Université Mohammed V
Faculté des Sciences
Département de Mathématiques. Rabat.

Année 2021-2022
Filière SMI M18
Statistique & Probabilité

Série Statistique Descriptive

Ex 1 : Indiquer le type des variables (qualitatives, quantitatives discrètes ou quantitatives continues) dans chacun des cas suivants :

a) L'état civil des habitants de Rabat. b) Le nombre de pages d'un annuaire téléphonique.

c) La durée de vie d'un appareil électronique. d) Le poids d'un nouveau-né.

e) Les catégories socioprofessionnelles au Maroc.

f) Le nombre de pièces défectueuses dans un lot de 1000 ampoules.

Ex 2 : Arrondir les nombres suivant :

Donnée	Niveau de précision	Décimale	Chiffre suivant	Résultat
45.723	au centième près			
37.5152	au centième près			
26.874	au dixième près			
92.14	au dixième près			
25.21350	au millième près			
59.4325	au millième près			
41.466	au centième près			
4.61521	au centième près			
2.39	au dixième près			
876.0245	au millième près			
7.56231	au millième près			

Ex 3 : Les téléspectateurs sont invités à évaluer une émission en envoyant un message contenant l'une des lettres A, B, C ou D qui représentent respectivement "très bonne émission", "bonne émission", "mauvaise émission" et "très mauvaise émission". Ci après les évaluations de 32 spectateurs :

B, B, A, C, A, D, A, A, B, C, D, D, C, A, B, B, C, A, D, C, A, A, B, A, C, D, B, B, C, D, B, A.

a) Identifier la variable et préciser son type.

b) Dresser le tableau de distribution des effectifs et des fréquences.

Ex 4 : La distribution du nombre d'enfants pour un échantillon de familles se présente comme suit :

x_i	0	1	2	3	4	5	6	7
n_i	5	15	30	24	18	8	4	1

a) Identifier la variable et préciser son type.

c) Déterminer le nombre de familles qui ont au moins 3 enfants.

d) Déterminer le nombre de familles qui ont au plus 4 enfants.

e) Déterminer la proportion de familles qui ont moins de 5 enfants et plus de un enfant.

Ex 5 : Les données suivantes sont les quantités de lait (en centilitre) vendues par le laitier du quartier durant les 20 derniers jours du mois de ramadan.

1068, 2123, 2012, 2490, 1647, 2312, 1065, 1207, 1290, 1708

1900, 1404, 1794, 1709, 1325, 1621, 2210, 1814, 2131, 1125

a) Identifier la variable et préciser son type.

b) Dresser le tableau de distribution des effectifs, des fréquences et des fréquences cumulées.

c) Construire la courbe des fréquences cumulées.

d) Donner une valeur approchée de x telle que 45% des quantités de lait vendues soient inférieures ou égales à x.

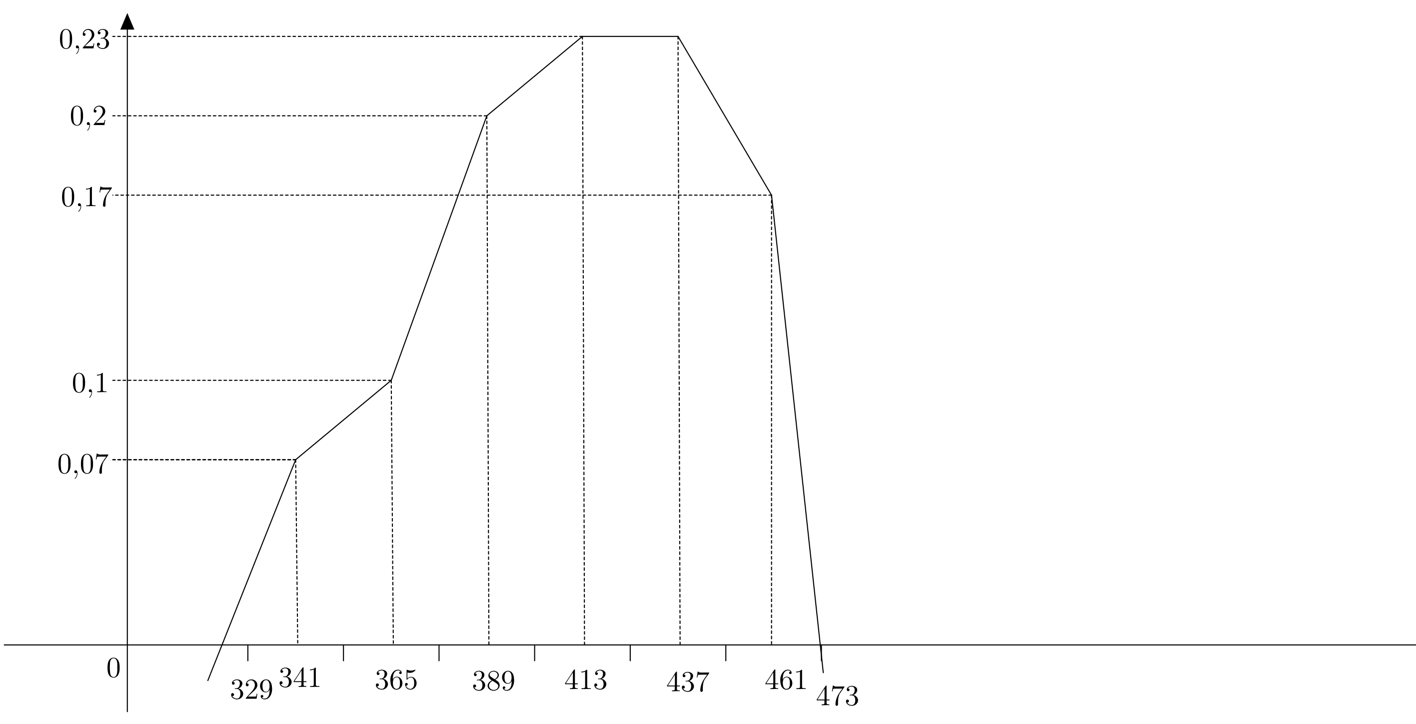
e) Calculer la valeur exacte de x telle que 45% des quantités de lait vendues soient inférieures ou égales à x.

Ex 6 : Pour la distribution du nombre d'enfants pour un échantillon de familles exercice 4, évaluer le mode et construire le diagramme en boîte.

Ex 7 : Voici les mesures de poids de 30 élèves (arrondis au dixième de kilogramme) qui ont été enregistrés : 59.2, 64.2, 63.0, 61.5, 62.3, 61.4, 57.5, 60.9, 59.8, 60.5, 59.0, 61.1, 60.7, 63.1, 61.6, 56.3, 61.9, 65.7, 60.4, 57.5, 58.9, 59.0, 61.2, 62.1, 61.4, 63.0, 58.4, 60.8, 60.2, 62.7, 60.0, 59.3, 61.9, 61.7, 64,2, 58.4, 62.2. Construire le diagramme tige et feuille et déterminer à partir de ce diagramme l'étendue et les quartiles et en déduire l'intervalle et l'écart interquartile.

