1 Chapitre 03 : Etude d'une variable statistique continue

On appelle V.S continue (ou caractère continu) toute application de et à valeurs réelles et qui prend un nombre "important" de valeurs (Les caractères continus sont

ceux qui ont une infinité de modalités).

Question : Comment étudier ce caractère ?

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

1.1 Classe de valeurs

Definition 1 *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 $[a, b[$.*

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

1. L_k : borne inférieure, L_{k+1} : borne supérieure.
2. **L'amplitude** de classe k notée a_k donnée par :
$$a_k = L_{k+1} - L_k$$
3. **Le centre** de classe k notée C_k donnée par :
$$C_k = \frac{L_k + L_{k+1}}{2}.$$
4. **Nombre de classes :** Soit N l'effectif total. Nous pouvons considérons dans ce chapitre trois formules suivant :

- $\sqrt{N}$, $[\sqrt{N}]$ (partie entière) ou $[\sqrt{N}] + 1$. Donc, le nombre de classes
$$k \simeq \sqrt{N}.$$
- La formule de sturge
$$k = 1 + 3.3 \log_{10}(N).$$
- La formule de yule
$$k = 2.5 \sqrt[4]{N}.$$

Remarque : On peut avoir plusieurs tableaux statistiques selon le nombre de classes.

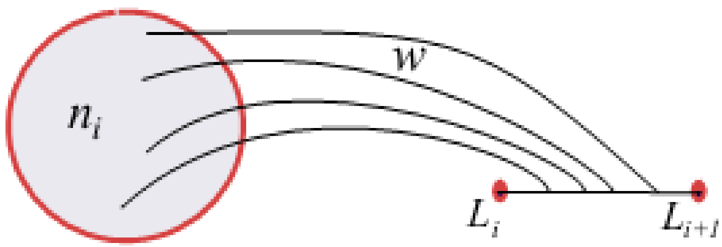
Nous rappellons maintenant la définition de effectif et fréquence d'une classe.

1.2 Effectif et fréquence d'une classe

Definition 2 *La quantité*

$$n_i = Card \{\omega \in \Omega \,:\, X(\omega) \in C_i\}\,.$$

S'appelle effectif partiel de C_i.



Nombre d'individus qui prennent des valeurs x_i dans C_i.

Definition 3 *Le nombre*

$$f_i = \frac{n_i}{N}$$

est appelé la fréquence partielle de C_i.

Definition 4 *On appelle l'effectif cumulé du C_i la quantité*

$$N_i = \sum_{j=1}^{i} n_j.$$

Definition 5 *On appelle la fréquence cumulé du C_i la quantité*

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

1.3 Répresentation graphique d'un caractère quantitatif continue :

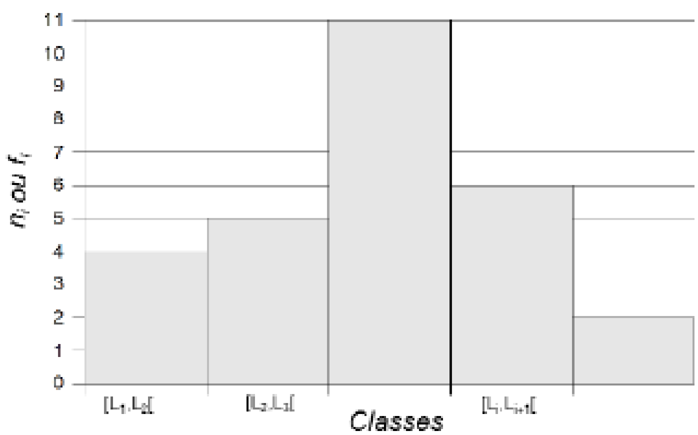
Comme dans le cas dans variable statistique discret la répresentation graphique d'un caractère quantitatif continue se presente sous deux formes (diagramme differentiel (Histogramme),diagramme integral).

Example 6 *Tableau statistique :*

Les classes	n_i	f_i	a_i	C_i	$d_i = \frac{n_i}{a_i}$	N_i	F_i
$[200, 400[$	45	0.225	200	300	0.225	45	0.225
$[400, 600[$	60	0.300	200	500	0.3	105	0.525
$[600, 800[$	57	0.285	200	700	0.285	162	0.810
$[800, 1000[$	23	0.115	200	900	0.115	185	0.925
$[1000, 1200[$	15	0.175	200	1100	0.075	200	1
Total	200	1					

1.3.1 Répresentation graphique : (Histogramme)

A chaque classe est associée un rectangledont la longeur est l'amplitude de la classe et dont la hauteur est $\dfrac{n_i}{a_i}$ où $\dfrac{f_i}{a_i}$.



Histogramme des fréquences ou des éffctifs.

