

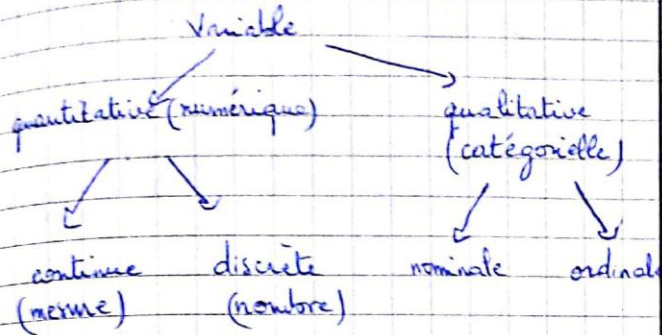
Série I

Rapports

statistique : ensemble de méthodes qui

permettent de :

- collecter (sondage)
- résumer (stat. descriptive)
- analyser (inférence stat.)



Ex 1 :

a) l'état civil des habitants de Rabat :

variable qualitative nominale.

b) Nombre de pages : variable quantitative
discrete.

c) Durée de vie d'un appareil : variable
quantitative continue.

d) le poids d'un nouveau né : variable
quantitative continue.

e) les catégories socio-professionnelles
au Maroc : variable qualitative nominale.

f) Nombre de pièces défectueuses : variable
quantitative discrete.

Ex 2 :

Rappel : nombre décimal à arrondir

① niveau de précision :

dixième près : un chiffre après la virgule
centième près : 2 chiffres " "
millième près : 3 chiffres " "

② chiffre suivant la décimale appelée

Exemple :

arrondir au dixième

2,3 2,35 2,4

2,35 → 2,4 (la décimale doit être
paire)

- voir série.

Ex 3 :

stat. descriptive

- tableaux
- représentations graphiques
- indicateurs

a) la variable est la qualité d'une
émission : qualitative ordinale

b.

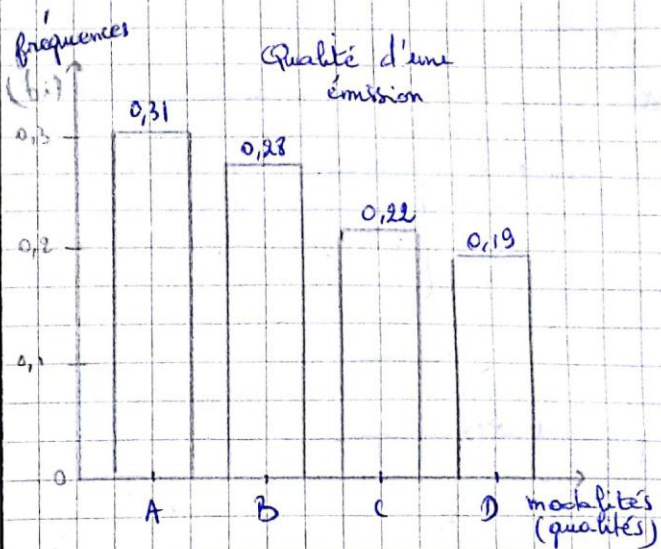
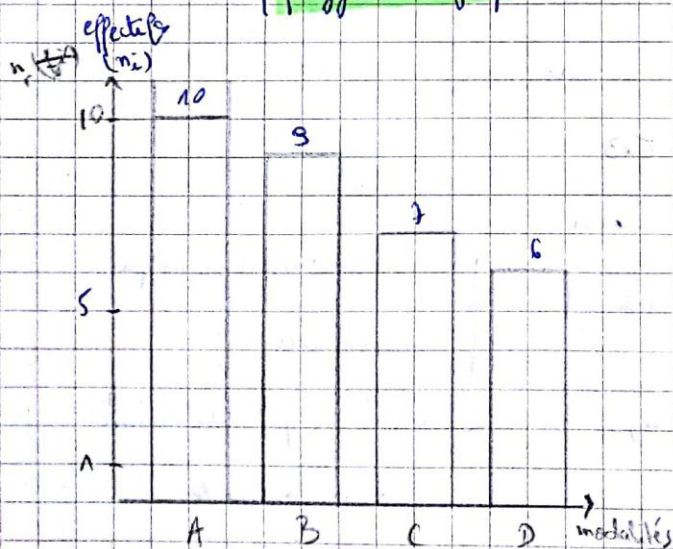
x_i	n_i	$f_i = \frac{n_i}{N}$
A	10	$\frac{10}{32} = 0,3125 = 0,31$
B	9	$\frac{9}{32} = 0,28125 = 0,28$
C	7	$\frac{7}{32} = 0,21875 = 0,22$
D	6	$\frac{6}{32} = 0,1875 = 0,19$
Total	32	1

