Chapitre 1

Statistiques descriptives

Une étude statistique a pour but d'obtenir une information, appelée caractère, sur une population à partir de données recueillies sur un échantillon de cette population.

Une étude statistique a pour but d'obtenir une information, appelée caractère, sur une population à partir de données recueillies sur un échantillon de cette population.

Définition

Le caractère étudié peut être :

Une étude statistique a pour but d'obtenir une information, appelée caractère, sur une population à partir de données recueillies sur un échantillon de cette population.

Définition

Le caractère étudié peut être :

quantitatif : les valeurs du caractère s'expriment avec des nombres (ex : températures, pointures, salaires...) ;

Une étude statistique a pour but d'obtenir une information, appelée caractère, sur une population à partir de données recueillies sur un échantillon de cette population.

Définition

Le caractère étudié peut être :

quantitatif : les valeurs du caractère s'expriment avec des nombres (ex : températures, pointures, salaires. . .) ;

qualitatif : les valeurs ne s'expriment pas par des nombres (ex : couleurs, type d'essence. . .) ;

Une étude statistique a pour but d'obtenir une information, appelée caractère, sur une population à partir de données recueillies sur un échantillon de cette population.

Définition

Le caractère étudié peut être :

- quantitatif : les valeurs du caractère s'expriment avec des nombres (ex : températures, pointures, salaires. . .) ;
- qualitatif : les valeurs ne s'expriment pas par des nombres (ex : couleurs, type d'essence. . .) ;
- discret : les valeurs du caractère sont isolés (ex : notes. . .) ;

Une étude statistique a pour but d'obtenir une information, appelée caractère, sur une population à partir de données recueillies sur un échantillon de cette population.

Définition

Le caractère étudié peut être :

quantitatif : les valeurs du caractère s'expriment avec des nombres (ex : températures, pointures, salaires. . .) ;

qualitatif : les valeurs ne s'expriment pas par des nombres (ex : couleurs, type d'essence. . .) ;

discret : les valeurs du caractère sont isolés (ex : notes. . .) ;

continu : les valeurs sont regroupées par classes (ou intervalles de nombre) (par ex : durée, distance parcourue,

Remarque.

Dans la suite du cours, on considère que le caractère est quantitatif.

Définition

Lorsque les valeurs d'une série statistique sont regroupés par classe de type $[a\,;b[$, on appelle centre de classe le nombre défini par $\dfrac{a+b}{2}$.

Définition

Lorsque les valeurs d'une série statistique sont regroupés par classe de type $[a\,;b[$, on appelle centre de classe le nombre défini par $\dfrac{a+b}{2}$.

Exemple

Le centre de la classe $[150\,;300[$ est égal à

$$\frac{150+300}{2}=\frac{450}{2}=225.$$

Définition

On considère une série qui possède des valeurs différentes $x_1, x_2, \ldots, x_p$ chacune affectée de leur effectif. L'effectif total est égal à N.
On appelle moyenne d'une série le nombre $\overline{m}$ tel que :

$$\overline{m} = \frac{n_1 x_1 + n_2 x_2 + n_3 x_3 + \cdots + n_p x_p}{n_1 + n_2 + \cdots + n_p} =$$

Exemple

Un élève souhaite calculer la moyennes de ses notes sur 20 :
8; 10; 14; 13; 10; 14; 10.

Exemple

Un élève souhaite calculer la moyennes de ses notes sur 20 :
8; 10; 14; 13; 10; 14; 10.

Parfois, les valeurs sont regroupées dans un tableau :

Notes	8	10	13	14
Effectifs	1	3	1	2

On peut alors utiliser la formule :

Remarques

Remarques

1. En règle générale, la moyenne n'est pas une valeur de la série. On peut la considérer comme le point d'équilibre des valeurs. Par conséquent, la moyenne est sensible aux valeurs extrêmes.
   Pour reprendre l'exemple, si la prochaine note du contrôle est très élevée alors la moyenne va augmenter.

Remarques

1. En règle générale, la moyenne n'est pas une valeur de la série. On peut la considérer comme le point d'équilibre des valeurs. Par conséquent, la moyenne est sensible aux valeurs extrêmes.
   Pour reprendre l'exemple, si la prochaine note du contrôle est très élevée alors la moyenne va augmenter.
2. Lorsque les valeurs sont regroupés par classe, au lieu d'utiliser les x_i, on utilise les centres de classe.

