

~ COURS ~
STATISTIQUES DESCRIPTIVES

Table des matières

I) Présentation des données	2
I.1. Les diagrammes	2
I.2. Histogramme	2
II) Diagramme en boîte	3
II.1. Définitions	3
II.2. Interprétation des indicateurs	4
III) Regroupement par classe	8

I) Présentation des données



Travail de l'élève 1 : Math'x 2nde p 164 "Choisir et interpréter un graphique"

I.1. Les diagrammes

On les utilise pour représenter graphiquement une série statistique dont le caractère est quantitatif et prend des valeurs isolées ou est qualitatif.

↪ **Pour le diagramme en bâton** : on représente sur l'axe des abscisses les différentes valeurs du caractère dans l'ordre croissant (aléatoirement pour un caractère qualitatif) et sur l'axe des ordonnées les effectifs.

On peut également représenter un nuage de point à la place des bâtons.

↪ **Pour le diagramme circulaire** : l'angle d'ouverture de chaque secteur est proportionnel à l'effectif (ou encore à la fréquence) :

$$\text{Angle} = \frac{\text{Effectif de la valeur}}{\text{Effectif total}} \times 360 = \text{Fréquence} \times 360$$

I.2. Histogramme

On l'utilise pour représenter graphiquement une série statistique dont le caractère est quantitatif et regroupé en classes par intervalles. Il est constitué de rectangles dont les bases correspondent aux intervalles et dont les aires sont proportionnelles aux effectifs ou aux fréquences des classes.

Pour lire un histogramme, il faut :

↪ déterminer les aires des rectangles correspondants à chaque classe ;

↪ interpréter ces aires en termes d'effectifs à l'aide de la légende.



Exemple :

Exemples du livre Hachette p 122-123



Exercice(s) du livre : Hachette : 1 p 123 + 8-11 p 134 + 14-16-18 p 135

En interro : 1 p 118 + 17 p 135

(Prévoir un TP tableur les autres années, par exemple le TP4 p 93 du Foucher)

II) Diagramme en boîte

II.1. Définitions

 **Travail de l'élève 2 :** Mettre les élèves pieds nus et les ranger dans l'ordre croissant de taille.

- ↪ Demander pour quelle taille on a autant d'élèves en dessous qu'au dessus ? Se mettre ou non dans le groupe pour avoir un effectif impair, puis changer pour avoir un effectif pair.
- ↪ Demander quelle est la taille pour laquelle un quart des élèves sont plus petits ? 3 quarts ? 10% ? 90% ?
- ↪ Résumer les définitions
- ↪ Dessiner le diagramme en boîte correspondant et schématiser les proportions.

Définition 1.

Dans une série statistique de n termes *classés par ordre croissant*, la **médiane** Me est :

- ↪ le terme du milieu si n est impair
- ↪ la moyenne des deux termes du milieu si n est pair

La médiane partage les termes de la série rangés par ordre croissant en deux groupes de "même effectif".

Faire un schéma

Définition 2.

Dans une série statistique de n termes *classés par ordre croissant* :

Le **premier quartile** $Q1$ est le plus petit terme de la série tel qu'au moins 25% des termes soient inférieurs ou égaux à $Q1$. Les **deuxième** et **troisième quartile** $Q2$ et $Q3$ sont définis de même en remplaçant 25% par 50% (25×2) et 75% (25×3).

On définit de manière analogue les **déciles** pour 10%, 20%, 30% ..., ou encore les centiles ...


Ainsi, sur le même principe que la médiane, les quartiles partagent la série en 4 groupes d'environ le "même effectif" (un quart chacun) et les déciles en 10 groupes.

Faire des schémas

L'**intervalle interquartile** [$Q1$; $Q3$] contient donc environ 50% des termes de la série statistique.


L'**intervalle interdécile** [$D1$; $D9$] contient donc environ 80% des termes de la série statistique.

Les diagrammes en boîtes résument graphiquement ces indicateurs, auquel on rajoute les termes minimum et maximum de la série.

