

10

Statistiques descriptives

I Vocabulaire

- La _____ est un ensemble d' _____ sur lesquels porte une étude. L'ensemble des données collectées s'appelle une _____ .
- On appelle _____ l'aspect de la population étudié. Il peut être :
 - _____ si ses valeurs (ou modalités) sont des nombres ;
 - _____ si ses valeurs (ou modalités) ne sont pas des nombres.
- Un caractère quantitatif peut être :
 - _____ s'il peut prendre un nombre fini de valeurs ;
 - _____ s'il peut prendre un nombre infini de valeurs.
- Les valeurs d'un caractère peuvent être regroupées par _____ .
- L' _____ d'une valeur (ou d'une classe) est le nombre d'individus dont la valeur du caractère est égal à cette valeur (ou incluse dans la classe). La _____ de cette valeur est la proportion d'individus dont le caractère est égal à cette valeur (ou incluse dans cette classe).

Exemple 10.1 :

1. Sur 78 % des couvertures de livres pour enfants figure un personnage masculin^a.
On s'intéresse à la statistique "78 %".
 - (a) Quelle est la population étudiée? Quels sont les individus?
 - (b) Quel est le caractère? Est-il quantitatif ou qualitatif?
2. On s'intéresse à la pointure des chaussures des élèves d'une classe.
 - (a) Quel est le caractère étudié?
 - (b) Est-il quantitatif ou qualitatif?

(c) Est-il continu ou discret?

3. La température mondiale moyenne a augmenté de $0,74^{\circ}\text{C}$ au cours du XX^e siècle, un réchauffement qui touche plus les zones émergées que les océans^b. On s'intéresse à la moyenne " $0,74^{\circ}\text{C}$ ".

(a) Quel est le caractère étudié?

(b) Est-il quantitatif ou qualitatif?

(c) Est-il continu ou discret?

^a. Source : Rapport sur l'égalité entre les filles et les garçons dans les modes d'accueil de la petite enfance, IGAS, Décembre 2012.

^b. Source : Aperçu sur le changement climatique, Organisation des Nations Unies.

Définition 10.1 – Effectifs et fréquences cumulé(e)s croissant(e)s

L' _____ (resp. la _____) d'une valeur est la somme des effectifs associés (resp. des fréquences associées) à toutes les valeurs qui sont inférieures ou égales à cette valeur.

Exemple 10.2 :

1. Dresser le tableau des effectifs cumulés croissants et des fréquences cumulées croissantes associé à la série suivante :

Âge (en années)	0	3	1	5	8	6	2	9	4	7	10
Effectif	1	3	2	5	1	8	1	2	5	4	2

2. Combien d'individus ont 7 ans ou moins?

3. Quelle est la part des individus ayant strictement moins de 5 ans?

Quelle est la médiane de cette série?

II Caractéristiques d'une série statistique

II.1 Médiane et écart interquartile

a) Indicateur de tendance centrale : la médiane

Définition 10.2 – Médiane

Une _____ d'une série statistique est une valeur telle que :

- au moins des valeurs lui sont inférieures ou égales
- au moins des valeurs lui sont supérieures ou égales.



Méthode

Soit une série statistique d'effectif total N ($N \neq 0$).

Une fois la série _____, alors :

- Si N est impair, alors la médiane est la valeur associée à l'individu de rang
- Si N est pair, alors la médiane est la moyenne des valeurs associées aux individus de rang ... et

Remarque(s) :

- La méthode ci-dessus dans le cas N pair permet de définir une valeur unique pour la médiane, car il pourrait alors y avoir plusieurs valeurs possibles. Ex : la série de valeurs « 4 ; 6 » admet 5 comme médiane, mais aussi 4,1, ou encore 5,3...

Exemple 10.3 :

1. Voici les notes sur 20 obtenues à un contrôle : 16 ; 11 ; 19 ; 9 ; 11.
Quelle est la médiane de cette série?

2. Voici les notes obtenues dans une autre classe au même contrôle : 8 ; 15 ; 11 ; 19.

Définition 10.3

Le premier (resp. troisième) _____ d'une série statistique, noté ... (resp. ...) est la plus petite valeur telle qu'au moins (resp.) des valeurs sont inférieures ou égales à ... (resp. ...).

