

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 opérations 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).

Exemple 16

Soit Ω l'ensemble des nouveaux nés au C.H.U d'une ville pendant les 3 premiers mois de 2017. Nous désignons par X le poids des nouveaux nés. On suppose que

$$x_{\min} = 2.701 \quad \text{et} \quad x_{\max} = 5.001.$$

Remarque 9

Comment étudier ce caractère ?

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

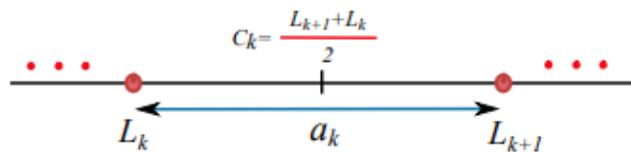
3.1.1 Classe de valeurs

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 $[a, b[$.

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

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



Remarque 10

On supposera dans tous les cas étudiés que la distribution à l'intérieur des classes est uniforme (voir Figure 3.1). Cette hypothèse permet de justifier le fait qu'on choisisse le centre des classes comme représentant.

3.1.2 Nombre de classes

En combien de classes partageons-nous les valeurs ? la réponse n'est pas unique. Soit N l'effectif total. Nous pouvons considérer dans ce cours trois réponses à titre d'exemple.

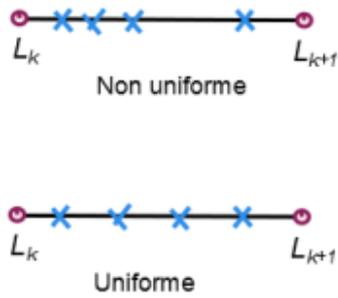


FIGURE 3.1: Une représentation de la distribution des valeurs à l'intérieur d'une classe.

1. Une réponse : \sqrt{N} , $[\sqrt{N}]$ (partie entière) ou $[\sqrt{N}] + 1$. Donc, le nombre de classes

$$k \simeq \sqrt{N}.$$

Exemple 17

Considérons 30 valeurs entre 56.5 cm et 97.8 cm. Dans ce cas, $k = \sqrt{30}$ et on prend $k = 6$.

2. Une réponse : la formule de Sturge

$$k = 1 + 3.3 \log_{10}(N).$$

3. Une réponse : la formule de Yule

$$k = 2.5 \sqrt[4]{N}.$$

Remarque 11

De ce fait, on peut avoir plusieurs tableaux statistiques selon le nombre de classes.

Exemple 18

Si on prend $N = 30$, alors le nombre de classes est donné, par exemple, par

- soit la formule de Sturge $k = 1 + 3.3 \log_{10}(30) \simeq 6$,

- soit la formule de Yule $k = 2.5 \sqrt[3]{30} \simeq 6$.

Nous mentionnons que les deux formules sont presque pareils si $N \ll 200$.

Nous rappelons maintenant la définition de l'étendu. De plus, dans le cas continue nous parlons aussi du pas ou de la longueur de la classe.

Définition 13

Le nombre

$$e = x_{max} - x_{min}$$

s'appelle étendu de X . Dans ce cas, on peut définir le pas par

$$a_i := \frac{\text{étendu}}{\text{nombre de classes}} = \frac{x_{max} - x_{min}}{k}.$$

3.1.3 Effectif et fréquence d'une classe

Définition 14

La quantité

$$n_i := \text{Card}\{w \in \Omega : X(w) \in C_i\}$$

s'appelle effectif partiel de C_i .

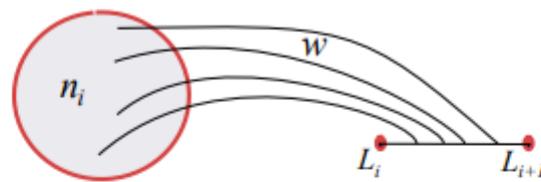


FIGURE 3.2: Le nombre d'individus qui prennent des valeurs x_i dans C_i .

Définition 15

Le nombre

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

est appelé la fréquence partielle de C_i .

