Chapitre 3

Étude d'une variable statistique continue

Nous rappelons qu'une variable statistique (V.S) quantitative concerne une grandeur mesurable. Ses valeurs sont des nombres exprimant une quantité et sur lesquelles les opé- rations arithmétiques (addition, multiplication, etc,...) ont un sens. Nous allons dans ce chapitre se focaliser sur la V.S quantitative continue.

3.1 Caractère continu

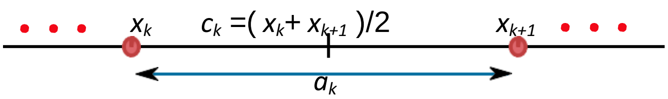
Définition 11
On appelle V.S continue (ou caractère continu) toute application de Ω *et à valeurs réelles et qui prend un nombre "important" de valeurs (Les caractères continus sont ceux qui ont une infinité de modalités).*

3.1.1 Classe de valeurs

Définition 12
On appelle classe de valeurs de X *un intervalle de type* $[a, b[$ *tel que* $X \in [a, b[$ *si et seulement si* $a \leq X(w) < b$*, c'est à dire, que les valeurs du caractère sont dans la classe* $[a, b[$.

Réponse : *Partager les valeurs prises par* X *en classes de valeurs.*

Dès qu'un caractère est identifié en tant que continu, ces modalités $C_k = [L_k, L_{k+1}[$ sont des intervalles avec



- x_k : borne inférieure.
- x_{k+1} : borne supérieure.
- $a_k = x_{k+1} - x_k$: son amplitude, son pas ou sa longueur.
- $C_k = (x_{k+1} + x_k)/2$: son centre.

3. Caractère quantitatif continu :

a. Histogramme - Polygone des fréquences :

Les modalités sont présentées sous forme de classes
[xi ; xi+1[, 1 ≤ i ≤ k .
Dans l'histogramme, chaque modalité [xi ; xi+1[du caractère est représentée par un rectangle dont la base ai est égale a` l'amplitude de la classe (ai = xi+1-xi) et dont la hauteur hi .

Remarque :

On distingue deux cas :

1er cas : Toutes les classes ont une même amplitude *a*.

Dans ce cas, on représente chaque modalité par un rectangle dont la hauteur est égale à l'effectif (ou à la fréquence) de la modalité. La base de tous les rectangles étant la même (égale à l'amplitude *a*).

Exemple3 : En mesurant la taille de 50 étudiants de la FST, on a obtenu les résultats suivants (en cm)

Taille (en cm)	Effectif n_i	Eff. cum. N_i	Fréquence f_i	Fréq. cum. F_i	...
[151; 155[	10	10	0.2	0.2	
[155; 159[	12	22	0.24	0.44	
[159; 163[	11	33	0.22	0.66	
[163; 167[	7	40	0.14	0.8	
[167; 171[	10	50	0.2	1	
Total	50		1		

- Population étudiée : Echantillon de 50 étudiants.
- Caractère étudié : La taille.
- Modalités : Le nombre des modalités observées étant élevé, il est donc nécessaire de les grouper en classes.
- Nature du caractère : Quantitatif continu. On peut

considérer les classes suivantes :

[151; 155[
[155; 159[
[159; 163[
[163; 167[
[167; 171[

Histogramme : Toutes les classes ont une même amplitude.

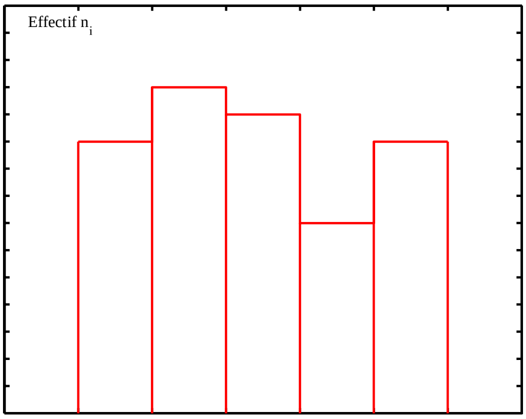


Fig. Histogramme

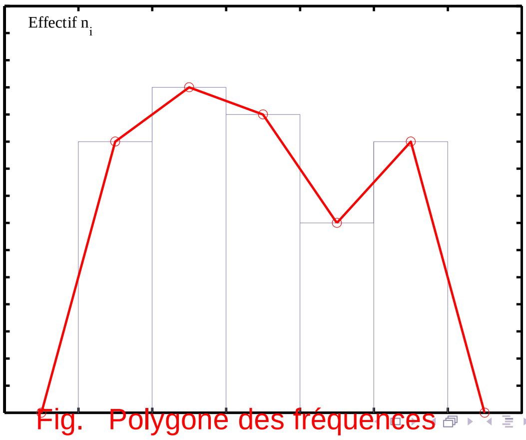


Fig. Polygone des fréquences

Polygone des fréquences :
Lorsque toutes les classes ont une même amplitude, le polygone des fréquences est construit en joignant par des segments de droites les milieux des côtés supérieurs des rectangles dans l'histogramme.

