in months for patients who arrive at the office with a request for emergency service.

2, 5, 10, 12, 4, 4, 5, 17, 11, 8, 9,
12, 21, 6, 8, 7, 13, 18, 3, 19

Show the freq., r.f, P.f distributions. (Use 6 classes)

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

$= \frac{21-2}{6} = 3.16 \rightarrow 4$

Class	f	r.f	P.f
1 – 4	4	4/20 = 0.2	20%
5 – 8	6	6/20 = 0.3	30%
9 – 12	5	5/20 = 0.25	25%
13 – 16	1	1/20 = 0.05	5%
17 – 20	3	3/20 = 0.15	15%
21 – 24	1	1/20 = 0.05	5%

*Rules:

① Class midpoint = $\frac{\text{lower limit} + \text{upper limit}}{2}$ مركز الفئة

② True limits (الحدود الفعلية)
- true lower limit = lower limit − 0.5
- true upper limit = upper limit + 0.5

③ Class width (طول الفئة) = U − L + 1

II Comulative frequency distribution المتجمع الصاعد (التكراري التراكمي)

the comulative frequency distribution shows the number of data items with values less than or equal to the upper limit of each class. عدد التكرارات التي تكون أقل أو تساوي الحد العلوي.

-Ex: The following data represents the weights of 15 students

49, 60, 53, 52, 70, 81, 58, 63, 87,
91, 65, 74, 85, 67, 87

a) Construct a frequency, relative frequency, and percent frequency. (Use 4 classes).

→ Class width $= \frac{91-49}{4} = 10.5 \rightarrow 11$

a) b)

Class	f	r.f	p.f	c.f	c.r.f	c.p.f
49–59	4	$\frac{4}{15}=0.267$	26.7%	4	$\frac{4}{15}=0.267$	26.7%
60–70	5	0.333	33.3%	9	0.6	60%
71–81	2	0.133	13.3%	11	0.733	73.3%
82–92	4	0.267	26.7%	15	1	100%

b) Construct a comulative frequency, comulative relative freq. , and percent comulative freq.

→ Comulative relative freq. $= \frac{c.f}{n}$ or by summing the relative frequency. المتجمع النسبي الصاعد

→ Comulative percent freq. $= \frac{c.f}{n} \times 100\% = c.r.f \times 100\%$
المتجمع المئوي الصاعد.

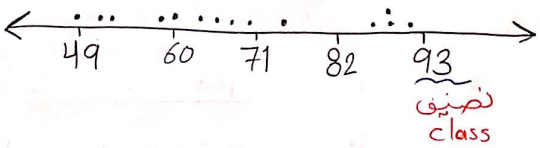
(2) Graphs for quantitative data: التمثيل بالرسم للبيانات الرقمية

a) Dot plot: النقاط

→ horizontal axis: the range for the data (lower limits).

Each data value is represented by a dot placed above the axis.

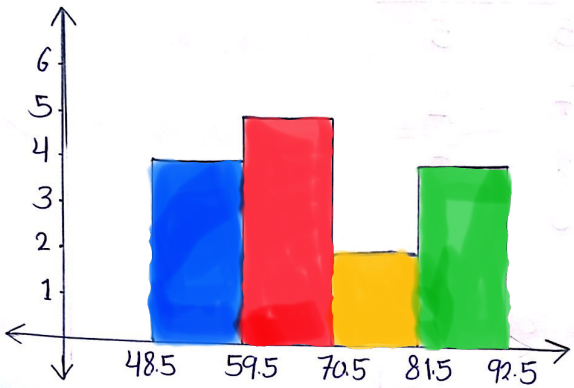
-Ex: For the previous example, construct a dot plot.



b) Histogram: المدرج التكراري

→ horizontal axis: Classes (true limits) الحدود الفعلية
vertical axis: freq., r.f, or p.f.
no natural separation between rectangles.
لا يوجد فراغات بين الأعمدة.