Ex 8 : RatJ17

Soit la série statistique définie par le polygone des fréquences.



Tous les résultats seront arrondis à 10^{-2} près.

1) Dresser le tableau de distribution des fréquences et des fréquences cumulées.

2) Calculer l'étendue de cette série.

3) Calculer la moyenne arithmétique et l'écart-type.

4) Calculer le pourcentage minimum des observations appartenant à l'intervalle $[327,8; 496,28]$.

Indication : Calculer $\dfrac{\overline{x} - 327,8}{\sigma(x)}$.

Ex 9 : L'office National des Pêches a reçu du port de Mehdia les résultats de la pêche de sardines pour le printemps 2005. Le tableau qui suit nous indique le nombre de tonnes de sardines pêchées par jour durant cette période.

Nombre de tonnes de sardines	[0, 6[	[6, 8[	[8, 10[	[10, 12[	[12, 16[	[16, 20[
Nombre de jours	10	14	32	24	10	10

a) Définir pour ce problème la population et la variable statistique.

b) Construire la tableau des effectifs, des fréquences et des fréquences cumules.

c) Calculer le mode, la moyenne et l'écart-type. (Préciser les formules utilisées)

d) Calculer x, sachant que seulement 25% des jours de pêche du printemps 2005 ont un tonnage de sardines pêchées supérieur à x. (Préciser la formule utilisée)

e) Tracer le diagramme en boîte et étudier la symétrie de cette série statistique.

f) Construire la courbe de Lorenz, calculer la médiale et calculer l'indice de Gini.

Ex 10 : La moyenne semestrielle des notes (de 0 à 20) d'une classe d'éléves de terminale est de 8,5 et leur l'écart-type est de 2,5. Il n'ya pas de notes supérieures à 18. Le professeur veut changer les notes afin d'obtenir un moyenne égale à 10 et un écart-type égal à 2. Il utilise la transformation $y = ax + b$ où $a > 0$ et b sont deux nombres réels. On note x l'ancienne note et y la nouvelle. Déterminer a et b et vérifier que ce changement est possible. Après le changement des notes, donner le pourcentages d'élèves qui ont une note comprise entre 6 et 14.

Ex 11 : La moyenne d'âges de N_1 hommes est 35, celle de N_2 femmes est 50 et la moyenne d'âges de ces hommes et ces femmes est 40. Calculer le rapport $\dfrac{N_2}{N_1}$.

Exercices Facultatifs

Ex 12 :

Une étude de l'influence de la température sur un processus est menée. Les observations sont les suivantes :

Température X	-5	-4	-3	-2	-1	0	1	2	3	4	5
Phénomène Y	1	5	4	7	10	8	9	13	14	13	18

1) Calculer la covariance entre X et Y ainsi que la coefficient de corrélation

2) peut-on ajuster y par $a\,x + b$? justifier la réponse. Si oui, calculer a et b les paramètres de la droite de régression et représenter le nuage de points et la droite de régression.

Ex 13 : Une garderie a compté pour chacun de ses enfants, combien de fois ils ont été absents de la garderie le mois de septembre 2012. Voici les résultats : 0 1 0 5 2 3 0 4 5 1 3 2 0 3 0 4 5 4 3 2 0 0 1 0 1 2 1 2 0 0 3 4 5 0 0 1 0 1 0 2 0 1 3 4 5 4 0 2 0 0.

a) Quelle est la variable étudiée ? donner son type.

b) Présenter les données sous forme d'un tableau de distribution.

c) Construire le diagramme en bâton, le polygone et la courbe des fréquences cumulées.

c) Construire le diagramme en boîte et étudier la symétrie.

Ex 14 : On considère une série de taux d'hémoglobine dans le sang (en $g.l^{-1}$) mesurés chez 41 adultes présumés en bonne santé. La série ordonnée est :

105 110 112 112 120 120 125 125 126 132 132 133 134 135 138 139 141 142 144 145 145 146 148 148 149 149 150 150 152 153 158 158 159 160 160 164 165 165 170 172 172

1) Construire le diagramme en tiges et feuilles.

2) Donner les quartiles de cette série.

3) Que peut-on dire de la symétrie.(Justifier la réponse)

Ex 15 : Soit le tableau statistique suivant :

Classe	$[2\,,\,4]$	$]4\,,\,6]$	$]6\,,\,8]$	$]8\,,\,10]$	$]10\,,\,12]$	$]12\,,\,14]$	$]14\,,\,18]$	$]18\,,\,20]$
Effectifs n_i	5	8	15	20	25	12	10	5

1) Calculer la moyenne arithmétique et l'écart-type et déterminer les quartiles de la série : Q_1 le premier quartile, m_e la médiane et Q_3 le troisième quartile.

2) On résume la série aux quatre intervalles

$$[2\,,\,Q_1]\ ,\ ]Q_1\ ,\ m_e]\ ,\ ]m_e\,,\,Q_3]\ \text{et}\ ]Q_3\,,\,20]$$

a) Calculer alors la moyenne arithmétique et l'écart-type de cette nouvelle série.

b) Comparer ces résultats à ceux de la deuxième question.

CFD17 :

Exercice : les résultats d'examen d'un groupe d'étudiants sont répartis comme suit :

Classe de notes	$]0,\,4]$	$]4,\,a]$	$]a,\,b]$	$]b,\,17]$	$]17,\,18,5]$
Pourcentage	$12,5\%$	$22,5\%$	$37,5\%$	$17,5\%$	10%

On prendra 4 chiffres après la virgule et on précisera les formules utilisées.

1) Dresser le tableau de distribution des fréquences et des fréquences cumulées.

2) Sachant que le premier quartile est égal 7,33334, calculer la borne manquante a. (On prendra a l'entier le plus proche.)

3) Sachant que le centre de la troisième classe est égal 12,

α) Calculer x telque 25% des notes sont supérieur ou égal à x.

β) Calculer la moyenne arithmétique et l'écart-type.

γ) Que peut-on dire de la symétrie ? (Justifier la réponse)

4) Sachant que $\displaystyle\sum_{k=1}^{5} n_k \times c_k^2 = 11126$, calculer l'effectif total.

CFD15 : Une étude sur le budget consacré aux fournitures scolaires auprès des ménages a donné les résultats suivants :

Budget en centaine de DH	$]8,\,10]$	$]10,\,14]$	$]14,\,16]$	$]16,\,a]$	$]a,\,24]$	$]24,\,40]$
Pourcentages cumulés	6%	15%	35%	F%	77%	100%

1/ Calcul de la borne manquante a et du pourcentage F avec $F \in]50,\,75[$

α) Calculer le 1^{er} quartile Q_1.

β) Exprimer la médiane en fonction de a et F et le $3^{ème}$ quartile en fonction de a et F.

γ) Sachant que l'écart-interquartile est égal à 8 et la médiane vaut 17, Déterminer a et F.

2) Calcul de l'effectif total.

α) Dresser le tableau de distribution des fréquences et calculer l'écart-type.

β) Sachant que $\displaystyle\sum_{k=1}^{5} n_k.c_k^2 = 5.25552 \times 10^5$, calculer l'effectif total.

