Introduction à la statistique descriptive

Karim Saïd

Ecole technique

Année scolaire 2020-2021

Introduction à la statistique descriptive

Karim Saïd

Ecole technique

Année scolaire 2020-2021

Statistique vient du mot latin *status* qui signifie état, situation. Les premières ébauches de la statistique remontent aux recensements qui furent mis sur pied dès les premiers siècles de notre ère. Ce n'est pourtant qu'au 18$^{\text{ème}}$ siècle qu'elle se constitue comme une discipline scientifique autonome. Aujourd'hui, la statistique est une branche des mathématiques appliquées intervenant dans divers domaines de la pensée humaine tels que la démographie, l'économie, la médecine, l'agronomie ou encore l'industrie.

Statistique vient du mot latin *status* qui signifie état, situation. Les premières ébauches de la statistique remontent aux recensements qui furent mis sur pied dès les premiers siècles de notre ère. Ce n'est pourtant qu'au $18^{\text{ème}}$ siècle qu'elle se constitue comme une discipline scientifique autonome. Aujourd'hui, la statistique est une branche des mathématiques appliquées intervenant dans divers domaines de la pensée humaine tels que la démographie, l'économie, la médecine, l'agronomie ou encore l'industrie. La démarche statistique peut se décomposer en cinq étapes.

Statistique vient du mot latin *status* qui signifie état, situation. Les premières ébauches de la statistique remontent aux recensements qui furent mis sur pied dès les premiers siècles de notre ère. Ce n'est pourtant qu'au 18$^{\text{ème}}$ siècle qu'elle se constitue comme une discipline scientifique autonome. Aujourd'hui, la statistique est une branche des mathématiques appliquées intervenant dans divers domaines de la pensée humaine tels que la démographie, l'économie, la médecine, l'agronomie ou encore l'industrie. La démarche statistique peut se décomposer en cinq étapes.Premièrement, il s'agit d'identifier précisément la population et le (les) caractère(s) à étudier.

Statistique vient du mot latin *status* qui signifie état, situation. Les premières ébauches de la statistique remontent aux recensements qui furent mis sur pied dès les premiers siècles de notre ère. Ce n'est pourtant qu'au 18$^{\text{ème}}$ siècle qu'elle se constitue comme une discipline scientifique autonome. Aujourd'hui, la statistique est une branche des mathématiques appliquées intervenant dans divers domaines de la pensée humaine tels que la démographie, l'économie, la médecine, l'agronomie ou encore l'industrie. La démarche statistique peut se décomposer en cinq étapes.Premièrement, il s'agit d'identifier précisément la population et le (les) caractère(s) à étudier.Suite à cela, des données seront récoltées par recensement ou échantillonnage.

Statistique vient du mot latin *status* qui signifie état, situation. Les premières ébauches de la statistique remontent aux recensements qui furent mis sur pied dès les premiers siècles de notre ère. Ce n'est pourtant qu'au 18ème siècle qu'elle se constitue comme une discipline scientifique autonome. Aujourd'hui, la statistique est une branche des mathématiques appliquées intervenant dans divers domaines de la pensée humaine tels que la démographie, l'économie, la médecine, l'agronomie ou encore l'industrie. La démarche statistique peut se décomposer en cinq étapes.Premièrement, il s'agit d'identifier précisément la population et le (les) caractère(s) à étudier.Suite à cela, des données seront récoltées par recensement ou échantillonnage.Ensuite, il faudra regrouper, classifier et présenter les données (*statistique descriptive*).

Statistique vient du mot latin *status* qui signifie état, situation. Les premières ébauches de la statistique remontent aux recensements qui furent mis sur pied dès les premiers siècles de notre ère. Ce n'est pourtant qu'au 18$^{\text{ème}}$ siècle qu'elle se constitue comme une discipline scientifique autonome. Aujourd'hui, la statistique est une branche des mathématiques appliquées intervenant dans divers domaines de la pensée humaine tels que la démographie, l'économie, la médecine, l'agronomie ou encore l'industrie. La démarche statistique peut se décomposer en cinq étapes.Premièrement, il s'agit d'identifier précisément la population et le (les) caractère(s) à étudier.Suite à cela, des données seront récoltées par recensement ou échantillonnage.Ensuite, il faudra regrouper, classifier et présenter les données (*statistique descriptive*).Il conviendra alors de comparer les résultats avec des modèles théoriques (*calcul des probabilités*).

Statistique vient du mot latin *status* qui signifie état, situation. Les premières ébauches de la statistique remontent aux recensements qui furent mis sur pied dès les premiers siècles de notre ère. Ce n'est pourtant qu'au $18^{\text{ème}}$ siècle qu'elle se constitue comme une discipline scientifique autonome. Aujourd'hui, la statistique est une branche des mathématiques appliquées intervenant dans divers domaines de la pensée humaine tels que la démographie, l'économie, la médecine, l'agronomie ou encore l'industrie. La démarche statistique peut se décomposer en cinq étapes.Premièrement, il s'agit d'identifier précisément la population et le (les) caractère(s) à étudier.Suite à cela, des données seront récoltées par recensement ou échantillonnage.Ensuite, il faudra regrouper, classifier et présenter les données (*statistique descriptive*).Il conviendra alors de comparer les résultats avec des modèles théoriques (*calcul des probabilités*).Enfin, il s'agira d'interpréter les résultats et d'établir des hypothèses plausibles en vue de prévisions (*statistique inférentielle*) concernant des circonstances analogues.

Statistique vient du mot latin *status* qui signifie état, situation. Les premières ébauches de la statistique remontent aux recensements qui furent mis sur pied dès les premiers siècles de notre ère. Ce n'est pourtant qu'au $18^{\text{ème}}$ siècle qu'elle se constitue comme une discipline scientifique autonome. Aujourd'hui, la statistique est une branche des mathématiques appliquées intervenant dans divers domaines de la pensée humaine tels que la démographie, l'économie, la médecine, l'agronomie ou encore l'industrie.
La démarche statistique peut se décomposer en cinq étapes.Premièrement, il s'agit d'identifier précisément la population et le (les) caractère(s) à étudier.Suite à cela, des données seront récoltées par recensement ou échantillonnage.Ensuite, il faudra regrouper, classifier et présenter les données (*statistique descriptive*).Il conviendra alors de comparer les résultats avec des modèles théoriques (*calcul des probabilités*).Enfin, il s'agira d'interpréter les résultats et d'établir des hypothèses plausibles en vue de prévisions (*statistique inférentielle*) concernant des circonstances analogues.
Ce chapitre se borne à une introduction à la statistique descriptive en présentant, sur la base de deux exemples illustratifs, les quelques mesures qui caractérisent un ensemble fini de données.

Definition

1. On appelle *population* l'ensemble de référence sur lequel vont porter les observations. Il est d'usage de désigner par la lettre N la taille d'une population.

Definition

1. On appelle *population* l'ensemble de référence sur lequel vont porter les observations. Il est d'usage de désigner par la lettre N la taille d'une population.
2. On appelle *échantillon* une partie de la population que l'on détermine par sondage lorsque la population est trop nombreuse à étudier ou impossible à observer dans sa totalité.

Definition

1. On appelle *population* l'ensemble de référence sur lequel vont porter les observations. Il est d'usage de désigner par la lettre N la taille d'une population.
2. On appelle *échantillon* une partie de la population que l'on détermine par sondage lorsque la population est trop nombreuse à étudier ou impossible à observer dans sa totalité.
3. On appelle *individu* tout élément de la population.

Definition

1. On appelle *population* l'ensemble de référence sur lequel vont porter les observations. Il est d'usage de désigner par la lettre N la taille d'une population.
2. On appelle *échantillon* une partie de la population que l'on détermine par sondage lorsque la population est trop nombreuse à étudier ou impossible à observer dans sa totalité.
3. On appelle *individu* tout élément de la population.
4. Lorsque l'on peut ainsi étudier une caractéristique que possède chacun des individus, on appelle cela une *variable statistique* ou *caractère*.

Definition

1. On appelle *population* l'ensemble de référence sur lequel vont porter les observations. Il est d'usage de désigner par la lettre N la taille d'une population.
2. On appelle *échantillon* une partie de la population que l'on détermine par sondage lorsque la population est trop nombreuse à étudier ou impossible à observer dans sa totalité.
3. On appelle *individu* tout élément de la population.
4. Lorsque l'on peut ainsi étudier une caractéristique que possède chacun des individus, on appelle cela une *variable statistique* ou *caractère*.
5. Les différentes valeurs que peut prendre une variable statistique sont les *modalités* de cette variable.

Definition

1. On appelle *population* l'ensemble de référence sur lequel vont porter les observations. Il est d'usage de désigner par la lettre N la taille d'une population.
2. On appelle *échantillon* une partie de la population que l'on détermine par sondage lorsque la population est trop nombreuse à étudier ou impossible à observer dans sa totalité.
3. On appelle *individu* tout élément de la population.
4. Lorsque l'on peut ainsi étudier une caractéristique que possède chacun des individus, on appelle cela une *variable statistique* ou *caractère*.
5. Les différentes valeurs que peut prendre une variable statistique sont les *modalités* de cette variable.
6. Le nombre d'individus vérifiant une modalité donnée est appelé *l'effectif*.

Définitions

Definition

1. On appelle *population* l'ensemble de référence sur lequel vont porter les observations. Il est d'usage de désigner par la lettre N la taille d'une population.
2. On appelle *échantillon* une partie de la population que l'on détermine par sondage lorsque la population est trop nombreuse à étudier ou impossible à observer dans sa totalité.
3. On appelle *individu* tout élément de la population.
4. Lorsque l'on peut ainsi étudier une caractéristique que possède chacun des individus, on appelle cela une *variable statistique* ou *caractère*.
5. Les différentes valeurs que peut prendre une variable statistique sont les *modalités* de cette variable.
6. Le nombre d'individus vérifiant une modalité donnée est appelé *l'effectif*.
7. La *fréquence* d'une modalité est le rapport entre l'effectif et le nombre d'observations. On l'exprime souvent en pour cent.

Definition

1. On appelle *population* l'ensemble de référence sur lequel vont porter les observations. Il est d'usage de désigner par la lettre N la taille d'une population.
2. On appelle *échantillon* une partie de la population que l'on détermine par sondage lorsque la population est trop nombreuse à étudier ou impossible à observer dans sa totalité.
3. On appelle *individu* tout élément de la population.
4. Lorsque l'on peut ainsi étudier une caractéristique que possède chacun des individus, on appelle cela une *variable statistique* ou *caractère*.
5. Les différentes valeurs que peut prendre une variable statistique sont les *modalités* de cette variable.
6. Le nombre d'individus vérifiant une modalité donnée est appelé *l'effectif*.
7. La *fréquence* d'une modalité est le rapport entre l'effectif et le nombre d'observations. On l'exprime souvent en pour cent.

Notation

On note une variable statistique par une lettre majuscule X, Y, ... et ses modalités par la même lettre minuscule affectée d'indices : x_1, x_2, ... pour la variable X ou y_1, y_2, ... pour la variable Y.

Exemple

On fait une étude auprès des étudiants de l'EPC.

Exemple

On fait une étude auprès des étudiants de l'EPC.On aimerait connaître le sexe, l'âge au 1^{er} janvier, la taille et le taux de satisfaction par rapport aux études (satisfait (S), insatisfait (I) et sans réponse (SR)) de chaque étudiant.

Exemple

On fait une étude auprès des étudiants de l'EPC.On aimerait connaître le sexe, l'âge au 1er janvier, la taille et le taux de satisfaction par rapport aux études (satisfait (S), insatisfait (I) et sans réponse (SR)) de chaque étudiant.
La population considérée est "les étudiants de l'EPC". Un échantillon est par exemple l'ensemble des étudiants en dernière année de formation. Tout étudiant en dernière année de formation est un individu.

Exemple

On fait une étude auprès des étudiants de l'EPC.On aimerait connaître le sexe, l'âge au 1er janvier, la taille et le taux de satisfaction par rapport aux études (satisfait (S), insatisfait (I) et sans réponse (SR)) de chaque étudiant.
La population considérée est "les étudiants de l'EPC". Un échantillon est par exemple l'ensemble des étudiants en dernière année de formation. Tout étudiant en dernière année de formation est un individu.

Variable statistique	Modalités
X : sexe	x_1 = homme, x_2 = femme
Y : âge	$y_1 = 18$, $y_2 = 19$, ...
Z : taille	$z_i \in [150; 200]$
U : taux de satisfaction	$u_1 = S$, $u_2 = I$, $u_3 = SR$

Definition

1. Une variable statistique est dite *qualitative*, respectivement *quantitative* si ses valeurs peuvent être comptées, respectivement qu'elles ne peuvent pas l'être.

Définitions

Definition

1. Une variable statistique est dite *qualitative*, respectivement *quantitative* si ses valeurs peuvent être comptées, respectivement qu'elles ne peuvent pas l'être.
2. On dit d'une variable statistique qualitative X qu'elle est *nominale*, respectivement *ordinale*, si ses valeurs ne possèdent pas d'ordre, respectivement si ses valeurs possèdent une certain ordre.

Definition

1. Une variable statistique est dite *qualitative*, respectivement *quantitative* si ses valeurs peuvent être comptées, respectivement qu'elles ne peuvent pas l'être.
2. On dit d'une variable statistique qualitative X qu'elle est *nominale*, respectivement *ordinale*, si ses valeurs ne possèdent pas d'ordre, respectivement si ses valeurs possèdent une certain ordre.
3. On dit d'une variable statiatique quantitative X qu'elle est *discrète*, respectivement *continue*, si l'ensemble des valeurs de celle-ci est fini ou infini dénombrable, respectivement infini non dénombrable.

Definition

1. Une variable statistique est dite *qualitative*, respectivement *quantitative* si ses valeurs peuvent être comptées, respectivement qu'elles ne peuvent pas l'être.
2. On dit d'une variable statistique qualitative X qu'elle est *nominale*, respectivement *ordinale*, si ses valeurs ne possèdent pas d'ordre, respectivement si ses valeurs possèdent une certain ordre.
3. On dit d'une variable statiatique quantitative X qu'elle est *discrète*, respectivement *continue*, si l'ensemble des valeurs de celle-ci est fini ou infini dénombrable, respectivement infini non dénombrable.

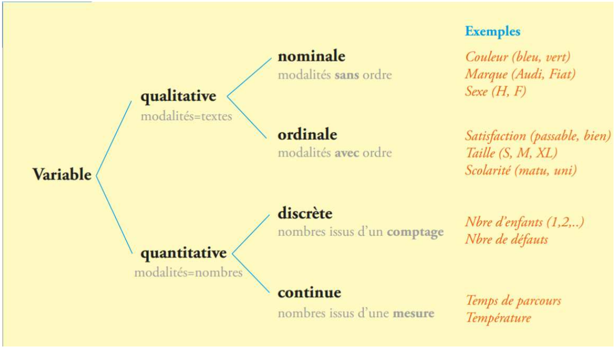


FIGURE – Caractérisation des différentes variables statistiques.

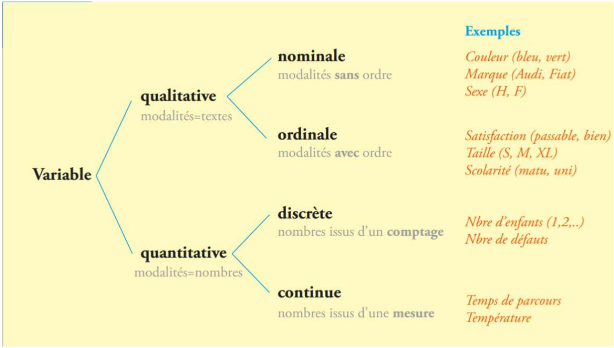


FIGURE – Caractérisation des différentes variables statistiques.

Exemple

Dans notre exemple précédent, X est une variable statistique qualitative nominale, Y est une une variable statistique quantitative discrète, Z est une une variable statistique quantitative continue et U est une une variable statistique qualitative ordinale.

Exemple

On étudie l'état civil des 40 employés d'une compagnie.

Exemple

On étudie l'état civil des 40 employés d'une compagnie.
Dans un premier temps, il s'agit de collecter l'information, dans ce cas l'état civil de chacun des individus de la population (les 40 employés de la compagnie) : les données brutes. La variable statistique est l'état civil. Elle est qualitative nominale et les modalités sont : marié(e), célibataire, divorcé(e) et veuf(ve).

