

Chapitre 5. Statistiques

Yann Barsamian

École Européenne de Bruxelles 1

Année scolaire 2023–2024



- Vocabulaire de base
- Représentations de séries statistiques
- Caractéristiques de séries statistiques

<http://www.barsamian.am/2023-2024/S4P6/TG9.pdf>

- Quand une personne vous montre des statistiques : il peut cacher une partie de la vérité, ou essayer de vous induire en erreur.
- Même quand on utilise des données issues de sources dont on ne peut pas contester l'honnêteté, il y a souvent un travail à faire avant de pouvoir interpréter ces données.

On va analyser des données brutes, qui proviennent d'études sur un sujet. Les données peuvent être quantitatives (ce sont des nombres, par ex. votre meilleur temps au 100m en course à pied) ou qualitatives (ce ne sont pas des nombres, par ex. votre matière préférée l'an dernier).

Pour avoir une vue d'ensemble de la série statistique (l'ensemble des données), on peut calculer les effectifs de chaque valeur : c'est le nombre de fois que la valeur apparaît dans la série. L'effectif total est le nombre total de valeurs.

Par exemple, une étude sur 30 élèves concernant le temps de travail journalier (en minutes) à la maison donne les résultats suivants :

Temps	5	10	15	20	30	40	50	70
Effectif	1	4	4	3	7	7	3	1

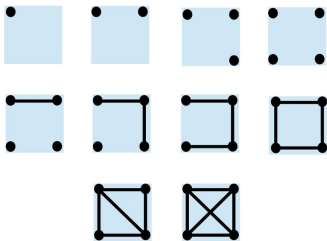
Comment compter des effectifs ?

Méthode 1) ne pas regrouper.

Méthode 2) regrouper par paquets de 5.

a)¹ : | ⊐ ⊓ □ ⊔ ; b)² : | || ||| |||| ~~|||||~~

Méthode 3) regrouper par paquets de 10^3 :



¹En France, en Espagne

²Dans le reste de l'Europe

³En foresterie

Afin de comparer des séries, on utilise les fréquences :

$$\text{fréquence d'une valeur} = \frac{\text{effectif de cette valeur}}{\text{effectif total}}$$

(c'est un nombre dans $[0; 1]$, souvent exprimé en pourcentage).

Les fréquences cumulées (d'abord ordonner les valeurs) : la première fréquence cumulée est la fréquence de la valeur, et la fréquence cumulée d'une autre valeur est égale à la fréquence cumulée de la valeur précédente, plus la fréquence de la valeur.

Temps	5	10	15	20	30	40	50	70
Eff.	1	4	4	3	7	7	3	1
Fr.	1/30	4/30	4/30	3/30	7/30	7/30	3/30	1/30
Fr. cum.	1/30	5/30	9/30	12/30	19/30	26/30	29/30	30/30

L'effectif de la valeur "15" est 4 (pour un effectif total de 30). Sa fréquence est donc 4/30. Sa fréquence cumulée est donc 5/30 (fréquence cumulée précédente) + 4/30 (sa fréquence) = 9/30.

La fréquence cumulée de la dernière valeur est toujours 1.

Vidéo d'introduction: <https://www.lumni.fr/video/representation-des-donnees-tableaux-histogrammes-diagrammes-circulaires-26-mai>

Diagramme en bâton : pour chaque valeur du critère, on dessine un bâton dont la hauteur est l'effectif de cette valeur.

Diagramme circulaire (“camembert”) : on dessine un disque, et chaque valeur du critère est représentée par un secteur angulaire de ce disque, d'angle $360^\circ \times$ sa fréquence⁴.

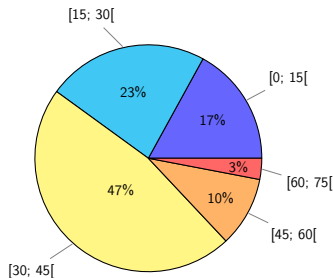
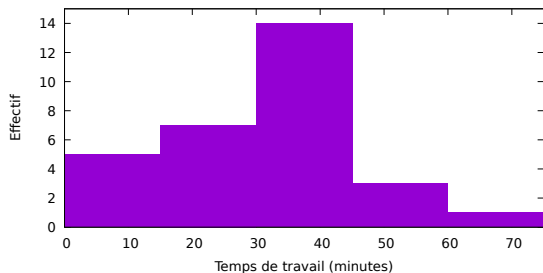
⁴360° correspond à un tour complet, donc à la totalité du disque.

Quand il y a trop de valeurs différentes...

Si le critère est quantitatif, on peut regrouper les valeurs par intervalles. On peut alors tracer un histogramme⁵.

