

Statistiques: exemples qui illustrent le cours (voir travaux sur tableur)

L'étude du vocabulaire et des notions de statistique du programme de 1^{ère} L sera illustrée par un exemple: La série statistique de la moyenne annuelle des notes de mathématiques des élèves qui m'avaient comme professeur de mathématiques pendant l'année scolaire 2007-2008.

Les données, rassemblées sur la feuille annexe, ont été structurées en trois séries statistiques:

Série A: elle contient les données brutes: moyennes annuelles (arrondies à un dixième de point près) des notes de mathématiques rangées dans l'ordre croissant.

Série B: elle contient les données regroupées dans des intervalles de un demi point d'amplitude, centrés sur les arrondis des notes à un demi point près.

Série C: elle contient les données regroupées dans des intervalles de un point d'amplitude, centrés sur les arrondis des notes à un point près.

I) Vocabulaire élémentaire

- **Population**: Ensemble étudié. *Exemple: Classe de 1^{ère} L de M. Sicard en 2007-2008.*
- **Individus**: Éléments de la population. *Exemple: Élèves de 1^{ère} L de M. Sicard en 2007-2008.*
- **Caractère étudié ou variable statistique**: Propriété étudiée dans la population.

Exemple: Moyenne annuelle des notes de mathématiques des élèves qui m'avaient comme professeur de mathématiques pendant l'année scolaire 2007-2008. Il s'agit des moyennes arrondies à 0,1 point près pour la série A, arrondies à 0,5 près pour la série B et à 1 près pour la série C.

- **L'effectif d'une modalité** est le nombre d'individus de la population possédant cette valeur du caractère.

Exemple: Pour la série C, l'effectif de la classe $[12,5 ; 13,5[$ de centre 13 est: 6.

- **L'effectif total** est le nombre d'individus de la population: C'est la somme des effectifs de chaque modalité.

Exemple: L'effectif total est le nombre d'élèves de cette classe 1^{ère} L, c'est à dire : 34.

- **La série statistique des effectifs** est la fonction qui, à chaque valeur du caractère (modalité), associe l'effectif de cette modalité.

Elle est le plus souvent définie à l'aide d'un tableau.

Exemple: voir tableau des données statistiques.

Comme pour les fonctions, les séries statistiques peuvent être représentées graphiquement.

Voir fichier graphiques.

- **La série statistique des effectifs cumulés** est la fonction qui à chaque modalité associe la somme des effectifs des modalités de valeurs inférieures ou égale à cette modalité.

Voir fichier graphiques.

- **La série statistique des fréquences** est la fonction qui, à chaque valeur du caractère, associe la fréquence de la classe de ce caractère.

Voir fichier graphiques.

- **La série statistique des fréquences cumulées** est la fonction qui, à chaque valeur du caractère, associe la fréquence cumulée de la classe de ce caractère. Même méthode que pour les effectifs cumulés.

Voir fichier graphiques.

II) Indicateurs de position

- **Le mode** d'une série statistique est la (ou les) modalités ayant le plus grand effectif.

Exemples:

Série A : possède 5 modes d'effectifs 2: 11,8 ; 12,2 ; 12,6 ; 13,8 et 15,4.

Série B : intervalle [11,75 ; 12,25 [de centre 12 (effectif: 6).

Série C : intervalle [11,5 ; 12,5 [de centre 12 (effectif: 8).

- **La médiane**

On ordonne la série des données statistiques par ordre croissant.

Si l'effectif total de la série est impair (de taille: $2n + 1$), la médiane est la valeur du terme de rang $n + 1$ dans cette série ordonnée.

Si l'effectif total de la série est pair (de taille: $2n$), la médiane est la moyenne des valeurs des termes de rang n et $n + 1$ dans cette série ordonnée.

*Ici l'effectif total est pair: $34 = 2 * 17$. $n = 17$ et $n+1 = 18$. On fait la moyenne entre la 17ème et la 18ème modalité. Pour cela, le tableau des effectifs cumulés est très pratique.*

Série A : 12,15.

Série B : intervalle [11,75 ; 12,25 [de centre 12.

Série C : intervalle [11,5 ; 12,5 [de centre 12.

- **La moyenne**

Utilisation du tableur pour être efficace

Série A : 12,01.

Série B : 12.

Série C : 12,06.

III) Indicateurs de dispersion

- **L'étendue** d'une série statistique est la différence entre la valeur minimum et la valeur maximum du caractère étudié.

Série A : $16,2 - 4,9 = 11,3$.

Série B : $16 - 5 = 11$ en prenant les centres.

$16,25 - 4,75 = 11,5$ en prenant les extrémités des intervalles.

Série C : $16 - 5 = 11$ en prenant les centres.

$16,5 - 4,5 = 12$ en prenant les extrémités des intervalles.

- **Les quartiles:**

Pour cela, le tableau des fréquences cumulées est très efficace.

Série A : $q1 = 11$ $q3 = 13,1$.

Série B : $q1 =$ intervalle [10,75 ; 11,25 [de centre 11.