-Ex: Construct a Histogram for the previous example.



c) Ogive: منحنى المتجمع الصاعد/المنحنى التراكمي

A graph of a comulative distribution.

→ horizontal axis: data values (true upper limit).
vertical axis: Comulative frequency, relative comulative frequency, or percent comulative frequency.

→ It's increasing. متزايد

-Ex: Consider the following table:

Class	c.f
10-14	5
15-19	7
20-24	14
25-29	20

a) Construct a frequency distribution.

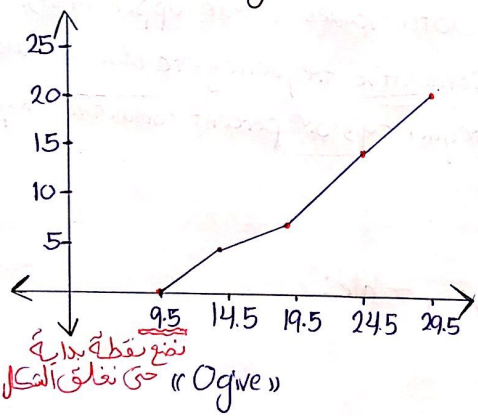
Class	c.f	f
10-14	5	5
15-19	7	2
20-24	14	7
25-29	20	6

b) What is the class width.

→ class width = U − L + 1

= 14 − 10 + 1 = 5

c) Construct an ogive.



نضع نقطة بداية حتى نغلق الشكل.

« Ogive »

* Distribution Shapes: شكل التوزيع

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

نحدد شكل التوزيع بناء على رسمة histogram.

• Distribution shape:

describes how tha data values of a variable are distributed.

يصف لنا كيفية توزيع البيانات.

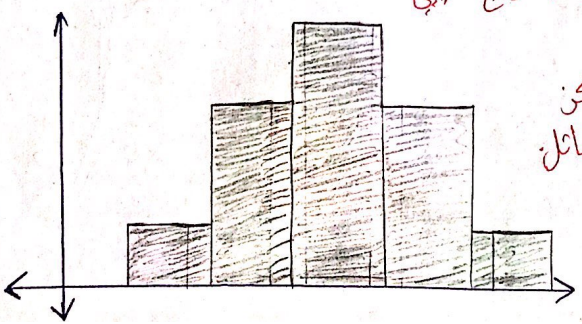
-Types of distribution shapes:

① Normal distribution:- التوزيع الطبيعي

→ Data values cluster around the center «middle». تتمركز في المنتصف

→ 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 والأطول تطبيقات على التوزيع الطبيعي

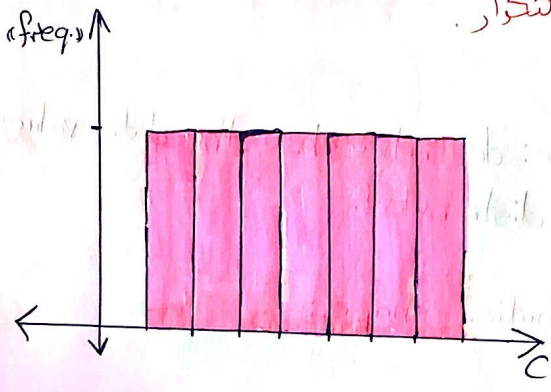


معظم البيانات تتمركز في المنتصف والشكل متماثل.

② The uniform distribution: التوزيع المنتظم

→ data values have same frequencies.

← جميع الفئات لها نفس التكرار.



③ Positively skewed: المنحاز يميناً

→ data values cluster on left. تتمركز في اليسار

→ its tail (ذيل) extends to the right.

→ An example of this type: ① housing prices, ② scores in a difficult exam.

