Biostatistique

khamesyacine@gmail.com

Biostatistique

1. Statistiques descriptives

2. Lois de probabilité

Biostatistique

1. Statistiques descriptives

2. Lois de probabilité

3. Tests d’hypothèse: Comparaison de moyennes

4. Tests d’hypothèse :Comparaison de pourcentages

Biostatistique

5. Corrélation & régression

6. Anova

7. Test non paramétrique

8. Test de Chi 2

9. Risque relatif & odds ratio

Statistiques descriptives

Les données	sont des observations et des variables
Les statistiques	organiser, résumer, analyser, interpréter, présenter et de tirer des conclusions à partir de ces données.
Une population	est la collection complète de tous les éléments (notes, personnes, mesures,…..) à étudier.
Un échantillon	est une sous-collecte d’individu sélectionnés dans une partie de la population.

Type de données

Un paramètre: est une mesure qui décrit une caractéristique d'une population

Une statistique: est une mesure qui décrit une caractéristique d'un échantillon

Type de données

Un paramètre

On remplit un lac artificiel avec 500 truites pour un total de 950 kg. Si on divise le poids total par le nombre de truites, on obtient une moyenne de 1,9 kg.

Si on considère la collection des 500 truites représente la population du lac, alors 1,9 kg est un paramètre, non une statistique.

Une statistique

A partir d'un échantillon de 877 directeurs enquêtés, il apparaît que 45 % d'entre eux n'embaucheraient pas quelqu'un qui fait une faute typographique dans sa demande d'embauche.

Ce chiffre de 45 % est une statistique parce qu'il est basé sur un échantillon, et non sur la population totale de tous les directeurs.

Type de données

Variable quelque chose qui peut être mesuré

Type de données

Quantitatives se composent de nombres représentant des comptages ou des mesures.

Qualitatives (ou catégorielles) peuvent être séparées en différentes catégories qui se distinguent par certaines caractéristiques non numériques.

Type de données

Quantitatives discrètes (discontinues)

Lorsque le nombre possible de valeurs est soit fini soit dénombrable (c'est-à-dire que le nombre de valeurs est 0 ou 1 ou 2 et ainsi de suite).

Quantitatives continue

On parle de données continues lorsqu'on a un nombre infini de valeurs possibles qui correspondent à une échelle continue de valeurs

Type de données

Qualitatives nominales

Le niveau nominal de mesure est caractérisé par des données qui consistent en noms, labels ou catégories seulement.

Exemple:

- **Oui/non/ne sait pas : comme les réponses à une enquête.**
- **Couleurs : la couleur des pois (vert, jaune) utilisés dans une expérience de génétique.**

Type de données

Qualitatives ordinales

Les catégories sont ordonnées mais la différence n’a pas de sens

Exemple:
Les ours selon leur agressivité:
5 non agressifs
20 un peu agressifs
40 fortement agressifs

Dans les exercices 1 et 2, Déterminez si la valeur donnée est une statistique ou un paramètre.

Dans une étude du comportement des oies près d'un aéroport, un échantillon aléatoire d'oies canadiennes incluait 12 mâles.

Dans une étude on attache des altimètres individuels à des oiseaux (frégates) et l'altitude moyenne est de 226 m.

Dans les exercices 3 et 4, déterminez si les données sont discrètes ou continues

Dans une étude sur des oiseaux de l'île Buldir en Alaska, 312 kittiwakes à pattes rouges adultes ont été bagués.

Dans une enquête sur 1 059 adultes, on trouve que 39 % d'entre eux ont des armes à la maison.

Kittiwakes (références)

Plans d'expériences

Une bonne utilisation des statistiques requiert typiquement plus de bon sens que d'expertise mathématique

Comme nous disposons de calculatrices et d'ordinateurs, les applications modernes des statistiques ne nous demandent plus de maîtriser des algorithmes complexes pour les calculs mathématiques.

À la place, nous pouvons nous focaliser sur l'interprétation des données et des résultats.

Plans d'expériences

Les méthodes statistiques sont dépendantes des données. On obtient classiquement des données à partir de deux sources distinctes: les études observationnelles et les études expérimentales

Dans une étude observationnelle

On observe et on mesure des caractéristiques spécifiques mais on n'essaie pas de modifier les sujets de l'étude.

Dans une étude expérimentale

On applique un certain traitement et on passe ensuite à l'observation de son effet sur les sujets.

- Un sondage est un bon exemple d'étude observationnelle.

- Un bon exemple d'étude expérimentale est celui d'un essai clinique où une expérience est planifiée et organisée avec un groupe traitement

Etude observationnelle

- Dans une étude transversale: les données sont observées, mesurées et collectées à un instant donné.

- Dans une étude rétrospective (ou cas témoins): les données sont collectées dans le passé (en relisant des examens médicaux, des interviews...).

- Dans une étude prospective (ou longitudinale ou de cohorte), les données sont collectées dans des groupes futurs (nommés cohortes) partageant des facteurs communs.

L’objectif principal :

Mesurer et décrire les différentes caractéristiques d’un jeu de données

- Tendance centrale
- Dispersion
- Mesures de positionnement relatifs
- Valeurs extrêmes

Mesure de tendance centrale

Mesurer et décrire les différentes caractéristiques d'un jeu de données

- Tendance centrale
- Dispersion
- Mesures de positionnement relatifs
- Valeurs extrêmes

est une valeur au centre ou au milieu du jeu de données

Mesure de tendance centrale

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Tendance centrale

Est une valeur au centre ou au milieu du jeu de données

Moyenne

La moyenne (arithmétique) d'un ensemble de valeurs est la mesure de tendance centrale obtenue en additionnant les valeurs et en divisant par le nombre totale de valeurs

Mesure de tendance centrale

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Tendance centrale

Est une valeur au centre ou au milieu du jeu de données

Moyenne

La moyenne (arithmétique) d'un ensemble de valeurs est la mesure de tendance centrale obtenue en additionnant les valeurs et en divisant par le nombre totale de valeurs

Formule :

$$Moyenne = \frac{\sum x_i}{n}$$

Mesure de tendance centrale

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Tendance centrale

Est une valeur au centre ou au milieu du jeu de données

Exemple

Mesure du taux de plomb dans l'aire:

5,40 1,10 0,42 0,73 0,48 1,10

$$\bar{x} = \frac{\sum x_i}{n} = \frac{5{,}40 + 1{,}10 + 0{,}42 + 0{,}48 + 1{,}10}{5} = 1{,}538$$

Mesure de tendance centrale

Tendance centrale

Un défaut de la moyenne est qu'elle est sensible à chacune des valeurs, donc une seule valeur exceptionnelle peut affecter la moyenne de façon importante

Glycémie normale : Entre 0.7 et 1 g/L de sang

Diabète : Supérieur à 1.26 g/L de sang

Hypoglycémie : Inférieur à 0.7 g/L de sang

N°	Glycémie normale (g/L)
1	0,95
2	0,80
40	0,75
…	…
100	0 99
1000	0,96
10000	0,99

Mesure de tendance centrale

Tendance centrale

Un défaut de la moyenne est qu'elle est sensible à chacune des valeurs, donc une seule valeur exceptionnelle peut affecter la moyenne de façon importante

Glycémie normale : Entre 0.7 et 1 g/L de sang

Diabète : Supérieur à 1.26 g/L de sang

Hypoglycémie : Inférieur à 0.7 g/L de sang

N°	Glycémie normale (g/L)
1	0,95
2	0,80
40	0,75
…	…
100	0 99
1000	0,96
10000	0,99

$$\bar{x} = \frac{\sum x_i}{n} = \frac{0{,}95+0{,}80+0{,}75+\cdots+099+0{,}96+0{,}99}{10000} = 0{,}0099$$

Mesure de tendance centrale

La médiane corrigé ce défaut (robuste), la médiane est vue comme une vraie valeur du milieu car n'est pas affectée par les valeurs exceptionnelles

Médiane

La médiane d'un ensemble de données est la mesure de tendance centrale qui est la valeur du milieu quand les données de départ sont triées par ordre croissant.