 **Exercice 1 :** Pour chacune des séries statistiques suivantes, calculer la moyenne et donner la médiane, les quartiles et les déciles.

1. 36,37,38,39,40,41,42,43,44 ;
2. 1, 2, 2, 4, 4, 5, 8, 8, 8, 80, 80, 90
3. 10, 10, 10, 1100, 2500, 3000, 4000, 5000
4. Les élèves des 1ST2S ont obtenus les notes suivantes à un devoir de mathématiques :

Note	8	9	10	11	12	13	14
Effectif	6	2	7	5	9	19	12

 **Exercice 2** : Ce tableau donne les salaires mensuels, en euros, des 185 employés d'une entreprise.

Salaire brut	1405	1480	1554	1870	2739	4215
Effectifs	40	15	81	35	9	5
Effectifs cumulés croissants						

1. Compléter le tableau ci-dessus.
2. Déterminer pour cette entreprise le salaire médian et les quartiles Q1 et Q3.
3. Représenter le diagramme en boîte de cette série.
4. Déterminer l'étendue et l'écart interquartile.
5. Déterminer le salaire moyen de cette série.
6. Utiliser les indicateurs précédents pour répondre aux questions suivantes et justifier :
 - a. Les salaires sont-ils dispersés ?
 - b. Si les salaires étaient égaux, quelle somme recevrait chaque employé ?
 - c. Peut-on affirmer ici que la moitié des salariés gagnent moins que la moyenne ?
 - d. Au moins 25% des employés ont un salaire inférieur à 1450€.
 - e. Au moins 50% des employés ont un salaire supérieur à 1500€.
 - f. Au moins 25% des employés ont un salaire entre 1500 et 1800€.
 - g. Au moins 25% des employés ont un salaire supérieur à 1800€.



Solution :

moy=1705.13, Min=1405, Q1=1480, Me=1554, Q3=1870, Max=4215



Exercice(s) du livre : 7 p 134 + 22-25 p 136

II.2. Interprétation des indicateurs



Travail de l'élève 3 : Les tableaux ci-dessous donnent les séries des âges de premiers investiture des présidents de la République en France (de Adolphe Thiers en 1871 à François Hollande en 2012) et aux Etats Unis (de Georges Washington en 1789 à Donald Trump en 2016).

Comparer ces deux séries avec un tableur ou une calculatrice.

F	74	65	72	50	47	54	61	65	53	65	61	61
	74	61	63	72	69	58	48	65	63	52	57	
USA	57	62	58	58	59	58	62	55	68	51	50	65
	50	49	66	52	57	47	55	50	51	48	56	56
	54	43	52	57	56	51	55	51	61	63	44	55
	56	61	53	70	64	46	54	47	70			

Pour comparer deux séries, on compare :

↪ leur tendance centrale avec moyenne et médiane

↪ leur dispersion le rapport l'étendue/l'écart interquartile.

Grâce aux indicateurs calculés, expliquer pourquoi on peut affirmer que le pourcentage de présidents ayant moins de 57 ans lors de leur investiture est plus important aux USA qu'en France.



Solution :

Pour la France, on trouve moy=61.3, Min=47, Q1=54, Me=61, Q3=65, Max=74

Pour les USA, on trouve moy=55.4, Min=43, Q1=51, Me=55, Q3=59, Max=70

L'âge moyen et l'âge médian sont tous les deux plus faibles aux USA qu'en France.

L'étendue est la même pour les deux séries, mais l'écart interquartile est plus faible aux USA, donc la dispersion des âges est moindre.

Remarques :

↪ La moyenne et la médiane sont des caractéristiques de **position**. Elles donnent des tendances centrales.

La moyenne est par contre sensible aux valeurs extrêmes de la série (la médiane non).

↪ L'étendue, l'écart interquartile, et l'écart interdécile sont des caractéristiques de **dispersion**.

On les compare souvent pour voir l'homogénéité de position et de dispersion d'une série (ou son hétérogénéité) :

— plus l'écart interquartile est grand par rapport à l'étendue, plus la série est dispersée.