Exemple

On étudie l'état civil des 40 employés d'une compagnie.
Dans un premier temps, il s'agit de collecter l'information, dans ce cas l'état civil de chacun des individus de la population (les 40 employés de la compagnie) : les données brutes. La variable statistique est l'état civil. Elle est qualitative nominale et les modalités sont : marié(e), célibataire, divorcé(e) et veuf(ve).
On donne l'état civil des employés identifiés par un numéro :

Exemple

On étudie l'état civil des 40 employés d'une compagnie.
Dans un premier temps, il s'agit de collecter l'information, dans ce cas l'état civil de chacun des individus de la population (les 40 employés de la compagnie) : les données brutes. La variable statistique est l'état civil. Elle est qualitative nominale et les modalités sont : marié(e), célibataire, divorcé(e) et veuf(ve).
On donne l'état civil des employés identifiés par un numéro :

1	Marié	11	Veuf	21	Célibataire	31	Célibataire
2	Mariée	12	Marié	22	Mariée	32	Divorcée
3	Célibataire	13	Célibataire	23	Marié	33	Divorcé
4	Divorcé	14	Célibataire	24	Marié	34	Marié
5	Marié	15	Mariée	25	Divorcée	35	Mariée
6	Célibataire	16	Célibataire	26	Mariée	36	Marié
7	Célibataire	17	Marié	27	Célibataire	37	Marié
8	Mariée	18	Veuve	28	Célibataire	38	Mariée
9	Marié	19	Marié	29	Marié	39	Célibataire
10	Divorcée	20	Divorcé	30	Veuf	40	Mariée

Exemple

On obtient avec ces données brutes une information personnalisée. On va sacrifier le caractère individuel de l'information pour obtenir un portrait d'ensemble.

Exemple

On obtient avec ces données brutes une information personnalisée. On va sacrifier le caractère individuel de l'information pour obtenir un portrait d'ensemble.On calcule pour chaque modalité le nombre d'individus ayant cette modalité.

Exemple

On obtient avec ces données brutes une information personnalisée. On va sacrifier le caractère individuel de l'information pour obtenir un portrait d'ensemble.On calcule pour chaque modalité le nombre d'individus ayant cette modalité.Il s'agit de l'effectif de la modalité :

Exemple

On obtient avec ces données brutes une information personnalisée. On va sacrifier le caractère individuel de l'information pour obtenir un portrait d'ensemble.On calcule pour chaque modalité le nombre d'individus ayant cette modalité.Il s'agit de l'effectif de la modalité :

20 individus mariés
11 individus célibataires
6 individus divorcés
3 individus veufs

Regroupement des données par modalités

Exemple

On obtient avec ces données brutes une information personnalisée. On va sacrifier le caractère individuel de l'information pour obtenir un portrait d'ensemble.On calcule pour chaque modalité le nombre d'individus ayant cette modalité.Il s'agit de l'effectif de la modalité :

20 individus mariés
11 individus célibataires
6 individus divorcés
3 individus veufs

Il est d'usage de présenter la distribution des effectifs sous la forme d'un tableau :

Exemple

On obtient avec ces données brutes une information personnalisée. On va sacrifier le caractère individuel de l'information pour obtenir un portrait d'ensemble.On calcule pour chaque modalité le nombre d'individus ayant cette modalité.Il s'agit de l'effectif de la modalité :

20 individus mariés
11 individus célibataires
6 individus divorcés
3 individus veufs

Il est d'usage de présenter la distribution des effectifs sous la forme d'un tableau :

Modalités	Effectifs	Fréquences
Mariés	20	50%
Célibataires	11	$27,5\%$
Divorcés	6	15%
Veufs	3	$7,5\%$
Total	40	100%

Exemple

On obtient avec ces données brutes une information personnalisée. On va sacrifier le caractère individuel de l'information pour obtenir un portrait d'ensemble.On calcule pour chaque modalité le nombre d'individus ayant cette modalité.Il s'agit de l'effectif de la modalité :

20 individus mariés
11 individus célibataires
6 individus divorcés
3 individus veufs

Il est d'usage de présenter la distribution des effectifs sous la forme d'un tableau :

Modalités	*Effectifs*	*Fréquences*
Mariés	*20*	50%
Célibataires	*11*	$27,5\%$
Divorcés	*6*	15%
Veufs	*3*	$7,5\%$
Total	*40*	100%

Pour trouver qu'il y a $27,5\%$ *de célibataires, il suffit de calculer*

$$\frac{11}{40} = 0,275 = 27,5\%.$$

Exemple

Dans un quartier composé de 50 ménages, on étudie le nombre de personnes par ménage.

Exemple

Dans un quartier composé de 50 ménages, on étudie le nombre de personnes par ménage.
Dans un premier temps, il s'agit de collecter les données brutes de chacun des individus de la population (les 50 ménages).

Exemple

Dans un quartier composé de 50 ménages, on étudie le nombre de personnes par ménage.
Dans un premier temps, il s'agit de collecter les données brutes de chacun des individus de la population (les 50 ménages).La variable statistique est le nombre de personnes par ménage. Elle est quantitative discrète et les modalités sont : 1, 2, 3, 4, 5, 6 et 8.

Exemple

Dans un quartier composé de 50 ménages, on étudie le nombre de personnes par ménage.
Dans un premier temps, il s'agit de collecter les données brutes de chacun des individus de la population (les 50 ménages).La variable statistique est le nombre de personnes par ménage. Elle est quantitative discrète et les modalités sont : 1, 2, 3, 4, 5, 6 et 8.
Les données brutes sont :

Exemple

Dans un quartier composé de 50 ménages, on étudie le nombre de personnes par ménage.
Dans un premier temps, il s'agit de collecter les données brutes de chacun des individus de la population (les 50 ménages).La variable statistique est le nombre de personnes par ménage. Elle est quantitative discrète et les modalités sont : 1, 2, 3, 4, 5, 6 et 8.
Les données brutes sont :

1	1	1	1	1	2	2	2	2	2
2	2	2	2	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	4
4	4	4	4	4	4	4	4	4	5
5	5	5	5	5	6	6	6	8	8

Exemple

On obtient avec ces données brutes une information personnalisée. Cette liste n'étant pas commode à lire, il convient à nouveau de sacrifier le caractère individuel de l'information pour obtenir un portrait d'ensemble. On va donc déterminer pour chaque modalité le nombre d'individus ayant cette modalité : l'effectif de la modalité.

Exemple

On obtient avec ces données brutes une information personnalisée. Cette liste n'étant pas commode à lire, il convient à nouveau de sacrifier le caractère individuel de l'information pour obtenir un portrait d'ensemble. On va donc déterminer pour chaque modalité le nombre d'individus ayant cette modalité : l'effectif de la modalité.

Modalités x_i	Effectifs n_i	Fréquences f_i
1	5	10%
2	9	18%
3	15	30%
4	10	20%
5	6	12%
6	3	6%
8	2	4%

Exemple

On obtient avec ces données brutes une information personnalisée. Cette liste n'étant pas commode à lire, il convient à nouveau de sacrifier le caractère individuel de l'information pour obtenir un portrait d'ensemble. On va donc déterminer pour chaque modalité le nombre d'individus ayant cette modalité : l'effectif de la modalité.

Modalités x_i	Effectifs n_i	Fréquences f_i
1	5	10%
2	9	18%
3	15	30%
4	10	20%
5	6	12%
6	3	6%
8	2	4%

Dans le tableau ci-dessus, on a $x_1 = 1$, $x_2 = 2$, *etc. Les* x_i *représentent le nombre de personnes par ménage. On a de plus* $n_1 = 5$, $n_2 = 9$, *etc. Les* n_i *indiquent le nombre de ménages comportant* x_i *personnes. Ainsi, on a par exemple 10 ménages comportant 4 personnes.*

Souvent, lors d'une étude statistique portant sur une variable statistique quantitative discrète ou continue, les données recueillies diffèrent à peu près toutes les unes des autres et sont étalées sur un large intervalle de valeurs. L'objectif de la statistique descriptive étant de résumer de la façon la plus adéquate possible cet ensemble de données, nous procédons alors à un regroupement de ces dernières à l'intérieur de *classes*, c'est-à-dire de sous-intervalles de valeurs. Les règles suivantes permettent de choisir judicieusement ces classes :

Souvent, lors d'une étude statistique portant sur une variable statistique quantitative discrète ou continue, les données recueillies diffèrent à peu près toutes les unes des autres et sont étalées sur un large intervalle de valeurs. L'objectif de la statistique descriptive étant de résumer de la façon la plus adéquate possible cet ensemble de données, nous procédons alors à un regroupement de ces dernières à l'intérieur de *classes*, c'est-à-dire de sous-intervalles de valeurs. Les règles suivantes permettent de choisir judicieusement ces classes :

- On fixe un nombre de classes entre 5 et 15. Le nombre de classes dépend de la taille de la population et il faut éviter, si possible, des fréquences de classes trop petites.

Souvent, lors d'une étude statistique portant sur une variable statistique quantitative discrète ou continue, les données recueillies diffèrent à peu près toutes les unes des autres et sont étalées sur un large intervalle de valeurs. L'objectif de la statistique descriptive étant de résumer de la façon la plus adéquate possible cet ensemble de données, nous procédons alors à un regroupement de ces dernières à l'intérieur de *classes*, c'est-à-dire de sous-intervalles de valeurs. Les règles suivantes permettent de choisir judicieusement ces classes :

- On fixe un nombre de classes entre 5 et 15. Le nombre de classes dépend de la taille de la population et il faut éviter, si possible, des fréquences de classes trop petites.
- Les intervalles sont du type $[b_{i-1}; b_i[$ ou $]b_{i-1}; b_i]$.

  b_{i-1} est la *borne inférieure* de la classe i ;

Souvent, lors d'une étude statistique portant sur une variable statistique quantitative discrète ou continue, les données recueillies diffèrent à peu près toutes les unes des autres et sont étalées sur un large intervalle de valeurs. L'objectif de la statistique descriptive étant de résumer de la façon la plus adéquate possible cet ensemble de données, nous procédons alors à un regroupement de ces dernières à l'intérieur de *classes*, c'est-à-dire de sous-intervalles de valeurs. Les règles suivantes permettent de choisir judicieusement ces classes :

- On fixe un nombre de classes entre 5 et 15. Le nombre de classes dépend de la taille de la population et il faut éviter, si possible, des fréquences de classes trop petites.
- Les intervalles sont du type $[b_{i-1}; b_i[$ ou $]b_{i-1}; b_i]$.

  b_{i-1} est la *borne inférieure* de la classe i ;
  b_i est la *borne supérieure* de la classe i ;

Représentation des données à l'intérieur des classes

Souvent, lors d'une étude statistique portant sur une variable statistique quantitative discrète ou continue, les données recueillies diffèrent à peu près toutes les unes des autres et sont étalées sur un large intervalle de valeurs. L'objectif de la statistique descriptive étant de résumer de la façon la plus adéquate possible cet ensemble de données, nous procédons alors à un regroupement de ces dernières à l'intérieur de *classes*, c'est-à-dire de sous-intervalles de valeurs. Les règles suivantes permettent de choisir judicieusement ces classes :

- On fixe un nombre de classes entre 5 et 15. Le nombre de classes dépend de la taille de la population et il faut éviter, si possible, des fréquences de classes trop petites.
- Les intervalles sont du type $[b_{i-1}; b_i[$ ou $]b_{i-1}; b_i]$.

  b_{i-1} est la *borne inférieure* de la classe i ;
  b_i est la *borne supérieure* de la classe i ;
  $c_i = \dfrac{b_{i-1} + b_1}{2}$ est le *centre* de la classe i ;

Souvent, lors d'une étude statistique portant sur une variable statistique quantitative discrète ou continue, les données recueillies diffèrent à peu près toutes les unes des autres et sont étalées sur un large intervalle de valeurs. L'objectif de la statistique descriptive étant de résumer de la façon la plus adéquate possible cet ensemble de données, nous procédons alors à un regroupement de ces dernières à l'intérieur de *classes*, c'est-à-dire de sous-intervalles de valeurs. Les règles suivantes permettent de choisir judicieusement ces classes :

- On fixe un nombre de classes entre 5 et 15. Le nombre de classes dépend de la taille de la population et il faut éviter, si possible, des fréquences de classes trop petites.
- Les intervalles sont du type $[b_{i-1}; b_i[$ ou $]b_{i-1}; b_i]$.

  b_{i-1} est la *borne inférieure* de la classe i ;
  b_i est la *borne supérieure* de la classe i ;
  $c_i = \dfrac{b_{i-1} + b_1}{2}$ est le *centre* de la classe i ;
  $L_i = b_i - b_{i-1}$ est la *largeur* (ou *étendue* ou *amplitude*) de la classe i.

Souvent, lors d'une étude statistique portant sur une variable statistique quantitative discrète ou continue, les données recueillies diffèrent à peu près toutes les unes des autres et sont étalées sur un large intervalle de valeurs. L'objectif de la statistique descriptive étant de résumer de la façon la plus adéquate possible cet ensemble de données, nous procédons alors à un regroupement de ces dernières à l'intérieur de *classes*, c'est-à-dire de sous-intervalles de valeurs. Les règles suivantes permettent de choisir judicieusement ces classes :

- On fixe un nombre de classes entre 5 et 15. Le nombre de classes dépend de la taille de la population et il faut éviter, si possible, des fréquences de classes trop petites.
- Les intervalles sont du type $[b_{i-1}; b_i[$ ou $]b_{i-1}; b_i]$.

  b_{i-1} est la *borne inférieure* de la classe i ;
  b_i est la *borne supérieure* de la classe i ;
  $c_i = \dfrac{b_{i-1} + b_1}{2}$ est le *centre* de la classe i ;
  $L_i = b_i - b_{i-1}$ est la *largeur* (ou *étendue* ou *amplitude*) de la classe i.
- En principe, on fixe les bornes des intervalles de telle sorte que ces derniers soient d'égales largeurs. Les bornes doivent permettre des calculs simples.

Souvent, lors d'une étude statistique portant sur une variable statistique quantitative discrète ou continue, les données recueillies diffèrent à peu près toutes les unes des autres et sont étalées sur un large intervalle de valeurs. L'objectif de la statistique descriptive étant de résumer de la façon la plus adéquate possible cet ensemble de données, nous procédons alors à un regroupement de ces dernières à l'intérieur de *classes*, c'est-à-dire de sous-intervalles de valeurs. Les règles suivantes permettent de choisir judicieusement ces classes :

- On fixe un nombre de classes entre 5 et 15. Le nombre de classes dépend de la taille de la population et il faut éviter, si possible, des fréquences de classes trop petites.
- Les intervalles sont du type $[b_{i-1}; b_i[$ ou $]b_{i-1}; b_i]$.

  b_{i-1} est la *borne inférieure* de la classe i ;
  b_i est la *borne supérieure* de la classe i ;
  $c_i = \dfrac{b_{i-1} + b_1}{2}$ est le *centre* de la classe i ;
  $L_i = b_i - b_{i-1}$ est la *largeur* (ou *étendue* ou *amplitude*) de la classe i.
- En principe, on fixe les bornes des intervalles de telle sorte que ces derniers soient d'égales largeurs. Les bornes doivent permettre des calculs simples.
- Si on doit vraiment utiliser des classes de largeurs inégales, on place les classes de largeur égale au centre de la distribution.

Exemple

Dans une région française, on étudie la superficie de chacune des 500 exploitations agricoles exprimées en hectares.

Exemple

Dans une région française, on étudie la superficie de chacune des 500 exploitations agricoles exprimées en hectares.
Dans cet exemple, la population est l'ensemble des exploitations agricoles d'une région française, tandis qu'un individu est ici une exploitation agricole donnée. La population étant définie, elle est observée selon certains critères. Le critère retenu, c'est-à-dire la variable statistique, est ici la superficie. Elle est de nature quantitative continue et les modalités sont des nombres représentant des superficies compris entre 0 ha et 40 ha.

Exemple

Dans une région française, on étudie la superficie de chacune des 500 exploitations agricoles exprimées en hectares.
Dans cet exemple, la population est l'ensemble des exploitations agricoles d'une région française, tandis qu'un individu est ici une exploitation agricole donnée. La population étant définie, elle est observée selon certains critères. Le critère retenu, c'est-à-dire la variable statistique, est ici la superficie. Elle est de nature quantitative continue et les modalités sont des nombres représentant des superficies compris entre 0 ha et 40 ha.
Les données brutes que l'on recueille sur cette population sont inutilisables telles quelles. En vue de synthétiser l'information, on procède à des regroupements, à des classements et à l'établissement de tableaux statistiques. Le tableau ci-dessous constitue déjà une première simplification de l'information complète contenue dans un registre officiel comportant une ligne pour chacune des 500 exploitations.