4) Que peut-on dire de la symétrie. (Justifier la réponse).

Corrigé de la série Statistique

Ex 1 : a) Qualitative. b) Quantitative discrète. c) et d) Quantitative continue. e) Qualitative. f) Quantitative discrète.

Ex 2 : Arrondir les nombres suivant :

Donnée	Niveau de précision	Décimale	Chiffre suivant	Résultat
45.723	au centième près	2	$3 < 5$	45.72
37.5152	au centième près	1	5	37.52
26.874	au dixième près	8	$7 > 5$	26.9
92.14	au dixième près	1	$4 < 5$	92.1
25.21350	au millième près	3	5	25.214
59.4325	au millième près	2	5	59.432
41.466	au centième près	6	$6 > 5$	41.47
4.61521	au centième près	1	5	4.62
2.39	au dixième près	3	$9 > 5$	2.4
876.0245	au millième près	4	5	876.024
7.56231	au millième près	2	$3 < 5$	7.562

Ex 3 : a) X = Qualité d'une certaine émission. C'est une variable qualitative.
b)

x_i	n_i	f_i	$\alpha_i = f_i \times 360$
A	10	0.31	111.6
B	9	0.28	100.8
C	7	0.22	79.2
D	6	0.19	68.4
Total	32	1	360

Ex 4 : a) X = Nombre d'enfants par famille. C'est une variable quantitative discrète.
b)

x_i	0	1	2	3	4	5	6	7	Total
n_i	5	15	30	24	18	8	4	1	105
f_i	0.05	0.14	0.28	0.23	0.17	0.08	0.04	0.01	1
F_i	0.05	0.19	0.47	0.7	0.87	0.95	0.99	1	

c) Le nombre de familles qui ont au moins 3 enfants est $55 = 24+18+8+4+1$.

d) Le nombre de familles qui ont au plus 4 enfants est $92 = 5+15+30+24+18$..

e) Le nombre de de familles qui ont moins de 5 enfants et plus de 1 enfant est $72 = 30 + 24 + 18$.

La proportion de familles qui ont moins de 5 enfants et plus de 1 enfant est 69%,
$\frac{72}{105} = 0.6857 \simeq 0.69$.

Ex 5 : a) X = Quantité de litre vendues par le laitier. C'est une variable quantitative continue.

Le nombre de classe est p :

On détermine P par l'une des deux méthodes,

régle de Sturge, $P = 1 + 3.3\ \log_{10}(20) = 5.293$, règle de Yule, $P = 2,5\sqrt[4]{20} = 5.286$

p est le nombre entier le plus proche de P, $\Longrightarrow p = 5$

$e = x_{max} - x_{min} = 2490 - 1065 = 1425$ et $a = \dfrac{e}{p} = \dfrac{1425}{5} = 285$

Les classes sont : $[1065,\, 1350]\,,\ ]1350,\, 1635]\,,\ ]1635,\, 1920]\,,\ ]1920,\, 2205]\,,\ ]2205,\, 2490]$

b)

$]x_i,\, x_{i+1}]$	n_i	f_i	$F(x)$
$[1065,\, 1350]$	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	

c)

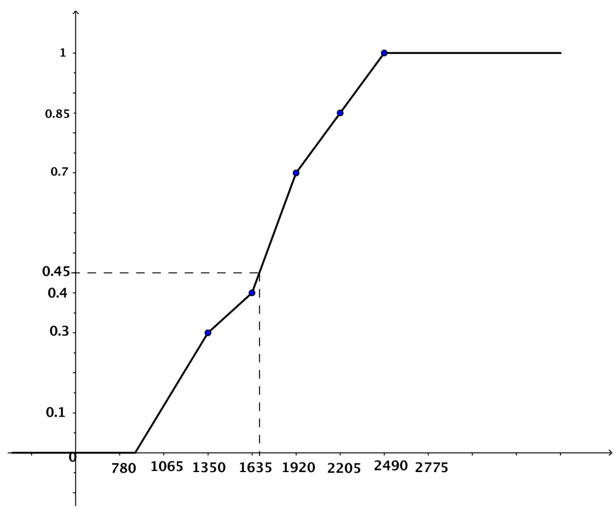


FIGURE 1 – Courbe des fréquences cumulées

d) $x \simeq 1700$ centilitres.

e) La classe contenant x est la première classe dont la fréquence cumulée F_{i+1} est supérieure ou égale à $r = 0.45$ on a $F_i < 0.45 \leq F_{i+1}$.

La première fréquence cumulée supérieure ou égale à 0.45 est $F_{i+1} = 0.7 \Longrightarrow]x_i,\, x_{i+1}] =]1635,\, 1920]$

La valeur de x est calculé par interpolation linéaire, on a,

$$x = x_i + (x_{i+1} - x_i)\frac{0.45 - F_i}{F_{i+1} - F_i)}$$

$x_i = 1635$, $x_{i+1} = 1920$, $F_i = 0.4$ et $F_{i+1} = 0.7$ d'où

$$x = 1635 + (1920 - 1635)\frac{0.45 - 0.4}{0.7 - 0.4)} = 1682.5$$

$x = 16.825$ litres.

<u>Ex 6 :</u> Le mode est la valeur de la variable qui correspond à l'effectif maximal et la plus grande fréquence. Le plus grand effectif est 30 et la plus grande fréquences est 0.28, donc le mode est 2.

x_i	0	1	2	3	4	5	6	7	Total
n_i	5	15	30	24	18	8	4	1	105
eff cum	5	20	50	74	92	100	104	105	

b) Pour la construction du diagramme en boîte, On a besoin de x_{min}, x_{max}, les quartiles c'est à dire le 1^{er} quartile Q_1, le $2^{ème}$ quartile Q_2 qui n'est autre que la médiane, le $3^{ème}$ quartile Q_3 et l'écart interquartile $R(Q)$ ainsi que les nombres $Q_1 - 1.5 \times R(Q)$ et $Q_3 + 1.5 \times R(Q)$

Pour la médiane, on a $N = 105$ est impair$\Longrightarrow m_e = x_{(\frac{N+1}{2})} = x_{(53)} = 3$

Pour le calcul du 1^{er} et le $3^{ème}$ quartile, déterminons les séries inférieur et supérieur.

on a $N = 105$ est impair$\Longrightarrow$ l'effectif total des séries inférieur et supérieur est :

$N_1 = \dfrac{N-1}{2} = \dfrac{105-1}{2} = 52$ (Si N était pair l'effectif total des séries inférieure et supérieure serait $N_1 = \dfrac{N}{2}$)

Les valeurs de la série inférieure sont toutes inférieure ou égale à la médiane tandis que les valeurs de la série supérieure sont toutes supérieure ou égale à la médiane.

x_i	0	1	2	3
n_i	5	15	30	2
effectifs cumulés	5	20	30	52

série inférieure avec $N_1 = 52$

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

série supérieure avec $N_1 = 52$

donc N_1 est pair d'où $\dfrac{N_1}{2} = 26 \Longrightarrow Q_1 = \dfrac{x_{(\frac{N_1}{2})} + x_{(\frac{N_1}{2}+1)}}{2} = \dfrac{x_{(26)} + x_{(27)}}{2} = \dfrac{2+2}{2} = 2$ et

