

3

Ch. 2 - Part 1

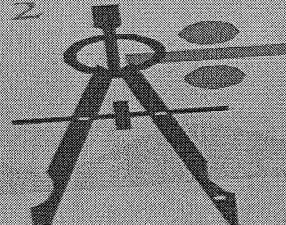
- Organizing Data.
- Histogram , frequency Polygons, and Ogives.

STAT.110

جمال السعدي
رياضيات - إحصاء

6.15%

7²



CH. 2 PART 1

جمال السعدي

Frequency distributions and graphs

التوزيعات
التكرارية
والرسومات

Organizing data: for the raw data

تنظيم البيانات
لبيانات الخام

When data are collected in original form, they are called raw data.

* عندما يجمع البيانات بشكلها الأصلي تسمى البيانات الخام.

When the raw data are organized into a table which is called frequency distribution, the frequency will be the number of values in a specific class of the distribution.

* عندما تنظم البيانات من جدول يسمى التوزيع التكراري

* التكرار: هو عدد القيم داخل فئة محددة من التوزيع.

A frequency distribution is the organization of raw data in table form, using classes and frequencies.

* التوزيع التكراري:

هو تنظيم البيانات الخام من جدول باستخدام الفئات والتكرار.

Types of frequency distributions are categorical frequency distribution, ungrouped frequency distribution and grouped frequency distribution.

* أنواع التوزيعات التكرارية

توزيعات تكرار نوعيه (ليست رقميه)
 (1) Categorical frequency distributions:

Is used for: nominal data or ordinal data

Example:

Twenty – five army inductees were given a blood test to determine their blood type. The data set is

A	B	B	AB	O
O	O	B	AB	B
B	B	O	A	O
A	O	O	O	AB
AB	A	O	B	A

Construct a frequency distribution for the data

Solution:

Class	Tally	Frequency F	Percent P
A		5	20
B		7	28
O		9	36
AB		4	16
		$n = \sum F = 25$	100

$$P = \frac{F}{n} \cdot 100\%$$

- For the sample, people have type O blood more than any other type

التوزيعات التكرارية المجمعه

(2) Grouped frequency distributions (للبيانات الكمية)

Is used when the range of the data is large.

The data must be grouped into classes.

- The width of the class must be > 1
- The number of the classes should be between $\underline{5}$ and $\underline{20}$

Example:

These data represent the record high temperatures in each 50 states construct a grouped frequency distribution for the data using 7 classes.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Solution

$$* \text{ Range} = H - L = 134 - 100 = 34$$

$$* \text{ Width} = \frac{R}{\text{Number of classes}} = \frac{34}{7} = 4.9$$

$$\text{Rounded up} = 5$$

Class limits	Class boundaries	Tally	Frequency	Cumulative frequency
100 - 104	99.5 - 104.5	//	2	2
105 - 109	104.5 - 109.5	////	8	10
110 - 114	109.5 - 114.5	//////	18	28
115 - 119	114.5 - 119.5	//////	13	41
120 - 124	119.5 - 124.5	////	7	48
125 - 129	124.5 - 129.5	/	1	49
130 - 134	129.5 - 134.5	/	1	50

$$n = \sum f = 50$$

Number of classes can be found by using the formula:

$$\text{Number of classes} \approx 1 + 3.3 \times \log (n)$$

→ قانون

Where n is sample size.

Example:

If sample size $n = 50$, Find the number of classes.

Solution

$$\text{Number of classes} \approx 1 + 3.3 \times \log (n)$$

$$\approx 1 + 3.3 \times \log (50) \quad \text{بالآلة}$$

$$\approx 6.6066$$

$$\approx \boxed{7}$$

(3) Ungrouped frequency distributions:

are used for data that can be enumerated and when the range of values in the data set is small (discrete data) and the sample size (n) is large.

- Example: * number of children per family,
* number of cars in a parking.
- Example: Number of patients in the waiting rooms of 16 clinics within a hospital at a specific.