Les extrémités rejoignent l'axe des abscisses.

Polygone des fréquences : Toutes les classes ont une même amplitude a.

2ème cas :
Les classes n'ont pas toutes la même amplitude.Dans ce cas, on utilise les densités d'effectif

$$d = \frac{n_i}{a_i}$$

ou les densités $d^f = \frac{f_i}{ai}$

Les classes n'ont pas toutes la même amplitude :

Pour construire l'histogramme, chaque modalité $[x_i\,;\,x_{i+1}[$ est représentée par un rectangle dont la base est égale à l'amplitude a_i de la classe $[x_i\,;\,x_{i+1}[$ et dont la hauteur h_i est proportionnelle à la densité d'effectif d_i (ou à la densité de fréquence d_i^f) correspondante.

$$h_i = C \times \frac{n_i}{a_i}$$

ou

$$h_i = C \times \frac{f_i}{a_i}$$

C : est une constante de proportionnalité.

h_i : est appelée effectif corrigé (ou fréquence corrigée) de la modalité $[x_i\,;\,x_{i+1}[$.

Le choix de la constante C est arbitraire (on peut prendre $C = 1$ par exemple).

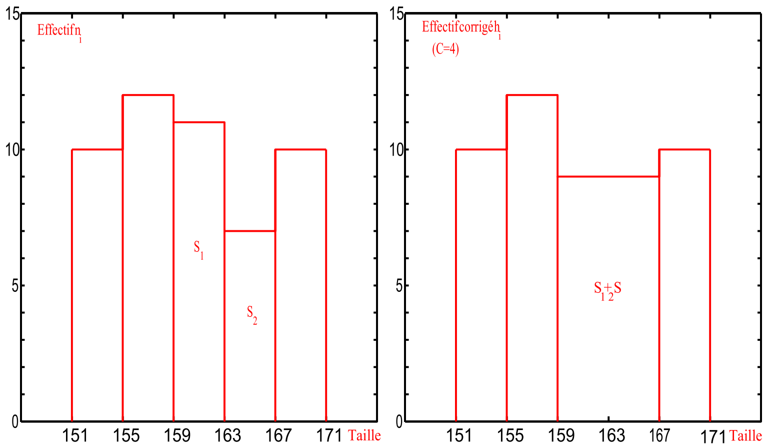
Cependant, pour simplifier les calculs et/ou pour tracer le polygone des fréquences, il faut choisir C égale au plus grand diviseur commun des amplitudes a_i .

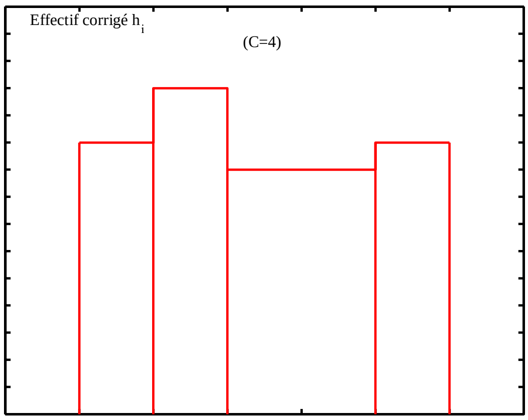
Exemple3 : Taille des étudiants (les classes n’ont pas toutes la même amplitude)

Taille (en cm)	a_i (en cm)	Effectif n_i	Eff. corrigé h_i	Fréq. f_i	Fréq. corrigée h_i
[151; 155[	4	10	10	0.2	0.2
[155; 159[	4	12	12	0.24	0.24
[159; 167[	8	18	9	0.36	0.18
[167; 171[	4	10	10	0.2	0.2
Total		50		1	

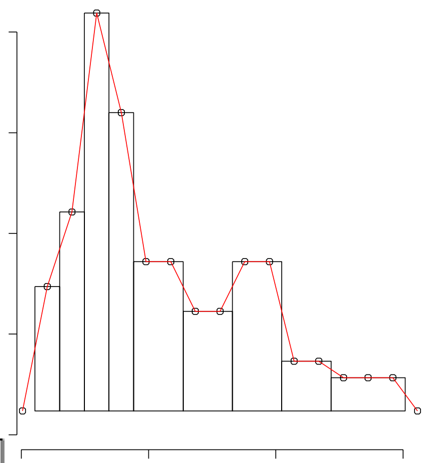
$C = PGDC(a_i) = 4$, $(h_i = C \times \frac{n_i}{ai})$