Mesure de tendance centrale

La médiane corrigé ce défaut (robuste), la médiane est vue comme une vraie valeur du milieu car n'est pas affectée par les valeurs exceptionnelles

Exemple

Mesure du taux de plomb dans l'aire:

5,40	1,10	0,42	0,73	0,48	1,10
0,42	0,48	0,73	1,10	1,10	5,40

$$Médiane = \frac{0{,}73 + 1{,}10}{2} = 0{,}915$$

Mesure de tendance centrale

La médiane 0,915 $\mu g/m^3$ est très différente de la moyenne 1,538 $\mu g/m^3$. Cette différence est due à l'effet de la valeur 5,40 sur la moyenne.

Si cette valeur est ramenée à 1,20 par exemple, la moyenne descendrait à 0,83 alors que la médiane ne changerait pas.

0,42 0,48 0,73 1,10 1,10 5,40

$$Moyenne = \frac{\sum x_i}{n}$$

$$= \frac{5{,}40 + 1{,}10 + 0{,}42 + 0{,}48 + 1{,}10}{5}$$

$$= 1{,}538$$

$$Médiane = \frac{0{,}73 + 1{,}10}{2} = 0{,}915$$

Mesure de tendance centrale

Mode

Le mode d’un jeu de données est la valeur qui est présente le plus grand nombre de fois

Mesure de tendance centrale

Mode

Le mode d'un jeu de données est la valeur qui est présente le plus grand nombre de fois

Quand deux valeurs apparaissent avec la même plus grande fréquence (nombre de fois), chacune est une mode et le jeu de données est bimodal

Mesure de tendance centrale

Mode

Le mode d'un jeu de données est la valeur qui est présente le plus grand nombre de fois

Quand deux valeurs apparaissent avec la même plus grande fréquence (nombre de fois), chacune est une mode et le jeu de données est bimodal

Quand plus de deux valeurs apparaissent avec la même plus grande fréquence (nombre de fois), chacune est une mode et le jeu de données est multimodal

Mesure de tendance centrale

Mode

Le mode d'un jeu de données est la valeur qui est présente le plus grand nombre de fois

Quand deux valeurs apparaissent avec la même plus grande fréquence (nombre de fois), chacune est une mode et le jeu de données est bimodal

Quand plus de deux valeurs apparaissent avec la même plus grande fréquence (nombre de fois), chacune est une mode et le jeu de données est multimodal

Quand aucune valeur n'est répétée, on dit qu'il n'y a pas de mode

Mesure de tendance centrale

Midrange (centre de classe) est la mesure de tendance centrale qui est la valeur à mi chemin entre la plus grande est la plus petite valeur du jeu de donnée ou une classe.

$$Midrange = \frac{Min\ - Max}{2}$$

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Tendance centrale	est une valeur au centre ou au milieu du jeu de données
Dispersion	mesurer la variation
Mesures de positionnement relatifs	
Valeurs extrêmes	

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Dispersion — mesurer la variation

- **Etendue** — est la différence entre la valeur maximale et la valeur minimale
- **Variance**
- **Ecart-type**

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Dispersion — mesurer la variation

- **Etendue** — est la différence entre la valeur maximale et la valeur minimale
- **Variance** — est une mesure de dispersion des valeurs autour de la moyenne
- **Ecart-type**

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Dispersion — mesurer la variation

- **Etendue** — est la différence entre la valeur maximale et la valeur minimale
- **Variance** — est une mesure de dispersion des valeurs autour de la moyenne
- **Ecart-type** — est une mesure de dispersion des valeurs autour de la moyenne

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Dispersion

mesurer la variation

Etendue

$$Etendue = (Valeur\ maximale) - (Valeur\ minimale)$$

Variance

$$\sigma^2 = \frac{\sum n_i (x_i - \bar{x})^2}{(n-1)} = \frac{(n_i \sum x_i^2) - (n \times \bar{x}^2)}{(n-1)}$$

Ecart-type

$$\sigma = \sqrt{\frac{\sum n_i (x_i - \bar{x})^2}{(n-1)}}$$

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Dispersion

mesurer la variation

Variance

$$\sigma^2 = \frac{\sum n_i (x_i - \bar{x})^2}{(n-1)} = \frac{(n_i \sum x_i^2) - (n \times \bar{x}^2)}{(n-1)}$$

Ecart-type

$$\sigma = \sqrt{\frac{\sum n_i (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{(n_i \sum x_i^2) - (n \times \bar{x}^2)}{(n-1)}}$$

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Dispersion

mesurer la variation

Ecart-type

La valeur de l'écart-type est en générale positive. Elle est nulle uniquement si toutes les données ont la même valeurs

Un écart-type grand indique une plus grande variation

La valeur de l'écart-type peut augmenter de façon importante si on inclut une ou plusieurs valeurs extrêmes

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Dispersion

mesurer la variation

Ecart-type

$$\sigma = \sqrt{\frac{\sum n_i (x_i - \bar{x})^2}{n}}$$

Ecart-type d'une population

Dispersion mesurer la variation

- **Etendue** est la différence entre la valeur maximale et la valeur minimale
- **Variance** est une mesure de dispersion des valeurs autour de la moyenne
- **Ecart-type** est une mesure de dispersion des valeurs autour de la moyenne
- **Coefficient de variation** comparer la dispersion dans des population différentes

Dispersion mesurer la variation

Coefficient de variation comparer la dispersion dans des population différentes

$$CV = \frac{\sigma}{\bar{x}} \times 100$$

Dispersion — mesurer la variation

- **Variance** — est une mesure de dispersion des valeurs autour de la moyenne
- **Ecart-type** — est une mesure de dispersion des valeurs autour de la moyenne
- **Coefficient de variation** — comparer la dispersion dans des population différentes

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Tendance centrale	est une valeur au centre ou au milieu du jeu de données
Dispersion	mesurer la variation
Mesures de positionnement relatifs	On introduit des mesures qui peuvent être utilisées pour comparer des valeurs issues de jeux de données différents
Valeurs extrêmes	

Mesurer et décrire les différentes caractéristiques d’un jeu de données

Mesures de positionnement relatifs

Score - Z

Quantiles et percentiles

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Mesures de positionnement relatifs

Score - Z

Score – Z (normalisé) est obtenu en convertissant une valeur sur une échelle normalisée

$$z = \frac{x - \bar{x}}{\sigma}$$

Exemple: comparaison de taille

Les hommes font en moyenne 1,75 m avec un écart type de 7,11 cm.

les femmes font en moyenne 1,61 m avec un écart type de 6,35 cm

(d'après des données de l'enquête nationale américaine de santé).

L'ancienne star du basket Michael Jordan fait 1,98 m et la basketteuse Rebecca Lobo fait 1,93 m.

Jordan est évidemment plus grand de 5 cm, mais qui est relativement plus grand ?.

