STATISTIQUES

Médiane, quartiles et déciles; Diagrammes en boîtes

Dans ce chapitre, on suppose qu'on dispose d'une série ordonnée: les valeurs ont été rangées dans l'ordre croissant, de la plus petite à la plus grande.

Médiane

> La médiane sépare une série statistique en deux groupes de même effectif, l'un contient les valeurs les plus petites et l'autre les valeurs les plus grandes.

Comment déterminer la médiane d'une série de *N* valeurs si *N* est pair:
- on vérifie que les valeurs sont rangées par ordre croissant
- on prend la moyenne des deux valeurs situées au milieu: pour cela on calcule l'entier n = N:2, et on calcule la moyenne entre la $n^{ième}$ et la $(n+1)^{ème}$ valeur

Exemple
Prenons les valeurs (notes à un DS, vitesses de vents, ...) rangées dans l'ordre croissant :
*1-3-3-3-5-5-6-7-7-8-8-8-9-**9-10**-10-10-10-11-11-12-13-13-13-14-15-16-19*
Il y a N = 28 valeurs; N:2 = 14; les deux valeurs du milieu sont la 14ème et la 15 ème qui sont 9 et 10 ; la médiane est la moyenne entre la 14ème et le 15ème valeur de la série rangée dans l'ordre croissant, cad Me = 9,5

Comment déterminer la médiane d'une série de *N* valeurs si *N* est impair:
- on vérifie que les valeurs sont rangées par ordre croissant
- on prend la valeur située au centre de la série: pour cela on calcule le décimal N : 2 , sa partie entière n est l'effectif des deux sous groupes encadrant la médiane, la médiane est donc la (n+1)ème valeur

Exemple
Prenons les valeurs (notes à un DS, vitesses de vents, ...) rangées dans l'ordre croissant :
*3-5-5-6-7-8-8-9-9-10-10-**10**-10-11-11-12-13-13-13-14-15-16-19*
Il y a N = 23 valeurs; N:2 = 11,5, la médiane est la 12ème valeur de la série rangée dans l'ordre croissant, cad Me = 10

Comment interpréter une médiane donnée?
si on connait la médiane d'une série, que peut-on en déduire?
Au moins **la moitié (50%) des valeurs sont inférieures** ou égales **à la médiane.**
Au moins **la moitié (50%) des valeurs sont supérieures** ou égales **à la médiane**.

Exemple: dans une classe, la médiane des notes à un contrôle est 11. On peut dire que:
*au moins **la moitié des élèves a** une note **inférieure** ou égale **à 11***
*au moins **la moitié des élèves a** pour note 11 ou **moins de 11***

*au moins **la moitié des élèves a** une note **supérieure** ou égale **à 11***
*au moins **la moitié des élèves** pour note 11 ou **plus de 11***

Quartiles

> Les quartiles permettent de séparer une série statistique en quatre groupes de même effectif (à une unité près).

> Un quart des valeurs sont inférieures au premier quartile Q1.
> Un quart des valeurs sont supérieures au troisième quartile Q3.

On appelle **intervalle interquartile l'intervalle]Q1; Q3[**.

On appelle **écart interquartile la différence Q3 – Q1**.

Comment déterminer les quartiles Q1 et Q3 d'une série de *N* valeurs ?
on calcule la quantité ¼ de N = $\frac{1}{4} \times N$ =N:4
Deux cas sont possibles: soit le résultat est entier (la division tombe juste), soit non

cas n°1: le résultat est entier (la division tombe juste)
- on vérifie que les valeurs sont rangées par ordre croissant
- Q1 est la $n^{ème}$ valeur où n = N:4
- Q3 est le n' ^{ème} valeur où l'**entier** n' = ¾ de N = $\frac{3}{4} \times N = 3 \times N : 4$

Exemple
Prenons les valeurs rangées dans l'ordre croissant :
*1-3-3-3-5-5-**6**-7-7-8-8-8-9-9-10-10-10-10-11-11-12-**13**-13-13-14-15-16-19*
Il y a N = 28 valeurs, qui est divisible par 4 car 28:4=7 qui est entier
n=N:4 = 7 donc Q1 = la 7ème valeur de la série rangée dans l'ordre croissant= 6
et n' = 3N:4 = 21 donc Q3 = la 21ème valeur de la série rangée dans l'ordre croissant= 13

cas n°2: le résultat n'est pas entier
- on vérifie que les valeurs sont rangées par ordre croissant
- on arrondit l**e décimal** N:4 à l'**entier supérieur** : l'entier n ; Q1 est la $n^{ème}$ valeur
- on arrondit l**e décimal** ¾ de N = $\frac{3}{4} \times N$ = 3N:4 à l'**entier supérieur** : l'entier n' ; Q3 est la n' ^{ème} valeur