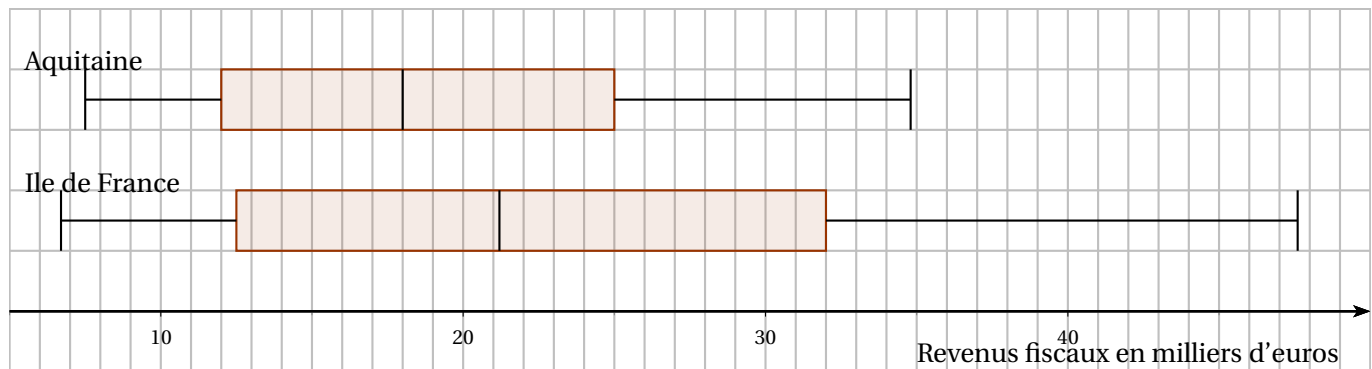
— plus l'écart interquartile est décalé (vers la gauche ou la droite de l'axe), moins la série est symétrique.



Exemple :

Les deux diagrammes en boîte ci-dessous représentent les revenus fiscaux des ménages en 2008 en Ile de France et en Aquitaine.

Les extrémités des diagrammes correspondent aux premier et neuvième déciles.



1. Expliquer le choix de ne pas faire apparaître les valeurs min et max des salaires ?


2. Quelles conclusions peut-on tirer de l'observation de ces diagrammes ?

La médiane permet de dire qu'il y a plus de salaires élevés en Ile de France qu'en Aquitaine.

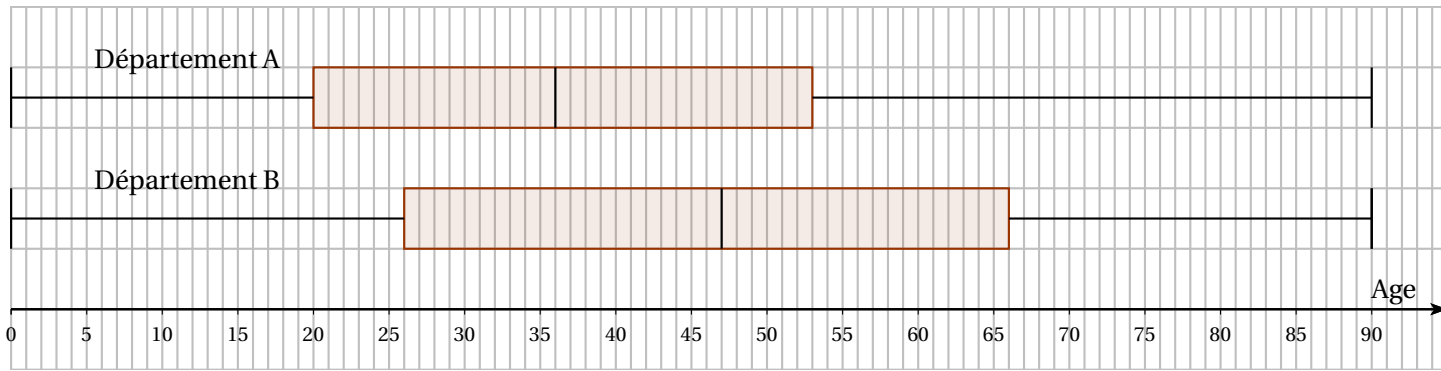
Les deux premiers quartiles sont sensiblement identiques, alors que le 3e quartile est beaucoup plus important en Ile de France, le 9e décile aussi. Ainsi, les salaires plus élevés sont plus nombreux, mais également plus élevés globalement en Ile de France.

