Statistics

File: All required material for First Exam

Concept:

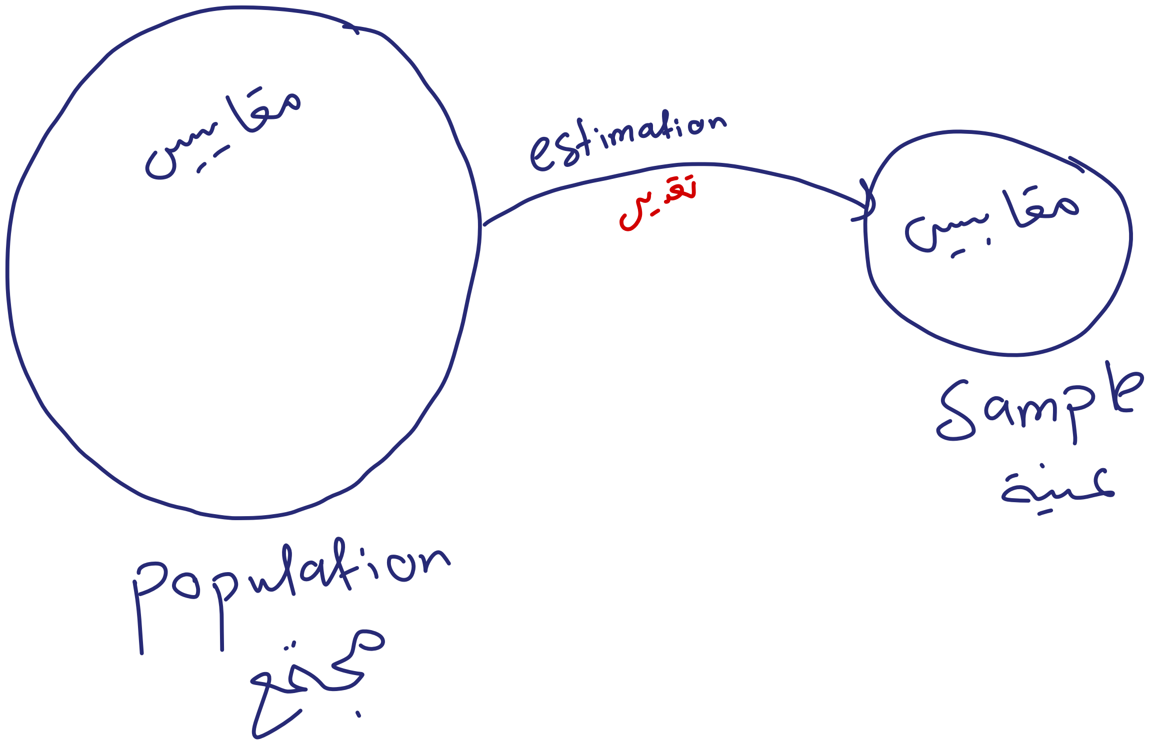
Chapter 1

Introduction to statistics

إحصاء
⇒ statistics ⇒ علم $\xrightarrow{\text{يدرس}}$ ظواهر معينة → Random عشوائية
→ deterministic محددة

⇓

استنتاجات
(Inferences)



* * *

* Mathematical statistics
↳ Branch

Techniques أساليب + statistical methods أساليب

* Applied statistics

↳ Application of statistical methods and Techniques

BIOSTASTICS

↙ medicine

↘ Biology

البيانات

(Data)

question → Data collection → Data entry

جمع البيانات.

↓

Data editing → Data Analysis

تحليل البيانات

↓

Inferences → publication

استنتاجات

∴

تلخيص

Data ⇒ Summarization

∴

Descriptive statistice → summarize, descripe

الإحصاء الوصفي

Inferential statistics → Conclusions, predictions

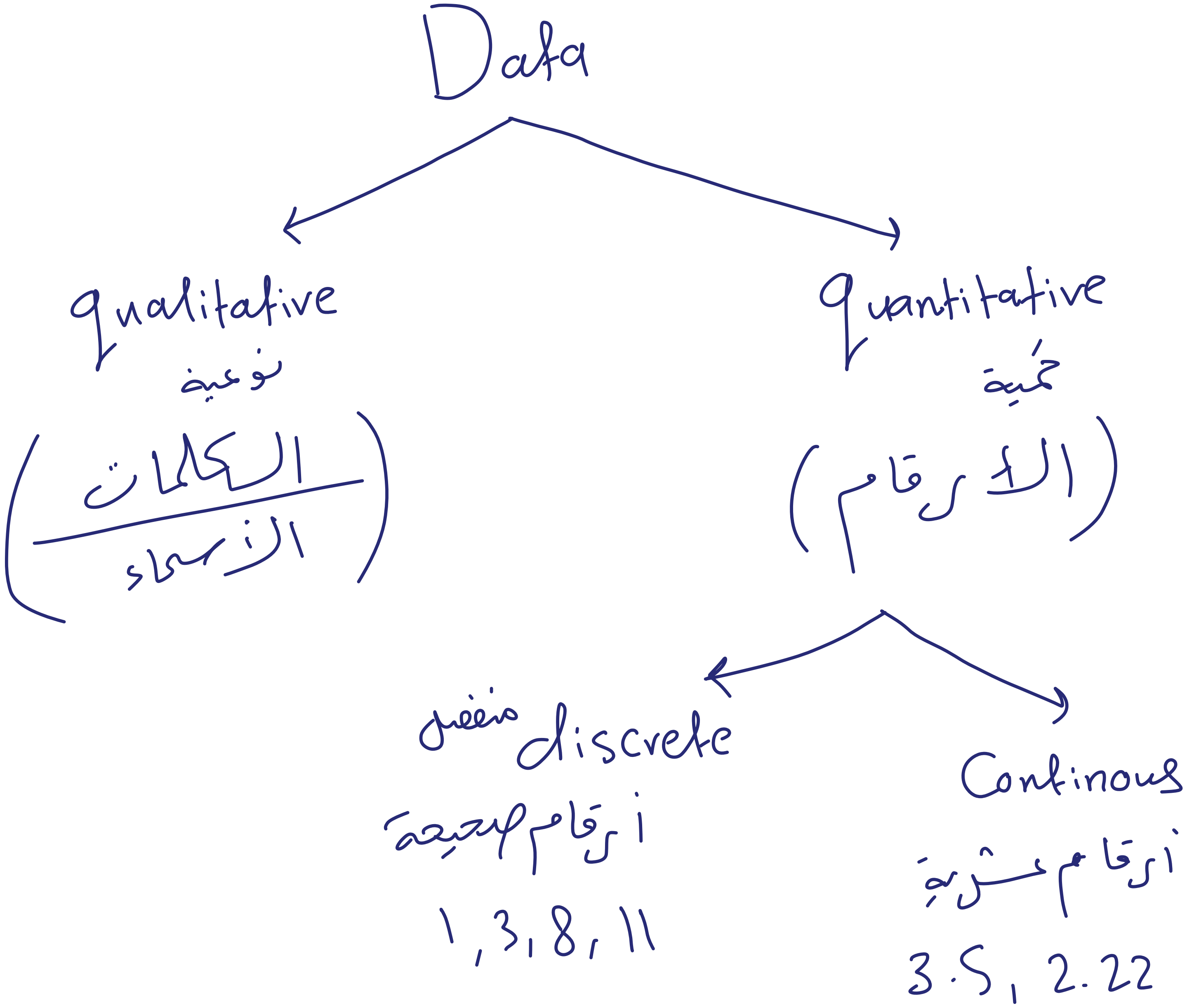
الإحصاء الاستنتاجي

∴

☆ كيف نعرض البيانات ؟

① numeric أرقام

② graphic رسومات


Data
qualitative
نوعية
الكلمات
الأسماء
quantitative
كمية
الأرقام
منفصل discrete
أرقام صحيحة
1, 3, 8, 11
Continous
أرقام عشرية
3.5, 2.22

Chapter 2

Descriptive statistics

* Measures of central tendency

مقاييس النزعة المركزية

Mean Median Mode

3 M's

① Mean (Arithmetic mean)

$\bar{X} = \frac{\sum X}{n}$ = مجموع البيانات / عددها

∴

Example find the mean for:

(2, 7, 5, 11, 5)

$$\bar{x} = \frac{2+7+5+11+5}{5} = 6$$

Example If the mean of (a,a,7,11,2) is 6, find the ⓐ?

$$\frac{6}{1} = \frac{a+a+7+11+2}{5}$$

$$30 = 2a + 20$$

$$2a = 30 - 20$$

$$2a = 10$$

$$\boxed{a = 5}$$

* * *

Example If the mean of (x,y,12) is 10, find the mean of (x,y)?

$$\frac{10}{1} = \frac{x+y+12}{3}$$

$$30 = x+y+12$$

$$\boxed{x+y = 18}$$

$$\frac{x+y}{2} = \frac{18}{2} = 9$$

Example find the mean of

$(-9, -7, -11, 8, 2)$

$$\bar{X} = \frac{-9 + -7 + -11 + 8 + 2}{5}$$

$$= -3.4$$

NOTE $\bar{X}$ could be negative

$$\frac{\bar{X}}{1} = \frac{\sum X}{n}$$

$$\boxed{\sum X = \bar{X} \cdot n}$$

عندما يحدث تغيير على مجموع قيم X

طالب يدخل على الصف / يطلع

معلم يغير صنف / يشيل

Example If the mean mark of 9 students is 15. Ahmad with mark 20 joined the class, find the new mean ?

الحل

$n = 9$

$\bar{X} = 15$

$X = 20$

Ahmad

$n = 10$

$\bar{X} = ?$

$$\sum X_{\text{قديم}} = \bar{X} \cdot n$$

$$= 15 * 9 = 135$$

$$\sum X_{\text{جديد}} = 135 + 20 = 155$$

$$\bar{X}_{new} = \frac{\sum X_{new}}{n_{new}}$$

$$\bar{X}_{new} = \frac{\sum X_{new}}{10}$$

$$= \frac{155}{10}$$

$$\boxed{\bar{X}_{new} = 15.5}$$

Example If the mean mark of 10 Boys is 12, and the mean mark of 12 Girls is 10 find the mean mark of students altogether?

الحل

$n = 10$

$\bar{X} = 12$

Boys

$$\sum X_B = \bar{X} \cdot n$$

$$= 12 * 10$$

$$= 120$$

$n = 12$

$\bar{X} = 10$

Girls

$$\sum X_G = \bar{X} \cdot n$$

$$= 10 * 12$$

$$= 120$$

$$\bar{X}_{total} = \frac{\Sigma X_B + \Sigma X_G}{22} = \frac{120+120}{22}$$

$$= 10.9$$

* * *

② The mode " المنوال "

⇐ العنصر الأكثر تكرارا

Example find the mode for:

1) (2, 7, 5, 11, 5)

⇒ The mode is 5

2) (2, 7, 5, 11, 5, 2)

The mode is 5, 2

3) (2, 7, 5, 11, 5, 2, 5)

The mode is 5

NOTE

1 mode ⇒ unimodal

2 modes ⇒ Bimodal

3 modes ⇒ Trimodal

③ The median (Q_2)

① الترتيب

② نعد البيانات

زوجي

← نأخذ القيمتان بالمنتصف

2

فردي

← نأخذ القيمة التي بالمنتصف

Example ~~5~~, ~~5~~, ~~8~~, ~~10~~, 12, ~~12~~, ~~13~~, ~~13~~, ~~15~~

find the median ?

⇒ The median is 12

Example 2, 9, 11, 5, 6, 27

find Q_2 ?

الترتيب ~~2~~, ~~5~~, $\underline{6, 9}$, ~~11~~, ~~27~~

The median is $\frac{6+9}{2} = \boxed{7.5}$

Example 2, 6, 16, 9, 3, 8, 11

الترتيب ~~2~~, ~~3~~, ~~6~~, $\boxed{8}$, ~~9~~, ~~11~~, ~~16~~

The median is 8

Example $\cancel{5}, \cancel{4}, \cancel{7}, \underline{9, 10}, \cancel{16}, \cancel{12}, \cancel{14}$

The median is $\dfrac{9+10}{2}$

$$= \boxed{9.5}$$

Example For the following ORDERED DATA $(\cancel{a}, \cancel{2}, \cancel{3}, 5, b, \cancel{9}, \cancel{10}, \cancel{c})$. IF the mean is 7, the median is 6 and the mode is 2 Find the a,b,c?

$\bar{x} = 7$

$Q_2 = \underline{\underline{6}}$

mode $= 2$

from mode $\Rightarrow \boxed{a = 2}$

from median $\Rightarrow \dfrac{5+b}{2} = \dfrac{6}{1}$

$$\boxed{b = 7}$$

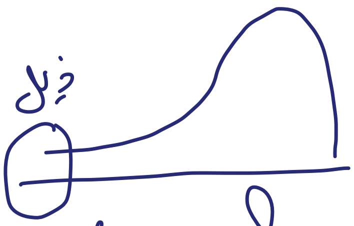
from mean $\Rightarrow$

$$7 = \frac{2+2+3+5+7+9+10+C}{8}$$

$$\boxed{C = 18}$$

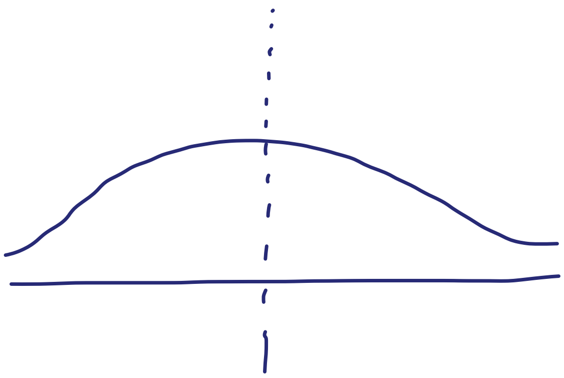
Skewness

التواء



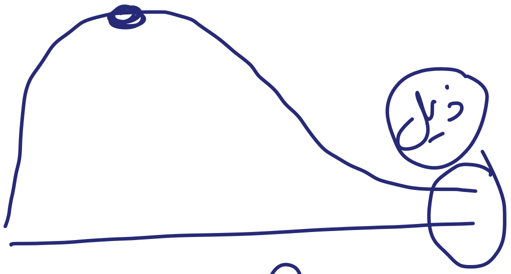
Skewed to Left

$\bar{x} < Q_2 < \text{mode}$



Symmetric

$\text{mode} = Q_2 = \bar{x}$



Skewed to Right

$\text{mode} < Q_2 < \bar{x}$

Example Determine the shape of the following:

