

THÈME: LES STATISTIQUES ET PROBABILITÉS



SÉQUENCE 4: LES STATISTIQUES DESCRIPTIVES

CAPACITÉS:

- ÉTUDIER UNE SÉRIE STATISTIQUE OU MENER UNE COMPARAISON PERTINENTE DE DEUX SÉRIES STATISTIQUES À L'AIDE D'UN LOGICIEL OU D'UNE CALCULATRICE.
- UTILISER DE FAÇON APPROPRIÉE LES DEUX COUPLES USUELS QUI PERMETTENT DE RÉSUMER UNE SÉRIE STATISTIQUE : (MOYENNE, ÉCART-TYPE) ET (MÉDIANE, ÉCART INTERQUARTILE).

1

MEVEL CHRISTOPHE



1°) Paramètres de position

L'activité Dégainer plus vite que ton ombre a permis de se remémorer les différents paramètres de position. Nous allons les généraliser maintenant.

a) La moyenne

Soit les deux séries statistiques issue de l'activité.
Soit N l'effectif total : $N = n_1 + n_2 + \dots + n_p$.

Variable	x_i	x_1	x_2	...	x_p
Effectif	n_i	n_1	n_2	...	n_p

Définition : La moyenne d'une série statistique est le nombre, noté \bar{x} , définie par :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N} \text{ où } N = n_1 + \dots + n_p.$$

$$\bar{x} = \sum_{i=1}^p n_i x_i \text{ où } \Sigma \text{ dit sigma traduit la somme de nombres.}$$

Exemples: Les moyennes des deux séries sont: $\bar{x}_1 = \dots$
et $\bar{x}_2 = \dots$

On peut aussi calculer la moyenne \bar{x} , en utilisant les fréquences f_i : $\bar{x} = f_1x_1 + \dots + f_px_p = \sum_{i=1}^p f_ix_i$.

Remarque : Pour une série regroupée en classes, on obtient une valeur approchée de la moyenne de la série en prenant pour x_i les centres (ou milieux) des classes.

b) Médiane et quartiles

Définition: La médiane d'une série statistique est le nombre noté M_e , tel que:

- 50 % des individus ont une valeur du caractère inférieure ou égale à M_e .
- 50 % des individus une valeur du caractère supérieure ou égale à M_e .

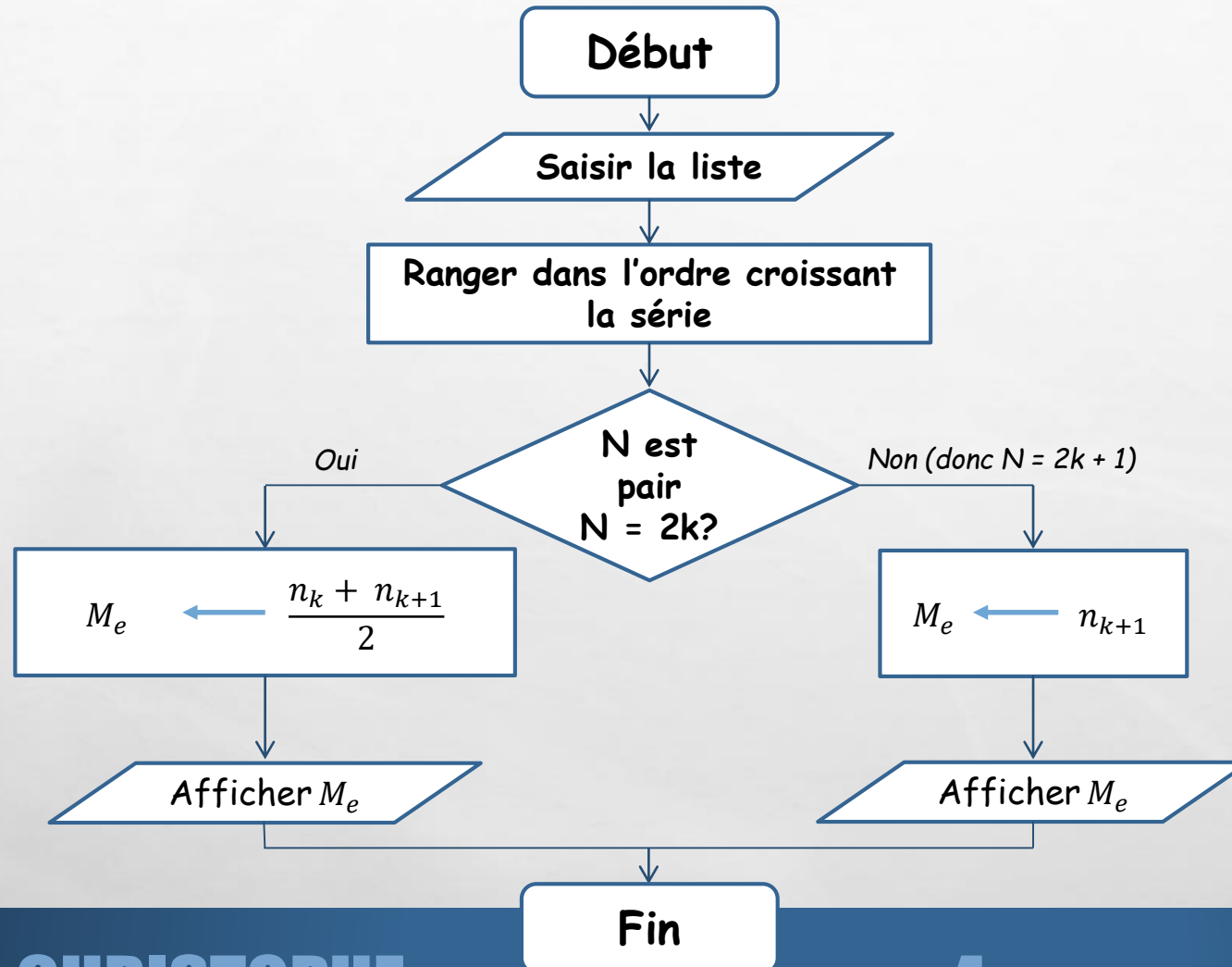
Pour la déterminer : on range la liste des N données par ordre croissant. Soit k un entier naturel.

