

Statistiques

I) Le vocabulaire des statistiques:

1. Définition 1 :

- ❖ Une étude statistique porte sur un ensemble (de personnes, d'animaux, d'objets, ...) appelé **population**. Chaque élément de la population est un **individu**. L'aspect étudié est nommé **caractère** ou **variable**. Les résultats obtenus après observation donnent une **série statistique**. Il existe des séries à **une** ou **plusieurs** variables.
- ❖ Lorsque les variables prennent des valeurs numériques (exemple : notes, tailles, âges, ...), les variables sont dites **quantitatives** (si la variable prend n'importe quelle valeur dans un intervalle donné, la variable est dite **continue** , si elle prend des valeurs isolées, la variable est dite **discrète**). Dans le cas contraire, les variables sont dites **qualitatives** (nationalité, couleurs, ...) ; les différentes possibilités du caractère sont appelées **modalités** .
- ❖ L'**effectif** d'une « valeur » d'un caractère est le nombre d'individus ayant cette valeur.

2. Définition 2 :

Une série statistique est définie par la donnée d'un tableau statistique :

- ❖ D'une population déterminée
- ❖ D'un caractère (x_i)

Un caractère peut être quantitatives ou qualitatif selon qu'il est mesurable ou non

- ❖ De l'effectif (n_i) (ou de la fréquence (F_i)) de chaque valeur (ou classe ou modalité) du caractère
- Donc l'ensemble des couples (x_i, n_i) (ou (x_i, F_i)) s'appelle une série statistique à un seul caractère.

- Série statistique
- 1) Série statistique à caractère quantitatif : (mesurable)
 - a) Série statistique à caractère quantitatif continue (ou regroupé par classe)
 - b) Série statistique à caractère quantitatif discret (ou discontinue)
 - 2) Série statistique à caractère qualitatif : (non mesurable)
 - 3) Série chronologique

1) Série statistique à caractère quantitatif :

a) Série statistique à caractère quantitatif discret (ou discontinue)

Exemple : Le tableau ci-dessous représente le nombre d'enfants par famille dans un quartier de 112 familles

Nombre d'enfants (x_i)	0	1	2	3	4	5	6
Effectifs (n_i)	6	20	40	24	15	5	2

b) Série statistique à caractère quantitatif regroupé par classe (ou continue)

Exemple : Soit le tableau ci-dessous donne la moyenne annuelle de 64 élèves de 2^{ème} année d'un établissement scolaire.

Moyenne (x_i)	[0,4[[4,8[[8,12[[12,16[[16,20[
Effectifs (n_i)	6	12	24	14	8

2) Série statistique à caractère qualitatif :

Exemple : Le tableau suivant est une série statistique à caractère qualitatif. Le tableau ci-dessous donne indique la production mondiale d'huile d'olive en 2002/2003 (en millier de tonnes).

Pays de monde	Portugal	Grèce	Turque	Tunisie	Espagne	Italie	Maroc	Syrie	Reste du monde
Effectifs (n_i)	36	468	252	72	1584	720	72	252	144

3) Série chronologique :

Définition : Une série chronologique est une série de valeurs provenant d'une même variable observée à des instants régulièrement espacés dans le temps.

Exemple : Le tableau ci-dessous indique les résultats du baccalauréat de l'année 2000 Jusqu'à l'année 2005 dans un établissement A.

Année (x_i)	2000	2001	2002	2003	2004	2005
Nombre de réussites (n_i)	350	767	840	590	880	1000

3. Paramètres d'un caractère statistique : On distingue deux types de paramètres sur un caractère statistique. Les paramètres d'un caractère statistique

- Les paramètres de position
- les paramètres de dispersion

i) Paramètres de position : Les paramètres de position

- Le mode
- La médiane
- La moyenne
- Quartile

ii) Paramètres de dispersion : Les paramètres de dispersion

- L'étendue
- L'écart type
- Variance

II) Série statistique simple :

I- Séries statistiques à une variable :

1. Exemples :

➤ Tableau 1 :

Dans un groupe de dix élèves, voici les notes à un devoir : 12, 4, 16, 16, 10, 7, 9, 12, 9, 12.