Graphiques :

var. qualitative : $\left\{ \begin{array}{l} \text{diagramme en barres d'orgues} \\ \text{diag en secteurs (circulaire)} \end{array} \right.$

var. discrète : diag. en barres.

var. continue : $\left\{ \begin{array}{l} \text{histogramme} \\ \text{polygone de fréquences} \end{array} \right.$



$$\alpha_i = f_i \times 360.$$

$$A : \alpha_i = 0,31 \times 360 = 111,6^\circ$$

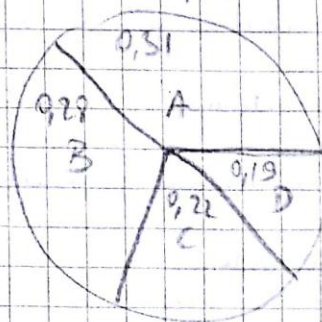
$$B : \alpha_i = 0,28 \times 360 = 100,8^\circ$$

$$C : \alpha_i = 0,22 \times 360 = 79,2^\circ$$

$$D : \alpha_i = 0,19 \times 360 = 68,4^\circ$$

$$\text{Total d'angles} = 360^\circ.$$

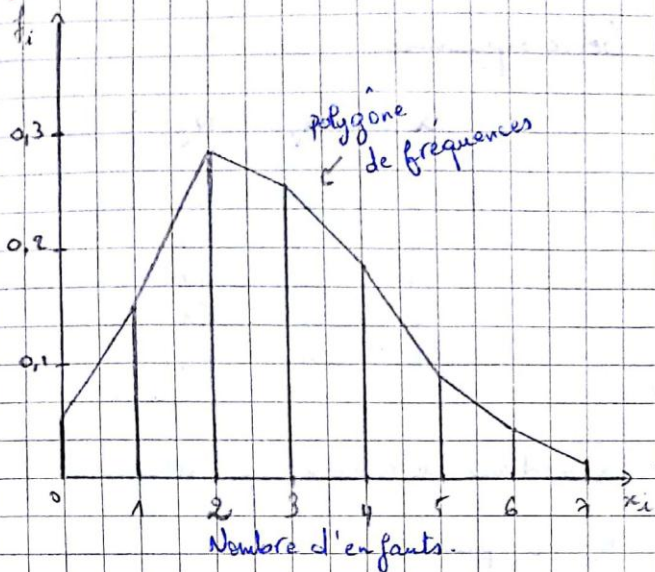
Diag circulaire : qualité d'émission.



Ex 4:

a) - variable : nombre d'enfants d'une famille
type : variable quantitative discrète

b) -

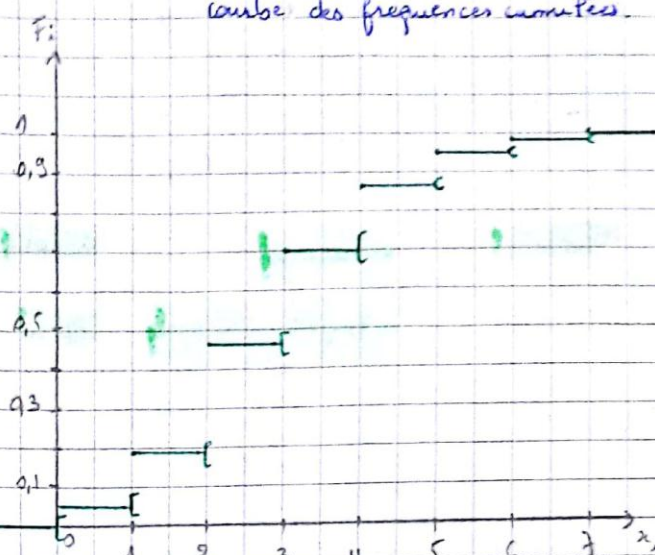


x_i	0	1	2	3	4	5	6	7
n_i	5	15	30	24	18	8	4	1
$f_i = \frac{n_i}{n}$	0,05	0,15	0,28	0,23	0,17	0,08	0,04	0,01
F_i	0,05	0,20	0,47	0,7	0,87	0,95	0,99	1