Source

Source

Est-ce que l'écart entre la taille de Jordan et celle des hommes dépasse l'écart entre la taille de Lobo et celle des femmes ?

Source

Source

Pour comparer les tailles de Michael Jordan et de Rebecca Lobo relativement aux tailles des hommes et des femmes, on doit normaliser ces tailles en les convertissant en scores z

Jordan : $$z = \frac{x - \bar{x}}{\sigma} = \frac{1{,}98 - 1{,}75}{0{,}711} = 3{,}23$$

Lobo: $$z = \frac{x - \bar{x}}{\sigma} = \frac{1{,}93 - 1{,}61}{0{,}635} = 5{,}04$$

La taille de Michael Jordan est à 3,23 écart types au-dessus de la moyenne, mais Rebecca Lobo est à 5,04 écarts types. La taille de Rebecca Lobo parmi les femmes est relativement plus importante que celle de Michael Jordan parmi les hommes.

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Mesures de positionnement relatifs

Score - Z

Score – Z (normalisé) est obtenu en convertissant une valeur sur une échelle normalisé

Quantiles et percentiles

On introduit des mesures qui peuvent être utilisées
pour comparer des valeurs issues de jeux de données différents

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Mesures de positionnement relatifs

Quantiles

Il y a trois quartiles séparent les données en quatre parties égales

- Q1
- Q2
- Q3

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Mesures de positionnement relatifs

Quantiles

Il y a trois quartiles séparent les données en quatre parties égales

- **Q1** — Sépare les premiers 25% des données triées des autres 75%
- **Q2**
- **Q3**

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Mesures de positionnement relatifs

Quantiles

Il y a trois quartiles séparent les données en quatre parties égales

Q1

Sépare les premiers 25% des données triées des autres 75%

Q2

C'est la même chose que la médiane, sépare les premiers 50% des données triées des autres 50%

Q3

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Mesures de positionnement relatifs

Quantiles

Il y a trois quartiles séparent les données en quatre parties égales

- **Q1** : Sépare les premiers 25% des données triées des autres 75%
- **Q2** : C'est la même chose que la médiane, sépare les premiers 50% des données triées des autres 50%
- **Q3** : Sépare les premiers 75% des données triées des autres 25%

Mesure de tendance centrale

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Mesures de positionnement relatifs

Percentiles

Il y a 99 percentiles qui partitionnent les données en 100 groupes

$$Percentile\ de\ valeur\ x = \frac{Nombre\ de\ valeurs\ inférieurs\ à\ x}{Nombre\ total\ de\ valeurs} \times 100$$

Exemple: Niveau de cotinine des fumeurs

0	1	1	3	17
12	18	103	112	121
173	173	198	208	210
253	265	266	277	284

Trouver le percentile correspondant au niveau de cotinine 112

$$Percentile\ de\ valeur\ x = \frac{8}{20} \times 100 = 40$$

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Tendance centrale	est une valeur au centre ou au milieu du jeu de données
Dispersion	mesurer la variation
Mesures de positionnement relatifs	On introduit des mesures qui peuvent être utilisées pour comparer des valeurs issues de jeux de données différents
Valeurs extrêmes	Une valeur extrême est une valeur située très loin de toutes les autres valeurs

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Valeurs extrêmes

Une valeur extrême est une valeur située très loin de toutes les autres valeurs

Quand on explore un jeu de données, on doit considérer les valeurs extrêmes parce qu'elles peuvent révéler des informations importantes et affecter fortement la moyenne et l'écart-type

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Valeurs extrêmes

Une valeur extrême est une valeur située très loin de toutes les autres valeurs

Une valeur extrême peut avoir un effet important sur la moyenne

Une valeur extrême peut avoir un effet important sur l'écart-type

Une valeur extrême peut avoir un effet un effet important sur un histogramme ou un nuage de point

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Valeurs extrêmes

Une valeur extrême est une valeur située très loin de toutes les autres valeurs

Une valeur extrême peut avoir un effet important sur la moyenne

Une valeur extrême peut avoir un effet important sur l'écart-type

Une valeur extrême peut avoir un effet un effet important sur un histogramme ou un nuage de point

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Valeurs extrêmes

Une valeur extrême est une valeur située très loin de toutes les autres valeurs

Une valeur extrême peut avoir un effet important sur la moyenne

Une valeur extrême peut avoir un effet important sur l'écart-type

Une valeur extrême peut avoir un effet un effet important sur un histogramme ou un nuage de point

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Valeurs extrêmes

Procédure pour trouver les valeur extrêmes est d'examiner la liste des données. En particulier le minimum et le maximum

Mesurer et décrire les différentes caractéristiques d'un jeu de données

Valeurs extrêmes

Si on n'est sur que la valeur extrême est une erreur, il faut la corriger ou la supprimer

L'effet du tabagisme passif est-il un *mythe*?

Source 🌐

Est-ce que les non fumeurs sont vraiment affectés par ceux qui fument des cigarettes

Est-ce que les non fumeurs doivent être préoccupés par leur santé parce qu'ils sont en présence de fumeurs ?

La cotinine est un métabolite de la nicotine, ce qui signifie que quand la nicotine est absorbée par le corps, la cotinine est produite De nombreuses lois ont été décrétées pour restreindre le droit de fumer dans les lieux publics.

Table 2-1 Measured Cotinine Levels in Three Groups

Smoker: Subjects reporting tobacco use.

ETS: (Environmental Tobacco Smoke) Subjects are nonsmokers who are exposed to environmental tobacco smoke ("secondhand smoke") at home or work.

NOETS: (No Environmental Tobacco Smoke) Subjects are nonsmokers who are not exposed to environmental tobacco smoke at home or work. That is, the subjects do not smoke and are not exposed to secondhand smoke.

Smoker:	1	0	131	173	265	210	44	277	32	3
	35	112	477	289	227	103	222	149	313	491
	130	234	164	198	17	253	87	121	266	290
	123	167	250	245	48	86	284	1	208	173
ETS:	384	0	69	19	1	0	178	2	13	1
	4	0	543	17	1	0	51	0	197	3
	0	3	1	45	13	3	1	1	1	0
	0	551	2	1	1	1	0	74	1	241
NOETS:	0	0	0	0	0	0	0	0	0	0
	0	9	0	0	0	0	0	0	244	0
	1	0	0	0	90	1	0	309	0	0
	0	0	0	0	0	0	0	0	0	0

Est-ce que ces lois sont justifiées si on se base sur des arguments de santé ou est-ce que ce sont seulement des tracas non nécessaires pour les fumeurs?

Est-ce que les non fumeurs doivent être préoccupés par leur santé parce qu'ils sont en présence de fumeurs ?

fumeurs

Sujets non fumeurs exposés à la fumée de cigarette

Sujets non fumeurs non exposés à la fumée de cigarette

Table 2-1 Measured Cotinine Levels in Three Groups

Smoker: Subjects reporting tobacco use.

ETS: (Environmental Tobacco Smoke) Subjects are nonsmokers who are exposed to environmental tobacco smoke ("secondhand smoke") at home or work.

NOETS: (No Environmental Tobacco Smoke) Subjects are nonsmokers who are not exposed to environmental tobacco smoke at home or work. That is, the subjects do not smoke and are not exposed to secondhand smoke.