Cette série de note est une série statistique $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ avec x_i les valeurs prises par X, à chaque valeur x_i on associe un effectif n_i égale au nombre d'élèves ayant obtenu la note

x_i . La fréquence de la valeur d'une série est : $F_i = \frac{n_i}{N}$ avec N = la somme des effectifs (l'effectif total)

x_i (Notes)	4	7	9	10	12	16
n_i (Effectif)	1	1	2	1	3	2
f_i (Fréquence)	0,1		0,2	0,1		0,2
effectifs cumulés croissants	1		4		8	
$n_i x_i$	4		18	10		32
$n_i x_i^2$	16	49		100	432	

➤ **Tableau 2**

Le tableau suivant donne la durée (en heures) passée chaque jour devant la télévision pour 1000 personnes.

Durées (en heures)	[0;2[[2;4[[4;6[[6;8[[8;10[
C_i (centre de classe)	1		5		9
n_i (Effectif)	100	200	500	150	50
effectifs cumulés croissants	100		800		1000
$n_i C_i$		600		1050	450
$n_i C_i^2$	100		12500		4050

Définition 1 :

- ❖ La première série est une série.....
- ❖ La deuxième série est une série.....

Les paramètres de position :

1. Mode (ou classe modale)

On appelle Mode ou classe Modale la valeur de X correspondant à l'effectif le plus haut.

- Pour la série quantitative discrète le Mode est :
- Pour la série quantitative continue la classe Modale est :

2. Moyenne (X) :

➤ **Tableau 1** : La moyenne est : $\bar{x} = \dots\dots\dots$

La moyenne est On appelle moyenne de cette série le nombre : $\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N} = \sum_{i=1}^n \frac{n_i x_i}{N}$ ou

N : l'effectif total $N = n_1 + n_2 + n_3 + \dots + n_p$

➤ **Tableau 2** : La moyenne est : $\bar{x} = \dots\dots\dots$

Pour calculer la moyenne , on utilise la formule précédente en remplaçant x_i par le centre c_i de

l'intervalle $[a_i , a_{i+1}[$. La moyenne de cette série est alors : $\bar{x} = \frac{n_1c_1 + n_2c_2 + \dots + n_pc_p}{N} = \sum_{i=1}^n \frac{n_i c_i}{N}$ ou

$$c_i = \frac{a_i + a_{i+1}}{2}$$

3. Médiane d'une série statistique : On note (M_e)

On appelle médiane d'une série statistique d'effectif total N la valeur de variable qui correspond a

l'effectif cumule $\frac{N}{2}$ si N est pair et $\frac{N+1}{2}$ si N est impair .

➤ **Tableau 1** =

$N = \dots$ est Donc la médiane est la valeur de variable qui correspond à l'effectif cumulé $\frac{N}{2}$ alors $M_e = \dots$

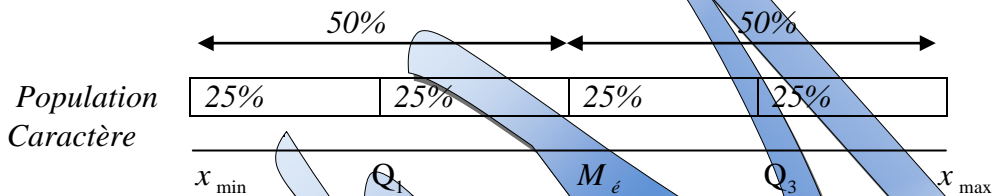
➤ **Tableau 2** =

$\frac{N}{2} = 500 ; 300 < 500 < 800$ alors $\langle M_e \rangle$

alors $\frac{M_e - 4}{500 - 300} = \frac{6 - 4}{800 - 300}$ D'où $M_e = \dots$

4. Quartiles :

❖ La médiane partage la série en deux groupes de même effectif. Les quartiles partagent la série en quatre groupes de même effectif. Ils sont donc au nombre de trois Q_1, Q_2 et Q_3 ou Q_2 est la médiane, les intervalles qu'ils définissent contiennent chacun 25% des observations, soit un quart de l'effectif, $\frac{N}{4}$ comme le montre le schéma :



❖ 2^{ème} méthode: Au moyen d'une interpolation linéaire en utilisant le tableau des effectifs cumulés.