Remarque(s) :

- On a donc un quart des valeurs qui sont inférieures à Q_1 et trois quarts des valeurs qui sont inférieures à Q_3 .



Méthode

Soit une série statistique d'effectif total N .

- Pour trouver Q_1 , on calcule $\frac{N}{4}$. Si ce n'est pas un nombre entier, on arrondit à l'entier supérieur. Q_1 est la valeur associée au rang ainsi trouvé.
- Pour trouver Q_3 , on calcule $\frac{3N}{4}$. Si ce n'est pas un nombre entier, on arrondit à l'entier supérieur. Q_3 est la valeur associée au rang ainsi trouvé.

Exemple 10.4 :

Dans un entreprise de 10 employés les salaires bruts sont répartis comme suit :

- 2 employés ont un salaire de 6 000 CHF.
- 1 employé a un salaire de 7 500 CHF.
- 3 employés ont un salaire de 4 500 CHF
- 4 employés on un salaire de 5 000 CHF.

Déterminer le salaire médian et les premier et troisième quartiles dans cette entreprise.

Définition 10.4

L' _____ est la différence entre le troisième et le premier quartile de la série.

L' _____ est l'intervalle $[Q_1; Q_3]$, il contient au moins 50 % des valeurs de la série.

Remarque(s) :

- Plus l'écart interquartile est grand, plus les données sont à priori dispersées autour de la médiane. Inversement, un intervalle interquartile petit indique que les données sont plutôt concentrées autour de la médiane.

Exemple 10.5 :

On donne la répartition des âges des français en 2012 (source : INSEE).

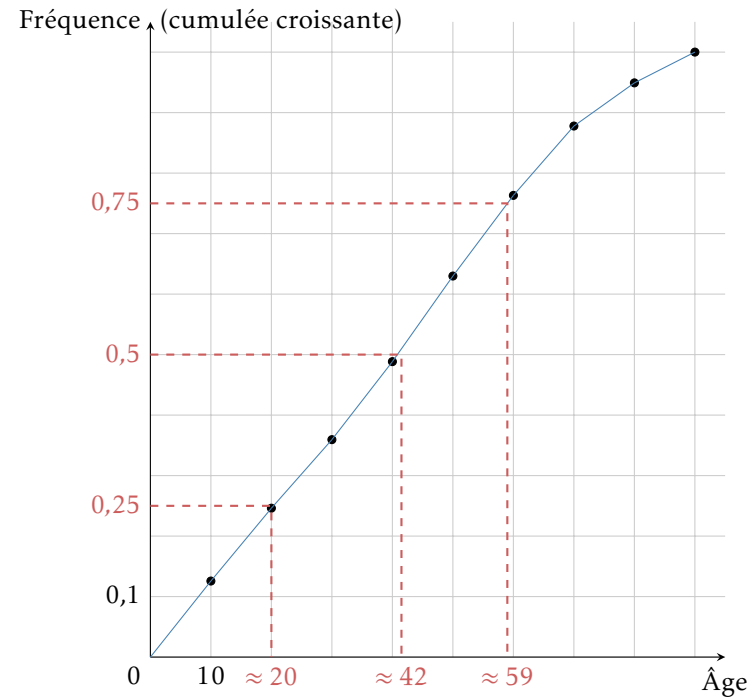
1. Compléter le tableau ci-dessous avec les fréquences cumulées croissantes.

Âge	Effectif (en milliers)
Moins de 10 ans	7720
De 10 à 19 ans	7397
De 20 à 29 ans	6959
De 30 à 39 ans	7922
De 40 à 49 ans	8689
De 50 à 59 ans	8189
De 60 à 69 ans	7040
De 70 à 79 ans	4384
80 ans ou plus	3125

On a tracé ci-contre la courbe des fréquences cumulées croissantes.

2. À partir de la courbe, déterminer la médiane, et les quartiles de la série statistique.
3. On parle aussi parfois de "classe médiane". Quelle est la classe médiane ?

4. Compléter : Au moins les trois quarts des français ont moins de ... ans.
5. Compléter : Au moins la moitié des français a plus de ... ans.
6. Calculer l'écart interquartiles de la série.



c) Représentation

II.2 Moyenne et écart type

a) Indicateur de tendance centrale : la moyenne

Définition 10.5

La moyenne ... d'une série statistique dont les valeurs du caractère sont x_1, x_2, \dots, x_k et les effectifs associés sont n_1, n_2, \dots, n_k est égale à :

.....

ou :

.....

où f_i est la fréquence associée à la valeur x_i .