Exemple du 23 p.163, avec regroupement par quarts d'heure:

Temps	[0; 15[[15; 30[[30; 45[[45; 60[[60; 75[
Effectif	5	7	14	3	1
Fréquence	17%	23%	47%	10%	3%
Angle	60°	84°	168°	36°	12°



⁵Avec des intervalles de même amplitude : comme un diagramme en bâtons.

Quand il y a trop de valeurs différentes. . .

Pour éviter de perdre des valeurs, on peut également construire un diagramme à tiges et feuilles.

Exemple :



Quand il y a trop de valeurs différentes. . .

Pour éviter de perdre des valeurs, on peut également construire un diagramme à tiges et feuilles.

On choisit une unité pour les tiges (les heures dans l'exemple précédent) et on regroupe les valeurs de la série dans la tige qui leur correspond. Le diagramme contient une ligne par tige, et deux colonnes : la colonne des tiges, la colonne des feuilles. Dans la colonne des feuilles, on note la valeur moins la valeur de la tige correspondante. Par ex. on note 5h10 dans la ligne de la tige 5, en notant seulement 10 dans la colonne des feuilles (on n'écrit pas 5h10 en entier).

Vidéo d'introduction (26 minutes):

<https://www.lumni.fr/video/statistiques-moyenne-et-mediane>

1) La moyenne⁶ :

C'est la somme des valeurs divisée par le nombre de valeurs. Lorsque l'on a des coefficients pour les valeurs (ou des effectifs), il faut pondérer la moyenne : on multiplie chaque valeur par l'effectif, et on divise par l'effectif total.

2) La médiane⁷ :

C'est un nombre qui sépare la série en deux parties de même effectif. Il y a au moins 50% des valeurs qui sont inférieures ou égales à la médiane, et au moins 50% des valeurs qui sont supérieures ou égales à la médiane.

⁶https://www.youtube.com/watch?v=88_16UbkdZM

⁷<https://www.youtube.com/watch?v=g1OCTw--VYQ>

Pour calculer la médiane d'une série ordonnée de n nombres :

$$u_1 \leq u_2 \leq \dots \leq u_n$$

- si n est impair, c'est la valeur centrale ; son rang est $\frac{n+1}{2}$

Ex.: 2, 5, 7, 8, 9 : la médiane est la valeur de rang $\frac{5+1}{2} = 3$:
c'est 7

- si n est pair, il y a deux valeurs centrales ; leur rang est $\frac{n}{2}$ et $\frac{n}{2} + 1$... on prend la valeur au milieu

Ex.: 2, 5, 7, 8, 9, 12 : la médiane est la demi-somme des valeurs
de rang $\frac{6}{2} = 3$ et $\frac{6}{2} + 1 = 4$: c'est $\frac{7+8}{2} = 7,5$

III/ Caractéristiques d'une série statistique

3) Les quartiles⁸ :

Le 1^{er} quartile Q1 (3^{ème} quartile Q3) : la plus petite valeur de la série supérieure ou égale à au moins 25% (75%) des valeurs.

Le rang de Q1 (Q3) est le premier entier supérieur ou égal à $\frac{n}{4}$ ($\frac{3n}{4}$), c'est-à-dire 25% (75%) de n .

Temps	5	10	15	20	30	40	50	70
Effectif	1	4	4	3	7	7	3	1
Effectif cum.	1	5	9	12	19	26	29	30

- $\frac{30}{4} = 7,5$ donc Q1 est la 8e valeur. C'est 15.
- $\frac{3 \times 30}{4} = 22,5$ donc Q3 est la 23e valeur. C'est 40.

⁸<https://www.youtube.com/watch?v=Yjh-9nMVmEw>,
<https://www.youtube.com/watch?v=IjsDKOODwIw>

4) La boîte à moustaches⁹ :

La boîte à moustache est un diagramme représentant les différentes caractéristiques que nous avons déjà calculées :

- il faut une échelle (un axe gradué horizontal, qui couvre les valeurs minimale et maximale de la série)
- ensuite, on trace une boîte rectangulaire qui va de Q_1 à Q_3 , avec une ligne verticale à la médiane
- enfin, on trace deux traits horizontaux (deux moustaches) : le premier de la gauche de la boîte jusqu'à la valeur minimale, l'autre de la droite de la boîte jusqu'à la valeur maximale (qu'on finit par des traits verticaux)

⁹<https://www.youtube.com/watch?v=la7c0Yf8VyM>

III/ Caractéristiques d'une série statistique

Exemple sur notre série à 30 valeurs :

- valeur minimale : 5
- valeur maximale : 70
- Q1 : 15
- Q3 : 40
- médiane : il y a 30 valeurs, c'est la demi-somme de la 15e et de la 16e valeur (qui valent chacune 30) : la médiane vaut 30.