Cas d'une série à caractère discret

- ❖ Q_1 : La valeur du caractère de rang $\frac{N}{4}$
- ❖ Q_2 : M_e
- ❖ Q_3 : La valeur du caractère de rang $\frac{3N}{4}$
- ❖ L'intervalle $[Q_1, Q_3]$ s'appelle intervalle interquartile
- ❖ Le réel $Q_3 - Q_1$ s'appelle l'écart interquartile

Remarque :

- ❖ Si N n'est pas multiple de 4 alors Q_1 (respectivement Q_3) est la valeur du caractère dont le rang est immédiatement supérieur à $\frac{N}{4}$ (respectivement $\frac{3N}{4}$).
- ❖ Pour Une série continue : Q_1 est la valeur correspondant à la fréquence cumulée croissante égale à 0,25.
- ❖ Q_3 est la valeur correspondant à la fréquence cumulée croissante égale à 0,75.

➤ **Tableau 1** =

- $\frac{N}{4} = 2,5$ alors $Q_1 = \dots$
- $\frac{3}{4}N = 7,5$ alors $Q_3 = \dots$

➤ Tableau 2 =

• $\frac{N}{4} = 250$; $100 < 250 < 300$ alors • $< Q_1 < \cdot$

alors $\frac{Q_1 - 2}{250 - 100} = \frac{4 - 2}{300 - 100}$ D'où $Q_1 = \dots\dots\dots$

• $\frac{3}{4}N = 750$; $300 < 750 < 800$ alors • $< Q_3 < \cdot$

alors $\frac{Q_3 - 4}{750 - 300} = \frac{6 - 4}{800 - 300}$ D'où $Q_3 = \dots\dots\dots$

➤ Tableau 1: L'écart interquartile est :

➤ Tableau 2: L'écart interquartile est :

Les paramètres de dispersion :

1. Etendue : L'étendue d'une série statistique est la différence entre ses deux valeurs extrêmes (la plus grande et la plus petite valeur du caractère)

➤ Tableau 1 = L'étendue est :

➤ Tableau 2 = L'étendue est :

2. Variance :

Cas d'une série à caractère discret : Quand la série statistique est discrète, on a un tableau du type :

Valeur du caractère (x_i)	x_1	x_2	x_3	... x_p	Total
Effectifs (n_i)	n_1	n_2	n_3	... n_p	N

La variance de la série est le nombre noté V et défini par

$$V = \frac{n_1(x_1)^2 + n_2(x_2)^2 + \dots + n_p(x_p)^2}{N} - (\bar{x})^2$$
 ou N : l'effectif total $N = n_1 + n_2 + n_3 + \dots + n_p$

Cas d'une série à caractère continue : on a un tableau du type :

Classe du caractère (x_i)	$[a_1, a_2[$	$[a_2, a_3[$	$[a_3, a_4[$... $[a_p, a_{p+1}[$
centre de l'intervalle	$c_1 = \frac{a_1 + a_2}{2}$	$c_2 = \frac{a_2 + a_3}{2}$	$c_3 = \frac{a_3 + a_4}{2}$... $c_p = \frac{a_p + a_{p+1}}{2}$
Effectifs (n_i)	n_1	n_2	n_3	... n_p

La variance de la série est le nombre noté V et défini par

$$V = \frac{n_1(c_1)^2 + n_2(c_2)^2 + \dots + n_p(c_p)^2}{N} - (\bar{x})^2$$
 ou N : l'effectif total $N = n_1 + n_2 + n_3 + \dots + n_p$

3. L'écart type :

L'écart type de la série est le nombre noté σ et défini par $\sigma = \sqrt{V}$

➤ Tableau 1 = l'écart type σ et la variance V de cette série statistique :

❖ Variance V : $V(X) = \dots\dots\dots$

❖ Écart type : $\sigma(X) = \dots\dots\dots$

➤ Tableau 2 = l'écart type σ et la variance V de cette série statistique :

❖ Variance V : $V(X) = \dots\dots\dots$

❖ Écart type : $\sigma(X) = \dots\dots\dots$

➤ **Remarque :** On peut obtenir N , \bar{x} , V et σ directement à partir de la calculatrice

Utilisation des calculatrices

Série statistique à une variable:

A titre d'exemple, on donne le mode de fonctionnement d'une calculatrice.