N_i 5 20 50 74 92 100 104 105

$$F_i = \sum_{j=1}^i f_j$$

courbe des fréquences cumulées



c) - Nombre de familles ayant au moins

$$3 \text{ enfants} = 24 + 18 + 8 + 4 + 1 = 55 \text{ familles.}$$

d) - Nombre de familles ayant au plus

$$4 \text{ enfants} = 5 + 15 + 30 + 24 + 18 = 92 \text{ familles.}$$

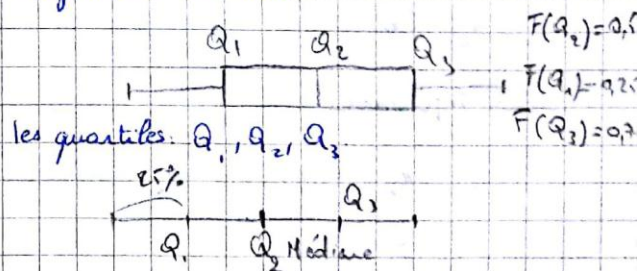
e) - proportion de familles ayant moins de

$$5 \text{ enfants et plus de 1 enfant} = \frac{30 + 24 + 18}{105} \approx 0,61$$

Ex 6 (suite).

le mode est 2 enfants (c'est la valeur la plus fréquente).

Diagramme en boîte :



Cas discret: 1 3 5 9 13

$N=5$
impair

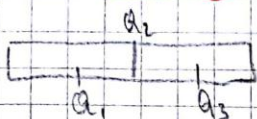
$$Q_2 = x_{(3)} = x_{\left(\frac{n+1}{2}\right)}$$

cas 2: $1 \ 3 \ 5 \ 9 \quad N=4$

$$\frac{3+5}{2} = \frac{x_{(2)} + x_{(3)}}{2}$$

$$= \frac{x_{(\frac{4}{2})} + x_{(\frac{4}{2}+1)}}{2}$$

Médiane $M_e = Q_2 = \begin{cases} x_{(\frac{N+1}{2})} & N \text{ impair} \\ \frac{x_{(\frac{N}{2})} + x_{(\frac{N}{2}+1)}}{2} & N \text{ pair} \end{cases}$



la médiane de la série inférieure est Q_1

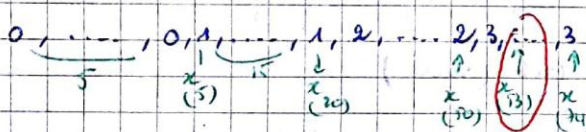
" " " " supérieure est Q_3

Diagramme en boîte:

on a: $N=105$ est impair.

$$\Rightarrow M_e = Q_2 = x_{(\frac{N+1}{2})} = x_{(\frac{105+1}{2})} = x_{(53)}$$

$M_e = 3$ enfants.



On divise la série en 2 séries
(inférieure et supérieure) de même effectifs

$$N_1 = N_2 = 52$$

Série inférieure:

x_i	n_i	N_i
0	5	5
1	10	15
2	15	30
3	2	32

La médiane de la série inférieure est Q_1

$N_1 = 52$ pair.

$$Q_1 = \frac{x_{(\frac{N_1}{2})} + x_{(\frac{N_1}{2}+1)}}{2} = \frac{x_{(26)} + x_{(27)}}{2}$$

$$Q_1 = \frac{2+2}{2} = 2$$

Série supérieure:

x_i	n_i	N_i
3	21	21
4	18	39
5	8	47
6	4	51
7	1	52

la médiane de la série sup. est Q_3

$N_2 = 52$ pair.

$$\Rightarrow Q_3 = \frac{x_{26} + x_{(27)}}{2} = \frac{4+4}{2} = 4$$



On calcule:

$$* Q_1 - 1,5 (Q_3 - Q_1)$$

avec: $Q_3 - Q_1 = R(Q)$ écart interquartile

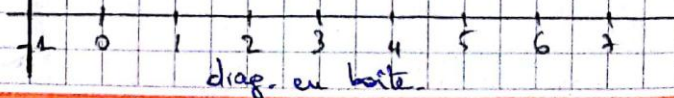
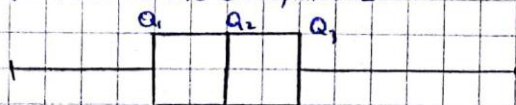
$$2 - 1,5 (4 - 2) = -1$$

$$* Q_3 + 1,5 \cdot R(Q)$$

$$4 + 1,5 (4 - 2) = 7$$

(On a pas de valeurs observées)

et on a: $\min = 0$, $\max = 7$.



Ex 5:

a) - variable : quantité de lait vendue.

type : variable quantitative continue.

b) - tableau de distribution : pour une var.

continue.

pour déterminer le nombre et l'amplitude des classes.

