

PRINCIPLES OF PROBABILITY & STATISTICS

CHAPTER 1

Descriptive Statistics 1

- ❖ Basic Concepts and definitions
- ❖ Organization and Graphical Representation of the Data
- ❖ Measures of Central Tendency
- ❖ Measures of Dispersion
- ❖ Exercises

INTRODUCTION

Data:

A collection of information collected by means of experiments or real life events and stored in proper format degree

مجموعة المعلومات التي جمعها من التجارب أو الأحداث الحياتية وضعها في تصميم مناسب تسمى (بيانات)

سوف يشمل هذا الكورس نوعان من الإحصاء:

إحصاء وصفي Descriptive stat

Those statistical methods which are used for presenting and summarizing data

وفيه تستخدم الطرق الإحصائية لعرض وتلخيص البيانات في صورة جداول أو أشكال بيانية وتشمل دراسة المتوسطات والمائيات والتشتت

Example

- a) Average height, average weight
- b) Human population data (gender, proportion of population)

إحصاء إستدلالي Inferential stat

Those statistical methods which are used for making conclusion about population

وهو ذلك الفرع من فروع الإحصاء الذي يهتم بعمل استدلالي إحصائي للمجتمع عن طريق عينة مأخوذة من ذلك المجتمع ويشمل

- Point estimate
- interval estimat
- hypothesis testing

Example

لدراسة دخل المواطنين في بلد ما بهذه الطريقة يتم سحب عينة عشوائية ومن خلالها يتم تعميم نتائج لعينة على المجتمع

Population (المجتمع):

It's is the set of all out comes such as individuals, objects, things.

هو مجموعة كل المدخلات من أشخاص وعناصر وأشياء

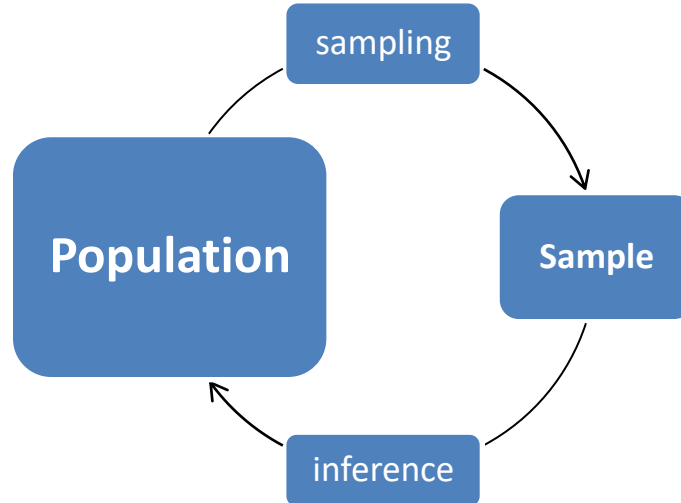
Example

- ☒ The set of all KSU students
- ☒ The set of income of all citizens of a country

Sample (العينة):

Is a subset of population

ويتم سحب العينات عن طريق العينة العشوائية البسيطة أو الطبقية أو العنقودية أو المنتظمة وتكون العلاقة بين العينة والمجتمع كما بالشكل التالي:



Parameter (المعلمة):

It's a numerical characteristics of population

هي خاصية مميزة للمجتمع

Example حساب الوسط الحسابي للمجتمع – حساب الانحراف المعياري للمجتمع

Statistic (الإحصاءة):

It is a function of sample or it is numerical characteristic of the sample

هي خاصية مميزة للعينة

Example حساب الوسط الحسابي للعينة – حساب الانحراف المعياري للعينة

Variable (المتغير):

Is a characteristic, feature or factor that varies from one individual to another.

هو خاصية أو سمة أو عامل يتغير من شخص إلى آخر داخل المجتمع

Example

✚ for countries [gross domestic product, sex ratio, birth rates]

الناتج القومي الكلي، نسبة الجنس – معدل المواليد

✚ for humans [height, weight, sex, marital status,]

الارتفاع والوزن والجنس والحالة الاجتماعية

Types of variable

Qualitative variable: these variable can only take values which are non-numerical (بيانات وصفية (ليست أرقام)

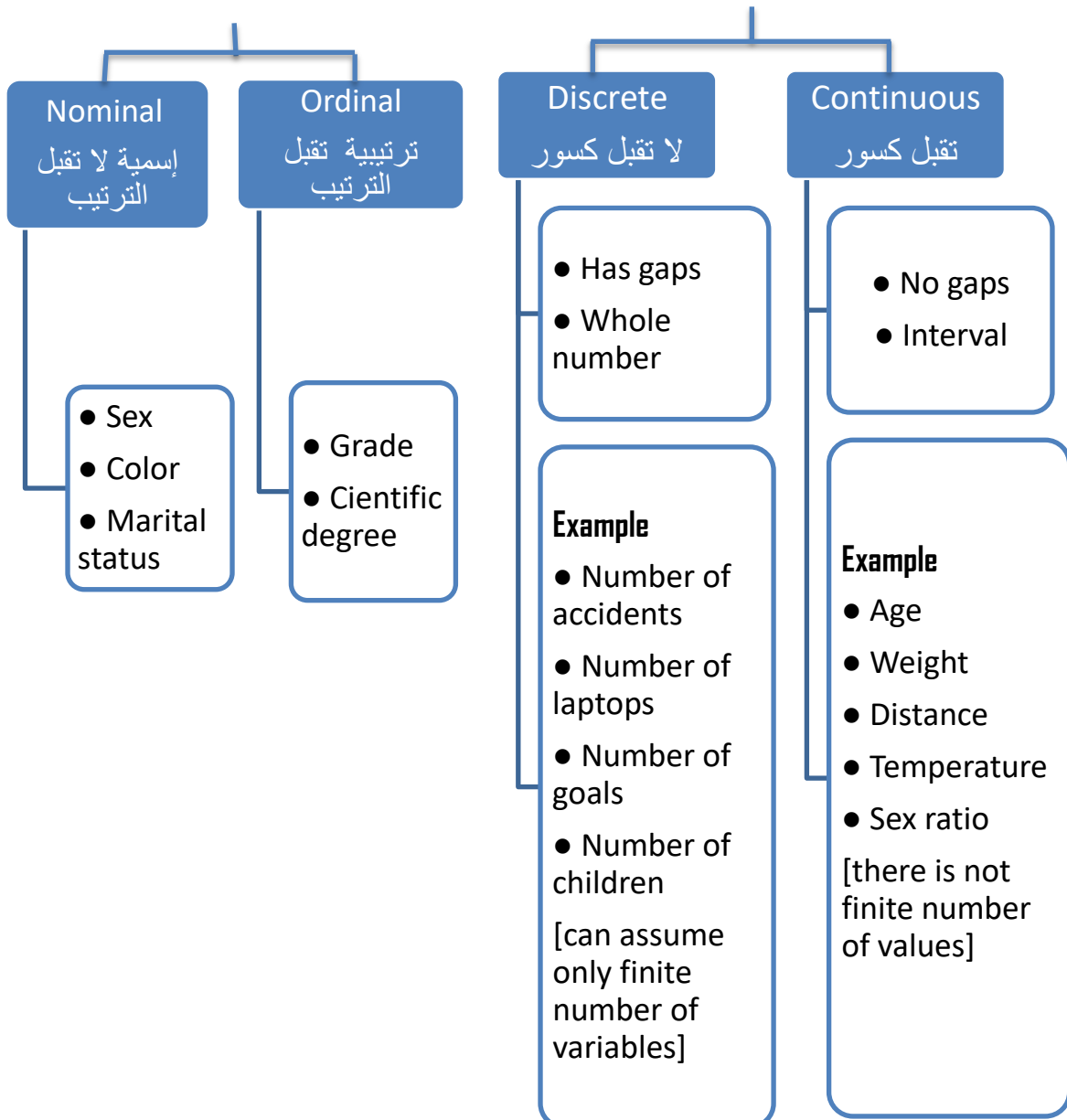
Example

marital status, gender
Can not be ordered
No mathematical operation
Nominal scale مقياس

Quantitative variable: these are variable which take numerical values (بيانات كمية (أرقام)

Example

height, temperature
Can be ordered in increasing or decreasing
Can under go mathematical operations



ORGANIZATION AND GRAPHICAL REPRESENTATION OF THE DATA

في هذا الباب سوف نستعرض إنشاء جداول التوزيع التكراري للبيانات بأنواعها وطريقة عرضها بيانيا

Example consider the blood groups of 40 persons below

O, O, A, B, A, O, A, A, A, O, B, OB, O, O, A, O, O, A, A, A, AB, A, B, A, A, O, O, A, O, O, A, A, A, O, A, O, O, AB

- Construct the frequency distribution for the above data

 **Solution**

Blood group	f	R.f	p.f	C.f	C.R.f
O	16	0.4	40%	16	0.4
A	18	0.45	45%	34	0.85
B	4	0.1	10%	38	0.95
AB	2	0.05	5%	40	1
Total	40	1	100%		

- The sum of all relative frequency equal one
- The sum of all percent frequency equal 100
- القيمة الأخيرة في التكرار التراكمي هي Sample size, n
- القيمة الأخيرة في التكرار النسبي التراكمي دائما واحد
- Relative frequency (R.f) = $\frac{\text{frequency}}{\text{total}}$
- Percent frequency = R. f \times (100)

خليك فإكر

Example Consider the sample of 40 students as follows

0, 1, 2, 3, 1, 2, 2, 1, 1, 2, 0, 2, 1, 0, 1, 0, 1, 1, 2, 1, 2, 1, 3, 1, 2, 1, 1, 0, 0, 2, 1, 1, 0, 1, 2, 2, 2, 1, 0, 1

Construct the frequency table

✓ Solution

No of subjects	f	R.f	p.f	C.f	C.R.f
0	8	0.2	20%	8	0.2
1	18	0.45	45%	26	0.65
2	12	0.3	30%	38	0.95
3	2	0.05	5%	40	1
Total	40	1	100%		

Example The following data is the price of accessories for girls

4, 1, 7, 9, 12, 16, 17, 7, 12, 19, 22, 24, 3, 2, 8, 6, 13, 24, 14, 11, 18, 16, 23, 20, 1, 2, 6, 25, 15, 7, 11, 12, 16, 17, 21, 22, 15, 17, 14, 5, 7, 8, 12, 13, 20, 23, 13, 19, 18, 12

Construct the frequency distribution table for the data by using 5 classes

✓ Solution

البيانات السابقة هي بيانات متصلة "تقبل كسور"

لعمل جدول التوزيع التكراري في حالة البيانات المتصلة لابد من :

① حساب المدى "Range" وهو عبارة عن أكبر قيمة ناقص أصغر قيمة

$$R = \max - \min = 25 - 1 = 24$$

② إيجاد عدد الفئات "K" وغالبا ما يعطى في السؤال وهو يتراوح بين 20-5 فئة ولكن العدد الأمثل يتراوح بين 10-4 فئات

③ نحسب قيمة "C" حيث $C \approx 5$ حيث $C = \frac{R}{K} = \frac{24}{5} = 4.8$ وبالتالي فإن طول الفئة هو

$$\text{Class width} = C - 1 = 5 - 1 = 4$$

④ لحساب حدود الفئات نطرح 0.5 من الحد الأدنى لكل فئة ونضيف 0.5 إلى الحد الأعلى لكل فئة

⑤ لحساب منتصف الفترة $\text{midpoint} = \frac{\text{upper} + \text{lower}}{2}$

⑥ لحساب التكرار المتجمع الصاعد (ACF) ascending cumulative frequency وتكون القيمة الأولى هي نفس قيمة التكرار

القيمة الثانية مجموع التكرار الأول + الثاني

القيمة الثالثة مجموع (الأولى + الثانية + الثالثة) ... وهكذا



إذا كانت البيانات الخام raw data أقل من 32 في هذه الحالة ليست هناك حاجة إلى جدول البيانات في جدول تكراري
إذا كانت عدد الفئات غير معلوم فإنه يمكن إيجادها من العلاقة $K = [3.32 \log n]$

Where n is the simple size

$[x]$ is the greatest integer less than or equal x

Example $[3.2]=3$ & $[3.9]=3$

Class limit	Class Boundaries	Midpoint	F	R.F	LESS Than	ACF
1-5	0.5 → 5.5	3	7	0.14	5.5	7
6-10	5.5 → 10.5	8	9	0.18	10.5	16
11-15	10.5 → 15.5	13	14	0.28	15.5	30
16-20	15.5 → 20.5	18	12	0.24	20.5	42
21-25	20.5 → 25.5	23	8	0.16	25.5	50
Total			50	1		

Example Consider The mileage of 40 cars per liter

12, 16, 15, 15, 12, 19, 17, 18, 16, 14, 13, 12, 20, 12, 15, 16, 20, 16, 15, 12, 18, 16, 17, 19, 15, 16, 17, 15, 16, 15, 14, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20

Construct the frequency distribution table

Solution

$$Rang = max - min = 20 - 12 = 8$$

$$\text{Number of classes } K = [3.32 \log n] = [3.32 \log 40] = [5.3] = 5$$

$$C = \frac{R}{K} = \frac{8}{5} = 1.6 \approx 2$$

Therefore length of class = $C-1=2-1=1$

Class limit	Class Boundaries	Midpoint	F	R.F	ACF	ACRF	ACPF
12-13	11.5 → 13.5	12.5	8	0.20	8	0.20	20%
14-15	13.5 → 15.5	14.5	10	0.25	18	0.45	45%
16-17	15.5 → 17.5	16.5	12	0.30	30	0.75	75%
18-19	17.5 → 19.5	18.5	6	0.15	36	0.90	90%
20-21	19.5 → 21.5	20.5	4	0.10	40	1	100%
Total			40	1			

GRAPHICAL REPRESENTATION

Pie-charts (القطاع الدائري):

Is a simple way of representing the proportion of each class of data on a circular disk.

Example For the frequency distribution table below

Blood group	F	R.F	P.F
O	16	0.40	40%
A	18	0.45	45%
B	4	0.10	10%
AB	2	0.05	5%
Total	40	1	100%

Construct the pie-chart

لرسم القطاع الدائري لابد من معرفة الزاوية المركزية لكل قطاع

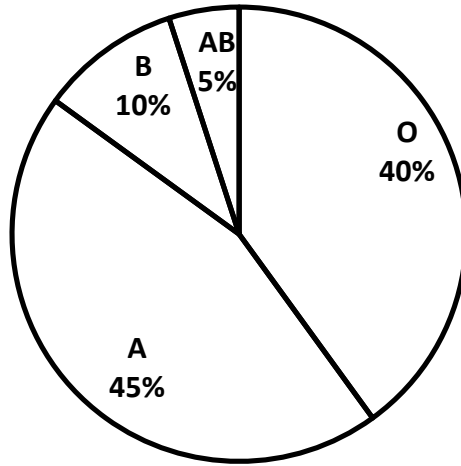
$$\text{measure central angle} = R.F \times 100$$

$$\text{measure central angle of sector O} = 0.40 \times 360 = 144^\circ$$

$$\text{measure central angle of sector A} = 0.45 \times 360 = 162^\circ$$

$$\text{measure central angle of sector B} = 0.10 \times 360 = 36^\circ$$

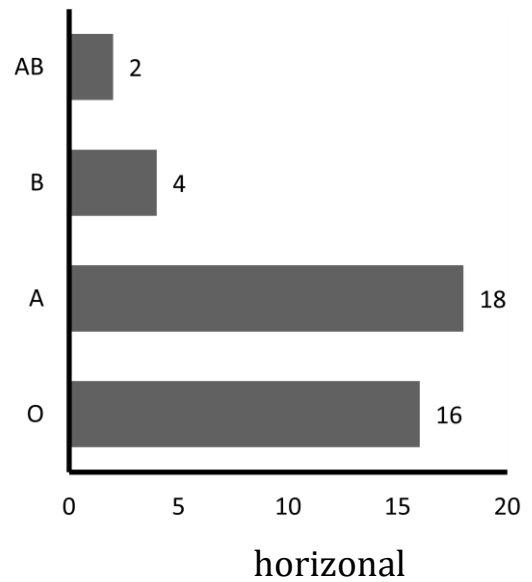
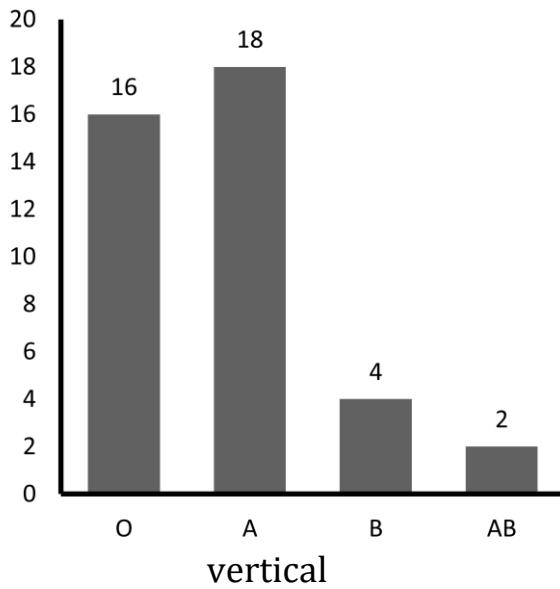
$$\text{measure central angle of sector AB} = 0.05 \times 360 = 18^\circ$$



Bar charts (الأعمدة البسيطة):

The frequency of each class is represented by a bar the height of the bar corresponds to the frequency

وهناك نوعان من الأعمدة البسيطة:

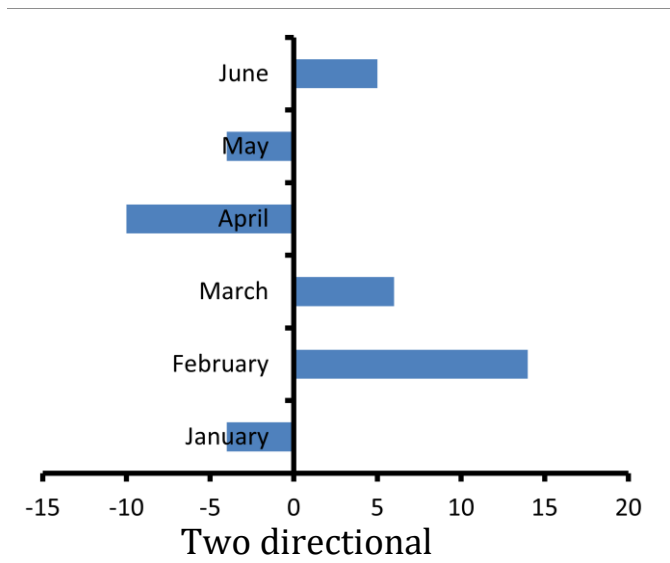


Two directional Bar charts (الأعمدة التكرارية ذات الاتجاهين)