Exemple
Prenons les valeurs rangées dans l'ordre croissant :
*3-5-5-6-7-**8**-8-9-9-10-10-10-10-11-11-12-13-**13**-13-14-15-16-19*
Il y a N = 23 valeurs;
N:4 = 5,75 donc Q1 est la 6ème valeur de la série rangée dans l'ordre croissant donc Q1= 8,
3N:4 = 17,25 donc Q3 est la 18 ème valeur de la série rangée dans l'ordre croissant donc Q3= 13

Comment interpréter des quartiles donnés?
si on connait les quartiles Q1 et Q3 d'une série, que peut-on en déduire?
Au moins **un quart (25%) des valeurs sont inférieures** ou égales **à Q1.**
Au moins **trois quarts (75%) des valeurs sont inférieures ou** égales **à Q3**.
Environ la moitié des valeurs se trouvent dans l'intervalle interquartile **[Q1 ; Q3]**.

Exemple: dans une classe, les notes présentent un premier quartile Q1 égal à 10 et un troisième quartile égal à 14. On peut dire que:
*au moins **un quart des élèves a** une note **inférieure** ou égale **à 10***
*au moins **un quart des élèves a** pour note 10 ou **moins de 10***
*En pratique: **environ un quart des élèves a moins de 10, (et environ trois quarts des élèves ont plus)***

*au moins **trois quarts des élèves a** une note **inférieure** ou égale **à 14***
*au moins **trois quarts des élèves a** pour note 14 ou **plus de 14***
*En pratique: **environ trois quarts des élèves a moins de 14, (et environ un quart des élèves ont plus)***

L'intervalle interquartile est l'intervalle]10 ; 14[.
Environ la moitié des élèves a une note entre 10 et 14

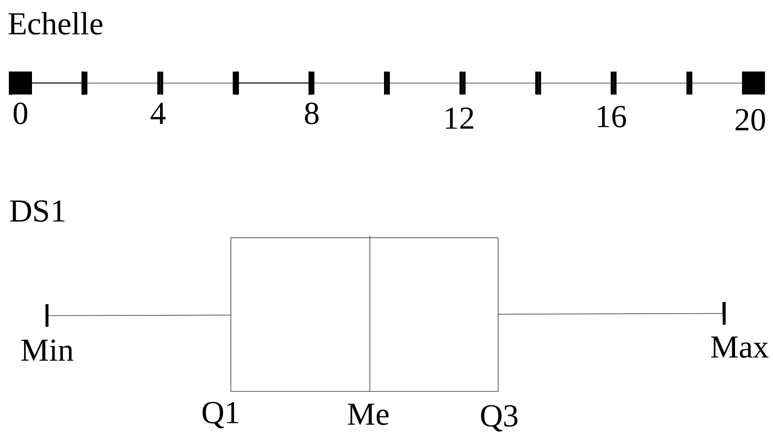
Diagramme en boites

La médiane comme paramètre de position et l'intervalle interquartile comme paramètre de dispersion fournissent une bonne description d'une série statistique.

On utilise ces deux données pour construire un diagramme en boîte de la série

Soit une série de valeurs qui se résume en:
- le minimum Min = 1
- le 1er quartile Q1 = 6
- la médiane Me = 9,5
- le 3ème quartile Q3 = 13
- le maximum Max = 19

Ces 5 données permettent de construire un diagramme en boites :



Déciles

> Les déciles permettent de séparer une série statistique en dix groupes de même effectif (à une unité près).
> Un dixième des valeurs sont inférieures au premier décile D1.
> Un dixième des valeurs sont supérieures au neuvième décile D9.

