III - Moyenne et écart-type

a) Un paramètre de position : la moyenne

Définition : La **moyenne** d'une série statistique, notée $\overline{x}$, est la somme des N valeurs de la série divisée par l'effectif total N.
Pour un caractère quantitatif discret (comme l'exemple de la partie I), si on note x_1, x_2, ..., x_p les valeurs de la série (on considère ici qu'il y a p valeurs différentes), n_1, n_2, ..., n_p leurs effectifs respectifs (avec $n_1 + n_2 + \ldots + n_p = N$) et f_1, f_2, ..., f_p leurs fréquences d'apparition alors on peut représenter la situation sous la forme de ce tableau :

Valeurs	x_1	x_2	...	x_p
Effectifs	n_1	n_2	...	n_p
Fréquences	f_1	f_2	...	f_p

On obtient la moyenne :
$\overline{x} = \dfrac{x_1 \times n_1 + x_2 \times n_2 + \ldots + x_p \times n_p}{N} = x_1 \times f_1 + x_2 \times f_2 + \ldots + x_p \times f_p$.

Remarque : Dans le cas d'une série de données sans effectifs (donc où chaque valeur apparait avec un effectif de 1), le calcul de la moyenne se fait tout simplement en ajoutant les valeurs et en divisant cette somme par le nombre total de valeurs.

Remarque : Le calcul de votre moyenne sur un bulletin se fait comme si toutes les notes avaient un effectif de 1. Mais lorsqu'on souhaite mettre des coefficients aux différentes disciplines, ces coefficients agissent comme des effectifs et on se retrouve dans la configuration de la définition.

Exemple : Dans le cas de l'exemple de la partie I, le score moyen obtenu par l'ensemble des 31 tireurs vaut :

$$\frac{7 \times 2 + 8 \times 1 + 10 \times 3 + 11 \times 3 + 12 \times 5 + 13 \times 4 + 14 \times 7 + 15 \times 3 + 16 \times 2 + 18 \times 1}{31} \simeq 12{,}58$$

Cela peut également se faire en python où deux étapes seront nécessaires : calculer la somme totale, puis diviser par l'effectif total (une fois sorti de la boucle).
Je vais le faire ci-dessous sous la forme d'une fonction nommée `moyAvecEffectifs` prenant en paramètres deux listes (la première pour les valeurs, la deuxième pour les effectifs) :

```python
def moyAvecEffectifs(listeVal, listeEff):
    sommeTotale=0
    effTotal=0
    for i in range(len(listeVal)):
      effTotal=effTotal+listeEff[i]
      sommeTotale=sommeTotale+listeVal[i]*listeEff[i]
    return sommeTotale/effTotal
```

En reprenant les listes de la partie I, il suffit alors d'écrire `moyAvecEffectifs(scores, effectifs)` et on obtient environ 12,58.

Propriété : **Linéarité de la moyenne**

- Si on multiplie chaque valeur d'une série par un réel k, la moyenne est multipliée par k.
- Si on ajoute un réel k à chaque valeur d'une série, la moyenne augmente de k.

Propriété : Soit $\overline{x}$ la moyenne d'une série statistique d'effectif total N.
Si $\overline{x_1}$ et $\overline{x_2}$ sont les moyennes respectives de deux sous-groupes de la série d'effectifs respectifs N_1 et N_2, alors :
$\overline{x} = \dfrac{N_1 \times \overline{x_1} + N_2 \times \overline{x_2}}{N}$.

Remarque : Cela revient à interpréter cette série de données, comme un tableau avec leurs et effectifs :

Valeurs	$\overline{x_1}$	$\overline{x_2}$
Effectifs	N_1	N_2

Exercice 1 : On a observé le prix au kg de différentes variétés de pommes dans un magasin :

Variété	Golden	Boskop	Reinette	Canada
Prix au kg	2,90 €	2,60 €	2,75 €	2,95 €

1. Calculer le prix moyen des pommes au kg.
2. Si on applique une augmentation de 0,8% à tous les fruits, quel sera le prix moyen au kg des pommes de ce magasin ?

Exercice 2 : La taille moyenne des 20 garçons d'une classe de seconde est 170,4 cm, celle des 14 filles est 168,5 cm. Calculer la taille moyenne des élèves de la classe.

. . . .

b) Un paramètre de dispersion : l'écart-type

Définition : On dispose d'une série statistique sous la même forme que dans la définition de la moyenne et dont la moyenne des valeurs vaut $\overline{x}$. Pour déterminer **l'écart-type** de cette série statistique, on calcule dans un premier temps la **variance** V :

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

L'écart-type σ vaut alors : $\sigma = \sqrt{V}$.

Exemple : Dans l'exemple de la partie I, le calcul de la variance donne :

$$V = \frac{(7-12{,}58)^2 \times 2 + (8-12{,}58)^2 \times 1 + (10-12{,}58)^2 \times 3 + \ldots + (18-12{,}58)^2 \times 1}{31} \simeq 6{,}37$$

et $\sigma = \sqrt{V} \simeq 2{,}52$.

Comme l'écart interquartile, l'écart-type sert à estimer la dispersion des données, notamment pour comparer des séries statistiques de deux populations différentes.
Plus l'écart-type est élevé, plus les données seront considérées comme dispersées (ou hétérogènes).

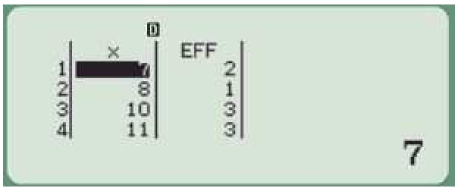
Remarque : Le couple (moyenne ; écart-type) est très utilisé pour comparer des séries statistiques (comme les résultats de deux classes de 2^{nde} du lycée), mais il faut se méfier d'une interprétation un peu trop rapide. En effet, la moyenne et l'écart-type sont très sensibles aux valeurs extrêmes, ce qui peut fausser l'interprétation générale.
En effet, deux élèves avec une moyenne extrêmement faible font chuter la moyenne de la classe sans qu'on puisse dire pour autant que le niveau global de la classe est plus faible.
On retrouve cette situation quand on s'intéresse à la moyenne des salaires en France où les plus gros salaires influencent la valeur de cette moyenne.
Dans ces cas où figurent des valeurs extrêmes, il est sans doute plus judicieux d'utiliser le couple (médiane ; écart interquartile) pour comparer deux séries de données.

c) Utilisation de la calculatrice (Casio fx92)

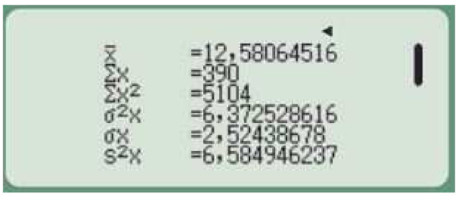
Reprenons l'exemple qui a été le fil conducteur de ce cours.

- Avec la Casio fx92 collège :

  Menu 2 : Statistiques puis 1 : variable
  Remplir la colonne x par les valeurs et la deuxième colonne par les effectifs.

  Appuyer sur la touche $\boxed{\text{OPTN}}$ puis $\boxed{3}$ pour sélectionner Calc à 1 variable.
  On obtient l'affichage ci-contre où $\overline{x}$ désigne la moyenne, ΣX la somme de toutes les valeurs, $\sigma^2 X$ la variance, σX l'écart-type (à ne pas conforme avec sX), n l'effectif total, . . .

ATTENTION : lorsqu'on demande de déterminer la médiane et les quartiles d'une série statistique, on attend les explications données dans le cours. La calculatrice ne sert alors que pour vérifier.