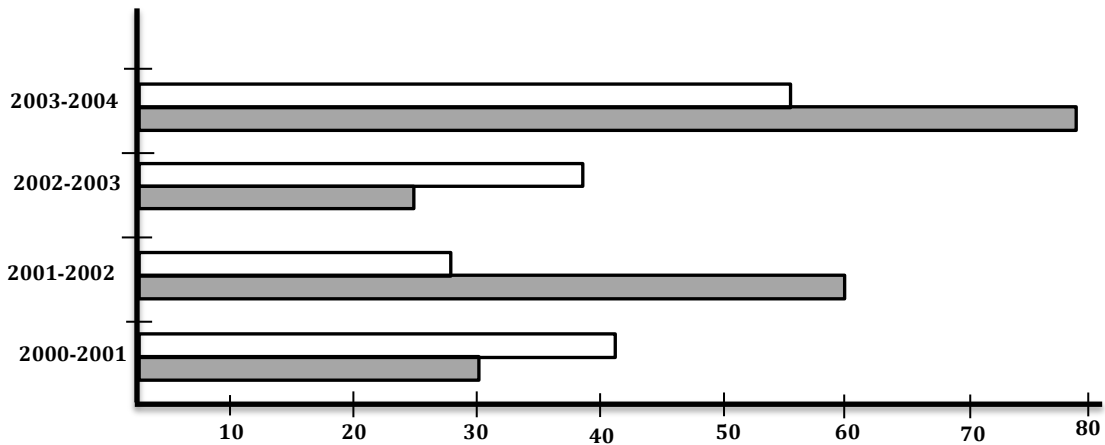
وتستخدم في الحالات التي تعبر عن قيم موجبة وسالبة للفئات المختلفة

Example Consider the following changes in income of a company

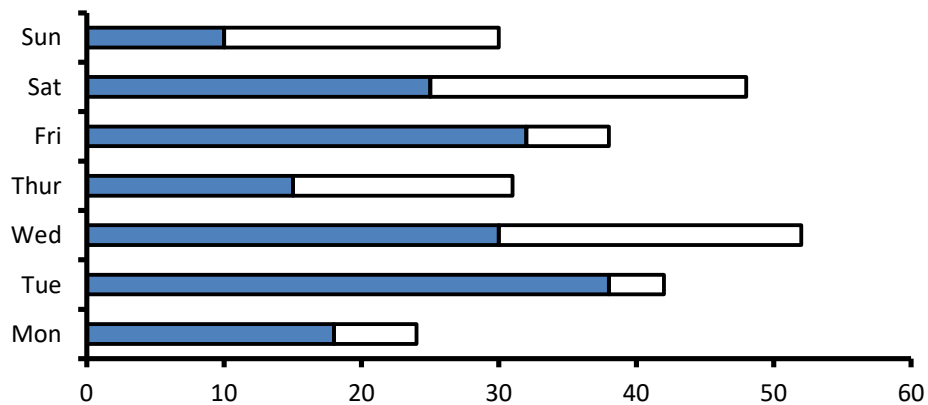
Month	Change in income
January	-4%
February	14%
March	6%
April	-10%
May	-4%
June	5%



Multiple Bar Chart



Stacked Bar Charts

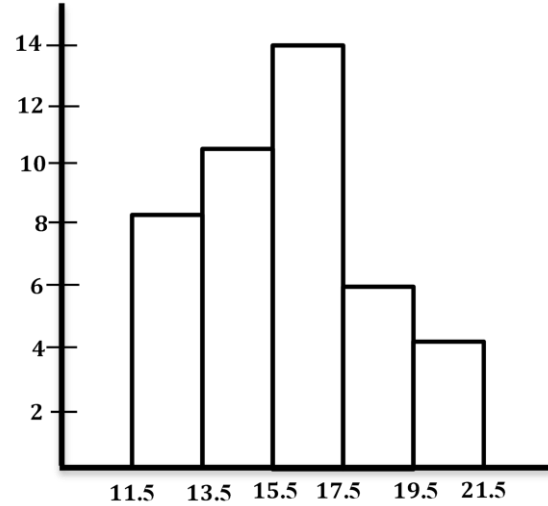


Histogram المدرج التكراري

- It is bar chart but have basic difference
- In histogram classes are adjacent
- Histograms are used to represent the quantitative continuous data
- Every class interval is called bin (خانة)

Example Consider the data

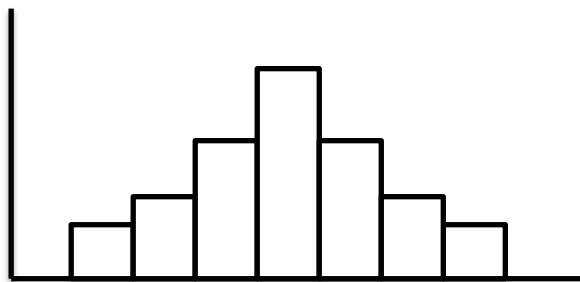
Class boundaries	Frequency
11.5 → 13.5	8
13.5 → 15.5	10
15.5 → 17.5	12
17.5 → 19.5	6
19.5 → 21.5	4
Total	40

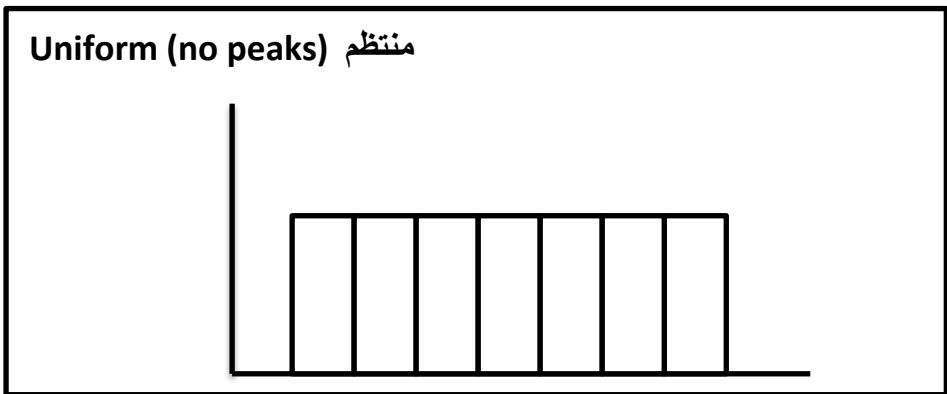
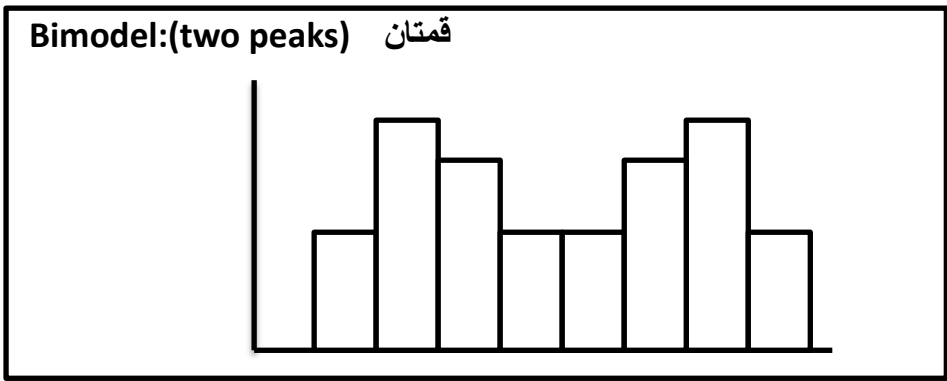


TYPES OF HISTOGRAMS

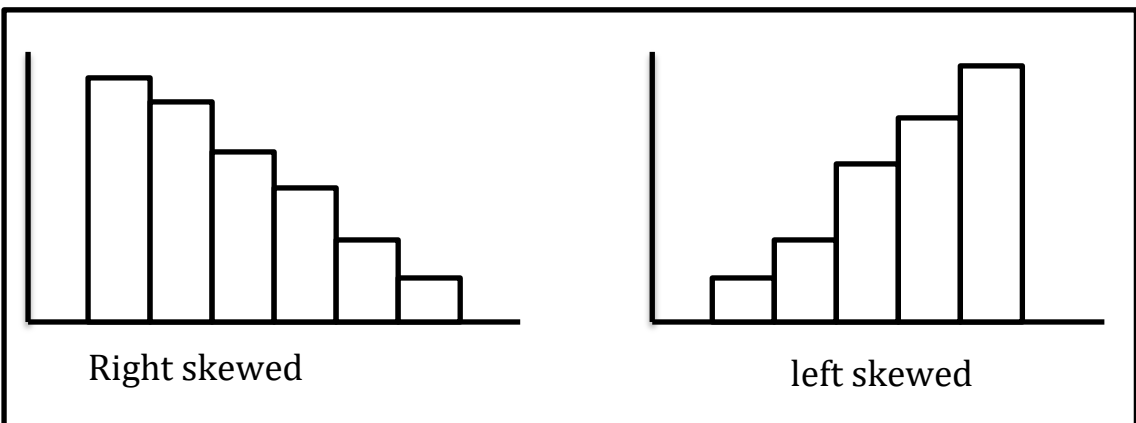
① Symmetric histogram (المدرجات المتماثلة)

Unimodal:(one peak) قمة واحدة





② *Symmetric skewed Histogram* (مدرجات ملتوية متماثلة)

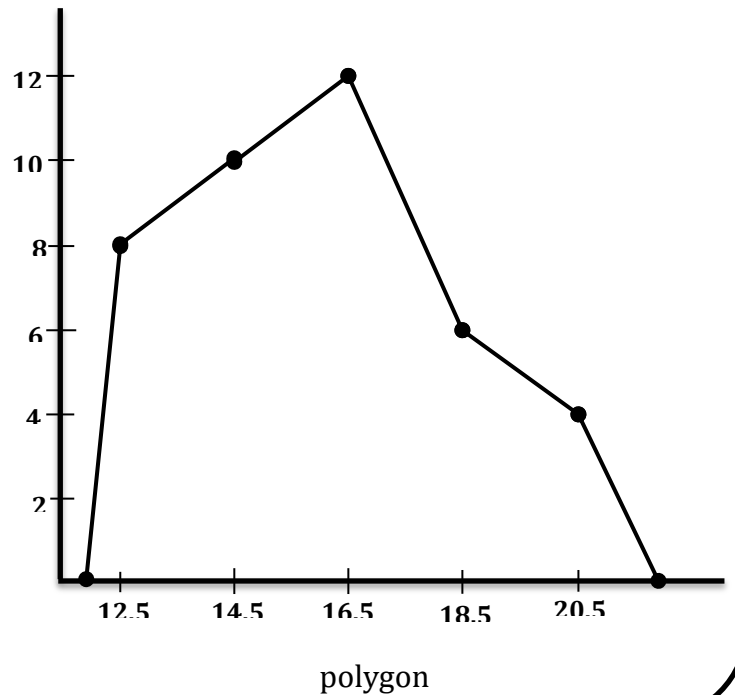


POLYGON (المضلع التكراري)

- ✚ Depend on two points (x_i, f_i) Where x_i is midpoint and f_i is frequency
- ✚ It is useful to compare two or more variables

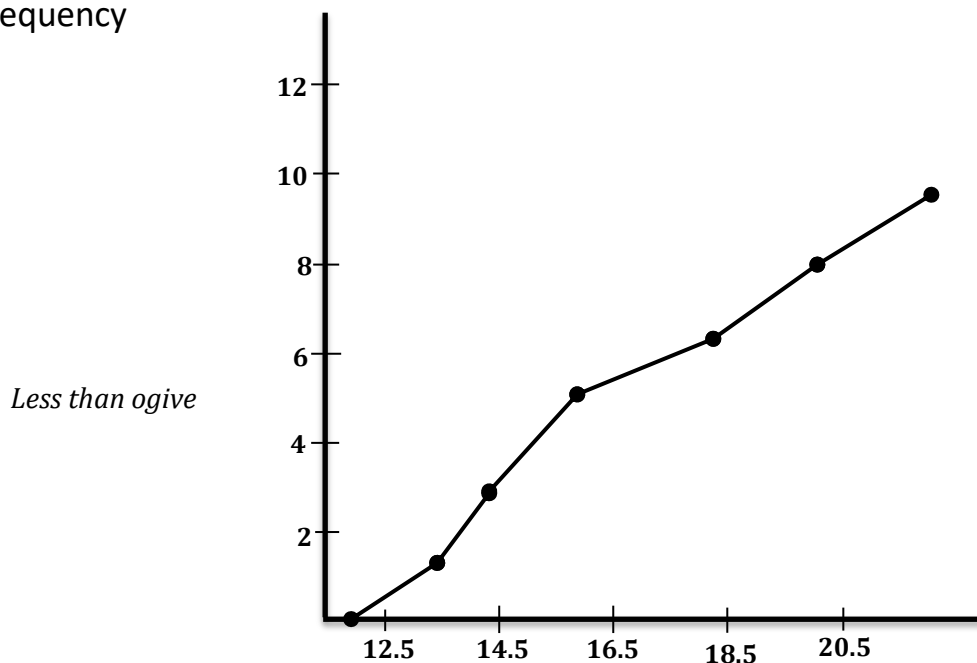
Example Consider the frequency table below

Class Boundaries	Frequency	Midpoint
11.5 → 13.5	8	12.5
13.5 → 15.5	10	14.5
15.5 → 17.5	12	16.5
17.5 → 19.5	6	18.5
19.5 → 21.5	4	20.5
Total	40	



Ogive (المنحنى المتجمع الصاعد)

Depend on two point (b_i, f_i) , where b_i is the upper bound and f_i is cumulative frequency



The measure of central tendency

mean

median

mode

percentiles

deciles

quartiles

1 mean \bar{X} (الوسط الحسابي)

mean is the sum of values divided by the number of values, so $\bar{X} = \frac{\sum x}{n}$

Example

Calculate the mean for the data **20, 18, 15, 15, 14, 12, 11, 9, 7, 6, 4, 1**

Solution

$$\begin{aligned}\bar{X} &= \frac{\sum x}{n} = \frac{20 + 18 + 15 + 15 + 14 + 12 + 11 + 9 + 7 + 6 + 4 + 1}{12} \\ &= \frac{132}{12} = 11\end{aligned}$$

الحاسبة: يمكن حساب الوسط الحسابي للبيانات السابقة بالخطوات التالية

لإدخال البيانات $AC \rightarrow 1 \rightarrow 3 \rightarrow mod$

للنواتج $Shift \rightarrow 2 \rightarrow 4 \rightarrow 1$

البيانات المجدولة يمكن إيجاد الوسط الحسابي لها من القانون

$$\bar{X} = \frac{\sum Fx}{\sum F}$$

Example find the mean for the data below

No of student	0	1	2	3	Total
Frequency	8	18	12	2	40

✓ **Solution**

No of student	F	Fx
0	8	0
1	18	18
2	12	24
3	2	6
Total	$\sum F = 40$	$\sum Fx = 48$

$$\bar{X} = \frac{\sum Fx}{\sum F} = \frac{48}{40} = 1.2$$

إذا كانت البيانات متصلة يمكن إيجاد الوسط الحسابي لها بالقانون التالي :

$$\bar{X} = \frac{\sum FM}{\sum F}$$

حيث M تمثل منتصف الفترة midpoint

Example calculate the mean for the mileage of the cars

Class boundaries	11.5 → 13.5	13.5 → 15.5	15.5 → 17.5	17.5 → 19.5	19.5 → 21.5	Total
Frequency	8	10	12	6	4	40

✓ **Solution**

Class boundary	Midpoint M	Frequency F	F.M
11.5 → 13.5	12.5	8	100
13.5 → 15.5	14.5	10	145
15.5 → 17.5	16.5	12	198
17.5 → 19.5	18.5	6	111
19.5 → 21.5	20.5	4	82
Total		40	636

$$\bar{X} = \frac{\sum FM}{\sum F} = \frac{636}{40} = 15.9$$

الحاسبة: يمكن إيجاد الوسط الحسابي للبيانات المتصلة كالتالي:

shift → **mod** → **4(stat)** → **1(on)**

إدخال القيم في الحاسبة وتكون قيمة x هي الـ midpoint

mod → **3** → **1** (القيم) → **AC**

للنواتج **shift** → **1** → **4** → **2(\bar{x})**

Advantages

- quick and easy
- all values are considered
- unique

Disadvantages

- can not found for qualitative
- highly affected by extreme value
- not applicable if a data is lost

Weighted mean \bar{X}_w (المتوسط المرجح)

$$\bar{X}_w = \frac{\sum Wx}{\sum W}$$

Example a person wants to decide which car better, where look 20%, mileage 30%, engine 50%

Car A: 7 for look, 6 for mileage and 8 for engine

Car B: 6 for look, 4 for mileage and 9 for engine

✓ **Solution**

$$\bar{X}_A = \frac{(0.2 \times 7) + (0.3 \times 6) + (0.5 \times 8)}{0.2 + 0.3 + 0.5} = 7.2$$

$$\bar{X}_B = \frac{(0.2 \times 6) + (0.3 \times 4) + (0.5 \times 9)}{0.2 + 0.3 + 0.5} = 6.9$$

Car A is better

2 Median (الوسيط)

Is that value which divides the data in half after ordering them

هو القيمة التي تنقسم البيانات بعد ترتيبها تصاعديا أو تنازليا

Example calculate the median

First sample: 28, 22, 26, 29, 21, 23, 24

Second sample: 28, 22, 26, 29, 21, 23, 24, 35

 **Solution**

إيجاد الوسيط: 1- ترتيب البيانات

2- إيجاد القيمة التي تقسم البيانات إلى 50% قبلها و50% بعدها

First sample (odd) we ordered data as ~~21~~, ~~22~~, ~~23~~, 24, 26, 28, ~~29~~


median


Second sample (even) ordered data as ~~21~~, ~~22~~, ~~23~~, 24, 26, 28, 29, ~~35~~

$$\text{median} = \frac{24 + 26}{2} = 25$$

median for frequency table \longrightarrow discrete

حساب التكرار التراكمي 

نوجد أصغر تكرار تراكمي أكبر من $\frac{n}{2}$ 

القيمة التي تناظر هذا التكرار تكون هي الوسيط 



Example consider the number of student in the following

No. student	Frequency	ACF
0	8	8
1	18	8+18 = 26
2	12	26+12 = 38
3	2	38+2 = 40
Total	40	

