

# 10

## Statistiques descriptives

### I Vocabulaire

- La **population** est un ensemble d'**individus** sur lesquels porte une étude. L'ensemble des données collectées s'appelle une **série statistique**.
- On appelle **caractère** l'aspect de la population étudié. Il peut être :
  - **quantitatif** si ses valeurs (ou modalités) sont des nombres;
  - **qualitatif** si ses valeurs (ou modalités) ne sont pas des nombres.
- Un caractère quantitatif peut être :
  - **discret** s'il peut prendre un nombre fini de valeurs;
  - **continu** s'il peut prendre un nombre infini de valeurs.
- Les valeurs d'un caractère peuvent être regroupées par **classe**.
- L'**effectif** 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 **fréquence** 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<sup>a</sup>.  
On s'intéresse à la statistique "78 %".
  - (a) Quelle est la population étudiée? Quels sont les individus?  
Les couvertures de livres pour enfants.
  - (b) Quel est le caractère? Est-il quantitatif ou qualitatif?  
Genre du personnage figurant sur la couverture.  
Il s'agit d'un caractère qualitatif.
2. On s'intéresse à la pointure des chaussures des élèves d'une classe.
  - (a) Quel est le caractère étudié?  
La pointure des chaussures.
  - (b) Est-il quantitatif ou qualitatif?  
Il s'agit d'un caractère quantitatif.
  - (c) Est-il continu ou discret?  
Il est discret car il y a un nombre fini de valeurs possibles.
3. La température mondiale moyenne a augmenté de 0,74°C au cours du XX<sup>e</sup> siècle, un réchauffement qui touche plus les zones émergées que les océans<sup>b</sup>.  
On s'intéresse à la moyenne "0,74°C".

- (a) Quel est le caractère étudié?  
La température mondiale moyenne.
- (b) Est-il quantitatif ou qualitatif?  
Il est quantitatif.
- (c) Est-il continu ou discret?  
Il s'agit d'un caractère quantitatif continu (il peut y avoir une infinité de valeurs possibles).

- 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'**effectif cumulé croissant** (resp. la **fréquence cumulée croissante**) 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

Âge (en années)	0	1	2	3	4	5	6	7	8	9	10
Effectif	1	2	1	3	5	5	8	4	1	2	2
ECC	1	3	4	7	12	17	25	29	30	32	34
FCC	$\frac{1}{34}$	$\frac{3}{34}$	$\frac{4}{34}$	$\frac{7}{34}$	$\frac{12}{34}$	$\frac{17}{34}$	$\frac{25}{34}$	$\frac{29}{34}$	$\frac{30}{34}$	$\frac{32}{34}$	$\frac{34}{34}$

2. Combien d'individus ont 7 ans ou moins?  
L'effectif cumulé croissant associé à 7 est 29. Donc 29 individus ont 7 ans ou moins.
3. Quelle est la part des individus ayant strictement moins de 5 ans?  
 $\frac{12}{34} \approx 0,3529 = 35,29\%$ .



**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 :**

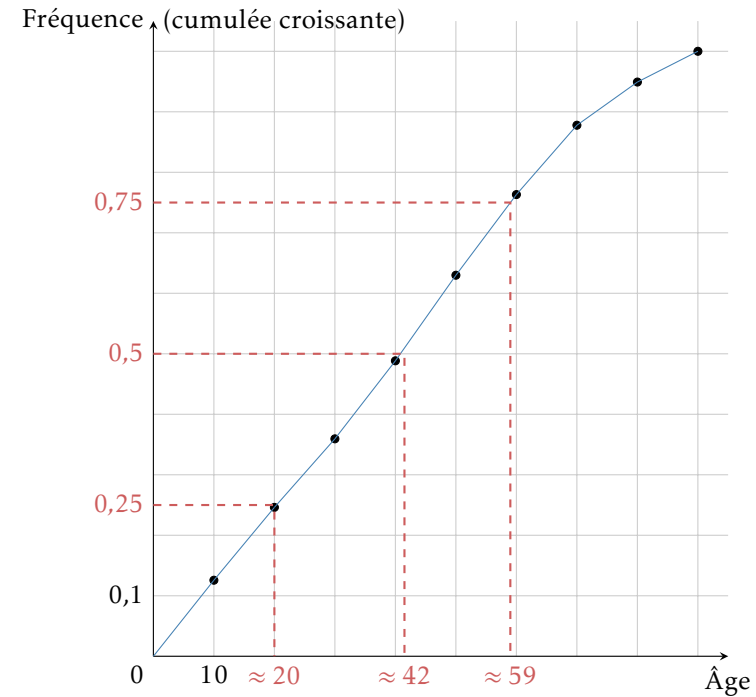
On donne la répartition des âges des français en 2012 (source : INSEE).