Par contre, le 1er décile indique que les salaires les plus faibles sont légèrement plus bas en Ile de France.

Enfin, l'écart interquartile (comme l'interdécile) indique que la dispersion des salaires est bien plus importante en Ile de France.

 **Exercice 3 :** On donne ci-après les diagrammes en boîtes représentant les âges des habitants de moins de 90 ans en 2012 pour les deux départements français, nommés ici A et B.

Par lecture graphique, avec la précision permises par les diagrammes, justifier les affirmations suivantes en précisant les paramètres statistiques utilisés.

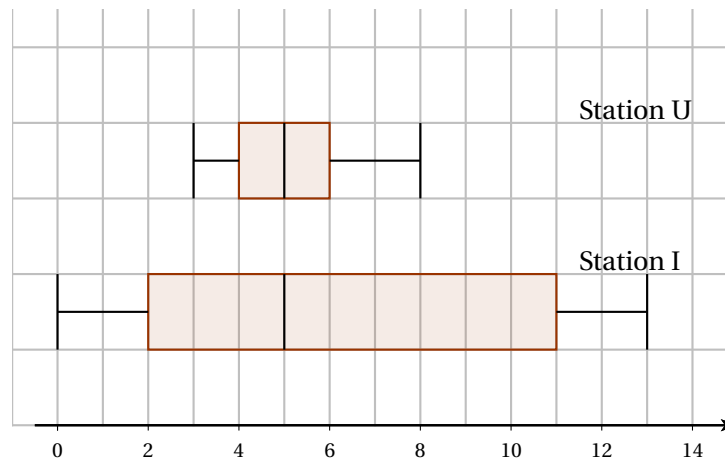


1. La proportion de personnes âgées de moins de 25 ans est inférieure dans le département B à la proportion analogue dans le département A.
2. La dispersion des âges des habitants autour de l'âge médian est plus importante pour le département B que pour le département A.

Exercice 4 : Dans une région de l'est de la France, la pollution atmosphérique est contrôlée quotidiennement, heure par heure, par un réseau de 21 stations de mesures. Parmi celles-ci, nous considérons la station notée U, qui est située en zone urbaine et la station I, en zone industrialisée et la station R, en zone rurale de moyenne montagne.

1. Pour cette partie, on compare les mesures obtenues aux stations U et I pour le dioxyde de soufre, un des polluants mesurés. Les concentrations de ce polluants sont exprimées en millièmes de gramme par mètre cube d'air (noté $\mu\text{g}/\text{m}^3$).

Les diagrammes en boîtes dessinés ci-après concernent les mesures horaires du polluants aux stations U et I, pour la journée du 16 novembre 2006. Les extrémités du diagrammes correspondent aux valeurs minimales et maximales.




- a. Pour chacune des deux stations, indiquer la médiane et calculer l'écart interquartile, ainsi que l'étendue de la série de mesures.
- b. Indiquer par lecture graphique et en précisant les paramètres statistiques utilisés, sur quelle(s) station(s) ce jour là :
 - ↪ la dispersion des mesures a été la plus importante ?
 - ↪ la moitié des mesures au moins ont été inférieures ou égales à 5 ?
 - ↪ 75% des mesures au moins ont été inférieures ou égales à 6 ?

2. Dans cette partie, on considère la station R. Le tableau ci-dessous donne les relevés horaires, pour la même journée du 16 novembre 2006 en ce qui concerne le polluant ozone. Les concentrations sont exprimées en millièmes de gramme par mètre cube d'air.