✓ **Solution**

$$n = 40, \text{ so } \frac{n}{2} = \frac{40}{2} = 20$$

Then, the smallest cumulative frequency greater than 20 is 26

The corresponding value to 26 is 1

Hence the median is 1

median for frequency distribution \longrightarrow continuous

أولاً نوجد الفئة الوسيطة وهي الفئة التي تكرارها التراكمي أكبر من $\frac{\sum F}{2}$ من خلال المعادلة

$$\tilde{X} = L + \frac{\frac{\sum f}{2} - (F - f)}{f} \times C$$

حيث

- L الحد الأدنى للفئة الوسيطة
- F التكرار التراكمي للفئة الوسيطة
- f تكرار الفئة الوسيطة
- C طول الفئة

Note that:

The median class is the first class whose cumulative frequency is greater than or equal $\frac{\sum f}{2}$

Example

Class boundary	Midpoint	Frequency	Relative frequency	ACF
11.5 → 13.5	12.5	8	0.2	8
13.5 → 15.5	14.5	10	0.25	18
15.5 → 17.5	16.5	12	0.30	30
17.5 → 19.5	18.5	6	0.15	36
19.5 → 21.5	20.5	4	0.10	40
Total		40	1	

Solution

$$\frac{\Sigma f}{2} = \frac{40}{2} = 20 \Rightarrow \text{median class is } 15.5 \rightarrow 17.5$$

$$L = 15.5 \quad \& \quad F = 30 \quad \& \quad f = 12 \quad \& \quad C = 2$$

$$\tilde{X} = L + \frac{\frac{\Sigma f}{2} - (F - f)}{f} \times C = 15.5 + \frac{20 - (30 - 12)}{12} \times 2 = 15.83$$

3 Mode (المنوال)

The value which has occurred maximum number



∞ يمكن إيجاد المنوال لجميع أنواع البيانات لذلك هو المقياس الأوسع انتشارا

∞ إذا تكررت كل البيانات بنفس المقدار فلا يوجد منوال

∞ يمكن للبيانات أن تحتوى أكثر من منوال

21	22	24	23	29	No mode
----	----	----	----	----	---------

21	28	24	28	28	Mode is 28
----	----	----	----	----	------------

Blood group	O	A	B	AB	Mode is
Frequency	16	8	16	4	O&B

21	22	21	22	23	23	No mode
----	----	----	----	----	----	---------

mode for frequency distribution table \longrightarrow continuous

$$\hat{X} = L + \frac{d_1}{d_1 + d_2} \times C$$

L الحد الأدنى للفئة المنوالية
 f_m تكرار الفئة المنوالية
التكرار اللاحق f_n & التكرار السابق f_p
 $d_1 = f_m - f_p$ & $d_2 = f_m - f_n$

Example

Seconds20	Frequency
50.5 \rightarrow 55.5	2
55.5 \rightarrow 60.5	7 $\rightarrow f_p$
60.5 \rightarrow 65.5	8 $\rightarrow f_m$
65.5 \rightarrow 70.5	4 $\rightarrow f_n$
Total	21

Solution

$$L = 60.5 \quad d_1 = 8 - 7 = 1 \quad d_2 = 8 - 4 = 4 \quad C = 5$$

$$\hat{X} = 60.5 + \left(\frac{1}{1+4} \right) \times 5 = 61.5$$

العلاقة بين الوسط والوسيط والمنوال

في حالة البيانات المتماثلة يمكن استخدام المقاييس الثلاثة وتقع في منتصف البيانات في هذه الحالة.

i.e (for symmetric distribution mean=mode=median)

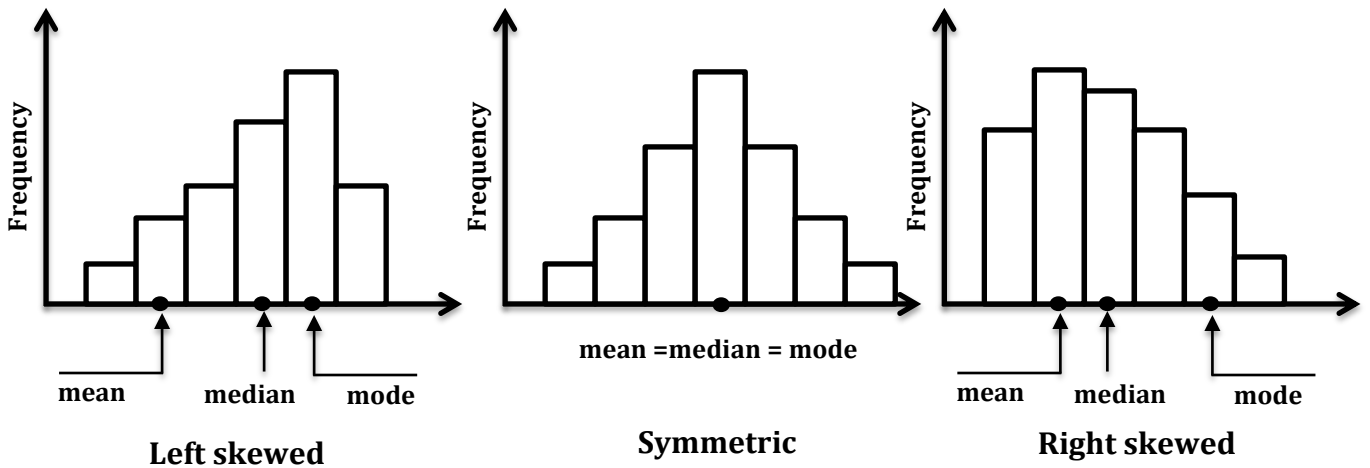
في حالة البيانات الملتوية يسار يكون المنوال اكبر الوسيط اكبر من الوسط لان وجود القيمة الشاذة (extreme value) يزيح الوسط ناحية اليسار وبالتالي يكون الوسط هو القيمة الأصغر في حالة البيانات ملتوية لليسار.

في حالة البيانات الملتوية يمين يكون الوسط هو القيمة الأكبر ثم الوسيط وأخيرا المنوال وهنا وجود القيم المتطرفة يزيح الوسط ناحية اليمين وبالتالي يكون الوسط هو القيمة الأكبر هنا

خليك فإكر

- الوسط الحسابي هو أدق مقاييس النزعة المركزية
- الوسط الحسابي يتأثر بكل العمليات التي تجرى عليه " جمع - طرح - ضرب - قسمة "
- الوسط الحسابي يمكن أن يكون قيمة سالبة
- أكثر المقاييس تأثرا بالقيم الشاذة هو الوسط الحسابي

The relationship among the three measures of central tendency



4 Percentiles (المنينيات)

The percentiles of a variable divide the value into 100 equal parts and we denote by P_1, P_2 and P_{99}

خليك فاكّر

- P_{50} is the median which divides the data into two equal part
- P_1 divides the data into 1% from bottom and 99% from top
- $p_r = \frac{r(n+1)}{100}$ where $r=1, 2, \dots, 99$ n is the sample size
- $P_r = X_K + S(X_{K+1} - X_K)$ K is the integer , S the reset of p_r

Example Calculate the 35th percentile of the data

40, 51, 92, 10, 36, 60, 70, 36, 36, 40, 80, 39, 53, 56, 60, 60, 70, 72, 88, 92, 50, 92, 20, 70, 38, 95, 56, 60, 88, 70

 **Solution**

نقوم بترتيب البيانات تصاعديا

10, 20, 36, 36, 36, 38, 39, 40, 40, 50, 51, 53, 56, 56, 60, 60, 60, 60, 70, 70, 70, 70, 72, 80, 88, 88, 92, 92, 92, 95

$$p_r = \frac{r(n+1)}{100} \quad \text{where} \quad r = 35 \quad n = 30$$

$$P_{35} = \frac{35(30+1)}{100} = 10.85$$

أي أن القيمة التي تقسم البيانات إلى 35% قبلها و65% بعدها تقع بين القيمة العاشرة والحادية عشر ويمكن إيجاد قيمتها

$$P_r = X_K + S(X_{K+1} - X_K) \quad \text{where} \quad K = 10 \quad S = 0.85$$

$$\begin{aligned} P_{35} &= X_{10} + 0.85(X_{11} - X_{10}) \\ &= 50 + 0.85(51 - 50) = 50.85 \end{aligned}$$

5 Deciles (العشيرات)

The deciles divide the data into 10 equal parts

خديك فإكر

- D_1 is the 10th percentile, D_2 is the 20th percentiles
- $d_r = \frac{r(n+1)}{10}$ where $r=1,2,\dots,9$
- $D_r = X_K + S(X_{K+1} - X_K)$ where K is the integer S is the rest

Example For the above example calculate the sixth decile

 **Solution**

$$d_r = \frac{r(n+1)}{10} \quad r = 6, \quad n = 30$$

$$d_6 = \frac{6(30+1)}{10} = 18.6$$

أن موقع القيمة التي تقسم البيانات إلى 0.6 قبلها و0.4 بعدها تقع بين القيمة الثامنة عشر والتاسعة عشر في البيانات ويمكن إيجاد قيمتها تحديدا من خلال القانون

$$D_r = X_K + S(X_{K+1} - X_K)$$

$$D_6 = 60 + 0.6(70 - 60) = 66$$

6 Quartiles (الربيعات)

The quartiles divide the data into 4 equal parts

➤ Q_1 is 25th percentile, Q_2 is the median and Q_3 is the 75th percentile

➤ $q_r = \frac{r(n+1)}{4}$ where $r = 1, 2, 3$

➤ $Q_r = X_K + S(X_{K+1} - X_K)$

Example For the above example find the first quartile

✓ **Solution**

$$q_r = \frac{r(n+1)}{4} \quad \text{where } r = 1, n = 30$$

$$q_1 = \frac{1 \times (30+1)}{4} = 7.75$$

$$K = 7 \quad S = 0.75$$

$$\begin{aligned} Q_1 &= X_K + S(X_{K+1} - X_K) \\ &= 39 + 0.75 \times (40 - 39) = 39.75 \end{aligned}$$

Example Find the quartiles of the data 28, 22, 26, 29, 21, 23, 24

✓ **Solution**

$$r = 1 \quad n = 7$$

$$q_1 = \frac{1 \times (7+1)}{4} = 2 \Rightarrow Q_1 = 22 + 0 \times (23 - 22) = 22$$

$$q_2 = \frac{3 \times (7+1)}{4} = 4 \Rightarrow Q_2 = 24 \quad \text{which is median}$$

$$q_3 = \frac{3 \times (7+1)}{4} = 6 \Rightarrow Q_3 = 28$$

Example Find the quartiles of the data 7, 8, 15, 36, 39, 40, 41, 56

 **Solution**

$$q_1 = 2.25 \Rightarrow Q_1 = 8 + 0.25 \times (15 - 8) = 9.75$$

$$q_2 = 4.5 \Rightarrow Q_2 = 36 + 0.5 \times (39 - 36) = 37.5$$

$$q_3 = 6.75 \Rightarrow Q_3 = 40 + 0.75 \times (41 - 40) = 40.75$$

Extreme values (القيم المتطرفة)

The value of x is said to be extreme if $\left\{ \begin{array}{l} X < Q_1 - 1.5(Q_3 - Q_1) \\ \text{or} \\ X > Q_3 + 1.5(Q_3 - Q_1) \end{array} \right.$

For the preceding example we have

$$\blacktriangleright Q_1 - 1.5(Q_3 - Q_1) = 9.75 - 1.5(40.75 - 9.75) = -36.75$$

$$\blacktriangleright Q_3 + 1.5(Q_3 - Q_1) = 40.75 + 1.5(40.75 - 9.75) = 87.25$$

وبالتالي فإن أي قيمة أقل من -36.75 أو أكبر من 87.25 تعتبر قيمة شاذة ولذلك لا توجد قيم شاذة في المثال أعلاه

خليك فإكر

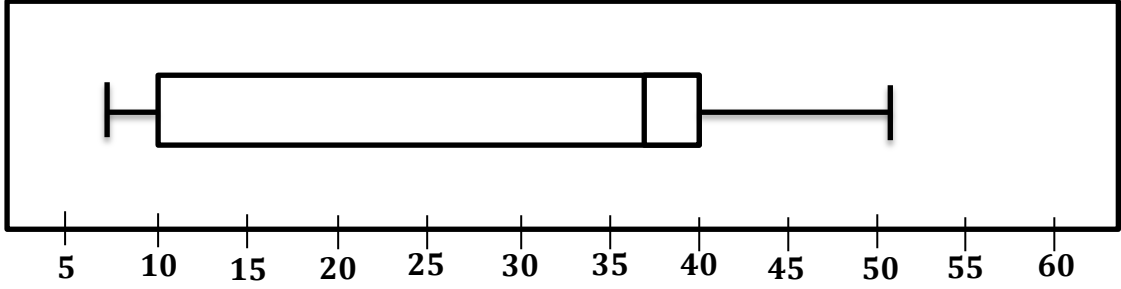
Five Numbers are:

Smallest value , Q_1 , Q_2 , Q_3 , largest value for the above example, the five numbers are 7, 9.75, 37.5, 40.75, 56



Example Draw the box plot for the data

7, 9.75, 37.5, 40.75 and 56

 **Solution**



خليك فاكّر

إذا لم تحتوي البيانات قيمة متطرفة كبيرة نرسم الذراع على أكبر عدد بالبيانات 
إذا وجدت قيمة متطرفة نرسم الذراع على الحد الأعلى ناحية اليمين للسور وكذلك ناحية اليسار. 

Measure of dispersion

variance

standard deviation

Range

Coefficient of variation

1 Variance (التباين)

وهو يعبر عن مدى تباعد أو تقارب القيمة الكبرى والصغرى عن مركز البيانات \bar{X} ويرمز له بالرمز S^2

$$S^2 = \frac{\sum(X - \bar{X})^2}{n - 1}$$

or

$$S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1}$$

خليك فاكّر

➤ The variance always more than or equal zero

❖ أي أن التباين وكذلك قيم التشتت لا يمكن أن تكون قيم سالبة

➤ The variance unit is square unit

❖ وحدة التباين دائما وحدة مربعة ومعامل الاختلاف ليس له وحدة أما باقي المقاييس لها نفس الوحدة محل الدراسة

❖ التباين والانحراف المعياري لا يتأثر بإضافة أو طرح أي قيمة على البيانات يتأثر فقط بالضرب أو القسمة

❖ الوسط الحسابي يتأثر بجميع العمليات (ضرب ، قسمة ، جمع ، طرح)

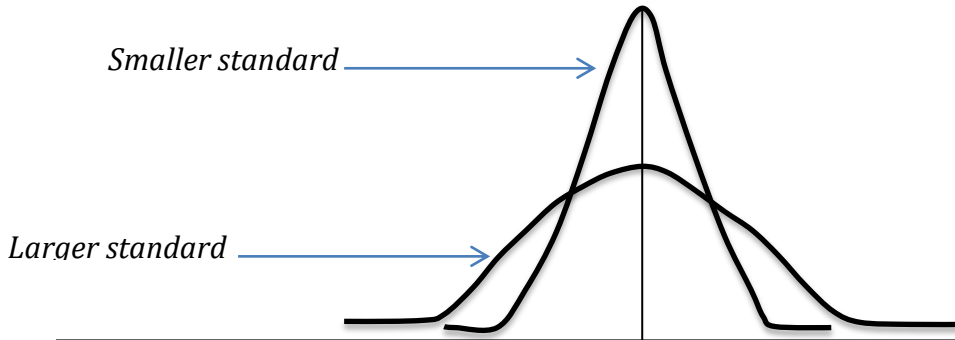
2 Standard deviation (الانحراف المعياري)

The standard deviation is the square root of the variance and denoted by (S)

خليك فاكّر

يعتبر مقياس الانحراف المعياري هو أفضل مقاييس التشتت في حالة عدم وجود قيمة شاذة وهو أفضل من التباين لأنه له نفس وحدة محل الدراسة
الوسط الحسابي هو أفضل مقاييس النزعة المركزية "شرط عدم وجود قيمة شاذة"

➤ A small value of standard deviation indicates that the value of the variable tend to be close to the center where a large value indicates that they tend to be far from the mean (center)



الحاسبة

يمكن إيجاد الانحراف المعياري والتباين كما يلي
 $mod \rightarrow 3 \rightarrow 1 \rightarrow (القيمة إدخال) \rightarrow AC$
المعياري الانحراف $shift \rightarrow 1 \rightarrow 4 \rightarrow 4(S)$
وبتربيع القيمة الأخير نحصل على التباين

Example let 2, 3, 6, 8, 10, 13 and 14 calculate the variance and standard deviation

 **Solution**

في هذا المثال سوف نقوم بثلاث طرق:

الأولى

$$S = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} = \sqrt{\frac{130}{6}} = 4.65$$

$$S^2 = 21.7$$

X	X^2	\bar{X}	$(X - \bar{X})^2$
2	4	8	36
3	9	8	25
6	36	8	4
8	64	8	0
10	100	8	4
13	169	8	25
14	196	8	36
$\sum X = 56$	$\sum X^2 = 578$		$\sum(X - \bar{X})^2 = 130$

الثانية

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} = \sqrt{\frac{578 - \frac{56^2}{7}}{6}} = 4.65$$

$$S^2 = 21.7$$

وبتربيع القيمة الأخيرة نحصل

الثالثة

$$\text{mod} \rightarrow 3 \rightarrow 1$$

$$2 = 3 = 6 = 8 = 10 = 13 = 14 = \boxed{\text{AC}}$$

$$\text{shift} \rightarrow 1 \rightarrow 4 \rightarrow 4(S) = 4.65$$

Example find the variance and the standard deviation for the data

6, 6, 6, 6, 6, 6, 6

 **Solution**

باستخدام الحاسبة

$mode \rightarrow 3 \rightarrow 1$

$6 = 6 = 6 = 6 = 6 = 6 = 6 =$ AC

$shift \rightarrow 1 \rightarrow 4 \rightarrow 4(S) = Zero$

وهذه هي الحالة الوحيدة التي يمكن للانحراف المعياري أن يساوي صفر عندما تكون جميع القيم متساوية