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

Âge	Effectif (en milliers)	ECC	FCC
Moins de 10 ans	7 720	7 720	0,1257
De 10 à 19 ans	7 397	15 117	0,2461
De 20 à 29 ans	6 959	22 076	0,3594
De 30 à 39 ans	7 922	29 998	0,4884
De 40 à 49 ans	8 689	38 687	0,6298
De 50 à 59 ans	8 189	46 876	0,7631
De 60 à 69 ans	7 040	53 916	0,8778
De 70 à 79 ans	4 384	58 300	0,9491
80 ans ou plus	3 125	61 425	1

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  $\bar{x}$  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 :

$$\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \dots + n_k \times x_k}{N}$$

ou :

$$\bar{x} = f_1 \times x_1 + f_2 \times x_2 + \dots + f_k \times x_k$$

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	10	30	50	70	90	110
Effectifs (milliers)	15 533	15 699	17 114	11 671	3 621	19

Déterminer la moyenne de cette série.

$N = 63657$ .  $\bar{x} \approx 41$ .

**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  $a\bar{x} + b$ .

**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 ?

$1600 \times 1,05 + 50 = 1730$ .

Donc le salaire moyen dans cette entreprise après augmentations est de 1 730 €.

**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 **variance**, notée  $V$ , de cette série est le nombre positif défini par :

$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + n_3(x_3 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N}$$

**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.

vée.

**Définition 10.7 – Écart-type**

L'**écart-type** d'une série statistique, noté  $\sigma$  (« sigma »), est le nombre positif défini par :

$$\sigma = \sqrt{V}$$

**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.

$$\begin{aligned} \bar{x}_A &= \frac{1 \times 9 + 1 \times 11 + 1 \times 15 + 1 \times 16 + 1 \times 19}{5} \\ &= 14 \end{aligned}$$

$$\begin{aligned} V_A &= \frac{1 \times (9 - 14)^2 + 1 \times (11 - 14)^2 + 1 \times (15 - 14)^2 + 1 \times (16 - 14)^2 + 1 \times (19 - 14)^2}{5} \\ &= 12,8 \end{aligned}$$

$$\begin{aligned} \bar{x}_B &= \frac{1 \times 10 + 1 \times 13 + 1 \times 14 + 1 \times 16}{4} \\ &= 13,25 \end{aligned}$$

$$\begin{aligned} V_B &= \frac{1 \times (10 - 13,25)^2 + 1 \times (13 - 13,25)^2 + 1 \times (14 - 13,25)^2 + 1 \times (16 - 13,25)^2}{4} \\ &\approx 4,69 \end{aligned}$$

- Classe A :  $\bar{x}_A = 14$  et  $\sigma_A \approx 3,58$ .
- Classe B :  $\bar{x}_B = 13,25$  et  $\sigma_B \approx 2,17$ .

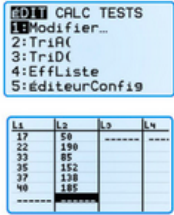
2. Dans quelle classe le niveau est-il le plus élevé ?

3. Dans quelle classe les niveaux sont-ils les plus homogènes ?

### Avec une calculatrice Texas

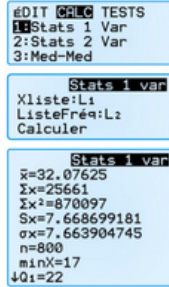
**1** Saisie des données

- 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, ( `2ND` `1` pour obtenir L1 et `2ND` `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`).