① - ordonner les observations :

1065 - 1068 - 1125 - 1207 - 1290 - 1325 - 1404 -

1621 - 1647 - 1708 - 1709 - 1794 - 1814 - 1900 -

2012 - 2123 - 2131 - 2210 - 2312 - 2490.

② - Calculer le nombre de classes :

Formule de Sturges : $1 + 3,3 \log_{(10)}(N) = 5,293$

" de Yule : $2,5 \sqrt[4]{N} = 5,286$ $\left\{ \begin{array}{l} P=5 \\ \text{classes} \end{array} \right.$

③ - l'étendue : $e = x_{\max} - x_{\min} = 2490 - 1065 = 1425$

④ - l'amplitude des classes :

$$a = \frac{e}{P} = \frac{1425}{5} = 285$$

⑤ - Construire les classes :

$[1065, 1350]$, $]1350, 1635]$; $]1635, 1920]$,

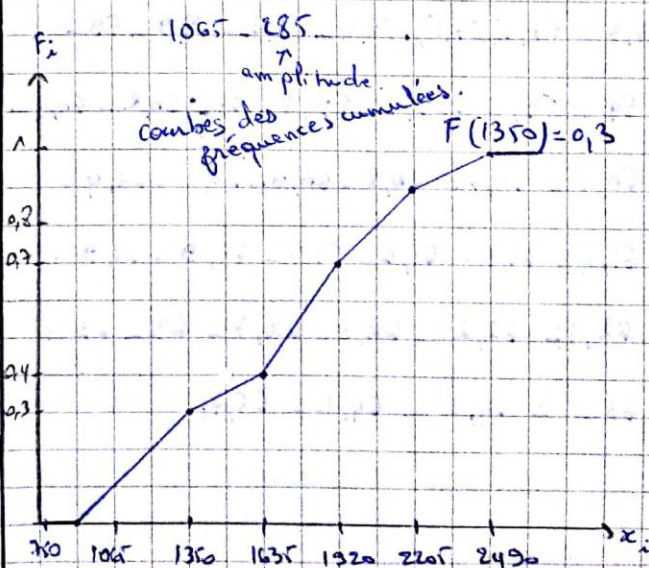
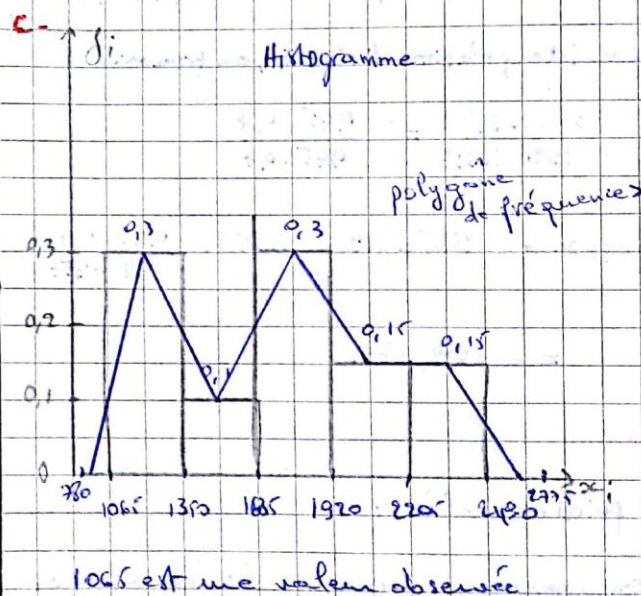
1065 est une valeur observée

$]1920, 2205]$; $]2205, 2490]$

2490 est une valeur observée.

classes	n_i	f_i	F_i
$[1065, 1350]$	6	0,3	0,3
$]1350, 1635]$	2	0,1	0,4
$]1635, 1920]$	6	0,3	0,7
$]1920, 2205]$	3	0,15	0,85
$]2205, 2490]$	3	0,15	1
Total	20	1	X

c -



d. x tq 45% des quantités de lait sont inférieures ou égales à x .

$\Rightarrow x$ tq $F(x) = 0,45$.

$\Rightarrow x$ est dans la classe $]1635, 1920]$

$\Rightarrow 1635 < x < 1920$.

Fest croissante $\Rightarrow F(1635) < F(x) < F(1920)$

$0,4 < 0,45 < 0,7$.

on a: $1635 < x < 1920$ et $0,4 < 0,45 < 0,7$.

par interpolation linéaire, on trouve:

$$\frac{x - 1635}{1920 - 1635} = \frac{0,45 - 0,4}{0,7 - 0,4}$$

$$\Rightarrow x = 1635 + \underbrace{(1920 - 1635)}_{q = 285} \cdot \underbrace{\frac{0,45 - 0,4}{0,7 - 0,4}}_{0,3 = \frac{p}{q}}$$

$x = 1682,5 \text{cl}$.