Le tableau ci-dessous donne la répartition des élèves d'une classe selon leurs âges.

âges (x_i)	14	15	16	17	18
Effectifs (n_i)	4	15	6	10	5

Pour choisir le mode de fonctionnement en statistique appuyer sur

2ndF MODE 1

Pour entrer les données taper

x_i	STO	n_i	M^+
14	STO	4	M^+
15	STO	15	M^+
16	STO	6	M^+
17	STO	10	M^+
18	STO	5	M^+

La calculatrice affiche l'effectif total $N = 40$

Pour afficher la moyenne \bar{X} appuyer sur :

RCL \bar{X}

On obtient :

$\bar{X} = 15,925$

Pour afficher l'écart type σ appuyer sur :

RCL σ_x

On obtient :

$\sigma_x = 1,232629304$

Pour afficher la variance V appuyer sur :

RCL σ_x^2

On obtient :

$V = 1,519375$

Pour afficher l'effectif total N appuyer sur :

RCL n

On obtient :

$N = 40$

4. Diagrammes en boîte ou boîte de dispersion (boîtes à moustaches)

Un diagramme en boîte est un rectangle délimité par le premier quartile et le troisième quartile.

Pour l'obtenir, on trace un axe horizontal (ou vertical) sur lequel on place les valeurs de Q_1 , Q_3 et Me .

L'un des côtés du rectangle a pour longueur l'écart interquartile $Q_3 - Q_1$, l'autre est quelconque.

On complète ce diagramme en traçant deux traits horizontaux : l'un joignant Q_1 au minimum de la série et l'autre joignant Q_3 au maximum de la série

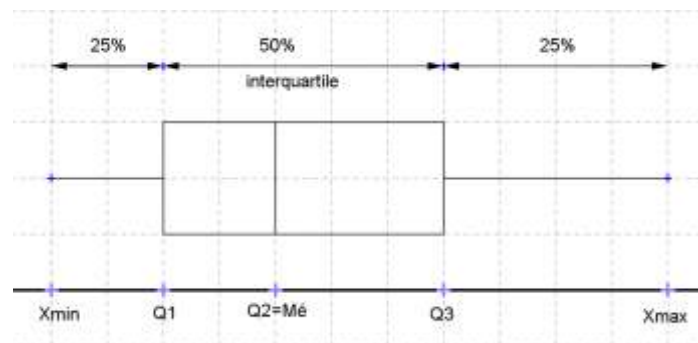


Tableau 1=



Tableau 2=



Activité N°1 :

Compléter les phrases ci –dessous en utilisant l'un de ces mots .

un histogramme , les individus, Le mode, L'étendue, une population , qualitatif, quantitatif , continue , discret , un diagramme en bâton ,

1. L'ensemble sur le quel porte une étude statistique s'appelle Les éléments de cet ensemble sont de la population
2. Un caractère est dit..... s'il n'est pas mesurable (exemples : couleur, moyen de transport....).
3. Un caractère est dit..... s'il est mesurable(exemples : Nombre d'enfant, poids, taille,...)
4. Graphiquement, on représente une série statistique à caractère quantitatif , par et on représente une série statistique à caractère qualitatif , par
5.est une valeur du caractère x_i (ou modalité) qui correspond à l'effectif le plus grand.
6.d'une série statistique est la différence entre ses deux valeurs extrêmes (la plus grande et la plus petite valeur du caractère)

Activité N°2 : (Cas d'une série à caractère discret)

Le tableau suivant indique la répartition de 128 appartements d'une cité suivant le nombre de leur pièces