2ème cas : Histogramme avec effectif corrigé

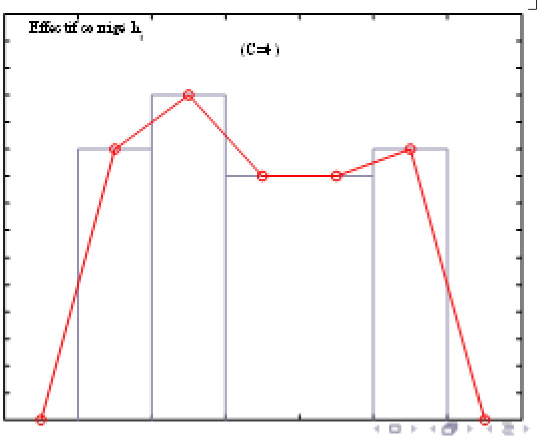




Polygone des fréquences : ´ les classes n’ont pas la meme ˆ amplitude.
Soit $C = PGCD(a_i)$.
Toutes les a_i sont des multiples de C ($a_i = k_i \times C$).
Pour construire le polygone des fréquences, on partage ´ chaque classe d’amplitude $a_i = k_i \times C$ en k_i classes de même amplitude C. Puis, on trace les milieux des sommets des rectangles de base C et de hauteur h_i et on joint ces milieux par des segments de droites.
Les extremit ´ es du polygone rejoignent l’axe des abscisses.



Polygone des fréquences :
Les classes n’ont pas la même amplitude



200

Exemple3 : Taille des étudiants

Taille (en cm)	Effectif n_i	Eff. cum. N_i	Fréquence f_i	Fréq. cum. F_i	...
[151; 155[	10	10	0.2	0.2	
[155; 159[	12	22	0.24	0.44	
[159; 163[	11	33	0.22	0.66	
[163; 167[	7	40	0.14	0.8	
[167; 171[	10	50	0.2	1	
Total	50		1		

1- Le mode

Définition

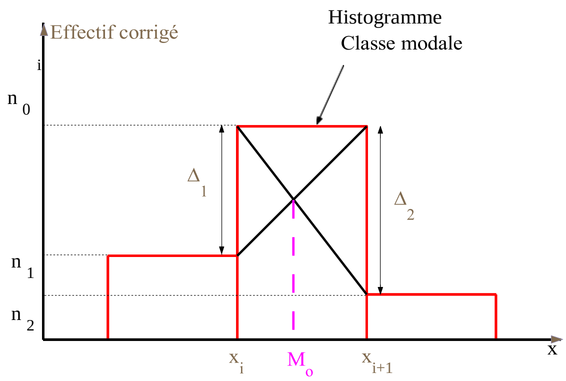
Le mode, noté M_o, est la valeur du caractère qui admet le plus grand effectif.
C'est la valeur du caractère la plus fréquente.

Remarque :

- Le mode peut être calculé pour tous les types de caractère.
- Pour un caractère continu, on définit d'abord la classe modale qui correspond à l'effectif corrigé le plus élevé.

On définit d'abord la classe modale qui correspond à l'effectif corrigé le plus élevé.
La classe modale correspond donc au maximum de l'histogramme.

La classe modale $[x_i , x_{i+1}[$ étant déterminée, On détermine la valeur du mode M_o en tenant compte des effectifs corrigés des deux classes adjacentes à la classe modale par la méthode suivante :



Définition

Nous définissions la classe modale comme étant la classe des valeurs de X qui a le plusgrand effectif partiel (ou la plus grande fréquence partielle). La quantité :

$\Delta 1$

$$M_o = x_i + a_i \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

s'appelle le mode avec (voir Figure)

– x_i *: la borne inférieure de la classe modale.*

– $a_i = x_{i+1} - x_i$: *le pas de la classe modale.*

– $\Delta_1 = n_0 - n_1$, $\Delta_2 = n_0 - n_2$ *ou bien* $\Delta_1 = f_0 - f_1$, $\Delta_2 = f_0 - f_2$.

