

1 Quelques rappels sur les statistiques à une variable

1.1 Moyenne ,Médiane et écart type

1.1.1 Moyenne

Définition

On la note \bar{x} et elle est définie par :
$$\bar{x} = \frac{\sum_{i=1}^k n_i x_i}{\sum_{i=1}^k n_i} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n_1 + n_2 + \dots + n_k}$$

$n_1, n_2, n_3, \dots, n_k$ sont les effectifs correspondants aux valeurs $x_1, x_2, x_3, \dots, x_k$, si la série est discrète , ou les centres de chaque classe, si la série est continue.

1.1.2 Médiane

Définition :

La **médiane** M_e est un paramètre de position

Il permet de couper la population étudiée en deux groupes contenant le même nombre d'individus.

Par définition 50 % de la population étudiée a des valeurs inférieure à M_e et 50 % une valeur supérieure à la M_e .

Ce paramètre est utile pour donner la répartition du caractère étudié.

1.1.3 Variance et écart type

Variance :

Pour calculer la variance V d'une série statistique on utilise la formule :

$$V = \frac{\sum_{i=1}^k n_i (x_i - \bar{x})^2}{\sum_{i=1}^k n_i} = \frac{n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + \dots + n_k (x_k - \bar{x})^2}{n_1 + n_2 + \dots + n_k}$$

Pour calculer la variance , il faut donc calculer d'abord la moyenne.

Ecart-type

L'**écart-type** est le nombre noté σ tel que : $\sigma = \sqrt{V}$

1.2 les Quartiles

Définitions :

On appelle **premier quartile** d'une série la plus petite valeur Q_1 de la série pour laquelle au moins un quart (25%) des données sont inférieures ou égales à Q_1 .

On appelle **troisième quartile** d'une série la plus petite valeur Q_3 de la série pour laquelle au moins trois quarts (75%) des données sont inférieures ou égales à Q_3 .

On appelle **intervalle interquartile** l'intervalle $[Q_1; Q_3]$.

On appelle **écart interquartile** l'amplitude de l'intervalle $[Q_1; Q_3]$ c'est-à-dire le nombre $Q_3 - Q_1$.

1.3 Déciles

Définitions :

On appelle **premier décile** d'une série la plus petite valeur D_1 des termes de la série pour laquelle au moins un dixième (10%) des données sont inférieures ou égales à D_1 .

On appelle **neuvième décile** d'une série la plus petite valeur D_9 des termes de la série pour laquelle au moins neuf dixièmes (90%) des données sont inférieures ou égales à D_9 .

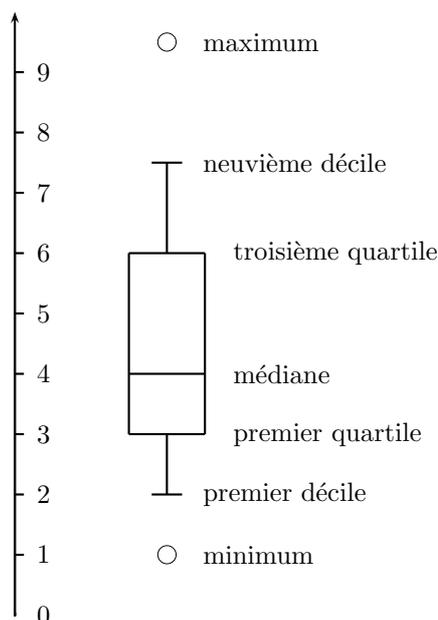
On appelle **intervalle interdécile** l'intervalle $[D_1; D_9]$.

On appelle **écart interdécile** l'amplitude de l'intervalle $[D_1; D_9]$, c'est-à-dire le nombre $D_9 - D_1$.

1.4 Diagramme en boîte

Ce type de diagramme est aussi appelé diagramme de Tuckey, boîte à moustaches ou boîte à pattes.

Il utilise le 1er et le 3ème quartile, les valeurs extrêmes, le 1er et le 9ème décile et éventuellement la médiane d'une série.



Le corps du diagramme, c'est-à-dire la boîte est formée d'un rectangle ayant pour extrémité inférieure le 1er quartile et pour extrémité supérieure le 3ème quartile.
A l'intérieur de ce rectangle on pourra tracer un segment représentant la médiane.

La largeur du rectangle n'est pas fixée, elle sera choisie de façon à obtenir un graphique harmonieux. Ce rectangle représente les données contenues dans l'intervalle interquartile.

On repère ensuite les hauteurs correspondant au 1er et au 9ème décile, et on trace deux pattes représentant les données contenues dans l'intervalle interdécile.

(la largeur des pattes n'a pas d'importance).

On peut ensuite terminer le graphique, en faisant figurer par des points les données qui sont en dehors de l'intervalle interdécile.

Si certaines données, sont manifestement très éloignées, on ne les représentera pas, mais on écrira leurs valeurs au dessous du diagramme.

Remarques :

*Une boîte avec des pattes courtes indique que la série est assez concentrée autour de sa médiane.
Au contraire des pattes longues indique que la série est assez dispersée.*

Un des avantages de cette représentation, est qu'elle nécessite très peu de calculs.

La représentation peut aussi se faire horizontalement, la graduation se trouvant alors sur l'axe horizontal, d'où l'appellation de "boîte à moustaches".

Le graphique est parfois fait en dessinant des pattes correspondant au 1er et au 99ème centile, ou même aux valeurs extrêmes.

1.5 Exercice

On considère la série suivante :

3, 3, 3, 4, 5, 5, 5, 6, 6, 7, 8, 8, 9, 9, 9, 9, 9, 10, 10, 11, 11, 11, 11, 12, 13, 13, 13, 14, 14, 15, 15, 17,

1. Déterminer la moyenne et l'écart type de cette série.
2. Déterminer la médiane et les quartiles de cette série.
3. Déterminer le premier et le neuvième décile de cette série.
4. construire la boîte à moustache de cette série.

Solution :

1. $\bar{x} \approx 9,26$ et $\sigma \approx 3,85$.
2. L'effectif total est $N=31$ et 50% de $31 = 15,5$ donc la médiane est la 16^{me} valeur de cette série ordonnée de manière croissante donc la médiane est $M_e = 9$.
 25% de $31 = 7,75$ donc le premier quartile est la $8^{ième}$ valeur de cette série ordonnée de manière croissante donc le premier quartile est $Q_1 = 6$.
 75% de $31 = 23,25$ donc le troisième quartile est la $24^{ième}$ valeur de cette série ordonnée de manière croissante donc le troisième quartile est $Q_3 = 13$.
3. 10% de $31 = 3,1$ donc le premier décile est la $4^{ième}$ valeur de cette série ordonnée de manière croissante donc le premier décile est $D_1 = 4$.
 90% de $31 = 27,9$ donc le neuvième décile est la $28^{ième}$ valeur de cette série ordonnée de manière croissante donc le neuvième décile est $D_9 = 14$.
4. On en déduit le diagramme suivant :