Nombre des pièces (x_i)	1	2	3	4	5	6
Nombre d'appartement (n_i)	10	34	50	24	8	2

- 1) Trouver le mode, l'étendue et la médiane de cette série
- 2) Représenter le diagramme en bâton à effectifs de cette série
- 3) Donner les effectifs cumulés croissants de cette série
- 4) a) Calculer les fréquences cumulées croissantes de cette série
a) représenter le polygone des fréquences cumulées croissantes
- 5) Calculer la moyenne \bar{x} de cette série
- 6) Calculer l'écart type σ et la variance V de cette série statistique
- 7) Déterminer les quartiles de cette série
- 8) Représenter la boîte de dispersion de cette série

Activité N°3 : (Cas d'une série à caractère continu)

Un grossiste a relevé le montant des factures de livraison reçue au cours d'un mois.

Montant des factures (En dinars)	[0,100[[100,200[[200,300[[300,400[[400,500[[500,600[
Nombre des factures	10	20	18	14	12	6

- 1) Tracer le histogramme de cette série
- 2) a) Donner sa classe modale
b) Déterminer pour cette série le centre de chaque classe.
c) Calculer la moyenne \bar{x} de cette série
- 3) Calculer l'écart type σ et la variance V de cette série statistique
- 4) Calculer les fréquences cumulées croissantes de cette série
- 5) a) Donner les effectifs cumulés croissants de cette série
b) représenter le polygone des effectifs cumulés croissants de cette série
c) déduire graphiquement la médiane de cette série
- 6) Calculer par la méthode d'interpolation linéaire les quartiles Q_1 , Q_2 et Q_3 de cette série
- 7) Représenter la boîte de dispersion de cette série

CORRECTION

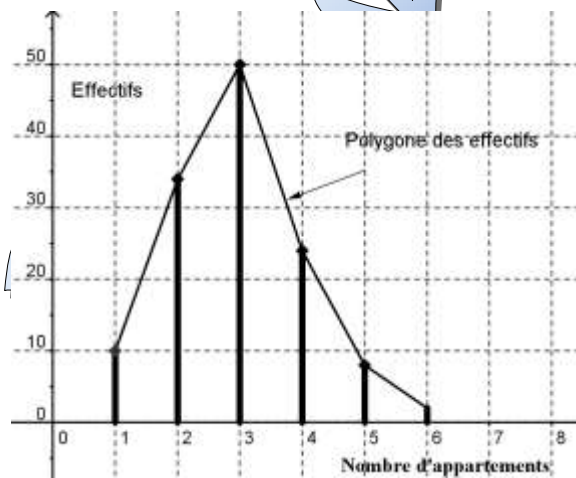
Activité N°1 :

- 1) L'ensemble sur le quel porte une étude statistique s'appelle une **population**. Les éléments de cet ensemble sont les individus de la population
- 2) Un caractère est dit **qualitatif**, s'il n'est pas mesurable (exemples : couleur, moyen de transport....).
- 3) Un caractère est dit **quantitatif** s'il est mesurable (exemples : Nombre d'enfant, poids, taille,...)
- 4) Graphiquement, on représente une série statistique à caractère quantitatif continue, par un **histogramme** et on représente une série statistique à caractère quantitatif discret, par un **diagramme en bâton**.
- 5) Le **mode** est une valeur du caractère x_i (ou modalité) qui correspond à l'effectif le plus grand.
- 6) L'**étendue** d'une série statistique est la différence entre ses deux valeurs extrêmes (la plus grande et la plus petite valeur du caractère)

Activité N°2 : (Cas d'une série à caractère discret)

- 1)
 - a) Le mode de cette série est : $M_o = 3$
 - b) L'étendue de cette série est : $6-1=5$
 - c) $N = 128$ est pair alors la valeur moyenne des deux valeurs du caractère des deux individus de rang $\frac{N}{2} = \frac{128}{2} = 64$ et $\frac{N}{2} + 1 = 65$. Donc la médiane $M_e = \frac{3+3}{2} = 3$