Ex 7:

ponds de 37 élèves:

on commence par ordonner les observations.

56,3 - 57,2 - 57,5 - 58,4 - 58,4 - 58,9 - 59 -
59 - 59,2 - 59,3 - 59,8 - 60 - 60,2 - 60,4 -
60,5 - 60,7 - 60,8 - 60,9 ^{61,1} 61,2 - 61,4 -
61,4 - 61,5 - 61,6 - 61,7 - 61,9 - 61,9 -
62,1 - 62,2 - 62,3 - 62,7 - 63 - 63 -
63 - 64,2 - 64,2 - 65,7.

Tiges	feuilles	n_i
56	3	1
57	5 5	2
58	4 4 9	3
59	0 0 2 3 8	5
60	0 2 4 5 7 8 9	7
61	(1) 2 4 4 5 6 7 9 9 9	10
62	1 2 3 7	4
63	0 0 1	3
64	2 2	2
65	7	1
		37

étendue: $e = x_{\max} - x_{\min} = 65,7 - 56,3 = 9,4$

Quantiles.

$Q_2 = Me = \text{médiane}$.

on a $N = 37$: impair

19^{ème} obs.
ordonnée

$$\Rightarrow Q_2 = x_{\left(\frac{N+1}{2}\right)} = x_{\left(\frac{37+1}{2}\right)} = x_{(19)}$$

$$\Rightarrow x_{(19)} = 61,1$$

on divise la série en 2 séries supérieures et inférieures de même effectifs.

$N_1 = N_2 = 18$.

on a: $N_1 = 18$ pair

$$Q_1 = \frac{x_{\left(\frac{N_1}{2}\right)} + x_{\left(\frac{N_1}{2} + 1\right)}}{2} = \frac{x_{(9)} + x_{(10)}}{2}$$

$$Q_1 = \frac{59,2 + 59,3}{2} = 59,25$$

on a $N_2 = 18$ pair.

$$\Rightarrow Q_3 = \frac{x_{(N_2/2)} + x_{(N_2/2+1)}}{2} = \frac{x_{(9)} + x_{(10)}}{2}$$

$$Q_3 = \frac{62,1 + 62,2}{2} = 62,15$$

L'intervalle interquartile: $[Q_1, Q_3] = [59,25, 62,15]$
= 50% des observations centrales.

L'écart interquartile: $R(Q) = Q_3 - Q_1 = 2,9$

Ex 8:

$$\bar{x} = \begin{cases} \frac{1}{N} \sum x_i & (\text{données brutes } x_1, x_2, \dots) \\ \frac{1}{N} \sum x_i \cdot n_i = \sum f_i \cdot x_i & (\text{Var. discrète}) \\ \frac{1}{N} \sum n_i \cdot c_i = \sum f_i \cdot c_i & (\text{Var. continue}). \end{cases}$$

c_i : centre de la classe associé à n_i

Variance $V(x)$ $\begin{cases} = \frac{1}{N} \sum (x_i - \bar{x})^2 & \text{données brutes} \\ = \frac{1}{N} \sum n_i (x_i - \bar{x})^2 = \sum f_i (x_i - \bar{x})^2 & \text{Var.} \\ = \frac{1}{N} \sum n_i (c_i - \bar{x})^2 = \sum f_i (c_i - \bar{x})^2 & \text{Var} \end{cases}$

L'écart-type: $\sigma = \sqrt{V(x)}$

$$V(x) = \frac{1}{N} \sum n_i (c_i - \bar{x})^2$$

$$V(x) = \left[\frac{1}{N} \sum n_i c_i^2 \right] - \bar{x}^2$$

$$V(x) = \left(\frac{12018}{100} \right) - (10,26)^2 = 14,91$$

$\Rightarrow$ L'écart-type: $\sigma = \sqrt{V(x)} = 3,86$

on a: $F(Q_1) = 0,25$; $F(Q_2) = 0,5$

et $F(Q_3) = 0,75$.

on a: Q_1 est tq: $F(Q_1) = 0,25$.