Exemple

Le tableau ci-dessous regroupe le temps de parcours des habitants d'un village entre leur domicile et leur lieu de travail. Le maire cherche à calculer le temps moyen.

Durée en min	$[0; 10[$	$[10; 20[$	$[20; 30[$	$[30; 50[$	$[50; 70[$	Total
Effectif	18	35	25	112	80	270
Centre de classe	5	15	25	40	60	

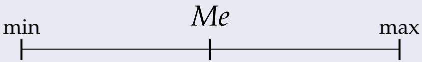
Définition

Définition

On considère une série statistique dont l'effectif total est égal à N. Les valeurs sont rangées dans l'ordre croissant.
Une médiane Me est un nombre réel qui permet de partager la série statistique en deux séries de même valeur.
Autrement dit, la moitié (50%) des valeurs de la série est inférieure ou égale à Me et l'autre moitié est supérieure ou égale à Me.

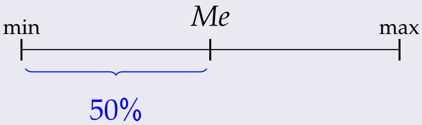
Définition

On considère une série statistique dont l'effectif total est égal à N. Les valeurs sont rangées dans l'ordre croissant.
Une médiane *Me* est un nombre réel qui permet de partager la série statistique en deux séries de même valeur.
Autrement dit, la moitié (50%) des valeurs de la série est inférieure ou égale à *Me* et l'autre moitié est supérieure ou égale à *Me*.



Définition

On considère une série statistique dont l'effectif total est égal à N. Les valeurs sont rangées dans l'ordre croissant.
Une médiane Me est un nombre réel qui permet de partager la série statistique en deux séries de même valeur.
Autrement dit, la moitié (50%) des valeurs de la série est inférieure ou égale à Me et l'autre moitié est supérieure ou égale à Me.

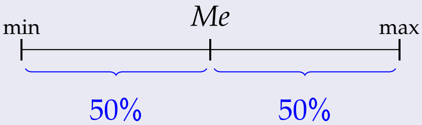


Définition

On considère une série statistique dont l'effectif total est égal à N. Les valeurs sont rangées dans l'ordre croissant.

Une médiane Me est un nombre réel qui permet de partager la série statistique en deux séries de même valeur.

Autrement dit, la moitié (50%) des valeurs de la série est inférieure ou égale à Me et l'autre moitié est supérieure ou égale à Me.



Méthode de calcul :

Méthode de calcul :

Par définition, la médiane dépend de l'effectif de la série :

Méthode de calcul :

Par définition, la médiane dépend de l'effectif de la série :

- Si N est impair, alors on calcule $\frac{N+1}{2}$ et le résultat correspond à la position de la médiane choisie dans la série.

Méthode de calcul :

Par définition, la médiane dépend de l'effectif de la série :

- Si N est impair, alors on calcule $\frac{N+1}{2}$ et le résultat correspond à la position de la médiane choisie dans la série.
- Si N est pair, alors la médiane choisie est égale à la moyenne de la valeur situé à la position $\frac{N}{2}$ et la valeur suivante.

Exemple

On considère le relevé des températures en janvier et en février dans une ville :

Janvier									
Valeurs	$-3°$	$-2°$	$-1°$	$0°$	$1°$	$2°$	$3°$	$4°$	Total
Effectifs	3	5	8	5	4	3	2	1	31
Février									
Valeurs	$-3°$	$-2°$	$-1°$	$0°$	$1°$	$2°$	$3°$	$4°$	Total
Effectifs	1	2	3	3	5	9	3	2	28

Exemple

En janvier : $N = 31$

Exemple

En janvier : $N = 31$

En févier : $N = 28$

Remarque

Contrairement à la moyenne, la médiane n'est pas sensible aux valeurs extrêmes. En effet, dans notre exemple, si la dernière température était égale à 20° au lieu de 4°, la moyenne augmenterait alors que la médiane resterait identique car l'effectif total n'a pas changé.

Définition

On considère une série statistique S dont l'effectif total est égal à N. Les valeurs sont rangées dans l'ordre croissant.