- 2) Le diagramme en bâton à effectifs de cette série



- 3) Les effectifs cumulés croissants de cette série

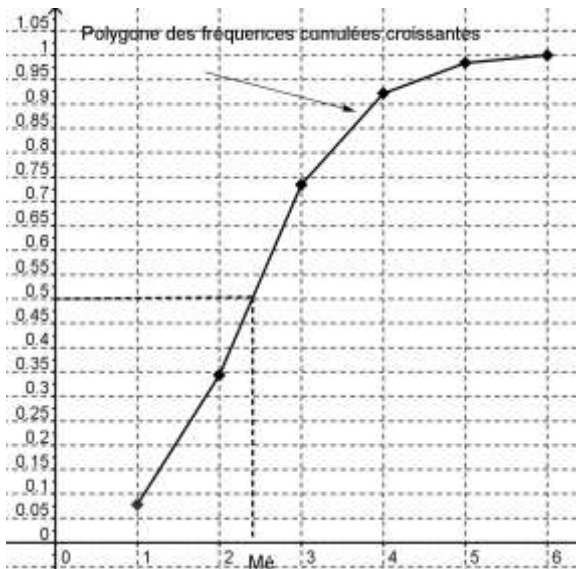
Nombre de pièces (x_i)	1	2	3	4	5	6	Total
Nombre d'appartement (n_i)	10	34	50	24	8	2	$N=128$
Effectifs cumulés croissants	10	44	94	118	126	128=N	

CORRECTION

4) a) Les fréquences cumulées croissantes de cette série

Nombre des pièces (x_i)	1	2	3	4	5	6
Nombre d'appartements (n_i)	10	34	50	24	8	2
Fréquences F_i	0,078125	0,265625	0,390625	0,1875	0,0625	0,015625
Fréquences cumulées croissantes	0,078125	0,34375	0,734375	0,921875	0,984375	F = 1

Le polygone des fréquences cumulées croissantes correspondant



la médiane est l'abscisse du point de la courbe dont l'ordonnée est égale à **0,5**
donc $M_e = 3$

5) Moyenne : $\bar{x} = 2,9375$

6) Écart type : $\sigma = 1,073472752$

Variance V : $V = 1,15234375$

7) Les quartiles de cette série :

Q_1 : La valeur du caractère de l'individu de rang $\frac{N}{4} = \frac{128}{4} = 32$ ($n_i \nearrow =$ Effectif cumulés croissante)

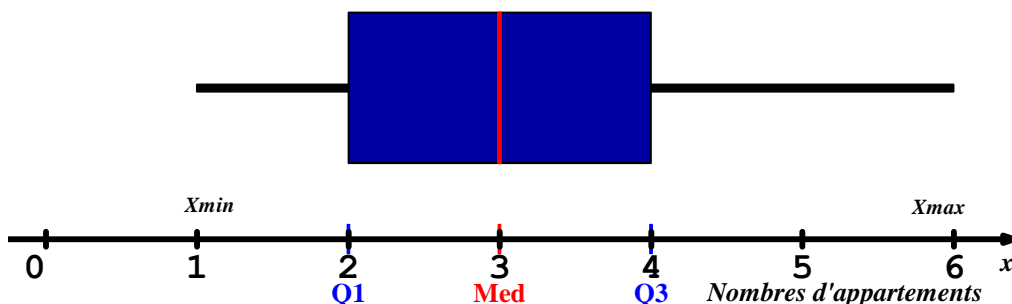
donc $Q_1 = 2$

$Q_2 = M_e = 3$

Q_3 = La valeur du caractère de l'individu de rang $\frac{3N}{4} = 96$ ($n_i \nearrow =$ Effectif cumulés croissante) donc

$Q_3 = 4$

8) Le diagrammes en boîte ou la boîte de dispersion de cette série :