Smoker:	1	0	131	173	265	210	44	277	32	3
	35	112	477	289	227	103	222	149	313	491
	130	234	164	198	17	253	87	121	266	290
	123	167	250	245	48	86	284	1	208	173
ETS:	384	0	69	19	1	0	178	2	13	1
	4	0	543	17	1	0	51	0	197	3
	0	3	1	45	13	3	1	1	1	0
	0	551	2	1	1	1	0	74	1	241
NOETS:	0	0	0	0	0	0	0	0	0	0
	0	9	0	0	0	0	0	0	244	0
	1	0	0	0	90	1	0	309	0	0
	0	0	0	0	0	0	0	0	0	0

Jeux de données

Table 2-1 Measured Cotinine Levels in Three Groups

Smoker: Subjects reporting tobacco use.

ETS: (Environmental Tobacco Smoke) Subjects are nonsmokers who are exposed to environmental tobacco smoke ("secondhand smoke") at home or work.

NOETS: (No Environmental Tobacco Smoke) Subjects are nonsmokers who are not exposed to environmental tobacco smoke at home or work. That is, the subjects do not smoke and are not exposed to secondhand smoke.

Smoker:	1	0	131	173	265	210	44	277	32	3
	35	112	477	289	227	103	222	149	313	491
	130	234	164	198	17	253	87	121	266	290
	123	167	250	245	48	86	284	1	208	173
ETS:	384	0	69	19	1	0	178	2	13	1
	4	0	543	17	1	0	51	0	197	3
	0	3	1	45	13	3	1	1	1	0
	0	551	2	1	1	1	0	74	1	241
NOETS:	0	0	0	0	0	0	0	0	0	0
	0	9	0	0	0	0	0	0	244	0
	1	0	0	0	90	1	0	309	0	0
	0	0	0	0	0	0	0	0	0	0

Déterminer le nombre de classes

Table 2-1 Measured Cotinine Levels in Three Groups

Smoker: Subjects reporting tobacco use.

ETS: (Environmental Tobacco Smoke) Subjects are nonsmokers who are exposed to environmental tobacco smoke ("secondhand smoke") at home or work.

NOETS: (No Environmental Tobacco Smoke) Subjects are nonsmokers who are not exposed to environmental tobacco smoke at home or work. That is, the subjects do not smoke and are not exposed to secondhand smoke.

Smoker:	1	0	131	173	265	210	44	277	32	3
	35	112	477	289	227	103	222	149	313	491
	130	234	164	198	17	253	87	121	266	290
	123	167	250	245	48	86	284	1	208	173
ETS:	384	0	69	19	1	0	178	2	13	1
	4	0	543	17	1	0	51	0	197	3
	0	3	1	45	13	3	1	1	1	0
	0	551	2	1	1	1	0	74	1	241
NOETS:	0	0	0	0	0	0	0	0	0	0
	0	9	0	0	0	0	0	0	244	0
	1	0	0	0	90	1	0	309	0	0
	0	0	0	0	0	0	0	0	0	0

Pour déterminer le nombre de classes, on utilise la règle de Sturges qui dit que

$$k \simeq 1 + 3{,}22 \times \log10(n)$$

Déterminer le nombre de classes

Table 2-1 Measured Cotinine Levels in Three Groups

Smoker: Subjects reporting tobacco use.

ETS: (Environmental Tobacco Smoke) Subjects are nonsmokers who are exposed to environmental tobacco smoke ("secondhand smoke") at home or work.

NOETS: (No Environmental Tobacco Smoke) Subjects are nonsmokers who are not exposed to environmental tobacco smoke at home or work. That is, the subjects do not smoke and are not exposed to secondhand smoke.

Smoker:	1	0	131	173	265	210	44	277	32	3
	35	112	477	289	227	103	222	149	313	491
	130	234	164	198	17	253	87	121	266	290
	123	167	250	245	48	86	284	1	208	173

n: nombre d'observation (fumeurs) = 40

Pour déterminer le nombre de classes, on utilise la règle de Sturges qui dit que

$$k \simeq 1 + 3{,}22 \times \log10(n)$$
$$\simeq 6{,}15$$

Le pas

Table 2-1 Measured Cotinine Levels in Three Groups

Smoker: Subjects reporting tobacco use.

ETS: (Environmental Tobacco Smoke) Subjects are nonsmokers who are exposed to environmental tobacco smoke ("secondhand smoke") at home or work.

NOETS: (No Environmental Tobacco Smoke) Subjects are nonsmokers who are not exposed to environmental tobacco smoke at home or work. That is, the subjects do not smoke and are not exposed to secondhand smoke.

Smoker:	1	0	131	173	265	210	44	277	32	3
	35	112	477	289	227	103	222	149	313	491
	130	234	164	198	17	253	87	121	266	290
	123	167	250	245	48	86	284	1	208	173

n: nombre d'observation (fumeurs) = 40

Pour déterminer le pas

$$\Delta = \frac{Max - Min}{nombre\ de\ classe} =$$

$$= \frac{491 - 0}{6} = 81{,}81$$

La distribution de fréquences listes les valeurs des données (classe ou intervalle) et les fréquences correspondantes

Distributions de fréquences des niveaux de cotinine des fumeurs	
Niveaux de cotinine (ng/ml)	**Effectif ou Fréquence (nombre de fumeurs)**
0-99	11
100-199	12
200-299	14
300-399	1
400-499	2

On pose :

Nombre de classe = 5

Pas =

La distribution de fréquences listes les valeurs des données (classe ou intervalle) et les fréquences correspondantes

Distributions de fréquences des niveaux de cotinine des fumeurs	
Niveaux de cotinine (ng/ml)	**Effectif ou Fréquence (nombre de fumeurs)**
0-99	11
100-199	12
200-299	14
300-399	1
400-499	2

$$\text{Fréquence relative} = \frac{Fréquence\ de\ classe}{Somme\ de\ toutes\ les\ fréquences}$$

Niveaux de cotinine (ng/ml)	Fréquence relative
0-99	
100-199	30
200-299	35
300-399	2,5
400-499	5

Fréquence cumulées

Niveaux de cotinine (ng/ml)	Fréquence relative	Fréquence cumulée
0-99	27,5	27,5
100-199	30	27,5 + 30 = 57,5
200-299	35	57,5 + 35 = 92,5
300-399	2,5	92,5 + 2,5 = 95
400-499	5	95 + 5 = 100

Niveaux de cotinine (ng/ml) Classes	Centre de classe	Effectifs (ni)	Effectif cumulé croissant (nic)	
0-99	49.5	11	11	27.5
100-199	149.5	12	23	57.5
200-299	249.5	14	37	92.5
300-399	349.5	1	38	95
400-499	449.5	2	40	100

Exemple

On a mesuré le temps de saignement sanguin en secondes pour un échantillon de 50 personnes atteintes d'hémophilie après 4mn de saignement (valeur considérée comme normale). On a obtenu les résultats suivants:

50	74	69	63	80
53	74	70	63	80
55	74	70	63	81
56	74	70	66	81
57	75	71	66	81
58	75	72	67	82
61	75	72	67	83
61	76	73	68	85
62	77	73	68	87
62	78	73	69	88

Paramètres Statistiques

Exemple

On a mesuré le temps de saignement sanguin en secondes pour un échantillon de 50 personnes atteintes d'hémophilie après 4mn de saignement (valeur considérée comme normale). On a obtenu les résultats suivants:

