

Chapitre 4 : Statistiques

1 – Généralités

a. Série statistique

Une **série statistique** est un ensemble de données que le statisticien cherche à décrire quantitativement.

Exemple 1 : La couleur des yeux des Français.



On a relevé ci-dessous la couleur des yeux de 30 français. Légende : M=Marron, B=Bleu, V=Vert.

Données : M; B; M; B; B; B; V; M; M; M; M; B; B; M; V; M; M; B; B; B; M; V; V; M; M; M; M; V; B; M

Exemple 2 : Les notes de la classe



Voici les notes d'une classe de 20 élèves pour un DS de maths.

Données : 14 ; 8 ; 8 ; 10 ; 9 ; 14 ; 10 ; 11 ; 11 ; 11 ; 12 ; 9 ; 8 ; 13 ; 9 ; 11 ; 9 ; 10 ; 14 ; 13 ;

Exemple 3 : La taille des nouveaux nés



Voici le relevé de la taille des 12 bébés nés le jour de Noël dans une maternité de La Rochelle

Données (en cm) : 47,8 ; 50,2 ; 49,0 ; 46,7 ; 52,1 ; 50,8 ; 47,9 ; 48,3 ; 52,4 ; 53,4 ; 50,9 ; 51,7.

Activité 1 : Pour chacune des 3 séries statistiques précédentes :

- 1) Déterminer la population, l'échantillon et le caractère étudié.
- 2) Réaliser un tableau d'effectifs et de fréquences
- 3) Représenter la série statistique en utilisant un graphique approprié



b. Vocabulaire statistique

- **Population** : Ensemble que l'on étudie et dont chaque élément est appelé **individu**.
- **Échantillon** : Partie de la population que l'on observe.
- **Caractère** : Propriété observée dans la population. Un caractère est dit **quantitatif** si les valeurs prises par le caractère sont numériques et **qualitatif** sinon. Un caractère quantitatif est dit **discret** s'il ne prend qu'un nombre fini de valeurs, et **continu** s'il peut prendre à priori toutes les valeurs d'un intervalle.
- **Classe** : Ensemble des individus correspondant à une même valeur, ou à un même intervalle de valeurs.

c. Effectifs & Fréquences

On considère une série statistique divisée en p classes.

Définition 1 :

- L'**effectif** n_i d'une classe C_i est le nombre d'éléments qu'elle contient.
- L'**effectif total**, noté N , est la taille de la population (ou de l'échantillon considéré).
- La **fréquence** f_i d'une classe C_i est la proportion d'individu qu'elle contient : $f_i = \frac{n_i}{N}$

Remarque : Lorsque le caractère est quantitatif on peut ranger ses valeurs dans l'ordre croissant. On peut alors parler d'**Effectif cumulé croissant** (ECC) et de **Fréquence cumulée croissante** (FCC).

Propriété 1 :

- On a $N = n_1 + \dots + n_p$
- La somme des fréquences vaut 1 : $f_1 + \dots + f_p = 1$

d. Représentation graphique

Divers **graphiques** permettent d'illustrer une série statistique. Ils permettent de représenter et de l'effectif ou la fréquence de chacune des classes.

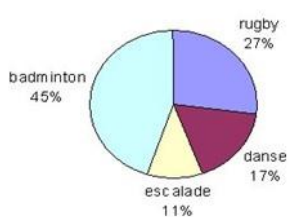
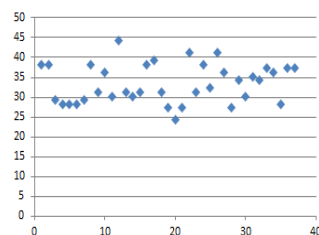
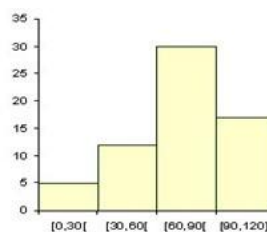


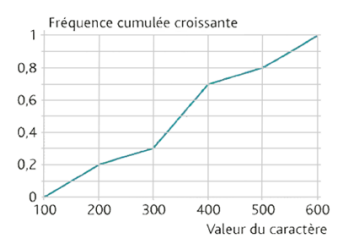
Diagramme circulaire



Nuage de points



L'histogramme



La courbe des fréquences cumulées



2 – Moyenne, Ecart-type

On considère une série statistique prenant p valeurs différentes $x_1, \dots, x_p$ avec les effectifs $n_1, \dots, n_p$.