3 Range (المدى)

- | | |
|---|-------------------------------------|
| وهو الفرق بين أكبر قيمة وأصغر قيمة في البيانات | <input checked="" type="checkbox"/> |
| أسرع مقياس يعطي فكرة عن مدى تفاوت البيانات | <input checked="" type="checkbox"/> |
| لا يفضل استخدامه في حالة وجود قيم شاذة | <input checked="" type="checkbox"/> |
| قيمته موجبة دائما حتى لو كانت كل البيانات سالبة | <input checked="" type="checkbox"/> |
| له نفس وحدة البيانات محل الدراسة | <input checked="" type="checkbox"/> |

$$R = \max - \min$$

Example find the Range for

X: 4, 8, 7, 3, 5, 10, 24, 5

Y: 10, 7, 9, 11, 11, 8, 9, 7

✓ Solution

For X: Range = max - min = 24 - 3 = 21

For Y: Range = 11 - 7 = 4

المدى لبيانات Y أصغر من المدى لبيانات X وبالتالي فإنه يمكننا القول أن البيانات في Y أكثر تجانسا أو البيانات في X أكثر تشتتا

For frequency distribution

$R = X_K - X_1$ where

X_K is the midpoint for last class

X_1 is the midpoint for first class

Interquartile Range (المدى الربيعي)

Also called as mid-spread IQR = $Q_3 - Q_1$

4 Coefficient variation (معامل الاختلاف)

$$C.V = \frac{s}{\bar{x}} \times 100$$

Has no unite

دائما يستخدم لمقارنة تشتت مجموعتين من البيانات ليس لها نفس الوحدة

Z-score (الدرجة المعيارية)

$$Z = \frac{X - \bar{X}}{s} \quad \text{where mean} = 0 \quad \text{and} \quad \text{standard deviation} = 1$$

قاعدة تشابيشيف The CHEBYSHEV'S Rule

At least $1 - \frac{1}{K^2}$ of the data lies within K standard deviation of the mean

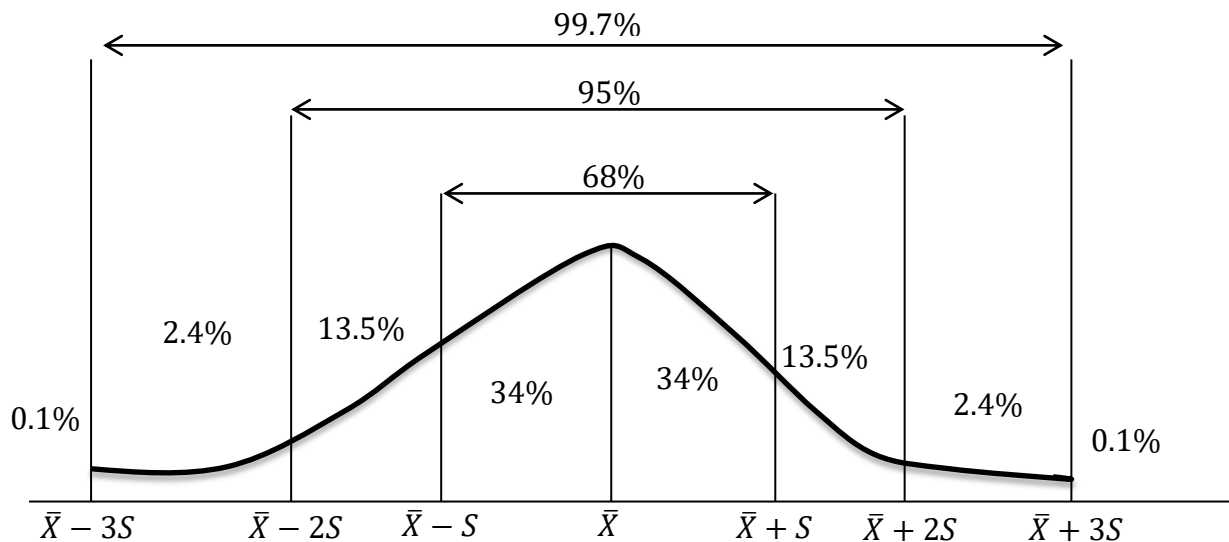
i.e in the interval $\bar{X} \pm KS$, where K whole number > 1

◆ At least 75% of data lie within two standard deviation, in the interval with end points $\bar{X} \pm 2S$

◆ At least 88.89% of the data lie within three standard deviation, in the interval with end points $\bar{X} \pm 3S$

The empirical rule (القاعدة التجريبية)

- 1- 68% of data lies within one standard deviation, in $\bar{X} \pm S$ interval
- 2- 95% of data lies within one standard deviation, in $\bar{X} \pm 2S$ interval
- 3- 99.7% of data lies within one standard deviation, in $\bar{X} \pm 3S$ interval



Example score of some tests have a bell-shaped distribution with mean $\bar{X} = 100$ and standard deviation $S=10$ discuss what the empirical rule concerning individuals with scores of 110, 120 and 130

Solution

- ① 68% $\rightarrow \bar{X} \pm S \Rightarrow$ scores lie between $(100-10, 100+10)$
 $\Rightarrow (90,110)$

② $95\% \rightarrow \bar{X} \pm 2S \Rightarrow$ scores lie between $(100-20, 100+20)$

$\Rightarrow (80,120)$

③ $99.7\% \rightarrow \bar{X} \pm 3S \Rightarrow$ scores lie between $(100-30, 100+30)$

$\Rightarrow (70,130)$

Example the mean price of apartments in certain city 50000 with standard deviation 10000 , we will determine the price range for which at least 75% of the houses will sell

 **Solution**

By Chebychev's Rule \Rightarrow at least 75% of the data fall within 2 standard deviation hence $(\bar{X} - 2S, \bar{X} + 2S) = (50000 - 2 \times 10000, 50000 + 2 \times 10000)$

$= (30000, 70000)$

i.e at least 75% of all apartments sold will have a price range from 30000 to 70000

EXERCISES

QUESTION1: Classify each variable as quantitative or qualitative

1. Blood group
2. Time to finish exam
3. Height of student
4. Colors of flowers

- Qualitative (1)
- Quantitative (2)
- Quantitative (3)
- Qualitative (4)

QUESTION2: Classify each variable as discrete or continuous

1. Weight of children
2. Number of students
3. Age of cats
4. Number of cars

- Continuous (1)
- Discrete (2)
- Continuous (3)
- Discrete (4)

QUESTION3: Answer with true or false to the following sentence

- ⊞ *Mode is defined for qualitative data*
- ⊞ *The mean is sensitive to extreme value*
- ⊞ *For a skewed distribution we have mode = mean = median*
- ⊞ *Histogram with two peaks is multimodal*

- *True*
- *True*
- *False*
- *False*

QUESTION4: Classify each variable as qualitative or quantitative

- ⊞ The variable that record ID of students in an exam
- ⊞ The variable that record weight of children in a school
- ⊞ The variable that record phone number
- ⊞ The variable that record sugar level

- Qualitative
- Quantitative
- Qualitative
- Qualitative

- ☒ The variable that record blood pressure
- ☒ The variable that record colors of cars
- ☒ The variable that record number of flights
- ☒ The variable that record time to finish exam

- Quantitative
- Qualitative
- Quantitative
- Quantitative

QUESTIONS: Classify each variable as Continuous or Discrete

- ☒ The variable of heights of people
- ☒ The age of baby
- ☒ The number of children in school
- ☒ The weight of books
- ☒ The distance between cities
- ☒ Number of accidents
- ☒ The number of goals
- ☒ Temperature in a city
- ☒ Sex ratio in a country

- Continuous
- Continuous
- Discrete
- Continuous
- Continuous
- Discrete
- Discrete
- Continuous
- Continuous

QUESTION6: Put right word or symbol in its proper position:

statistic , variable , bar chart , Descriptive statistic.

is those statistical methods or techniques which are used for pie charting and summarizing

is a function of a sample

is characteristic, feature or factor that varies from one individual to another in a population.

the frequency of each class is represented by a bar

The height of the bar doesn't matter

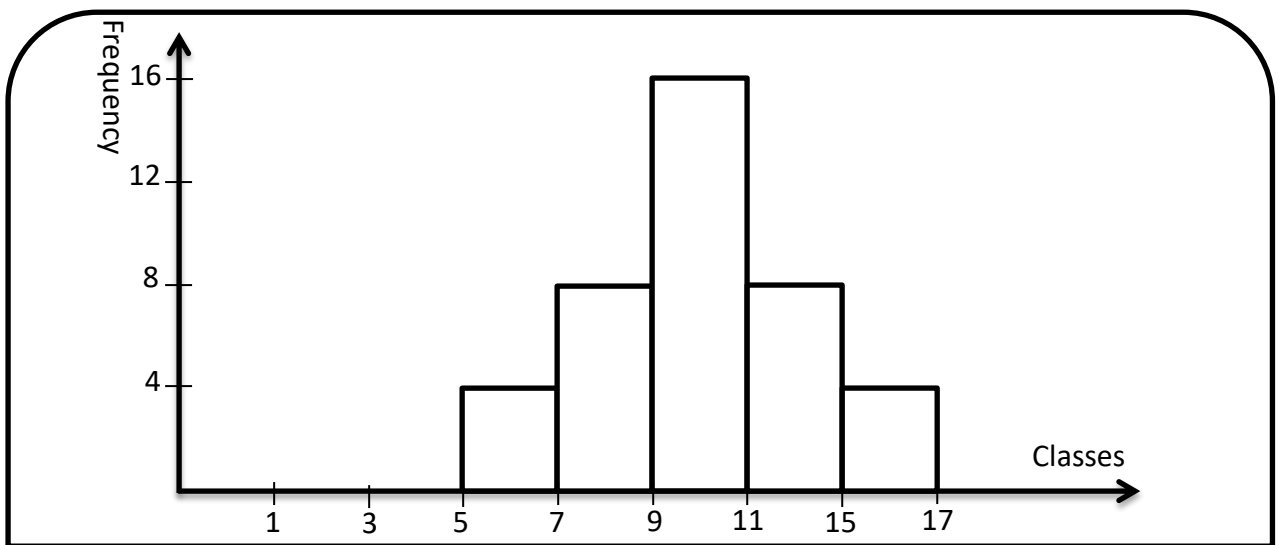
QUESTION7: Consider the following data.

5.5	7.5	6	6.5	7	6.75	7	7.25	7.5	7.75	8
8.25	9	9.75	10	10.25	10.5	10.75	9	9.25	9.5	9.75
10.25	10.5	11	12	12.5	11.5	11.25	11.75	12.75	12.99	13.01
13.25	13.5	14.89	9.01	9.55	10.10	10.89				

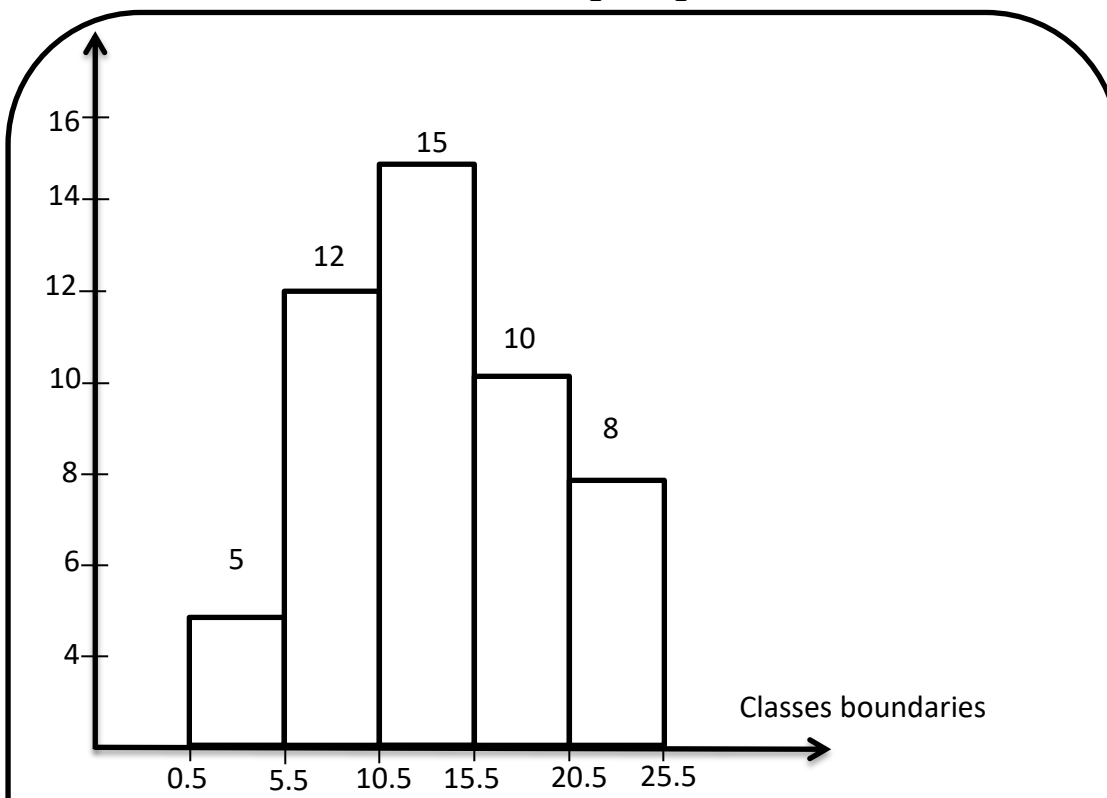
Construct the frequency distribution table

Class boundaries	Midpoint	Frequency	R.F	P.F	A.C.F
5-7		4			
7-9		8			
9-11		16			
11-13		8			
13-15		4			
Total		40			

Draw the histogram for the data above



QUESTION8: If we have data with the following histogram.



A) Complete the following frequency distribution

 **Solution**

Class limit	Class boundaries	Midpoint	Frequency	R.F	P.F	A.C.F
			5			
			12			
			15			
			10			
			8			
Total			50			

B) Calculate the median for the given data

$$\frac{n}{2} = \frac{50}{2} = 25 \quad \bar{X} = L + \frac{\frac{\sum f}{2} - (F-f)}{f} \times c$$

$$\text{Median class } 10.5 - 15.5 \Rightarrow L = 10.5$$

$$F = 32 \quad f = 15 \quad c = 5$$

$$\bar{X} = 10.5 + \frac{25 - (32 - 15)}{15} \times 5 \approx 13.17$$

C) calculate the range for the given data

 **Solution**

$$R = X_K - X_1 \quad , \text{ where } X_K \text{ is midpoint for last , } X_1 \text{ midpoint for first}$$

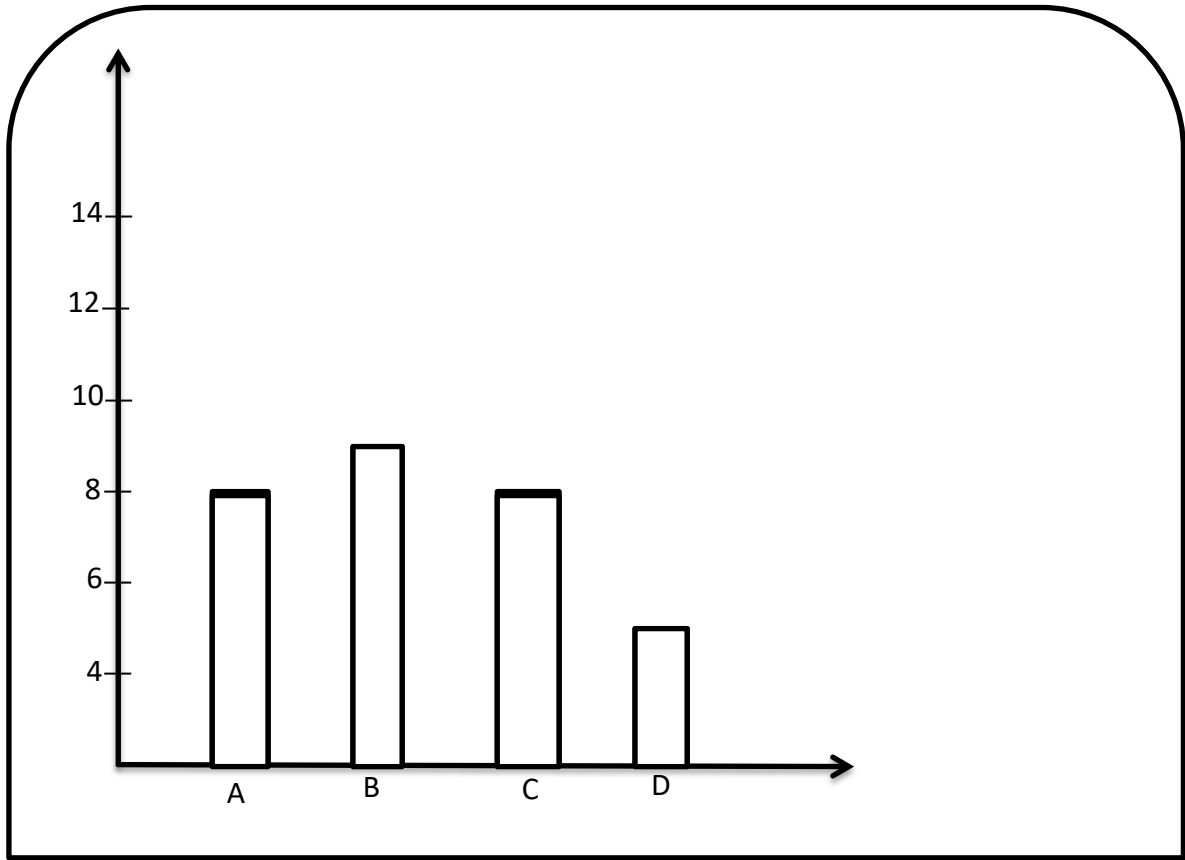
$$R = 23 - 3 = 20$$

QUESTIONS: consider the following data

B	A	A	B	C	A	B	A	C	A	B
C	D	C	B	C	D	A	A	D	C	B
A	C	B	D	C	D	B	B			

Draw the bar graph for the given data

 **Solution**



QUESTION: Using the data shown in the following table

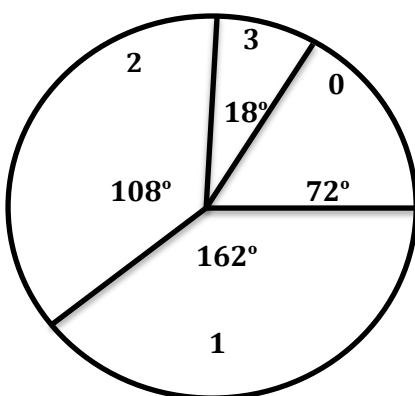
No. of subjects	Frequency
0	8
1	18
2	12
3	2
Total	40

- A. Represent them graphically using pie chart and bar chart
- B. Compute the range of this data
- C. Compute the mean of this data

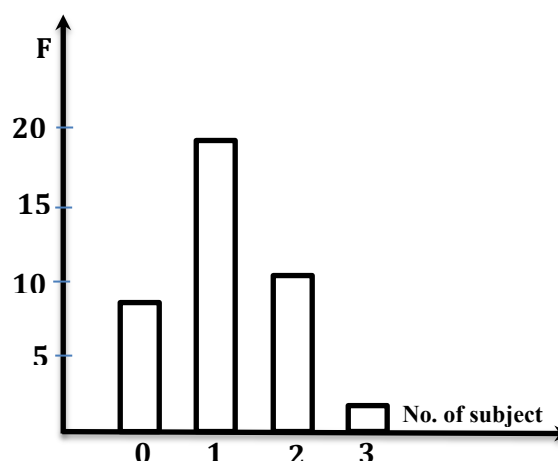
 **Solution**

A.