50	74	69	63	80
53	74	70	63	80
55	74	70	63	81
56	74	70	66	81
57	75	71	66	81
58	75	72	67	82
61	75	72	67	83
61	76	73	68	85
62	77	73	68	87
62	78	73	69	88

a. La moyenne arithmétique

$$\overline{X} = \frac{\sum n_i(x_i)}{N} = 73{,}82$$

A. Déterminer l’échantillon statistique, le caractère étudié et sa nature
B. Regrouper ces résultats en classes de même amplitude, puis dresser un tableau d’effectifs, en précisant les centres de classes, les pourcentages, les effectifs cumulés croissants et décroissants.
C. Quel est le pourcentage des personnes ayant un temps de saignement inférieur à 70 secondes.
D. Représenter la série graphiquement, puis tracer les deux courbes cumulatives.
E. Déterminer graphiquement et par le calcul le mode, la médiane (Quartile 2), quartile 1 et quartile 3. ch

A. Déterminer la population statistique, le caractère étudié et sa nature

L'échantillon étudié est 50 personnes atteintes d'hémophilie
Le caractère étudié : le temps
sa nature : quantitative

B. Regrouper la série en classes d'égales amplitudes

Déterminer le nombre de classe :

Nombre de classe = $1 + (3{,}3 \times \log 50)$ = 6,18

n: nombre de sujets ou d'observation

Déterminer l'amplitude (Δ) :

Max = 88

Min = 50

Nombre de classe égale à 7

L'amplitude $(\Delta) = \frac{Min - Max}{nombre\ de\ classe}$ =6

B. Regrouper ces résultats en classes de même amplitude, puis dresser un tableau d'effectifs, en précisant les centres de classes, les pourcentages, les effectifs cumulés croissants et décroissants.

Classes	Centre de classe	Effectifs (ni)	Effectif cumulé croissant (nic)	
50-56	53	3	3	6
56-62	59	5	8	16
62-68	65	9	17	34
68-74	71	13	30	60
74-80	77	10	40	80
80-86	83	8	48	96
86-92	89	2	50	100

a. La moyenne arithmétique pondérée

$$\overline{X}_{pondérée} = \frac{\sum n_i(Centre\ de\ classe)}{N}$$

a. Le mode

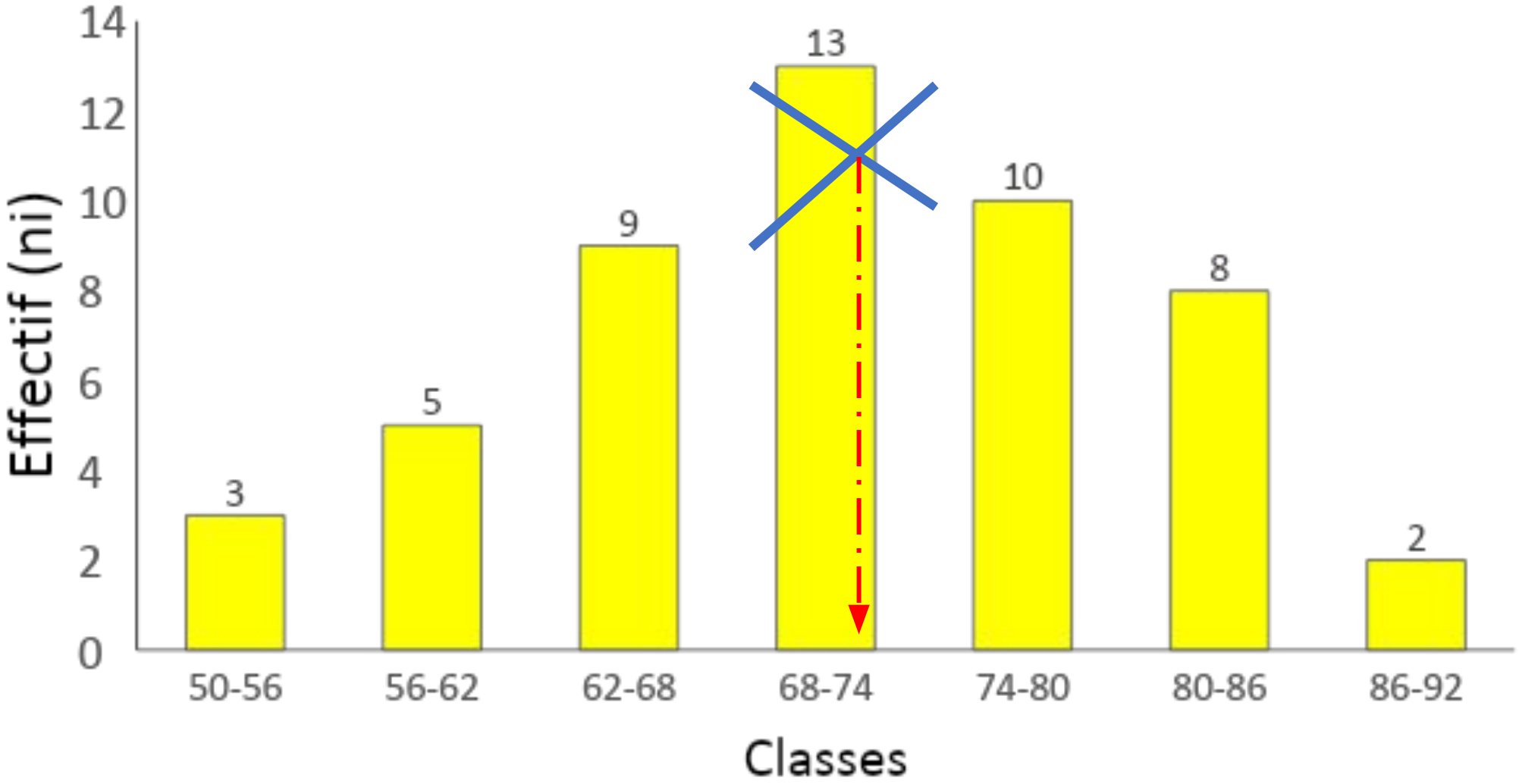
$$M_0 = Borne\ \text{inférieure} + \left((\Delta)\frac{(\Delta_1)}{(\Delta_1+\Delta_2)}\right)$$