$Q_3 = \dfrac{x_{(\frac{N_1}{2})} + x_{(\frac{N_1}{2}+1)}}{2} = \dfrac{x_{(26)} + x_{(27)}}{2} = \dfrac{4+4}{2} = 4.$

L'écart interquartile est : $\mathbf{R}(Q) = 4 - 2 = 2$, $Q_1 - 1.5 \times \mathbf{R}(Q) = -1$ et $Q_3 + 1.5 \times \mathbf{R}(Q) = 7$

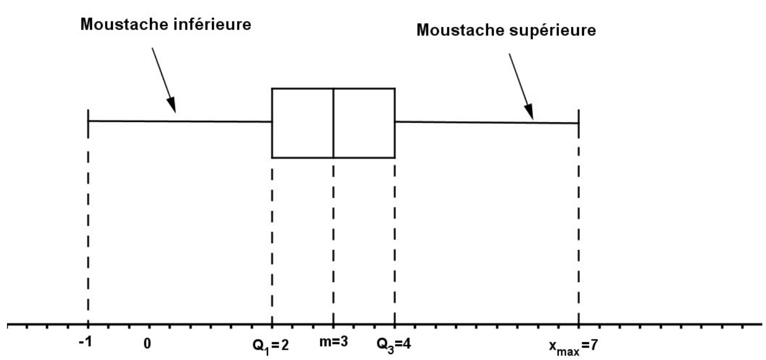


FIGURE 2 – Boîte à moustache

<u>Ex 7 :</u>le série ordonnée est : 56.3, 57.5, 57.5, 58.4, 5 8.4, 58.9, 59.0, 59.0, 59.2, 59.3, 59.8, 60.0, 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, 63.0, 63.0, 63.1, 64.2, 64.2, 65.7.

Tige	Feuille									Effectifs
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
62	1	2	3	7						4
63	0	0	1							3
64	2	2								2
65	7									1

a) L'étendue $e = 65.7 - 56.3 = 9.4$.

On a $N = 37$ est impair et $\dfrac{N+1}{2} = 19 \Longrightarrow$ la médiane est

$$m = x_{(\frac{N+1}{2})} = x_{(19)} = 61.1$$

On a $N_1 = 18$ est pair et $\dfrac{N_1}{2} = 9 \Longrightarrow$ le 1^{er} quartile est

$$Q_1 = \frac{x_{(\frac{N_1}{2}+1)} + x_{(\frac{N_1}{2})}}{2} = \frac{x_{(10)} + x_{(9)}}{2} = \frac{59.2 + 59.3}{2} = 59.25$$

On a $N_1 = 18$ est pair et $\dfrac{N_1}{2} = 9 \Longrightarrow$ le $3^{ème}$ quartile est

$$Q_3 = \frac{x_{(\frac{N_1}{2}+1)} + x_{(\frac{N_1}{2})}}{2} = \frac{x_{(10)} + x_{(9)}}{2} = \frac{62.1 + 62.2}{2} = 62.15$$

L'intervalle inter-quartile est $[Q_1, Q_3] = [59.25, 62.15]$.

l'écart inter-quatile est $\mathbf{R}(Q) = 62.15 - 59.25 = 2.90$

Ex 8 :

1) Le tableau des fréquences et des fréquences cumulées

Calculons l'amplitude. On a que les classes ont la même amplitude, d'où l'amplitude est/ $a = 365 - 341 = 24$.

$x_{min} = 341 - \dfrac{a}{2} = 341 - \dfrac{24}{2} = 341 - 12 = 329$

la $1^{ère}$ classe est $[329, 353[$, la $2^{èrme}$ classe est $[353, 377[$,

la $3^{èrme}$ classe est $[377, 401[$, la $4^{èrme}$ classe est $[401, 425[$,

la $5^{èrme}$ classe est $[425, 449[$, la $6^{èrme}$ classe est $[449, 473[$

Classe	[329, 353[	[353, 377[	[377, 401[	[401, 425[	[425, 449[	[449, 473[
Centre Classe c_i	341	365	389	413	437	461
Fréquence f_i	$0,07$	$0,1$	$0,2$	$0,23$	$0,23$	$0,17$
Fréquence cumulée F_i	$0,07$	$0,17$	$0,37$	$0,6$	$0,83$	1
$c_i \times f_i$	$23,87$	$36,5$	77.8	94.99	100.51	$78,37$
$c_i^2 \times f_i$	$8139,67$	$13322,5$	$30264,2$	$39230,87$	$43922,87$	$36128,57$

2) L'étendue est $e = x_{max} - x_{min} = 473 - 329 = 144$.

3) $\overline{x} = \sum_{k=1}^{n} f_k\, c_k = 341 \times 0,07 + 365 \times 0,1 + 389 \times 0,2 + 413 \times 0,23 + 437 \times 0,23 + 461 \times 0,17 = 412,04$.

$$\begin{aligned} V(x) &= \sum_{k=1}^{n} f_k\, c_k^2 - \overline{x}^2 \\ &= 341^2 \times 0,07 + 365^2 \times 0,1 + 389^2 \times 0,2 + 413^2 \times 0,23 + 437^2 \times 0,23 + 461^2 \times 0,17 - 412,04^2 \\ &= 171008,68 - 169776,96 = 1231,72 \Longrightarrow \sigma(x) = \sqrt{V(x)} = \sqrt{1231,72} = 35,0958. \end{aligned}$$

D'où $\sigma(x) = 35,1$

4) d'après le Théorème de Tchebychev le pourcentage minimum d'observations appartenant à l'intervalle $[\overline{x} - k\sigma(x), \overline{x} + k\sigma(x)]$ est $100(1 - \frac{1}{k^2})\%$ pour $k \geq 2$

On a $\overline{x} = 412,04$ et $\sigma(x) = 35,1$

$$\frac{\overline{x} - 327,8}{\sigma(x)} = \frac{412,04 - 327,8}{35,1} = 2,4 \Longrightarrow 327,8 = \overline{x} - 2,4 \times \sigma(x)$$

$$\frac{496,28 - \overline{x}}{\sigma(x)} = \frac{496,28 - 412,04}{35,1} = 2,4 \Longrightarrow 496,28 = \overline{x} + 2,4 \times \sigma(x)$$ D'où d'après le théorème de tchebychev le pourcentage minimum est :

$$100 \times (1 - \frac{1}{2,4^2})\% = 82,64\%$$

<u>Ex 9 :</u> a) la variable statistique X = nombre de tonnage de sardines. La population est les jours de pêches.

b) le tableau de distribution est :

Nbre de tonnes de sardines	[0, 6[	[6, 8[	[8, 10[	[10, 12[	[12, 16[	[16, 20[	Total
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	–

c) L'amplitude des classes n'est pas la même, il faut corriger les données.