Exemple

Dans une région française, on étudie la superficie de chacune des 500 exploitations agricoles exprimées en hectares.
Dans cet exemple, la population est l'ensemble des exploitations agricoles d'une région française, tandis qu'un individu est ici une exploitation agricole donnée. La population étant définie, elle est observée selon certains critères. Le critère retenu, c'est-à-dire la variable statistique, est ici la superficie. Elle est de nature quantitative continue et les modalités sont des nombres représentant des superficies compris entre 0 ha et 40 ha. Les données brutes que l'on recueille sur cette population sont inutilisables telles quelles. En vue de synthétiser l'information, on procède à des regroupements, à des classements et à l'établissement de tableaux statistiques. Le tableau ci-dessous constitue déjà une première simplification de l'information complète contenue dans un registre officiel comportant une ligne pour chacune des 500 exploitations.

Superfice en ha	*Nombre d'exploitations*	*Fréquences en %*
$]0;10]$	*48*	$9,6$
$]10;15]$	*62*	$12,4$
$]15;20]$	*107*	$21,4$
$]20;25]$	*133*	$26,6$
$]25;30]$	*84*	$16,8$
$]30;40]$	*66*	$13,2$

Exemple

Les individus étant rassemblés par classes, on dit qu'on a affaire à un ensemble de données groupées.

Regroupement des données par modalités

Exemple

Les individus étant rassemblés par classes, on dit qu'on a affaire à un ensemble de données groupées.Ce qu'on gagne en simplicité par ce regroupement, on le perd en information.

Exemple

Les individus étant rassemblés par classes, on dit qu'on a affaire à un ensemble de données groupées.Ce qu'on gagne en simplicité par ce regroupement, on le perd en information.On sait par exemple que la classe $]20; 25]$ *comporte 133 exploitations dont les superficies sont comprises entre 20 et 25.*

Exemple

Les individus étant rassemblés par classes, on dit qu'on a affaire à un ensemble de données groupées.Ce qu'on gagne en simplicité par ce regroupement, on le perd en information.On sait par exemple que la classe $]20;25]$ *comporte 133 exploitations dont les superficies sont comprises entre 20 et 25.Mais on ne connaît rien de la répartition de ces 133 individus à l'intérieur de leur classe.*

Exemple

Les individus étant rassemblés par classes, on dit qu'on a affaire à un ensemble de données groupées.Ce qu'on gagne en simplicité par ce regroupement, on le perd en information.On sait par exemple que la classe $]20;25]$ *comporte 133 exploitations dont les superficies sont comprises entre 20 et 25.Mais on ne connaît rien de la répartition de ces 133 individus à l'intérieur de leur classe.Il est alors commode de formuler l'hypothèse d'une répartition uniforme au sein de chaque classe.*

Exemple

Les individus étant rassemblés par classes, on dit qu'on a affaire à un ensemble de données groupées.Ce qu'on gagne en simplicité par ce regroupement, on le perd en information.On sait par exemple que la classe $]20;25]$ *comporte 133 exploitations dont les superficies sont comprises entre 20 et 25.Mais on ne connaît rien de la répartition de ces 133 individus à l'intérieur de leur classe.Il est alors commode de formuler l'hypothèse d'une répartition uniforme au sein de chaque classe.On assigne ainsi à l'individu occupant la place* x *sur 133 dans la classe* $]20;25]$ *(d'étendue 5), la valeur* $20+\dfrac{x}{133}\cdot 5$.

Exemple

Les individus étant rassemblés par classes, on dit qu'on a affaire à un ensemble de données groupées.Ce qu'on gagne en simplicité par ce regroupement, on le perd en information.On sait par exemple que la classe $]20; 25]$ *comporte 133 exploitations dont les superficies sont comprises entre 20 et 25.Mais on ne connaît rien de la répartition de ces 133 individus à l'intérieur de leur classe.Il est alors commode de formuler l'hypothèse d'une répartition uniforme au sein de chaque classe.On assigne ainsi à l'individu occupant la place x sur 133 dans la classe* $]20; 25]$ *(d'étendue 5), la valeur* $20 + \frac{x}{133} \cdot 5$*.Avec cette convention, le dernier individu (le* $133^{\grave{e}me}$*) est affecté de la valeur 25, borne supérieure de l'intervalle.*

La répartition d'une population et sa distribution de fréquences sont parfois plus expressives sur le plan visuel lorsqu'on les représente à l'aide d'un *diagramme circulaire* (appelé également *diagramme en secteurs*) .

Diagramme circulaire

La répartition d'une population et sa distribution de fréquences sont parfois plus expressives sur le plan visuel lorsqu'on les représente à l'aide d'un *diagramme circulaire* (appelé également *diagramme en secteurs*) .Un diagramme circulaire consiste à représenter la population totale par un disque et à le diviser en tranches, de façon proportionnelle aux effectifs de la variable statistique considérée. On obtient ainsi une représentation graphique de la répartition relative de la population, autrement dit de la distribution de fréquences.

La répartition d'une population et sa distribution de fréquences sont parfois plus expressives sur le plan visuel lorsqu'on les représente à l'aide d'un *diagramme circulaire* (appelé également *diagramme en secteurs*) .Un diagramme circulaire consiste à représenter la population totale par un disque et à le diviser en tranches, de façon proportionnelle aux effectifs de la variable statistique considérée. On obtient ainsi une représentation graphique de la répartition relative de la population, autrement dit de la distribution de fréquences.

Exemple

Reprenons notre exemple des exploitations agricoles. Ce qui caractérise "la taille d'une tranche" est l'angle au centre. Pour le trouver, il suffit de faire une règle de trois avec la relation 360° *correspond à une fréquence de* 100% *ou, de manière équivalente, à un effectif de 500.*

La répartition d'une population et sa distribution de fréquences sont parfois plus expressives sur le plan visuel lorsqu'on les représente à l'aide d'un *diagramme circulaire* (appelé également *diagramme en secteurs*) .Un diagramme circulaire consiste à représenter la population totale par un disque et à le diviser en tranches, de façon proportionnelle aux effectifs de la variable statistique considérée. On obtient ainsi une représentation graphique de la répartition relative de la population, autrement dit de la distribution de fréquences.

Exemple

Reprenons notre exemple des exploitations agricoles. Ce qui caractérise "la taille d'une tranche" est l'angle au centre. Pour le trouver, il suffit de faire une règle de trois avec la relation 360° *correspond à une fréquence de* 100% *ou, de manière équivalente, à un effectif de 500.*

Superfice en ha	*Effectifs*	*Fréquences en %*	*Angles en °*
$]0;10]$	*48*	$9,6$	$34,56$
$]10;15]$	*62*	$12,4$	$44,64$
$]15;20]$	*107*	$21,4$	$77,04$
$]20;25]$	*133*	$26,6$	$95,76$
$]25;30]$	*84*	$16,8$	$60,48$
$]30;40]$	*66*	$13,2$	$47,52$

FIGURE – *Données avec angles.*

Exemple

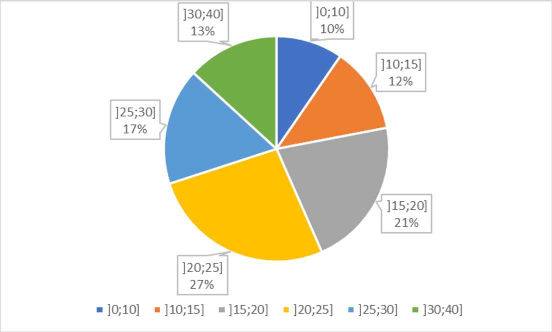


FIGURE – *Diagramme en secteurs.*

Exemple

Reprenons notre exemple relatif à l'état civil des employés d'une compagnie. Pour représenter le diagramme en secteurs, il convient de déterminer l'angle de chaque tranche.

Exemple

Reprenons notre exemple relatif à l'état civil des employés d'une compagnie. Pour représenter le diagramme en secteurs, il convient de déterminer l'angle de chaque tranche.

Etats civils	*Effectifs*	*Fréquences en %*	*Angles en °*
Mariés	*20*	50	180
Célibataires	*11*	$27,5$	99
Divorcés	*6*	15	54
Veufs	*3*	$7,5$	27

FIGURE – *Données avec angles.*

Diagramme circulaire

Exemple

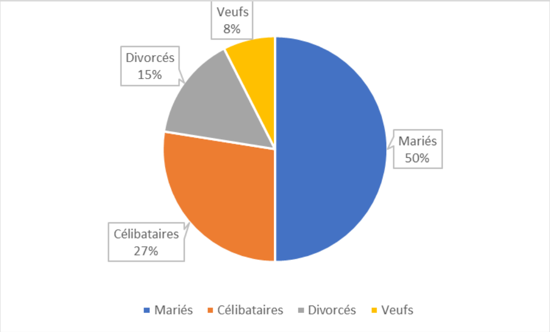


FIGURE – *Diagramme en secteurs.*

Diagramme en bâtons

Lorsque la variable statistique est quantitative discrète, la distribution des effectifs peut être représentée visuellement par un *diagramme en bâtons*. Il s'agit d'un diagramme dans lequel les modalités se trouvent sur l'axe horizontal et chaque bâton monte jusqu'à hauteur de l'effectif (ou de la fréquence) correspondant(e).

Diagramme en bâtons

Lorsque la variable statistique est quantitative discrète, la distribution des effectifs peut être représentée visuellement par un *diagramme en bâtons*. Il s'agit d'un diagramme dans lequel les modalités se trouvent sur l'axe horizontal et chaque bâton monte jusqu'à hauteur de l'effectif (ou de la fréquence) correspondant(e).

Exemple

Reprenons notre exemple relatif à l'état civil des employés d'une compagnie. Le diagramme en bâtons de cette distribution est représenté ci-dessous.

Lorsque la variable statistique est quantitative discrète, la distribution des effectifs peut être représentée visuellement par un *diagramme en bâtons*. Il s'agit d'un diagramme dans lequel les modalités se trouvent sur l'axe horizontal et chaque bâton monte jusqu'à hauteur de l'effectif (ou de la fréquence) correspondant(e).

Exemple

Reprenons notre exemple relatif à l'état civil des employés d'une compagnie. Le diagramme en bâtons de cette distribution est représenté ci-dessous.

Modalités	*Effectifs*
Mariés	*20*
Célibataires	*11*
Divorcés	*6*
Veufs	*3*
Total	*40*

FIGURE – *Données.*

Exemple

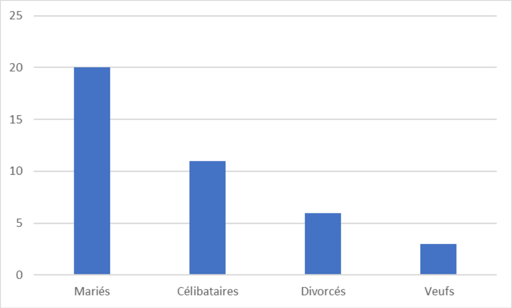


FIGURE – *Diagramme en bâtons.*

Exemple

Reprenons notre exemple relatif au nombre de personnes par ménage.

Diagramme en bâtons

Exemple

Reprenons notre exemple relatif au nombre de personnes par ménage.

Modalités	*Effectifs*
1	*5*
2	*9*
3	*15*
4	*10*
5	*6*
6	*3*
7	*0*
8	*2*

FIGURE – *Données.*

Exemple

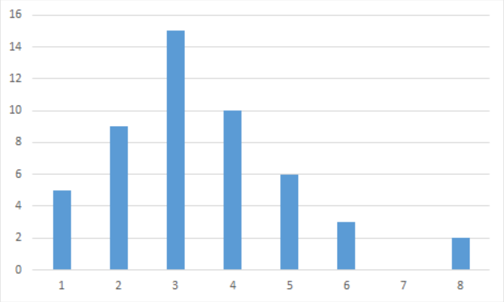


FIGURE – *Diagramme en bâtons.*

Lorsque la variable statistique est quantitative continue ou discrète, mais que les données sont regroupées en classes, la distribution peut être représentée visuellement par un *histogramme*, qui est un diagramme en colonnes où les rectangles sont juxtaposés. En effet, les modalités sont ici remplacées par des classes et celles-ci sont formées d'intervalles successifs de sorte qu'il n'y a plus lieu de séparer ces rectangles.

Histogramme

Lorsque la variable statistique est quantitative continue ou discrète, mais que les données sont regroupées en classes, la distribution peut être représentée visuellement par un *histogramme*, qui est un diagramme en colonnes où les rectangles sont juxtaposés. En effet, les modalités sont ici remplacées par des classes et celles-ci sont formées d'intervalles successifs de sorte qu'il n'y a plus lieu de séparer ces rectangles.

Exemple

Dans notre exemple, les classes de superficie n'ont pas toutes la même amplitude.

Lorsque la variable statistique est quantitative continue ou discrète, mais que les données sont regroupées en classes, la distribution peut être représentée visuellement par un *histogramme*, qui est un diagramme en colonnes où les rectangles sont juxtaposés. En effet, les modalités sont ici remplacées par des classes et celles-ci sont formées d'intervalles successifs de sorte qu'il n'y a plus lieu de séparer ces rectangles.

Exemple

Dans notre exemple, les classes de superficie n'ont pas toutes la même amplitude.Certaines classes ont une amplitude de 10 ha, d'autres 5 ha.

Histogramme

Lorsque la variable statistique est quantitative continue ou discrète, mais que les données sont regroupées en classes, la distribution peut être représentée visuellement par un *histogramme*, qui est un diagramme en colonnes où les rectangles sont juxtaposés. En effet, les modalités sont ici remplacées par des classes et celles-ci sont formées d'intervalles successifs de sorte qu'il n'y a plus lieu de séparer ces rectangles.

Exemple

Dans notre exemple, les classes de superficie n'ont pas toutes la même amplitude.Certaines classes ont une amplitude de 10 ha, d'autres 5 ha.Pour être fidèle, une représentation graphique doit tenir compte de ces différences.

Lorsque la variable statistique est quantitative continue ou discrète, mais que les données sont regroupées en classes, la distribution peut être représentée visuellement par un *histogramme*, qui est un diagramme en colonnes où les rectangles sont juxtaposés. En effet, les modalités sont ici remplacées par des classes et celles-ci sont formées d'intervalles successifs de sorte qu'il n'y a plus lieu de séparer ces rectangles.

Exemple

Dans notre exemple, les classes de superficie n'ont pas toutes la même amplitude.Certaines classes ont une amplitude de 10 ha, d'autres 5 ha.Pour être fidèle, une représentation graphique doit tenir compte de ces différences.Si, dans un histogramme, on représente les classes par des rectangles, alors, la surface totale représentant l'ensemble de la population, il faut que chaque rectangle ait une aire proportionnelle à l'effectif de la classe que ce dernier représente.

Exemple

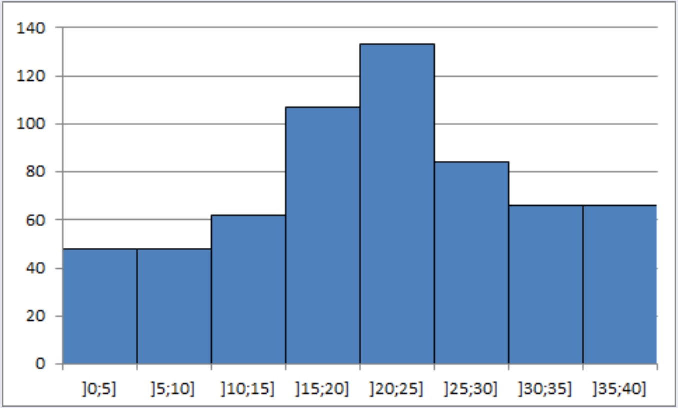


FIGURE – *Histogramme trompeur.*

Exemple

L'histogramme représenté ci-dessus est trompeur dans la mesure où il donne l'impression erronée que la classe initiale $[0; 10[$ *contient 96 exploitations : 48 d'une surface de 0 à 5 ha et le même nombre d'une surface de 5 à 10 ha.*

Exemple

L'histogramme représenté ci-dessus est trompeur dans la mesure où il donne l'impression erronée que la classe initiale $[0; 10[$ *contient 96 exploitations : 48 d'une surface de 0 à 5 ha et le même nombre d'une surface de 5 à 10 ha.Pour éviter cette déformation, il y a lieu de choisir une amplitude de référence (par exemple 5 ha) et de procéder à une correction des effectifs.*

Exemple

L'histogramme représenté ci-dessus est trompeur dans la mesure où il donne l'impression erronée que la classe initiale $[0; 10[$ *contient 96 exploitations : 48 d'une surface de 0 à 5 ha et le même nombre d'une surface de 5 à 10 ha.Pour éviter cette déformation, il y a lieu de choisir une amplitude de référence (par exemple 5 ha) et de procéder à une correction des effectifs.*
Avec cette correction, on obtient alors le tableau et l'histogramme correspondant suivants.