1) $\text{mode} = 16 \quad , \quad \bar{x} = 14$

$\text{mode} > \bar{x}$ skewed to left

2) $\text{mode} = 7630 \quad , \quad \bar{x} = 7630$

mode = mean Symmetric

3) mode = 2 , $\bar{x} = 4$

mean > mode   skewed to Right

Characteristics of mean

① mean is affected by an outlier

outlier   القيم الشاذة

2, 2, 2, 2, 20   outlier

2, 2, 2, 2 $\boxed{\bar{x} = 2}$

outlier   2, 19, 19, 18, 20

2, 2, 2, 2, 20

$$\bar{x} = \frac{2+2+2+2+20}{5}$$

$$= \boxed{5.6}$$

2) 2, 2, 2, 2 $\bar{x} = 2$

③   5, 5, 5, 5 $\bar{y} = 5$ (+3)

$$\boxed{\bar{y} = \bar{x} + b}$$

③

$$\begin{tikzcd} 2, 2, 2, 2 \arrow[d, "*3"'] & \bar{x} = 2 \arrow[d, "*3"] \\ 6, 6, 6, 6 & \bar{y} = 6 \end{tikzcd}$$

* Measures of variability (Spread)

مقاييس التشتت

مستحيل يكونوا سوالب

① The Range = Max - Min

المدى

Example find the range for:

1) (2, 7, 5, $\underline{11}$, ②)

The Range is = 11 - 2 = $\boxed{9}$

② The variance and standard deviation

والانحراف المعياري

① $$S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

② $$S^2 = \frac{\Sigma x^2}{n-1} - \frac{(\Sigma(x))^2}{n(n-1)}$$

$$S = \sqrt{\text{variance}} = \sqrt{\frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}}$$

$$\Sigma x^2 \quad Vs \quad (\Sigma(x))^2$$

نربع القيم ثم نجمع

2, 2, 2, 2

4, 4, 4, 4 $\Sigma x^2 = 16$

نجمع القيم ثم نربع

$\underline{2, 2, 2, 2}$

$(\Sigma x)^2 \quad (8)^2 = \underline{\underline{64}}$

Example find variance for $\underline{(2, 7, 5, 11, 5)}$

$$\bar{x} = \frac{2+7+5+11+5}{5} = 6$$

طريقة (1)

$$S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

x	2	7	5	11	5
$(x-\bar{x})$	-4	1	-1	5	-1
$(x-\bar{x})^2$	16	1	1	25	1

Always $\Sigma(x-\bar{x}) = 0$

$\Sigma(x-\bar{x})^2 = 44$

$$S^2 = \frac{44}{5-1} = \frac{44}{4} = 11$$

$$S = \sqrt{11} = 3.31$$

طريقة (2)

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

X	2	7	5	11	5
X^2	4	49	25	121	25

$\sum X = 30$

$\sum X^2 = 224$

$$S^2 = \frac{224}{5-1} - \frac{(30)^2}{5(5-1)} = 11$$

$$S = \sqrt{11} = 3.31$$

Example find the variance for (5, 7, 1, 2, 4)

طريقة (1)

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

$\bar{X} = 3.8$

X	5	7	1	2	4
$(X - \bar{X})$	1.2	3.2	-28	-1.8	0.2
$(X - \bar{X})^2$	1.44	10.24	7.84	3.24	0.04

$\sum (X - \bar{X})^2 = 22.8$

$$S^2 = \frac{22.8}{5-1} = 5.7$$

$$S = \sqrt{5.7} = 2.38$$

(2) طريقة

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

X	5	7	1	2	4
X^2	25	49	1	4	16

$\sum X = 19$

$\sum X^2 = 95$

$$S^2 = \frac{95}{4} - \frac{(19)^2}{5(4)} = 5.7$$

$$S = \sqrt{5.7} = 2.38$$

✱ Characteristics of variance

① مستحيل يكون سالب

② 2, 2, 2, 2 $\quad S_x^2 = 0$

$\xrightarrow{+3}$ $\xleftarrow{-2}$

5, 5, 5, 5 $\quad S_y^2 = 0$

لا يتأثر بالجمع والطرح

③ 2, 2, 2, 2 $S_x^2 = a$

$\times 3 \downarrow$

6, 6, 6, 6 $S_y^2 = C^2 * S_x^2$

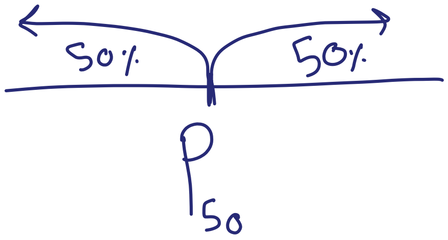
$S_y^2 = 9 * a$

S_x^2 : الvariance القديم

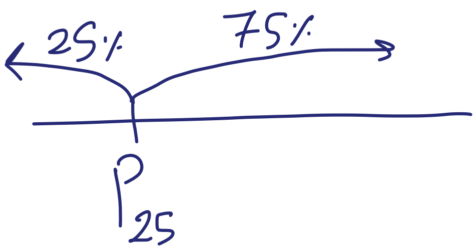
S_y^2 : الvariance الجديد

☆ Quartiles and percentiles

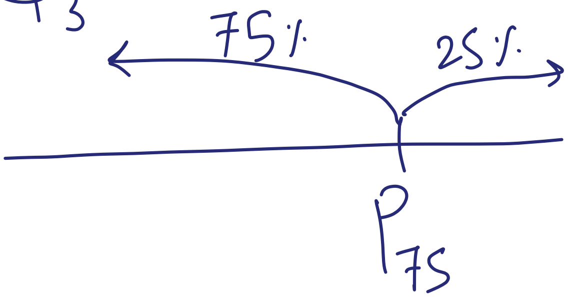
Q_2 : median

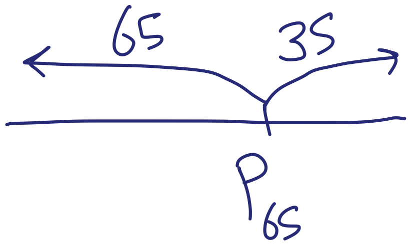


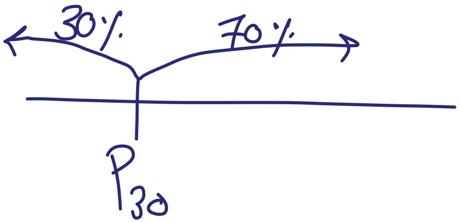
Q_1 : Lower quartile



Q_3 : upper quartile







Percentile مسان

$$P_p = \frac{n.P}{100}$$

n: sample size

P: percentile

كسر ← → عدد صحيح

→ next integer

$2.2 \Rightarrow (3)^{th}$

$3.9 \Rightarrow (4)^{th}$

$$\frac{\text{العدد} + (\text{العدد} + 1)}{2}$$

$$\frac{(3)^{th} + (4)^{th}}{2}$$

$$\frac{(5)^{th} + (6)^{th}}{2}$$

Inter-quartile Range (IQR)

$$\boxed{IQR = Q_3 - Q_1}$$

Example 2, 7, 5, 11, 3, 8, 10 Find IQR:

الترتيب → 2, 3, 5, 7, 8, 10, 11

$$Q_1 = P_{25} = \frac{n.P}{100} = \frac{7 * 25}{100} = \frac{175}{100} = (1.75)^{th}$$

$$= (2)^{th}$$

$$\boxed{Q_1 = 3}$$

$$Q_3 = P_{75} = \frac{7 * 75}{100} = \frac{525}{100} = (5.25)^{th}$$

$$= (6)^{th}$$

$$Q_3 = 10$$

$$IQR = 10 - 3 = \boxed{7}$$

Example 260, $\boxed{290}$, $\boxed{300}$, 320, 330, $\boxed{340}$, $\boxed{340}$, 550

find IQR?

$$Q_1 = P_{25} = \frac{8 * 25}{100} = \frac{200}{100} = \frac{(2)^{th} + (3)^{th}}{2}$$

$$\frac{290 + 300}{2} = \boxed{295}$$

$$Q_3 = P_{75} = \frac{8 * 75}{100} = \frac{600}{100} = \frac{(6)^{th} + (7)^{th}}{2}$$

$$\frac{340 + 340}{2} = \boxed{340}$$

$$IQR = 340 - 295 = 45$$

Example 2, 3, 5, 7, 9, 10, 11, 12 find the value that 65% of data lie below it?

$$P_{65} = \frac{8 * 65}{100} = \frac{520}{100} = (5.2)^{th} = (6)^{th}$$

$$\boxed{P_{65} = 10}$$

Example 3, 5, 5, 6, 8, 8, 9, 10, 11, 12

find the value that 30% of data lie above it?

الحل

$$P_{70} = \frac{10 * 70}{100} = \frac{(7)^{th} + (8)^{th}}{2}$$

$$= \frac{9 + 10}{2} = \boxed{9.5}$$

Outliers

القيم الشاذة

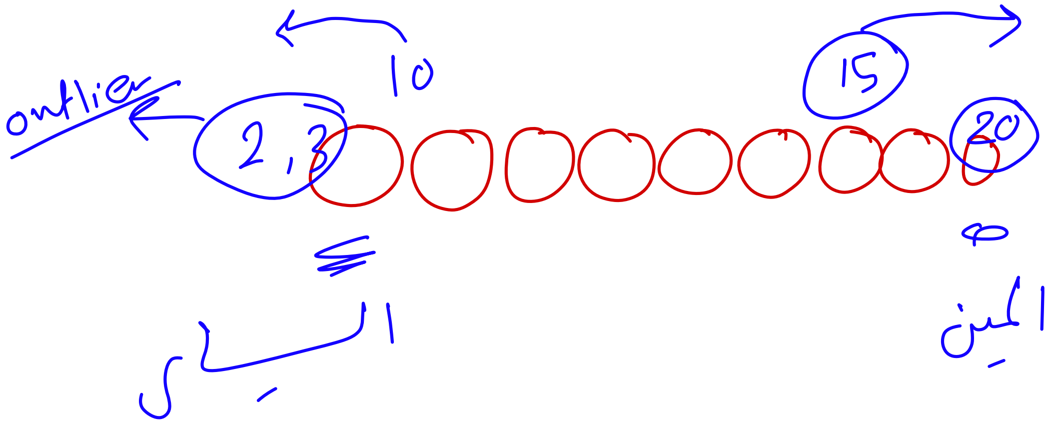
مهم

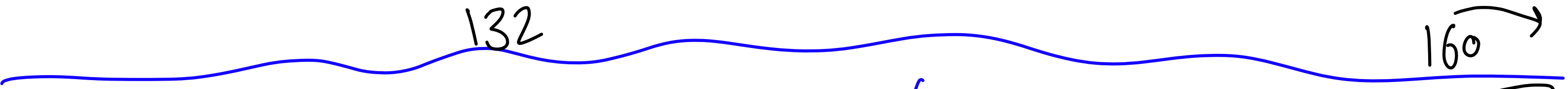
حد اعلى

$$Q_3 + 1.5 * IQR$$

من الشمال

$$Q_1 - 1.5 * IQR$$





Example 123, 140, 145, 146, 147, 149, 150, 172

find an outliers (If there was)
إن وجد !

من اليمين

$$Q_3 + 1.5 * IQR$$

$$149.5 + 1.5 * 7 = 160$$

so 172 is an outlier

من اليسار

$$Q_1 - 1.5 * IQR$$

$$142.5 - 1.5 * 7 = 132$$

so 123 is an outlier

$$Q_1 : P_{25} = \frac{8 * 25}{100} = \frac{(2)^{th} + (3)^{th}}{2}$$

$$= \frac{140 + 145}{}$$

$$2$$

$$= 142.5$$

$$Q_3 : P_{75} = \frac{8 * 75}{100} = \frac{(6)^{th} + (7)^{th}}{2}$$

$$= \frac{149 + 150}{2}$$

$$= 149.5$$

$$IQR = 149.5 - 142.5 = 7$$

Example 340, 300, 520, 340, 320, 290, 260, 330

find outliers (If there was) ?

من اليمين

$$Q_3 + 1.5 * IQR$$

$$340 + 1.5 * 45 = 407.5$$

$\therefore$ 520 is an outlier

من اليسار

$$Q_1 - 1.5 * IQR$$

$$295 - 1.5 * 45 = 227.5$$

X

الترتيب 227.5 407.5

$260, 290, 300, 320, 330, 340, 340, \underline{\boxed{520}}$

$$Q_1 : P_{25} = \frac{8 * 25}{100} = \frac{(2)^{th} + (3)^{th}}{2}$$

$$= \frac{290 + 300}{2}$$

$$= 295$$

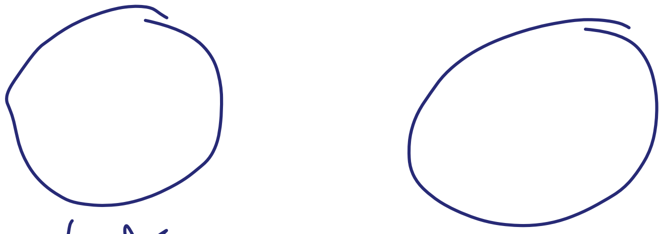
$$Q_3 : P_{75} = \frac{8 * 75}{100} = \frac{(6)^{th} + (7)^{th}}{2}$$

$$= \frac{340 + 340}{2}$$

$$= 340$$

$$IQR = 340 - 295 = 45$$

* Coefficent of variation (CV)



ثمن سيارة

$S = 0$ \$

ثمن خبز

$S = 0$ JD

$$CV = \frac{S}{\bar{X}} * 100\%$$

Example In a class, if the mean is 30 and standard deviation is 2, find the Coefficient of variation for this Class?

~~الحل~~ $CV = \frac{S}{\bar{X}} * 100\% = \frac{2}{30} * 100\%$

$= 6.67\%$

Example If the mean and Coefficient of variation of a data are 15 and 48 Respectively, then find the value of stand. deviation

$\bar{X} = 15$

$CV = 48$

$$CV = \frac{S}{\bar{X}} * 100$$

$$48 = \frac{S}{15} * 100$$

$$S = 48 * 0.15 = 7.2$$

~~حل:~~

Example If $n = 5$, $\bar{X} = 6$, $\sum x^2 = 765$, then find the Coefficient of variation ?