Nbre de tonnes	[0, 2[	[2, 4[	[4, 6[	[6, 8[	[8, 10[	[10, 12[	[12, 14[	[14, 16[	[16, 18[	[18, 20[	Total
f_i	0.033	0.033	0.034	0.14	0.32	0.24	0.05	0.05	0.05	0.05	1

La classe modale est la classe qui a la plus grande fréquence. La plus grande fréquence est 0.32, donc la classe modale [8, 10[. Le mode est

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

x_{i+1} = est la borne supérieure de la classe modale = 10,

a = l'amplitude commune à toutes les classes = 2,

f_{i+1} = la plus grande fréquence = 0.32,

f_i = la fréquence de la classe au dessus ou avant la classe modale = 0,14 et

f_{i+2} = la fréquence de la classe en dessous ou après la classe modale = 0,24. Le mode est

$m_d = x_{i+1} - a \times \dfrac{(f_{i+1} - f_{i+2})}{(f_{i+1} - f_{i+2}) + (f_{i+1} - f_i)} = 10 - 2\dfrac{(0.32 - 0.24)}{(0.32 - 0.24) + (0.32 - 0.14} = 9.38.$
La moyenne est

$$\begin{aligned}
\overline{x} &= \frac{1}{N}\sum_{k=1}^{N} n_k \times c_k \\
&= \frac{3 \times 10 + 7 \times 14 + 9 \times 32 + 11 \times 24 + 14 \times 10 + 18 \times 10}{100} \\
&= \frac{1000}{100} \\
&= 10
\end{aligned}$$

D'où $\overline{x} = 10$
La variance et l'écart-type sont :

$$\begin{aligned}
V(x) &= \frac{1}{N}\sum_{k=1}^{N} n_k \times c_k^2 - \overline{x}^2 \\
&= \frac{3^2 \times 10 + 7^2 \times 14 + 9^2 \times 32 + 11^2 \times 24 + 14^2 \times 10 + 18^2 \times 10}{100} \\
& -10^2 \\
&= \frac{11472}{100} - 100 \\
&= 14,72
\end{aligned}$$

l'écart-type est $\sigma_x = \sqrt{V(x)} = \sqrt{14,72} = 3.84$

d) x n'est autre que le troisième quartile
la classe du troisième quartile est [10, 12[car la $1^{ère}$ fréquence cumulée supérieure ou égale à 0.75 est 0.8

$$Q_3 = x_i + (x_{i+1} - x_i)\frac{0.75 - F_i}{F_{i+1} - F_i}$$

$x_{i+1} = 12$, $x_i = 10$ $F_{i+1} = 0,8$ et $F_i = 0.56$

$$Q_3 = x_i + (x_{i+1} - x_i)\frac{0.75 - F_i}{F_{i+1} - F_i} = 10 + 2 \times \frac{0.75 - 0.56}{0.8 - 0.56} = 11.583$$

e) la classe médiane est [8, 10[car la $1^{ère}$ fréquence cumulée supérieure ou égale à 0.5 est 0.56

$$m_e = x_i + (x_{i+1} - x_i) \times \frac{0.5 - F_i}{F_{i+1} - F_i}$$

$x_{i+1} = 10$, $x_i = 8$, $F_{i+1} = 0,56$ et $F_i = 0.24$

$$m_e = x_i + (x_{i+1} - x_i) \times \frac{0.5 - F_i}{F_{i+1} - F_i} = 8 + 2 \times \frac{0.5 - 0.24}{0.56 - 0.24} = 9.625$$

la classe du premier quartile est [8, 10[car la $1^{ère}$ fréquence cumulée supérieure ou égale à 0.25 est 0.56

$$Q_1 = x_i + (x_{i+1} - x_i)\frac{0.25 - F_i}{F_{i+1} - F_i}$$

$x_{i+1} = 10$, $x_i = 8$, $F_{i+1} = 0,56$ et $F_i = 0.24$

$$Q_1 = x_i + (x_{i+1} - x_i)\frac{0.25 - F_i}{F_{i+1} - F_i} = 8 + 2 \times \frac{0.25 - 0.24}{0.56 - 0.24} = 8.062$$

$\mathbf{R}$(Q)=11.583-8.062=3.521, $Q_1 - 1.5 \times \mathbf{R}(Q) = 2.780$ et $Q_3 + 1.5 \times \mathbf{R}(Q) = 16.864$

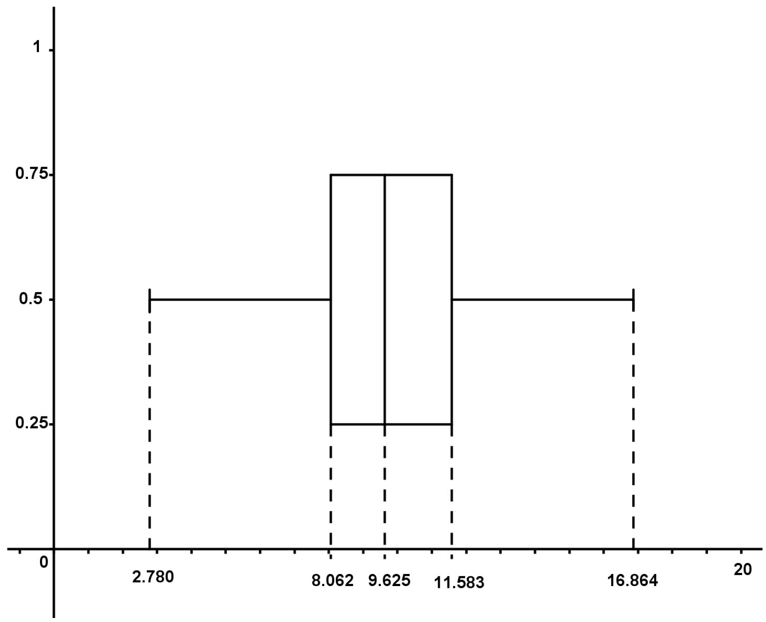


FIGURE 3 – Diagramme en Boîte

D'après le diagramme en boîte, on peut voir que la distribution n'est pas symétrique car la ligne vertical parallèle à l'axe des ordonnés et passant par la médiane ne divise pas la boite en deux parties égales.
Le premier coefficient d'asymétrie de Pearson est égal :

$$A_{P1} = 3 \times \frac{\overline{X} - m}{\sigma} = 3\frac{10 - 9.625}{3.84} = 0.29$$

la distribution statistique est unimodale, le second coefficient de Pearson est égal :

$$A_{P2} = \frac{\overline{X} - m_d}{\sigma} = \frac{10 - 9.38}{3.84} = 0.16$$

Le coefficient d'asymétrie de Yule est égal :

$$A_Y = \frac{Q_1 + Q_3 - 2 \times Q_2}{\mathbf{R}(Q)} = \frac{11.583 + 8.062 - 2 \times 9.625}{3.521} = 0.11$$

Le coefficient d'asymétrie de Fisher est égal :

$$A_F = \frac{\mu_3}{\sigma^3} = \frac{1}{N\,\sigma^3}\sum_{k=1}^{n} n_k\,(c_k - \overline{x})^3 = \frac{19.44}{3.84^3} = 0.34$$

tous les coefficient d'asymétrie sont positif, donc on a une asymétrie à droite.