No. of subjects X	Frequency F	R.F	FX	Degree
0	8	0.2	0	$0.2 \times 360 = 72$
1	18	0.45	18	$0.45 \times 360 = 162$
2	12	0.3	24	$0.3 \times 360 = 108$
3	2	0.05	6	$0.05 \times 360 = 18$
Total	40		48	



Pie-chart



B. The Range = max - min = 3 - 0 = 3

C. The mean = $\bar{X} = \frac{\sum FX}{\sum F} = \frac{48}{40} = 1.2$

QUESTION: the following data give the result of a sample survey A, B and C represent the three categories

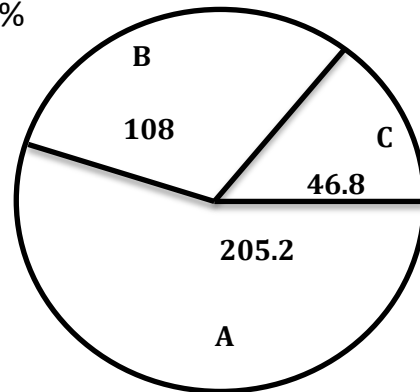
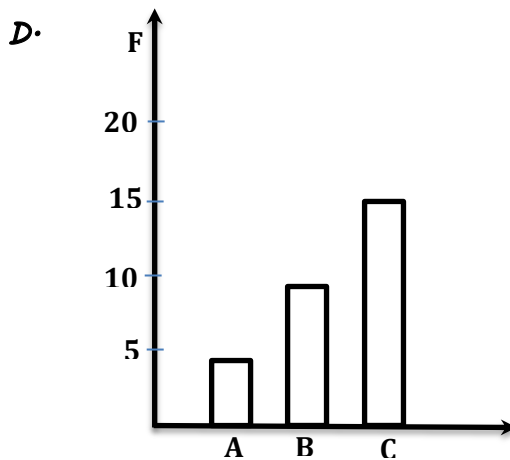
C	A	C	B	A	C	B	C	C	C	B	C	B	C	B
C	B	C	C	B	C	C	C	A	B	C	C	B	C	A

- A. Prepare the frequency table
- B. Calculate the relative frequency and percentage for all symbols
- C. What percentage of elements belongs to category B?
- D. Draw a bar chart and pie chart for frequency table

Solution

Letters	Frequency	Relative Frequency	Percentage Frequency
A	4	0.13	13%
B	9	0.3	30%
C	17	0.57	57%
Total	30	1	100%

C. The percentage of B Category is 30%



$$A = 0.13 \times 360 = 46.8$$

$$B = 0.3 \times 360 = 108$$

$$C = 0.57 \times 360 = 205.2$$

QUESTION: A Company manufactures car batteries of a particular type the lives in years of 40 such batteries were recorded as follows

2.6	3	3.7	3.2	2.2	4.1	3.5	4.5	4.6	3.8	3.5	2.3	3.2
3.4	3.8	3.2	4.6	3.7	2.9	3.6	2.5	4.4	3.4	3.3	2.9	3
4.3	2.8	3.5	4.2	3.5	3.2	3.9	3.2	3.2	3.1	3.7	3.4	3.2
2.6												

Construct Frequency distribution using class of size 0.5, starting 2-2.5

 **Solution**

Class interval	Frequency	Relative frequency	Percentage	ACF
2 → 2.5	2	0.05	5%	2
2.5 → 3.0	6	0.15	15%	8
3.0 → 3.5	14	0.35	35%	22
3.5 → 4.0	11	0.275	27.5%	33
4.0 → 4.5	4	0.10	10%	37
4.5 → 5	3	0.075	7.5%	40
Total	40	1	100%	

QUESTION: The distance (in km) of 40 engineers from their residence to place of work as follows

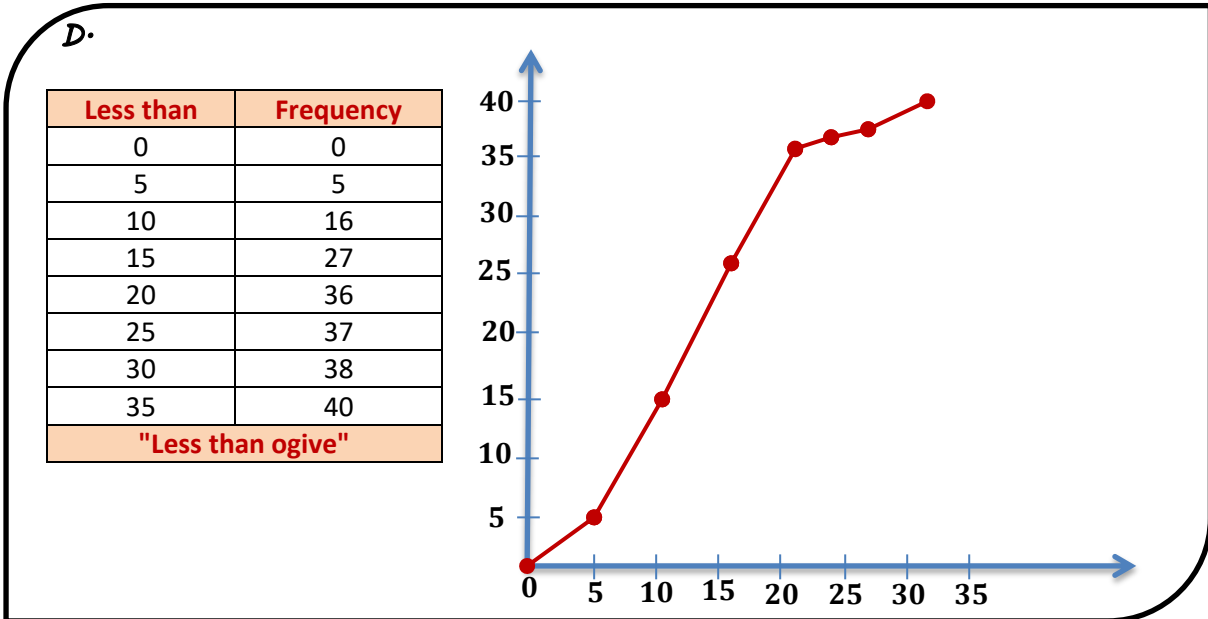
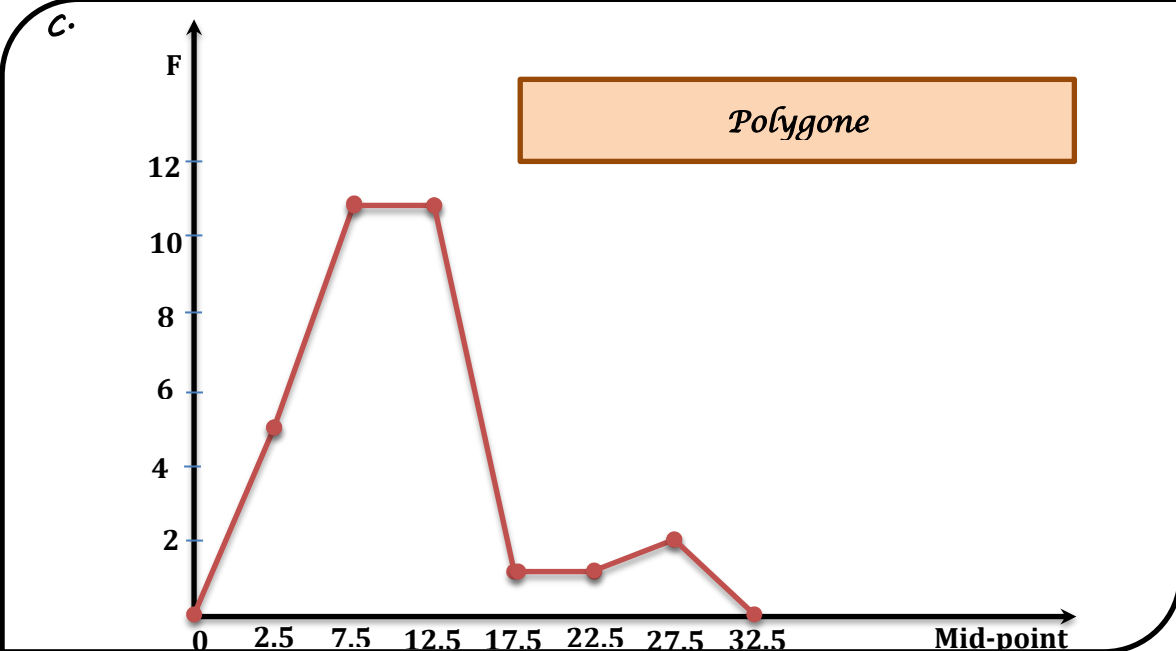
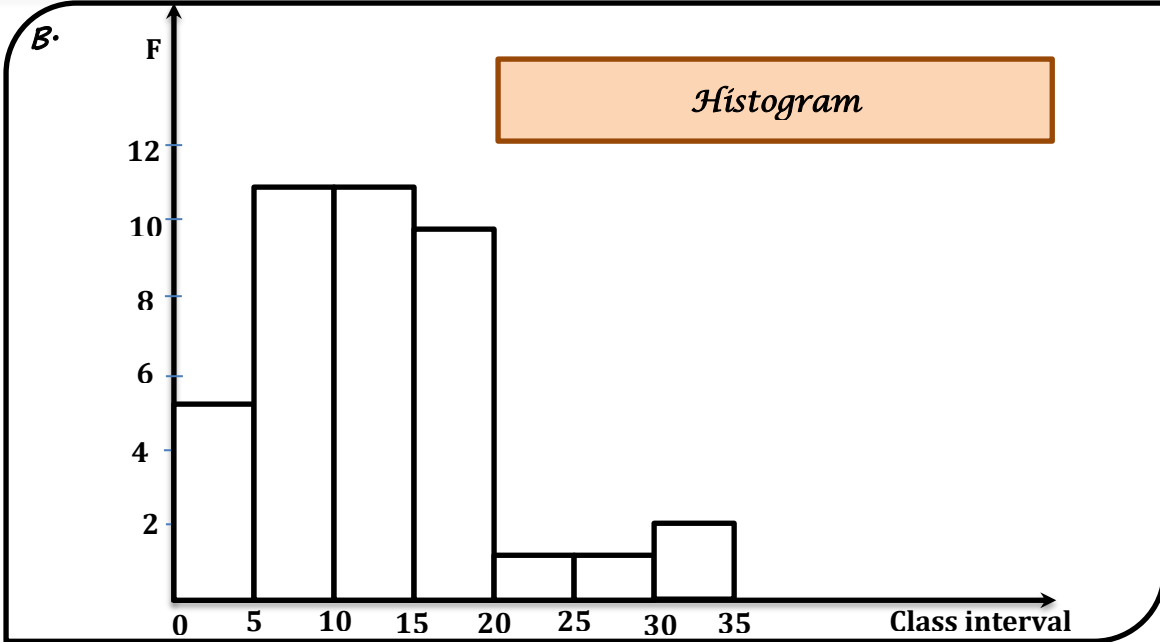
5	3	10	20	25	11	13	7	12	31
19	10	12	17	18	11	32	17	16	2
7	9	7	8	3	5	12	15	18	3
12	14	0.5	9	6	15	15	7	6	12

- A. Construct a frequency distribution table with class size 5
- B. Draw the histogram for the data
- C. Draw the polygon for the data
- D. Draw the ogive
- E. How many engineers have residence at distance more than or equal 15km
- F. How many engineers have residence at distance less than 20km

 **Solution**

A.

Class interval	Frequency	Relative Frequency	Percentage	ASF
0 → 5	5	0.125	12.5%	5
5 → 10	11	0.275	27.5%	16
10 → 15	11	0.275	27.5%	27
15 → 20	9	0.225	22.5%	36
20 → 25	1	0.025	2.5%	37
25 → 30	1	0.025	2.5%	38
30 → 35	2	0.05	5%	40
	40	1	100%	



E. Engineers have residence at distance more than or equal 15 km from their workplace are 13

F. Engineers have residence at distance less than 20km are 27

QUESTION: a sample of 100 children was asked how many times they play computers games for a period of one week

No. of times	No. of children
0 → 3	23
4 → 7	40
8 → 11	28
12 → 15	6
16 → 19	3
Total	100

A. Find the class midpoint

B. Do all classes have the same width? If so, what is this width?

C. Prepare the relative frequency and percentage distribution columns?

D. What percentage of these children play 8 or more times?

 **Solution**

A.

No. of times	Midpoint	F	R.F	P.F	A.C.F
0 → 3	1.5	23	0.23	23%	23
4 → 7	5.5	40	0.40	40%	63
8 → 11	9.5	28	0.28	28%	91
12 → 15	13.5	6	0.06	6%	97
16 → 19	17.5	3	0.03	3%	100
Total		100		100%	

B. All classes have the same width and it is equal 4

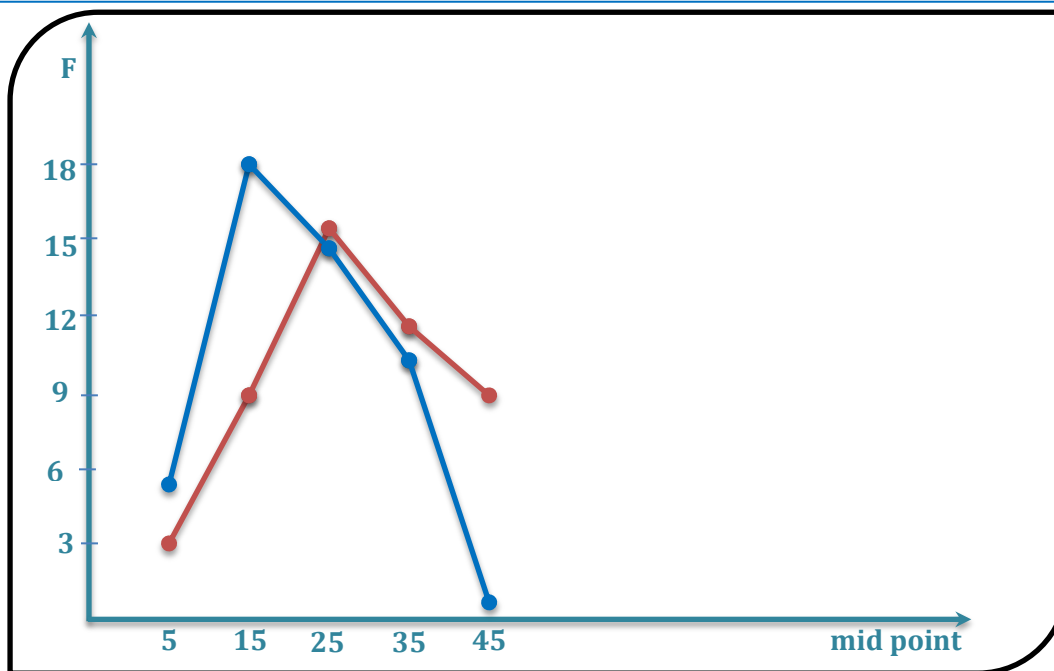
D. the percentage of children play 8 or more $\frac{37}{100} \times 100 = 37\%$

QUESTION: The table below gives the distribution of students of two sections according to the marks obtained

Marks	Section A	Section B
0 → 10	3	5
10 → 20	9	19
20 → 30	17	15
30 → 40	12	10
40 → 50	9	1