$]x_i, x_{i+1}]$	n_i	c_i	$n_i c_i$	$n_i c_i^2$	f_i	F_i	
$[2,4)$	5	3	15	45	0,05	0,05	$\Rightarrow Q_1$ est dans la classe $]6,8]$
$[4,6)$	8	5	40	200	0,08	0,13	$\Rightarrow 6 < Q_1 < 8$
$[6,8]$	15	7	105	735	0,15	0,28	F est croissante.
$[8,10]$	20	9	180	1620	0,2	0,48	$\Rightarrow F(6) < F(Q_1) < F(8)$
$[10,12]$	25	11	275	3025	0,25	0,73	et on a: $0,13 < 0,25 < 0,28$
$[12,14]$	12	13	156	2028	0,12	0,85	$\frac{Q_1 - 6}{8 - 6} = \frac{0,25 - 0,13}{0,28 - 0,13}$
$[14,18)$	10	16	160	2560	0,1	0,95	$\Rightarrow Q_1 = 6 + \frac{(8-6)}{2} \cdot \left(\frac{0,25-0,13}{0,28-0,13} \right)$
$[18,20]$	5	19	95	1805	0,05	1.	$\Rightarrow Q_1 = 7,6$
	$N=100$	Σ	1026	12018	1	Σ	Q_2 est tq: $F(Q_2) = 0,5 \Rightarrow Q_2 \in]10,12]$

$$\bar{x} = \frac{1}{N} \sum c_i n_i = \sum f_i c_i = \frac{1026}{100}$$

$$\bar{x} = 10,26$$

par interpolation linéaire:

$$Q_2 = 10 + (12-10) \cdot \frac{0,5-0,48}{0,73-0,48} = 10,16$$

Ex 9:

1) x = "nombre de tonnes de sardines pêchées"

population: nombre de jours.

c_i	3	7	9	11	14	18	
nombre de tonnes de sardines	$[0, 6[$	$[6, 8[$	$[8, 10[$	$[10, 12[$	$[12, 14[$	$[16, 20[$	
nombre de jours	10	14	32	24	10	10	100
f_i	0,1	0,14	0,32	0,24	0,1	0,1	1
F_i	0,1	0,24	0,56	0,8	0,9	1	x
$n_i \cdot c_i$	30	98	288	264	140	180	1000
$n_i \cdot c_i^2$	90	686	2592	2904	1960	3240	11472

$$3. \bar{x} = \frac{1}{N} \sum_i n_i \cdot c_i = \frac{1000}{100} = 10$$

$$v(x) = \frac{1}{N} \sum_i n_i \cdot c_i^2 - \bar{x}^2$$

$$= \frac{11472}{100} - 10^2$$

$$v(x) = 114,72 - 100 = 14,72$$

$$\sigma(x) = \sqrt{v(x)} = \sqrt{14,72} = 3,84$$

$[x_i, x_{i+1}[$	$[0, 2[$	$[2, 4[$	$[4, 6[$	$[6, 8[$	$[8, 10[$	$[10, 12[$	$[12, 14[$	$[14, 16[$	$[16, 18[$	$[18, 20[$	
n_i		10		14	32	24	5	5	5	5	100
f_i	0,033	0,033	0,034	0,14	0,32	0,24	0,05	0,05	0,05	0,05	1

l'effectif maximal est 32

→ la classe modale est $[8, 10[$

f_{i+1} : fréquence de la classe modale: 0,32

f_i : " " " qui précède

la classe modale = 0,14

f_{i+1} : l'effectif de la classe p_i est

la classe modale = 0,24.

x_i : la borne inférieure de la classe modale

x_{i+1} : " " sup. " " "

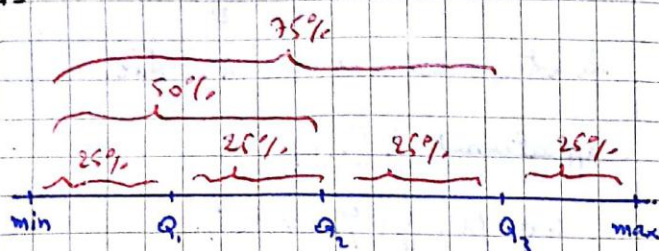
a : l'amplitude commune à toutes les classes.

$$m_d = x_{i+1} - a \cdot \frac{f_{i+1} - f_{i+2}}{(f_{i+1} - f_i) + (f_{i+1} - f_{i+2})}$$

$$= 10 - 2 \cdot \frac{0,32 - 0,24}{(0,32 - 0,14) + (0,32 - 0,24)}$$

$$m_d = 9,38$$

d-



x n'est autre que le 3^{ème} quantile

la classe du 3^{ème} quantile est $[10, 12[$

Car la première fréquence cumulée $\geq 0,75$ est 0,8.