Represent the flowing data by using:

Ungrouped frequency distributions

5	4	4	8
8	5	8	4
4	4	8	4
5	8	4	4

Solution

Class	Frequency	Cumulative Frequency	Percent
4	8	8	50%
5	3	11	19%
8	5	16	31%
Total	16	-	100%

Shown here are four frequency distributions each is incorrectly constructed state the reason why. *توزيعات تكرر غير متساوية*

a.	class	Frequency
6	27 - 32	1
6	33 - 38	0
6	39 - 44	6
6	45 - 49	4
5	50 - 55	2

منشأه بطريقة خاطئة
وضع سبب التوافق.

غير متساوي

⇒ class width is not uniform

b.	class	Frequency
	5 - 9	1
	9 - 13	2
	13 - 17	5
	17 - 20	6
	20 - 24	3

متداخلة

⇒ class limits overlap and class width is not uniform

c.	class	Frequency
	123 - 127	3
	128 - 132	7
	138 - 142	2
	143 - 147	19

فئة ناقصة

⇒ A class has been omitted

d.	class	Frequency
5	9 - 13	1
6	14 - 19	6
6	20 - 25	2
6	26 - 28	5
3	29 - 32	9

غير متساوي

⇒ class width is not uniform

Find: the class boundaries, Midpoints and width for each class:

(1) 12 -18

(2) 13.6 - 14.7

(3) 2.15 - 3.93

Solution

(1) 12 -18

* Class boundaries

11.5 - 18.5

$$* \text{Midpoint} = \frac{12 + 18}{2} = \frac{30}{2} = 15$$

$$* \text{Class width} = \text{upper boundary} - \text{lower boundary} = 18.5 - 11.5 = 7$$

(2) 13.6 - 14.7

* Class boundaries

13.55 - 14.75

$$* \text{Midpoint} = \frac{13.6 + 14.7}{2} = 14.15$$

$$* \text{Class width} = \text{upper boundary} - \text{lower boundary} = 14.75 - 13.55 = 1.2$$

(3) 2.15 - 3.93

* Class boundaries

2.145 - 3.935

$$* \text{Midpoint} = \frac{2.15 + 3.93}{2} = 3.04$$

$$* \text{Class width} = \text{upper boundary} - \text{lower boundary} = 3.935 - 2.145 = 1.79$$

Graphs

الرسومات البيانية

● For Continuous data:-

للبيانات المتصلة

(1) Histogram: المدرج

مدرجات على شكل مستطيلات متلاصقة رأسية
يستخدم لوصف الكميات المتصلة.

- On x – axis, put class boundaries.
- On y – axis, put frequency.

(2) The frequency polygon: المضلع

- On x – axis, put midpoint of classes.
- On y – axis, put frequency

(3) Ogive (cumulative frequency graph)

- On x – axis, class boundaries..
- On y – axis, put cumulative frequency.

