Chapitre 11. Cours

Statistiques

1. Vocabulaire et effectifs

Définition 1.1 — Série statistique

On étudie un **caractère** sur une **population**. Chaque individu fournit une **valeur**. L'**effectif** n_i d'une valeur x_i est le nombre d'individus qui la prennent ; l'effectif total est $N = \sum n_i$.
La **fréquence** de x_i est $f_i = \dfrac{n_i}{N}$; la somme des fréquences vaut 1.

Propriété 1.2 — Effectifs cumulés croissants

L'**effectif cumulé croissant** d'une valeur est le nombre d'individus dont la valeur lui est inférieure ou égale. La dernière colonne du tableau des effectifs cumulés vaut toujours N : c'est le contrôle de la ligne.

Exemple 1.3

Les notes d'un devoir dans une classe de 25 élèves :

Note x_i	6	8	10	12	14	16
Effectif n_i	2	4	7	6	4	2
Effectif cumulé	2	6	13	19	23	25

La dernière case vaut bien $25 = N$ ✓.

2. Indicateurs de position

Définition 2.1 — Moyenne

La **moyenne** d'une série de valeurs x_i d'effectifs n_i est

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_p x_p}{N} = \frac{1}{N}\sum n_i x_i.$$

Définition 2.2 — Médiane

On range les N valeurs par ordre croissant. La **médiane** M est une valeur qui partage la série en deux moitiés de même effectif :

- si N est **impair**, c'est la valeur de rang $\dfrac{N+1}{2}$;
- si N est **pair**, c'est la moyenne des valeurs de rangs $\dfrac{N}{2}$ et $\dfrac{N}{2}+1$.

Définition 2.3 — Quartiles

Le premier quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25 % des données lui soient inférieures ou égales.
Le troisième quartile Q_3 est la plus petite valeur telle qu'au moins 75 % des données lui soient inférieures ou égales.

Méthode 2.4 — Trouver Q_1 et Q_3

1. Calculer $\frac{N}{4}$ et $\frac{3N}{4}$.
2. Si le résultat n'est pas entier, l'**arrondir à l'entier supérieur** ; s'il est entier, le garder tel quel.
3. Lire la valeur de ce rang dans la série ordonnée, ou dans la ligne des effectifs cumulés.

Exemple 2.5

Reprenons la série de l'exemple 1.3, avec $N = 25$.
Moyenne : $\overline{x} = \dfrac{2\times 6 + 4\times 8 + 7\times 10 + 6\times 12 + 4\times 14 + 2\times 16}{25} = \dfrac{274}{25} = 10{,}96$.
Médiane : $N = 25$ est impair, le rang est $\dfrac{26}{2} = 13$. La colonne des effectifs cumulés atteint 13 à la note 10 : donc $M = 10$.
Quartiles : $\dfrac{25}{4} = 6{,}25$, arrondi à 7 : la 7$^{\text{e}}$ valeur est 10, donc $Q_1 = 10$.
$\dfrac{3\times 25}{4} = 18{,}75$, arrondi à 19 : la 19$^{\text{e}}$ valeur est 12, donc $Q_3 = 12$.

Attention — La moyenne n'est pas la médiane

Ici $\overline{x} = 10{,}96$ et $M = 10$: les deux indicateurs diffèrent. La moyenne tient compte de *toutes* les valeurs, donc une note extrême la déplace ; la médiane ne dépend que du *rang* central, elle est insensible aux valeurs extrêmes. C'est pourquoi on publie la médiane des salaires plutôt que leur moyenne.

3. Indicateurs de dispersion

Définition 3.1

L'**étendue** est la différence entre la plus grande et la plus petite valeur.
L'**écart interquartile** est $Q_3 - Q_1$: c'est l'amplitude de la moitié centrale des données.

Définition 3.2 — Variance et écart type

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

$$V = \frac{1}{N}\sum n_i\,(x_i - \overline{x})^2\,.$$

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

Propriété 3.3 — Formule pratique

$$V = \left(\frac{1}{N}\sum n_i x_i^2\right) - \overline{x}^2.$$

Cette écriture évite de recalculer tous les écarts : on cumule les carrés, on retranche le carré de la moyenne.