$$x = Q_3 = x_i + (x_{i+1} - x_i) \cdot \frac{0,75 - F_i}{F_{i+1} - F_i}$$

$$Q_3 = 10 + (12 - 10) \cdot \frac{0,75 - 0,58}{0,8 - 0,58} = 11,583$$

e- la classe du 1^{er} quantile est $[8, 10[$

car la première fréquence cumulée $\geq 0,25$ est 0,56.

$$Q_1 = x_i + (x_{i+1} - x_i) \cdot \frac{0,25 - F_i}{F_{i+1} - F_i}$$

$$Q_1 = 8 + (10 - 8) \cdot \frac{0,25 - 0,14}{0,56 - 0,14} = 8,062$$

la classe de la médiane est $[8, 10[$

$$m_e = Q_2 = x_i + (x_{i+1} - x_i) \cdot \frac{0,5 - F_i}{F_{i+1} - F_i}$$

$$= 8 + (10 - 8) \cdot \frac{0,5 - 0,24}{0,56 - 0,24} = 9,525$$

$R(Q)$ = écart interquantile:

$$R(Q) = Q_3 - Q_1 = 11,583 - 8,062$$

$$R(Q) = 3,521$$

$$y_1 = Q_1 - 1,5 R(Q) = 2,780$$

$$y_3 = Q_3 + 1,5 R(Q) = 16,864$$



D'après le diag. en boîte, on voit qu'on a une asymétrie à droite

Coefficient d'asymétrie de Yule

$$A_y = \frac{Q_3 + Q_1 - 2Q_2}{R(Q)} = \frac{11,583 + 8,062 - 2 \times 9,525}{3,521}$$

$$A_y = 0,11 > 0$$

Coefficient d'asymétrie de Pearson

$$A_p = 3 \cdot \frac{\bar{x} - m_e}{s} = 3 \cdot \frac{10 - 9,625}{3,84}$$

$$A_p = 0,29 > 0$$

$$A_{p2} = \frac{\bar{x} - m_d}{s} = \frac{10 - 9,38}{3,84} = 0,16 > 0$$

Coefficient d'asymétrie de Fisher

$$A_F = \frac{\mu^3}{\sigma^3} = \frac{1}{N \cdot \sigma^3} \sum_i n_i (c_i - \bar{x})^3 = 0,34 > 0$$

Ex 12:

$$\bar{x} = 8,5 \quad \text{et} \quad \sigma(x) = 2,5$$

$$\bar{y} = 10 \quad \sigma(y) = 2$$

$$y = ax + b$$

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

$$\sigma(y) = a\sigma(x) \Rightarrow \sigma(y) = |a| \cdot \sigma(x).$$

$$10 = a \times 8,5 + b$$

$$2 = |a| \cdot 2,5 = a \cdot 2,5 \Rightarrow a = \frac{2}{2,5} = 0,8$$

$$10 = 0,8 \times 8,5 + b \Rightarrow b = 10 - 0,8 \cdot 8,5 = 10,6,8$$

$$b = 3,2.$$

La nouvelle note est $y = 0,8x + 3,2$.

Il n'y a pas de notes sup à 18.

$$x = 18$$

$$y = 0,8 \times 18 + 3,2 = 14,4 + 3,2 = 17,6 < 20$$

La transformation est possible puisque toutes les notes vont être comprises entre 0 et 20.

Après le changement des notes:

$$\bar{y} = 10 \quad \text{et} \quad \sigma = 2$$

Le pourcentage des notes comprises entre 6 et 14.

Th. de Tchebychev : pour un entier $k \geq 2$

au au moins $100 \cdot \left(1 - \frac{1}{k^2}\right) \%$ sont

à l'écart-type de la moyenne.

