

Cours 2

Représentations graphiques d'une variable statistique

Ressource R2.08 - Statistique Descriptive

2023-2024

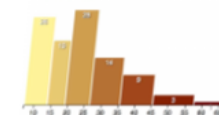
Plan du cours

1. Types de données, découpage en classes

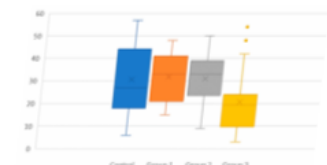
2. Diagrammes en bâtons



3. Histogrammes

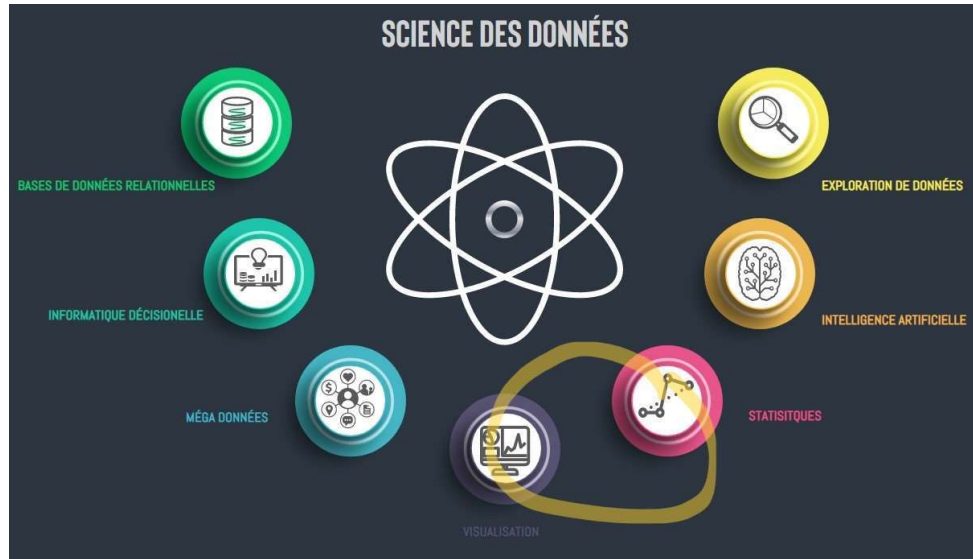


4. Boîtes à moustaches de Tukey



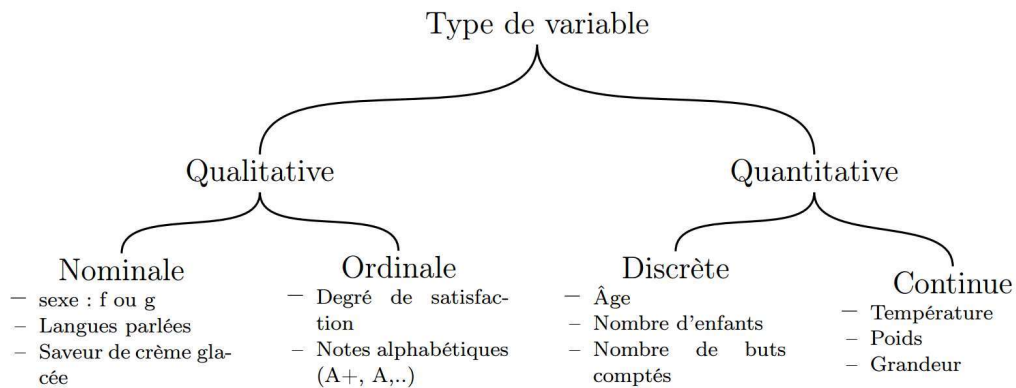


Les représentations graphiques font partie des outils de la statistique. En science des données, ces représentations font partie de la statistique mais aussi de la **visualisation des données (datavisualisation)**.



1. Types de données, découpage en classes

• Rappel : types de données statistiques



Dans cette ressource, on travaillera essentiellement avec des **variables quantitatives**, mais certaines des représentations graphiques de ce cours s'appliquent à d'autres types de variables.

• Découpage en classes (pour une variable continue)

Définition

Soit X une variable statistique quantitative continue.

Faire un **découpage en classes** pour la variable X , c'est donner un ensemble d'intervalles qui recouvrent l'ensemble de toutes les valeurs pouvant être prises par X .

Exemple 2 : Tailles des étudiants

Comme l'ensemble des valeurs prises par cette variable Y est grand, on peut choisir de le découper en classes en prenant une liste d'intervalles recouvrant toutes les tailles :

$[150, 160[$, $[160, 170[$, $[170, 180[$, $[180, 190[$, $[190, 200[$.