Définition

On considère une série statistique S dont l'effectif total est égal à N. Les valeurs sont rangées dans l'ordre croissant.

- Le premier quartile Q_1 de S est le plus petit élément a de S tel qu'au moins 25% des données soient inférieures ou égales à a.

Définition

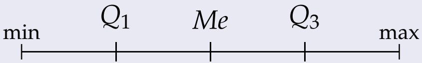
On considère une série statistique S dont l'effectif total est égal à N. Les valeurs sont rangées dans l'ordre croissant.

- Le premier quartile Q_1 de S est le plus petit élément a de S tel qu'au moins 25% des données soient inférieures ou égales à a.
- Le troisième quartile Q_3 de S est le plus petit élément b de S tel qu'au moins 75% des données soient inférieures ou égales à b.

Définition

On considère une série statistique S dont l'effectif total est égal à N. Les valeurs sont rangées dans l'ordre croissant.

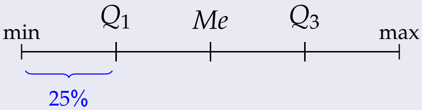
- Le premier quartile Q_1 de S est le plus petit élément a de S tel qu'au moins 25% des données soient inférieures ou égales à a.
- Le troisième quartile Q_3 de S est le plus petit élément b de S tel qu'au moins 75% des données soient inférieures ou égales à b.



Définition

On considère une série statistique S dont l'effectif total est égal à N. Les valeurs sont rangées dans l'ordre croissant.

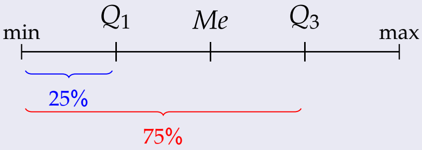
- Le premier quartile Q_1 de S est le plus petit élément a de S tel qu'au moins 25% des données soient inférieures ou égales à a.
- Le troisième quartile Q_3 de S est le plus petit élément b de S tel qu'au moins 75% des données soient inférieures ou égales à b.



Définition

On considère une série statistique S dont l'effectif total est égal à N. Les valeurs sont rangées dans l'ordre croissant.

- Le premier quartile Q_1 de S est le plus petit élément a de S tel qu'au moins 25% des données soient inférieures ou égales à a.
- Le troisième quartile Q_3 de S est le plus petit élément b de S tel qu'au moins 75% des données soient inférieures ou égales à b.



Méthode de calcul :

Méthode de calcul :

Par définition, les quartiles dépendent de l'effectif de la série :

Méthode de calcul :

Par définition, les quartiles dépendent de l'effectif de la série :

Premier quartile : On arrondit le nombre $\frac{N}{4}$ à l'unité par excès et cela donne la position de Q_1 dans la série S.

Méthode de calcul :

Par définition, les quartiles dépendent de l'effectif de la série :

Premier quartile : On arrondit le nombre $\frac{N}{4}$ à l'unité par excès et cela donne la position de Q_1 dans la série S.

Troisième quartile : On arrondit le nombre $3 \times \frac{N}{4}$ à l'unité par excès et cela donne la position de Q_3 dans la série S.

Exemple

On reprend les températures du moins de janvier de l'exemple précédent.

Exemple

On reprend les températures du moins de janvier de l'exemple précédent.

Premier quartile : $N = 31$

Exemple

On reprend les températures du moins de janvier de l'exemple précédent.

Premier quartile : $N = 31$

Troisième quartile : $N = 31$

Exemple

On reprend les températures du moins de janvier de l'exemple précédent.

Premier quartile : $N = 31$

Troisième quartile : $N = 31$

Exemple

On reprend les températures du moins de janvier de l'exemple précédent.

Premier quartile : $N = 31$

Troisième quartile : $N = 31$

Conclusion :

Dans les classes antérieures, l'étendue était un indicateur de dispersion utilisé dont voici rappelée la définition :

Définition

Dans une série statistique, l'étendue est la différence entre la valeur maximale et la valeur minimale.

Remarque

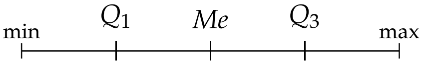
Une valeur élevée de l'étendue signifie qu'au moins une des valeurs extrêmes de la série est éloignée de la médiane et le risque de dispersion des valeurs est donc plus important.
Dans l'exemple précédent, l'étendue de la série janvier était identique à celle de la série février ce qui montre les besoins d'avoir d'autres indicateurs.
Les quartiles nous permettent d'obtenir d'autres indicateurs de dispersion lié à la médiane.