$$\Rightarrow [\bar{x} - k\sigma, \bar{x} + k\sigma].$$

$$6 = \bar{y} - 2 \cdot \sigma(y)$$

$$14 = \bar{y} + 2 \cdot \sigma(y)$$

donc d'après le th. de Tchebychev

on a au moins $100 \times \left(1 - \frac{1}{2^2}\right) \%$

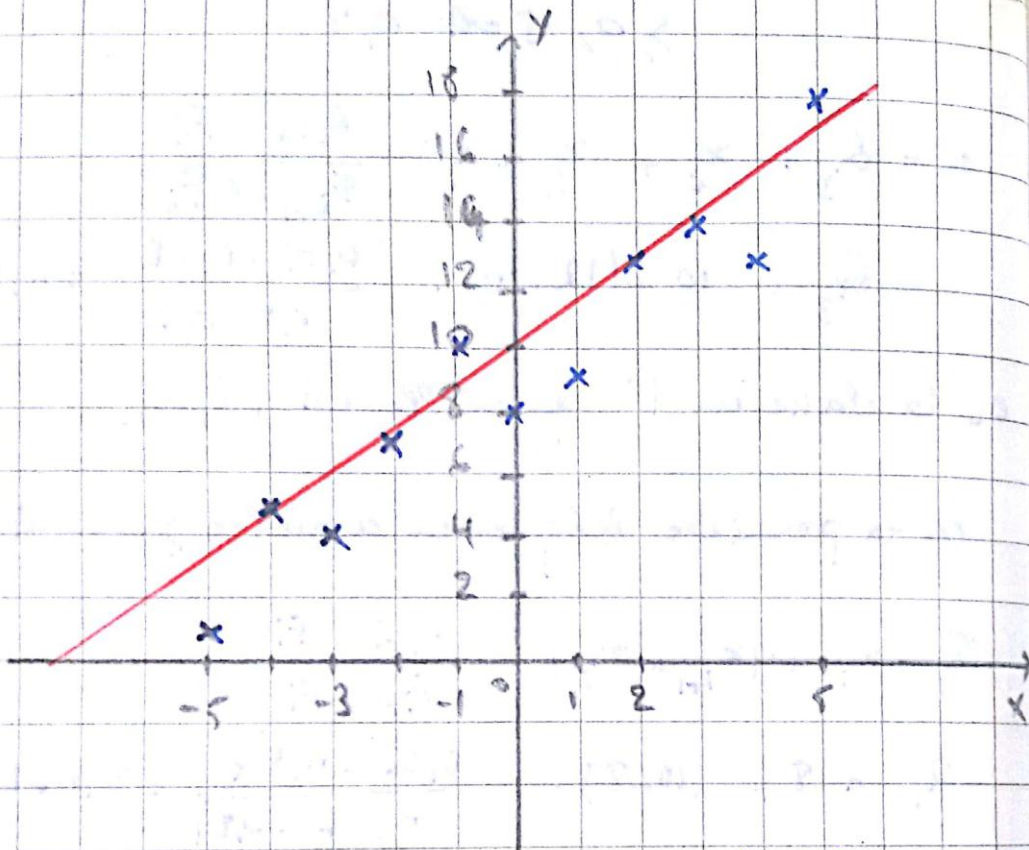
de notes entre 6 et 14 c'est à dire

appartenant à $[6, 14]$

$$100 \times \left(1 - \frac{1}{4}\right) \% = 75 \%$$

Ex 10:

x	-5	-4	-3	-2	-1	0	1	2	3	4	5
y	1	5	4	7	10	8	9	13	14	13	18



$$\text{cor}(x, y) = r(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$$

$r(x, y) > 0,8 \Rightarrow$ il existe une relation
linéaire entre x et y .

$$\begin{aligned} \text{cov}(x, y) &= \frac{1}{N} \sum (x_i - \bar{x})(y_i - \bar{y}) \\ &= \left(\frac{1}{N} \sum x_i y_i \right) - \bar{x} \bar{y} \end{aligned}$$

$$\bar{x} = \frac{1}{N} \sum x_i = 0, \quad \bar{y} = \frac{1}{N} \sum y_i = 9,27.$$

$$\text{cov}(x, y) = \frac{1}{N} \sum x_i y_i - \bar{x} \bar{y}$$

$$= \frac{1}{11} (-5 \times 1 + (-4 \times 5) + \dots + 5 \times 8) - 0 \times 9,27.$$

$$\text{cov}(x, y) = 14,36.$$

$$\text{si } \text{cov}(x, y) = \begin{cases} > 0 \Rightarrow x \text{ et } y \text{ varient dans le même sens.} \\ < 0 \Rightarrow x \text{ et } y \text{ varient dans des sens opposés.} \end{cases}$$

$$V(x) = \left(\frac{1}{N} \sum x_i^2 \right) - \bar{x}^2 = 10 \Rightarrow \sigma_x = 3,33$$

$$V(y) = \left(\frac{1}{N} \sum y_i^2 \right) - \bar{y}^2 = 22,61 \Rightarrow \sigma_y = 4,75.$$

le coefficient de corrélation :

$$r(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = 0,95 \approx 1$$

$\Rightarrow$ Il existe une relation linéaire entre

x et $y \Rightarrow$ on peut donc ajuster le modèle

par une droite de régression linéaire

$$y = ax + b$$

$$a = \frac{\text{cov}(x, y)}{V(x)}, \quad b = \bar{y} - a\bar{x}$$

$$= \frac{14,36}{10}$$

$$a = 1,436 \quad \Rightarrow \quad b = 9,27$$

la droite de régression est : $y = 1,436x + 9,27$