

Statistiques descriptives

Leçon

1. Indicateurs de position (rappels)

Définition

On considère une série statistique de valeurs $x_1, \dots, x_p$ d'effectifs $n_1, \dots, n_p$, d'effectif total $N = n_1 + \dots + n_p$.

- La **moyenne** est $\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N}$.
- La **médiane** Me partage la série *rangée* en deux groupes de même effectif.
- Le **premier quartile** Q_1 est la plus petite valeur telle qu'au moins 25 % des valeurs lui soient inférieures ou égales ; le **troisième quartile** Q_3 : idem avec 75 %.

Définition – Indicateurs de dispersion

L'**étendue** est $x_{\max} - x_{\min}$; l'**écart interquartile** est $Q_3 - Q_1$.

2. Linéarité de la moyenne

Propriété

Si on transforme chaque valeur x_i en $y_i = ax_i + b$, alors la moyenne est transformée de la même façon : $\bar{y} = a\bar{x} + b$.

Exemple. La moyenne d'une classe à un devoir est 11,5. Si on ajoute 1 point à chaque copie, la nouvelle moyenne est 12,5. Des températures de moyenne 15 °C ont pour moyenne $\frac{9}{5} \times 15 + 32 = 59$ °F.

3. Écart type

Définition

La **variance** d'une série est la moyenne des carrés des écarts à la moyenne :

$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N}$$

L'**écart type** est $\sigma = \sqrt{V}$. Il s'exprime dans la même unité que les valeurs.

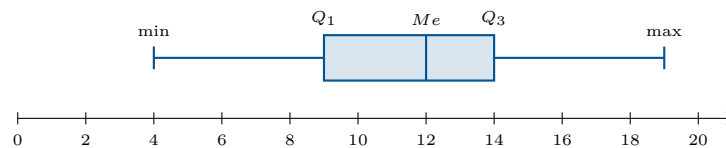
Remarque. L'écart type mesure la dispersion des valeurs *autour de la moyenne* : plus il est grand, plus les valeurs sont dispersées. On l'obtient en pratique avec la calculatrice ou un tableur.

Exemple. Série : 4 ; 8 ; 10 ; 12 ; 16. Moyenne $\bar{x} = 10$. Écarts : -6 ; -2 ; 0 ; 2 ; 6. $V = \frac{36 + 4 + 0 + 4 + 36}{5} = 16$ et $\sigma = 4$.

Méthode – Comparer deux séries

On compare un indicateur de position et un indicateur de dispersion :

- le couple (moyenne ; écart type) ;
- ou le couple (médiane ; écart interquartile), moins sensible aux valeurs extrêmes, visualisé par une **boîte à moustaches**.



4. Ajout ou suppression d'une valeur

Propriété

L'ajout ou la suppression d'une valeur extrême peut beaucoup modifier la moyenne, mais modifie peu la médiane : la médiane est un indicateur **robuste**.

Exemple. Série : 8 ; 10 ; 11 ; 13 ; 13 : moyenne 11, médiane 11. En ajoutant la valeur 40 : moyenne $\frac{95}{6} \approx 15,8$, médiane $\frac{11+13}{2} = 12$.

Méthode – Nouvelle moyenne après ajout d'une valeur

Si une série de N valeurs a pour moyenne $\bar{x}$, la somme des valeurs est $N\bar{x}$. Après ajout d'une valeur v , la moyenne devient $\frac{N\bar{x} + v}{N + 1}$.

5. Séries regroupées en classes

Définition

Lorsque les données sont nombreuses ou continues, on les regroupe en **classes** $[a ; b[$ de même amplitude. On les représente par un **histogramme** : rectangles accolés de hauteur proportionnelle à l'effectif.

Méthode – Moyenne d'une série en classes

En supposant la répartition uniforme dans chaque classe, on remplace chaque classe par son **centre** $\frac{a+b}{2}$, puis on calcule la moyenne pondérée. (Si on connaît la moyenne de chaque classe, on l'utilise à la place du centre.)

Méthode – Classe médiane et estimation de la médiane

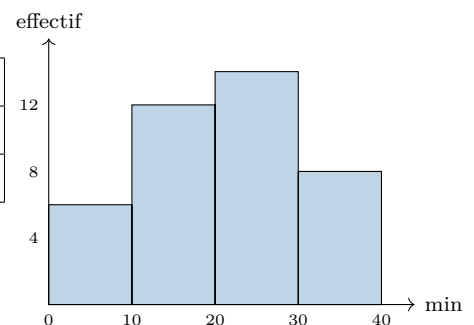
- On calcule les effectifs cumulés croissants ; la **classe médiane** est la première classe dont l'effectif cumulé atteint $\frac{N}{2}$.
- Si la répartition est uniforme dans la classe médiane $[a ; b[$ (effectif n , effectif cumulé avant elle c) : $Me \approx a + (b - a) \times \frac{\frac{N}{2} - c}{n}$.

Exemple.

Temps de trajet (en min) de 40 élèves :

Classe	$[0 ; 10[$	$[10 ; 20[$	$[20 ; 30[$	$[30 ; 40[$
Effectif	6	12	14	8
Eff. cumulé	6	18	32	40

Moyenne : $\frac{6 \times 5 + 12 \times 15 + 14 \times 25 + 8 \times 35}{40} = \frac{840}{40} = 21$ min.
 $\frac{N}{2} = 20$: classe médiane $[20 ; 30[$; $Me \approx 20 + 10 \times \frac{20 - 18}{14} \approx 21,4$ min.



Remarque. On peut aussi tracer le **polygone des fréquences cumulées** et lire graphiquement l'abscisse du point d'ordonnée 0,5 pour estimer la médiane.

6. Algorithmique

Exemple de fonctions renvoyant une moyenne et un écart type.

```
1 from math import sqrt
2
3 def moyenne(L):
4     s = 0
5     for x in L:
6         s = s + x
7     return s / len(L)
8
9 def ecart_type(L):
10    m = moyenne(L)
11    s = 0
12    for x in L:
13        s = s + (x - m)**2
14    return sqrt(s / len(L))
```

`ecart_type([4, 8, 10, 12, 16])` renvoie 4.0.