Définition

On appelle l'**intervalle interquartile** l'intervalle $[Q_1 ; Q_3]$.
On appelle l'**écart interquartile** le nombre $Q_3 - Q_1$.

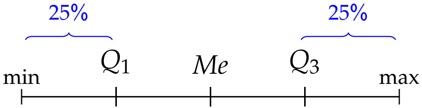
Définition

On appelle l'**intervalle interquartile** l'intervalle $[Q_1\,;Q_3]$.
On appelle l'**écart interquartile** le nombre $Q_3 - Q_1$.



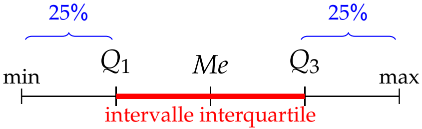
Définition

On appelle l'intervalle interquartile l'intervalle $[Q_1 \,;\, Q_3]$.
On appelle l'écart interquartile le nombre $Q_3 - Q_1$.



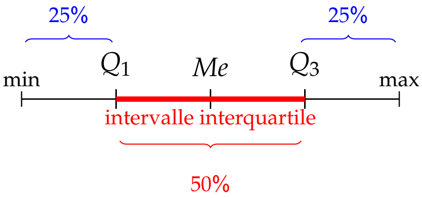
Définition

On appelle l'intervalle interquartile l'intervalle $[Q_1 ; Q_3]$.
On appelle l'écart interquartile le nombre $Q_3 - Q_1$.



Définition

On appelle l'intervalle interquartile l'intervalle $[Q_1 ; Q_3]$.
On appelle l'écart interquartile le nombre $Q_3 - Q_1$.



Remarque

Remarque

Un écart interquartile faible impose que les valeurs soient regroupées proches de la médiane et donc peu dispersée pour la moitié d'entre elles.

Définition

Soit S une série statistique de valeurs $x_1, x_2, \ldots, x_n$ et de moyenne $\overline{m}$. La variance est le nombre positif ou nul V défini par :

$$V =$$

Exemple

Prenons une série de notes :

Notes	7	9	12	13	15	18
Effectifs	1	2	1	2	4	2
$(x_i - \overline{m})$						
$(x_i - \overline{m})^2$						

Théorème (de König-Huygens (admis))

Soit S une série statistique de valeurs $x_1, x_2, \ldots, x_n$ et de moyenne $\overline{m}$.
On pose $\overline{n} = \frac{x_1^2+x_2^2+\cdots+x_n^2}{n}$ (moyenne des carrés des valeurs).
La variance V de la série statistique est alors égale à :

$$V = \overline{n} - (\overline{m})^2.$$

Exemple

Prenons une série de notes :

Notes	7	9	12	13	15	18
Effectifs	1	2	1	2	4	2
$(x_i)^2$	49	81	144	169	225	324

Cela n'est cependant pas exploitable car l'unité de mesure n'est plus la même depuis que l'on a utilisé des nombres au carré. On s'intéresse donc surtout à l'indicateur suivant :

Définition

Soit S une série statistique de variance V. L'écart-type σ est le nombre réel positif défini par :

Remarque

L'écart-type est un nombre réel positif qui caractérise la dispersion des valeurs d'une série statistiques par rapport à la moyenne.
Plus l'écart-type est petit et plus les valeurs sont proches de la moyenne (dispersion faible). Au contraire, si l'écart-type est grand alors les valeurs sont éloignés de la moyenne (dispersion élevée).

Remarque

L'écart-type est un nombre réel positif qui caractérise la dispersion des valeurs d'une série statistiques par rapport à la moyenne.
Plus l'écart-type est petit et plus les valeurs sont proches de la moyenne (dispersion faible). Au contraire, si l'écart-type est grand alors les valeurs sont éloignés de la moyenne (dispersion élevée).

Exemple

Dans l'exemple précédent, on obtient $\sigma = \sqrt{11,1875} \approx 3,34$.
Cela donne une indication de la répartition moyenne des notes autour de la moyenne $\overline{m}$.