~~حل~~

$$CV = \frac{S}{\bar{X}} * 100\% \Rightarrow \frac{12.09}{6} * 100\%$$

$$= \boxed{201.55\%}$$

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

$$= \frac{765}{5-1} - \frac{(30)^2}{5(4)}$$

$$= 146.25$$

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

$$= 6 * 5$$

$$= 30$$

$$S = \sqrt{146.25}$$

$$= 12.09$$

Example find the Coefficient of variation for $(24, 26, 33, 37, 29, 31)$?

$$CV = \frac{S}{\bar{x}} * 100\%$$

$(x-\bar{x})$	-6	-4	3	7	-1	1
$(x-\bar{x})^2$	36	16	9	49	1	1

$\Sigma(x-\bar{x})^2$

$= 112$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{180}{6} = 30$$

$$S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{112}{6-1} = 22.4$$

$$\boxed{S = 4.73}$$

$$CV = \frac{4.73}{30} * 100\% = 15.77\%$$

Example Two plants C and D of a factory show the following Results about the number of workers and the wages paid to them.

	C	D
no. of workers	5000	6000
Average monthly wages	2500	2500
Standard deviation	9	10

Using Coefficient of variation, find in which plant C or D, is there greater variability in individual wages

~~Sol~~ $$CV_C = \frac{9}{2500} * 100$$

$$= 0.36$$

$$CV_D = \frac{10}{2500} * 100 = 0.40$$

$CV_D > CV_C$ ∴ D has greater variability

CV	SD
Relative measure of variabity	absolute measure of variability
used to campare variability between 2 samples	used to measure the dispersion of data in a single set

Example The mean and SD of marks obtained by 40 students of a class in three subjects maths, science and social science are given

	Mean	SD
Maths	56	12
Science	65	14
Social Science	60	10

which of three subjects shows the heighst variation?

$$CV_{maths} = \frac{12}{56} * 100\% = 21.43$$

$$CV_{sci.} = \frac{14}{65} * 100\% = 21.54$$

$$CV_{\substack{soc\\sci.}} = \frac{10}{60} * 100\% = 16.67$$

∴ ∴

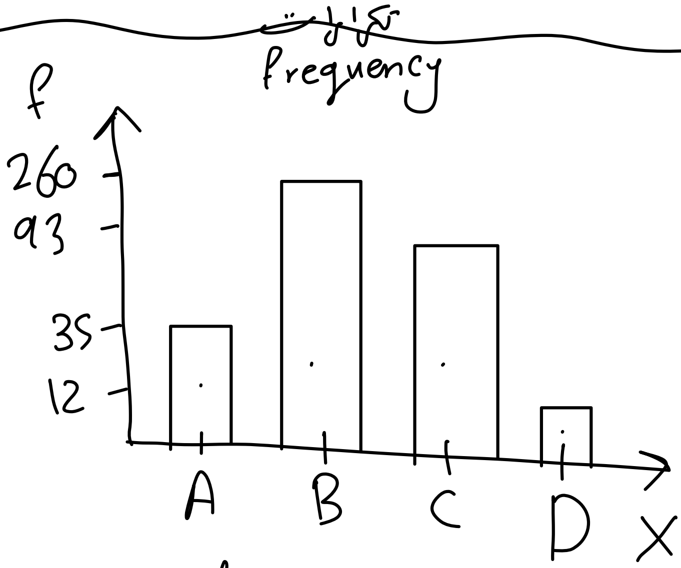
★ Graphical methods

① Bar Graph ② Stem and leaf plot

③ Box plot

① Bar Graph

X	A	B	C	D
f	35	260	93	12



- difficult to construct
- identify sample points is lost

② Stem and leaf plot

الساق والورقة

Example 23 71

58 71

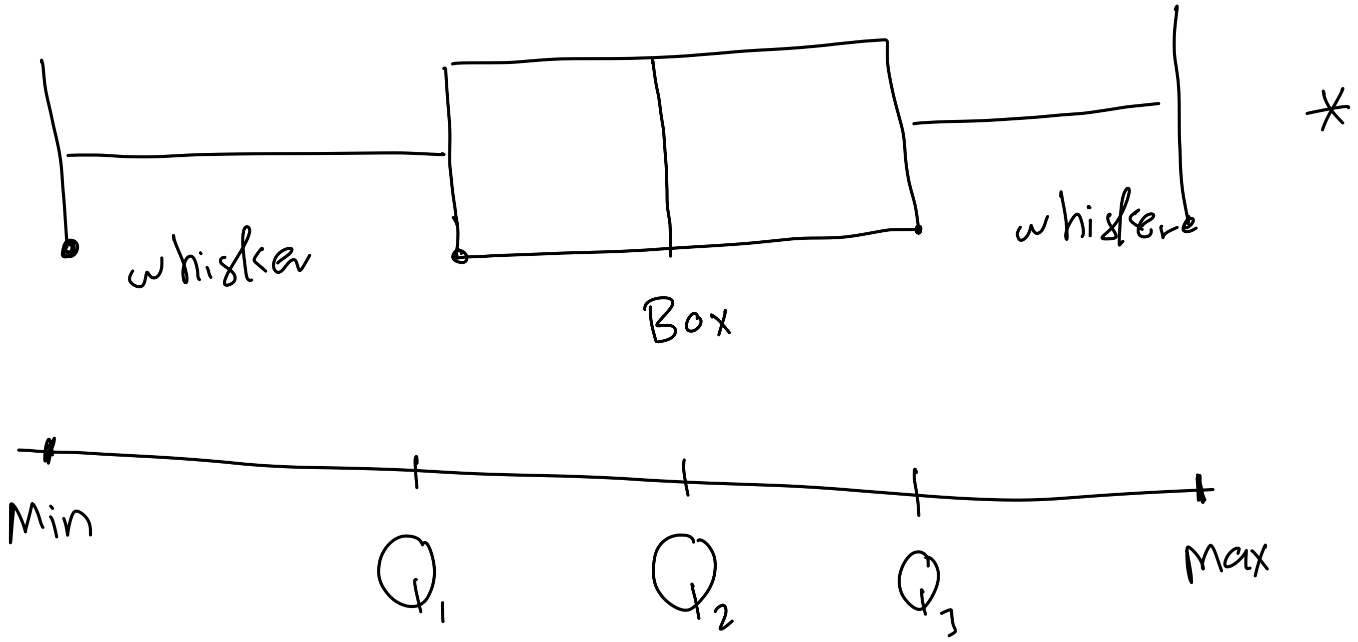
S	L
2	3
3	

62
62
63
65
67
72
80
82
82
82

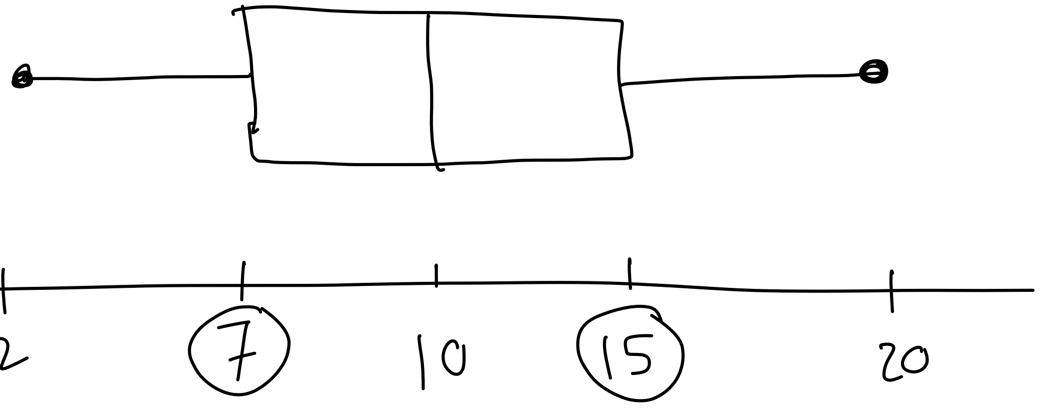
4 |
5 | 8
6 | 2 2 3 5 7
7 | 1 1 2
8 | 0 2 2

---

③ Box Plot "Box and whiskers"



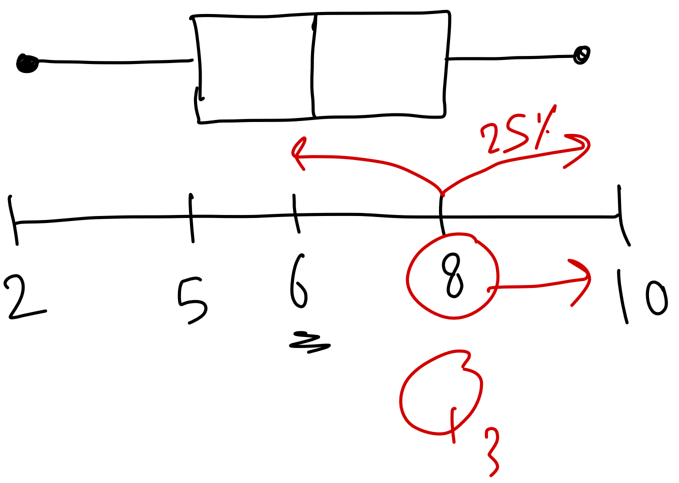
Example



find IQR ?

$$IQR = Q_3 - Q_1 = 15 - 7 = 8$$

Example If there are 32 students in the class, find:



① $Q_2 = 6$

② Range $= 10 - 2 = 8$

③ $IQR = 8 - 5 = 3$

④ The number of students achieved more than 8 ?

$0.25 * 32 = \boxed{8}$

* Central tendency = Location

→ useful to define the center or middle of sample

→ Could be negative

→ Mean (generally not part of data set)

① oversensitive to extreme values

② easy to calculate

③ each sample has only one sample mean

→ Median (maybe part of data set)

→ ① less affected by outliers

② less efficient than mean

→ mode (always part of data set)

Measures of variation (dispersion)
(spread)

① Range

- Simplest MOV
- quick summary of variation
- extremely affected by outlier

② IQR

not affected by outlier

---

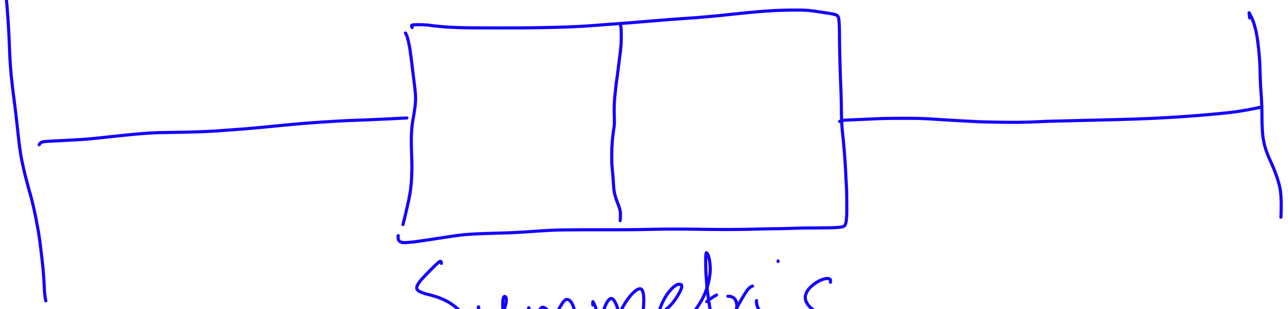
skewed to Right (+ skewness)

skewed to left (− skewness)

Variance
&
SD

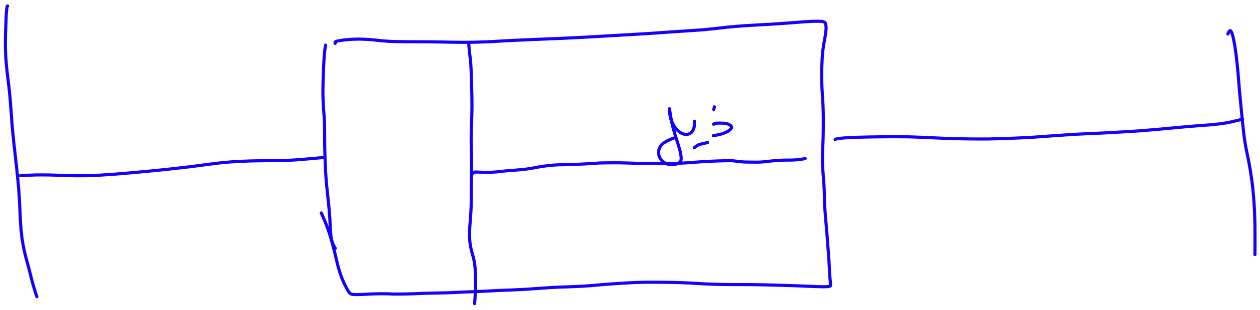
2, 2, 2, 2, 2 متشابهة

$S^2 = 0$



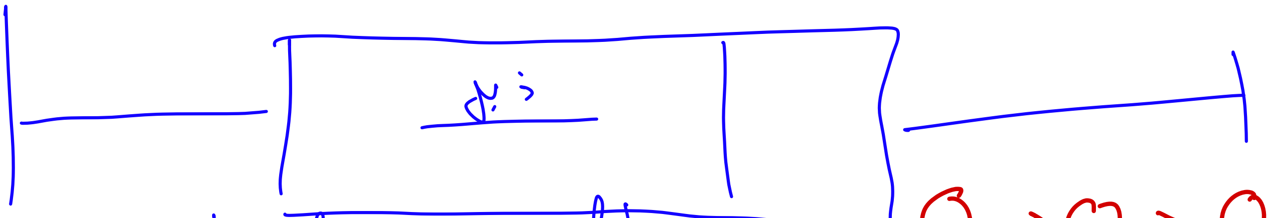
Symmetric

$Q_1 \rightarrow Q_2 = Q_2 \rightarrow Q_3$



skewed to Right

$Q_1 \rightarrow Q_2 < Q_2 \rightarrow Q_3$



skewed to left

$Q_1 \rightarrow Q_2 > Q_2 \rightarrow Q_3$

Chapter 3 probability الاحتمالات

تجارب

★ Experiment

① Deterministic محددة

② Random عشوائية

★ The sample space

(الفضاء العيني)

Ω : defined as the set of all possible outcomes

جميع نتائج التجربة العشوائية

Example Find the sample space for the following :

رمي عملة نقدية

① Tossing a fair coin 1 - time

$\Omega : \{H, T\}$ $\quad 2^1$

② Tossing a fair coin 2-times

$$\Omega : \left\{ \begin{matrix} (H, H), & (H, T) \\ (T, H), & (T, T) \end{matrix} \right\} \quad 2^2$$

③ Tossing a fair coin 3-times

$$\Omega : \left\{ \begin{matrix} (H, H, H), & (H, H, T), & (H, T, H), & (H, T, T) \\ (T, H, H), & (T, H, T), & (T, T, H), & (T, T, T) \end{matrix} \right.$$

$$2^3$$

NOTE when tossing a fair coin k-times, so the number of elements in a sample space is 2^k

1) Throwing a fair dice 1-time
حجر النرد

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

2) Throwing a fair dice 2-times
6^2

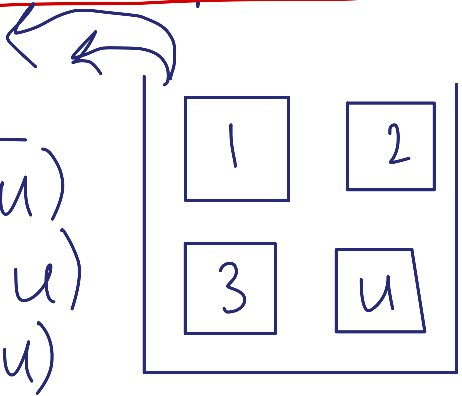
	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

NOTE when throwing a fair dice K times, so the number of elements in a sample space is 6^K

Example 2- cards are drawn from the Box that contained 4- Cards numbered as (1- 4). Find the sample space when:

① The drawn was with replacement

مع إرجاع



	1	2	3	4
1	(1,1)	(1,2)	(1,3)	(1,4)
2	(2,1)	(2,2)	(2,3)	(2,4)
3	(3,1)	(3,2)	(3,3)	(3,4)
4	(4,1)	(4,2)	(4,3)	(4,4)

② The drawn was without Replac.