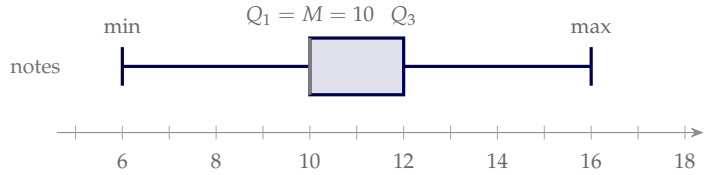
Remarque — Deux couples d'indicateurs

On associe toujours un indicateur de position à un indicateur de dispersion de la même famille : **moyenne et écart type** d'un côté, **médiane et écart interquartile** de l'autre. Mélanger les deux couples n'a pas de sens.

4. Diagramme en boîte

Définition 4.1

Le **diagramme en boîte** (ou boîte à moustaches) résume une série par cinq nombres : minimum, Q_1, médiane, Q_3, maximum. La boîte s'étend de Q_1 à Q_3 et contient donc environ la moitié des données ; les moustaches vont aux valeurs extrêmes.



La série de l'exemple 1.3 : la boîte va de 10 à 12, la médiane est confondue avec Q_1.

Méthode 4.2 — Comparer deux séries

1. Comparer les **positions** : la boîte la plus à droite correspond aux valeurs les plus élevées.
2. Comparer les **largeurs de boîte** : la plus étroite indique une série plus homogène.
3. Comparer les **moustaches** : elles signalent l'étendue et les valeurs isolées.

5. Linéarité de la moyenne

Théorème 5.1

Si toutes les valeurs d'une série sont transformées par $y_i = a\,x_i + b$, alors

$$\overline{y} = a\,\overline{x} + b \qquad \text{et} \qquad \sigma_y = |a|\,\sigma_x.$$

Démonstration

$\overline{y} = \frac{1}{N}\sum n_i(ax_i + b) = \frac{a}{N}\sum n_i x_i + \frac{b}{N}\sum n_i = a\overline{x} + b$, puisque $\sum n_i = N$.
Pour l'écart type, chaque écart à la moyenne devient $y_i - \overline{y} = a(x_i - \overline{x})$: les écarts sont tous multipliés par a, donc l'écart type par $|a|$. ■

Ajouter une constante déplace la série sans la disperser davantage ; la multiplier par a dilate les écarts d'un facteur $|a|$.

Théorème 5.2 — Moyenne de deux groupes

Si un groupe de N_1 individus a pour moyenne $\overline{x_1}$ et un second groupe de N_2 individus pour moyenne $\overline{x_2}$, la moyenne de la population réunie vaut

$$\overline{x} = \frac{N_1\,\overline{x_1} + N_2\,\overline{x_2}}{N_1 + N_2}.$$

Attention — Ce n'est pas la moyenne des moyennes

On ne fait $\dfrac{\overline{x_1} + \overline{x_2}}{2}$ que si les deux groupes ont *exactement* le même effectif. Sinon il faut pondérer : une classe de 30 élèves à 12 et une de 10 élèves à 16 donnent $\dfrac{30 \times 12 + 10 \times 16}{40} = \dfrac{520}{40} = 13$, et non 14.

6. Avec Python : les indicateurs d'une série

```
notes = [6, 6, 8, 8, 8, 8, 10, 10, 10, 10, 10, 10, 10,
       12, 12, 12, 12, 12, 12, 14, 14, 14, 14, 16, 16]
def moyenne(s):
  return sum(s) / len(s)
def mediane(s):
  t = sorted(s)
  n = len(t)
  if n % 2 == 1:
    return t[n // 2]
  return (t[n // 2 - 1] + t[n // 2]) / 2
def ecart_type(s):
  m = moyenne(s)
  return (sum((v - m)**2 for v in s) / len(s)) ** 0.5
print(len(notes), moyenne(notes), mediane(notes))
print(round(ecart_type(notes), 4))
print(max(notes) - min(notes))
```