Exemple

L'histogramme représenté ci-dessus est trompeur dans la mesure où il donne l'impression erronée que la classe initiale $[0;10[$ *contient 96 exploitations : 48 d'une surface de 0 à 5 ha et le même nombre d'une surface de 5 à 10 ha.Pour éviter cette déformation, il y a lieu de choisir une amplitude de référence (par exemple 5 ha) et de procéder à une correction des effectifs.*
Avec cette correction, on obtient alors le tableau et l'histogramme correspondant suivants.

Superfice en ha	*Nombre d'exploitations*
$]0;5]$	$\boxed{24}$
$]5;10]$	$\boxed{24}$
$]10;15]$	*62*
$]15;20]$	*107*
$]20;25]$	*133*
$]25;30]$	*84*
$]30;35]$	$\boxed{33}$
$]35;40]$	$\boxed{33}$

FIGURE – *Effectifs corrigés.*

Exemple

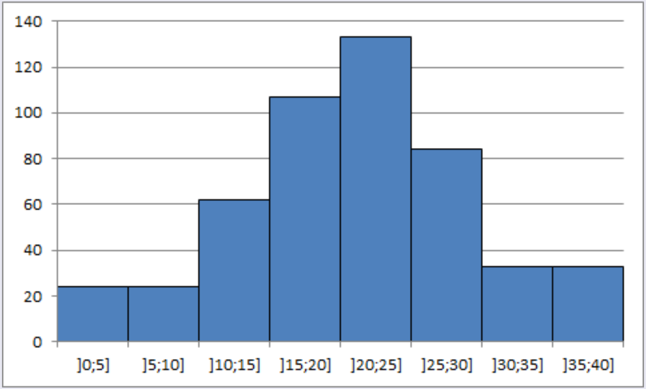


FIGURE – *Histogramme correct.*

Algorithme de correction des effectifs

1. On choisit une classe de référence de largeur l (en général la plus fréquente).

Algorithme de correction des effectifs

1. On choisit une classe de référence de largeur l (en général la plus fréquente).
2. Pour une classe quelconque de largeur L et d'effectif E, on calcule le rapport $x = \dfrac{E}{L}$.

Algorithme de correction des effectifs

1. On choisit une classe de référence de largeur l (en général la plus fréquente).
2. Pour une classe quelconque de largeur L et d'effectif E, on calcule le rapport $x = \frac{E}{L}$.
3. On attribue alors à cette classe l'effectif corrigé $c = x \cdot l = \frac{E}{L} \cdot l$. Notons que cet effectif n'est pas forcément un nombre entier.

Algorithme de correction des effectifs

1. On choisit une classe de référence de largeur l (en général la plus fréquente).
2. Pour une classe quelconque de largeur L et d'effectif E, on calcule le rapport $x = \frac{E}{L}$.
3. On attribue alors à cette classe l'effectif corrigé $c = x \cdot l = \frac{E}{L} \cdot l$. Notons que cet effectif n'est pas forcément un nombre entier.

Exemple

Dans notre exemple, la classe de référence ayant pour largeur $l = 5$*, la classe* $]0;10]$ *a pour largeur* $L = 10$*, on calcule* $x = \frac{E}{L} = \frac{48}{10} = 4,8$*,*

Algorithme de correction des effectifs

1. On choisit une classe de référence de largeur l (en général la plus fréquente).
2. Pour une classe quelconque de largeur L et d'effectif E, on calcule le rapport $x = \frac{E}{L}$.
3. On attribue alors à cette classe l'effectif corrigé $c = x \cdot l = \frac{E}{L} \cdot l$. Notons que cet effectif n'est pas forcément un nombre entier.

Exemple

Dans notre exemple, la classe de référence ayant pour largeur $l = 5$*, la classe* $]0; 10]$ *a pour largeur* $L = 10$*, on calcule* $x = \frac{E}{L} = \frac{48}{10} = 4,8$*,ce qui conduit à l'effectif corrigé* $c = x \cdot l = 4,8 \cdot 5 = 24$*.*

Diagrammes trompeurs ou faux

Dans la presse, à la télévision ou dans des tracts à caractère politique, il n'est pas rare d'y découvrir des diagrammes ou des graphes déformant la réalité, voire complètement faux. Le but de cette section consiste à mettre en avant les techniques utilisées pour déformer la réalité au travers de quelques exemples.

Exemple

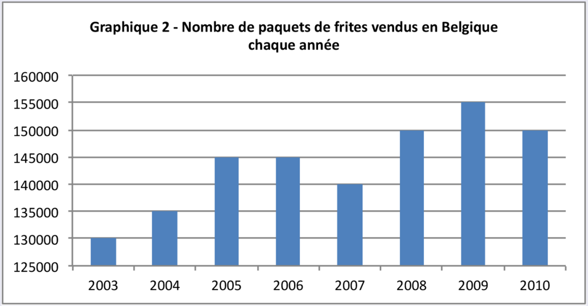
Dans cet exemple relatif à l'évolution du nombre de paquets de frites vendus en Belgique, nous allons voir comment présenter les données pour donner trois messages radicalement différents.

Exemple

Dans cet exemple relatif à l'évolution du nombre de paquets de frites vendus en Belgique, nous allons voir comment présenter les données pour donner trois messages radicalement différents.
En dépit des chiffres de 2007, le diagramme en bâtons ci-dessous semble indiquer une augmentation des ventes du nombre de paquets de frites.

Exemple

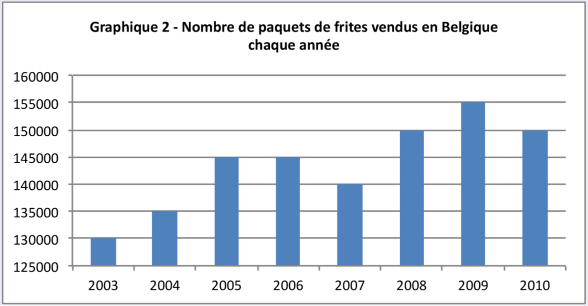
Dans cet exemple relatif à l'évolution du nombre de paquets de frites vendus en Belgique, nous allons voir comment présenter les données pour donner trois messages radicalement différents.
En dépit des chiffres de 2007, le diagramme en bâtons ci-dessous semble indiquer une augmentation des ventes du nombre de paquets de frites.



Cependant, on y regardant de plus près, on observe que l'axe des y ne part pas à 0, mais à 125'000 !

Exemple

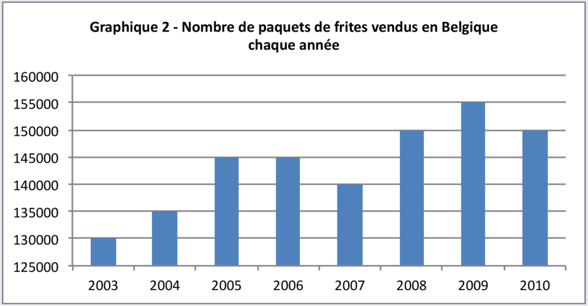
Dans cet exemple relatif à l'évolution du nombre de paquets de frites vendus en Belgique, nous allons voir comment présenter les données pour donner trois messages radicalement différents.
En dépit des chiffres de 2007, le diagramme en bâtons ci-dessous semble indiquer une augmentation des ventes du nombre de paquets de frites.



Cependant, on y regardant de plus près, on observe que l'axe des y ne part pas à 0, mais à 125'000 !Sur ce diagramme, on peut y lire que 130'000 paquets de frites ont été vendus en 2003, contre 135'000 en 2004, ce qui correspond à une augmentation de 5'000 paquets en une année, soit d'environ $3,85\%$.

Exemple

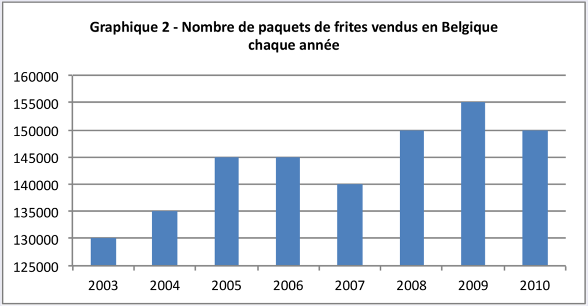
Dans cet exemple relatif à l'évolution du nombre de paquets de frites vendus en Belgique, nous allons voir comment présenter les données pour donner trois messages radicalement différents.
En dépit des chiffres de 2007, le diagramme en bâtons ci-dessous semble indiquer une augmentation des ventes du nombre de paquets de frites.



Cependant, on y regardant de plus près, on observe que l'axe des y ne part pas à 0, mais à 125'000 !Sur ce diagramme, on peut y lire que 130'000 paquets de frites ont été vendus en 2003, contre 135'000 en 2004, ce qui correspond à une augmentation de 5'000 paquets en une année, soit d'environ $3,85\%$*.Or, l'effet visuel du diagramme laisse supposer au premier abord que les ventes ont doublé durant cette période,*

Exemple

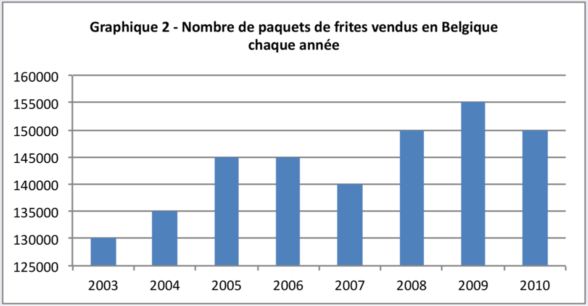
Dans cet exemple relatif à l'évolution du nombre de paquets de frites vendus en Belgique, nous allons voir comment présenter les données pour donner trois messages radicalement différents.
En dépit des chiffres de 2007, le diagramme en bâtons ci-dessous semble indiquer une augmentation des ventes du nombre de paquets de frites.



Cependant, on y regardant de plus près, on observe que l'axe des y ne part pas à 0, mais à 125'000 !Sur ce diagramme, on peut y lire que 130'000 paquets de frites ont été vendus en 2003, contre 135'000 en 2004, ce qui correspond à une augmentation de 5'000 paquets en une année, soit d'environ $3,85\%$*.Or, l'effet visuel du diagramme laisse supposer au premier abord que les ventes ont doublé durant cette période,*

Exemple

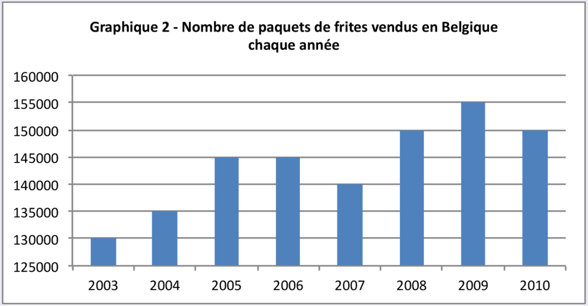
Dans cet exemple relatif à l'évolution du nombre de paquets de frites vendus en Belgique, nous allons voir comment présenter les données pour donner trois messages radicalement différents.
En dépit des chiffres de 2007, le diagramme en bâtons ci-dessous semble indiquer une augmentation des ventes du nombre de paquets de frites.



Cependant, on y regardant de plus près, on observe que l'axe des y ne part pas à 0, mais à 125'000 !Sur ce diagramme, on peut y lire que 130'000 paquets de frites ont été vendus en 2003, contre 135'000 en 2004, ce qui correspond à une augmentation de 5'000 paquets en une année, soit d'environ $3,85\%$*.Or, l'effet visuel du diagramme laisse supposer au premier abord que les ventes ont doublé durant cette période,*

Exemple

Dans cet exemple relatif à l'évolution du nombre de paquets de frites vendus en Belgique, nous allons voir comment présenter les données pour donner trois messages radicalement différents.
En dépit des chiffres de 2007, le diagramme en bâtons ci-dessous semble indiquer une augmentation des ventes du nombre de paquets de frites.



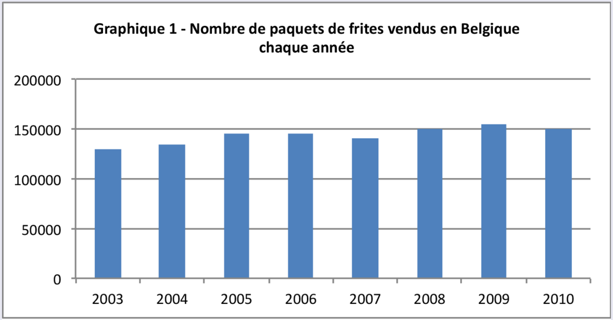
Cependant, on y regardant de plus près, on observe que l'axe des y *ne part pas à 0, mais à 125'000 !Sur ce diagramme, on peut y lire que 130'000 paquets de frites ont été vendus en 2003, contre 135'000 en 2004, ce qui correspond à une augmentation de 5'000 paquets en une année, soit d'environ* $3,85\%$*.Or, l'effet visuel du diagramme laisse supposer au premier abord que les ventes ont doublé durant cette période,*

Exemple

En présentant les mêmes données, mais en faisant partir l'axe des y de l'origine, le diagramme en bâtons ci-dessous semble indiquer une tendance de la vente des paquets de frites plutôt stable.

Exemple

En présentant les mêmes données, mais en faisant partir l'axe des y de l'origine, le diagramme en bâtons ci-dessous semble indiquer une tendance de la vente des paquets de frites plutôt stable.



Exemple

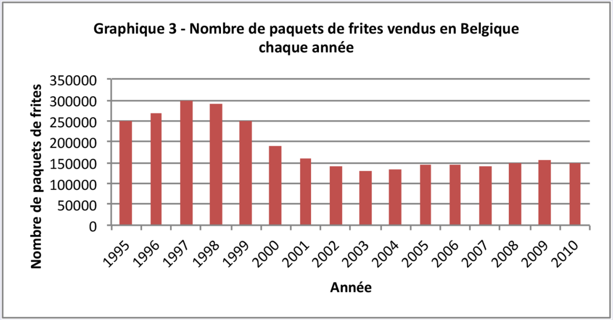
Les deux diagrammes ci-dessus contiennent uniquement les chiffres des ventes entre 2003 et 2010. Qu'en est-il si on considère les chiffres des années précédent 2003 ?

Exemple

Les deux diagrammes ci-dessus contiennent uniquement les chiffres des ventes entre 2003 et 2010. Qu'en est-il si on considère les chiffres des années précédent 2003 ? Le diagramme ci-dessous présente l'évolution du nombre de paquets de frites entre 1995 et 2010. En tenant compte de ces chiffres, il semble que les ventes de paquets de frites ont tendance à diminuer !

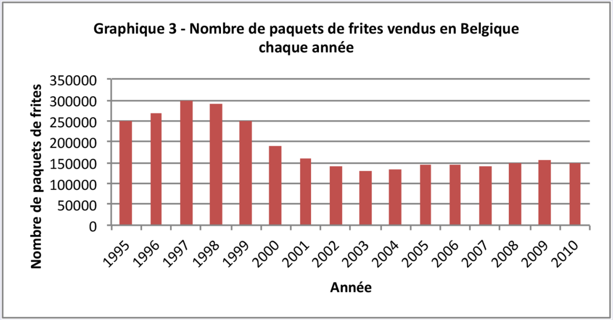
Exemple

Les deux diagrammes ci-dessus contiennent uniquement les chiffres des ventes entre 2003 et 2010. Qu'en est-il si on considère les chiffres des années précédent 2003 ?
Le diagramme ci-dessous présente l'évolution du nombre de paquets de frites entre 1995 et 2010. En tenant compte de ces chiffres, il semble que les ventes de paquets de frites ont tendance à diminuer !



Exemple

Les deux diagrammes ci-dessus contiennent uniquement les chiffres des ventes entre 2003 et 2010. Qu'en est-il si on considère les chiffres des années précédent 2003 ? Le diagramme ci-dessous présente l'évolution du nombre de paquets de frites entre 1995 et 2010. En tenant compte de ces chiffres, il semble que les ventes de paquets de frites ont tendance à diminuer !



Avec une même étude, il est donc possible de faire passer trois messages complétement différents selon la manière dont on présente l'information.