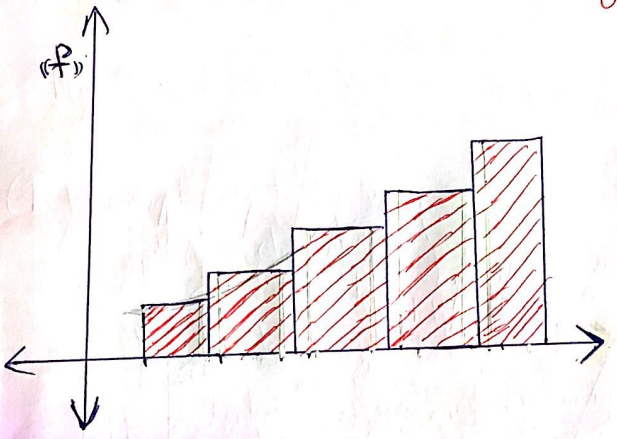
أمثلة عليها: أسعار المساكن، علامات في امتحان صعب.

(4) Negatively skewed: المنحنى يسارًا

→ data values cluster on right. تتمركز في اليمين

→ its tail extends to the left.

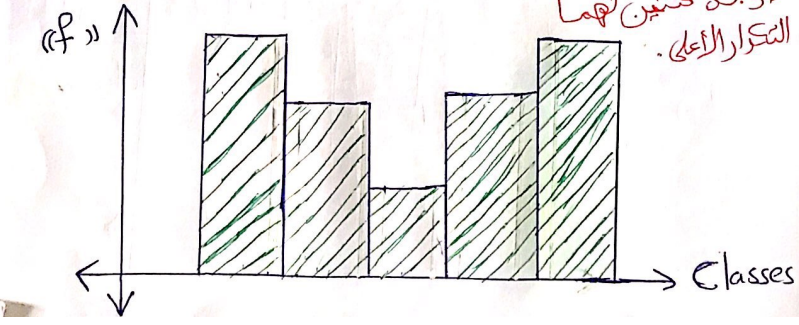
→ An example of this type: scores in an easy exam.
العلامات في امتحان سهل



(5) Bimodal: ثنائي المنوال

«استطرق له في Ch5»

→ We have 2 modes. لهما منوالين أي أنه يوجد قيمتين لهما التكرار الأعلى.



- Sec 2.4 : Crosstabulations and Scatter diagrams:-

1 **Crosstabulation:** الجداول المتقاطعة

is a tabular Summary of data for 2 variables.
تمثيل متغيرين

→ The two variables may be:-

1. 2 variables qualitative.
2. 2 variables quantitative.
3. 1 variable qualitative and 1 variable quantitative.

- Ex: The following data represents the quality rating and meal price for 300 restaurants:

Restaurant	Quality rating	Meal price.
1	Good	18
2	Very good	22
3	Good	28
4	Excellent	38
5	Very good	33
⋮	⋮	⋮
300	Good	13

a) Construct a crosstabulation of quality rating and meal price:

Quality rating \ Meal price	10–19	20–29	30–39	40–49	Total
Good	42	40	2	0	84
Very good	34	64	46	6	150
Excellent	2	14	28	22	66
Total	78	118	76	28	300

Note that we use the class width = 10

b) how many elements.

# of elements = sample size = 300

sample since it is, study type of is survey.

c) find the frequency for the good restaurant and its meal price less than or equal 19 ?

$$f = 42$$

$$r.f = \frac{42}{300} = 0.14$$

$$P.f = \frac{42}{300} \times 100\% = 14\%$$

- Note: the relative frequency = probability.

d) find the percentage of very good restaurant.

$$\frac{150}{300} \times 100\% = 50\%$$

e) find the percentage of restaurants with meal price greater than 19?

$$\frac{118+76+28}{300} \times 100\% = 74\%$$

f) find the percentage of restaurants who are Excellent and with price greater than 39?

$$\frac{22}{300} \times 100\% = 7.3\%$$

h) find the probability of restaurants with meal price greater than 39? r.f

$$\frac{28}{300} = 0.0933$$

* Row percentages $= \frac{f}{\text{total raw}} \times 100\%$

m) Construct a row percentage.