f)

$$P_0 = 0 \quad \text{et} \quad P_i = 100 \times \sum_{k=1}^{i} f_k = \times 100\, F_i,\ 1 \leq i \leq p$$

$$q_0 = 0 \quad \text{et} \quad q_i = 100 \times \frac{\displaystyle\sum_{k=1}^{i} n_k\, x_k}{\displaystyle\sum_{k=1}^{p} n_k\, x_k} = 100 \times \frac{\displaystyle\frac{1}{n}\sum_{k=1}^{i} n_k\, x_k}{\overline{x}},\ 1 \leq i \leq p$$

Nbre de tonnes de sardines	[0, 6[	[6, 8[	[8, 10[	[10, 12[	[12, 16[	[16, 20[	Total
F_i	0.1	0.24	0.56	0.8	0.9	1	–
P_i	10	24	56	80	90	100	–
$n_k \times c_k$	30	98	288	264	140	180	1000
$\displaystyle\sum_{k=1}^{i} n_k \times c_k$	30	128	416	680	820	1000	–
q_i	3	12.8	41.6	68	82	100	–

calculons la médiale :

La classe médiale est la classe correspondant au premier pourcentage de masse des valeurs q_i supérieur ou égal à 50.

$m_d = x_i - (x_{i+1} - x_i) \times \dfrac{(50 - q_i)}{(q_{i+1} - q_i)}$

$x_{i+1} =$ est la borne supérieure de la classe médiale,

$x_i =$ est la borne inférieure de la classe médiale,

$q_{i+1} =$ le pourcentage de masse de la classe médiale et

$q_i =$ le pourcentage de masse de la classe au dessus ou avant la classe médiale.

la classe médiale est [10, 12] car le 1^{er} pourcentage de la masse des valeurs q_i supérieure ou égal à 50 est 68.

$$m_L = x_i + (x_{i+1} - x_I)\frac{50 - q_i}{q_{i+1} - q_i}$$

$x_I = 10$, $x_{i+1} = 12$, $q_{i+1} = 68$ et $q_i = 41.6$

$$\begin{aligned} m_L &= x_i + (x_{i+1} - x_I)\frac{50 - q_i}{q_{i+1} - q_i} \\ &= 10 + (12 - 10)\frac{50 - 41.6}{68 - 41.6} \\ &= 10.636 \end{aligned}$$

La médiane est $m_e = 9.625 \Longrightarrow m_L - m_e = 10.636 - 9.625 = 1.011$ l'écart entre la médiale et la médiane est faible par rapport à l'étendue donc la concentration est faible

Indice de Gini : $I_G = \dfrac{S}{5000}$ avec $S = 5000 - (S_1 + S_2 + S_3 + S_4 + S_5 + S_6)$ et

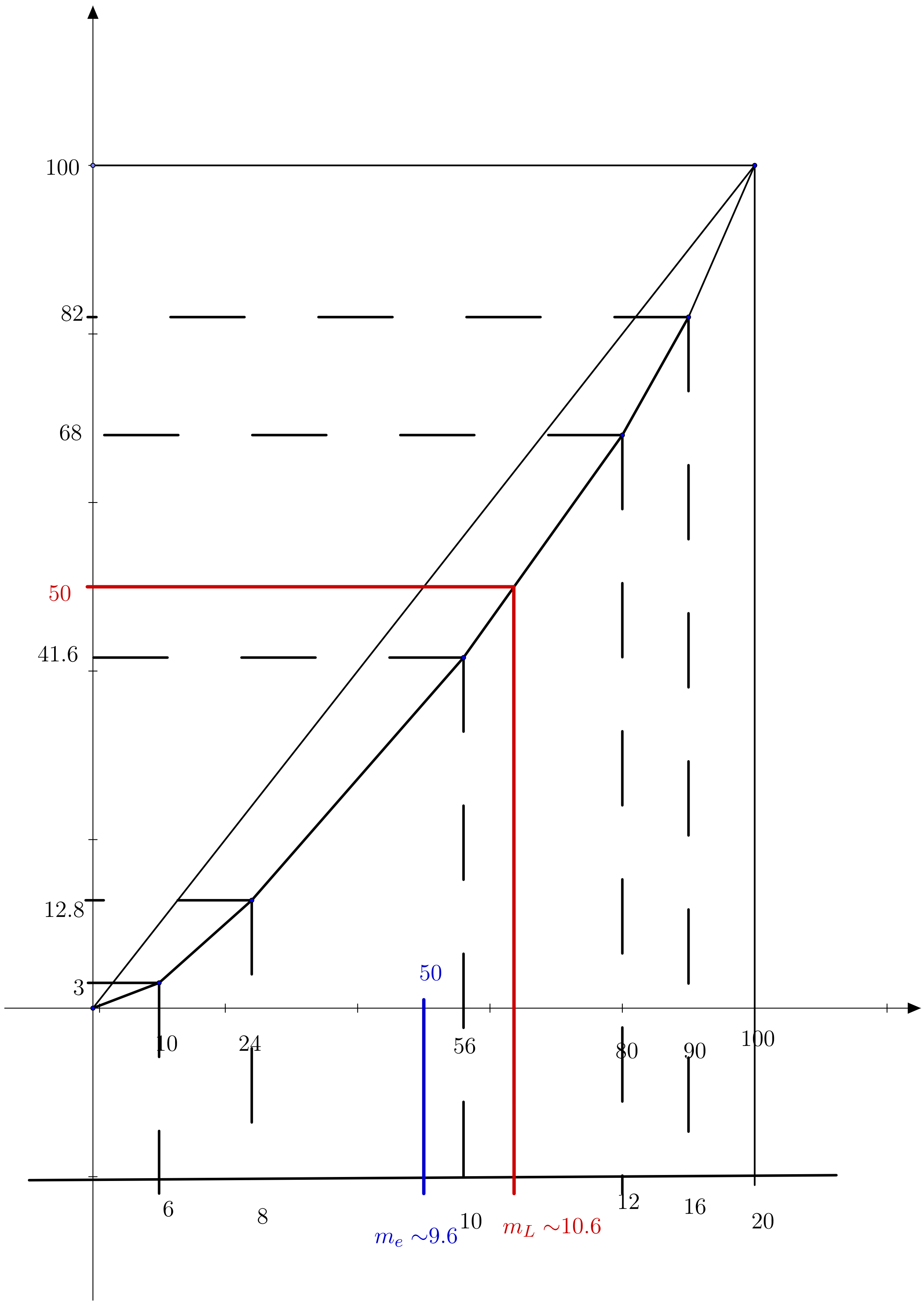
$$S_k = \frac{q_k + q_{k-1}}{2} \times (P_k - P_{k-1}) = \frac{q_k + q_{k-1}}{2} \times 100 \times f_k$$

$S_1 = \dfrac{3 + 0}{2} \times 100 \times 0.1 = 15$ $\qquad$ $S_2 = \dfrac{12.8 + 3}{2} \times 100 \times 0.14 = 110.6$