بدون إرجاع

	1	2	3	4
1	X	(1,2)	(1,3)	(1,4)
2	(2,1)	X	(2,3)	(2,4)
3	(3,1)	(3,2)	X	(3,4)
4	(4,1)	(4,2)	(4,3)	X

NOTE

★ إذا السؤال ما حدد مع إرجاع أو بدون إرجاع

نعتبر "بدون إرجاع"

* * *

★ The Event الحادث

★ أي مجموعة من عناصر الفضاء العيني

① Simple event : Consist of 1 element of sample space
الحادث البسيط

② Comosite "combined" event : Consist of more than 1 element of sample space
الحادث المركب

③ Certain event : Consist of all elements of sample space
الحادث الأكيد

④ Impossible event : Consist of no elements of sample space
الحدث المستحيل

Example Throwing a dice 1-time, define

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

A : {getting a number divisible by 5}

A : {5} ⟶ Simple event

B : {getting a prime number}
عدد أولي

عدد أولي : يقبل القسمة على نفسه و على الواحد فقط
عدا رقم (١)

B : {2, 3, 5} ⟶ Composite event

C : getting a number less than 7

C : {1, 2, 3, 4, 5, 6} ⟶ Certain event

D: getting a number more than 6

D: $\{ \} = \phi \longrightarrow$ impossible event

* * *

* The probability of events

احتمال الحوادث

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

* * *

Example In previous question:

① $p(A) = \frac{1}{6}$

③ $P(C) = \frac{6}{6} = 1$

② $p(B) = \frac{3}{6} = \frac{1}{2}$

④ $p(D) = \frac{0}{6} = 0$

NOTES

~~مهم~~

① Certain events probability = 1

② impossible events probability = 0

③ $P(\Omega) = 1$

④ $P(\phi) = 0$

مهم ⑤ $0 \leq P(A) \leq 1$

Rules of probability

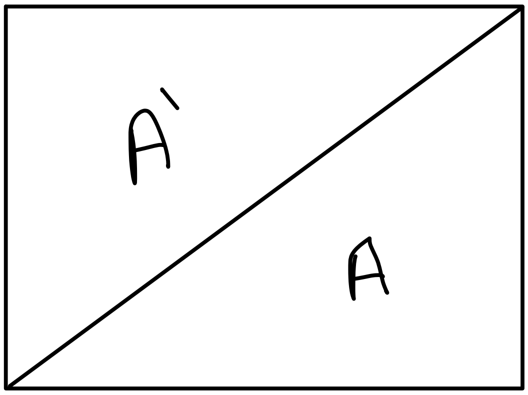
قوانين الاحتمالات

A: الاحتمال الأول

B: الاحتمال الثاني

intersection	$\cap$	تقاطع
union	$\cup$	اتحاد

$= 1$

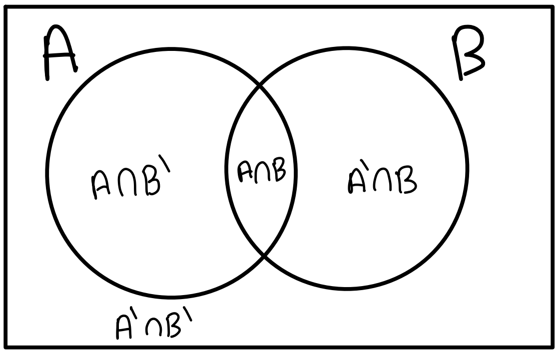


$$P(A) + \underbrace{P(A`)}_{} = 1$$

① $P(A') = 1 - P(A)$

② $P(A \cap B') = P(A) - P(A \cap B)$

$P(A' \cap B) = P(B) - P(A \cap B)$



③ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

" Addition Rule "

④ $P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B)$

$P(A' \cup B') = P(A \cap B)' = 1 - P(A \cap B)$

" De morgan's Law "

⑤ A and B are mutually exclusive

" disjoint "

if : ① $P(A \cap B) = 0$

and ② $P(A \cup B) = P(A) + P(B)$

⑥ A and B are independent if

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

NOTE A & B & C independent

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

" multiplication Rule "

⑦ $$P(A \mid B) = \frac{P(A\cap B)}{P(B)}$$

given

if

"Conditional probability"

NOTES

① $P(A|B)$ and A & B are independen

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

$$P(A|B) \equiv P(A)$$

$$P(A|B') = P(A)$$

② $P(A\cap B) \equiv P(B\cap A)$ A and B

③ $P(A\cup B) \equiv P(B\cup A)$ A or B

Example If $P(A) = 0.8$, $P(B) = 0.7$
$P(A \cap B) = 0.6$ find:

i) $P(A') = 1 - P(A) = 1 - 0.8 = 0.2$

ii) $P(B') = 1 - P(B) = 1 - 0.7 = 0.3$

iii) $P(A \cap B') = P(A) - P(A \cap B)$
$= 0.8 - 0.6 = 0.2$

iv) $P(A' \cap B) = P(B) - P(A \cap B)$
$= 0.7 - 0.6 = 0.1$

v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0.8 + 0.7 - 0.6 = 0.9$

vi) $P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B)$
$= 1 - 0.9$
$= 0.1$

vii) $P(A' \cup B') = P(A \cap B)' = 1 - P(A \cap B)$
$= 1 - 0.6 = 0.4$

[illegible]

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

$= 0.2 + 0.7 - 0.1 = 0.8$

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

$= 0.8 + 0.3 - 0.2 = 0.9$

x) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.6}{0.7} = \frac{6}{7}$

xi) $P(A'|B) = \frac{P(A' \cap B)}{P(B)} = \frac{0.1}{0.7} = \frac{1}{7}$

xii) $P(A|B') = \frac{P(A \cap B')}{P(B')} = \frac{0.2}{0.3} = \frac{2}{3}$

xiii) $P(A'|B') = \frac{P(A' \cap B')}{P(B')} = \frac{0.1}{0.3} = \frac{1}{3}$

xiv) $P(A|B)' = 1 - P(A|B) = 1 - \frac{6}{7}$

$= \frac{1}{7}$

NOTE $P(A|B)' = P(A'|B)$

Example If $P(A) = 0.6$, $P(B) = 0.5$ and $P(A \cup B) = 0.8$ are A & B mutually exclusive? independent? neither?

Mutually exclusive

$$P(A \cup B) \overset{?}{=} P(A) + P(B)$$

$$0.8 \overset{?}{=} 0.6 + 0.5$$

$$0.8 \overset{?}{=} 1.1$$

(Not ME)

independent

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

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

$$0.8 = 0.5 + 0.6 - P(A \cap B)$$

$$0.8 = 1.1 - P(A \cap B)$$

$$P(A \cap B) = 1.1 - 0.8$$

$$\boxed{P(A \cap B) = 0.3}$$

$$0.3 \overset{?}{=} 0.5 * 0.6$$

$$0.3 = 0.3 \quad ✓$$

So A & B are independent events

Example If A and B are independent events such that $P(A) = 2 * P(B)$ and $P(A \cup B) = 0.8$ then find $P(A)$?

حل

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

$$P(A) = 2 * P(B)$$

$$P(A \cup B) = 0.8$$

$$0.8 = P(A) + P(B) - P(A \cap B)$$

$$0.8 = P(A) + P(B) - P(A) \cdot P(B)$$

$$0.8 = 2 * P(B) + P(B) - 2P(B) * P(B)$$

$$\boxed{X = P(B)}$$

$$0.8 = 2X + X - 2X^2$$

$$2X^2 - 3X + 0.8 = 0$$

$X = 1.53$ ✗ $\quad \boxed{X = 0.34}$ ✓ $\quad \boxed{P(B) = 0.34}$

$$P(A) = 2 * 0.34$$

$$= 0.68$$

* Relative Risk (RR)

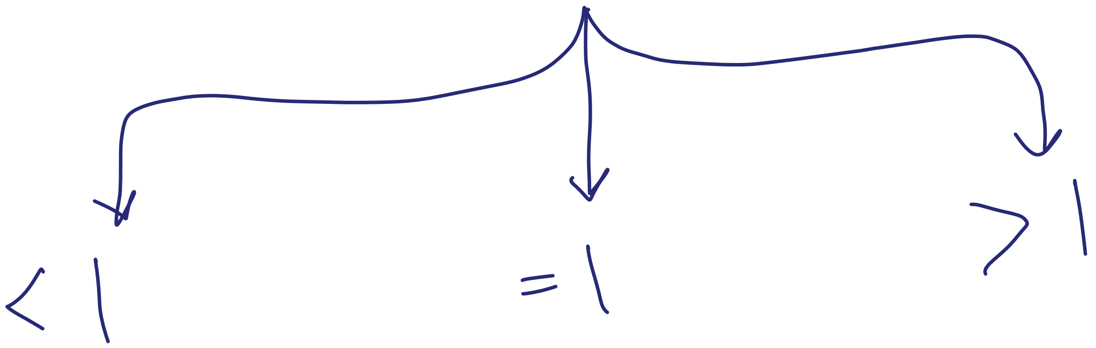
خطر

Risk factor
عامل خطر

"Risk ratio"

متعرض

$$RR = \frac{P(B|A)}{P(B|A')} = \frac{\text{Risk for exposure group}}{\text{Risk for unexposure group}}$$



* Some interpretations

⇒ RR = 1.5 exposure group risk is higher than unexposure groupe in about 0.50

⇒ RR = 3 exposure group risk is higher than unexposure groupe in about 3 times

⇒ RR = 0.8 exposure group risk is less than unexposure groupe in about 0.20

NOTE If A and B are independent then the RR is 1

ضروري

$$RR = \frac{P(B|A)}{P(B|A')} = \frac{\cancel{P(B)}}{\cancel{P(B)}} = \textcircled{1}$$

* * *

Example A study enrolls a 100 smokers and 100 non-smokers. They are followed for next years for developing lung CA. 30 of smokers and 10 of non-smokers developed lung CA. Calculate the RR?

	Smokers	non-Smokers
	100	100
CA	30	10

$$RR = \frac{P(B|A)}{P(B|A')} = \frac{30/100}{10/100} = \textcircled{3}$$

Example If 1 in 10 people exposed to a substance gets sick. If 1 in 100 people who are not exposed get sick. find the RR?

	تعرضوا	لم يتعرضوا
الكل	10	100
Sick	1	1

$$RR = \frac{1/10}{1/100} = 10$$

Example suppose we want to know if exercise affects the risk of developing some disease we collect data and find that 28% of people who exercise regularly develop this disease while 50% of people who do not exercise regularly develop this disease, find the RR?

الحل $RR = \frac{0.28}{0.50} = 0.56$

* * *

Example Suppose we want to know if some new studying program affects the ability of students to pass a particular exam. we collect a data and find that 40% of students who use the new studying program pass the exam while 40% of students who do not use the studying program also pass the exam, calculate the RR?

الحل $RR = \frac{0.40}{0.40} = 1$

* * *

Example suppose 50 basketball players use a new training program, and 50 players use an old training program. At the end of the program we test each player to see if they pass a certain skills test. Find RR?

	passed	failed	
New program	34	16	= 50
old program	39	11	= 50

(1) $$RR = \frac{34/50}{39/50} = 0.872$$

B: pass

A: new program

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

$$= \frac{\frac{34}{34+16}}{\frac{39}{39+11}} = \frac{34/50}{39/50}$$

$$= 0\ 872$$

Example suppose that among 100.000 women with negative mammograms 20 will be diagnosed with breast CA within years, whereas 1 women in 10 with positive mammograms will be diagnosed by breast CA within years find RR ?

