

1. Rappels

Généralités, vocabulaire.

L'objet de la statistique est l'étude de certaines caractéristiques des éléments d'un ensemble fini. Cet ensemble est appelé population, et les éléments sont les individus.

Les caractéristiques étudiées sont appelées caractères ou variables.

Une variable est:

- qualitative si elle ne peut pas faire l'objet de calculs
- quantitative si elle est numérique et si elle peut faire l'objet de calculs
- continue si elle peut prendre toute valeur dans un intervalle
- discrète dans le cas contraire

Les modalités d'une variable discrète sont les valeurs que peut prendre cette variable.

Dans le cas d'une variable continue, ou lorsqu'il y a trop de modalités, on regroupe les valeurs selon des intervalles disjoints. Chaque intervalle est une classe. Pour les calculs on utilise le centre de chaque classe.

On parle de série statistique pour désigner un recueil de données relatives à une variable numérique, ordonnées par ordre croissant des modalités.

Dans ce qui suit, $x_1, x_2, \dots, x_p$ désignent les valeurs numériques prises par une série statistique, avec les effectifs respectifs $n_1, n_2, \dots, n_p$.

L'effectif d'une modalité est le nombre d'individus présentant cette modalité.

L'effectif total de la population est :
$$N = \sum_{i=1}^p n_i = n_1 + n_2 + \dots + n_p .$$

La fréquence d'une modalité est le rapport entre l'effectif de cette modalité et l'effectif total de la population.

Une fréquence est un nombre compris entre 0 et 1.

La fréquence de la modalité 1 est donc :
$$f_1 = \frac{n_1}{N}$$

L'effectif cumulé d'une modalité est la somme des effectifs des modalités qui lui sont inférieures ou égales.

L'effectif cumulé de la modalité 3 est donc : $N_3 = n_1 + n_2 + n_3 .$

On définit de même la fréquence cumulée.

Les représentations de séries statistiques.

Il existe de nombreuses façons de représenter des données statistiques.

- Diagramme circulaire. (variable qualitative ou quantitative discrète)

Un disque est partagé en secteurs dont les angles sont proportionnels aux effectifs des modalités qu'ils représentent.

- Diagramme en bâtons. (variable qualitative ou quantitative discrète)

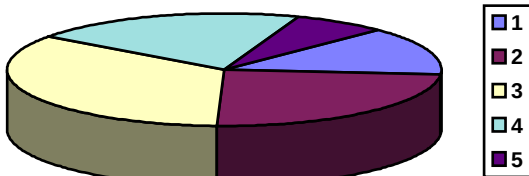
La hauteur du segment représentant une modalité est proportionnel à l'effectif de la modalité.

Exemple

On considère la série statistique suivante:

modalités	1	2	3	4	5
effectifs	4	7	10	6	2

Représentations possibles:



secteurs circulaires

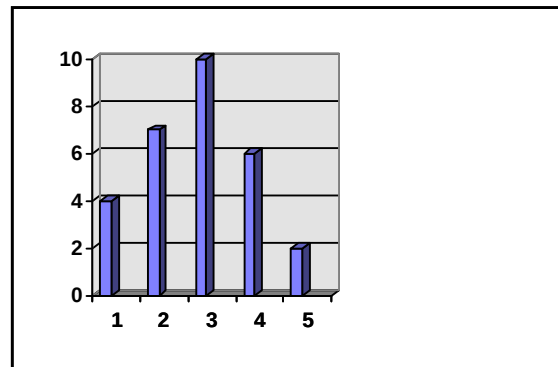


diagramme en bâtons

- Histogramme. (variable quantitative continue)

Dans le cas où les valeurs de la variable sont regroupées en classes, on utilise un histogramme où chaque classe est représentée par un rectangle dont l'aire est proportionnelle à l'effectif de la classe.

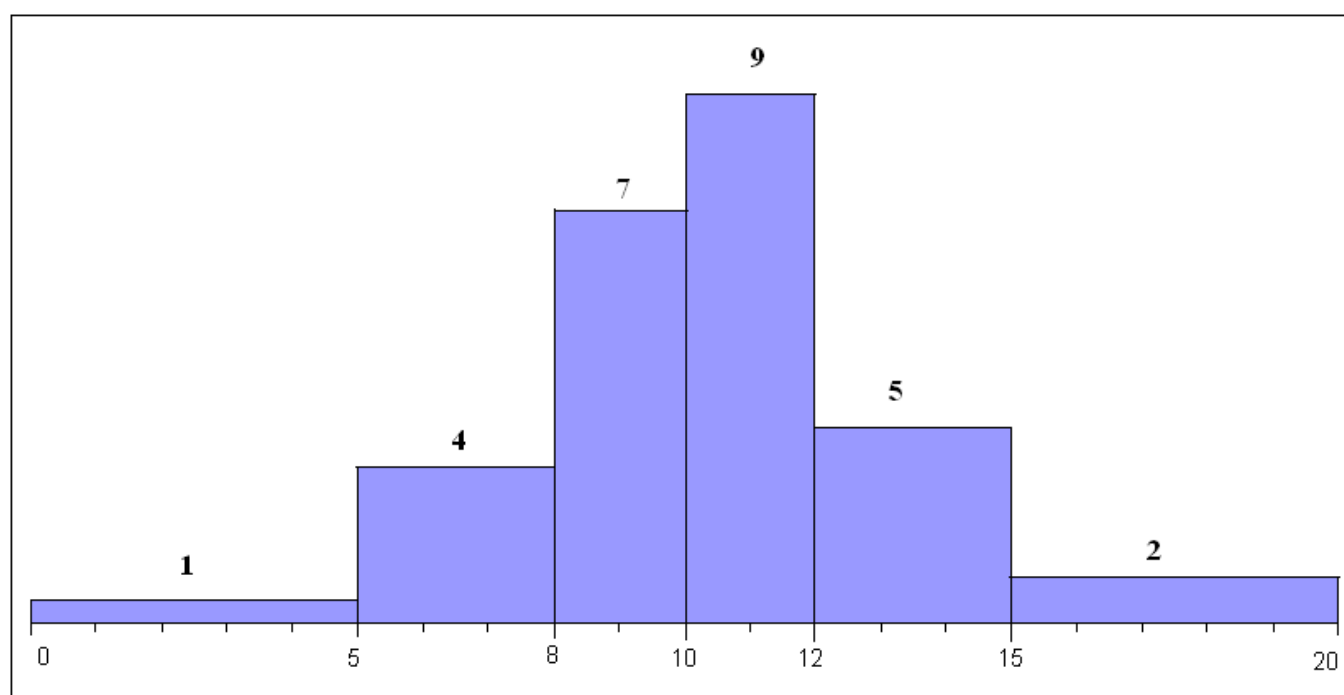
On appelle polygone des effectifs une ligne brisée rejoignant les sommets des bâtons ou les milieux des côtés supérieurs des rectangles .