$S_3 = \dfrac{41.6 + 12.8}{2} \times 100 \times 0.32 = 870.4$ $\qquad$ $S_4 = \dfrac{68 + 41.6}{2} \times 100 \times 0.24 = 1315.2$

$S_5 = \dfrac{82+68}{2} \times 100 \times 0.1 = 750 \qquad S_6 = \dfrac{100+82}{2} \times 100 \times 0.1 = 910$

$S = 5000 - 3971.2 = 1028.8 \Longrightarrow I_G = \dfrac{1028.8}{5000} = 0.20$



Ex 10 :

$\overline{x} = 8.5$, $\sigma = 2.5$ et $0 \leq 18 \leq 20$. $y = ax + b$ avec $a > 0$ et $b \in \mathbb{R}$

$$\left\{\begin{array}{lcl} \overline{y} & = & a\,\overline{x} + b \\ \sigma(y) & = & |a|\,\sigma(x) \end{array}\right. \Longrightarrow \left\{\begin{array}{lcl} 8.5\,a + b & = & 10 \\ 2.5\,a & = & 2 \end{array}\right. \Longrightarrow \left\{\begin{array}{l} a = \dfrac{2}{2.5} = 0.8 \\ b = 3.2 \end{array}\right.$$

Ce changement est possible car :pour $x = 0$,on a : $y = 0.8 \times 0 + 3.2 = 3.2 \geq 0$, et la première note étant inférieure ou égale à 18 alors, $0.8 \times 18 + 3.2 = 17.6 \leq 20$. donc $0 \leq Y \leq 20$

Th. Tchebytchev

Le pourcentage minimum d'observation appartenant à l'intervalle $[\overline{x} - k \times \sigma(x), \overline{x} + k \times \sigma(x)]$ pour $k \geq 2$ est $100 \times (1 - \dfrac{1}{k^2})\%$

Après changement des notes, on a

La moyenne $\overline{y} = 10$ avec un écart-type $\sigma(y) = 2$.

Remarquons que

$6 = \overline{y} - 2\ \sigma(y) = 10 - 2 \times 2$ et que

$14 = \overline{y} + 2\ \sigma(y) = 10 + 2 \times 2$.

D'après le théorème de Tchebychev, le pourcentage d'étudiants ayant obtenue une note entre 6 et 14 est supérieur ou égal à $100 \times (1 - \dfrac{1}{2^2})\% = 75\%$

<u>Ex 11 :</u> En effet

$\overline{x} = \dfrac{1}{N_1}\displaystyle\sum_{k=1}^{N_1} x_k = 35 \Longrightarrow \sum_{k=1}^{N_1} x_k = 35N_1$, $\overline{y} = \dfrac{1}{N_2}\displaystyle\sum_{k=1}^{N_2} y_k = 50 \Longrightarrow \sum_{k=1}^{N_2} y_k = 50N_2$ et

$\overline{z} = \dfrac{1}{N}\displaystyle\sum_{k=1}^{N} z_k = 40 \Longrightarrow \sum_{k=1}^{N} z_k = 40N = \sum_{k=1}^{N_1} x_k + \sum_{k=1}^{N_2} Y_k = 35N_1 + 50N_2 = (N_1 + N_2)40 \Longrightarrow$ $5N_1 = 10N_2 \Longrightarrow \dfrac{N_2}{N_1} = \dfrac{1}{2}$.

<u>Ex 12 :</u>

1)

$\overline{x}$	$\overline{y}$	$V(x)$	$V(y)$	$Cov(x,y)$	$\rho(x,y)$
0	9.27	10	22.61	14.36	0.955

$\overline{x} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{11} x_k = 0$, $\overline{y} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{11} y_k = 9.27$, $V(x) = \left(\dfrac{1}{n}\displaystyle\sum_{k=1}^{11} x_k^2\right) - \overline{x}^2 = 10$,

$V(y) = \left(\dfrac{1}{n}\displaystyle\sum_{k=1}^{11} y_k^2\right) - \overline{y}^2 = 22.61$, $Cov(x,y) = \left(\dfrac{1}{n}\displaystyle\sum_{k=1}^{11} x_k\, y_k\right) - \overline{x}\,\overline{y} = 14.36$

$\rho(x,y) = \dfrac{Cov(x,y)}{\sqrt{V(x), V(y)}} = 0.955 \simeq 1$. $a = \dfrac{Cov(x,y)}{V(x)} = 1.44$ et $b = \overline{y} - a \times \overline{x} = 9.27$

2) L'ajustement est correct car $\rho(x,y) = 0.955 \simeq 1$ est trés proche de 1.

$a = \dfrac{Cov(x,y)}{V(x)} = 1.44$, $b = \overline{y} - a \times \overline{x} = 9.27$ et $y = a\,x + b = 1.44\,x + 9.27$

3)

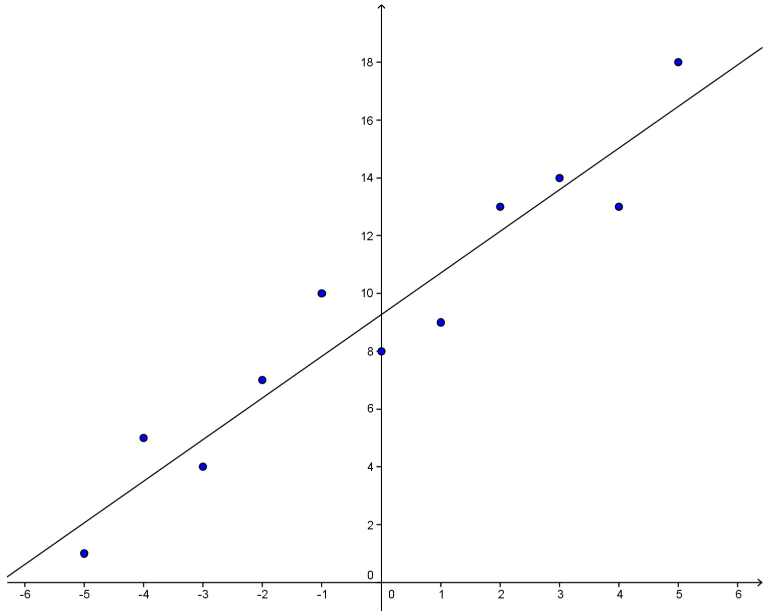


FIGURE 4 – nuage de points et droites de régression

<u>Ex 15 :</u>

1) Le tableau de la série est :

$]x_i\,,\,x_{i+1}]$	n_i	f_i	$F(x_i)$	c_i	$n_i \times c_i$	$n_i \times c_i^2$
$[2\,,\,4]$	5	0.05	0.05	3	15	45
$]4\,,\,6]$	8	0.08	0.13	5	40	200
$]6\,,\,8]$	15	0.15	0.28	7	105	735
$]8\,,\,10]$	20	0.2	0.48	9	180	1620
$]10\,,\,12]$	25	0.25	0.73	11	275	3025
$]12\,,\,14]$	12	0.12	0.85	13	156	2028
$]14\,,\,18]$	10	0.1	0.95	16	160	2560
$]18\,,\,20]$	5	0.05	1	19	95	1805
Total	$N = 100$	1			1026	12018