$P(B/A') = 0.002$

$P(B/A) = 0.1$

حل

	(+)	(−)
	10	100.000
Breast CA	1	20

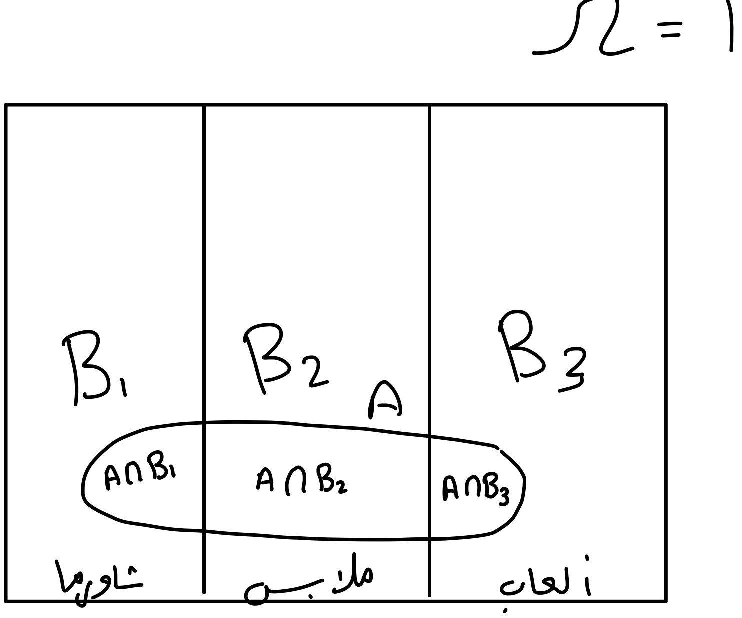
$$RR = \frac{1/10}{20/100.000} = 500$$

* Total probabity

* mutually exclusive events

A & B

⇒ حدوث A ⇒ عدم حدوث B

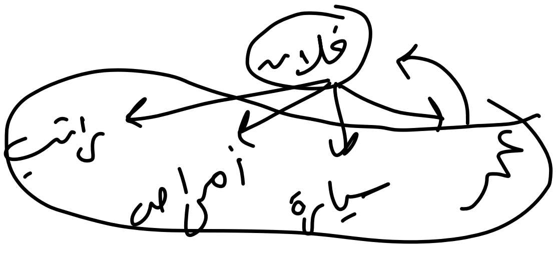


$$P(A) = P(A\cap B_1) + P(A\cap B_2) + P(A\cap B_3)$$

$$\frac{P(A|B_1)}{1} = \frac{P(A\cap B_1)}{P(B_1)} \Rightarrow \boxed{P(A\cap B_1) = P(A|B_1) * P(B_1)}$$

$$P(A) = P(A|B_1) * P(B_1) + P(A|B_2) * P(B_2) + P(A|B_3) * P(B_3)$$

$$\boxed{P(A) = \sum P(A|B_n) * P(B_n)}$$

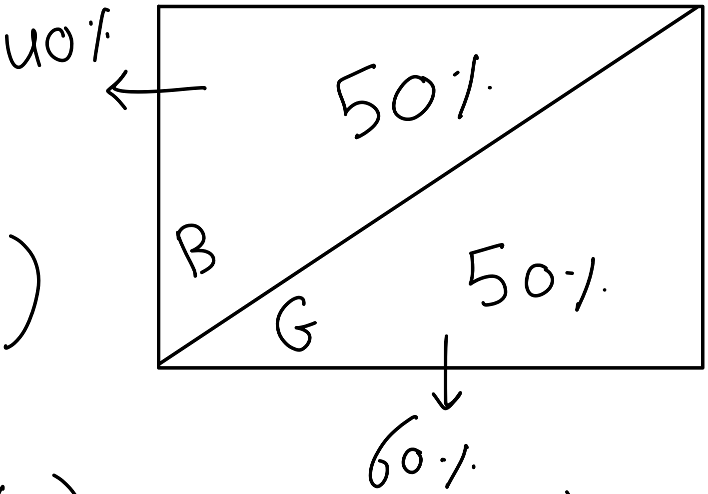


⇒ insurance

⇒ disease

Example If 40% of Boys opted for maths and 60% of Girls opted for maths, then what is the probability that math is chosen if half of class's population is girls?

$$P(\text{maths}) = P(\text{maths} \cap B) + P(\text{maths} \cap G)$$



$$= P(\text{math} \mid B) * P(B) + P(\text{math} \mid G) * P(G)$$

$$= 0.40 * 0.50 + 0.60 * 0.50$$

$$= 0.50$$

* * *

Example Company A produces 10% defective products, Company B produces 20% defective products and Company C produces 5% defectiv products. If choosing a company is an equally likely event, find the probability that the product chosen is defective?

الحل

A	B	C
10% D	20% D	5% D
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$P(\text{defective}) =$$

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

$$= P(D|A) * P(A) + P(D|B) * P(B) + P(D|C) * P(C)$$

$$= 0.10 * \frac{1}{3} + 0.20 * \frac{1}{3} + 0.05 * \frac{1}{3}$$

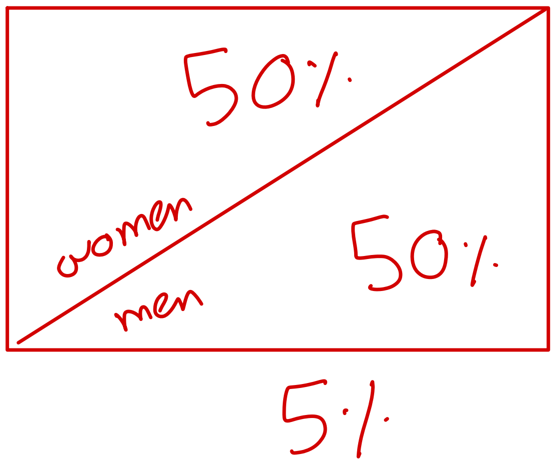
$$= 0.12$$

Example Suppose 5 men out of 100 and 10 women out of 250 are color Blind (عمى الوان), then find the total probability of color blind people?

(Assume that both men and women are equally in number)

$\frac{10}{250}$

$$P(CB) = P(CB \cap W) + P(CB \cap m)$$



$$= P(CB|W) * P(W) + P(CB|m) * P(m)$$

$$= \frac{10}{250} * 0.50 + \frac{5}{100} * 0.50$$

$$= 0.045$$

Example we are planning a 5 year study of cataract (المي البيضا/الزرق) in a population of 5000 people 60 years old and older. we know that :

A_1 : {ages 60 − 64} A_2 : {Ages 65 − 69}

A_3: { ages 70 - 74 } A_4: { Ages 75 + }

what is the probability of event B which is defined as the probability of developing cataract in the next 5 years, given:

$P(A_1) = 0.45$ $P(B|A_1) = 0.024$

$P(A_2) = 0.28$ $P(B|A_2) = 0.046$

$P(A_3) = 0.20$ $P(B|A_3) = 0.088$

$P(A_4) = 0.07$ $P(B|A_4) = 0.153$

الحل

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + P(B \cap A_4)$$

$$P(B) = P(B|A_1) * P(A_1) + P(B|A_2) * P(A_2) + P(B|A_3) * P(A_3) + P(B|A_4) * P(A_4)$$

$$= 0.024 * 0.45 + 0.046 * 0.28 + 0.088 * 0.2 + 0.153 * 0.07$$

$= 0.05$

∴ ∴

Example Suppose 2 doctors A & B, test all patients coming into clinic for syphilis. Let us define the following 2 events:

A+ (doctor A makes a positive diagnosis)

B+ (doctor B makes a positive diagnosis)

$P(A+) = 0.10$, $P(B+) = 0.17$, $P(A+ \cap B+) = 0.08$

Answer the following:

a) find the conditional probability that doctor B makes a positive diagnosis givin that doctor A makes a positive diagnosis?

$$P(B+ \mid A+) = \frac{P(B+ \cap A+)}{P(A+)} = \frac{0.08}{0.10}$$

$$= 0.8$$

اول

B) what is the conditional probability that doctor B makes a positive diagnosis given that doctor A makes a negative diagnosis?

$$P(B+ | A'+) = \frac{P(B+ \cap A'+)}{P(A'+)} = \frac{0.17 - 0.08}{0.90} = 0.1$$

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

c) what is the RR of B+ given A+?

$$RR = \frac{P(B+ | A+)}{P(B+ | A'+)} = \frac{0.8}{0.1} = 8$$

* * *

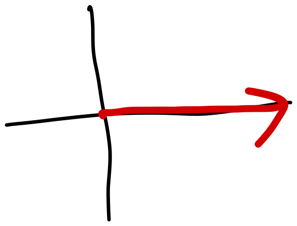
✱ Baye's Rule and screening tests

الحقيقة

		(+)	(−)
Test	(+)	True positive	false positive
	(−)	false negative	True negative

① positive predictive value

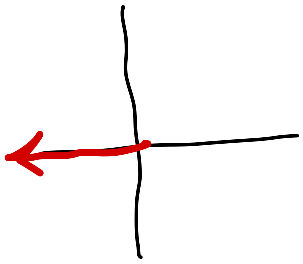
$$PPV \equiv PV+ = \frac{TP}{TP+FP}$$



$$= P\left(\text{disease} \,\middle|\, +\right)$$

② Negative predictive value

$$NPV \equiv PV(-) = \frac{TN}{TN+FN}$$



$$= P\left(\text{no disease} \,\middle|\, (-)\right)$$

③ sensitivity $= \dfrac{TP}{FN+TP}$

$$= P\left((+) \,\middle|\, \text{disease}\right)$$

④ specifity $= \dfrac{TN}{TN+FP}$

$$P((-) \mid \text{no disease})$$

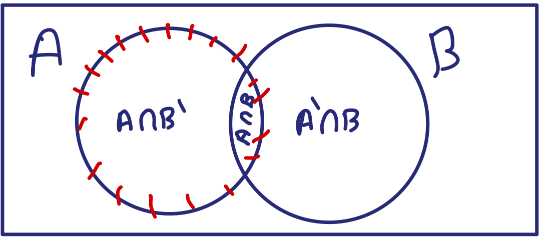
* Baye's Rule

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

$$\boxed{P(B\cap A) = P(A\cap B)}$$

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

$$\boxed{P(B|A) = \frac{P(A|B) * P(B)}{P(A|B) * P(B) + P(A|B') * P(B')}}$$



Example The table below shows the results from looking at the diagnostic accuracy of a new rapid test for HIV in 100.000

subjects, The rows of the table represent the test Result and the columns the true disease status

	HIV(+)	HIV(−)	Total
Test (+)	378	397	775
Test (−)	2	98823	98825
Total	380	99220	100000

① find PV(+) (PPV) ?

$$PPV = P\left(disease \,\middle|\, (+) \right) = \frac{P(disease \cap (+))}{P(+)}$$

$$= \frac{378}{775}$$

② find NPV (PV(−)) ?

$$NPV = P(\text{no disease} \mid (-)) = \frac{P(\text{no} \cap -)}{P(-)}$$

$$= \frac{98823}{98825}$$

③ find Sensitivity ?

$$P((+) \mid \text{disease}) = \frac{P(+ \cap \text{disease})}{P(\text{disease})} = \frac{378}{380}$$

④ find the specifity ?

$$P((-) \mid \text{no disease}) = \frac{P(- \cap \text{no})}{P(\text{no})} = \frac{98823}{99220}$$

Example In a hospital, If you know that probability to have certain disease is 0.15 and the probability to have a positive test if you have the disease is 0.95 and a negative test if you don't have the disease is 0.93, and a positive test if you don't have the disease is 0.7

find PV(+) ?

$$P(B') = 1 - P(B)$$
$$= 1 - 0.15$$
$$= 0.85$$

$$P(D) = 0.15$$

$P(A|B)$ $\quad P(\overset{A}{+}|\overset{B}{D}) = 0.95 \longrightarrow$ sensitivity

$P(A'|B')$ $\quad P(-|D') = 0.93 \longrightarrow$ specifity

$P(A|B')$ $\quad P(+|D') = 0.7 \longrightarrow$ false positive

A: (+)
B: disease

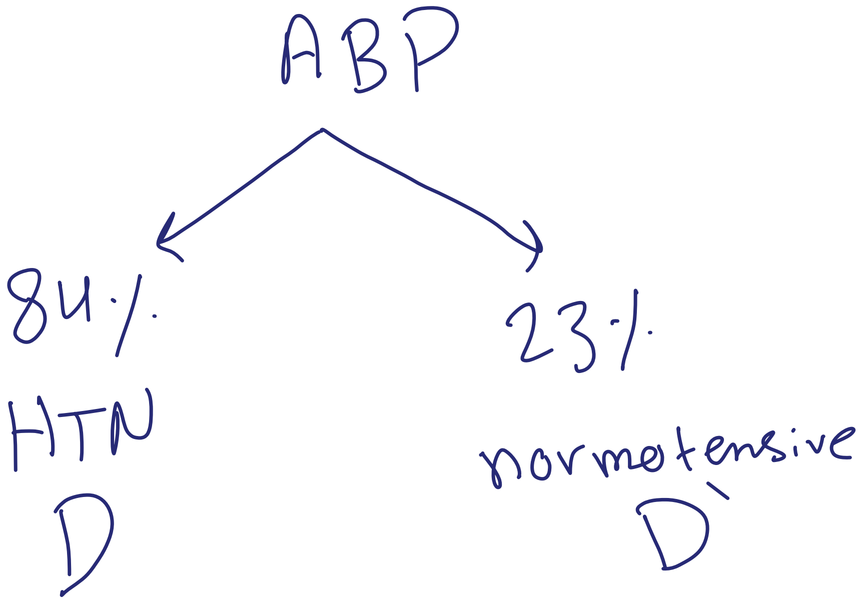
$$PV(+) = P(D|+)$$

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A|B) * P(B) + P(A|B') * P(B')}$$

$$= \frac{0.95 * 0.15}{0.95 * 0.15 + 0.7 * 0.85}$$

Example Suppose 84% of hypertensive and 23% of normotensive are classified by Automated blood pressure machine as hypertensive what are the (PV+) and (PV-) of the machine if you know that 20% of the adult population have the disease?

حل



$P(+|D`)$

$P(\overset{A}{+}|\overset{B}{D}) = 0.84$ → sensitivity

$P(\overset{A}{+}|\overset{B'}{D`}) = 0.23$

A: (+)
B: D

$P(A'|B') = 1 - P(A|B') = 1 - 0.23 = \boxed{0.77}$

$P(\underset{B}{D}) = 0.20$

$\boxed{\underline{\text{NOTE}} = P(A'|B') = 1 - P(A|B')}$

$$PV(+) = P(B|A) = \frac{P(A|B)*P(B)}{P(A|B)*P(B)+P(A|B')*P(B')}$$

$$= \frac{0.84*0.20}{0.84*0.20+0.23*0.80}$$

$$= 0.48$$

$$PV(-) = P(B'|A') = \frac{P(A'|B')*P(B')}{P(A'|B')*P(B')+P(A'|B)*P(B)}$$

$$= \frac{0.77*0.80}{0.77*0.80+0.16*0.20}$$

$$= 0.95$$

NOTE $P(A'|B) = P(A|B)' = 1 - P(A|B)$

~~Q/~~

Example A man is known to speak the truth 2 out of 3 times. He throws a dice and reports that the number obtained is 4. Find probability that actually the number obtained is 4?

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

$P(4) = \frac{1}{6}$

الحل

$$P\left(\overset{A}{\underset{4}{\text{صدق}}} \middle| \overset{B}{4}\right) = \frac{2}{3}$$

$$P\left(\overset{A}{\underset{4}{\text{صدق}}} \middle| \overset{B'}{4'}\right) = \frac{1}{3}$$

$$P(4) = \frac{1}{6}$$

$$P\left(\overset{B}{4} \middle| \overset{A}{\underset{4}{\text{صدق}}}\right) = \frac{P(A|B) * P(B)}{P(A|B) * P(B) + P(A|B') * P(B')}$$

$$= \frac{\frac{2}{3} * \frac{1}{6}}{\frac{2}{3} * \frac{1}{6} + \frac{1}{3} * \frac{5}{6}} = \frac{2}{7}$$

Example suppose 50 men out of 1000 men and 25 women out of 1000 are orators (خطباء). An orator is chosen at random. Find the probability that a male person is selected?