	10-19	20-29	30-39	40-49	
Good	42/84	40/84	2/84	0	100%
very good	34/150	64/150	46/150	6/150	100%
Excellent	2/66	14/66	28/66	22/66	100%

Note that: the sum of each row is 100%

* Column percentage = $\frac{f}{\text{total column}} \times 100\%$

(2) Scatter diagram and trendline:- الرسم المبعثر

is a graphical presentation of the relationship between 2 quantitative variables. لرسم متغيرين رقميين.

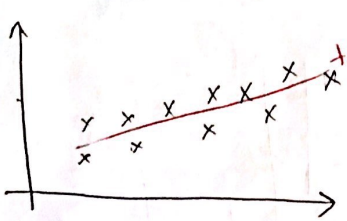
- trendline: the line that provides an approximation of the relationship. أفضل خط قريب على النقاط.

(x, y): orderd pair
x: the first variable
y: the second variable.

* Shape of relationship
  - lineaw. التركيز على linear
  - Quadratic.
  - erponential.

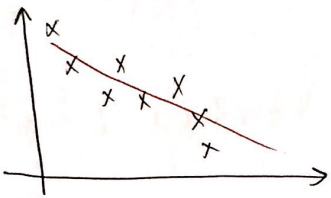
سيتم التعلم بطريقة أوسع في Ch.12

(1)



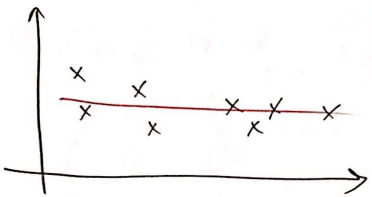
positive relationship.

2



negative relation.

3



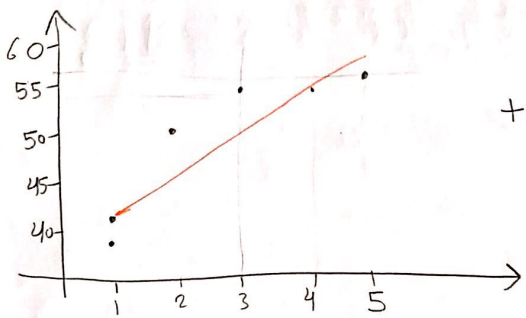
No relation.

-Ex:

Number of TV commercials	Sales
2	50
5	57
1	41
3	54
4	38
	54

نلاحظ أن المتغيرين quantitative

- The scatter diagram and trend line :-



+ve relationship

* Stat 2331

- Sheet 1: Ch1+2

Q1 The prices ($ million) of some football players are summarized below:-

Prices	F
6–15	24
16–25	26
26–3	10
36–45	8
46–55	5
56–65	7

① What is scale level of measurment of the data? المقياس

→ Interval لأن البيانات على شكل أرقام

② Are the data quantitative or qualitative?

quantitative. بيانات رقمية

③ What is the class width? طول الفئة

→ Class width = U − L + 1
= 15 − 6 + 1
= 10

أو الفرق بين أي حدين علويين أو سفليين.

④ What is the midpoint of the class: 56–65? مركز الفئة

→ midpoint = $\frac{U+L}{2}$ $\frac{\text{الحد العلوي + السفلي}}{2}$

= $\frac{65+65}{2}$ = 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 of players with prices between 15 and 56? التكرار بين

→ 26 + 10 + 8 + 5 = 49

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

(7) What is the probability (احتمال) = R.f that a player's price is more than (أكثر من) 35 million?

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

أكبر من 35

$8+5+7=20$ probability الاحتمال أي

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

(8) What is the percentage (النسبة المئوية) of players whose price are at most (على الأكثر) 45 million? = P.r.f

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

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

$= 85\%$

Classes	F
6-15	24
16-25	26
26-35	10
36-45	8
46-55	
56-65	

على الأكثر 45 أي 45 أو أقل

$24+26+10+8 = 68$

(9) Construct a relative frequency distribution?

(10) Construct a percent frequency distribution for these data?