Plusieurs découpages en classes sont possibles. On pourrait par exemple choisir aussi cette liste d'intervalles :

$[150, 165[$, $[165, 170[$, $[170, 175[$, $[175, 180[$, $[180, 190[$, $[190, 200[$.

• Tableau statistique par classes

On peut ensuite faire le tableau statistique de cette variable par classes.

Avec le premier découpage en classes :

Classe	$[150, 160[$	$[160, 170[$	$[170, 180[$	$[180, 190[$	$[190, 200[$
Effectif	4	18	38	32	8
Fréquence	0.04	0.18	0.38	0.32	0.08

Avec le second découpage en classes :

Classe	$[150, 165[$	$[165, 170[$	$[170, 175[$	$[175, 180[$	$[180, 190[$	$[190, 200[$
Effectif						
Fréquence						

La variable statistique continue Y est ainsi transformée en une variable discrète, avec un nombre fini de valeurs possibles : chaque valeur est une classe.



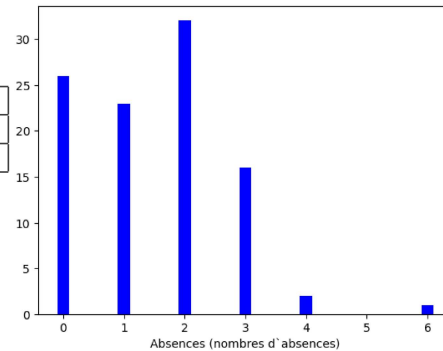
2. Diagrammes en bâtons

Définition

Soit X une variable statistique qualitative, ou quantitative discrète.
Un **diagramme en bâtons** est la représentation graphique des données en dessinant des colonnes de **hauteurs proportionnelles à l'effectif** pour chaque valeur prise par X .

Exemple 1 : Absences des étudiants

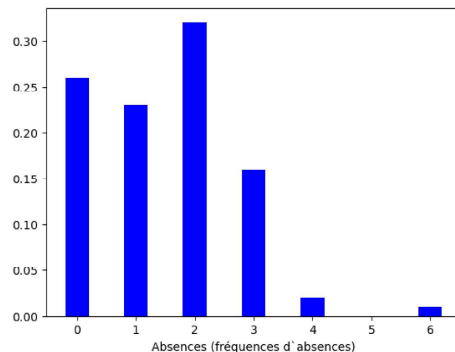
valeur	0	1	2	3	4	6
effectif	26	23	32	16	2	1
fréquence	0.26	0.23	0.32	0.16	0.02	0.01



5

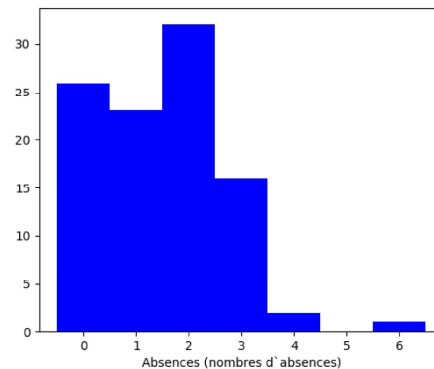
Les colonnes sont de **hauteurs proportionnelles à l'effectif**, on peut donc changer d'échelle comme on le souhaite.

Ci-dessous par exemple, les hauteurs des colonnes sont les fréquences des absences :



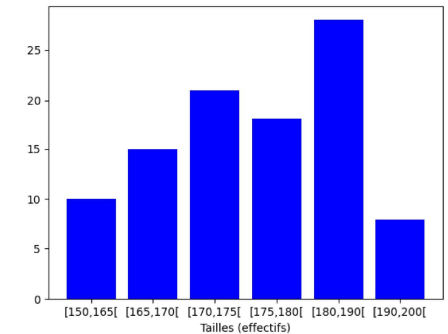
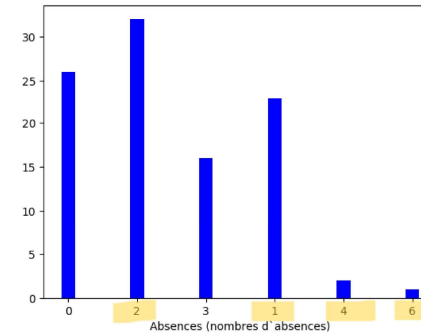
La largeur des colonnes n'a pas d'importance.

Ci-dessous par exemple, les colonnes remplissent l'espace... pour autant, ce n'est pas un histogramme mais bien toujours un diagramme en bâtons !