مثال 1

$$P(\overset{\textcircled{A}}{\text{orator}} \mid \overset{\textcircled{B'}}{\text{man}}) = \frac{50}{1000}$$

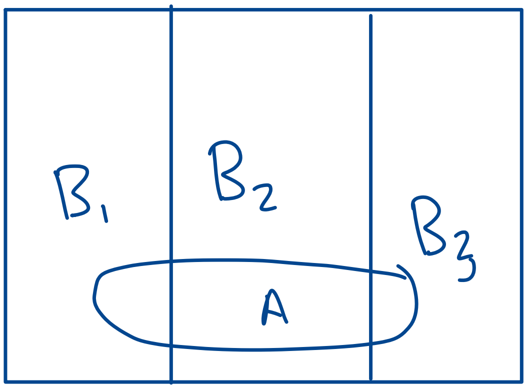
$$P(\overset{\textcircled{A}}{\text{orator}} \mid \overset{\textcircled{B'}}{\text{women}}) = \frac{25}{1000}$$

$$P(\overset{B}{\text{man}} \mid \overset{A}{\text{orator}}) = \frac{P(A|B) * P(B)}{P(A|B) * P(B) + P(A|B') * P(B')}$$

$$= \frac{\frac{50}{1000} * \frac{1}{2}}{\frac{50}{1000} * \frac{1}{2} + \frac{25}{1000} * \frac{1}{2}} = \underline{\qquad}$$

☆ Generalized Baye's Rule

$$P(A) = \sum P(A|B_n) * P(B_n)$$



$$P(B_n|A) = \frac{P(A|B_n) * P(B_n)}{P(A)} = \frac{P(A|B_n) * P(B_n)}{\sum P(A|B_n) * P(B_n)}$$

Example Three factories A, B and C of an electrical bulb produce Respectively 35% 35% and 30% of total product. Approximately, 1.5%, 1% and 2% of the bulbs produced by these factories are known to be defective. IF a randomly selected bulb manufactured by a company was found to be defective, what is the probability that the bulb was manufacters in factory A?

$$P(A) = 0.35 \qquad P(B) = 0.35$$

$$P(C) = 0.30$$

$$P(\overset{A}{D} \mid \overset{1}{A}) = 0.015$$

$$P(\overset{A}{D} \mid \overset{2}{B}) = 0.01 \qquad P(\overset{A}{D} \mid \overset{3}{C}) = 0.02$$

$$P(A|D) = \frac{P(D|A) * P(A)}{P(D|A) * P(A) + P(D|B) * P(B) + P(D|C) * P(C)}$$

$$= \frac{0.015 * 0.35}{0.015 * 0.35 + 0.01 * 0.35 + 0.02 * 0.30}$$

$$= 0.356$$

Chapter 4

Discrete probability distribution

المتغير العشوائي

* The Random variable

⇒ A function from the sample space to a set of real numbers (X, Y, Z)

Example when tossing a fair coin 2-times, the Random variable X is the number of heads obtaind, what is the RV?

$$\Omega : \left\{ \begin{matrix} \overset{2}{(H,H)}, & \overset{1}{(H,T)} \\ \underset{1}{(T,H)}, & \underset{0}{(T,T)} \end{matrix} \right\}$$

$$X : \{\underbrace{0, 1, 2}_{\text{support}}\}$$

NOTE

قيم RV سميتها Support

Countable
" integers "
" أرقام صحيحة "
⇒ Discrete Random variable

0, 1, 8, 9

non Countable
" intervals "
" فترات/كسور عشرية "
⇒ Continous Random variable

3.8 , 6.2

Example Find probability distribution for the previous question ?

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

* probability distribution for discrete Random variable (Discrete probability distribution)

Example Determine whether each Random variable X is discrete or Continous:

① Let X be the number of Fortune 500 Companies that lost money in previous year

② Let X Represent the volume of gasoline in a 21-gallon tank

③ Let X Represent the speed of Rockets

④ Let X Represent the number of Calves born on a farm in one-year

⑤ Let X Represent the number of days of rain for the next 3 days

NOTE

integers دائماً Discrete Random variable قيم -

integes ← Continous Random variable قيم -

non-integers ←

* Probability mass function (PMS)

① $0 \leq P(X=x) \leq 1$

② $\sum P(X=x) = 1$

Relative Frequency

$$\boxed{R.f = \frac{f}{\sum f} = P(X=x)}$$

X	0	1	2	3	4	5
f	10	5	1	3	8	3

Frequency distribution

⇓

X	0	1	2	3	4	5
$P(X=x)$	$\frac{10}{30}$	$\frac{5}{30}$	$\frac{1}{30}$	$\frac{3}{30}$	$\frac{8}{30}$	$\frac{3}{30}$

PMf

Example If $p(X=x) = K \cdot X$, $X = 1, 2, 3, 4$

and is a PMF, find:

$0 \leq P(X=x) \leq 1$ $\sum (P(X=x)) = 1$

x	1	2	3	4
$P(X=x)$	k	$2k$	$3k$	$4k$

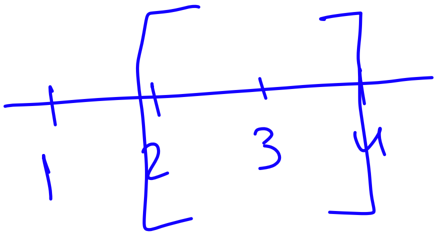
① The value of k

$$k + 2k + 3k + 4k = 1$$

$$10k = 1 \qquad \boxed{k = 0.1}$$

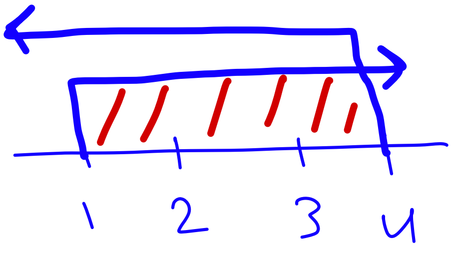
② $P(X=3) = 3k = 3 * 0.1 = 0.3$

③ $P(1 < X \leq 4)$



$$= P(X=2) + P(X=3) + P(X=4) = 0.9$$

④ $P(X > 1 \mid X < 4) = \dfrac{P(X>1 \cap X<4)}{P(X<4)}$



$$= \frac{P(X=2) + P(X=3)}{P(X=1) + P(X=2) + P(X=3)}$$

$$= \frac{0.5}{0.6} = \frac{5}{6}$$

⑤ $p(X = 3.4) = 0$

Example when throwing a fair dice 2 times, define the Random variable S to be the sum of 2 numbers obtained. Find the Probability distribution of S ?

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

S	2	3	4	5	6	7	8	9	10	11	12
$P(S=s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Symmetric

μ . population mean

$\bar{X}$. sample mean

$$\boxed{\underset{\text{mean}}{\mu} = 7}$$

* The Expected value $(E(x))$

$$E(x) = \mu = \text{mean} = \sum x \cdot p(X=x)$$

Example find the expected value for.

①

x	1	2	3	4
$p(X=x)$	0.4	0.3	0.2	0.1

$$E(x) = 1 * 0.4 + 2 * 0.3 + 3 * 0.2 + 4 * 0.1$$

$$= 2$$

②

x	1	2	3
$p(X=x)$	A	$\boxed{B}$	A

$\mu = E(x) = 2 \Rightarrow$ Symmetric

Example If the $E(x) = 2$, find a and b?

x	1	2	3	4
$p(X=x)$	a	0.3	b	0.1

ج/ $2 = 1 * a + 2 * 0.3 + 3 * b + 4 * 0.1$

$$2 = a + 0.6 + 3b + 0.4$$

$$2 = 1 + a + 3b$$

$$\boxed{a + 3b = 1} \text{ - - - - } ①$$

$$a + 0.3 + b + 0.1 = 1$$

$$a + b + 0.4 = 1$$

$$\boxed{a + b = 0.6} \text{ - - - } ②$$

$$\begin{array}{r} \cancel{a} + b = 0.6 \\ - \quad \cancel{a} + 3b = 1 \\ \hline \end{array}$$

$$a + 0.2 = 0.6$$

$$\boxed{a = 0\ 4}$$

$$-2b = -0.4$$

$$\boxed{b = 0.2}$$

* Properities of E(x):

① $E(\underset{\text{الثابت}}{a}) = \underset{\text{الثابت نفسه}}{a}$ $\qquad E(3) = 3$

$\qquad E(2.5) = 2.5$

② $E(ax) = a * E(x) \Rightarrow E(2x) = 2 * E(x)$

③ $E(x \pm y) = E(x) \pm E(y)$

Example

x	1	2	3	4
P(X=x)	0.4	0.3	0.2	0.1

find:

① $E(x) = 1*0.4 + 2*0.3 + 3*0.2 + 4*0.1$

$= 2$

② $E(x^2) = 1^2*0.4 + 2^2*0.3 + 3^2*0.2$

$$+ 4^2 * 0.1$$

$$= 5$$

③ $E\left(\frac{1}{x}\right) = \frac{1}{1} * 0.4 + \frac{1}{2} * 0.3$

$$+ \frac{1}{3} * 0.2 + \frac{1}{4} * 0.1$$

$$= \underline{\qquad}$$

④ $E\left(e^{x}\right) = e^{1} * 0.4 + e^{2} * 0.3 + e^{3} * 0.2$

$$+ e^{4} * 0.1 = \underline{\qquad}$$

العدد النيبري

$\simeq 2.7$

* * *

Example If the $\mu = E(x) = 10$, find:

① $E(\mu) = E(10) = 10$

② $E(E(x)) = E(10) = 10$

③ $E(2X-3) = E(2X) - E(3)$

$= 2 * E(X) - 3$

$= 2 * 10 - 3 = 17$

④ $E\left(1 - \frac{X}{2}\right) = E(1) - E\left(\frac{X}{2}\right)$

$= 1 - E\left(\frac{1}{2} * X\right)$

$= 1 - \frac{1}{2} * E(X) = -4$

Example If the probability mass function for the number of episodes of otitis media in the first 2 years of life are shown, find the Expected number of visits?

r	0	1	2	3	4	5	6
$P(R=r)$	0.129	0.264	0.271	0.185	0.095	0.039	0.017

$$E(x) = 0 * 0.129 + 1 * 0.264 + 2 * 0.271 + 3 * 0.185 + 4 * 0.095 + 5 * 0.039 + 6 * 0.017 \simeq 2.0 \ldots$$

حل

∴ ∴

Example Determine which one of the following tables Represent PMF:

A)

r	0	1	2	3
$P(X=r)$	0.18	0.34	0.36	0.13

B)

r	0	1	2	3
$P(X=r)$	0.10	0.42	0.03	0.44

C)

r	0	1	2	3
$P(X=r)$	0.12	0.13	0.61	0.14

D)

r	0	1	2	3
$P(X=r)$	0.26	0.44	−0.15	0.45

A)

r	0	1	2	3	4
$P(X=r)$	0.15	0 25	0.10	0.25	0.30

B)

r	0	1	2	3
$P(X=r)$	0.15	0.20	0.30	0.10

C)

r	0	1	2	3	4
$P(X=r)$	0 15	-0.20	0.30	0.20	0.15

D)

r	-1	0	1	2	3	4
$P(X=r)$	0.15	0.30	0.20	0.15	0.10	0.10

* The variance (σ^2)

$$\sigma^2 = \mathrm{var}(X) = E(X-\mu)^2$$

$$\sigma^2 = E(X^2) - (E(X))^2$$

$$\boxed{\sigma^2 = E(X^2) - \mu^2}$$

Example

x	1	2	3	4
$P(X=x)$	0.4	0.3	0.2	0.1

Find the variance and STD?

$$\sigma^2 = E(X^2) - \mu^2$$

$$\Rightarrow E(X^2) = 1^2 * 0.4 + 2^2 * 0.3 + 3^2 * 0.2 + 4^2 * 0.1$$

$$= 5$$

$$\Rightarrow E(X) = \mu = 1 * 0.4 + 2 * 0.3 + 3 * 0.2 + 4 * 0.1$$

$$= 2$$

$$\sigma^2 = 5 - (2)^2 = \boxed{1}$$

$$\sigma = \sqrt{1} = \boxed{1}$$

* Properities of variance:

① $Var(a) = 0$ $\qquad var(9) = 0 \;,\; Var(3) = 0$

② $var(ax) = a^2 * var(x)$

③ $var(ax \pm b) = Var(ax) \pm \cancelto{0}{var(b)}$

$= a^2 * var(x)$

④ $Var(x) = \underbrace{E(x^2)} - (E(x))^2$

$$E(x^2) = Var(X) + (E(X))^2$$

$$\boxed{E(x^2) = \sigma^2 + \mu^2}$$

Example If $\mu = 10$, $\sigma^2 = 3$ find:

i) $E(var(x)) = E(3) = 3$

ii) $Var(E(x)) = Var(10) = 0$

iii) $Var(var(x)) = var(3) = 0$

NOTE

$(a \pm b)^2 = a^2 \pm 2ab + b^2$

حلو

iv) $E(x - \overset{\mu}{\boxed{10}})^2 = E(x^2 - 20x + 100)$

$E(x^2) - 20 * E(x) + E(100)$

$\sigma^2 + \mu^2 - 20 * 10 + 100$

$3 + 100 - 200 + 100 = \boxed{3}$

حل آخر

$Var(x) = E(x - \mu)^2$

$= E(x - \mu)^2 = var(x) = \boxed{3}$

v) $E(x^2) = \sigma^2 + \mu^2 = 3 + 10^2 = 103$

vi) $E(x - 2)^2 = E(x^2 - 4x + 4)$

$$= E(x^2) - 4 * E(x) + E(4)$$

$$= 103 - 40 + 4$$

$$= \boxed{67}$$

r	0	1	2	3	4	5	6
$P(R=r)$	0.129	0.264	0.271	0.185	0.095	0.039	0.017

PMF

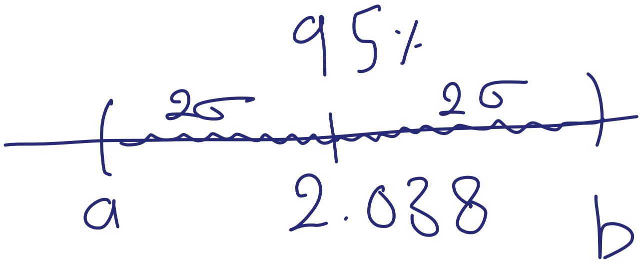
NOTE: تقريبا 95% من الاحتمالات

في PMF تقع بين 2σ

Ex $E(x) = 2.038$

$\sigma = 1.402$

find a & b such that Approxemitaly 95% of data lie within it ?