Heure	1	2	3	4	5	6	7	8	9	10	11	12
Concentration	78	79	77	59	57	65	65	67	68	67	59	54
Heure	13	14	15	16	17	18	19	20	21	22	23	24
Concentration	64	68	72	74	72	72	76	77	76	74	77	76

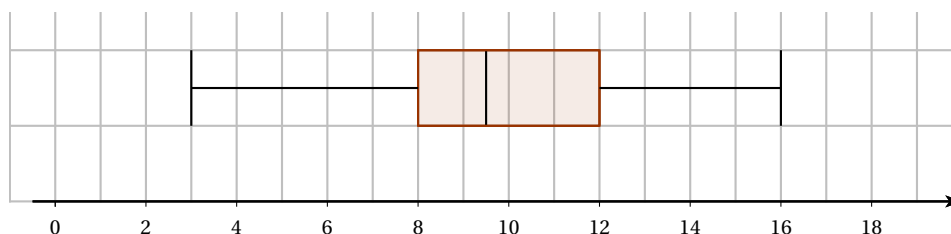
- Quelles sont les valeurs minimale et maximale de cette série?
- Déterminer la médiane et les quartiles de cette série.
- Calculer la moyenne de cette série.
- Construite le diagramme en boîte de cette série.

 **Exercice 5** : Dans un lycée, on étudie les moyennes trimestrielles du premier trimestre de deux classes appelées respectivement Jaune et Rouge.

- Les 25 élèves de la classe Jaune ont obtenu les moyennes trimestrielles suivantes au premier trimestre :

Notes	3	4	5	7	10	11	12	13	14	15	16	18
Effectifs	1	1	1	2	5	2	5	3	1	1	1	1
Effectifs cumulés croissants												

- Déterminer la médiane Me , le premier quartile $Q1$ et le troisième quartile $Q3$ de cette série statistique.
 - Représenter le diagramme en boîte correspondant.
 - Corriger l'une des deux phrases suivantes afin qu'elles soient toutes les deux vraies :
 - ↪ au moins 25% des élèves de la classe Jaune n'ont pas la moyenne
 - ↪ au moins 75% des élèves de la classe Jaune ont la moyenne
 - Calculer la moyenne trimestrielle de la classe Jaune (obtenu en faisant la moyenne des moyennes individuelles)
- On donne le diagramme en boîte suivant de la classe Rouge :




Parmi les affirmations suivantes, lesquelles sont vraies, fausses ou indécidables (ie que l'on ne peut pas conclure avec les éléments connus). **Justifier**

- au moins 50% des élèves de la classe Rouge ont une note comprise entre 10 et 12
- au moins 50% des élèves de la classe Rouge ont une note inférieure ou égale à 10
- au moins 75% des élèves de la classe Rouge ont une note inférieure ou égale à 12
- environ 12.5% des élèves de la classe Rouge ont une note supérieure ou égale à 14
- les notes de la classes Jaune sont plus dispersée que celle de la classe Rouge

- f. le pourcentage d'élèves ayant une note inférieure ou égale à 11 est plus petit dans la classe Rouge que dans la classe Jaune

 **Exercice(s) du livre** : Hachette : 12 p 134 + 26-27 p 137
Foucher : exo 48 p 112

III) Regroupement par classe

 **Travail de l'élève 4** : Hachette : Activité 4 p 121 (rajouter la médiane avec le polygone des effectifs cumulés croissants)

Dans le cas d'une variable continue ou de valeurs regroupées par classes, on obtient une valeur approchée de la médiane et des quartiles à l'aide du polygone des fréquences cumulées croissantes ou des effectifs cumulés croissants.

On obtient une valeur approchée de la moyenne en prenant pour valeurs prises par la variable les centres des classes.

 **Exemple :**

Celui du livre p 128, avec la moyenne.

 **Exercice(s) du livre** : 6 p 129 + 42 p 140