- si la série est de **taille impaire** ($2k + 1$), alors la médiane est la donnée de rang $k + 1$.
- si la série est de **taille paire** ($2k$), alors la médiane est la demi-somme des données de rang k et $k + 1$.

Exemples: Les médianes des deux séries de l'activité sont:

- ✓
- ✓

Algorithme de détermination de la médiane:



Définition: La série des N valeurs rangée par ordre croissant.

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

Remarque:

En France, la médiane est un nombre n'appartenant pas forcément à la série alors que le premier et troisième quartile le sont.

Ceci diffère du système anglophone où les quartiles sont considérés comme les médianes des deux demi-séries obtenues à partir de la médiane. C'est cette règle qui est programmée dans votre calculatrice.

Exemples: Le premier et le troisième quartile des séries de l'activités sont:

- Série 1:
 - ✓ Premier quartile:
 - ✓ Troisième quartile:
- Série 2:
 - ✓ Premier quartile:
 - ✓ Troisième quartile:

Définition: Les valeurs extrêmes, la médiane et le premier et troisième quartile permettent de partager une série en quatre parties contenant chacune environ un quart de l'effectif total.

Le diagramme en boîte ou « à moustache» est un diagramme regroupant ces différentes valeurs:

- la « boîte » est un rectangle limité par le premier et le troisième quartile où figure la médiane ;
- les « moustaches » en revanche peuvent s'achever aux valeurs extrêmes (le minimum et le maximum de la série).

Exemple: Voir les diagrammes imprimés de l'activité.

Remarque: Une vidéo permet de se rappeler la lecture sur le diagramme: <https://youtu.be/DRBi4Qnxo-s>

2°) Paramètres de dispersion

a) Etendue et écart interquartile

Définitions:

- L'étendue est la différence entre la plus grande et la plus petite valeur de la série.
- L'écart interquartile est la différence entre Q_3 et Q_1 .

b) Variance et écart-type

Définitions:

La variance d'une série statistique est le réel, notée V , défini par:

$$V = \frac{1}{N} [n_1(x_1 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2]$$

$$V = \frac{1}{N} \sum_{i=1}^p n_i(x_i - \bar{x})^2$$

Autre formule: $V = \frac{1}{N} \sum_{i=1}^p n_i x_i^2 - \bar{x}^2$

L'écart-type est le réel noté σ sert à mesurer la dispersion, ou l'étalement, d'un ensemble de valeurs autour de la moyenne. Plus l'écart-type est faible, plus l'ensemble est homogène (INSEE) se calcul par : $\sigma = \sqrt{V}$

Remarque: On peut aussi calculer V en utilisant les fréquences f_i :

$$V = \sum_{i=1}^p f_i (x_i - \bar{x})^2$$

Tous les calculs se font à la calculatrice, alors à quoi servons-nous?

A interpréter les résultats car la machine ne peut qu'engendrer un programme suivant un algorithme fait par l'homme et donc n'est pas en mesure de réfléchir par elle-même...mais cela ne va pas tarder!!!

Comment interpréter?

Deux choix s'offrent à vous:

- Le couple (moyenne; écart-type) qui a un inconvénient majeur, il associe deux paramètres sensibles aux valeurs extrêmes;
- Le couple (médiane; écart-interquartile) qui n'a pas de défaut mais dont la détermination est moins pratique qu'on associe à une représentation graphique en boîte à moustaches.