$$a = 2.038 - 2 * 1.402 = —$$

$$b = 2.038 + 2 * 1.402 = —$$

★ Binomial distribution

التوزيع الثنائي

① مستقلين ②

③ احتمالين

* If we have n-independent trials $n \geq 2$ and the outcomes in EACH trial are success or fail only. Let X be the number of success, then we say that X follows binomial distribution and is denoted by $X \sim Bin(n, p)$, where:

n: number of trials
p: probability of success

ال Binomial له ٣ شروط:

① $n \geq 2$

② outcomes → success
→ fail

نتيجتان فقط

③ independent

$$X \underset{\text{follows}}{\sim} Bin(n, P)$$

يتبع

n : number of trials

P : probability of success in each trial

q : Probability of fail in each trial

خواص

① $P + q = 1$

② $P(X = k) = \binom{n}{k} * P^{k} * q^{n-k}$

$\binom{n}{k}$: Combination

③ $\mu = E(x) = n \cdot P$

④ $var(x) = \sigma^{2} = n \cdot P \cdot q$

⑤ $STD(x) = \sigma = \sqrt{n \cdot p \cdot q}$

Example when tossing a fair coin 10 times, find:

① independent
② $n \geq 2$
③ outcomes $\begin{cases} \nearrow H \\ \searrow T \end{cases}$

$X \sim Bin(10, 0.5)$

A) the probability of getting:

i) exactly 8 heads

$$P(X=8) = \binom{10}{8} * 0.5^{8} * 0.5^{(10-8)}$$

$$= \underline{\underline{0.04}}$$

ii) at least 9 H $= P(X \geq 9)$

$$= P(X=9) + P(X=10)$$

$$= \binom{10}{9} * 0.5^{9} * 0.5^{(10-9)} + (- - - -$$

at least :	على الأقل	$P(X \geq k)$
at most :	على الأكثر	$P(X \leq k)$

iii) at least 2 H

$$P(X \geq 2) = 1 - P(X < 2)$$
$$= 1 - P(X \leq 1)$$
$$= 1 - \left(P(X=0) + P(X=1) \right)$$

$\vdots$

iv) at most 1 H

$$P(X \leq 1) = \left(P(X=0) + P(X=1) \right)$$

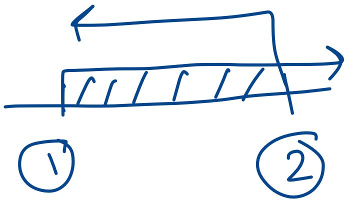
v) at most 8 H

$$P(X \leq 8) = 1 - P(X > 8)$$
$$= 1 - P(X \geq 9)$$
$$= 1 - \left(P(X=9) + P(X=10) \right)$$

vi) at most 2 H given that at least 1 H.

$$\Rightarrow P\left(\begin{matrix}\text{at most}\\ 2H\end{matrix}\,\middle|\,\begin{matrix}\text{at least}\\ 1H\end{matrix}\right) = \frac{P\left(\begin{matrix}\text{at}\\ \text{most}\end{matrix}\, 2H \cap \begin{matrix}\text{at least}\\ 1H\end{matrix}\right)}{P\left(\begin{matrix}\text{at least}\\ 1H\end{matrix}\right)}$$

$$= \frac{P(x \leq 2 \cap x \geq 1)}{P(x \geq 1)}$$



$$= \frac{P(X=1) + P(X=2)}{(1 - P(X<1))}$$

$$= \frac{P(X=1) + P(X=2)}{(1 - P(X=0))} = \text{- - - -}$$

B) The Expected number of heads and the variance and the standard deviation

$$\Rightarrow E(X) = \mu = n \cdot p$$

$$= 10 * 0.5 = 5$$

$$\Rightarrow \sigma^2 = n \cdot p \cdot q = 10 * \frac{1}{2} * \frac{1}{2} = 2.5$$

$$\Rightarrow \sigma = \sqrt{n \cdot p \cdot q} = \sqrt{2.5} = \text{—}$$

س1,

Example If $X \sim Bin(3, P)$ and $p(X \geq 1) = \frac{19}{27}$

find $var(X)$?

$$var(X) = n \cdot p \cdot q$$

$$= \cancel{3} * \frac{1}{\cancel{3}} * \frac{2}{3} = \boxed{\frac{2}{3}}$$

$$P(X \geq 1) = \frac{19}{27}$$

$$1 - P(X < 1) = \frac{19}{27}$$

$$1 - \frac{19}{27} = P(X = 0)$$

$$\frac{8}{27} = \underset{1}{\cancel{\binom{3}{0}}} * \underset{1}{\cancel{p^0}} * q^{(3-0)}$$

$$\frac{8}{27} = q^3 \Rightarrow \boxed{q = \frac{2}{3}} \quad \boxed{P = \frac{1}{3}}$$

Example If $X \sim Bin(n, P)$ and $\mu = 2, \sigma^2 = 1.6$ find n, P?

$$2 = n.P \longrightarrow \cancel{2} = n * \frac{\cancel{2}}{10}$$

$$\boxed{n = 10}$$

$$1.6 = \underbrace{n.P}_{\downarrow}.q$$

$$1.6 = 2 * q$$

$$\boxed{q = 0.8} \quad \boxed{P = 0.2}$$

* Using the Binomial tables

Example If $X \sim Bin(10, 0.4)$, find.

i) $P(X \leq 6) = 0.945$

ii) $P(X > 6) = 1 - P(X \leq 6)$

$= 1 - 0.945 = \_\_\_\_$

iii) $P(X < 7) = P(X \leq 6) = 0.945$

iv) $P(X \geq 7) = 1 - P(X < 7)$

$= 1 - P(X \leq 6)$

$= 1 - 0.945 = \_\_\_\_$

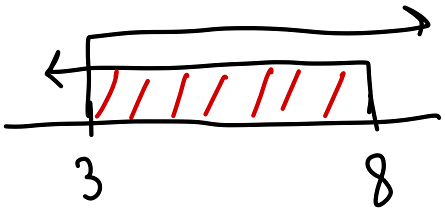
v) $P(3 \leq X \leq 8) =$

vi) $P(3 < X \leq 8) = P(4 \leq X \leq 8)$

vii) $P(3 < X < 8) = P(4 \leq X \leq 7)$

viii) $P(3 \leq X < 8) = P(3 \leq X \leq 7)$

ix) $P(X \geq 3 \mid X \leq 8) = \dfrac{P(X \geq 3 \cap X \leq 8)}{P(X \leq 8)}$



$= \dfrac{P(3 \leq X \leq 8)}{P(X \leq 8)}$

Example A family has 5 children, what is the probability that 3 children are females.

$$X \sim \text{Bin}(5, 0.5)$$

$$P(X=3) = \binom{5}{3} * 0.5^{3} * 0.5^{(5-3)} = 0.3125$$

Example In a multiple choice exam of 10-quest. each question had 5 answers only one of them is correct. Ahmad is answering the exam by guessing what is the probability that ahmad will answer:

1) 5 questions correctly

$$P(X=5) = \binom{10}{5} * 0.2^{5} * 0.8^{(10-5)} = 0.027$$

2) at most 5 questions

$$P(X \leq 5) = 0.994$$

Example what is the probability of obtaining 2 boys out of 5 children if the probability of a boy is 0.51 at each birth and the genders are considered independent Random variables?

$$X \sim Bin(5, 0.51)$$

$$P(X=2) = \binom{5}{2} * 0.51^{2} * 0.49^{(5-2)}$$

$$= 0.306$$

* * *

Example Evaluate the probability of 2 lymphocytes out of 10 white blood cells if the probability of any one cell being a lymphocyte is 0.2.

$$X \sim Bin(10, 0.2)$$

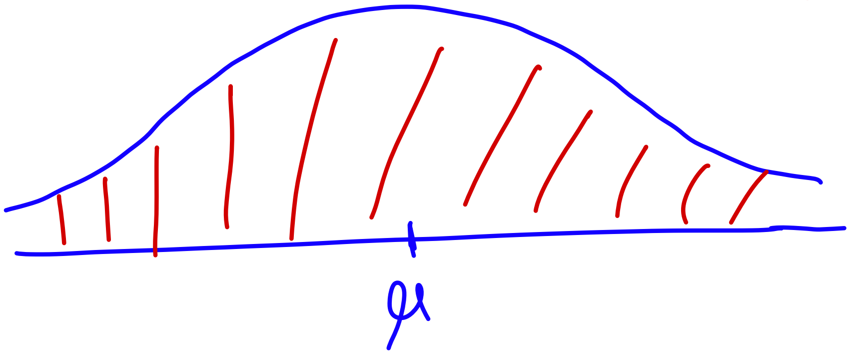
$$P(X=2) = \binom{10}{2} * 0.2^{2} * 0.8^{10-2} = 0.3020$$

Chapter 5

The normal distribution

التوزيع الطبيعي

→ Continous probability distribution



① Symmetric

② μ = mode = Q_2

③ Total area = 1

مهمة ④ Probability = area

⑤ $P(X=k)=0$

⑥ $P(X \le k) = P(X<k)$

ملاحظة المساواة ليس لها معنى في normal distribution

$$X \sim n(\mu, \sigma^2)$$

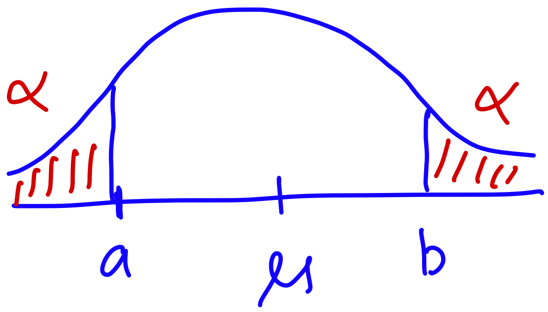
- follows
- is distributed as

Examples of normal distribution:

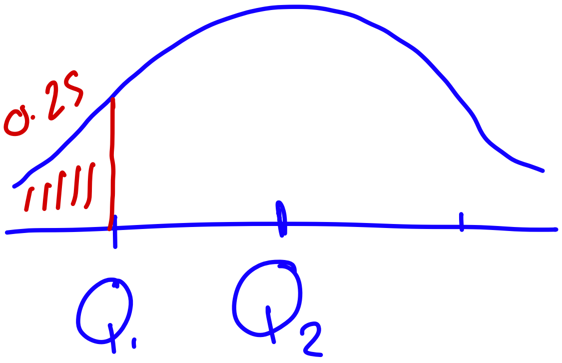
- weights
- speed
- Blood pressure
- height

NOTES

① $\mu = \frac{a+b}{2}$

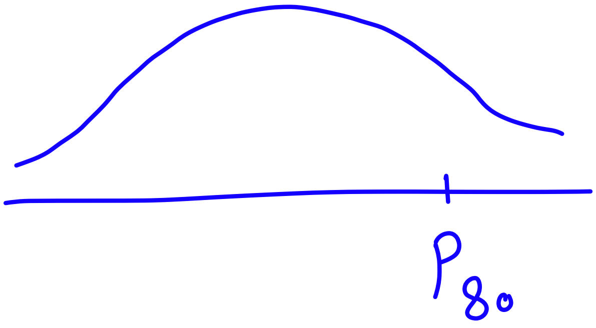


②

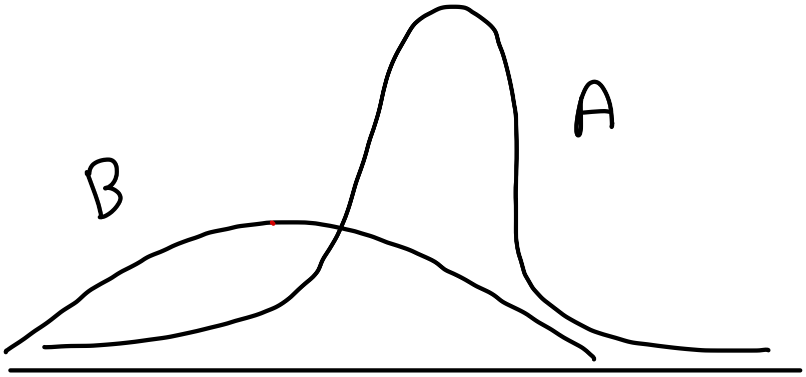


- $P(X \le Q_1) = 0.25$
- $P(X \le Q_3) = 0.75$

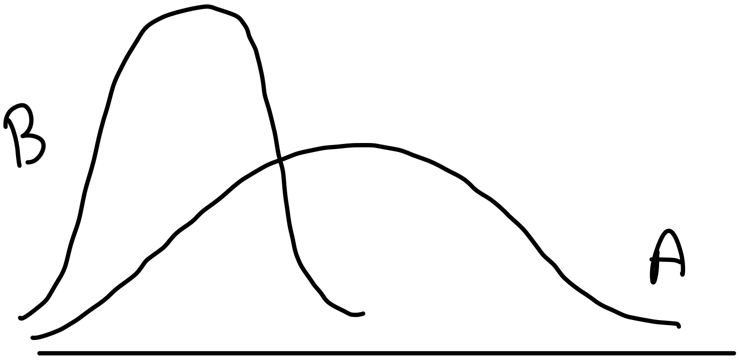
③ $P(X \le P_{80}) = 0.80$



Example A & B are normally distributed curves. which one has larger mean? which one has larger standard deviation?