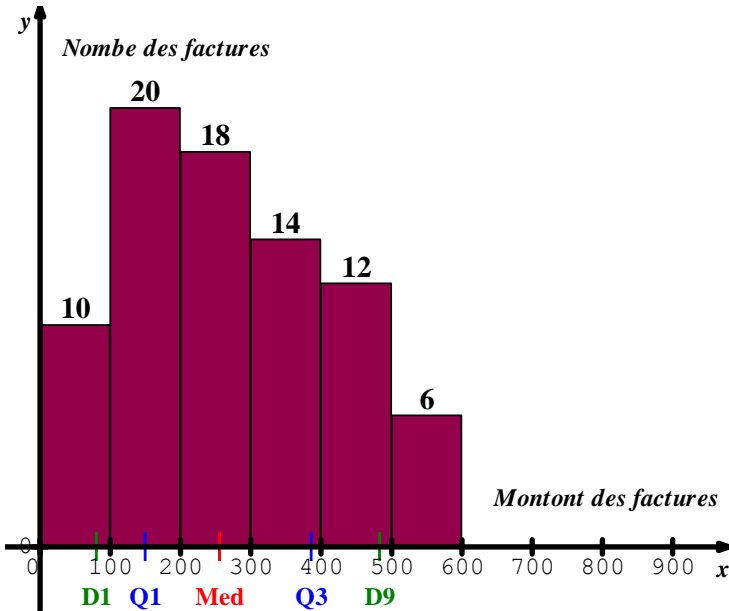
CORRECTION

Activité N°3 : (Cas d'une série à caractère continue)

Un grossiste a relevé le montant des factures de livraison reçue au cours d'un mois.

Montant des factures (En dinars)	[0,100[[100,200[[200,300[[300,400[[400,500[[500,600[
Nombre des factures	10	20	18	14	12	6

1. Le histogramme de cette série



2. a) La classe modale de cette série est [100,200[donc le mode de cette série est

$$M_o = \frac{100 + 200}{2} = 150$$

b) Le centre de chaque classe :

Montant des factures (En dinars)	[0,100[[100,200[[200,300[[300,400[[400,500[[500,600[
Nombre des factures	10	20	18	14	12	6
centre de l'intervalle c_i	50	150	250	350	450	550

c) La moyenne \bar{x} de cette série :

Moyenne : $\bar{x} = 270,063$; calculatrice : c_i STO n_i M⁺

3. l'écart type σ et la variance V de cette série statistique :

Écart type : $\sigma = 146,876$

Variance V : $V = 2157,55938$

4. Les fréquences cumulées croissantes de cette série

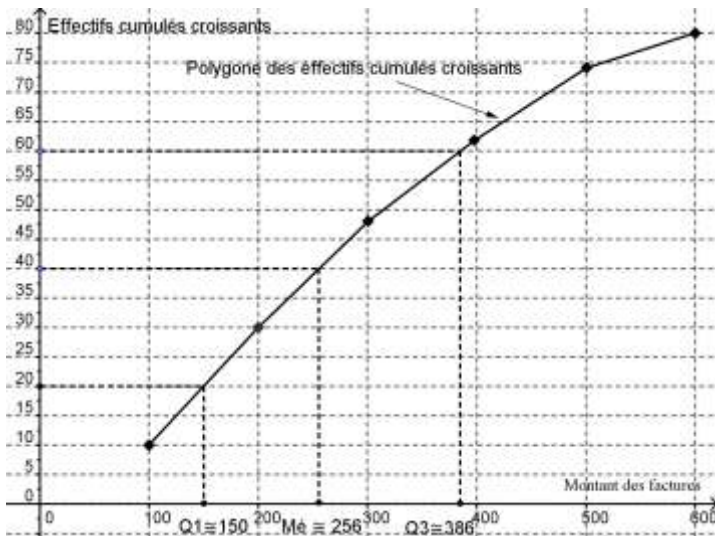
Montant des factures (En dinars)	[0,100[[100,200[[200,300[[300,400[[400,500[[500,600[Total
Nombre des factures	10	20	18	14	12	6	N=80
Fréquences F_i	0,125	0,250	0,225	0,175	0,150	0,075	
Fréquences cumulées croissantes	0,125	0,375	0,6	0,775	0,925	F = 1	

CORRECTION

5. a) Les effectifs cumulés croissants de cette série

Montant des factures (En dinars)	[0,100[[100,200[[200,300[[300,400[[400,500[[500,600[Total
Nombre des factures	10	20	18	14	12	6	N=80
Effectifs cumulés croissants	10	30	48	62	74	N= 80	

b) Le polygone des effectifs cumulés croissants de cette série



Le polygone des effectifs cumulés croissants correspondant
La moitié de l'effectif total est $\frac{N}{2} = \frac{80}{2} = 40$
d'après le graphique, la médiane est l'abscisse du point de la courbe dont l'ordonnée est égale à 40
donc le médiane $M_e \approx 255$