6

L'axe horizontal ne suit pas nécessairement une échelle. C'est pourquoi cette représentation est adaptée aussi aux variables qualitatives.



Le diagramme à barre donne la répartition des moyens de transport d'élèves pour venir au collège.

en voiture : 6 élèves
en bus : 17 élèves
à pied : 14 élèves
à vélo : 10 élèves

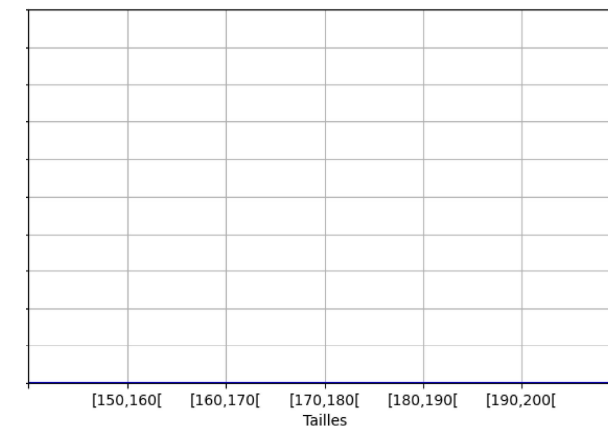


7

Exercice : Taille des étudiants

Avec le découpage en classes utilisé dans le tableau suivant, compléter le diagramme en bâtons ci-dessous :

Classe	[150, 160[[160, 170[[170, 180[[180, 190[[190, 200[
Effectif	4	18	38	32	8
Fréquence	0.04	0.18	0.38	0.32	0.08



8



3. Histogrammes

Définition

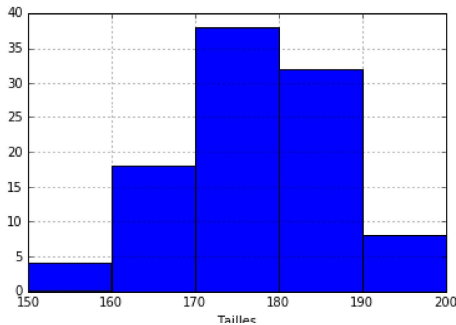
Soit X une variable statistique continue, avec un découpage en classes

$$[x_1, x_2[, [x_2, x_3[, \dots, [x_{n-1}, x_n]$$

Un **histogramme** est la représentation graphique des données en indiquant en abscisse les bords des classes x_1, x_2, \dots, x_n (à l'échelle) et en dessinant des rectangles de **surfaces proportionnelles à l'effectif** de chaque classe.

Exemple : Tailles des étudiants

[150, 160[[160, 170[[170, 180[[180, 190[[190, 200[
4	18	38	32	8



Les surfaces des rectangles ci-contre sont bien proportionnelles aux effectifs, car chaque rectangle est de largeur 10. Leurs surfaces respectives sont donc :

- 1e rectangle : $10 \times 4 = 40$
- 2e rectangle :
- 3e rectangle :
- 4e rectangle :
- 5e rectangle :

9



Dans le cas général, la hauteur des rectangles n'est pas égale aux effectifs (ni aux fréquences).

• Effectifs corrigés, fréquences corrigées

Définition

Soit n_i l'effectif de la classe $[x_i, x_{i+1}[$, et f_i sa fréquence.

On introduit

$$n_i^c = \frac{n_i}{|x_{i+1} - x_i|}, \quad \text{et} \quad f_i^c = \frac{f_i}{|x_{i+1} - x_i|}$$

- n_i^c est appelé l'**effectif corrigé** de la i -ème classe,
- f_i^c est appelée la **fréquence corrigée** de la i -ème classe.

Hauteur des rectangles

Si les intervalles ne sont pas tous de même longueur, pour que les surfaces soient proportionnelles à l'effectif de chaque modalité, la hauteur du i -ème rectangle de l'histogramme est donnée par :

- l'effectif corrigé n_i^c si l'unité choisie est l'effectif,
- la fréquence corrigée f_i^c si on a choisi les fréquences comme unités de l'histogramme.