Represent the marks on the same graph by two frequency polygons

 **Solution**



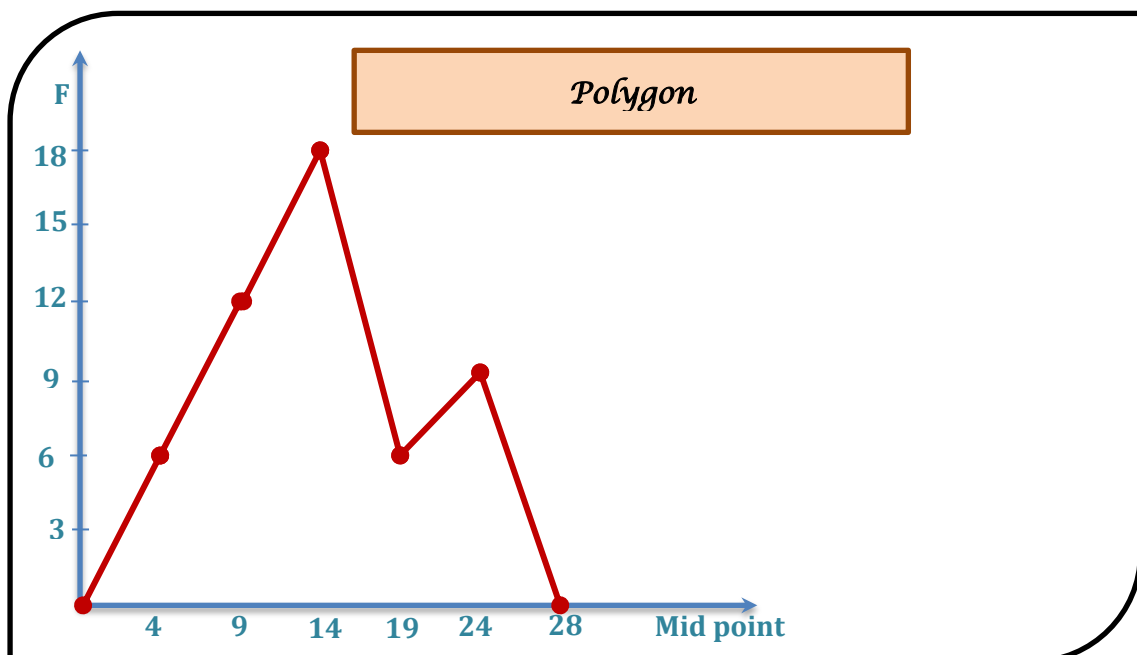
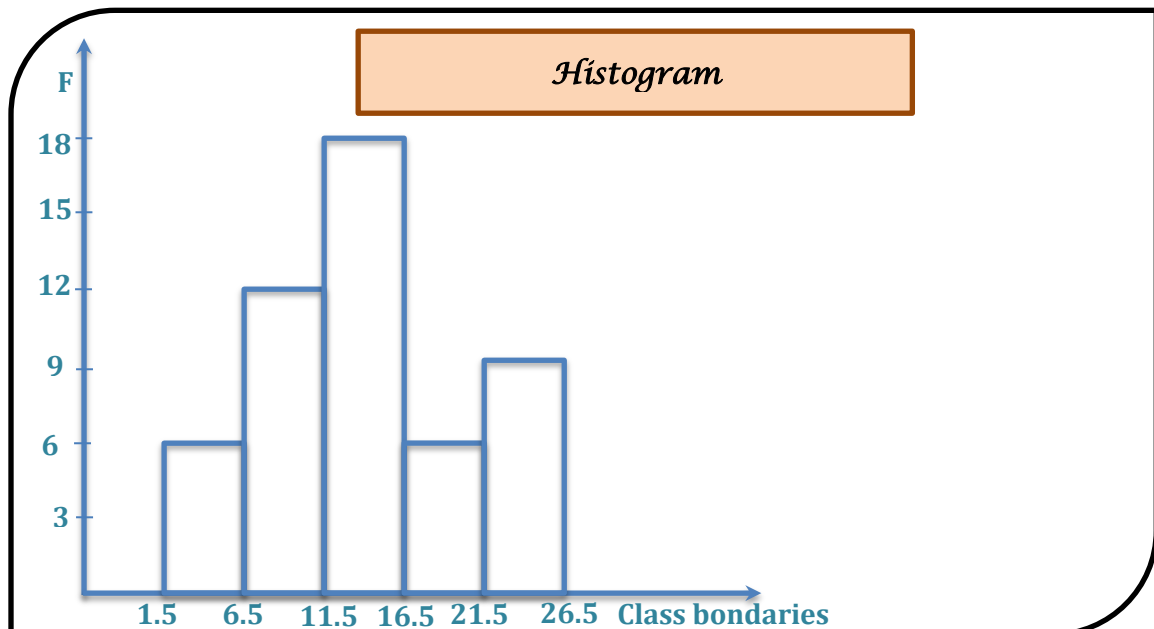
QUESTION: Consider the following frequency distribution, representing the degree of an examination of 50 student of a class.

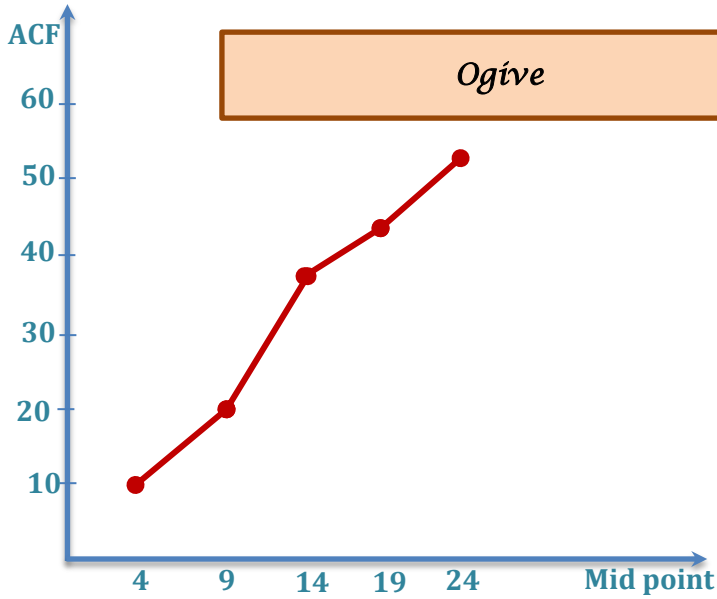
Class limit	Class Boundary	Class midpoint	Frequency	Relative Frequency	ACF
2-6			6		
7-11				0.24	
12-16					36
17-21				0.12	
22-26			8		
Total			50		

- A. Complete the frequency distribution table
- B. Draw the histogram, polygon and ogive
- C. Calculate the mean, the median and mod
- D. Calculate the standard deviation

Class limit	Class boundaries	Class midpoint	Frequency	Relative Frequency	ACF
2-6	1.5-6.5	4	6	0.12	6
7-11	6.5-11.5	9	12	0.24	18
12-16	11.5-16.5	14	18	0.36	36
17-21	16.5-21.5	19	6	0.12	42
22-26	21.5-26.5	24	8	0.16	50
Total			50		

B.





C.

M	4	9	14	19	24	Total
F	6	12	18	6	8	50
FM	24	108	252	114	192	690

$$\bar{X} = \frac{\sum FM}{\sum F}$$

$$\bar{X} = \frac{690}{50} = 13.8$$

Class boundaries	Frequency	ACF
1.5-6.5	6	6
6.5-11.5	12	18
11.5-16.5	18	36
16.5-21.5	6	42
21.5-26.5	8	50
Total	50	

$$\frac{\sum F}{2} = \frac{50}{2} = 25$$

Class med in 11.5-16.5

$$L=11.5 \quad f=18 \quad F=36 \quad C=5$$

$$\bar{X} = L + \frac{\frac{\sum f}{2} - (F - f)}{f} \times C$$

$$\bar{X} = 11.5 + \frac{25 - (36 - 18)}{18} \times 5 = 13.4$$

Class boundaries	Frequency
1.5-6.5	6
6.5-11.5	12 → f_p
11.5-16.5	18 → f_m
16.5-21.5	6 → f_n
21.5-26.5	8
Total	50

$$\hat{X} = L + \frac{d_1}{d_1 + d_2} + C$$

$$d_1 = f_m - f_b = 18 - 12 = 6$$

$$d_2 = f_m - f_n = 18 - 6 = 12$$

The class mode 11.5-16.5

$$L=11.5 \quad C=5$$

$$\bar{X} = 11.5 + \frac{6}{6+12} \times 5 = 13.2$$

D.

يمكن إيجاد الانحراف المعياري بالحاسبة عن طريق إدخال قيم

الـ $X \leftarrow$ midpoint والـ $F \leftarrow$ frequency


QUESTION: The point scored by a team are as follows


17, 2, 7, 27, 15, 5, 14, 8, 10, 24, 48, 10, 8, 7, 18, 28

- A. Calculate the mean and standard deviation
- B. Calculate the standard score of the value 7
- C. Calculate the coefficient of variation
- D. Calculate Q_1, Q_2, Q_3

 **Solution**

$$A. \bar{X} = \frac{\sum X}{n} = \frac{17+2+7+27+15+5+14+8+10+24+48+10+8+7+18+28}{16} = \frac{248}{16} = 15.5$$

 mod $\rightarrow 3 \rightarrow 1$ (البيانات) \rightarrow AC


 Shift $\rightarrow 1 \rightarrow 4 \rightarrow 2(\bar{X})$



...  الحاسبة



لإيجاد الانحراف المعياري ننشئ الجدول التالي:

X	17	2	7	27	15	5	14	8	10	24	48	10	8	7	18	28	$\sum X$ = 248
X^2	289	4	49	729	225	25	196	64	100	576	2304	100	64	49	324	784	$\sum X^2$ = 5882

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} = \sqrt{\frac{5882 - \frac{248^2}{16}}{15}} = 11.656$$

 mod $\rightarrow 3 \rightarrow 1$ (data) \rightarrow AC

 Shift $\rightarrow 1 \rightarrow 4 \rightarrow 2(\bar{X})$  الوسط $2(\bar{X})$

 Shift $\rightarrow 1 \rightarrow 4 \rightarrow 2(\bar{X})$  الانحراف $4(S)$

...  الحاسبة


B. standard score $Z = \frac{X - \bar{X}}{S}$

for $X = 7$ $\bar{X} = 15.5$ $S = 11.656$

$$Z = \frac{7 - 15.5}{11.656} = -0.729$$

C. $C.V = \frac{S}{\bar{X}} \times 100 = \frac{11.656}{15.5} \times 100 = 75.2\%$

 The coefficient variation has no unit

 The variance has unit square

...  خليك فاكتر

$$D. q_r = \frac{r(n+1)}{4}, \text{ where } r = 1, 2, 3$$

$$Q_r = X_K + S(X_{K+1} - X_K)$$

يجب ترتيب البيانات في البداية

2, 5, 7, 7, 8, 8, 10, 10, 14, 15, 17, 18, 24, 27, 28, 48

$$q_1 = \frac{1 \times (16 + 1)}{4} = 4.25$$

أي أن الربع الأول يقع بين القيمة الرابعة والخامسة وقيمتها هي:

$$Q_1 = 7 + 0.25 \times (8 - 7) = 7 + 0.25 = 7.25$$

$$q_2 = \frac{2 \times (16 + 1)}{4} = 8.5 \Rightarrow K = 8 \quad S = 0.5$$

$$Q_2 = 10 + 0.5 \times (14 - 10) = 10 + 2 = 12$$

$$q_3 = \frac{3 \times (16 + 1)}{4} = 12.75 \Rightarrow K = 12 \quad S = 0.75$$

$$Q_3 = 18 + 0.75 \times (24 - 18) = 18 + 4.5 = 22.5$$

QUESTION: The daily sale of sugar (kg) in a certain grocery shop

Mon	Tus	Wed	Thur	Fri	Sat
15	120	12	50	70.5	140.5

- A. Calculate the average
- B. Calculate the variance and standard deviation
- C. Determine the coefficient of variation

 **Solution**

A. The average (mean): $\bar{X} = \frac{\sum X}{n}$

$$\bar{X} = \frac{75 + 120 + 12 + 50 + 70.5 + 140.5}{6} = 78kg$$

B.

X	75	120	12	50	70.5	140.5	$\sum X = 468$
X ²	5625	14400	144	2500	4970.25	19740.25	$\sum X^2 = 4737905$

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} = \sqrt{\frac{47379.5 - \frac{468^2}{6}}{5}} = 46.637 \text{ kg}$$

$$S^2 = 2175.1 \text{ kg}^2$$

$$C \cdot \text{Coefficient variation } C.V = \frac{S}{\bar{X}} \times 100 = \frac{46.637}{78} \times 100 = 59.9\%$$

QUESTION: Let the following data be marks of 10 students

45, 45, 63, 76, 67, 84, 75, 84, 75, 48, 62, 65

- A. Calculate Q1, Q2 and Q3
- B. Calculate IQR
- C. Have the given data extreme value?
- D. Construct the box plot

 **Solution**

أولاً: ترتيب البيانات

45, 45, 48, 62, 63, 65, 67, 75, 76, 84

$$q_r = \frac{r(n+1)}{4} \quad \text{where } r = 1, 2, 3$$

$$Q_r = X_K + S(X_{K+1} - X_K)$$

$$q_1 = \frac{1 \times (10 + 1)}{4} = 2.75 \quad \Rightarrow K = 2 \quad S = 0.75$$

$$Q_1 = 45 + 0.75 \times (48 - 45) = 47.25$$

$$q_2 = \frac{2 \times (10 + 1)}{4} = 5.5 \quad \Rightarrow K = 5 \quad S = 0.5$$

$$Q_2 = 63 + 0.5 \times (65 - 63) = 64$$

$$q_3 = \frac{3 \times (10 + 1)}{4} = 8.25 \quad \Rightarrow K = 8 \quad S = 0.25$$

$$Q_3 = 75 + 0.25 \times (76 - 75) = 75.25$$

$$B \cdot \text{IQR} = Q_3 - Q_1 = 75.25 - 47.25 = 28$$

C. Extreme value are

$$X < Q_1 - 1.5 \times (Q_3 - Q_1)$$

$$X < 47.325 - 1.5 \times (28)$$

$$X < 5.25$$

$$X > Q_3 + 1.5 \times (Q_3 - Q_1)$$

$$X > 75.25 + 1.5 \times (28)$$

$$X > 117.25$$

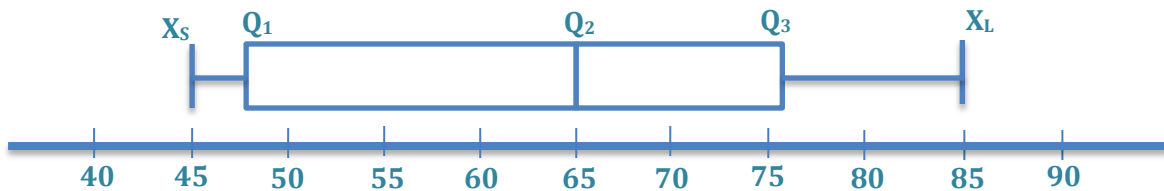
أي قيمة أقل من 5.25 أو أكبر من 117.25 هي قيمة شاذة وبالتالي فإن هذه البيانات لا تحتوي على قيم شاذة

D. To construct the box plot

Five numbers are

Smallest value, Q_1 , Q_2 , Q_3 largest value

45, 47.25, 64, 75.25, 84



QUESTION: Consider the following data

-15, 20, 40, 50, 65, 65, 70, 73, 75, 137

A. Have the given data extreme value

B. Use the suitable measure to calculate the average and dispersion for the give data

 **Solution**

لمعرفة ما إذا كانت توجد قيم شاذة أم لا نوجد الربيعيات حيث:

$$q_r = \frac{r(n+1)}{4} \quad \& \quad Q_r = X_K + S(X_{K+1} - X_K)$$

$$q_1 = \frac{1 \times (10 + 1)}{4} = 2.75 \quad \Rightarrow \quad K = 2 \quad S = 0.75$$

$$Q_1 = 20 + 0.75 \times (40 - 20) = 35$$

$$q_2 = \frac{2 \times (10+1)}{4} = 5.5 \Rightarrow K = 5 \quad S = 0.5$$

$$Q_2 = 65 + 0.5 \times (65 - 65) = 65$$

$$q_3 = \frac{3 \times (10 + 1)}{4} = 8.25 \Rightarrow K = 8 \quad S = 0.25$$

$$Q_3 = 73 + 0.25 \times (75 - 73) = 73.5$$

$$A) X < Q_1 - 1.5 \times (IQR) = 35 - 1.5 \times (38.5) = -22.5$$

$$X > Q_3 + 1.5 \times (IQR) = 73.5 + 1.5 \times (38.5) = 131.25$$

أي قيمة أقل من -22.75 أو أكبر من 131.25 هي قيمة شاذة

ولذلك 137 هي extreme value

لوجود قيمة شاذة في البيانات لا يفضل استخدام الوسط الحسابي لأنه يتأثر بوجود القيم الشاذة وكذلك لحساب التشتت لا يفضل استخدام الانحراف أو التباين لتأثرها بالقيم الشاذة أيضا.