<u>Comment déterminer les déciles D1 et D9 d'une série de *N* valeurs ?</u>
On calcule la quantité $\frac{1}{10}$ de N = $\frac{1}{10} \times N$ =N:10
Deux cas sont possibles: soit le résultat est entier (la division tombe juste), soit non

<u>cas n°1: le résultat est entier (la division tombe juste)</u>
- on vérifie que les valeurs sont rangées par ordre croissant
- D1 est la n$^{\text{ème}}$ valeur où n = N:10
- D9 est le n' $^{\text{ème}}$ valeur où l'**entier** n' = $\frac{9}{10}$ de N = $\frac{9}{10} \times N = 9 \times N:10$

Exemple
Prenons les valeurs rangées dans l'ordre croissant :
*1-3-**3**-3-5-5-6-7-7-8-8-8-9-9-10-10-10-10-11-11-12-12-13-13-13-13-**14**-15-16-19*
Il y a N = 30 valeurs, qui est divisible par 10 car 30:10=3 qui est entier
n=N:10 = 3 donc D1 est la 3ème valeur de la série rangée dans l'ordre croissant donc D1 = 3=
et n' = 9N:10 = 27 donc D9 est la 27ème valeur de la série rangée dans l'ordre croissant donc D9= 14

<u>cas n°2: le résultat n'est pas entier</u>
- on vérifie que les valeurs sont rangées par ordre croissant
- on arrondit l**e décimal** N:10 à l'**entier supérieur** : l'entier n ; D1 est la n$^{\text{ème}}$ valeur

- on arrondit l**e décimal** $\frac{9}{10}$ de N = $\frac{9}{10} \times N = 9 \times N : 10$ à l'**entier supérieur** : l'entier n' ; D9 est la n' ème valeur

Exemple
Prenons les valeurs rangées dans l'ordre croissant :
*3-5-**5**-6-7-8-8-9-9-10-10-10-10-11-11-12-13-13-13-14-**15**-16-19*
Il y a N = 23 valeurs;
N:10 = 2,3 donc D1 est la 3ème valeur de la série rangée dans l'ordre croissant donc D1= 5
9N:10 = 20,7 donc D9 est la 21 ème valeur de la série rangée dans l'ordre croissant donc D9= 15

Comment interpréter des déciles donnés?
si on connait les déciles D1 et D9 d'une série, que peut-on en déduire?
Au moins **un dixième (10%) des valeurs sont inférieures** ou égales **à D1.**
En pratique: environ 10% des valeurs sont inférieurs à D1

Au moins neuf dixièmes **(90%) des valeurs sont inférieures ou** égales **à Q3**.
En pratique: environ 10% des valeurs sont supérieurs à D9

Diagramme en boites

On remplace parfois les extrémités des pattes du diagramme en boîte par D1 et D9, comme dans l'EXERCICE 1 sur la mesure des vents
On rajoute alors parfois un petit rond pour le min et un petit rond pour le max, comme dans l'EXERCICE 2 sur la pluviométrie

Correction EXERCICE 2 *sur la pluviométrie :*

Partie A

1. m = 859,2 mm/m²

2. 38:2=19 donc la médiane est la demi-somme des 19ème et 20ème valeurs: Me =(1099,8+1101,0)/2 = 1100,4

38:4 = 9,5 donc Q1 = la 10ème valeur = 1029,7

*38:4*3 = 28,5 donc Q3 = la 29 ème valeur = 1233,3*

3. les extrémités des pattes sont D1 et D9.

Or 38:10= 3,8 donc D1 = la 4ème valeur = 972,5

*et 38:10*9 = 34,2 donc D9 = la 35ème valeur = 1313,4*

les « petits ronds » sont le min=782 et le max=1603,6

PartieB

1. L'écart interquartile le plus faible est à Dinard avec Q3-Q1 = 157,2

2. Lorient satisfait la condition puisque la médiane est 919,2, ce qui signifie que au moins la moitié des années ont eu une pluviosité supérieure ou égale à 919,2.

De même pour Montélimar et Brest

3. au moins un quart des années ont connu une pluviométrie inférieure ou égale à 447 mm/m²

4. a. Les deux villes ayant la pluviométrie la plus irrégulière sont Montélimar et Nice, puisque leurs boîtes sont les plus étendues. La région de ces deux villes est la Provence.

4. b. on peut comparer les étendues (écart entre le min et e max) ou les écarts interquartiles ou encore les écarts-types. Ce qui rejoint le résultat précédent.