Exemple

On considère les notes obtenues à un devoir de maths dans une classe de 28 élèves :

note	[0 ; 5 [	[5 ; 8 [	[8 ; 10 [	[10 ; 12 [	[12 ; 15 [	[15 ; 20]
effectif	1	4	7	9	5	2

Voici un histogramme permettant de représenter cette série statistique.



Citons aussi : le diagramme polaire, les pyramides, les représentations symboliques.

2. Caractéristiques centrales et de dispersion

Médiane, quartiles et diagramme en boîte

La médiane (M_e) d'une série statistique est une valeur du caractère qui partage la population en deux sous-ensembles de même effectif.

Si la série est continue, la médiane peut se préciser par interpolation

Si la série est discrète et de taille impaire, ($N = 2n + 1$), la médiane est la valeur du terme de rang $n + 1$.

Si la série est discrète et de taille paire, ($N = 2n$), la médiane est la demi-somme des valeurs des termes de rang n et $n + 1$.

Les quartiles partagent la population en quatre sous-ensembles de même effectif.

Le premier quartile d'une série statistique, noté Q_1 , correspond à la plus petite valeur de la série telle qu'au moins 25% des données soient inférieures ou égales à ce nombre

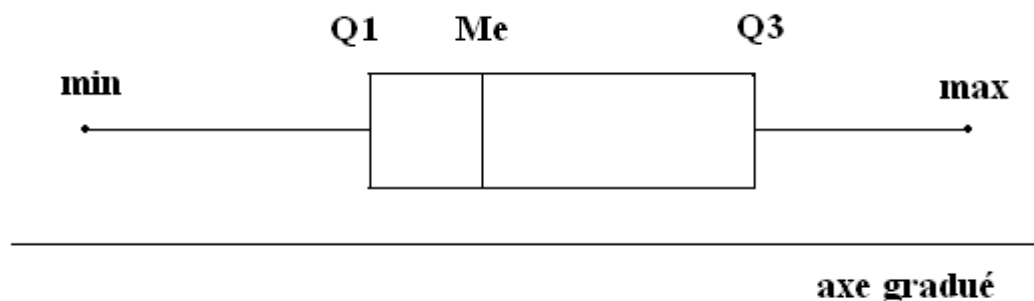
Le second quartile est la médiane

Le troisième quartile d'une série statistique, noté Q_3 , correspond à la plus petite valeur de la série telle qu'au moins 75% des données soient inférieures ou égales à ce nombre

$[Q_1, Q_3]$ est l'intervalle interquartile, $Q_3 - Q_1$ est l'écart interquartile, l'écart interquartile est un paramètre de dispersion.

On peut aussi définir les déciles (partage de la population en 10 parties contenant chacune 10% de l'effectif total) en vingtiles (partage de la population en 20 parties contenant chacune 5% de l'effectif total) ...

Le diagramme en boîte (ou en boîte à moustaches) permet de représenter graphiquement une série statistique avec ses principaux indicateurs. Il contient le maximum et le minimum de la série statistique, le médiane, le premier et le troisième quartiles. Parfois le diagramme se construit entre le premier et le neuvième décile.



Moyenne

$x_1, x_2, \dots, x_p$ désignent les valeurs numériques prises par une série statistique dont les termes sont rangés par ordre croissant, avec les effectifs respectifs $n_1, n_2, \dots, n_p$.

On définit la moyenne $\bar{x}$ par:
$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{n_1 + n_2 + \dots + n_p}$$

Variance et Ecart-type

La variance et l'écart-type permettent de rendre compte de la dispersion d'une série statistique.

La variance est définie par $V = \frac{n_1 x_1^2 + n_2 x_2^2 + \dots + n_p x_p^2}{n_1 + n_2 + \dots + n_p} - \bar{x}^2$, $\bar{x}$ étant la moyenne de la série.

L'écart-type est définie par $\sigma = \sqrt{V} = \sqrt{\frac{n_1 x_1^2 + n_2 x_2^2 + \dots + n_p x_p^2}{n_1 + n_2 + \dots + n_p} - \bar{x}^2}$

p.265 : 15, 20, p.267 : 25	p.261 :1, 2, p.265 :16, 18, p.268 : 30, 31, 33
----------------------------	--

Problèmes:

p.274 : 48, 52	p.272, 273
----------------	------------