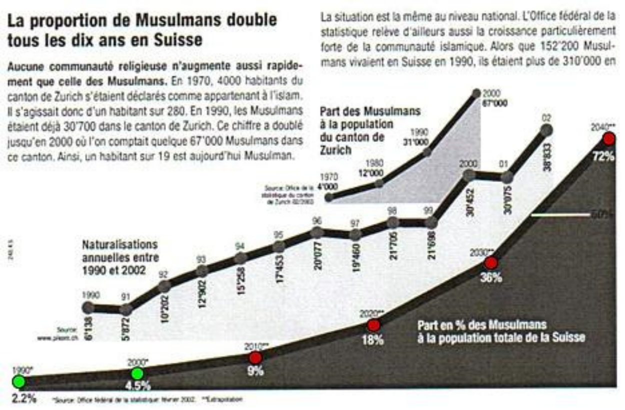
Exemple

En vue des votations du 26 septembre 2004, un comité proche d'un parti politique publie le document suivant.

Diagrammes trompeurs ou faux

Exemple

En vue des votations du 26 septembre 2004, un comité proche d'un parti politique publie le document suivant.



La proportion de Musulmans double tous les dix ans en Suisse

Aucune communauté religieuse n'augmente aussi rapidement que celle des Musulmans. En 1970, 4000 habitants du canton de Zurich s'étaient déclarés comme appartenant à l'islam. Il s'agissait donc d'un habitant sur 280. En 1990, les Musulmans étaient déjà 30'700 dans le canton de Zurich. Ce chiffre a doublé jusqu'en 2000 où l'on comptait quelque 67'000 Musulmans dans ce canton. Ainsi, un habitant sur 19 est aujourd'hui Musulman.

La situation est la même au niveau national. L'Office fédéral de la statistique relève d'ailleurs aussi la croissance particulièrement forte de la communauté islamique. Alors que 152'200 Musulmans vivaient en Suisse en 1990, ils étaient plus de 310'000 en

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.Or, en y regardant de plus près, on observe que les chiffres de 1990 et de 2000 ($2,2\%$ et $4,5\%$) sont munis d'une étoile indiquant qu'ils proviennent de l'Office fédéral de la statistique.

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.Or, en y regardant de plus près, on observe que les chiffres de 1990 et de 2000 ($2,2\%$ *et* $4,5\%$*) sont munis d'une étoile indiquant qu'ils proviennent de l'Office fédéral de la statistique.Les chiffres suivants (à partir de 2010) sont quant à eux munis de deux étoiles, pour préciser qu'il s'agit d'une extrapolation.*

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.Or, en y regardant de plus près, on observe que les chiffres de 1990 et de 2000 ($2,2\%$ *et* $4,5\%$*) sont munis d'une étoile indiquant qu'ils proviennent de l'Office fédéral de la statistique.Les chiffres suivants (à partir de 2010) sont quant à eux munis de deux étoiles, pour préciser qu'il s'agit d'une extrapolation.*
Mais comment arriver à un tel pronostic ? On observe que $4,5\%$ *représente à peu près le double de* $2,2\%$.

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.Or, en y regardant de plus près, on observe que les chiffres de 1990 et de 2000 ($2,2\%$ *et* $4,5\%$*) sont munis d'une étoile indiquant qu'ils proviennent de l'Office fédéral de la statistique.Les chiffres suivants (à partir de 2010) sont quant à eux munis de deux étoiles, pour préciser qu'il s'agit d'une extrapolation.*

Mais comment arriver à un tel pronostic ? On observe que $4,5\%$ *représente à peu près le double de* $2,2\%$*.Avec ces deux seules valeurs, on en conclut que le pourcentage de la communauté musulmane de Suisse double tous les 10 ans pour atteindre ainsi* 72% *en 2040, soit le dernier point représenté sur le graphe.*

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.Or, en y regardant de plus près, on observe que les chiffres de 1990 et de 2000 ($2,2\%$ et $4,5\%$) sont munis d'une étoile indiquant qu'ils proviennent de l'Office fédéral de la statistique.Les chiffres suivants (à partir de 2010) sont quant à eux munis de deux étoiles, pour préciser qu'il s'agit d'une extrapolation.
Mais comment arriver à un tel pronostic ? On observe que $4,5\%$ représente à peu près le double de $2,2\%$.Avec ces deux seules valeurs, on en conclut que le pourcentage de la communauté musulmane de Suisse double tous les 10 ans pour atteindre ainsi 72% en 2040, soit le dernier point représenté sur le graphe.On comprend mieux pourquoi le graphe s'arrête à ce point.

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.Or, en y regardant de plus près, on observe que les chiffres de 1990 et de 2000 ($2,2\%$ *et* $4,5\%$*) sont munis d'une étoile indiquant qu'ils proviennent de l'Office fédéral de la statistique.Les chiffres suivants (à partir de 2010) sont quant à eux munis de deux étoiles, pour préciser qu'il s'agit d'une extrapolation.*
Mais comment arriver à un tel pronostic ? On observe que $4,5\%$ *représente à peu près le double de* $2,2\%$*.Avec ces deux seules valeurs, on en conclut que le pourcentage de la communauté musulmane de Suisse double tous les 10 ans pour atteindre ainsi* 72% *en 2040, soit le dernier point représenté sur le graphe.On comprend mieux pourquoi le graphe s'arrête à ce point.En effet, le suivant indiquerait que le taux de musulmans s'élèverait à* 144% *en 2050 !*

Diagrammes trompeurs ou faux

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.Or, en y regardant de plus près, on observe que les chiffres de 1990 et de 2000 ($2,2\%$ *et* $4,5\%$*) sont munis d'une étoile indiquant qu'ils proviennent de l'Office fédéral de la statistique.Les chiffres suivants (à partir de 2010) sont quant à eux munis de deux étoiles, pour préciser qu'il s'agit d'une extrapolation.*
Mais comment arriver à un tel pronostic ? On observe que $4,5\%$ *représente à peu près le double de* $2,2\%$*.Avec ces deux seules valeurs, on en conclut que le pourcentage de la communauté musulmane de Suisse double tous les 10 ans pour atteindre ainsi* 72% *en 2040, soit le dernier point représenté sur le graphe.On comprend mieux pourquoi le graphe s'arrête à ce point.En effet, le suivant indiquerait que le taux de musulmans s'élèverait à* 144% *en 2050 !Cette projection est donc basée sur un doublement arbitraire, mais renforcée par les chiffres zurichois, qui indiquent une forte progression entre 1970 et 2000.*

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.Or, en y regardant de plus près, on observe que les chiffres de 1990 et de 2000 ($2,2\%$ *et* $4,5\%$*) sont munis d'une étoile indiquant qu'ils proviennent de l'Office fédéral de la statistique.Les chiffres suivants (à partir de 2010) sont quant à eux munis de deux étoiles, pour préciser qu'il s'agit d'une extrapolation.*
Mais comment arriver à un tel pronostic ? On observe que $4,5\%$ *représente à peu près le double de* $2,2\%$*.Avec ces deux seules valeurs, on en conclut que le pourcentage de la communauté musulmane de Suisse double tous les 10 ans pour atteindre ainsi* 72% *en 2040, soit le dernier point représenté sur le graphe.On comprend mieux pourquoi le graphe s'arrête à ce point.En effet, le suivant indiquerait que le taux de musulmans s'élèverait à* 144% *en 2050 !Cette projection est donc basée sur un doublement arbitraire, mais renforcée par les chiffres zurichois, qui indiquent une forte progression entre 1970 et 2000.Cette extrapolation basée sur un seul canton ne donne aucune raison de penser que ces chiffres sont transposables à toute la Suisse.*

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.Or, en y regardant de plus près, on observe que les chiffres de 1990 et de 2000 ($2,2\%$ *et* $4,5\%$*) sont munis d'une étoile indiquant qu'ils proviennent de l'Office fédéral de la statistique.Les chiffres suivants (à partir de 2010) sont quant à eux munis de deux étoiles, pour préciser qu'il s'agit d'une extrapolation.*
Mais comment arriver à un tel pronostic ? On observe que $4,5\%$ *représente à peu près le double de* $2,2\%$*.Avec ces deux seules valeurs, on en conclut que le pourcentage de la communauté musulmane de Suisse double tous les 10 ans pour atteindre ainsi* 72% *en 2040, soit le dernier point représenté sur le graphe.On comprend mieux pourquoi le graphe s'arrête à ce point.En effet, le suivant indiquerait que le taux de musulmans s'élèverait à* 144% *en 2050 !Cette projection est donc basée sur un doublement arbitraire, mais renforcée par les chiffres zurichois, qui indiquent une forte progression entre 1970 et 2000.Cette extrapolation basée sur un seul canton ne donne aucune raison de penser que ces chiffres sont transposables à toute la Suisse.Notons enfin, que selon l'OFS, il y avait* $4,9\%$ *de musulmans en Suisse en 2011 et* $5,3\%$ *en 2018.*

Exemple

Le texte explique qu'aucune communauté religieuse n'augmente autant rapidement que les musulmans. La courbe ci-dessus semble en effet indiquer que la croissance du nombre de musulmans en Suisse est exponentielle.Or, en y regardant de plus près, on observe que les chiffres de 1990 et de 2000 ($2,2\%$ *et* $4,5\%$*) sont munis d'une étoile indiquant qu'ils proviennent de l'Office fédéral de la statistique.Les chiffres suivants (à partir de 2010) sont quant à eux munis de deux étoiles, pour préciser qu'il s'agit d'une extrapolation.*
Mais comment arriver à un tel pronostic ? On observe que $4,5\%$ *représente à peu près le double de* $2,2\%$*.Avec ces deux seules valeurs, on en conclut que le pourcentage de la communauté musulmane de Suisse double tous les 10 ans pour atteindre ainsi* 72% *en 2040, soit le dernier point représenté sur le graphe.On comprend mieux pourquoi le graphe s'arrête à ce point.En effet, le suivant indiquerait que le taux de musulmans s'élèverait à* 144% *en 2050 !Cette projection est donc basée sur un doublement arbitraire, mais renforcée par les chiffres zurichois, qui indiquent une forte progression entre 1970 et 2000.Cette extrapolation basée sur un seul canton ne donne aucune raison de penser que ces chiffres sont transposables à toute la Suisse.Notons enfin, que selon l'OFS, il y avait* $4,9\%$ *de musulmans en Suisse en 2011 et* $5,3\%$ *en 2018.Soit des valeurs bien différentes des* 9% *et* 18% *prédites par les auteurs du document ci-dessus !*

Exemple

Quant à l'affiche ci-dessous, elle contient un certain nombre d'éléments forts discutables. Jörg Mäder, conseiller national zurichois depuis 2019, décortique les nombreux éléments controversés de cette affiche sur cette vidéo.

Exemple

Quant à l'affiche ci-dessous, elle contient un certain nombre d'éléments forts discutables. Jörg Mäder, conseiller national zurichois depuis 2019, décortique les nombreux éléments controversés de cette affiche sur cette vidéo.

Exemple

Le graphique publié par un quotidien en août 2008 (ci-dessous, à gauche) semble montrer que la consommation de viande s'est stabilisée ces dernières années.

Exemple

Le graphique publié par un quotidien en août 2008 (ci-dessous, à gauche) semble montrer que la consommation de viande s'est stabilisée ces dernières années.Cependant, on y regardant de plus près, on observe que que l'axe horizontal du graphique n'est pas linéaire :

Exemple

Le graphique publié par un quotidien en août 2008 (ci-dessous, à gauche) semble montrer que la consommation de viande s'est stabilisée ces dernières années.Cependant, on y regardant de plus près, on observe que que l'axe horizontal du graphique n'est pas linéaire :la moitié du graphique représente 50 ans, alors que l'autre moitié (la partie stable) ne concerne que 7 ans, donnant ainsi une impression erronée de la situation. Le graphique de droite montre les même données de façon correcte, et donne une impression différente.

Exemple

Le graphique publié par un quotidien en août 2008 (ci-dessous, à gauche) semble montrer que la consommation de viande s'est stabilisée ces dernières années.Cependant, on y regardant de plus près, on observe que que l'axe horizontal du graphique n'est pas linéaire :la moitié du graphique représente 50 ans, alors que l'autre moitié (la partie stable) ne concerne que 7 ans, donnant ainsi une impression erronée de la situation. Le graphique de droite montre les même données de façon correcte, et donne une impression différente.

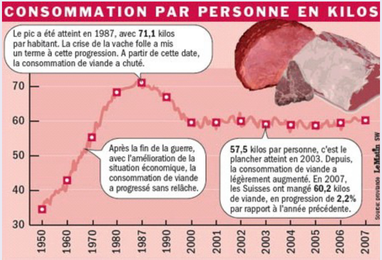


FIGURE – *Graphe faux.*

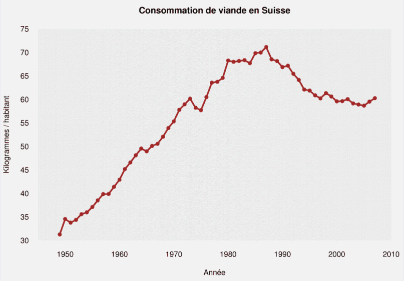


FIGURE – *Graphe correct.*

Exemple

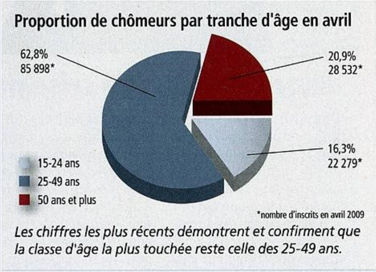
On peut voir une autre différence entre les deux graphiques : celui de gauche indique des variations à l'intérieur des années. En fait, il apparaît que ces variations ont été ajoutées pour éviter que le graphique ne soit trop lisse. On peut s'étonner que de telles considérations purement esthétiques prennent le pas sur le traitement correct de l'information.

Exemple

Le diagramme circulaire ci-dessous représente la proportion de chômeurs par tranche d'âge.

Exemple

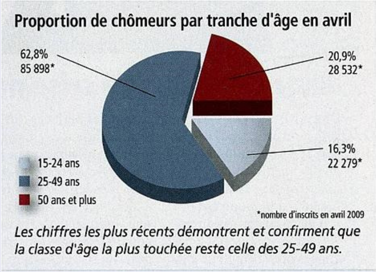
Le diagramme circulaire ci-dessous représente la proportion de chômeurs par tranche d'âge.



La légende conclut que la classe la plus touchée est celle des 25 à 49 ans. Les trois classes étant d'amplitudes différentes, il est difficile d'établir des comparaisons. Il n'est en effet pas surprenant que le plus grand nombre de chômeurs se trouve dans la classe la plus peuplée !

Exemple

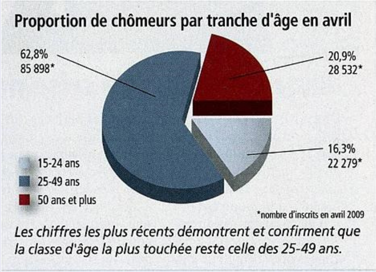
Le diagramme circulaire ci-dessous représente la proportion de chômeurs par tranche d'âge.



La légende conclut que la classe la plus touchée est celle des 25 à 49 ans. Les trois classes étant d'amplitudes différentes, il est difficile d'établir des comparaisons. Il n'est en effet pas surprenant que le plus grand nombre de chômeurs se trouve dans la classe la plus peuplée !
En fait, la valeur intéressante n'est pas la valeur absolue, mais le pourcentage de chômeurs à l'intérieur de chaque classe. Selon le rapport du SECO utilisé pour créer le graphique, ces taux sont de 4% *pour les 15-24 ans, de* $3,6\%$ *pour les 25-49 ans,*

Exemple

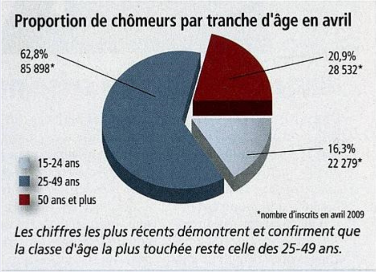
Le diagramme circulaire ci-dessous représente la proportion de chômeurs par tranche d'âge.



La légende conclut que la classe la plus touchée est celle des 25 à 49 ans. Les trois classes étant d'amplitudes différentes, il est difficile d'établir des comparaisons. Il n'est en effet pas surprenant que le plus grand nombre de chômeurs se trouve dans la classe la plus peuplée !
En fait, la valeur intéressante n'est pas la valeur absolue, mais le pourcentage de chômeurs à l'intérieur de chaque classe. Selon le rapport du SECO utilisé pour créer le graphique, ces taux sont de 4% *pour les 15-24 ans, de* $3,6\%$ *pour les 25-49 ans,et en-dessous de* 3% *pour les 50-65 ans.*

Exemple

Le diagramme circulaire ci-dessous représente la proportion de chômeurs par tranche d'âge.