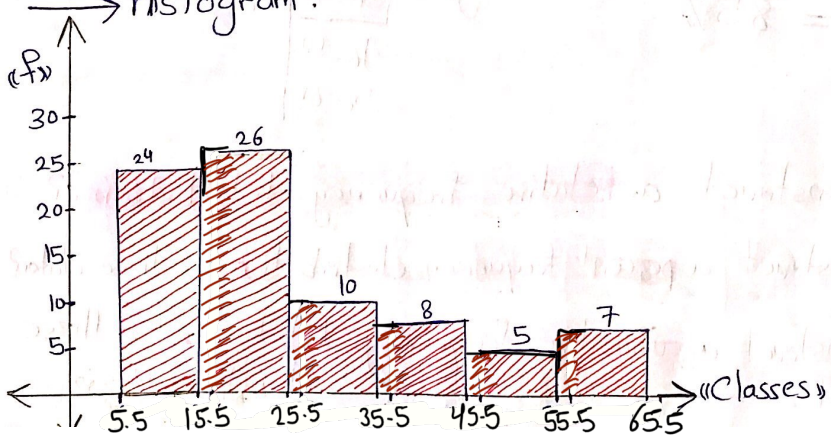
(11) Construct a cumulative frequency distribution for these data?

في جدول واحد (11) + (10) + (9) نوع

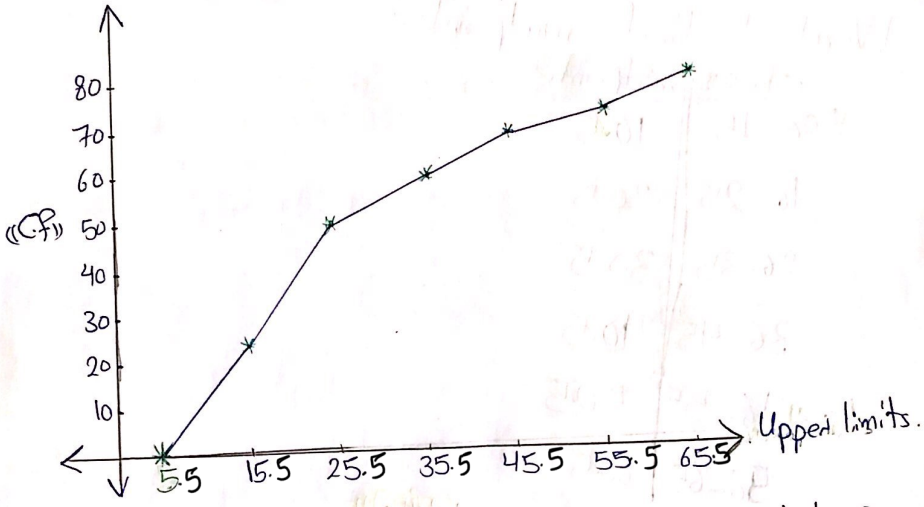
Classes	F	R.f = $\frac{f}{n}$	P.r.f = $\frac{f}{n}$ x100%	C.f
6 – 15	24	$\frac{24}{80}$ = 0.3	$\frac{24}{80}$ x100% = 30%	24
16 – 25	26	0.325	32.5%	24+26=50
26 – 35	10	0.125	12.5%	50+10=60
36 – 45	8	0.1	10%	60+8=68
46 – 55	5	0.0625	6.25%	68+5=73
56 – 65	7	0.0875	8.75%	73+7=80

(12) Construct a histogram, and ogive for these data?

→ Histogram:



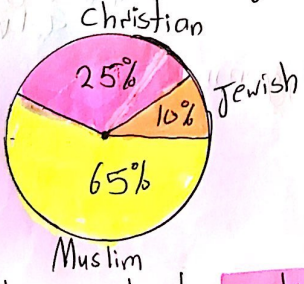
→ Ogive:



(13) What is the distribution shape for these data?
شكل التوزيع

→ Based on Histogram:
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

* 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 frequency distribution:-

Classes	f
M	325
J	50
C	125