لأن البيانات السابقة لها أكثر من منوال يفضل استخدام الوسيط

$$\text{median } \frac{\cancel{-15}, \cancel{20}, \cancel{40}, \cancel{50}, 65, 65, \cancel{70}, \cancel{73}, \cancel{75}, \cancel{137}}{2} = 65$$

أفضل مقاييس لتشتت في هذه الحالة هو المدى الربيعي

$$B) IQR = Q_3 - Q_1 = 38.5$$

QUESTION: The following data give the number of computer keyboards for a sample of 25 days

45	52	48	41	56
46	44	42	48	53
51	53	51	48	46
43	52	50	54	47
44	47	50	49	52

Prepare a box-plot and then comment the skewness of data

 **Solution**

يجب ترتيب البيانات في البداية وحساب Q_1, Q_2, Q_3 :

Arrange data

41, 42, 43, 44, 44, 45, 46, 46, 47, 47, 48, 48, 48, 49, 50, 50, 51, 51, 52, 52, 52, 53, 53, 54, 56

$$q_r = \frac{r(n+1)}{4} \quad \& \quad Q_r = X_K + S(X_{K+1} - X_K)$$

$$q_1 = \frac{1 \times (25 + 1)}{4} = 6.5 \quad \Rightarrow \quad K = 6 \quad S = 0.5$$

$$Q_1 = 45 + 0.5 \times (45 - 46) = 45.5$$

$$q_2 = \frac{2 \times (25 + 1)}{4} = 13 \quad \Rightarrow \quad K = 13 \quad S = 0$$

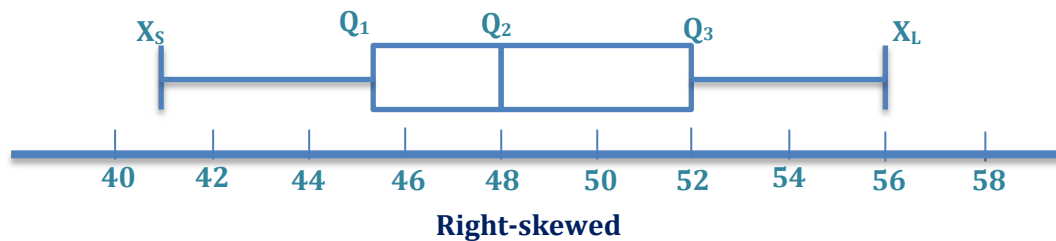
$$Q_2 = 48 + 0 \times (49 - 48) = 48$$

$$q_3 = \frac{3 \times (25 + 1)}{4} = 19.5 \quad \Rightarrow \quad K = 19 \quad S = 0.5$$

$$Q_3 = 52 + 0.5 \times (52 - 52) = 52$$

Five numbers:

Small value Q_1, Q_2, Q_3 large value **41, 45.5, 48, 52, 56**



QUESTION: The mean age of six persons is 49 years, the age of five of these six persons are 55, 39, 44, 51 and 45 find the age of the sixth

Solution

$$\bar{X} = \frac{\sum X}{n}$$

$$49 = \frac{55 + 39 + 44 + 51 + 45 + X}{6}$$

$$234 + X = 294$$

$$X = 294 - 234 = 60$$

QUESTION: The following observation have arranged in ascending order

29, 32, 48, x, x+2, 72, 78, 84, 95

If the median of the data is 63, then

- A. Calculate the value of X
- B. Calculate mean and standard deviation
- C. Find the five number and construct the box plot

 **Solution**

~~29, 32, 48, x, x+2, 72, 78, 84, 95,~~

A. the median $X + 2 = 63 \Rightarrow X = 61$

So the data will be

29, 32, 48, 61, 63, 72, 78, 84, 95

B. The mean $\bar{X} = \frac{\sum X}{n}$

$$\bar{X} = \frac{29+32+48+61+63+72+78+84+95}{9} = \frac{562}{9} = 62.4$$

To calculate standard deviation, we construct the table

X	29	32	48	61	63	72	84	95	$\sum X = 562$
X²	841	1024	2304	3721	3969	5184	7056	8836	$\sum X^2 = 39019$

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} = \sqrt{\frac{39019 - \frac{562^2}{9}}{85}} = 22.15$$

PRINCIPLES OF PROBABILITY & STATISTICS

CHAPTER 2



The fundamental principle of country

مبدأ العد التنازلي

(قاعدة الضرب) Multiplicative rule

إذا كان لدينا عدد K من ظاهرة O_1, O_2, \dots, O_K وهذه الظاهرة تحدث n_1, n_2, \dots, n_K من الطرق بالترتيب وبالتالي فإن عدد الطرق لحدوث كل هذه الظواهر في نفس الوقت هو $n_1 \times n_2 \times \dots \times n_K$ من الطرق.

(قاعدة الجمع) Addition rule

إذا كان لدينا عدد K من ظاهرة O_1, O_2, \dots, O_K وهذه الظاهرة تحدث بـ n_1 أو $n_2 \dots n_K$ من الطرق بالترتيب وبالتالي فإن عدد الطرق لحدوث كل هذه الظواهر في نفس الوقت هو $n_1 + n_2 + \dots + n_K$ من الطرق.

Example How many 6-digits Zip Codes are possible if

- Digits can be repeated?
- Digits can not be repeated?

✓ Solution

A- If digits can be repeated, then number of 6-digits equal to $10 \times 10 \times 10 \times 10 \times 10 \times 10 = 10^6$ codes

B- If digits can not be repeated, then the number of 6-digits equal to $10 \times 9 \times 8 \times 7 \times 6 \times 5 = 151200$ codes

Example How many book can we use Arabic or English characters?

✓ Solution

by addition rule we can index $28+26= 54$ book

(مضروب العدد) Factorial notation

$$n! = n \times (n - 1) \times (n - 2) \dots \dots 3 \times 2 \times 1$$

Note

$$0! = 1 \quad \& \quad 1! = 1$$

Example $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

Permutations (التباديل)

$${}_n P_r = \frac{n!}{(n-r)!} \text{ where } 0 \leq r \leq n$$

خليك فاكِر

Key هناك كلمات مفتاحية نعرف من خلالها أنه يجب استخدام التباديل حيث أن الترتيب فيها مهم
words \Rightarrow order, arrangement, array

Example How many ways one can arrange in order any three of the first 8 letters of L, m, o, p, q, r, s, t

$${}_8 P_3 = \frac{8!}{(8-3)!} = \frac{8!}{5!} = \frac{8 \times 7 \times 6 \times 5!}{5!} = 8 \times 7 \times 6 = 336$$

Combinations (التوافيق)

$${}_n C_r = \frac{n!}{r!(n-r)!} \text{ where } 0 \leq r \leq n$$

خليك فاكِر

في التوافيق الترتيب غير مهم
يمكن كتابة التوافيق أيضا على الصورة $\binom{n}{r}$

Example How many different unordered groups of any three of six letter L, m, n, o, p and q?

Solution

$${}_6 C_3 = \frac{6 \times 5 \times 4}{3 \times 2 \times 1} = 20$$

Example How many ways can you arrange 4 out of 7 books on a shelf.

Solution

لاحظ وجود كلمة arrange وهي من الكلمات الدالة على وجوب الترتيب لذلك نستخدم التبادل

$${}_7 P_4 = 7 \times 6 \times 5 \times 4 = 840$$

7 Shift X 4 = 840 بالحاسبة

Example How many possible different hands of 5 cards each can be dealt from a standard deck of 52 cards.

 **Solution**

This is a combination because order is not important

$$n = 52 , r = 5$$

$${}_{52}C_5 = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = 2598960$$

52

Shift

/

5

=

2598960

بالحاسبة

Example If a man owns 5 pairs of pants, 7 shirts and four pairs of shoes, how many outfits can be assembled?

 **Solution**

By multiplicative rule $5 \times 7 \times 4 = 140$ outfit

Example In a group of 10 people, a 20 \$, 10\$ and 5\$ prize will be given, How many ways can the prize be distributed

 **Solution**

لا حظ أن الترتيب هنا مهم

لأنه من العشرة أفراد سوف يحصل شخص عن جائزة الـ 20\$ ومن التسعة أفراد الباقية سوف الباقية سوف يحصل شخص على جائزة الـ 10\$ وبالتالي فإن فرد من الثمانية الباقين سوف يحصل على 5\$

$${}_{10}P_3 = 10 \times 9 \times 8 = 720$$

Example In a group of 10 people, three 5\$ prizes will be given, How many ways can the prizes be distributed

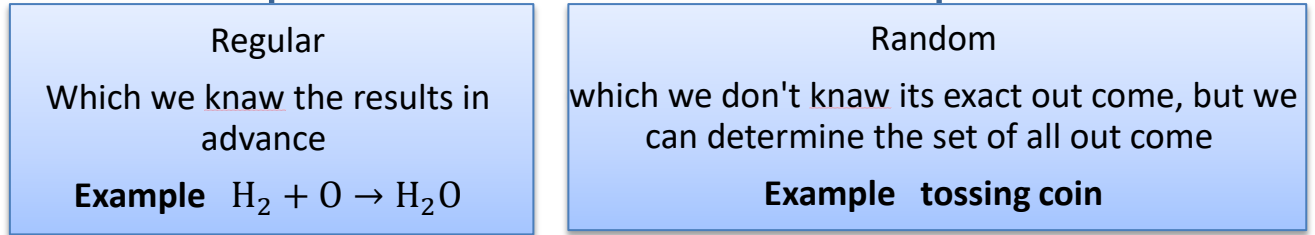
 **Solution**

Order does not matter ${}_{10}C_3 = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 120$

- ❖ and \rightarrow multiplicative rule
- ❖ or \rightarrow additional rule
- ❖ ordered matter \rightarrow permutation
- ❖ ordered does not matter \rightarrow combination

خليك فاكتر

Type of experiment



في التجربة العشوائية يقال لحدثان لهما نفس الفرصة في الظهور equally likely

خليك فاكسر

Probability science: is a branch of mathematics that deals with theoretical models of random experiments



Note probability space of any random experiment is a tripple have from

$$[\Omega, \mathcal{A}, p]$$

Ω is the set of all outcome

\mathcal{A} is algebra of events

p is called probability measure

أولاً: Ω هي مجموعة كل النواتج الممكنة لتجربة عشوائية

Example In the experiment of flip a coin and roll adie , The set of all outcome

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

Event is a subset of sample space

Example Tossing a coin three times and determine the Sample space, then determine the following events

E_1 : The event of two heads

E_2 : The event that at least two heads

E_3 : The event that at most two heads

E_4 : The event that a heads is the first toss

 **Solution**

$$\Omega = \{HHH, HHT, HTH, THH, TTT, TTH, THT, HTT\}$$

$$E_1 = \{HHT, HTH, THH\}$$

$$E_2 = \{HHT, HTH, THH, HHH\}$$

$$E_3 = \{HHT, HTH, THH, TTT, TTH, THT, HTT\}$$

$$E_4 = \{HHH, HHT, HTH, HTT\}$$

Algebra of events:

ثانياً:

A collection \mathcal{A} of subset of Ω is said to be an algebra on Ω if

- $\Omega \in \mathcal{A}$
- If A and $B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$
- If $A \in \mathcal{A}$, then $\bar{A} \in \mathcal{A}$

- ✓ هناك نوعان من الحوادث
- ✓ حوادث بسيطة تحتوى على عنصر واحد فقط
- ✓ حوادث مركبة تحتوي أكثر من عنصر

Example In tossing die , the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$

The events $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\} \Rightarrow$ Simple

Even events $\{2, 4, 6\} \Rightarrow$ Compound

Odd events $\{1, 3, 5\} \Rightarrow$ Compound

Note if Ω is a set, then the set of all subset in Ω is 2^Ω

Example $\Omega = \{a, b, c\} \Rightarrow |\Omega| = 3$

Then the set of all elements is $2^{|\Omega|} = 2^3 = 8$

$\phi, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}$

بعض العمليات على الأحداث

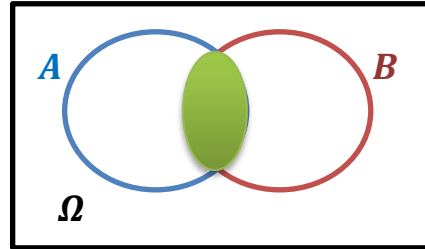
Union (الاتحاد)

Is denoted by $A \cup B$, is an event containing all element in A or B or both
i.e $A \cup B$ is the occurrence of at least one event

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

$A = \{1, 2\}$, $B = \{2, 5, 6\}$

Then $A \cup B = \{1, 2, 5, 6\}$



$A \cup B$

Intersection (التقاطع)

$A \cap B$ or A and B

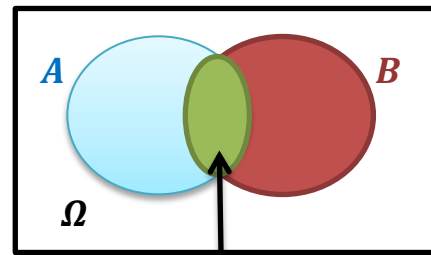
$A \cap B$ is an event containing all elements in A and B in the same time

Example tossing coin two time

$\Omega = \{HH, HT, TH, TT\}$

$A = \{TH, TT\}$, $B = \{HT, TT\}$

$A \cap B = \{TT\}$



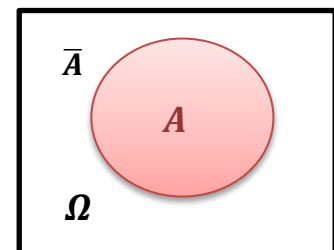
$A \cap B$

Complement (\bar{A}) (المتمة)

The complement of an event that occurs if A does not

Example Tossing adie , determine the following event

- E_1 : the event of even or prime
- E_2 : the event of prime
- E_3 : the event of prime does not occur



Solution

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

$$E_1 = \{2, 4, 6\} \cup \{2, 3, 5\} = \{2, 3, 4, 5, 6\}$$

$$E_2 = \{2, 3, 5\}$$

$$E_3 \text{ is the complement of } E_2 \Rightarrow E_3 = \{1, 4, 6\}$$

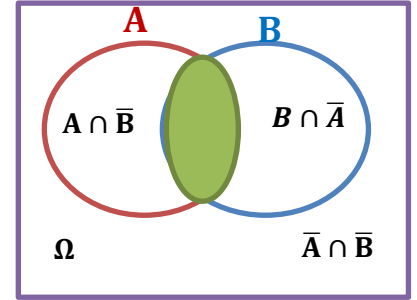
Difference between two events (الفرق)

يرمز لها بأحد الرموز $A \setminus B$ أو $A \cap \bar{B}$ وهي تعني العناصر الموجودة في A وغير موجود في B

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

$A = \{1, 2\}$ $B = \{2, 3, 5\}$

$$A \setminus B = \{1\}$$



Exactly one event

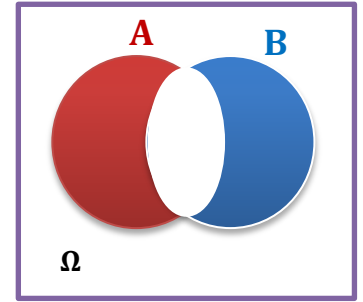
If A and B are two event from Ω , then:

$$(A \setminus B) \cup (B \setminus A) \text{ or } (A \cap \bar{B}) \cup (B \cap \bar{A})$$

Example let $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$, $C = \{2, 4, 6, 7\}$

Find $A \Delta B \Delta C$

Solution



First we find the symmetric of first two set A and B

$$A \Delta B = \{1, 2, 5, 6\}$$
 then

$$A \Delta B \Delta C = \{1, 4, 5, 7\}$$

Impossible event (الحدث المستحيل)

هو الحدث الذي لا يمكن وقوعه مثال ظهور العدد 7 عند القاء زهر أو ظهور عدد سالب لنفس التجربة كما أن تقاطع أي حدث مع متممه مثال على الحوادث المستحيلة

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

A the event of even number $A = \{2, 4, 6\}$ So that the event of odd number $\bar{A} = \{1, 3, 5\}$ then $A \cap \bar{A} = \emptyset \Rightarrow$ impossible event

Certain event (الحدث المؤكد)

هو ذلك الحدث الذي لا بد من حدوثه بشكل مؤكد ومثال ذلك احتمال وقوع الحدث اتحاد متممة ذلك الحدث هو sure event

$$A \cup \bar{A} = \Omega$$

for any event A and \bar{A}

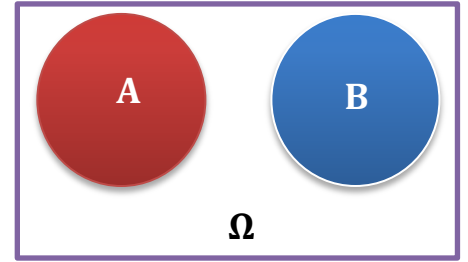
Mutually exclusive events (الأحداث المتنافية)

يقال لحدثان A و B أنهما متنافيان أو disjoint إذا كانت $A \cap B = \emptyset$

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

$A = \{1, 2\}$ $B = \{3, 4, 6\}$ then $A \cap B = \emptyset$

So we said A and B are disjoint or mutually



EXERCISES

QUESTION1: If an automobile license plate must consist of three letters followed by three single digit numbers.

how many different license plates are possible?

Solution

According to the fundamental principle of counting the possible number of license plate = $26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17576000$

QUESTION2: The science club has challenged the math club each club's team should be comprised of 2 boys and 3 girls, there are 20 boys and 15 girls in the science club and 25 boys and 30 girls in the math club

Solution

For science

$${}_{20}C_2 \times {}_{15}C_3 = 86450$$

For math

$${}_{25}C_2 \times {}_{30}C_3 = 1218000$$

QUESTION3: A manager must choose five secretaries from a many 12 applicants and assign them to different stations how many different arrangement are possible?

Solution

لاحظ أن الترتيب هنا مهم في هذا المثال ولذلك نستخدم التباديل

$$n = 12 \quad r = 5 \quad \Rightarrow nP_r = 12P_5 = 95040$$

CONCEPT OF PROBABILITY

Probability (الاحتمال)

Is the numerical measure of likelihood that a specific event will occur

الاحتمال هو مقياس عددي لوقوع حدث ما

التكرار النسبي هو تكرار لفئة مقسومة على المجموع الكلي للتكرار وهو نفس التعريف الكلاسيكي للاحتتمال

$$F_A = \frac{n(A)}{N} \Rightarrow P(A) = \frac{n(A)}{N} \text{ : حيث}$$

Laplace principle of probability

خليك فاكِر

For simple event $E \Rightarrow P(E) = \frac{1}{|\Omega|}$

For compound event $A \Rightarrow P(A) = \frac{|A|}{|\Omega|}$

In the experiment of rolling a fair die find the

A: probability of event number

B: probability of odd number

C: Probability of number less than two

D: probability of number more than six

Solution

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

$$A = \{2, 4, 6\} \Rightarrow |A| = 3 \quad \& \quad B = \{1, 3, 5\} \Rightarrow |B| = 3$$

$$C = \{1\} \Rightarrow |C| = 1 \quad \& \quad D = \phi \Rightarrow |D| = 0$$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2} \quad \text{حدث مركب}$$

$$P(B) = \frac{|B|}{|\Omega|} = \frac{3}{6} = \frac{1}{2} \quad \text{حدث مركب}$$

$$P(C) = \frac{|C|}{|\Omega|} = \frac{1}{6} \quad \text{حدث بسيط}$$

$$P(D) = \frac{|D|}{|\Omega|} = \frac{0}{6} = 0 \quad \text{حدث مستحيل}$$

- 1 $0 \leq P(A) \leq 1$
- 2 $\sum P(A) = 1$
- 3 $P(\emptyset) = 0$
- 4 If $A \cap B = \emptyset \Rightarrow P(A \cap B) = 0$ disjoint
- 5 $P(\Omega) = 1$
- 6 $P(A) + P(\bar{A}) = 1$
- 7 If $A \subset B, \Rightarrow P(A) \leq P(B)$
- 8 $P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$
- 9 $P(\overline{A \cap B}) = P(\bar{A} \cup \bar{B})$
- 10 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ additive rule

De_Mogan's laws

If A and B are disjoint $\Rightarrow P(A \cap B) = 0$ then,

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

EXAMPLE:

Let $P(A) = \frac{3}{8}$

$P(B) = \frac{1}{2}$

$P(A \cap B) = \frac{1}{4}$

Find: **A)** $P(A \cup B)$

B) $P(\bar{A}), P(\bar{B})$

C) $P(\bar{A} \cap \bar{B})$

D) $P(\bar{A} \cup \bar{B})$

E) $P(A \cap \bar{B})$

F) $P(\bar{A} \cap B)$

Solution

A) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{8} + \frac{1}{2} - \frac{1}{4} = \frac{5}{8}$

B) $P(\bar{A}) = 1 - P(A) = 1 - \frac{3}{8} = \frac{5}{8}$

$P(\bar{B}) = 1 - P(B) = 1 - \frac{1}{2} = \frac{1}{2}$

C) $P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - \frac{5}{8} = \frac{3}{8}$

D) $P(\bar{A} \cup \bar{B}) = P(\overline{A \cap B}) = 1 - P(A \cap B) = 1 - \frac{1}{4} = \frac{3}{4}$

E) $P(A \cap \bar{B}) = P(A) - P(A \cap B) = \frac{3}{8} - \frac{1}{4} = \frac{1}{8}$

F) $P(\bar{A} \cap B) = P(B) - P(A \cap B) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$

EXAMPLE: Three students A, B, and C are in a swimming race. A and B have the same probability of swimming and each is twice as likely to win as C.