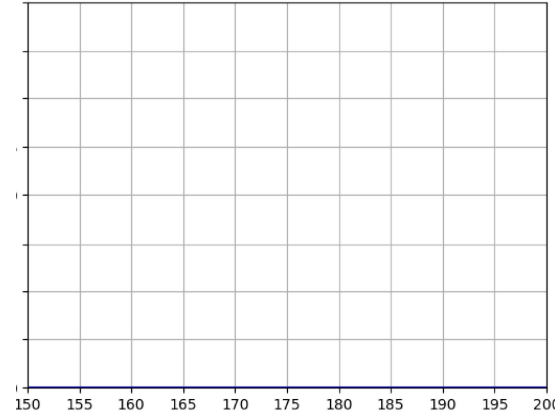
10

Exercice : Tailles des étudiants.es, avec classes de longueurs différentes

Compléter le tableau statistique suivant avec les effectifs corrigés, puis tracer l'histogramme associé.

Classes	[150, 165[[165, 170[[170, 175[[175, 180[[180, 190[[190, 200[
Effectifs n_i	10	15	21	18	28	8
n_i^c						

Surfaces des rectangles ci-contre :



- 1e rectangle : $15 \times \dots$
- 2e rectangle :
- 3e rectangle :
- 4e rectangle :
- 5e rectangle :
- 6e rectangle :

11

12

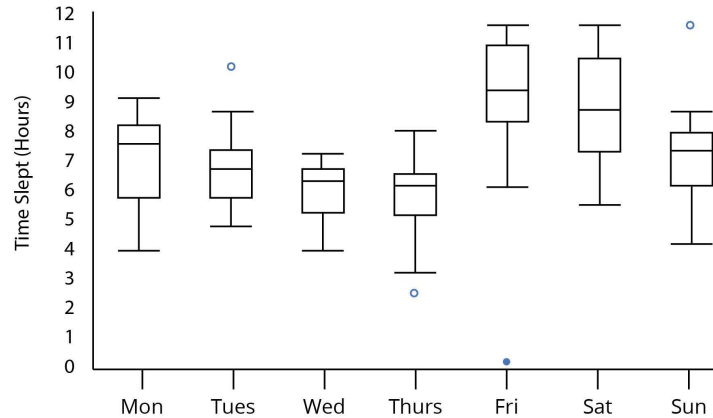


4. Boîte à moustaches de Tukey

La boîte à moustaches résume graphiquement les principales caractéristiques d'une variable statistique : médiane, quartiles, minimum, maximum, déciles.

Ce type de diagramme sert principalement à comparer des variables statistiques.

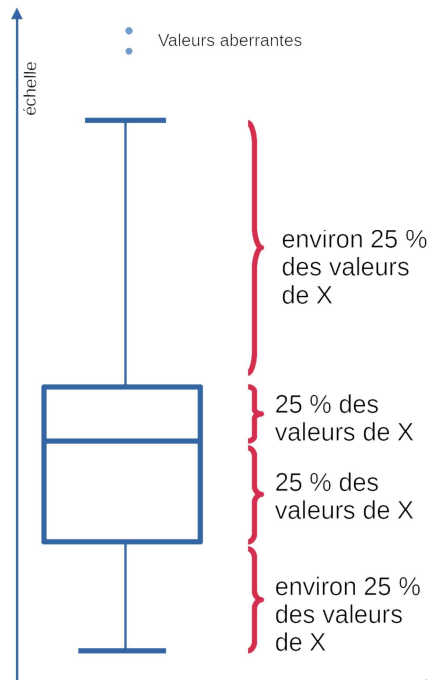
Exemple : Boîtes à moustaches représentant le nombre d'heures de sommeil selon les jours de la semaine pour un ensemble d'individus.



13

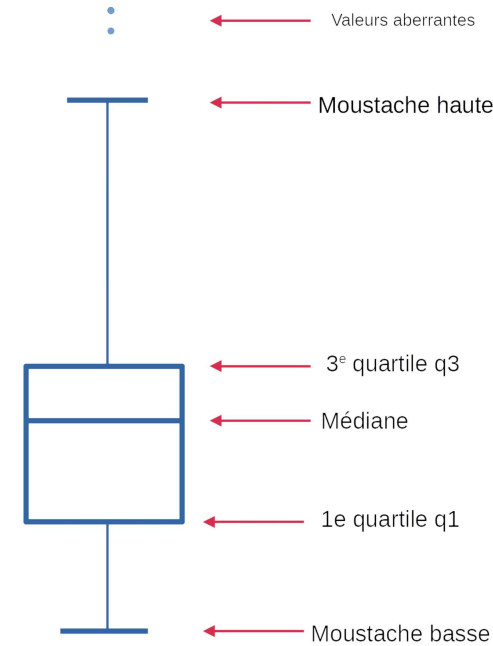
• Lecture d'une boîte à moustaches

La boîte à moustaches indique sur l'échelle verticale où se situent les 4 intervalles contenant 25% des valeurs de X .