La légende conclut que la classe la plus touchée est celle des 25 à 49 ans. Les trois classes étant d'amplitudes différentes, il est difficile d'établir des comparaisons. Il n'est en effet pas surprenant que le plus grand nombre de chômeurs se trouve dans la classe la plus peuplée !
En fait, la valeur intéressante n'est pas la valeur absolue, mais le pourcentage de chômeurs à l'intérieur de chaque classe. Selon le rapport du SECO utilisé pour créer le graphique, ces taux sont de 4% *pour les 15-24 ans, de* $3,6\%$ *pour les 25-49 ans,et en-dessous de* 3% *pour les 50-65 ans.La classe la plus touchée est donc celle des jeunes, contrairement à ce qu'en dit l'auteur du diagramme.*

Exemple

Une chaîne de télévision a présenté le diagramme ci-dessous en 2011. Celui-ci rend compte du taux de dépenses publiques en 2011 en % du PIB de trois pays et de l'Union européenne.

Exemple

Une chaîne de télévision a présenté le diagramme ci-dessous en 2011. Celui-ci rend compte du taux de dépenses publiques en 2011 en % du PIB de trois pays et de l'Union européenne.

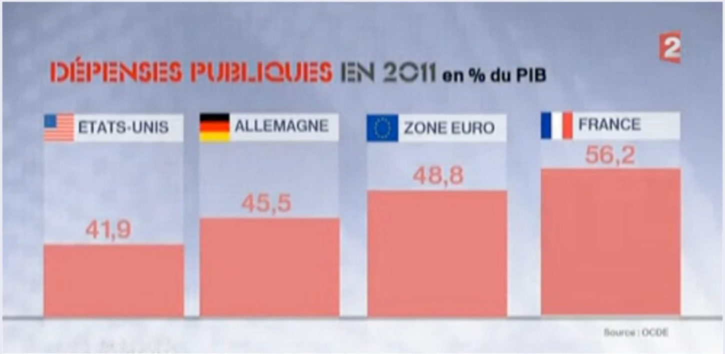


FIGURE – *Diagramme erroné.*

Exemple

Si le 100% *correspond à la zone comprise entre l'horizontale et à la parallèle passant juste en dessous du nom des pays, les* $41,9\%$ *de dépenses publiques de Etats-Unis semblent être correctement représentées.*

Exemple

Si le 100% *correspond à la zone comprise entre l'horizontale et à la parallèle passant juste en dessous du nom des pays, les* $41,9\%$ *de dépenses publiques de Etats-Unis semblent être correctement représentées.Cela n'est pas le cas pour les autres pays ! En effet, les* $56,2\%$ *de la France sont beaucoup trop hauts.*

Exemple

Si le 100% *correspond à la zone comprise entre l'horizontale et à la parallèle passant juste en dessous du nom des pays, les* $41,9\%$ *de dépenses publiques de Etats-Unis semblent être correctement représentées.Cela n'est pas le cas pour les autres pays ! En effet, les* $56,2\%$ *de la France sont beaucoup trop hauts.Le diagramme donne l'impression que les dépenses publiques de la France sont de l'ordre de* 80% *!*

Exemple

Si le 100% *correspond à la zone comprise entre l'horizontale et à la parallèle passant juste en dessous du nom des pays, les* $41,9\%$ *de dépenses publiques de Etats-Unis semblent être correctement représentées.Cela n'est pas le cas pour les autres pays ! En effet, les* $56,2\%$ *de la France sont beaucoup trop hauts.Le diagramme donne l'impression que les dépenses publiques de la France sont de l'ordre de* 80% *!Cette technique a pour objectif de susciter une émotion auprès de la population, en amplifiant la différence du taux de dépenses publiques par rapport à d'autres pays.*

Exemple

Si le 100% *correspond à la zone comprise entre l'horizontale et à la parallèle passant juste en dessous du nom des pays, les* $41,9\%$ *de dépenses publiques de Etats-Unis semblent être correctement représentées.Cela n'est pas le cas pour les autres pays ! En effet, les* $56,2\%$ *de la France sont beaucoup trop hauts.Le diagramme donne l'impression que les dépenses publiques de la France sont de l'ordre de* 80% *!Cette technique a pour objectif de susciter une émotion auprès de la population, en amplifiant la différence du taux de dépenses publiques par rapport à d'autres pays. Ci-dessous, figure le diagramme correct, tel qu'il aurait dû être présenté aux téléspectateurs.*

Exemple

Si le 100% correspond à la zone comprise entre l'horizontale et à la parallèle passant juste en dessous du nom des pays, les 41, 9% de dépenses publiques de Etats-Unis semblent être correctement représentées.Cela n'est pas le cas pour les autres pays ! En effet, les 56, 2% de la France sont beaucoup trop hauts.Le diagramme donne l'impression que les dépenses publiques de la France sont de l'ordre de 80% !Cette technique a pour objectif de susciter une émotion auprès de la population, en amplifiant la différence du taux de dépenses publiques par rapport à d'autres pays. Ci-dessous, figure le diagramme correct, tel qu'il aurait dû être présenté aux téléspectateurs.

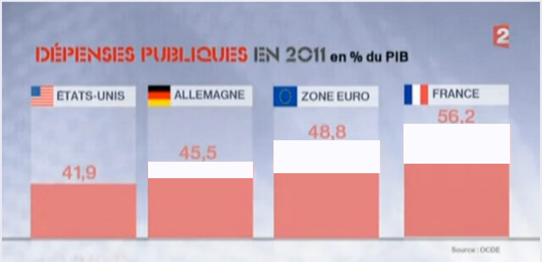


FIGURE – *Diagramme correct.*

Exemple

La figure ci-dessous illustre le fait que le prix des montres a augmenté de 40% *sur 7 ans.*

Exemple

La figure ci-dessous illustre le fait que le prix des montres a augmenté de 40% *sur* 7 *ans.*

Exemple

La figure ci-dessous illustre le fait que le prix des montres a augmenté de 40% *sur 7 ans.*

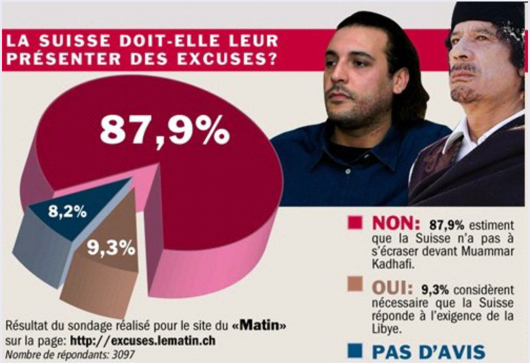
Le graphiste a voulu associer cette augmentation au diamètre des montres. On peut vérifier qu'ils augmentent bien de 40%*. Le lecteur voit l'augmentation de la surface des horloges qui, elle, n'est pas de* 40%*, mais proche de* 100% *! Enfin, les aiguilles ont été ajoutées dans un but purement esthétique, mais peuvent induire en erreur en donnant l'impression qu'elles contiennent de l'information.*

Exemple

Dans le diagramme circulaire ci-dessous, la somme des parties fait $105,4\%$ *! Le* $8,2\%$ *était probablement un* $2,8\%$ *à l'origine, ce qui donnerait la somme attendue de* 100%.

Exemple

Dans le diagramme circulaire ci-dessous, la somme des parties fait $105,4\%$ *! Le* $8,2\%$ *était probablement un* $2,8\%$ *à l'origine, ce qui donnerait la somme attendue de* 100%*.*



Polygone des effectifs

A l'histogramme, on associe souvent le *polygone des effectifs*. Il s'agit d'une courbe polygonale telle que la surface comprise entre cette courbe et l'axe des abscisses soit égale à la surface de l'histogramme. Elle est obtenue en joignant les milieux des sommets des rectangles de l'histogramme.

Polygone des effectifs

A l'histogramme, on associe souvent le *polygone des effectifs*. Il s'agit d'une courbe polygonale telle que la surface comprise entre cette courbe et l'axe des abscisses soit égale à la surface de l'histogramme. Elle est obtenue en joignant les milieux des sommets des rectangles de l'histogramme.Pour la première et la dernière classe, on crée à cet effet deux classes fictives d'effectifs nuls.

A l'histogramme, on associe souvent le *polygone des effectifs*. Il s'agit d'une courbe polygonale telle que la surface comprise entre cette courbe et l'axe des abscisses soit égale à la surface de l'histogramme. Elle est obtenue en joignant les milieux des sommets des rectangles de l'histogramme.Pour la première et la dernière classe, on crée à cet effet deux classes fictives d'effectifs nuls.

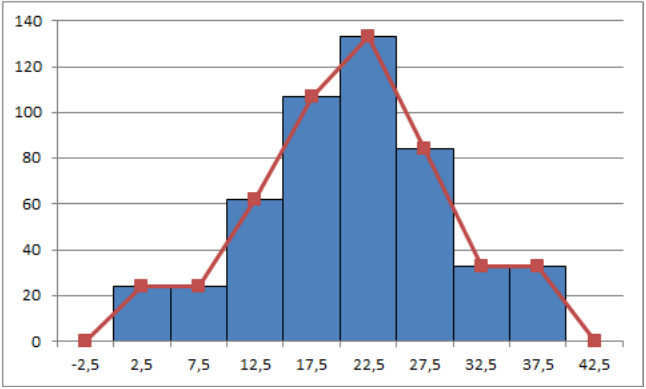


FIGURE – Polygone des effectifs.

Aux données de départ, on associe le tableau des effectifs cumulés croissants et cumulés décroissants. On interprète les données de ce tableau comme suit. On peut affirmer, par exemple, que 350 exploitations agricoles ont une superficie d'au plus 25 ha. Par ailleurs, 283 exploitations ont une superficie d'au moins à 20 ha.

Aux données de départ, on associe le tableau des effectifs cumulés croissants et cumulés décroissants. On interprète les données de ce tableau comme suit. On peut affirmer, par exemple, que 350 exploitations agricoles ont une superficie d'au plus 25 ha. Par ailleurs, 283 exploitations ont une superficie d'au moins à 20 ha.

Classes	Effectifs	Effectifs cumulés croissants	Effectifs cumulés décroissants
]0; 10]	48	48	500
]10; 15]	62	110	452
]15; 20]	107	217	390
]20; 25]	133	350	283
]25; 30]	84	434	150
]30; 40]	66	500	66

Aux données de départ, on associe le tableau des effectifs cumulés croissants et cumulés décroissants. On interprète les données de ce tableau comme suit. On peut affirmer, par exemple, que 350 exploitations agricoles ont une superficie d'au plus 25 ha. Par ailleurs, 283 exploitations ont une superficie d'au moins à 20 ha.

Classes	Effectifs	Effectifs cumulés croissants	Effectifs cumulés décroissants
]0; 10]	48	48	500
]10; 15]	62	110	452
]15; 20]	107	217	390
]20; 25]	133	350	283
]25; 30]	84	434	150
]30; 40]	66	500	66

Les données contenues dans ce tableau peuvent être représentées par deux courbes : *le polygone des effectifs cumulés croissants* et le *polygone des effectifs cumulés décroissants*.

Polygone des effectifs cumulés

Aux données de départ, on associe le tableau des effectifs cumulés croissants et cumulés décroissants. On interprète les données de ce tableau comme suit. On peut affirmer, par exemple, que 350 exploitations agricoles ont une superficie d'au plus 25 ha. Par ailleurs, 283 exploitations ont une superficie d'au moins à 20 ha.

Classes	Effectifs	Effectifs cumulés croissants	Effectifs cumulés décroissants
]0; 10]	48	48	500
]10; 15]	62	110	452
]15; 20]	107	217	390
]20; 25]	133	350	283
]25; 30]	84	434	150
]30; 40]	66	500	66

Les données contenues dans ce tableau peuvent être représentées par deux courbes : *le polygone des effectifs cumulés croissants* et le *polygone des effectifs cumulés décroissants*.Dans la représentation de ces courbes, on ne se préoccupe pas des différences d'amplitude des classes. Notons qu'il est également possible de réaliser un polygone des fréquences.

Polygone des effectifs cumulés

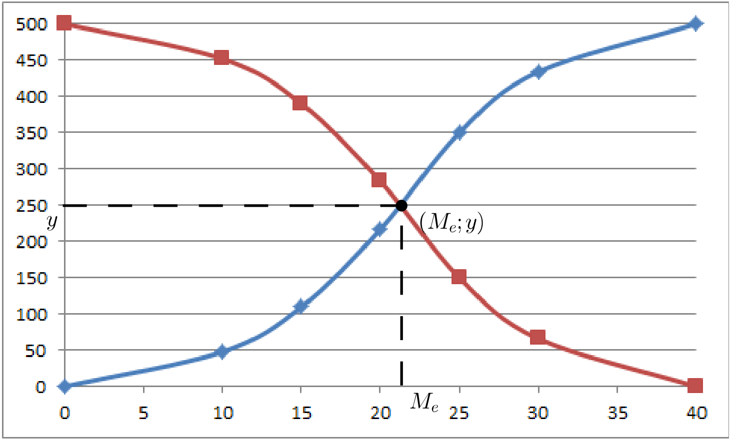


FIGURE – Polygones des effectifs cumulés.

Remarque

L'intersection de ces deux polygones est un point $(M_e; y)$, *dont la première coordonnée est appelée médiane* M_e *de la population. On observe, dans notre exemple, que* $M_e \cong 21$. *Cette valeur divise la population en deux parties d'effectifs égaux. En effet, soit* y *la seconde coordonnée du point d'intersection. Comme ce point est sur le polygone des effectifs cumulés croissants, on peut affirmer que* y *exploitations ont une superficie inferieure à* M_e *; le reste, c'est-a-dire* $500 - y$, *ont une superficie superieure à* M_e. *Ce point étant également sur le polygone des effectifs cumulés décroissants,* y *decrit le nombre d'exploitations ayant une superficie superieure à* M_e. *On en deduit que* $y = 500 - y$ *et donc, que* $y = \dfrac{500}{2} = 250$. *Ainsi la moitié des exploitations ont une superficie supérieure (respectivement inférieure) à* $M_e \cong 21$ *ha.*

Remarque

Dans une étude statistique, si on souhaite connaître la proportion de chaque valeur que peut prendre la variable statistique étudiée, on regarde sa fréquence f_i.

Remarque

L'intersection de ces deux polygones est un point $(M_e; y)$, *dont la première coordonnée est appelée médiane* M_e *de la population. On observe, dans notre exemple, que* $M_e \cong 21$. *Cette valeur divise la population en deux parties d'effectifs égaux. En effet, soit* y *la seconde coordonnée du point d'intersection. Comme ce point est sur le polygone des effectifs cumulés croissants, on peut affirmer que* y *exploitations ont une superficie inferieure à* M_e *; le reste, c'est-a-dire* $500 - y$, *ont une superficie superieure à* M_e. *Ce point étant également sur le polygone des effectifs cumulés décroissants,* y *decrit le nombre d'exploitations ayant une superficie superieure à* M_e. *On en deduit que* $y = 500 - y$ *et donc, que* $y = \frac{500}{2} = 250$. *Ainsi la moitié des exploitations ont une superficie supérieure (respectivement inférieure) à* $M_e \cong 21$ *ha.*

Remarque

Dans une étude statistique, si on souhaite connaître la proportion de chaque valeur que peut prendre la variable statistique étudiée, on regarde sa fréquence f_i.
Si par contre on souhaite connaître la proportion des individus qui présentent des valeurs inférieures à une valeur fixée, on regarde la fréquence cumulée croissante F_i.

Remarque

L'intersection de ces deux polygones est un point $(M_e; y)$, dont la première coordonnée est appelée médiane M_e de la population. On observe, dans notre exemple, que $M_e \cong 21$. Cette valeur divise la population en deux parties d'effectifs égaux. En effet, soit y la seconde coordonnée du point d'intersection. Comme ce point est sur le polygone des effectifs cumulés croissants, on peut affirmer que y exploitations ont une superficie inferieure à M_e ; le reste, c'est-a-dire $500 - y$, ont une superficie superieure à M_e. Ce point étant également sur le polygone des effectifs cumulés décroissants, y decrit le nombre d'exploitations ayant une superficie superieure à M_e. On en deduit que $y = 500 - y$ et donc, que $y = \dfrac{500}{2} = 250$. Ainsi la moitié des exploitations ont une superficie supérieure (respectivement inférieure) à $M_e \cong 21$ ha.