Définition 16

On appelle l'effectif cumulé de C_i la quantité

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

Définition 17

On appelle la fréquence cumulée de C_i la quantité

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

Remarque 12

Nous avons, comme dans le chapitre précédent, les interprétations suivantes :

- n_i : est le nombre d'individus dont les valeurs des caractères sont dans la classe C_i ,
- f_i : est le pourcentage des w tel que $X(w) \in C_i$,
- N_i : est égale au $\text{Card}\{w : X(w) \in C_1 \cup C_2 \cup \dots \cup C_i\}$,
- F_i : est le pourcentage des w tel que

$$X(w) \in C_1 \cup \dots \cup C_i.$$

3.2 Représentation graphique d'un caractère continu

3.2.1 Histogramme des fréquences (ou effectifs)

Nous pouvons représenter le tableau statistique par un histogramme. Nous reportons les classes sur l'axe des abscisses et, au-dessus de chacune d'elles, nous traçons un rectangle dont l'aire est proportionnelle à la fréquence f_i (ou l'effectif n_i) associée. Ce graphique est appelé l'histogramme des fréquences (voir Figure 3.3).

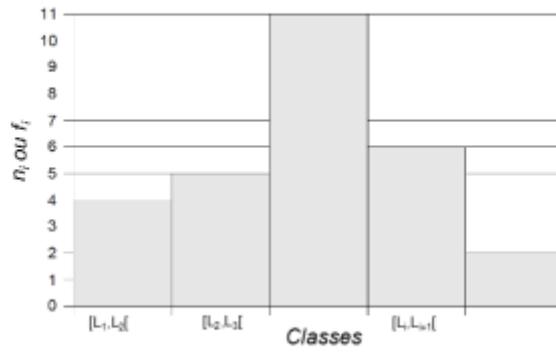


FIGURE 3.3: Histogramme des fréquences ou des effectifs.

3.2.2 Fonction de répartition

Notation : Nous allons noter par

$$C_i = [x_{\min} = a_0, x_{\min+1} = a_1[.$$

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

Remarque 13

Nous calculons $F_x(x)$ par extrapolation (voir Figure 3.4). Nous avons déjà $F(L_i) = F_i$. De plus,

$$\tan(\alpha) = \frac{F(L_{i+1}) - F(L_i)}{L_{i+1} - L_i} = \frac{F(x) - F(L_i)}{x - L_i}.$$

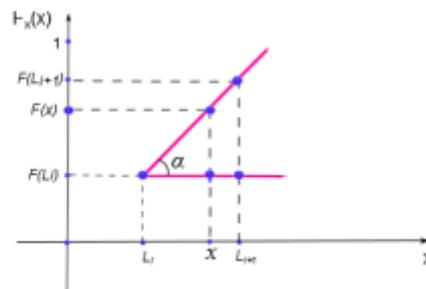


FIGURE 3.4: Le calcul de $F_x(x)$ par extrapolation.

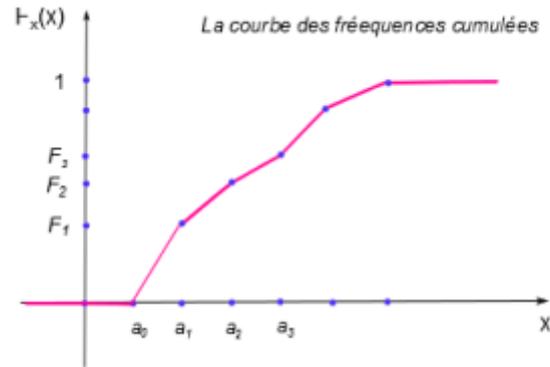


FIGURE 3.5: La courbe des fréquences cumulées.

3.3 Paramètres de tendance central

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



FIGURE 3.6: Le centre de la classe.

La moyenne

Définition 19

La quantité

$$\bar{x} = \sum_{i=1}^n f_i C_i$$

s'appelle la moyenne de X.

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".

Définition 20

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

$$M_0 = L_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} a_i$$

s'appelle le mode avec (voir Figure 3.7)

- L_i : la borne inférieure de la classe modale.
- a_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.