6. Les quartiles de cette série :

Graphiquement : sur le polygone des effectifs cumulés croissants

- ❖ Le point d'ordonnée est d'abscisse $\frac{N}{4} = \frac{80}{4} = 20$ est d'abscisse Q_1 , on lit $Q_1 \approx 150$
- ❖ Le point d'ordonnée est d'abscisse $\frac{N}{2} = \frac{80}{2} = 40$ est d'abscisse Q_2 , on lit $Q_2 = M_e \approx 256$
- ❖ Le point d'ordonnée est d'abscisse $\frac{3N}{4} = 60$ est d'abscisse Q_3 , on lit $Q_3 \approx 386$

Calcul des quartiles par interpolation linéaire : $N = 80$

- ❖ Q_1 appartient à la classe du caractère de l'individu de rang $\frac{N}{4} = 20$ ($n_i \nearrow$)

donc $Q_1 \in [100, 200[$

x_i	$n_i \nearrow$
100	10
$Q_1 = ?$	$\frac{N}{4} = 20$
200	30

$$\frac{Q_1 - 100}{20 - 10} = \frac{200 - 100}{30 - 10} = 5 \quad \text{donc } Q_1 = 150$$

- ❖ Q_2 appartient à la classe du caractère de l'individu de rang $\frac{N}{2} = 40$ ($n_i \nearrow$)

donc $Q_2 \in [200, 300[$

CORRECTION

x_i	$n_i \nearrow$
200	30
$Q_2 = ?$	$\frac{N}{2} = 40$
300	48

$$\frac{Q_2 - 200}{40 - 30} = \frac{300 - 200}{48 - 30} = \frac{100}{18} \quad \text{donc} \quad Q_2 = \frac{1000}{18} + 200 \approx 256$$

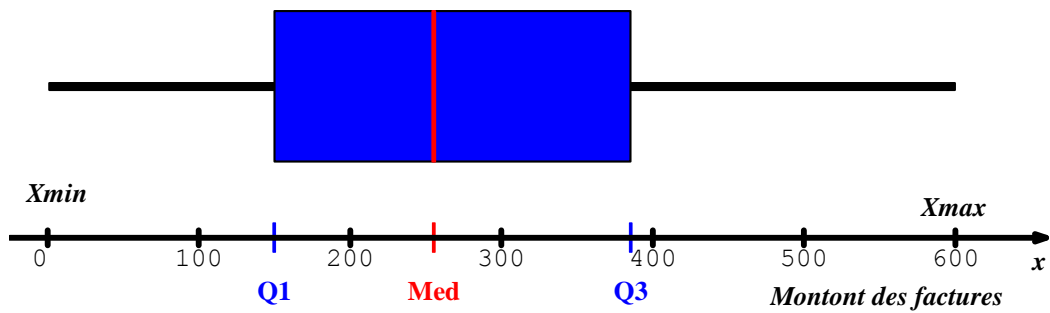
❖ Q_3 appartient à la classe du caractère de l'individu de rang $\frac{3N}{4} = 60$ ($n_i \nearrow$)
 donc $Q_3 \in [300, 400[$

x_i	$n_i \nearrow$
300	48
$Q_3 = ?$	$\frac{3N}{4} = 60$
400	62

$$\frac{Q_3 - 300}{60 - 48} = \frac{400 - 300}{62 - 48} = \frac{100}{14} \quad \text{donc} \quad Q_3 = \frac{1200}{14} + 300 \approx 386$$

7. Le diagrammes en boîte ou la boîte de dispersion de cette série :

- Effectif total : 80
- Minimum : 0
- Maximum : 600
- Premier quartile : $Q_1 = 150$
- Médiane : $Q_2 = 256$
- Troisième quartile : $Q_3 = 386$
- Intervalle interquartile : $[150, 386]$



*Naitar*medyassine*