Remarque

Dans une étude statistique, si on souhaite connaître la proportion de chaque valeur que peut prendre la variable statistique étudiée, on regarde sa fréquence f_i.
Si par contre on souhaite connaître la proportion des individus qui présentent des valeurs inférieures à une valeur fixée, on regarde la fréquence cumulée croissante F_i.
Pour visualiser la proportion des individus qui présentent des valeurs supérieures ou égales à une valeur fixée, on étudiera alors la fréquence cumulée décroissante F'_i.

Valeurs centrales

Une *valeur centrale* est une valeur caractéristique ou représentative d'un ensemble de données. Si cette valeur caractéristique a tendance à se situer au milieu d'un ensemble de données rangées par ordre de grandeur croissant, alors on dit qu'elle est une *mesure de tendance centrale* ou une *valeur centrale*.

Definition

La *moyenne arithmétique* $\overline{x}$ est la valeur centrale la plus connue. Elle est égale au quotient de la somme de toutes les valeurs observées du caractère par l'effectif total.

Definition

La *moyenne arithmétique* $\overline{x}$ est la valeur centrale la plus connue. Elle est égale au quotient de la somme de toutes les valeurs observées du caractère par l'effectif total.Ainsi

$$\boxed{\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_N x_N}{N}.}$$

Exemple

Reprenons notre exemple du nombre de personnes par ménage :

Modalités x_i	Effectifs n_i	Fréquences f_i
1	5	10%
2	9	18%
3	15	30%
4	10	20%
5	6	12%
6	3	6%
7	0	0%
8	2	4%

Exemple

Reprenons notre exemple du nombre de personnes par ménage :

Modalités x_i	Effectifs n_i	Fréquences f_i
1	5	10%
2	9	18%
3	15	30%
4	10	20%
5	6	12%
6	3	6%
7	0	0%
8	2	4%

la moyenne arithmétique $\overline{x}$ est donnée par

Exemple

Reprenons notre exemple du nombre de personnes par ménage :

Modalités x_i	Effectifs n_i	Fréquences f_i
1	5	10%
2	9	18%
3	15	30%
4	10	20%
5	6	12%
6	3	6%
7	0	0%
8	2	4%

la moyenne arithmétique $\overline{x}$ est donnée par

$$\overline{x}$$

Exemple

Reprenons notre exemple du nombre de personnes par ménage :

Modalités x_i	*Effectifs* n_i	*Fréquences* f_i
1	*5*	10%
2	*9*	18%
3	*15*	30%
4	*10*	20%
5	*6*	12%
6	*3*	6%
7	*0*	0%
8	*2*	4%

la moyenne arithmétique $\overline{x}$ *est donnée par*

$$\overline{x} = \frac{5 \cdot 1 + 9 \cdot 2 + 15 \cdot 3 + 10 \cdot 4 + 6 \cdot 5 + 3 \cdot 6 + 2 \cdot 8}{50}$$

Exemple

Reprenons notre exemple du nombre de personnes par ménage :

Modalités x_i	*Effectifs* n_i	*Fréquences* f_i
1	*5*	10%
2	*9*	18%
3	*15*	30%
4	*10*	20%
5	*6*	12%
6	*3*	6%
7	*0*	0%
8	*2*	4%

la moyenne arithmétique $\overline{x}$ *est donnée par*

$$\overline{x} = \frac{5 \cdot 1 + 9 \cdot 2 + 15 \cdot 3 + 10 \cdot 4 + 6 \cdot 5 + 3 \cdot 6 + 2 \cdot 8}{50} = 3,44.$$

Pour des séries de données groupées, se fondant sur une répartition uniforme au sein des classes, on convient d'affecter à tous les individus d'une classe $]b_{i-1}, b_i]$ le centre

$$c = \frac{b_{i-1} + b_i}{2}.$$

Pour des séries de données groupées, se fondant sur une répartition uniforme au sein des classes, on convient d'affecter à tous les individus d'une classe $]b_{i-1}, b_i]$ le centre

$$c = \frac{b_{i-1} + b_i}{2}.$$

Exemple

Pour notre exemple des exploitations agricoles, à l'aide du tableau suivant

Classes x_i	*Centres* c_i	*Effectifs* n_i
$]0; 10]$	*5*	*48*
$]10; 15]$	$12,5$	*62*
$]15; 20]$	$17,5$	*107*
$]20; 25]$	$22,5$	*133*
$]25; 30]$	$27,5$	*84*
$]30; 40]$	*35*	*66*
Total		*500*

Pour des séries de données groupées, se fondant sur une répartition uniforme au sein des classes, on convient d'affecter à tous les individus d'une classe $]b_{i-1}, b_i]$ le centre

$$c = \frac{b_{i-1} + b_i}{2}.$$

Exemple

Pour notre exemple des exploitations agricoles, à l'aide du tableau suivant

Classes x_i	*Centres* c_i	*Effectifs* n_i
$]0; 10]$	*5*	*48*
$]10; 15]$	$12,5$	*62*
$]15; 20]$	$17,5$	*107*
$]20; 25]$	$22,5$	*133*
$]25; 30]$	$27,5$	*84*
$]30; 40]$	*35*	*66*
Total		*500*

on tire la moyenne arithmétique des superficies de ces 500 exploitations agricoles, en calculant

Pour des séries de données groupées, se fondant sur une répartition uniforme au sein des classes, on convient d'affecter à tous les individus d'une classe $]b_{i-1}, b_i]$ le centre

$$c = \frac{b_{i-1} + b_i}{2}.$$

Exemple

Pour notre exemple des exploitations agricoles, à l'aide du tableau suivant

Classes x_i	*Centres* c_i	*Effectifs* n_i
$]0; 10]$	*5*	*48*
$]10; 15]$	$12,5$	*62*
$]15; 20]$	$17,5$	*107*
$]20; 25]$	$22,5$	*133*
$]25; 30]$	$27,5$	*84*
$]30; 40]$	*35*	*66*
Total		*500*

on tire la moyenne arithmétique des superficies de ces 500 exploitations agricoles, en calculant

$\overline{x}$

Pour des séries de données groupées, se fondant sur une répartition uniforme au sein des classes, on convient d'affecter à tous les individus d'une classe $]b_{i-1}, b_i]$ le centre

$$c = \frac{b_{i-1} + b_i}{2}.$$

Exemple

Pour notre exemple des exploitations agricoles, à l'aide du tableau suivant

Classes x_i	Centres c_i	Effectifs n_i
$]0; 10]$	*5*	*48*
$]10; 15]$	$12,5$	*62*
$]15; 20]$	$17,5$	*107*
$]20; 25]$	$22,5$	*133*
$]25; 30]$	$27,5$	*84*
$]30; 40]$	*35*	*66*
Total		*500*

on tire la moyenne arithmétique des superficies de ces 500 exploitations agricoles, en calculant

$$\overline{x} = \frac{5 \cdot 48 + 12,5 \cdot 62 + 17,5 \cdot 107 + 22,5 \cdot 133 + 27,5 \cdot 84 + 35 \cdot 66}{500} = \frac{10500}{500}$$

Pour des séries de données groupées, se fondant sur une répartition uniforme au sein des classes, on convient d'affecter à tous les individus d'une classe $]b_{i-1}, b_i]$ le centre

$$c = \frac{b_{i-1} + b_i}{2}.$$

Exemple

Pour notre exemple des exploitations agricoles, à l'aide du tableau suivant

Classes x_i	*Centres* c_i	*Effectifs* n_i
$]0; 10]$	*5*	*48*
$]10; 15]$	$12,5$	*62*
$]15; 20]$	$17,5$	*107*
$]20; 25]$	$22,5$	*133*
$]25; 30]$	$27,5$	*84*
$]30; 40]$	*35*	*66*
Total		*500*

on tire la moyenne arithmétique des superficies de ces 500 exploitations agricoles, en calculant

$$\overline{x} = \frac{5 \cdot 48 + 12,5 \cdot 62 + 17,5 \cdot 107 + 22,5 \cdot 133 + 27,5 \cdot 84 + 35 \cdot 66}{500} = \frac{10500}{500} = 21 \; ha.$$

Exemple

On constate que, dans un village de 500 habitants, il y a 490 personnes avec des cheveux noirs et 10 avec des cheveux blonds.

Exemple

On constate que, dans un village de 500 habitants, il y a 490 personnes avec des cheveux noirs et 10 avec des cheveux blonds.Comment résumer la couleur des cheveux "moyenne" des habitants de ce village ?

Exemple

On constate que, dans un village de 500 habitants, il y a 490 personnes avec des cheveux noirs et 10 avec des cheveux blonds.Comment résumer la couleur des cheveux "moyenne" des habitants de ce village ?On répondra sûrement "noir", en pensant que l'écrasante majorité des habitants a les cheveux noirs.

Exemple

On constate que, dans un village de 500 habitants, il y a 490 personnes avec des cheveux noirs et 10 avec des cheveux blonds.Comment résumer la couleur des cheveux "moyenne" des habitants de ce village ?On répondra sûrement "noir", en pensant que l'écrasante majorité des habitants a les cheveux noirs.En réfléchissant ainsi, on donne comme réponse la valeur qui apparaît le plus fréquemment.

Exemple

On constate que, dans un village de 500 habitants, il y a 490 personnes avec des cheveux noirs et 10 avec des cheveux blonds.Comment résumer la couleur des cheveux "moyenne" des habitants de ce village ?On répondra sûrement "noir", en pensant que l'écrasante majorité des habitants a les cheveux noirs.En réfléchissant ainsi, on donne comme réponse la valeur qui apparaît le plus fréquemment.Il s'agit du mode.

Definition

Le *mode*, noté M_o, est la valeur du caractère qui correspond à l'effectif le plus grand ou à la fréquence la plus importante.

Definition

Le *mode*, noté M_o, est la valeur du caractère qui correspond à l'effectif le plus grand ou à la fréquence la plus importante.Cette valeur centrale est simple à percevoir, mais elle ne tient pas compte de l'ensemble des valeurs du caractère étudié.

Exemple

Les nombres 3, 5, 7, 7, 7, 9, 9 ont pour mode $M_o = 7$.

Definition

Le *mode*, noté M_o, est la valeur du caractère qui correspond à l'effectif le plus grand ou à la fréquence la plus importante.Cette valeur centrale est simple à percevoir, mais elle ne tient pas compte de l'ensemble des valeurs du caractère étudié.

Exemple

Les nombres 3, 5, 7, 7, 7, 9, 9 ont pour mode $M_o = 7$*.Remarquons que le mode peut ne pas être unique.*

Definition

Le *mode*, noté M_o, est la valeur du caractère qui correspond à l'effectif le plus grand ou à la fréquence la plus importante.Cette valeur centrale est simple à percevoir, mais elle ne tient pas compte de l'ensemble des valeurs du caractère étudié.

Exemple

Les nombres 3, 5, 7, 7, 7, 9, 9 ont pour mode $M_o = 7$*.Remarquons que le mode peut ne pas être unique.Ainsi, l'ensemble 3, 5, 7, 7, 7, 9, 9, 9, qui a deux modes : 7 et 9, est dit bimodal.*

Exemple

Reprenons notre exemple du nombre de personnes par ménage.

Modalités x_i	*Effectifs* n_i
1	5
2	9
3	15
4	10
5	6
6	3
7	0
8	2

Exemple

Reprenons notre exemple du nombre de personnes par ménage.

Modalités x_i	Effectifs n_i
1	5
2	9
3	15
4	10
5	6
6	3
7	0
8	2

le mode est donné par $M_o = 3$*, car 3 est la modalité au plus grand effectif.*

Pour des séries de données groupées par classes, la détermination du mode s'effectue comme suit :

Pour des séries de données groupées par classes, la détermination du mode s'effectue comme suit :

1. On détermine les effectifs rectifiés.

Pour des séries de données groupées par classes, la détermination du mode s'effectue comme suit :

1. On détermine les effectifs rectifiés.
2. On identifie la classe ayant le plus grand des effectifs rectifiés. Elle porte le nom de *classe modale* et peut ne pas être unique.

Pour des séries de données groupées par classes, la détermination du mode s'effectue comme suit :

1. On détermine les effectifs rectifiés.
2. On identifie la classe ayant le plus grand des effectifs rectifiés. Elle porte le nom de *classe modale* et peut ne pas être unique.
3. On convient que le mode est déporté à l'intérieur de la classe modale, à droite de sa borne inférieure A, en fonction des effectifs rectifiés des classes voisines. Le mode est alors défini par la formule

Pour des séries de données groupées par classes, la détermination du mode s'effectue comme suit :

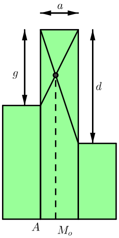
1. On détermine les effectifs rectifiés.
2. On identifie la classe ayant le plus grand des effectifs rectifiés. Elle porte le nom de *classe modale* et peut ne pas être unique.
3. On convient que le mode est déporté à l'intérieur de la classe modale, à droite de sa borne inférieure A, en fonction des effectifs rectifiés des classes voisines. Le mode est alors défini par la formule

$$\boxed{M_o = A + \frac{g}{g+d} \cdot a.}$$

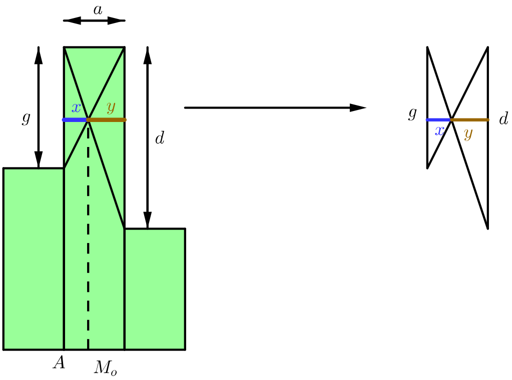
Pour des séries de données groupées par classes, la détermination du mode s'effectue comme suit :

1. On détermine les effectifs rectifiés.
2. On identifie la classe ayant le plus grand des effectifs rectifiés. Elle porte le nom de *classe modale* et peut ne pas être unique.
3. On convient que le mode est déporté à l'intérieur de la classe modale, à droite de sa borne inférieure A, en fonction des effectifs rectifiés des classes voisines. Le mode est alors défini par la formule

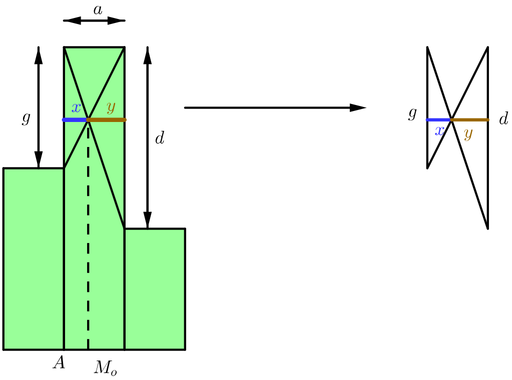
$$\boxed{M_o = A + \frac{g}{g+d} \cdot a.}$$



Preuve de la formule du mode :

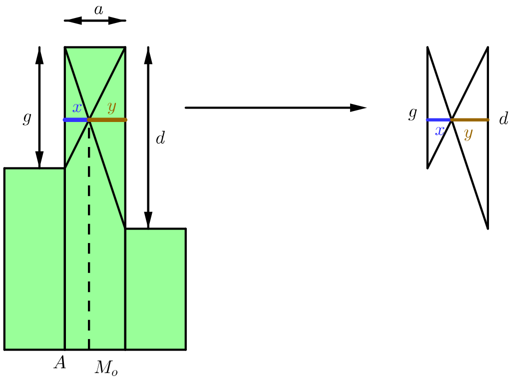


Preuve de la formule du mode :



Il est clair que

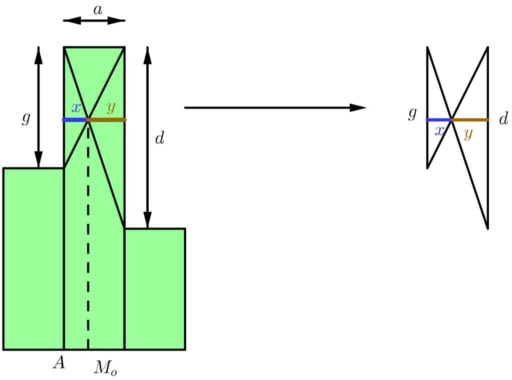
Preuve de la formule du mode :



Il est clair que

- $x = M_o - A$.

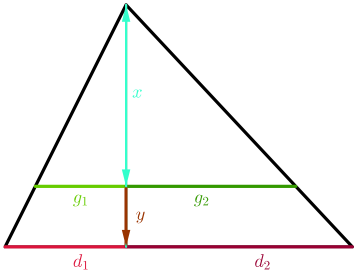
Preuve de la formule du mode :



Il est clair que

- $x = M_o - A$.
- $y = a - x = a - (M_o - A) = a - M_o + A$.