– n_0 *et* f_0 *sont l'effectif et la fréquence associés à la classe modale.*

– n_1 *et* f_1 *sont l'effectif et la fréquence de la classe qui précède la classe modale.*

– n_2 *et* f_2 *sont l'effectif et la fréquence de la classe qui suit la classe modale.*

Exemple3 : Taille des étudiants

Taille (en cm)	Effectif n_i
[151; 155[	10
[155; 159[	12
[159; 163[	11
[163; 167[	7
[167; 171[	10
Total	50

Dans cet exemple les classes ont toutes la même amplitude, on n'a pas besoin d'utiliser les densités. On utilise les effectifs
La classe modale est : [155, 159[. On a

$$\Delta_1 = 12 - 10 = 2$$

$$\Delta_2 = 12 - 11 = 1$$

M_o = 155+2/(2+1)*4=   157.66 cm

Exemple3 : Taille des étudiants (les classes n'ont pas toutes la même amplitude)

Taille (en cm)	Effectif n_i	Amplitude a_i (en cm)	Eff. corrigé h_i
[151; 155[	10	4	10
[155; 159[	12	4	12
[159; 167[	18	8	9
[167; 171[	10	4	10
Total	50		

La classe modale est [155; 159[(c'est la classe qui admet la plus grande densité). On a

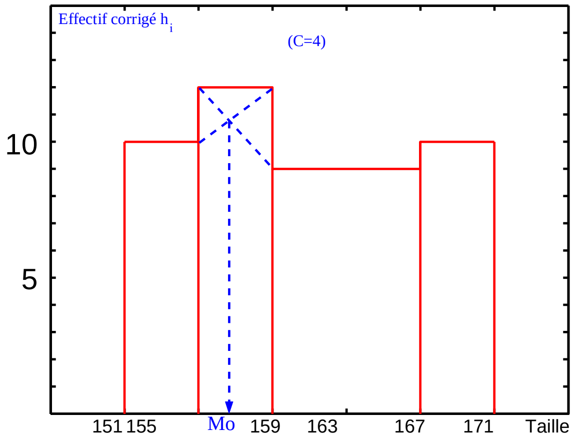
$$\Delta_1 = 12 - 10 = 2$$

$$\Delta_2 = 12 - 9 = 3$$

Donc

Mo = 155 + 2/(2+3)x4

Détermination graphique :



La médiane :

Définition

La médiane d'un caractère ordinal, notée M_e, est la valeur centrale de la série statistique ordonnée.

Cette définition n'a de sens que si les modalités sont toutes ordonnées par ordre croissant.

La classe médiane $[x_i, x_{i+1}[$ étant déterminée, la médiane

M_e vérifie

$x_i \longrightarrow N_{i-1}$

$M_e \longrightarrow N/2$

$x_{i+1} \longrightarrow N_i$

(Me – xi)/ (N/2 − Ni−1)= (xi +1 – xi) Ni − Ni−1

3.2 Paramètres de dispersion

Définition

La variance est la quantité

$$Var(x) = \sum_{i=}^{n} f_i(\bar{x} - C_i)^2.$$

Remarque

5*Pour le calcul, on utilise (voir Chapitre* 2*, Théorème*

$$Var(x) = \sum_{i=}^{n} f_i C_i^2 - \bar{x}^2.$$

Définition

La quantité

$$\sigma_x = \sqrt{Var(x)}$$

s'appelle l'écart type de la V.S X.

Nous généralisons la notion de la médiane dans la définition suivante.

Définition 24
Pour $i \in \{1,2,3\}$, *la quantité* Q_i *tel que* $F(Q_i) = \frac{i}{4}$ *s'appelle le* i^{em} *quartile.*

Exemple 19
Pour $i = 2$, Q_2 *tel que* $F(Q_2) = \frac{2}{4} = 0.5$. *Donc,* $Q_2 = Me$.

La détermination ou le calcul de Q_i se fait exactement comme le calcul de la médiane (graphiquement ou analytiquement).

Interprétation : Il y a 25 % d'individus dont la valeur du caractère est dans l'intervalle $[a_0, Q_1]$. De même pour les autres quartiles. Ces intervalles s'appellent "intervalles interquartiles".

$$Q_1 \longrightarrow 25\%,$$

$$Q_2 \longrightarrow 50\%,$$

$$Q_3 \longrightarrow 75\%.$$

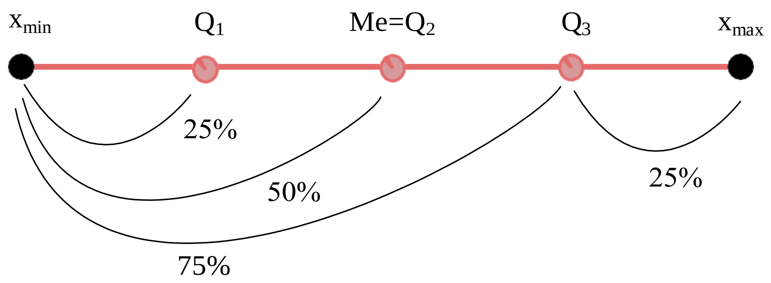


FIGURE : Les quartiles.