Cours 10 : Paramètres de position

Étant donnée une variable statistique quantitative, nous définissons deux **paramètres de position**, c'est-à-dire des valeurs autour desquelles la variable est centrée :
- la moyenne ;
- la médiane.

Moyenne

Supposons que l'on dispose d'une série statistique brute $(x_1, \ldots, x_N)$, où l'effectif total est N. La **moyenne** est définie comme :

$$\overline{x} := \frac{x_1 + x_2 + \ldots + x_N}{N}.$$

En général, avec des variables discrètes, on ne travaille pas avec la série brute mais avec la série triée par classes : pour chaque modalité i, on calcule l'effectif n_i ou la fréquence $f_i = n_i/N$. La moyenne est alors :

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{N} = f_1x_1 + f_2x_2 + \ldots + f_px_p.$$

Dans le cas de variables continues triées par classes, on ne dispose pas des valeurs exactes x_i. La méthode la plus simple consiste à utiliser pour chaque classe la valeur centrale de la classe, c'est-à-dire la moyenne de ses extrémités.

Exemple : considérons l'exemple, utilisé la semaine passée, de la taille dans une promotion d'étudiants :

Taille (en cm)	[120,150[	[150,160[	[160,170[	[170,180[	[180,190[	[190, 210[
Centre (en cm)	135	155	165	175	185	200
Fréquence (en %)	1,7	12,8	37,8	36,0	9,3	2,3

La taille moyenne vaut alors (attention à convertir les pourcentages en nombres décimaux !) :

$$0,017 \times 135 + 0,128 \times 155 + 0,378 \times 165 + 0,360 \times 175 + 0,093 \times 185 + 0,023 \times 200 = 169,31.$$

La taille moyenne de cette population est donc de 169 centimètres.

Médiane

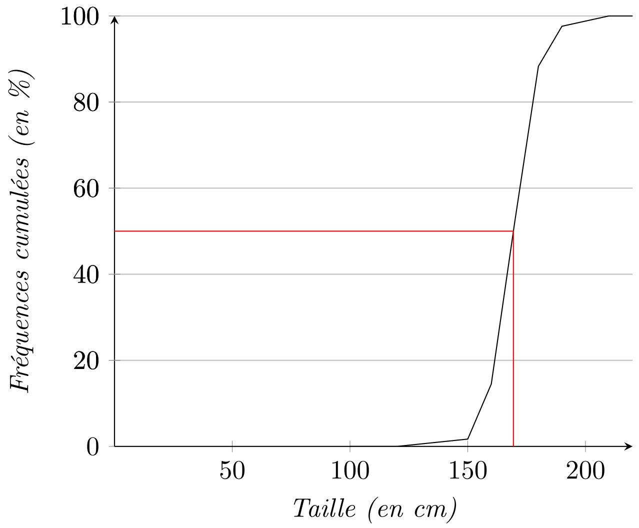
Supposons que l'on dispose d'une série statistique brute $(x_1, \ldots, x_N)$, classée en ordre croissant, où l'effectif total est N. La **médiane** est la valeur centrale[1] de la série.

En général, avec des variables discrètes, on ne travaille pas avec la série brute mais avec la série triée par classes. Il suffit alors de calculer les fréquences cumulées, et de choisir la première valeur x dont la fréquence cumulée associée F_x est supérieure ou égale à 50%.

1. Il y a des conventions – peu importantes – si l'effectif total est pair.

Dans le cas de variables continues triées par classes, on ne dispose pas des valeurs exactes x_i. La méthode la plus simple consiste à chercher la valeur de x dont la fréquence cumulée associée est 50%. Cela se détermine aisément graphiquement [2].

Exemple : reprenons l'exemple précédent :



Graphiquement, on trouve une médiane d'environ 170 centimètres (une valeur plus précise, accessible par le calcul, serait de $169,39$ centimètres).

Remarque : la moyenne peut être fortement affectée par des valeurs extrêmes, là où la médiane y est insensible. Si la distribution est à peu près symétrique, comme dans l'exemple utilisé ici, la moyenne sera très proche de la médiane. Cependant, dans le cas de distributions concentrées vers le bas ou vers le haut, la différence entre les deux peut être non négligeable. C'est le cas par exemple des revenus : une minorité de revenus élevés tire la moyenne vers le haut. Par exemple, en France en 2019, le salaire médian est de 1789€, alors que le salaire moyen est de 2238€ – une différence de 20% ! Ce phénomène se retrouve aussi pour les dépenses des clients dans certains secteurs.

2. Au vu de l'imprécision associée au tri par classes, cette détermination graphique est suffisante.