b. La médiane

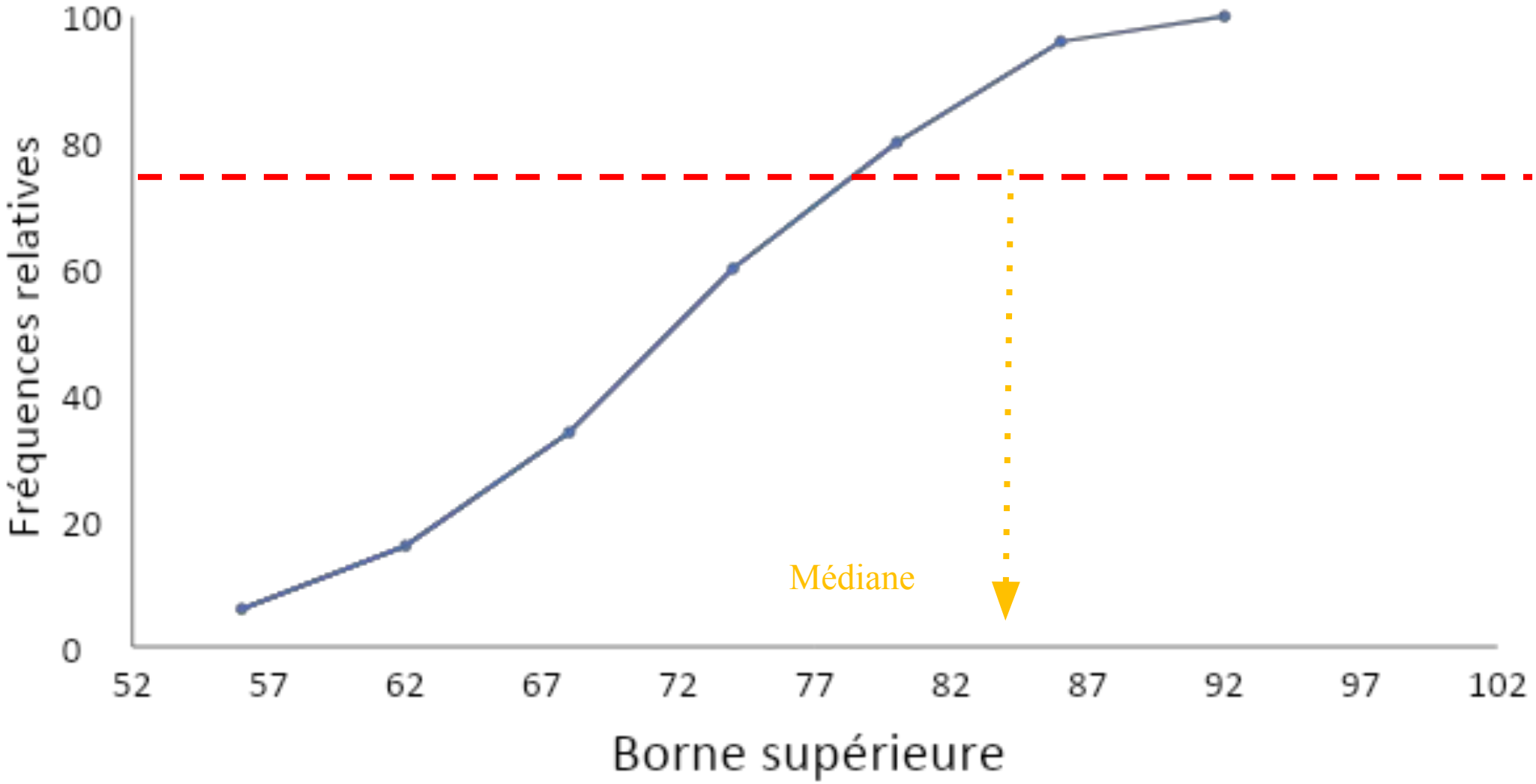
$$Médiane = Borne\ \text{inférieure} + \left((\Delta)\frac{(\frac{N}{2} - nicp)}{(ni)}\right)$$

a. Le mode

$$M_0 = Borne\ \text{inférieure} + \left((\Delta)\frac{(\Delta_1)}{(\Delta_1+\Delta_2)}\right) = \mathbf{68 + 6\left(\frac{(13-9)}{(13-9)+(13-10)}\right) = 71,42}$$



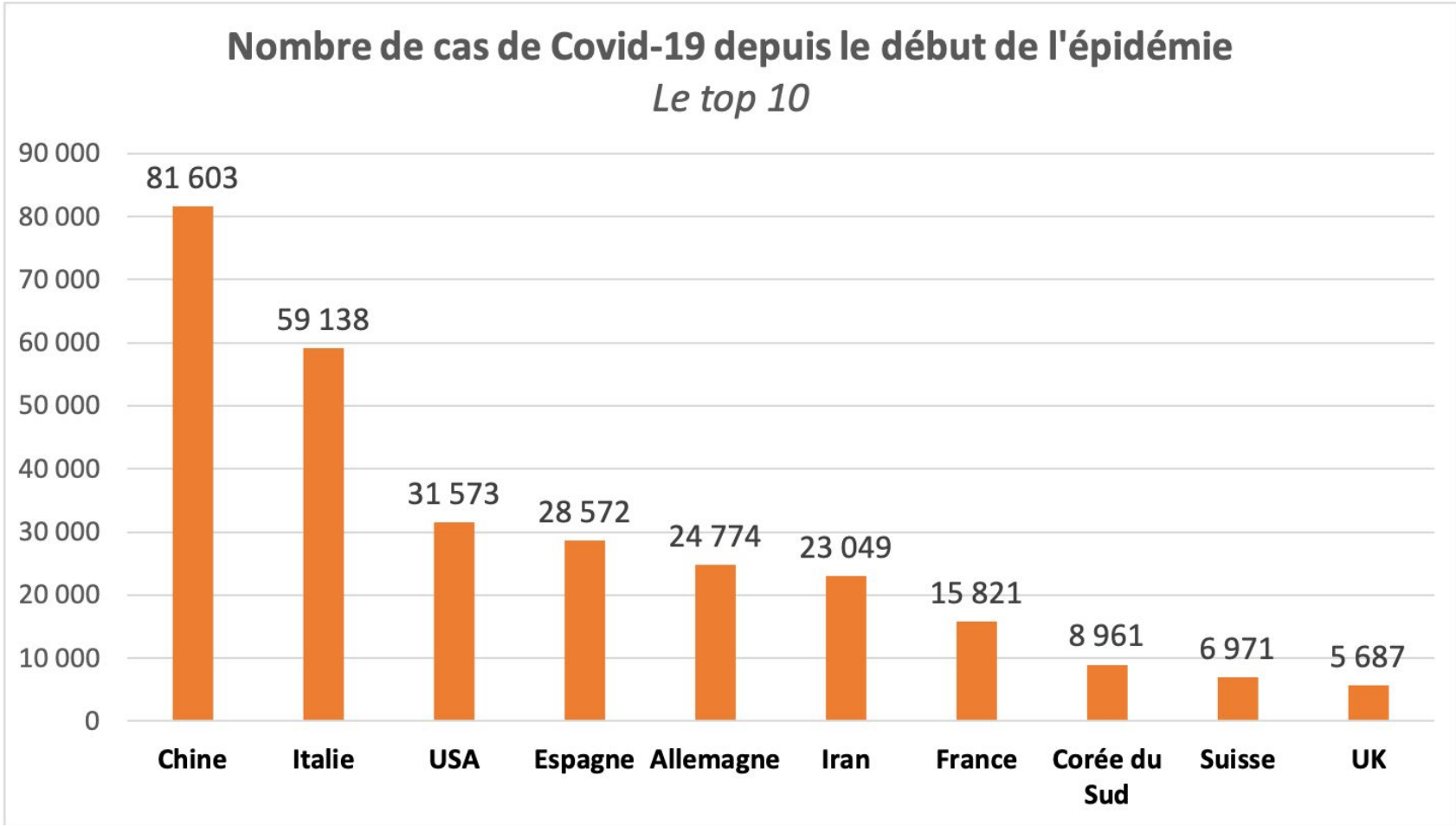
$$Médiane = Borne\ inférieure + \left((\Delta)\frac{(\frac{N}{2} - nicp)}{(nc)}\right) = 68 + 6\left(\frac{\frac{50}{2} - 17}{13}\right) = 71{,}69$$