Définition 2 : La **moyenne** de la série est le nombre : $\bar{x} = \frac{n_1x_1 + \dots + n_px_p}{N} = \frac{1}{N} \sum_{i=1}^p n_i x_i$.

Exemple 4 : Voici les notes en mathématiques d'Antoine et Mathilde au premier trimestre.

Antoine	3	18	19	4
Mathilde	10	12	11	11

Calculer la moyenne de leurs notes.

$$\text{Antoine : } \bar{x} = \frac{3+18+19+4}{4} = \frac{44}{4} = 11$$

$$\text{Mathilde : } \bar{x} = \frac{10+2 \times 11+12}{4} = \frac{44}{4} = 11$$

Activité 2 : Vers une nouvelle mesure statistique.

- 1) Comparer les notes d'Antoine et Mathilde dans l'exemple 4.
- 2) Pour mesurer la « dispersion » de leur résultat nous allons introduire un nouveau paramètre statistique
 - a. Compléter le tableau suivant :

Notes d'Antoine					Somme des carrés des écarts	Moyenne des carrés des écarts
Ecart avec la moyenne						
Carré de l'écart						

La valeur obtenue s'appelle la **variance** de la série. Elle mesure la dispersion des valeurs autour de la moyenne.

- b. Pour revenir à la même unité que les valeurs de la série, calculer la racine carrée de la variance.

La valeur obtenue s'appelle **l'écart-type** de la série.

- 3) On note $\bar{x}$ la moyenne des notes de Mathilde.
 - a. Montrer que la variance des notes de Mathilde est donnée par $V = \frac{(10-\bar{x})^2 + (12-\bar{x})^2 + 2 \times (11-\bar{x})^2}{4}$
 - b. Calculer cette Variance et en déduire l'écart type des notes de Mathilde.
 - c. Comparer l'écart-type des notes d'Antoine et de Mathilde.

Définition 3 :

- La **variance** de la série est le nombre : $V = \frac{n_1(x_1 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N} = \frac{1}{N} \sum_{i=1}^p n_i (x_i - \bar{x})^2$.
- L'**écart-type** de la série est le nombre $\sigma = \sqrt{V}$

Exemple 4 (Suite) : Thomas a eu les notes suivantes en mathématiques au 1^{er} trimestre : 14 – 9 – 15 – 12.

Calculer l'écart-type de ses notes, puis comparer ces résultats avec ceux d'Antoine et Mathilde.

$$\bar{x} = \frac{14+9+15+12}{4} = \frac{50}{4} = 12,5 \quad V = \frac{(14-12,5)^2 + (9-12,5)^2 + (15-12,5)^2 + (12-12,5)^2}{4} = 21 \quad \text{et} \quad \sigma = \sqrt{21} \simeq 4,58.$$

Les résultats de Thomas sont plus dispersés que ceux de Mathilde mais moins que ceux d'Antoine.



3 – Médiane & Quartiles

On considère une série statistique prenant N valeurs rangées par **ordre croissant**.

Exemple 5 : On a étudié la pointure de deux classes de Première.

1)

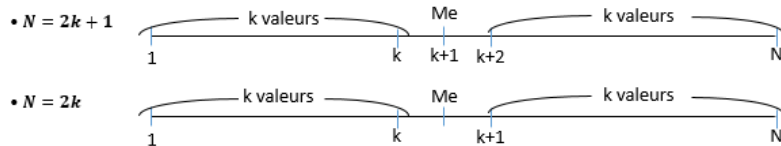
Pointure	37	38	39	40	41	42	43	44	45
Effectifs	1	2	1	2	4	5	7	10	3

2)

Pointure	36	37	38	39	40	41	42	43	44
Effectifs	2	4	4	3	3	2	1	0	1

Définition 4 : La **médiane** est une valeur qui sépare la série en deux parties égales :

- Si N est impair ($N = 2k + 1$), alors la médiane est la $k + 1^{\text{ème}}$ valeur de la série.
- Si N est pair ($N = 2k$), alors la médiane est la moyenne entre la $k^{\text{ème}}$ et la $k + 1^{\text{ème}}$ valeur de la série.



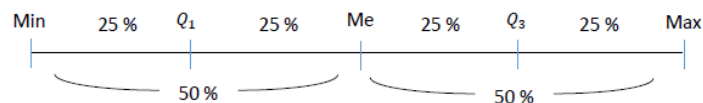
Exemple 5 (Suite) : Calculer la médiane des deux séries de pointure

1) $N = 35 = 2 \times 17 + 1$ est impair. La médiane est la $18^{\text{ème}}$ valeur : $Me = 43$.