* * *

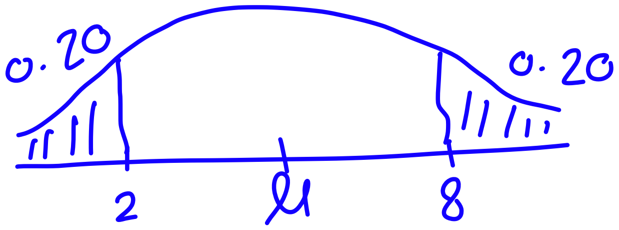


* * *

Example If $X \sim n(\mu, u)$, and

$P(X<2)=0.20$, $P(X>8)=0.20$, Find:

① $\mu = \frac{2+8}{2} = 5$



② $P(2 \le X < 8)$

$0.20 + X + 0.20 = 1$

$X + 0.40 = 1$ $\boxed{X = 0.60}$

③ 20th percentile (P_{20})

$P(X \le P_{20}) = 0.20$

$\boxed{P_{20} = 2}$

④ 80th percentile (P_{80})

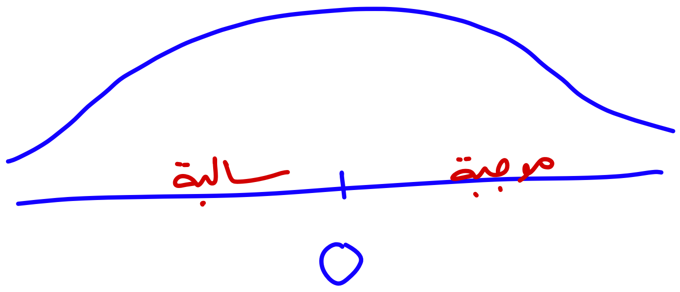
$P(X \le P_{80}) = 0.80$

$\boxed{P_{80} = 8}$

∴ ∴

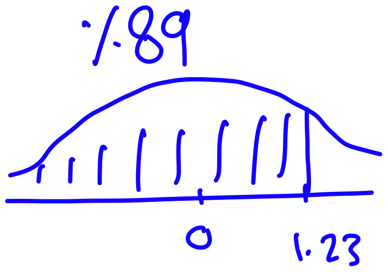
* The standard normal distribution

$Z \sim n(0,1)$



Example If $Z \sim n(0,1)$, then find:

i) $P(Z \leq 1.23) = 0.8907$



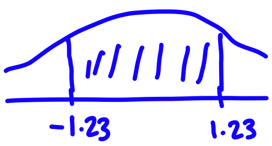
ii) $P(Z \geq 1.23) = 0.1093$

iii) $P(1.23 \leq Z \leq 2.12)$

$= P(Z \leq 2.12) - P(Z \leq 1.23)$

$= 0.9830 - 0.8907$

iv) $P(-1.23 < Z < 1.23)$



$= P(Z < 1.23) - P(Z < -1.23)$

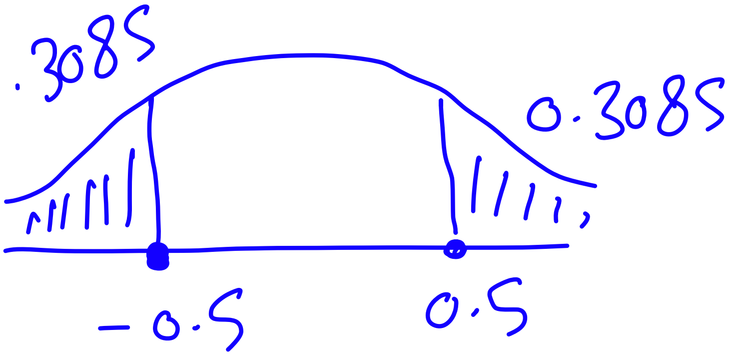
$= 0.8907 - 0.1093 = \_\_\_$

v) The z-score that correspond to a commulative area of 0.6915

z = 0.5

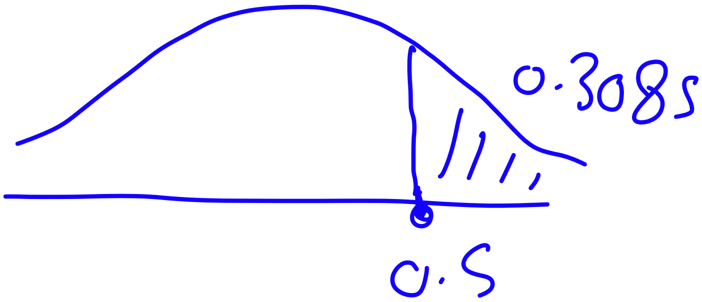
vi) If $p(z \leq a) = 0.3085$ find a?

$\boxed{a = -0.5}$



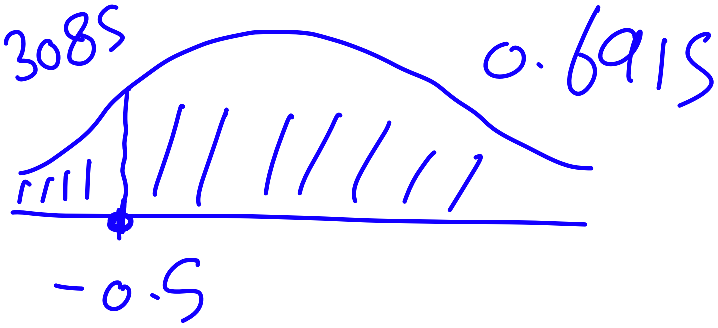
vii) a if $P(z > a) = 0.3085$?

a = 0.5



viii) a if $P(Z \geq a) = 0.6915$

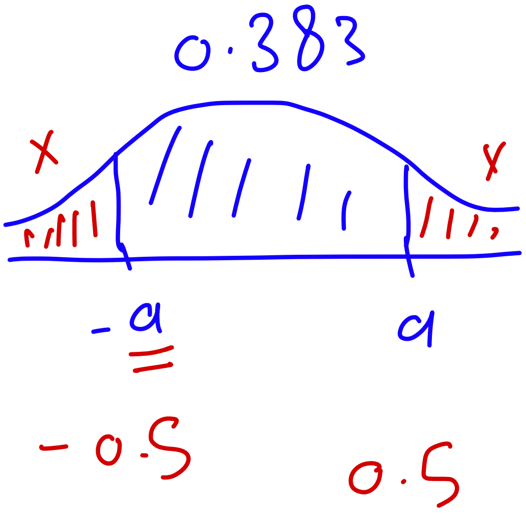
$$\boxed{a = -0.5}$$



ödev

ix) a if $P(-a < Z < a) = 0.383$

$P(Z<a) - P(Z<-a) = 0.383$



$x + x + 0.383 = 1$

$x = 0.3085$

$$\boxed{a = \pm 0.5}$$

* Standardization

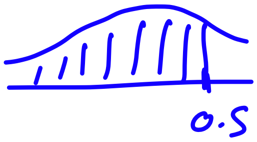
$$\boxed{n} \longrightarrow \boxed{z}$$

$$X \sim n(\mu, \sigma^2)$$

$$\Rightarrow Z = \frac{X-\mu}{\sigma} \sim n(0, 1)$$

$\boxed{\text{Example}}$ If $X \sim n(5, 4)$, find:

1) $P(X \le 6)$



$$= P\left(Z \le \frac{6-5}{2}\right) = P(Z \le 0.5)$$

$$= 0.6915$$

2) $P(X > 4) \Rightarrow P\left(Z > \frac{4-5}{2}\right)$

$$\Rightarrow P(Z > -0.5)$$

$$= 0.6915$$

3) $P(4 \leq X \leq 6)$

$= P(X \leq 6) - P(X \leq 4)$

$= P\left(Z \leq \frac{6-5}{2}\right) - P\left(Z \leq \frac{4-5}{2}\right)$

$= P(Z \leq 0.5) - P(Z \leq -0.5)$

$= 0.6915 - 0.3085$

4) a if $P(X > a) = 0.6915$

$P\left(Z > \frac{a-5}{2}\right) = 0.6915$

$P(Z > -0.5) = 0.6915$

$-0.5 = \frac{a-5}{2}$ $\boxed{a = 4}$

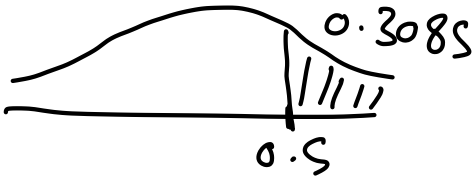
Example If $X \sim n(\mu, 4)$, $P(X \geq 6) = 0.3085$

find μ?

$$\Rightarrow P(X \geq 6) = 0.3085$$

$$= P\left(Z \geq \frac{6-\mu}{2}\right) = 0.3085$$

$$\frac{1}{\cancel{2}} = \frac{6-\mu}{\cancel{2}}$$



$$1 = 6 - \mu \qquad \boxed{\mu = 5}$$

* * *

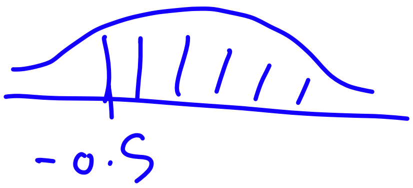
Example If $X \sim n(5, \sigma^2)$,

$P(X > 4) = 0.6915$ find σ^2?

$$\Rightarrow P(X > 4) = 0.6915$$

$$= P\left(Z > \frac{4-5}{\sigma}\right) = 0.6915$$

$$-\frac{1}{2} = \frac{-1}{\sigma}$$



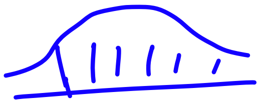
$\boxed{\sigma = 2}$ $\boxed{\sigma^2 = 4}$

∴ ∴ ∴

Example If $X \sim n(\mu, \sigma^2)$, $p(x>4) = 0.6915$, $p(x \le 6) = 0.6915$, find μ, σ^2?

$\Rightarrow P(X > 4) = 0.6915$

$P\left(Z > \frac{4-\mu}{\sigma}\right) = 0.6915$



$-\frac{1}{2} = \frac{(4-\mu)}{\sigma}$

$8 - 2\mu = -\sigma$

$\boxed{2\mu - \sigma = 8}$ …… ①

$\Rightarrow P(X \le 6) = 0.6915$

$= P\left(Z \le \frac{6-\mu}{\sigma}\right) = 0.6915$

$\frac{6-\mu}{\sigma} = \frac{1}{2}$ $\Rightarrow \sigma = 12 - 2\mu$

$\boxed{2\mu + \sigma = 12}$ --- ②

$$
\begin{array}{r}
\cancel{2}\,\cancel{\mu} - \sigma = 8 \\
- \quad \cancel{2\mu} + \sigma = 12 \\
\hline
-2\sigma = -4
\end{array}
$$

$$\boxed{\sigma = 2}$$

$$2\mu - 2 = 8$$

$$2\mu = 10$$

$$\boxed{\mu = 5}$$

$$\boxed{\begin{array}{l} \sigma^2 = 4 \\ \mu = 5 \end{array}}$$

* * *

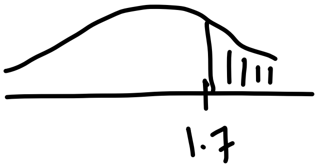
Example suppose that the grades in general exam are normally distributed with mean 68 and SD equal to 10, find:

a) the proportion of students that achieved more than 85

$$P(X > 85) = P\left(Z > \frac{85-68}{10}\right)$$

$$= P(Z > 1.7)$$

$$= 0.0446$$



B) the proportion of students that achieved between 60 and 90

$$= P(60 < X < 90)$$

$$= P(X < 90) - P(X < 60)$$

$$= P\left(Z < \frac{90-68}{10}\right) - P\left(Z < \frac{60-68}{10}\right)$$

$$= P(Z < 2.2) - P(Z < -0.8)$$

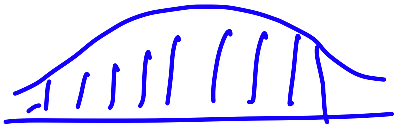
$$= 0.9861 - 0.2119$$

C) 95th percentile (P_{95})

$$P_{95} \Rightarrow P(X \le P_{95}) = 0.95$$

$$= P\left(Z \le \frac{P_{95} - 68}{10}\right) = 0.95$$

$$Z = \frac{1.64+1.65}{2}$$



$$= 1.645$$

$$1.645 = \frac{\overset{x}{P_{95}} - 68}{10}$$

Example If heights of students are normally distributed with mean 170 Cm and SD 10 Cm. Find:

A) a student is selected at random what is the probability that he will be shorter than 170

$$P(X < 170) \Rightarrow P\left(Z < \frac{170 - 170}{10}\right)$$

$$= P(Z < 0)$$

$$= 0.5$$

حل /

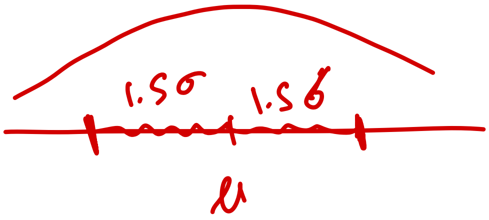
B) 10 students are selected at random what is the probability that exactly 4 of them are shorter than 170 ?

$$X \sim Bin(10, 0.5)$$

$$P(X=4) = \binom{10}{4} * 0.5^{4} * 0.5^{6}$$

[illegible]

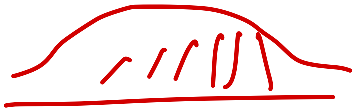
Example Suppose a child is considered to have normal lung growth if his/her standarized FVC is within 1.5 standard deviation of the mean. what is the Proportion of children are within the normal range ?

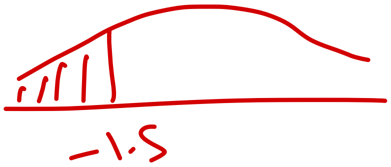


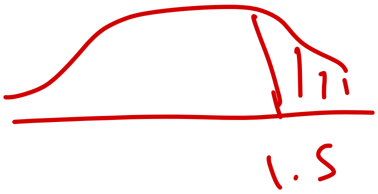
$\Rightarrow P(-1.5 < Z < 1.5)$

$= P(Z < 1.5) - P(Z < -1.5)$

$= 0.9332 \quad - \quad 0.0668 = 0.8664$

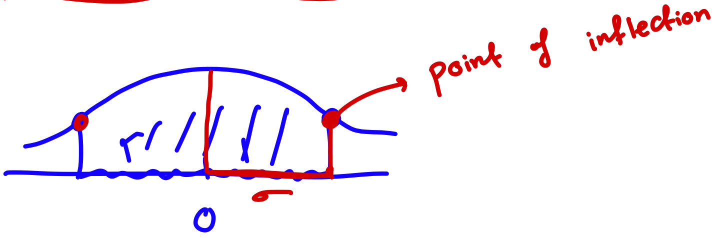






* * *

NOTE



① $P(-1 < Z < 1) = 0.6827$

② $P(-2 < Z < 2) = 0.95$

③ $P(-3 < Z < 3) = 0.99$

NOTE the height of normal distribution curve is always = $\frac{1}{\sqrt{2\pi\sigma}}$

$\downarrow\uparrow h \propto \frac{1}{\downarrow\sigma\uparrow}$

Example If $X \sim n(50, 4)$, find:

1) the mean = 50

2) the mode = 50

3) the median = 50

4) $IQR = Q_3 - Q_1 = 51.34 - 48.66$
$= 2.68$

5) variance and SD
④ ②

$IQR \Rightarrow Q_3 : P_{75} \Rightarrow P(X \leq P_{75}) = 0.75$

$= P\left(Z \leq \frac{P_{75} - 50}{2}\right) = 0.75$

$$0.67 = \frac{P_{75} - 50}{2}$$

$$\boxed{P_{75} = 51.34}$$

$$Q_1 : P_{25} \Rightarrow P(X < P_{25}) = 0.25$$

$$\Rightarrow P\left(Z \leq \frac{P_{25} - 50}{2}\right) = 0.25$$

$$-0.67 = \frac{P_{25} - 50}{2}$$

$$P_{25} = 48.66$$

اذكرونا بدعوة طيبة

همة الطب