Variables et représentations graphiques

Représentation d'une variable qualitative

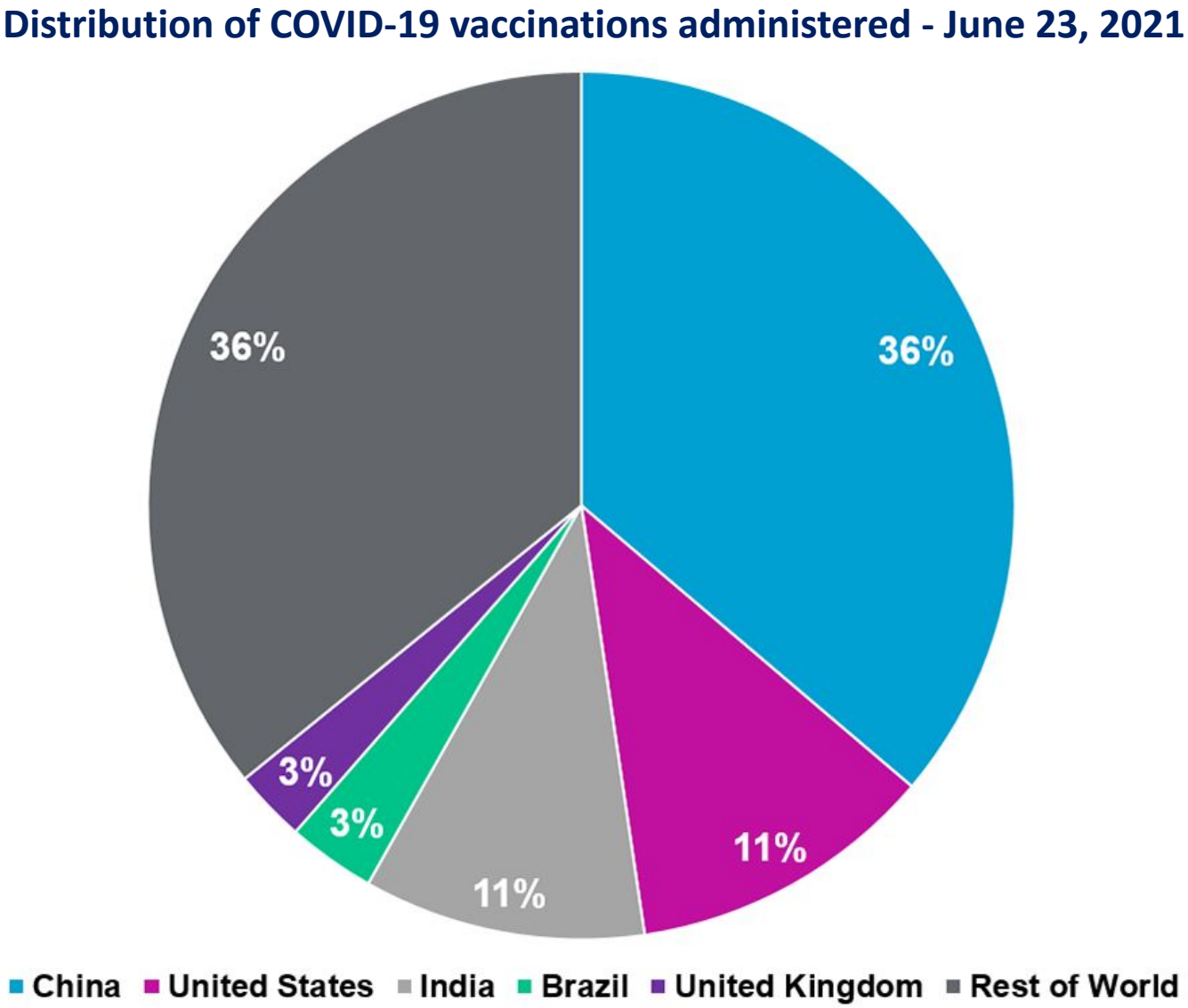
Diagramme en bâtons



Source 🌐

Représentation d'une variable qualitative

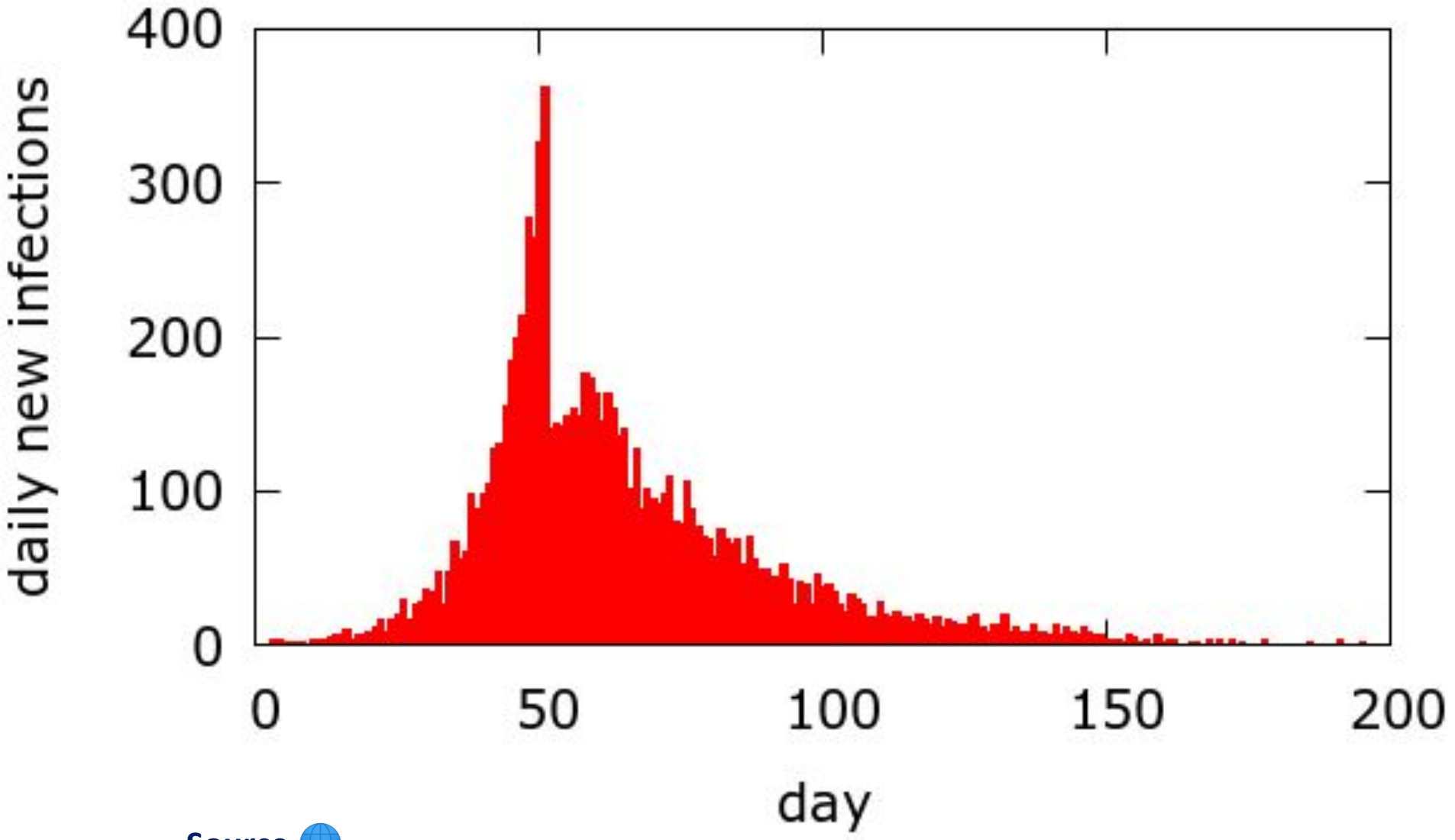
Camembert (pie)