Exemple 10.6 :

Âges	[0;20[[20;40[[40;60[[60;80[[80;100[[100;120[
Centre						
Effectifs (milliers)	15 533	15 699	17 114	11 671	3 621	19

Déterminer la moyenne de cette série.

Propriété 10.1

Soient a et b deux réels.
 Si une série de valeurs $\{x_i\}_{1 \leq i \leq p}$ a pour moyenne \bar{x} alors la série de valeurs $\{ax_i + b\}_{1 \leq i \leq n}$ a pour moyenne

Remarque(s) :

- Cela évite d'avoir à refaire tous les calculs lorsque toutes les valeurs subissent la même évolution.

Exemple 10.7 :

On observe les salaires dans une entreprise.
 Le salaire moyen au moment de l'étude est de 1 600 €.
 L'année suivante, la direction décide d'augmenter de 5 % l'ensemble des salaires, puis celle d'après, elle décide de les augmenter de 50 €.
 Quel est alors le salaire moyen dans cette entreprise ?

b) Indicateur de dispersion : l'écart type

Définition 10.6 – Variance

Soient $x_1, x_2, x_3, \dots, x_p$ les valeurs d'une série statistique et $n_1, n_2, n_3, \dots, n_p$ les effectifs associés.

La _____, notée V , de cette série est le nombre positif défini par :

$$V = \dots\dots\dots$$

Remarque(s) :

- La variance est la moyenne du carré des écarts à la moyenne.
- On élève tous les écarts au carré afin d'éviter que des écarts ne se compensent. Ainsi, plus les écarts à la moyenne sont importants, plus la variance est élevée.

Définition 10.7 – Écart-type

L' _____ d'une série statistique, noté σ (« sigma »), est le nombre positif défini par :

$$\dots\dots\dots$$

Exemple 10.8 :

Dans une école, le même contrôle est passé dans deux classes.
 Voici les résultats obtenus dans chacune des classes :

- Classe A : 9 ; 16 ; 11 ; 15 ; 19
- Classe B : 13 ; 10 ; 16 ; 14

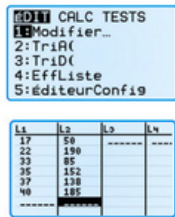
1. Calculer la moyenne et l'écart-type dans chacune des classes.

2. Dans quelle classe le niveau est-il le plus élevé?
3. Dans quelle classe les niveaux sont-ils les plus homogènes?

Avec une calculatrice Texas

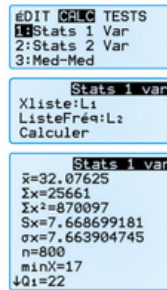
1 Saisie des données

- a. Accéder au mode statistiques (**stats**), choisir le menu **Édit**, sélectionner **1:Modifier...** puis appuyer sur **entrer**.
- b. Saisir les valeurs dans une liste, par exemple L1, et les effectifs dans une autre liste, par exemple L2.



2 Calcul de paramètres de la série statistique

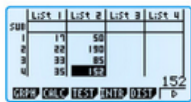
- a. Appuyer sur **stats**.
- b. Choisir le menu **CALC**, puis sélectionner **1:Stats 1 Var** suivi de **entrer**.
- c. Compléter l'écran comme ci-contre, (**2NDE** **1** pour obtenir L1 et **2NDE** **2** pour obtenir L2).
- d. Sélectionner **Calculer** suivi de **entrer**.



Avec une calculatrice Casio

1 Saisie des données

- a. Accéder au menu statistiques : **MENU** puis icône **STAT** et **EXE**.
- b. Saisir les valeurs dans une liste, par exemple List1, et les effectifs dans une autre liste, par exemple List2.



2 Calcul de paramètres de la série statistique

- a. Accéder au menu **CALC** (**F2**), puis au menu **SET** (**F6**).
- b. Choisir List 1 pour 1Var XList et List 2 pour 1Var Freq, suivi de **EXE**. List s'obtient à l'aide de **F1**, on complète ensuite par le numéro de la liste, puis **EXE**.
- c. Sélectionner le menu **1VAR** (**F1**).