2) $N = 20 = 2 \times 10$ est pair. La médiane est la moyenne entre la $10^{\text{ème}}$ et $11^{\text{ème}}$ valeur : $Me = \frac{38+39}{2} = 38,5$.

Définition 5 :

- Le **premier quartile**, noté Q_1 , est la plus petite valeur de la série tel qu'au moins 25% des valeurs de la série soit inférieures ou égales à Q_1 .
- Le **troisième quartile**, noté Q_3 , est la plus petite valeur de la série tel qu'au moins 75% des valeurs de la série soit inférieures ou égales à Q_3 .
- L'**écart interquartile** est la différence $Q_3 - Q_1$.



Exemple 5 (Suite) : Calculer les quartiles des deux séries de pointure.

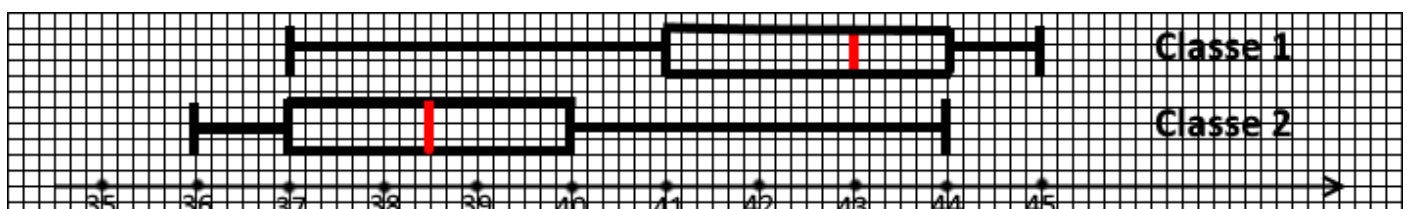
1) $\frac{1}{4} \times 35 = 8,75$ donc Q_1 est la $9^{\text{ème}}$ valeur : $Q_1 = 41$. $\frac{3}{4} \times 35 = 26,25$ donc Q_3 est la $27^{\text{ème}}$ valeur : $Q_3 = 44$.

2) $\frac{1}{4} \times 20 = 5$ donc Q_1 est la $9^{\text{ème}}$ valeur : $Q_1 = 39$. $\frac{3}{4} \times 20 = 15$ donc Q_3 est la $15^{\text{ème}}$ valeur : $Q_3 = 40$

Définition 6 : Le **diagramme en boîte** permet de représenter la médiane et les quartiles d'une série :



Exemple 5 (Suite) : Représenter sur un même axe le diagramme en boîte des deux séries de pointures.



Statistiques – Exercices

1 (Pointure de chaussure)

1) Construction d'une série statistique

Construire une série statistique regroupant les pointures de chaussures des élèves de la classe.

2) Description de la série

- Quelle est la population étudiée ?
- Quelle est l'échantillon prélevé ? Celui-ci est-il représentatif de la population ?
- Quelle est le caractère étudié ? Quel est l'ensemble des valeurs prises par le caractère ? Le caractère est-il qualitatif ou quantitatif ? Discret ou continu ?

3) Effectifs & Fréquences

- Réaliser un tableau d'effectifs et de fréquences de la série.
- Déterminer les effectifs cumulés croissants et fréquences cumulées croissantes de la série.

4) Mesures statistiques

- Calculer la pointure moyenne de la série.
- Calculer l'écart-type de la série.
- Déterminer la valeur médiane de la série
- Déterminer le premier et le troisième quartile de la série.
- En déduire l'écart interquartile de la série.
- Quelle est l'étendue de la série.

5) Représentation graphique

- A l'aide d'un nuage de points représenter les effectifs de chaque pointure.
- Réaliser le diagramme en boîte de la série.

6) Compléter les affirmations suivantes

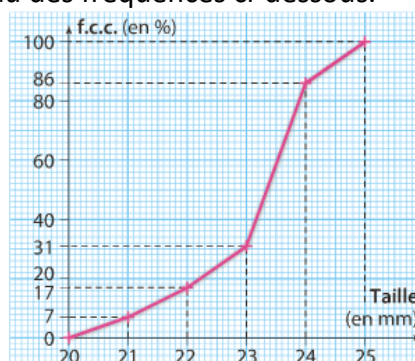
- Au moins 25% des élèves chaussent moins de
- La moitié des élèves chaussent plus de
- 50% des élèves chaussent entre et
- Le quart des élèves chaussent plus de

2 Une station-service a noté les achats de Gazole en Litre sur une journée donnée. Les données ont été rangés par classe.