Source 🌐

Représentation d'une variable quantitative continue

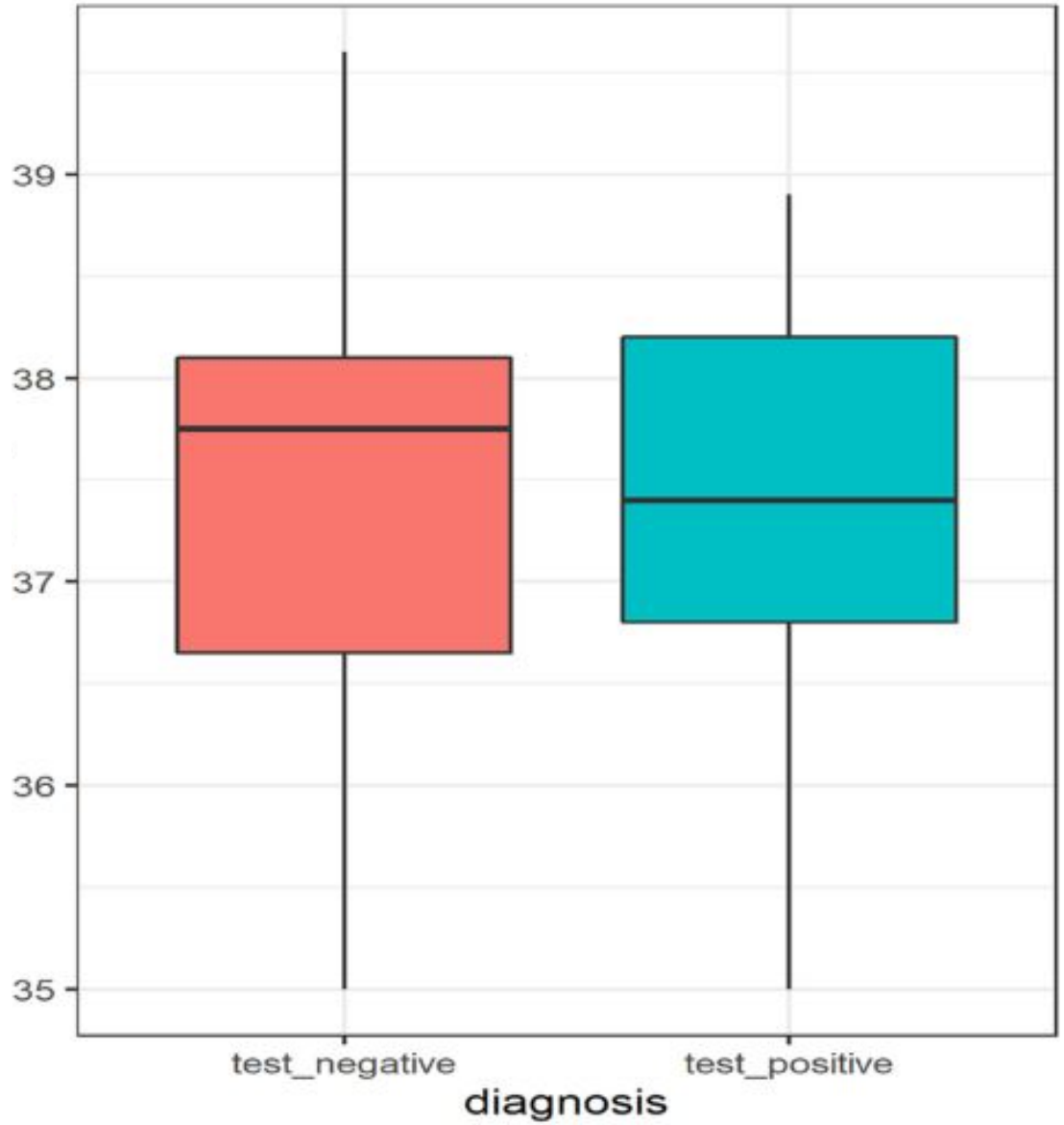
Histogramme



Source

Représentation d’une variable quantitative

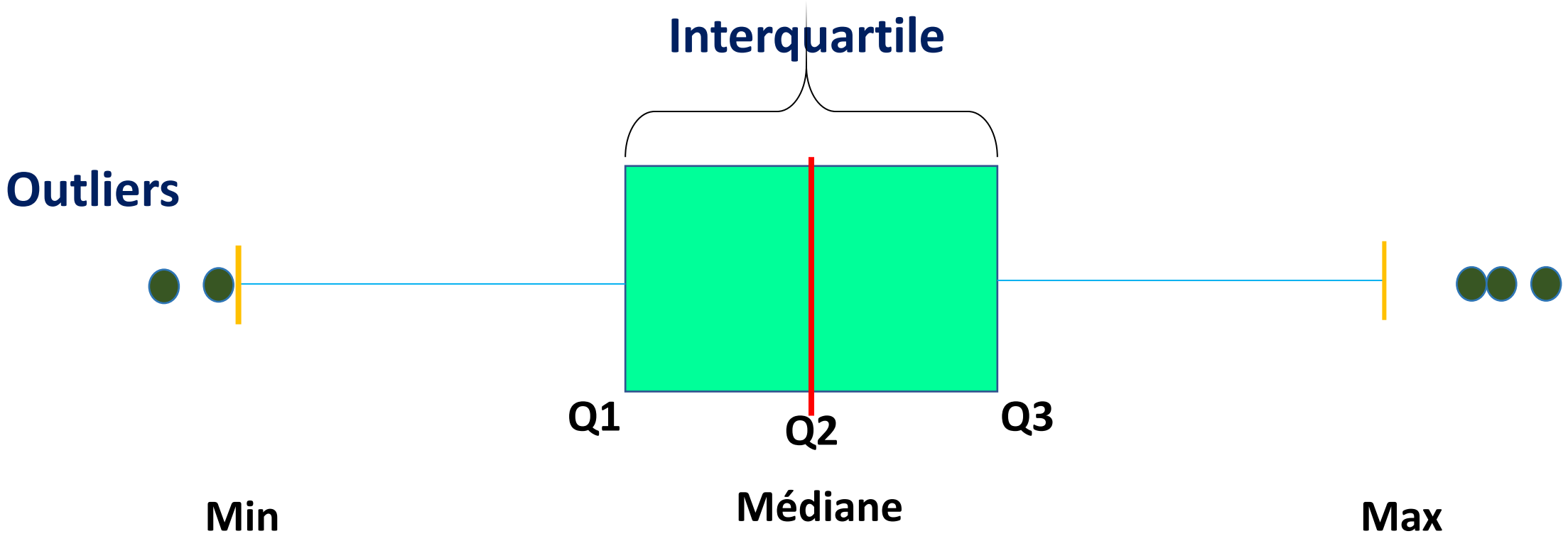
Boite à moustache



Source

Représentation d'une variable quantitative

Boite à moustache



Représentation d'une variable quantitative

Boite à moustache

une boîte à moustaches (ou boxplot) est un autre graphique fréquemment utilisé. Les boîtes à moustaches sont utiles pour révéler la tendance centrale des données, l'étendue des données et la présence de valeurs extrêmes.
La construction d'une boîte à moustaches requiert de calculer d'abord le minimum, le maximum et les quartiles