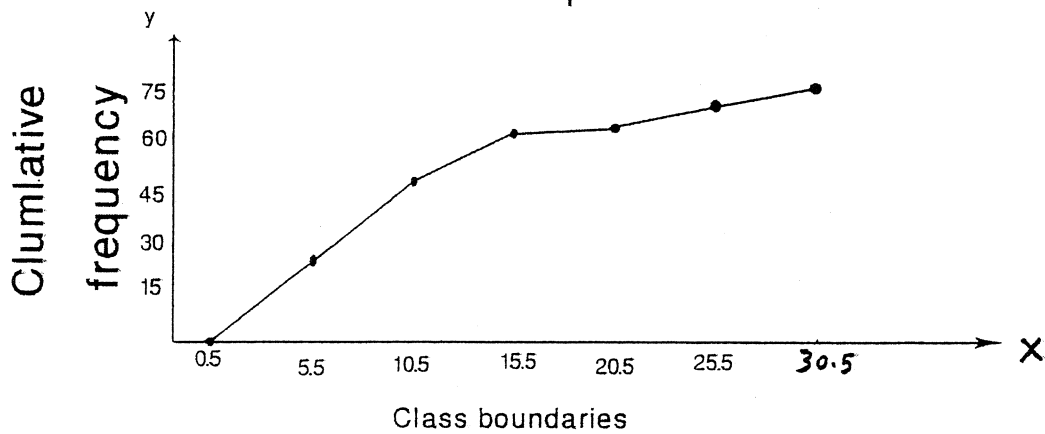
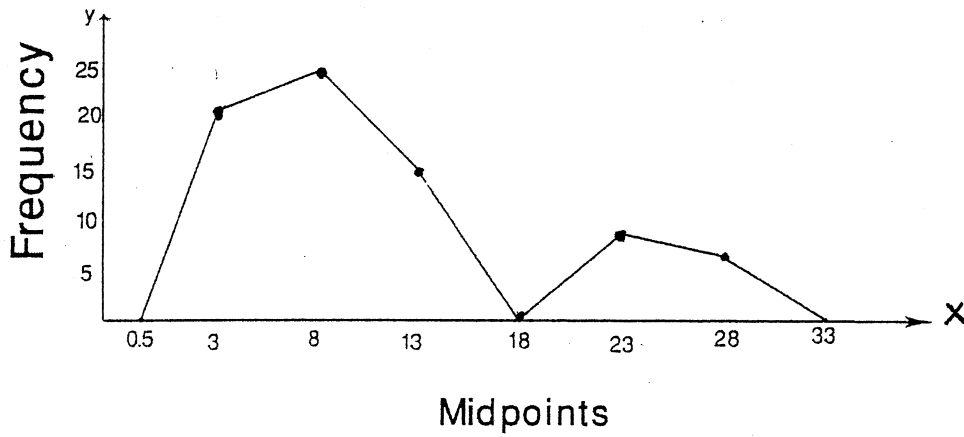
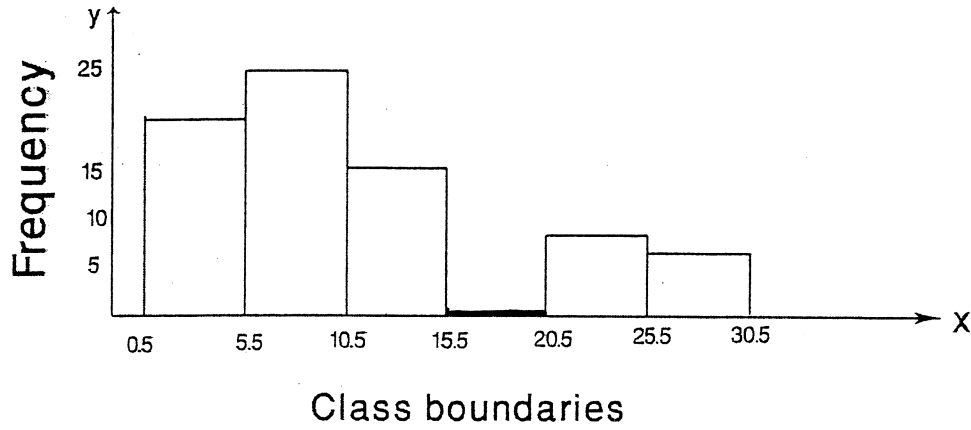
Example:

For 75 employees of a large department store, the following distribution for years of service was obtained construct a histogram. Frequency polygon and O give for the data.