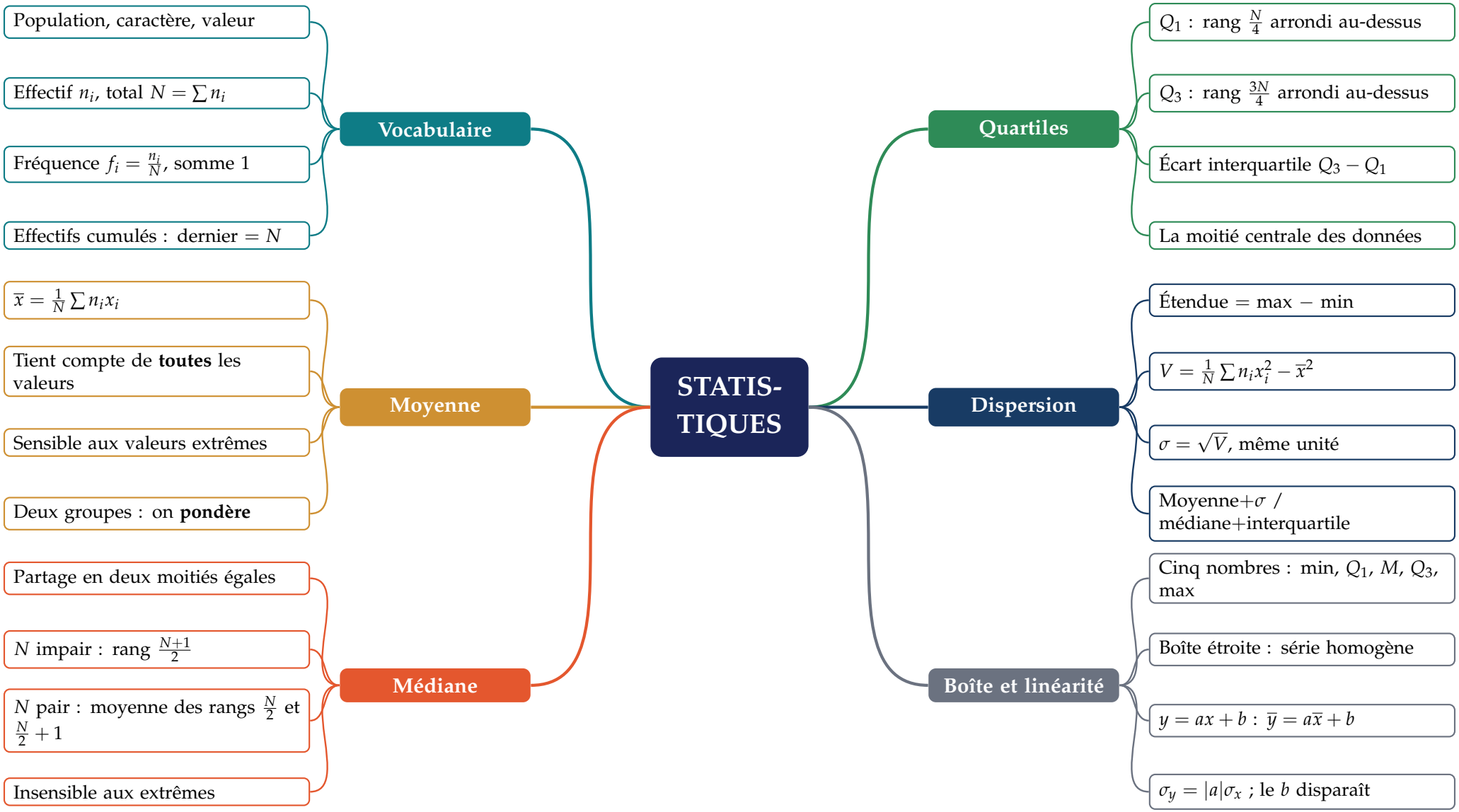
Trace d'exécution.

```
25 10.96 10
2.72
10
```

Interprétation. Le programme confirme les trois valeurs calculées à la main : moyenne 10,96, médiane 10, étendue 10. L'écart type vaut ici exactement 2,72 points — car $V = \dfrac{3\,188}{25} - 10{,}96^2 = 7{,}3984 = 2{,}72^2$ — ce qui indique que les notes s'écartent en moyenne de moins de trois points de 10,96 : la classe est assez homogène. Noter la ligne `n % 2 == 1` : c'est le test de parité qui distingue les deux cas de la définition 2.2 — le programme code exactement la règle du cours.

Remarque — Les trois réflexes du chapitre

1. Moyenne et médiane répondent à deux questions différentes ; la médiane résiste aux valeurs extrêmes. 2. Un indicateur de position s'accompagne d'un indicateur de dispersion de la même famille. 3. La moyenne de deux groupes se *pondère* par les effectifs, jamais par une simple demi-somme.



Chapitre 11 · Chaque branche renvoie à une section du cours et à une fiche méthode.