Volume	Effectifs	Fréquences	F.C.C
[0; 10[	47		
[10; 20[	108		
[20; 30[	173		
[30; 40[	182		
[40; 50[	245		
[50; 60[	168		
[60; 70[	175		
[70; 80[	68		
[80; 90[	22		
[90; 100[	3		

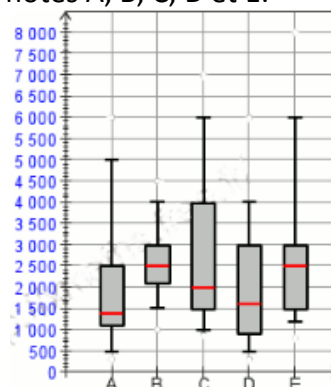
- Compléter les colonnes « Fréquence » et « FCC » du tableau précédent.
- Calculer le volume moyen acheté par les clients.
- Tracer la courbe des fréquences cumulées croissantes.
- A l'aide de cette courbe déterminer la médiane et les quartiles de la série.

3 On a relevé la taille d'une population de frelons. A l'aide de la courbe des f.c.c, compléter le tableau des fréquences ci-dessous.



Taille dans	[20 ; 21[	...	...	...	...
Fréquence	7 %	...	...	...	...

4 Les diagrammes en boîtes donnés ci-dessous représentent la répartition des salaires dans 5 entreprises notés A, B, C, D et E.



Pour chacune des affirmations suivantes, indiquer en justifiant votre réponse, une entreprise pour laquelle cette affirmation est vraie :

- La moitié au moins des salariés ont un salaire qui ne dépasse pas 1500€.
- 25% au moins des salariés gagnent moins de 1000€.
- Au moins 3 salariés sur 4 ont un salaire supérieur à 2000€
- Au moins la moitié des salariés ont un salaire compris entre 2000€ et 3000€
- Le PDG (qui a le plus gros salaire) gagne deux fois plus que 75% des salariés



TP – Population de la région Poitou-Charentes

Nous allons travailler sur le fichier *TP-Stats.ods* qui renseigne le nombre d'habitants des villes de la région Poitou-Charentes en 2006 et 2011. Ces données sont réelles et proviennent du site de l'INSEE.

➤ Copier/coller le fichier *TP-Stats.ods* dans votre dossier Travail puis l'ouvrir avec LibreOffice Calc.

Partie 1 : Lecture de la série

- 1) Rechercher la population de la ville de « La Rochelle » en 2006 et en 2011.
- 2) Dans les cellules C1463 et D1463 déterminer la population totale du Poitou-Charentes en 2006 et en 2011.
- 3) Déterminer le nombre de commune du département de la Charente-Maritime.
- 4) Déterminer la commune de la région qui possédait le moins d'habitants en 2006.
On pourra utiliser la fonction Données/Trier

Partie 2 : Étude statistique de la série.

- 1) A l'aide des fonctions MIN() ; MAX() ; MOYENNE() ; MEDIANE() ; QUARTILE() ; ECARTYPE() du tableur, compléter le tableau 1 : Étude statistique de la série.
- 2) Réaliser sur votre copie un diagramme en boîte pour chacune des deux séries 2006 et 2011.
- 3) Vrai ou Faux ? Justifier.
 - a. Le quart des communes de la région possède moins de 264 habitants en 2006.
 - b. Plus de la moitié des communes possèdent moins de 1000 habitants en 2006.
 - c. Plus de la moitié des communes possèdent moins de 1000 habitants en 2010.
 - d. La moitié des communes possèdent entre 271 et 1050 habitants en 2010.
 - e. Il y a moins de 25% de communes de plus de 1000 habitants en 2006.
 - f. « Saint-Savinien » fait partie des 25% de communes les plus peuplées de la région en 2006 et 2010
- 4) A l'aide des paramètres statistiques comparer la population de la région Poitou-Charentes entre 2006 et 2011.

Partie 3 : Evolution de la population

- 1) Quel est le pourcentage d'évolution de la population de La Rochelle entre 2006 et 2011 ?
- 2) A l'aide du recopiage automatique, compléter la colonne « Evolution en % entre 2006 et 2011 »
- 3) Quel est le pourcentage d'évolution de la population totale de la région entre 2006 et 2011

Partie 4 : Répartition de la population entre les départements en 2011

- 1) Compléter le tableau 2 : Répartition de la population entre les départements en 2011 qui donne les effectifs et les fréquences d'habitants de chaque département.
- 2) Illustrer la répartition de la population entre les départements en 2011, à l'aide d'un diagramme circulaire.