1.3.2 Fonction de répartition

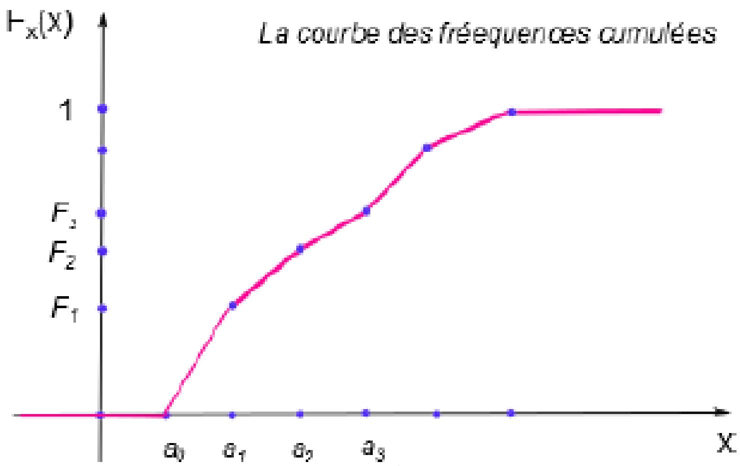
Definition 7 *La fonction $F_x : \mathbb{R} \to [0,1]$ définie par $F_x(x)$ représenter le pourcentage des individus tel que la valeur de leur caractère est inférieure ou égale à x. Elle est donnée par*

$$F_x(x) = \begin{cases} 0 & si \quad x \prec x_{\min} \\ F_1 & si \quad x_{\min} \prec x \prec x_{\min+1} \\ F_i & si \quad x_i \prec x \prec x_{i+1} \\ ... & \\ 1 & si \quad x \geq x_n \end{cases}$$

et elle s'appelle la fonction de répartition de X.

Nous expliquons cette formulation de la fonction de répartitionn dans cette remarque .

La courbe de F_x est nulle avant a_0, constant égale à 1 aprés a_n et joint les points $(a_0, 0), (a_1, F_1) \ldots\ldots (a_n, 1)$ par des segments de droites.



La courbe des fréquences cumulées.

1.3.3 Caractéristique de position centrales :

On note C_i le centre de classe C_i et nous considérons f_i la fréquence partielle.

1. **La moyenne arithmétique :**

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{m} n_i C_i = \sum_{i=1}^{m} f_i C_i.$$

(2) **Le mode**: La définition suivante permet de comprendre la démarche à suivre pour calculer le mode

d'une manière exacte et qui se trouve dans une des classes appelée "classe modale".

Definition 8 *Nous définissions la classe modale comme étant la classe des valeurs de X qui a le plus*

grand effectif partiel (où la plus grande fréquence partielle).
Dans l'example 5
La classe modale est $[400, 600[$ dans ce cas on peut prendre

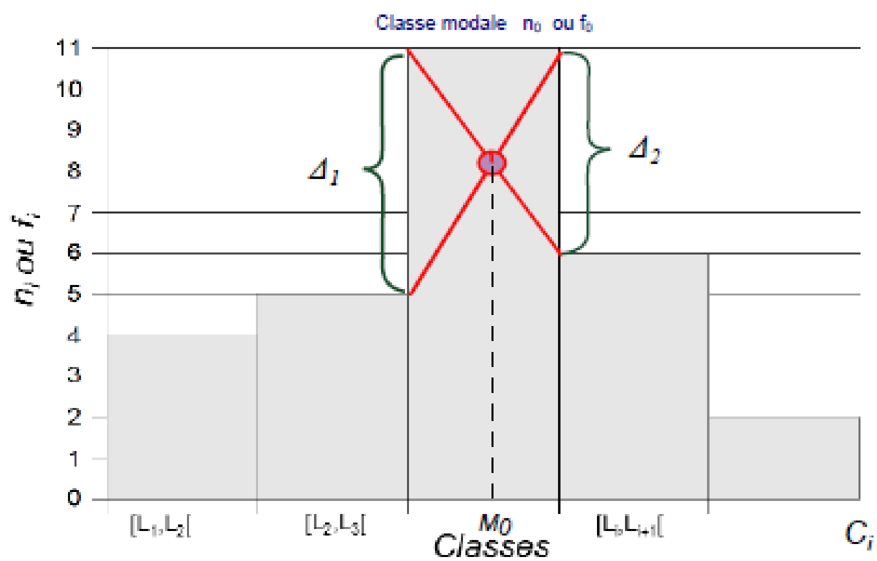
$$M_o = \frac{400 + 600}{2} = 500.$$

On peut aussi déterminer graphyquement le mode : La quantité

$$M_o = L_{i-1} + a_i * \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

Tel que :

- L_{i-1} : La borne inférieur de la classe modale.
- a_i : Le pas de classe modale.
- $\Delta_1 = n_0 - n_1; \Delta_2 = n_0 - n_2$ où bien $\Delta_1 = f_0 - f_1; \Delta_2 = f_0 - f_2$.
- $n_1$et $f_1$sont l'effectif et la fréquence de la classe qui précède la classe modale.