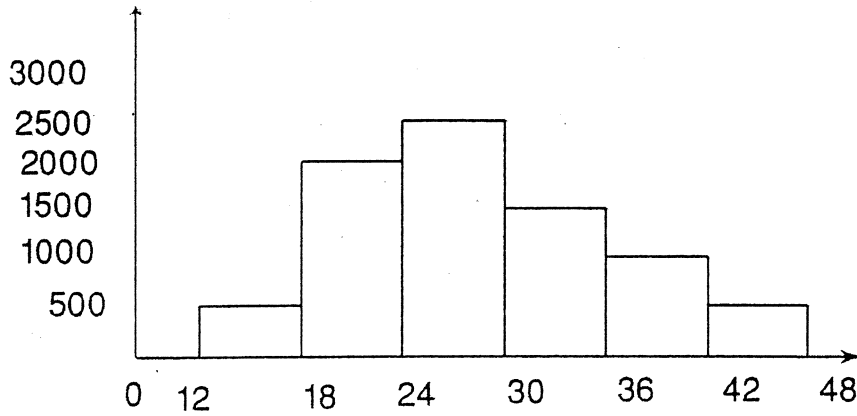
Class limits	Frequency
1-5	21
6-10	25
11-15	15
16-20	0
21-25	8
26-30	6

Solution

Class boundaries	Midpoint	Frequency	Cumulative freq
0.5 – 5.5	3	21	21
5.5 – 10.5	8	25	46
10.5 – 15.5	13	15	61
15.5 – 20.5	18	0	61
20.5 – 25.5	23	8	69
25.5 – 30.5	28	6	75



ALSAADI

Example:

In the above graph:

What type of this graph?

- (a) Ogive (b) Pie chart (c) pereto chart (d) Histogram

2. The class that has highest frequency is

- (a) 24-30- (b) 30-36 (c) 12-18 (d) 18-24

3. The class with frequency 2000 is

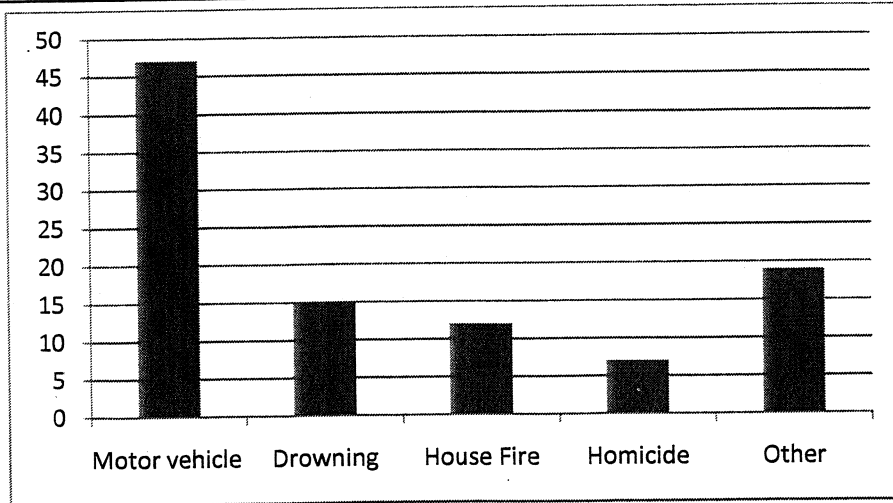
- (a) 24-30 (b) 30 -36 (c) 12- 18 (d) 18 - 24

4. The total frequency of the data shown is

- (a) 6000 (b) 8000 (c) 7500 (d) 1200

- **The bar charts:** display the data by using vertical bars of various heights to represent the frequencies of discrete or categorical variables.
- Example : Represent the data set by using **bar chart**:

Cause of Death (سبب الموت)	Frequency (التكرار)
Motor vehicle (حادث سيارة)	47
Drowning (الغرق)	15
House Fire (حريق منزل)	12
Homicide (القتل)	7
Other (أخرى)	19
Total	100



Notes

* Sum of relative frequencies = 1

مجموع التكرارات
النسبية يساوي 1

* Distributions that are used when the proportions of the data is more important than the actual number of the data are known as:

Relative frequency distribution.

* The graphs that their distributions as proportions instead of raw data as frequencies are called:

Relative frequency graph.

* Histogram is suitable for representing continuous data.

* المدرج التكراري مناسب لتمثيل البيانات المتصلة.

* The graph that represent the cumulative frequencies for the classes in freq. distribution is called a: Ogive

Or: cumulative frequency graph.