14

• Construction d'une boîte à moustaches



Définition : la boîte

Pour une variable statistique X , la boîte à moustaches est composée d'abord du rectangle central (la boîte) :

- le haut du rectangle correspond au 3e quartile q_3 de X
- le trait au milieu du rectangle est au niveau de la médiane de X
- le bas du rectangle correspond au 1e quartile q_1 de X

La largeur de la boîte est libre, elle n'a aucune signification.

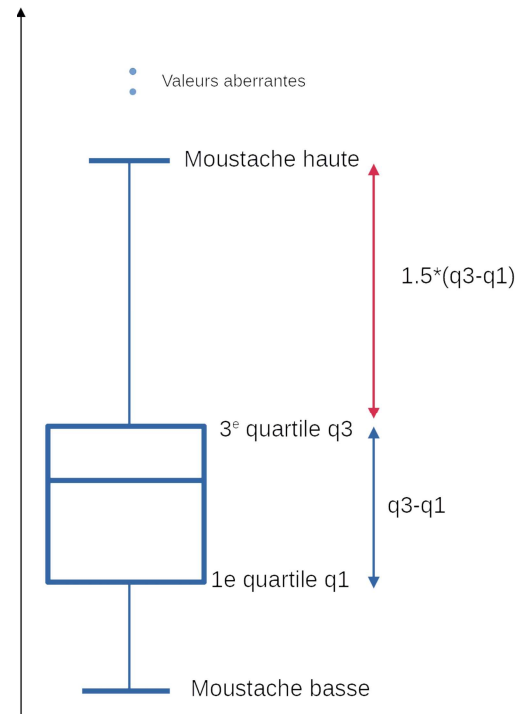
15

Définition : les moustaches

La boîte est ensuite prolongée de "moustaches". Leur longueur est au plus de $1.5 \times (q_3 - q_1)$. Plus précisément :

- la moustache haute est en $q_3 + 1.5 \times (q_3 - q_1)$, sauf si la plus grande valeur prise par X (= le maximum de X) est plus basse, auquel cas le tiret du haut est situé au maximum de X .
- la moustache basse est en $q_1 - 1.5 \times (q_3 - q_1)$, sauf si la plus petite valeur prise par X (= le minimum de X) est plus haute, auquel cas le tiret du bas est situé au minimum de X .

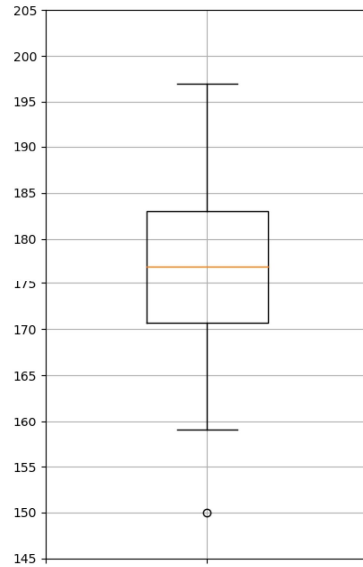
Si X a des valeurs au-delà d'une des moustaches, ces valeurs sont représentées par des ronds.



16

• Exemples

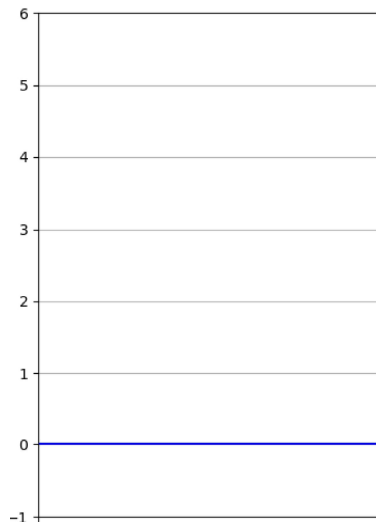
Pour les tailles des étudiant.es, on obtient la boîte à moustaches suivante :



17

• Exercice

Réaliser la boîte à moustaches pour l'exemple des absences des étudiant.es



18



2e QCM Moodle : à faire avant lundi 8/04 23h59

19

Vous trouverez ce QCM sur Moodle tout en bas de la section **Semaine 2**.

Il sera composé des questions suivantes :

1. On vous donne une série statistique et un découpage en classes.
Dire si on a besoin de calculer les effectifs (ou fréquences) corrigés pour tracer l'histogramme.
2. On vous donne le tableau statistique contenant les fréquences d'une variable statistique.
Vous devez calculer la fonction de répartition F en une valeur qui vous est donnée.

20