Find the probability that B or C wins.

 **Solution**

$$\text{Let } P(C) = x, \text{ so } P(A) = P(B) = 2x$$

$$\text{since } \sum P(x) = 1 \Rightarrow P(A) + P(B) + P(C) = 1$$

$$2x + 2x + x = 1 \Rightarrow 5x = 1$$

$$\text{So } P(A) = P(B) = \frac{2}{5} \text{ and } P(C) = \frac{1}{5}$$

$$P(B \cup C) = P(B) + P(C) = \frac{2}{5} + \frac{1}{5} = \frac{3}{5}$$

Example: Let die is weighted, so the even number have the same chance of appearing, the odd number have the same chance, and each even number is twice odd.

Find the probability that:

- A) An even number appears
- B) Odd number appears
- C) A prime number appears
- D) An odd number but not prime

 **Solution**

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

$$\text{Let } A = \{2, 4, 6\} \text{ \& } B = \{1, 3, 5\} \text{ \& } C = \{2, 3, 5\}$$

even *odd* *prime*

$$P(2) + P(4) + P(6) + 2P(1) + 2P(3) + 2P(5) = 1$$

$$x + x + x + 2x + 2x + 2x = 1$$

$$9x = 1 \Rightarrow x = \frac{1}{9}, \text{ then } P(1) = P(3) = P(5) = \frac{1}{9}$$

$$P(2) = P(4) = P(6) = \frac{2}{9}$$

$$\mathbf{A)} P(A) = P(2) + P(4) + P(6) = \frac{6}{9}$$

$$\mathbf{B)} P(B) = P(1) + P(3) + P(5) = \frac{3}{9}$$

$$C) P(C) = P(2) + P(3) + P(5) = \frac{4}{9}$$

$$D) P(B \cap \bar{C}) = P(B) - P(B \cap C) = \frac{3}{9} - \frac{2}{9} = \frac{1}{9}$$

Example: Let $P(A) = 0.8$ $P(B) = 0.55$ $P(A \cup B) = 0.9$

Find :

- A) Occurrence of A and B
- B) Occurrence of only A and not B
- C) Non occurrence of A and B
- D) Occurrence of only A or only B

 **Solution**

$$A) P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.8 + 0.55 - 0.9 = 0.45$$

$$B) P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.8 - 0.45 = 0.35$$

$$C) P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.9 = 0.1$$

$$D) P(A \cap \bar{B}) \cup P(\bar{A} \cap B) = P(A) - P(A \cap B) + P(B) - P(A \cap B) = 0.35 + 0.10 = 0.45$$

Example: $P(A) = \frac{1}{2}$ $P(B) = \frac{1}{8}$ $P(C) = \frac{1}{4}$

Where A, B and are mutually exclusive, Find :

- A) $P(A \cup B \cup C)$
- B) $P(\bar{A} \cap \bar{B} \cap \bar{C})$

 **Solution**

$$A) P(A \cup B \cup C) = P(A) + P(B) + P(C) = \frac{1}{2} + \frac{1}{8} + \frac{1}{4} = \frac{7}{8}$$

$$B) P(\bar{A} \cap \bar{B} \cap \bar{C}) = P(\overline{A \cup B \cup C}) = 1 - P(A \cup B \cup C) = 1 - \frac{7}{8} = \frac{1}{8}$$

Example: A certain family owns two television sets one color and one black and white, let A be the event of color and B is the event of black and white

if $P(A) = 0.4$ $P(B) = 0.3$ $P(A \cup B) = 0.5$ Find :

- A) Both sets are on
- B) The color set on and other is off
- C) Exactly one set is on

D) neither set is on

 **Solution**

$$A) P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.3 - 0.5 = 0.2$$

$$B) P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.4 - 0.2 = 0.2$$

$$C) P(A \cap \bar{B}) \cup P(\bar{A} \cap B) = 0.2 + P(B) - P(A \cap B) = 0.2 + 0.3 - 0.2 = 0.3$$

$$D) P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.5 = 0.5$$

EXAMPLE: Let $P(A) = 0.4$ $P(A \cup B) = 0.6$ find the value of $P(B)$ which make A and B mutually exclusive

 **Solution**

Since A and B are disjoint $P(A \cup B) = P(A) + P(B)$

$$0.6 = 0.4 + P(B) \Rightarrow P(B) = 0.2$$

Example: There are two traffic lights on the route used by pickup Andropov to go from home to work, let E denote the event that pickup must stop at the first light and F in a similar manner for second light, suppose $P(E) = 0.4$, $P(F) = 0.3$ and $P(E \cap F) = 0.15$ what is the probability that he:

- A) Must stop for at least one light?
- B) Doesn't stop at either light?
- C) Must stop at exactly one light
- D) Must stop just at the first light

 **Solution**

$$A) P(E \cup F) = P(E) + P(F) - P(E \cap F) = 0.4 + 0.3 - 0.15 = 0.55$$

$$B) P(\bar{E} \cap \bar{F}) = P(\overline{E \cup F}) = 1 - P(E \cup F) = 1 - 0.55 = 0.45$$

$$C) P(E \cap \bar{F}) \cup P(F \cap \bar{E}) = P(E) - P(E \cap F) + P(F) - P(E \cap F) = 0.4 - 0.15 + 0.3 - 0.15 = 0.4$$

$$D) P(E \cap \bar{F}) = P(E) - P(E \cap F) = 0.4 - 0.15 = 0.25$$

Example: In a hospital, there are 12 nurses and 4 doctors

- A) if 4 nurses and 2 doctors to be chosen, how many possibilities?
- B) If a committee contains 6 person, what is the probability that two doctors in this committee?

 **Solution**

الترتيب غير مهم بالسؤال ولذلك نستخدم التوافق

$$A) nC_r = \frac{n!}{r!(n-r)!}$$

$$12C_4 = \frac{12!}{4!(12-4)!} = \frac{12 \times 11 \times 10 \times 9}{4 \times 3 \times 2 \times 1} = 495$$

$$4C_2 = \frac{4!}{2!(4-2)!} = \frac{4 \times 3}{2 \times 1} = 6$$

All possibilities are $495 \times 6 = 2970$

$$B) P(A) = \frac{n(A)}{N} = \frac{12C_4 \times 4C_2}{16C_6} = \frac{2970}{8008}$$

Example: Let A, B, and C $\in 2^\Omega$, where

$P(A - B) = 0.15$, $P(B - A) = 0.30$, $P(C - A) = 0.35$
 $P(A \cap B) = 0.10$, $P(A \cap C) = 0.15$, $P(B \cap C) = 0.20$
and $P(A \cap B \cap C) = 0.05$ then find

A)

$P(A)$
 $P(C)$

$P(A - C)$
 $P(B - C)$

$P(B)$
 $P(A \setminus B)$

$P(C - B)$
 $P(A \setminus B \cap C)$

B) If you know that

$$P(A \cup B \cup C) =$$

$$P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Find $P(A \cup B \cup C)$, $P(\bar{A} \cap \bar{B} \cap \bar{C})$

C) Are the events A, B and C independent

 **Solution**

A)

$$P(A - B) = P(A) - P(A \cap B) \Rightarrow 0.15 = P(A) - 0.10 \Rightarrow P(A) = 0.25$$

$$P(A - C) = P(A) - P(A \cap C) = 0.25 - 0.15 \Rightarrow P(A - C) = 0.10$$

$$P(B - A) = P(B) - P(B \cap A) \Rightarrow 0.30 = P(B) - 0.10 \Rightarrow P(B) = 0.40$$

$$P(C - A) = P(C) - P(C \cap A) \Rightarrow 0.35 = P(C) - 0.15 \Rightarrow P(C) = 0.50$$

$$P(B - C) = P(B) - P(B \cap C) \Rightarrow P(B - C) = 0.4 - 0.2 = 0.2$$

$$P(C - B) = P(C) - P(C \cap B) = 0.5 - 0.2 \Rightarrow P(C - B) = 0.3$$

$$P(A \setminus B) = \frac{P(A \cap B)}{P(B)} = \frac{0.10}{0.40} = 0.25$$

$$P(A \setminus B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{0.05}{0.20} = 0.25$$

B)

$$P(A \cup B \cup C) = 0.25 + 0.40 + 0.5 - 0.10 - 0.15 - 0.20 + 0.05 = 0.75$$

$$P(\bar{A} \cap \bar{B} \cap \bar{C}) = P(\overline{A \cup B \cup C}) = 1 - P(A \cup B \cup C) = 1 - 0.75 = 0.25$$

C)

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

$$0.05 = (0.25)(0.40)(0.5)$$

$$0.05 = 0.05 \Rightarrow \text{The event are independent}$$

1 $P(A \cap B) = P(A) \cdot P(B)$ independent

$$P(A \cap B) = (0.03)(0.08) = 0.0024$$

2 at least one machines $\Rightarrow P(A \cup B)$

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) = 0.03 + 0.08 + 0.0024 \\ &= 0.1076 \end{aligned}$$

3 only first $\Rightarrow P(A \cap \bar{B})$

$$P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.03 - 0.0024$$

4 $P(A \setminus B) = \frac{P(A \cap \bar{B})}{P(A)} = \frac{P(A) \cdot P(\bar{B})}{P(A)} = P(\bar{B}) = 0.0800$

Conditional probability and independence of event

The probability of A given B defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$



① If A_1 , and A_2 are mutually exclusive, then

$$P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$$

② If $P(B) > 0$, then $P(\Omega | B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

③ If $B \subseteq A$, then $A \cap B = B$,

Therefore

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

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

$$\text{So, } P(B) = P(A) \cdot P(B|A)$$

Example: Let $\Omega = \{1, 2, 3, 4, 5, 6\}$

$$A_1 = \{1, 5\} \Rightarrow P(A_1) = \frac{1}{3} \quad \& \quad A_2 = \{2, 4, 6\} \Rightarrow P(A_2) = \frac{1}{2}$$

$$B = \{1, 2, 3\} \Rightarrow P(B) = \frac{1}{2}$$

$$A_1 \cap B = \{1\} \Rightarrow P(A_1 \cap B) = \frac{1}{6}$$

$$A_2 \cap B = \{2\} \Rightarrow P(A_2 \cap B) = \frac{1}{6}$$

So, $P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$

$$= \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} = \frac{1/6}{3/6} + \frac{1/6}{3/6} = \frac{2}{3}$$

Example: A student is randomly selected from a class where 35% is left hand, and 50% are sophomores, we know that 5% of the class consists of left handed sophomores, Given that a randomly selected student is a sophomore, what is the probability that he is left handed

✓ **Solution**

Let $P(A) = 0.35$, $P(B) = 0.5$, $P(A \cap B) = 0.05$

$$\text{Then } P(A \setminus B) = \frac{P(A \cap B)}{P(B)} = \frac{0.05}{0.05} = 0.1$$

يمكن استنتاج قواعد أخرى من الاحتمال الشرطي كما يلي

If $B(> 0)$

$$P(A \setminus B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(B) \cdot P(A \setminus B)$$

If $A(> 0)$

$$P(B \setminus A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(A \cap B) = P(A) \cdot P(B \setminus A)$$



Independent event (الأحداث المستقلة)

Example: In tossing coin two times, let A means get a head on the first flip, B means get a head on the second flip, what the probability that you get a head on the second flip given that you had a head on the first flip?

✓ **Solution**

$$\Omega = \{HH, HT, TH, TT\}$$

$$A = \{HH, HT\} \Rightarrow P(A) = \frac{2}{4} = \frac{1}{2}$$

$$B = \{HH, TH\} \Rightarrow P(B) = \frac{2}{4} = \frac{1}{2}$$

$$P(A \cap B) = \{HH\} \Rightarrow P(A \cap B) = \frac{1}{4}$$

$$P(A \setminus B) = \frac{P(B \cap A)}{P(A)} = P(A) \cdot P(B) \Rightarrow \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

So A and B are independent

Example: In tossing coin three times, let A get a head on first toss and B get a head on the second toss, C is get head in third toss, what is the probability of getting a head of third toss given that the previous two flips were heads?

✓ **Solution**

$$\Omega = \{HHH, HHT, HTH, THH, TTT, TTH, THT, HTT\}$$

$$A = \{HHH, HHT, HTH, HTT\} \Rightarrow P(A) = \frac{4}{8}$$

$$B = \{HHH, HHT, THH, THT\} \Rightarrow P(B) = \frac{4}{8}$$

$$A \cap B = \{HHH, HHT\} \Rightarrow P(A \cap B) = \frac{2}{8}$$

$$C = \{HHH, HTH, THH, TTH\} \Rightarrow P(C) = \frac{4}{8}$$

$$P(C \setminus A \cap B) = \frac{P(C \cap B \cap A)}{P(A \cap B)} = \frac{1/8}{2/8} = \frac{1}{2}$$

✚ **Mutually exclusive** $\Rightarrow P(A \cap B) = 0$

✚ **Independent** $\Rightarrow P(A \cap B) = P(A) \cdot P(B)$

خليك فاكِر

Example: Give that a student studied, the probability of passing quiz is 0.99. Given the student did not study, the probability of passing quiz is 0.05. Assume that the probability of studying is 0.7. A student flunks the quiz, what is the probability that he did not study

✓ **Solution**

$$P(\text{pass} \setminus \text{studied}) = 0.99 \Rightarrow P(A \setminus B) = 0.99$$

$$P(\text{pass} \setminus \text{not study}) = 0.05 \Rightarrow P(A \setminus \bar{B}) = 0.05$$

$$P(\text{study}) = 0.7 \Rightarrow P(B) = 0.7$$

$$P(\text{no study} \setminus \text{no pass}) \Rightarrow P(\bar{B} \setminus \bar{A})$$

$$P(\bar{B} \setminus \bar{A}) = \frac{P(\bar{B} \cap \bar{A})}{P(\bar{B})}$$

$$P(A \cap B) = P(B) \cdot P(A \setminus B) = 0.7 \times 0.99 = 0.693$$

$$P(A \setminus \bar{B}) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{P(A) - P(A \cap B)}{P(\bar{B})} \Rightarrow 0.05 = \frac{P(A) - 0.693}{0.3}$$

$$P(A) = 0.708$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.708 + 0.3 - 0.693$$

$$\text{Then } P(\bar{B} \setminus \bar{A}) = \frac{P(\bar{B} \cap \bar{A})}{P(\bar{A})} = \frac{P(\overline{B \cup A})}{P(\bar{A})} = \frac{1 - P(A \cup B)}{P(\bar{A})} = \frac{1 - 0.715}{1 - 0.708} = 0.976$$

Example: In a certain college 0.25 of students failed math, 0.15 of them failed in chemistry and 0.10 failed in both

- A)** If he failed chemistry, what the probability that he failed in math?
- B)** If he failed in math, what the probability that he failed in chemistry?
- C)** What is the probability that he failed in math or chemistry?

 **Solution**

$$P(M) = 0.25$$

$$P(C) = 0.15$$

$$P(M \cap C) = 0.10$$

$$\text{A) } P(M \setminus C) = \frac{P(M \cap C)}{P(C)} = \frac{0.10}{0.15} = \frac{2}{3}$$

$$\text{B) } P(C \setminus M) = \frac{P(C \cap M)}{P(M)} = \frac{0.10}{0.25} = \frac{2}{5}$$

$$\text{C) } P(M \cup C) = P(M) + P(C) - P(M \cap C) = 0.25 + 0.15 - 0.10 = 0.30$$

BAY'S THEOREM

$$P(A \setminus B) = \frac{P(A) \bullet P(B \setminus A)}{P(A) \bullet P(B \setminus A) + P(\bar{A}) \bullet P(B \setminus \bar{A})}$$

BAYE'S THEOREM

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})}$$

Example: In a statistics class it was found that 60% of student attend class on Thursday, from past data it was found 98% of those who went to class on Thursday pass the course, while 20% of those who did not go to class on Thursday passed the course.

- A)** What percentage of students is expected to pass the course?
B) Given a student passes the course, what is the probability that he attend classed on Thursday

Solution

A_1 : student attend class

A_2 : student did not attend

B_1 : student pass the course

B_2 : do not pass the course

A)

$$P(A_1) = 0.6 \quad P(A_2) = 0.4 \quad P(B_1|A_1) = 0.98 \quad P(B_2|A_2) = 0.2$$
$$P(B_1) = P(B_1 \cap A_1) + P(B_1 \cap A_2)$$

$$\text{Since } P(B_1|A_1) = \frac{P(B_1 \cap A_1)}{P(A_1)} \Rightarrow P(B_1 \cap A_1) = P(B_1|A_1) \cdot P(A_1)$$

$$P(B_1|A_2) = \frac{P(B_1 \cap A_2)}{P(A_2)} \Rightarrow P(B_1 \cap A_2) = P(B_1|A_2) \cdot P(A_2)$$

$$\text{So } P(B_1) = P(B_1|A_1) \cdot P(A_1) + P(B_1|A_2) \cdot P(A_2)$$

$$= 0.98 \times 0.6 + 0.2 \times 0.4 = 0.668$$

$$\text{B) } P(A_1|B_1) = \frac{P(B_1|A_1) \cdot P(A_1)}{P(B_1|A_1) \cdot P(A_1) + P(B_1|A_2) \cdot P(A_2)} = \frac{0.6 \times 0.98}{0.6 \times 0.98 + 0.4 \times 0.2} = 0.854$$