La classe du 1^{er} quartile est $]6\,,\,8]$,

$$Q_1 = x_i + (x_{i+1} - x_i)\frac{0.25 - F(x_i)}{F(x_{i+1}) - F(x_i)}\ ,\ F(x_i) = 0.13 \text{ et } F(x_{i+1}) = 0.28$$

$$\Longrightarrow Q_1 = 6 + 2 \times \frac{0.25 - 0.13}{0.28 - 0.13} = 7.6$$

La classe de la médiane est $]10\,,\,12]$,

$$m_e = x_i + (x_{i+1} - x_i)\frac{0.5 - F(x_i)}{F(x_{i+1}) - F(x_i)}\ ,\ F(x_i) = 0.48 \text{ et } F(x_{i+1}) = 0.73$$

$$\Longrightarrow m_e = 10 + 2 \times \frac{0.5 - 0.48}{0.73 - 0.48} = 10.16$$

La classe du $3^{\mathbf{er}}$ quartile est $]12\,,\,14]$,

$$Q_3 = x_i + (x_{i+1} - x_i)\frac{0.75 - F(x_i)}{F(x_{i+1}) - F(x_i)}\ ,\ F(x_i) = 0.73 \text{ et } F(x_{i+1}) = 0.85$$

$\Longrightarrow Q_3 = 12 + 2 \times \dfrac{0.75 - 0.73}{0.85 - 0.73} = 12.33333$

2) $\overline{X} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{n} n_i c_i = \dfrac{1026}{100} = 10.26$

$V(X) = \left(\dfrac{1}{N}\displaystyle\sum_{i=1}^{n} n_i c_i^2\right) - \overline{X}^2 = \dfrac{12018}{100} - (10.26)^2$

$= 120.18 - 105.2676 = 14.9124$

$\sigma = \sqrt{V(X)} = \sqrt{14.91} = 3.86$

3) Le tableau de cette distribution est :

classe	f_i	$F(x)$	c_i	$f_i \times c_i$	$c_i^2 \times f_i$
$[2\,,\,Q_1]$	0.25	0.25	$(2+Q_1)/2 = 4.8$	1.2	5.76
$]Q_1\,,\,m_e]$	0.25	0.5	$(Q_1+m_e)/2 = 8.88$	2.22	19.71
$]m_e\,,\,Q_3]$	0.25	0.75	$(m_e+Q_3)/2 = 11.245$	2.81	31.61
$]Q_3\,,\,20]$	0.25	1	$(Q_3+20)/2 = 16.165$	4.04	65.33
Total	1			10.27	122.41

$\overline{X1} = \displaystyle\sum_{i=1}^{4} C_i\, f_i = 10.27$ $V(X1) = \sum_{i=1}^{4} C_i^2\, f_i - \overline{X1}^2 = 122.41 - 10.27^2 = 16.94$, $\sigma_1 = \sqrt{16.94} = 4.11$

b)$\overline{X1} \simeq \overline{X} =$, $V(X1) \simeq V(X)$ et $\sigma \simeq \sigma_1$

4) $\overline{X} = 0.25 \times \left(\dfrac{a_0 + Q_1}{2} + \dfrac{Q_1 + m_e}{2} + \dfrac{m_e + Q_3}{2} + \dfrac{Q_3 + a_1}{2}\right)$

$= 0.25 \times \left(\dfrac{a_0 + a_1}{2} + Q_1 + m_e + Q_3\right).$

CFD17 :1) Le tableau des fréquences et des fréquences cumulées

Classe de notes	$]0,\,4]$	$]4,\,a]$	$]a,\,b]$	$]b,\,17]$	$]17,\,18.5]$
Fréquences f_i	0.125	0.225	0.375	0.175	0.1
Fréquencescumulées F_i	0.125	0.35	0.725	0.9	1

2) La classe du premier quartile est $]4,\,a]$.

$$\begin{aligned} Q_1 &= x_i + (x_{i+1} - x_i)\frac{0.25 - F_i}{F_{i+1} - F_i} \\ &= 4 + (a-4)\frac{0.25 - 0.125}{0.35 - 0.125} \\ &= 4 + (a-4)\frac{0.125}{0.225} = 7.33334 \\ \Longrightarrow\quad & a = \frac{0.225(7.33334 - 4)}{0.125} + 4 = 10.000012 \end{aligned}$$

d'où $a = 10$.

3)$\dfrac{b+a}{2} = \dfrac{b+10}{2} = 12 \Longrightarrow b = 14.$

α) x n'est autre que le troisième quartile. la classe du troisième est $]14,\,17]$.

$$\begin{aligned} Q_3 &= x_i + (x_{i+1} - x_i)\frac{0.75 - F - i}{F_{i+1} - F_i} \\ &= 14 + (17 - 14)\frac{0.75 - 0.725}{0.90 - 0.725} \\ &= 14 + 3 \times \frac{0.025}{0.175} = 14.42857 \end{aligned}$$

d'où $Q_3 = 14,4286$

β) $\overline{x} = \sum_{k=1}^{n} f_k \, x_k = (0.125 \times 2 + 0.225 \times 7 + 0.375 \times 12 + 0.175 \times 15.5 + 0.1 \times 17.75) = 10.8125.$

$V(x) = \sum_{k=1}^{n} f_k \, x_k^2 - \overline{x}^2 = (0.125 \times 2^2 + 0.225 \times 7^2 + 0.375 \times 12^2 + 0.175 \times 15.5^2 + 0.1 \times 17.75^2) -$
$10.8125^2 = 22.1648.$

$\sigma(X) = \sqrt{V(x)} = \sqrt{22.1648} = 4.7080$

γ) Calculons la médiane. la classe médiane est [10, 14].

$$\begin{aligned} Q_2 &= x_i + (x_{i+1} - x_i)\frac{0.5 - F_i}{F_{i+1} - F_i} \\ &= 10 + (14 - 10)\frac{0.5 - 0.35}{0.725 - 0.35} \\ &= 10 + 4 \times \frac{0.15}{0.375} = 11.6 \end{aligned}$$

D'après le coefficient d'asymétrie de Pearson, on a :

$$A_{P1} = 3 \times \frac{\overline{x} - Q_2}{\sigma} = 3 \times \frac{10.8125 - 11.6}{4.7080} = -0.5018 < 0.$$

D'autre part d'après le coefficient d'asymétrie de Yule, on a :

$$A_Y = \frac{Q_1 + Q_3 - 2 \times Q_2}{Q_3 - Q_1} = \frac{7.33334 + 14.2886 - 2 \times 11.6}{14.2886 - 7.33334} = -0.2269 < 0$$

D'où la distribution est asymétrique à gauche.

4) $V(x) = \sum_{k=1}^{5} n_k \; c_k^2 - \overline{x}^2 \implies N = \frac{\sum_{k=1}^{5} n_k \; c_k^2}{V(x) + \overline{x}^2} = \frac{11126}{22.1648 + 10.8125^2} = 80.00002.$D'où $N = 80$