Représentation ou détermination graphique du mode (cascontinu).

Dans l'example 5 :

$$M_o = 400 + 200 * \frac{0.300 - 0.225}{(0.300 - 0.225) + (0300 - 0.285)} \simeq 567$$

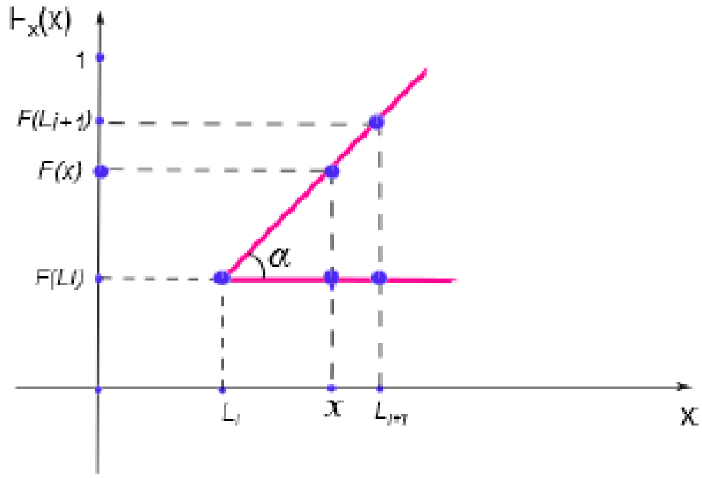
(2) **La médiane** M_e **:** C'est la valeur M_e tel que $F(M_e) = 0.5$ où F est la fonction cumulative la classe $[L_{i-1}, L_i[$.

Nous pouvons la déterminer graphiquement où par calcule.

- Graphiquement à partir de la formule

$$\tan\alpha = \frac{F\left(L_{i+1}\right) - F\left(L_i\right)}{L_{i+1} - L_i} = \frac{F\left(M_e\right) - F\left(L_i\right)}{M_e - L_i} = \frac{0.5 - F\left(L_i\right)}{M_e - L_i}$$

Plus précisément, dans la figure ci-dessous, nous mettons $F(x) = 0.5$ et $x = Me$.



Dans l'example 5 : $M_e \in [400, 600[$

$$M_e = 400 + 200 * \frac{0.5 - 0.225}{0.525 - 0.225} \approx 584 \in [400, 600[\,.$$

1.3.4 Caractéristique de dispersion :

(a) **L'étendu** : Le nombre

$$E = e_{\max} - e_{\min}.$$

S'appelle étendu de X.
Example 5 : $E = 1200 - 200 = 1000$

(b) **La Variance** : est la quantité

$$Var\left(X\right) = \frac{1}{N}\sum_{i=1}^{m} n_i\left(c_i - \bar{X}\right)^2 = \sum_{i=1}^{m} f_i\left(c_i - \bar{X}\right)^2.$$

(c) **L'écart type** σ_x : est la quantité

$$\sigma_x = \sqrt{Var(X)} = \sqrt{\sum_{i=1}^{m} f_i \left(c_i - \bar{X}\right)^2}.$$

(d) **L'écart absolue moyen :**

$$E_{ab} = \frac{1}{N}\sum_{i=1}^{m} n_i \left|c_i - \bar{X}\right|$$

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

(e) **Les quartiles :** Pour $i \in \{1.2.3\}$, la quantité Q_i tel que

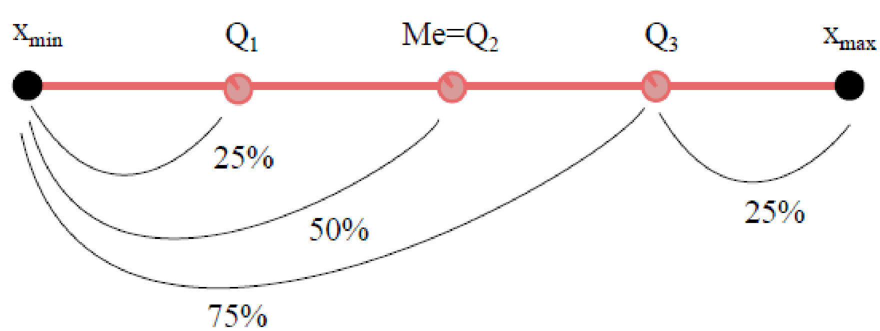
$$F(Q_i) = \frac{i}{4}$$

s'appelle $i^{ème}$ quartile.

$$\text{Pour } i = 2,\ Q_2 \text{ telque } F(Q_2) = \frac{2}{4} = 0.5. \text{ Donc } Q_2 = M_e.$$

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".

$$\begin{array}{ccc} Q_1 & \rightarrow & 25\%, \\ Q_2 & \rightarrow & 50\%, \\ Q_3 & \rightarrow & 75\%. \end{array}$$



Les quartiles.