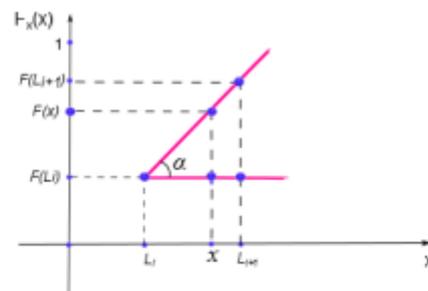


FIGURE 3.8: Le calcul de la médiane par extrapolation.

3.4 Paramètres de dispersion

Définition 22

La variance est la quantité

$$Var(x) = \sum_{i=1}^n f_i(\bar{x} - C_i)^2.$$

Remarque 15

Pour le calcul, on utilise (voir Chapitre 2, Théorème 1)

$$Var(x) = \sum_{i=1}^n f_i C_i^2 - \bar{x}^2.$$

Définition 23

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.

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".

$$Q_1 \rightarrow 25\%,$$

$$Q_2 \rightarrow 50\%,$$

$$Q_3 \rightarrow 75\%.$$

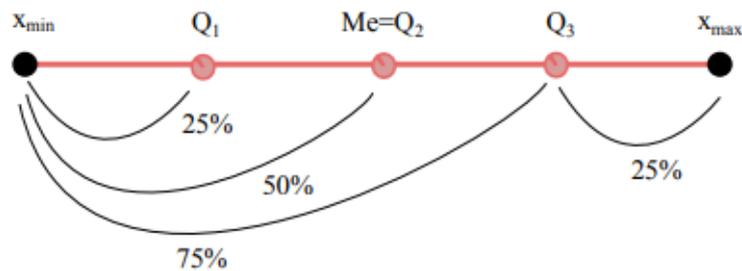


FIGURE 3.9: Les quartiles.

Exercice 16

- Classer ces statistiques selon leurs natures (indicateur de position ou de dispersion)

	Minimum	Moyenne	Écart-type	Mode	Médiane	Premier quartile
Position						
Dispersion						

Exercice 17

- Chez un fabricant de tubes de plastiques, on a prélevé un échantillon de 100 tubes dont on a mesuré le diamètre en décimètre.

1.94	2.20	2.33	2.39	2.45	2.50	2.54	2.61	2.66	2.85
1.96	2.21	2.33	2.40	2.46	2.51	2.54	2.62	2.68	2.87
2.07	2.26	2.34	2.40	2.47	2.52	2.55	2.62	2.68	2.90
2.09	2.26	2.34	2.40	2.47	2.52	2.55	2.62	2.68	2.91
2.09	2.28	2.35	2.40	2.48	2.52	2.56	2.62	2.71	2.94
2.12	2.29	2.36	2.41	2.49	2.52	2.56	2.63	2.73	2.95
2.13	2.30	2.37	2.42	2.49	2.53	2.57	2.63	2.75	2.99
2.14	2.31	2.38	2.42	2.49	2.53	2.57	2.65	2.76	2.99
2.19	2.31	2.38	2.42	2.49	2.53	2.59	2.66	2.77	3.09
2.19	2.31	2.38	2.42	2.50	2.54	2.59	2.66	2.78	3.12

1. Identifier la population, les individus, le caractère et son type.
2. En utilisant la méthode de Yule puis de Sturge, établir le tableau statistique (Faites débiter la première classe par la valeur 1.94).
3. Tracer l'histogramme de cette variable statistique.
4. Déterminer par le calcul la valeur du diamètre au-dessous de laquelle se trouvent 50% des tubes de plastique. Que représente cette valeur.
5. Déterminer par le calcul le pourcentage de tubes ayant un diamètre inférieur à 2.58.

Exercice 18

- Une étude sur le budget consacré aux vacances d'été auprès de ménages a donné les résultats suivants

Budget X	Fréquence cumulée	Fréquences
$[800, 1000[$	0.08	
$[1000, 1400[$	0.18	
$[1400, 1600[$	0.34	
$[1600, \beta[$	0.64	
$[\beta, 2400[$	0.73	
$[2400, \alpha[$	1	

Le travail demandé :

- Certaines données sont manquantes. Calculer la borne manquante α sachant que l'étendue de la série est égale à 3200.
- Calculer les fréquences dans le tableau.
- Calculer la borne manquante β dans les deux cas suivants :
 1. Le budget moyen est égal à 1995.
 2. Le budget médian est égal à 1920.