$q3 =$ intervalle [12,75 ; 13,25 [de centre 13.

Série C : $q1 =$ intervalle [10,5 ; 11,5 [de centre 11.

$q3 =$ intervalle [12,5 ; 13,5 [de centre 13..

- **L'intervalle inter quartiles:**

Intervalle qui regroupe la moitié centrale de l'effectif total, c'est à dire situé entre le premier et le troisième quartile.

Série A : [11 ; 13,1].

Série B : [11 ; 13].

Série C : [11 ; 13].

- **L'écart inter quartiles:**

C'est l'étendue entre le premier et le troisième quartile, c'est à dire l'amplitude de l'intervalle inter quartiles.

Série A : $13,1 - 11 = 2,1$ Série B et C : $13 - 11 = 2$.

• **Les déciles d'une série statistique:**

Pour cela, le tableau des fréquences cumulées est très efficace.

Série A : $d1 = 9$ $d9 = 15,4$.

Série B : $d1 =$ intervalle $[8,75 ; 9,25 [$ de centre 9.
 $d9 =$ intervalle $[15,25 ; 15,75 [$ de centre 15,5.

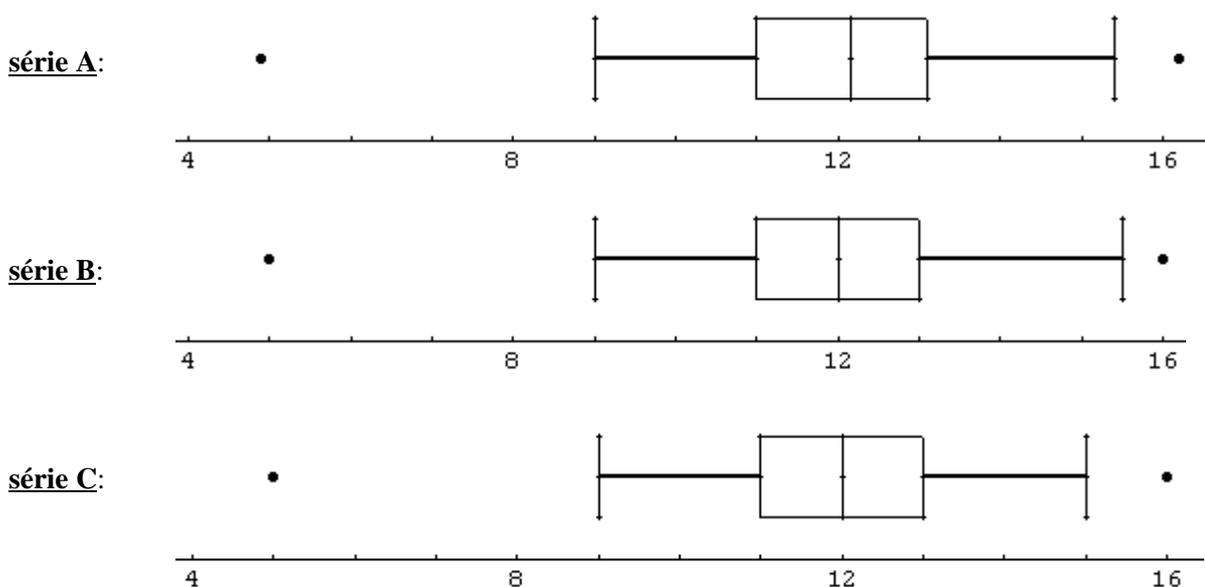
Série C : $d1 =$ intervalle $[8,5 ; 9,5 [$ de centre 9.
 $d9 =$ intervalle $[14,5 ; 15,5 [$ de centre 15..

• **Le diagramme en boîte ou « Boîte à moustache » ou « Boîte à pattes » :**

Récapitulons les résultats obtenus:

	Série A	Série B	Série C
Minimum	4,9	5	5
décile 1	9	9	9
quartile 1	11	11	11
médiane	12,15	12	12
quartile 3	13,1	13	13
décile 9	15,4	15,5	15
maximum	16,2	16	16

On obtient les boîtes à moustache:



• **L'écart-type:**

Ici, il est indispensable d'utiliser un tableur. On obtient:

Série B : variance: 5,29 écart-type: 2,3.

Série C : variance: 5,35 écart-type: 2,31.

• **Plages de normalité:**

En prenant pour moyenne 12 et pour écart-type: 2,3, on obtient:

plage de normalité à 95 %: $[12 - 2 \times 2,3 ; 12 + 2 \times 2,3] = [7,4 ; 16,6]$

plage de normalité à 99 %: $[12 - 3 \times 2,3 ; 12 + 3 \times 2,3] = [5,1 ; 18,9]$

Dans $[7,4 ; 16,6]$ il y a 33 modalités sur 34, soit $\frac{33}{34} \approx 97 \%$

Dans $[5,1 ; 18,9]$ il y a 33 modalités sur 34, soit $\frac{33}{34} \approx 97 \%$