Posons alors d_1, d_2, g_1 et g_2 comme le montre la figure ci-dessous.



Puisque l'on a deux triangles semblables, il est possible d'appliquer le théorème de Thalès :

Puisque l'on a deux triangles semblables, il est possible d'appliquer le théorème de Thalès :

- $\dfrac{y}{x} = \dfrac{d_1}{g_1} \Rightarrow y \cdot g_1 = x \cdot d_1.$

Puisque l'on a deux triangles semblables, il est possible d'appliquer le théorème de Thalès :

- $\dfrac{y}{x} = \dfrac{d_1}{g_1} \Rightarrow y \cdot g_1 = x \cdot d_1.$
- $\dfrac{y}{x} = \dfrac{d_2}{g_2} \Rightarrow y \cdot g_2 = x \cdot d_2.$

Puisque l'on a deux triangles semblables, il est possible d'appliquer le théorème de Thalès :

- $\dfrac{y}{x} = \dfrac{d_1}{g_1} \Rightarrow y \cdot g_1 = x \cdot d_1.$
- $\dfrac{y}{x} = \dfrac{d_2}{g_2} \Rightarrow y \cdot g_2 = x \cdot d_2.$

On a alors

Puisque l'on a deux triangles semblables, il est possible d'appliquer le théorème de Thalès :

- $\frac{y}{x} = \frac{d_1}{g_1} \Rightarrow y \cdot g_1 = x \cdot d_1$.
- $\frac{y}{x} = \frac{d_2}{g_2} \Rightarrow y \cdot g_2 = x \cdot d_2$.

On a alors

$$x \cdot d_1 + x \cdot d_2 \quad = \quad y \cdot g_1 + y \cdot g_2$$

Puisque l'on a deux triangles semblables, il est possible d'appliquer le théorème de Thalès :

- $\dfrac{y}{x} = \dfrac{d_1}{g_1} \Rightarrow y \cdot g_1 = x \cdot d_1.$
- $\dfrac{y}{x} = \dfrac{d_2}{g_2} \Rightarrow y \cdot g_2 = x \cdot d_2.$

On a alors

$$\begin{aligned} x \cdot d_1 + x \cdot d_2 &= y \cdot g_1 + y \cdot g_2 \\ x \cdot (d_1 + d_2) &= y \cdot (g_1 + g_2) \end{aligned}$$

Puisque l'on a deux triangles semblables, il est possible d'appliquer le théorème de Thalès :

- $\dfrac{y}{x} = \dfrac{d_1}{g_1} \Rightarrow y \cdot g_1 = x \cdot d_1$.
- $\dfrac{y}{x} = \dfrac{d_2}{g_2} \Rightarrow y \cdot g_2 = x \cdot d_2$.

On a alors

$$\begin{aligned} x \cdot d_1 + x \cdot d_2 &= y \cdot g_1 + y \cdot g_2 \\ x \cdot (d_1 + d_2) &= y \cdot (g_1 + g_2) \\ x \cdot d &= y \cdot g \end{aligned}$$

Puisque l'on a deux triangles semblables, il est possible d'appliquer le théorème de Thalès :

- $\dfrac{y}{x} = \dfrac{d_1}{g_1} \Rightarrow y \cdot g_1 = x \cdot d_1.$
- $\dfrac{y}{x} = \dfrac{d_2}{g_2} \Rightarrow y \cdot g_2 = x \cdot d_2.$

On a alors

$$\begin{aligned} x \cdot d_1 + x \cdot d_2 &= y \cdot g_1 + y \cdot g_2 \\ x \cdot (d_1 + d_2) &= y \cdot (g_1 + g_2) \\ x \cdot d &= y \cdot g \\ x &= \frac{y \cdot g}{d}. \end{aligned}$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$M_o - A \quad = \quad \frac{g}{d} \cdot (a - M_o + A)$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$\begin{aligned} M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\ M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \end{aligned}$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$\begin{aligned} M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\ M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\ M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \end{aligned}$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A)
\end{aligned}$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$
\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d}{d} + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A)
\end{aligned}
$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d}{d} + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d + g}{d}\right) &= A + \frac{g}{d} \cdot (a + A)
\end{aligned}$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$
\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d}{d} + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d + g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o &= \left(A + \frac{g}{d} \cdot (a + A)\right) \cdot \left(\frac{d}{d + g}\right)
\end{aligned}
$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d}{d} + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d+g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o &= \left(A + \frac{g}{d} \cdot (a + A)\right) \cdot \left(\frac{d}{d+g}\right) \\
M_o &= A \cdot \frac{d}{d+g} + \frac{g}{d} \cdot (a + A) \cdot \frac{d}{d+g}
\end{aligned}$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d}{d} + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d + g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o &= \left(A + \frac{g}{d} \cdot (a + A)\right) \cdot \left(\frac{d}{d + g}\right) \\
M_o &= A \cdot \frac{d}{d + g} + \frac{g}{d} \cdot (a + A) \cdot \frac{d}{d + g} \\
M_o &= A \cdot \frac{d}{d + g} + (a + A) \cdot \frac{g}{d + g}
\end{aligned}$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d}{d} + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d+g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o &= \left(A + \frac{g}{d} \cdot (a + A)\right) \cdot \left(\frac{d}{d+g}\right) \\
M_o &= A \cdot \frac{d}{d+g} + \frac{g}{d} \cdot (a + A) \cdot \frac{d}{d+g} \\
M_o &= A \cdot \frac{d}{d+g} + (a + A) \cdot \frac{g}{d+g} \\
M_o &= \frac{Ad + ag + Ag}{d+g}
\end{aligned}$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$
\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d}{d} + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d+g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o &= \left(A + \frac{g}{d} \cdot (a + A)\right) \cdot \left(\frac{d}{d+g}\right) \\
M_o &= A \cdot \frac{d}{d+g} + \frac{g}{d} \cdot (a + A) \cdot \frac{d}{d+g} \\
M_o &= A \cdot \frac{d}{d+g} + (a + A) \cdot \frac{g}{d+g} \\
M_o &= \frac{Ad + ag + Ag}{d+g} \\
M_o &= \frac{Ad + Ag}{d+g} + \frac{ag}{d+g}
\end{aligned}
$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$
\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d}{d} + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d+g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o &= \left(A + \frac{g}{d} \cdot (a + A)\right) \cdot \left(\frac{d}{d+g}\right) \\
M_o &= A \cdot \frac{d}{d+g} + \frac{g}{d} \cdot (a + A) \cdot \frac{d}{d+g} \\
M_o &= A \cdot \frac{d}{d+g} + (a + A) \cdot \frac{g}{d+g} \\
M_o &= \frac{Ad + ag + Ag}{d+g} \\
M_o &= \frac{Ad + Ag}{d+g} + \frac{ag}{d+g} \\
M_o &= \frac{A \cdot (d+g)}{d+g} + \frac{ag}{d+g}
\end{aligned}
$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$
\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d}{d} + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d+g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o &= \left(A + \frac{g}{d} \cdot (a + A)\right) \cdot \left(\frac{d}{d+g}\right) \\
M_o &= A \cdot \frac{d}{d+g} + \frac{g}{d} \cdot (a + A) \cdot \frac{d}{d+g} \\
M_o &= A \cdot \frac{d}{d+g} + (a + A) \cdot \frac{g}{d+g} \\
M_o &= \frac{Ad + ag + Ag}{d+g} \\
M_o &= \frac{Ad + Ag}{d+g} + \frac{ag}{d+g} \\
M_o &= \frac{A \cdot (d+g)}{d+g} + \frac{ag}{d+g}
\end{aligned}
$$

Comme $x = M_o - A$ et $y = a - M_o + A$, cette dernière égalité s'écrit

$$
\begin{aligned}
M_o - A &= \frac{g}{d} \cdot (a - M_o + A) \\
M_o - A &= \frac{g}{d} \cdot a - \frac{g}{d} \cdot M_o + \frac{g}{d} \cdot A \\
M_o + \frac{g}{d} \cdot M_o &= A + \frac{g}{d} \cdot a + \frac{g}{d} \cdot A \\
M_o \cdot \left(1 + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d}{d} + \frac{g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o \cdot \left(\frac{d+g}{d}\right) &= A + \frac{g}{d} \cdot (a + A) \\
M_o &= \left(A + \frac{g}{d} \cdot (a + A)\right) \cdot \left(\frac{d}{d+g}\right) \\
M_o &= A \cdot \frac{d}{d+g} + \frac{g}{d} \cdot (a + A) \cdot \frac{d}{d+g} \\
M_o &= A \cdot \frac{d}{d+g} + (a + A) \cdot \frac{g}{d+g} \\
M_o &= \frac{Ad + ag + Ag}{d+g} \\
M_o &= \frac{Ad + Ag}{d+g} + \frac{ag}{d+g} \\
M_o &= \frac{A \cdot (d+g)}{d+g} + \frac{ag}{d+g}
\end{aligned}
$$

Exemple

Dans notre exemple des exploitatations agricoles, après rectification des effectifs, on obtient le tableau suivant :

Exemple

Dans notre exemple des exploitatations agricoles, après rectification des effectifs, on obtient le tableau suivant :

Superfice en ha	*Nombre d'exploitations*
$]0; 10]$	*24*
$]10; 15]$	*62*
$]15; 20]$	*107*
$]20; 25]$	*133*
$]25; 30]$	*84*
$]30; 40]$	*33*

La classe modale est donc la quatrième classe.

Exemple

Dans notre exemple des exploitatations agricoles, après rectification des effectifs, on obtient le tableau suivant :

Superficie en ha	*Nombre d'exploitations*
]0; 10]	*24*
]10; 15]	*62*
]15; 20]	*107*
]20; 25]	*133*
]25; 30]	*84*
]30; 40]	*33*

La classe modale est donc la quatrième classe.Ainsi $A = 20$, $g = 133 - 107 = 26$, $d = 133 - 84 = 49$ *et* $a = 5$.

Exemple

Dans notre exemple des exploitatations agricoles, après rectification des effectifs, on obtient le tableau suivant :

Superfice en ha	*Nombre d'exploitations*
$]0; 10]$	*24*
$]10; 15]$	*62*
$]15; 20]$	*107*
$]20; 25]$	*133*
$]25; 30]$	*84*
$]30; 40]$	*33*

La classe modale est donc la quatrième classe.Ainsi $A = 20$, $g = 133 - 107 = 26$, $d = 133 - 84 = 49$ *et* $a = 5$.*Il s'ensuit que*

$$M_o = 20 + \frac{26}{26 + 49} \cdot 5$$

Exemple

Dans notre exemple des exploitatations agricoles, après rectification des effectifs, on obtient le tableau suivant :

Superficie en ha	*Nombre d'exploitations*
$]0; 10]$	*24*
$]10; 15]$	*62*
$]15; 20]$	*107*
$]20; 25]$	*133*
$]25; 30]$	*84*
$]30; 40]$	*33*

La classe modale est donc la quatrième classe.Ainsi $A = 20$, $g = 133 - 107 = 26$, $d = 133 - 84 = 49$ *et* $a = 5$.*Il s'ensuit que*

$$M_o = 20 + \frac{26}{26 + 49} \cdot 5 \cong 21,73 \; ha.$$

Exemple

En 2016, un Suisse apprend par la presse que l'OFS estime le salaire brut moyen à 7491 francs. Il le compare avec son salaire qui se monte à 6942 francs et peste contre la pingrerie de son employeur chez qui il court réclamer une augmentation.

Exemple

En 2016, un Suisse apprend par la presse que l'OFS estime le salaire brut moyen à 7491 francs. Il le compare avec son salaire qui se monte à 6942 francs et peste contre la pingrerie de son employeur chez qui il court réclamer une augmentation.Mais le salaire moyen est-il un indicateur pertinent dans ce cas ? Sûrement pas. Il est basé sur un grand nombre de personnes gagnant peu et un nombre restreint de managers gagnant des salaires indécents se montant à plusieurs millions, entraînant ainsi une distorsion vers le haut du salaire moyen.

Exemple

En 2016, un Suisse apprend par la presse que l'OFS estime le salaire brut moyen à 7491 francs. Il le compare avec son salaire qui se monte à 6942 francs et peste contre la pingrerie de son employeur chez qui il court réclamer une augmentation.Mais le salaire moyen est-il un indicateur pertinent dans ce cas ? Sûrement pas. Il est basé sur un grand nombre de personnes gagnant peu et un nombre restreint de managers gagnant des salaires indécents se montant à plusieurs millions, entraînant ainsi une distorsion vers le haut du salaire moyen.Il faudrait plutôt que notre individu se pose la question de savoir s'il gagne plus ou moins que la plupart de ses compatriotes.

Exemple

En 2016, un Suisse apprend par la presse que l'OFS estime le salaire brut moyen à 7491 francs. Il le compare avec son salaire qui se monte à 6942 francs et peste contre la pingrerie de son employeur chez qui il court réclamer une augmentation.Mais le salaire moyen est-il un indicateur pertinent dans ce cas ? Sûrement pas. Il est basé sur un grand nombre de personnes gagnant peu et un nombre restreint de managers gagnant des salaires indécents se montant à plusieurs millions, entraînant ainsi une distorsion vers le haut du salaire moyen.Il faudrait plutôt que notre individu se pose la question de savoir s'il gagne plus ou moins que la plupart de ses compatriotes.Pour répondre à cette interrogation, on va considérer la médiane.

Exemple

En 2016, un Suisse apprend par la presse que l'OFS estime le salaire brut moyen à 7491 francs. Il le compare avec son salaire qui se monte à 6942 francs et peste contre la pingrerie de son employeur chez qui il court réclamer une augmentation.Mais le salaire moyen est-il un indicateur pertinent dans ce cas ? Sûrement pas. Il est basé sur un grand nombre de personnes gagnant peu et un nombre restreint de managers gagnant des salaires indécents se montant à plusieurs millions, entraînant ainsi une distorsion vers le haut du salaire moyen.Il faudrait plutôt que notre individu se pose la question de savoir s'il gagne plus ou moins que la plupart de ses compatriotes.Pour répondre à cette interrogation, on va considérer la médiane.Cette indice coupe la population en deux parties égales.

Exemple

En 2016, un Suisse apprend par la presse que l'OFS estime le salaire brut moyen à 7491 francs. Il le compare avec son salaire qui se monte à 6942 francs et peste contre la pingrerie de son employeur chez qui il court réclamer une augmentation.Mais le salaire moyen est-il un indicateur pertinent dans ce cas ? Sûrement pas. Il est basé sur un grand nombre de personnes gagnant peu et un nombre restreint de managers gagnant des salaires indécents se montant à plusieurs millions, entraînant ainsi une distorsion vers le haut du salaire moyen.Il faudrait plutôt que notre individu se pose la question de savoir s'il gagne plus ou moins que la plupart de ses compatriotes.Pour répondre à cette interrogation, on va considérer la médiane.Cette indice coupe la population en deux parties égales.La médiane des salaires bruts en Suisse étant de 6502 francs en 2016 selon l'OFS, il est plutôt favorisé puisqu'il fait partie de la moitié de la population qui gagne le plus !

Definition

La *médiane*, notée M_e, est la valeur du caractère qui partage en deux l'effectif total.

Definition

La *médiane*, notée M_e, est la valeur du caractère qui partage en deux l'effectif total.C'est la valeur du caractère qui correspond à une fréquence cumulée égale à 50%.

Definition

La *médiane*, notée M_e, est la valeur du caractère qui partage en deux l'effectif total.C'est la valeur du caractère qui correspond à une fréquence cumulée égale à 50%.Dans une population, il y a ainsi autant d'individus possédant une valeur du caractère inférieure au caractère médian que d'individus possédant une valeur du caractère supérieure à la médiane.

Definition

La *médiane*, notée M_e, est la valeur du caractère qui partage en deux l'effectif total.C'est la valeur du caractère qui correspond à une fréquence cumulée égale à 50%.Dans une population, il y a ainsi autant d'individus possédant une valeur du caractère inférieure au caractère médian que d'individus possédant une valeur du caractère supérieure à la médiane.
La *classe médiane* d'une variable continue est la première classe où la fréquence cumulée atteint ou dépasse 50%.

Definition

La *médiane*, notée M_e, est la valeur du caractère qui partage en deux l'effectif total.C'est la valeur du caractère qui correspond à une fréquence cumulée égale à 50%.Dans une population, il y a ainsi autant d'individus possédant une valeur du caractère inférieure au caractère médian que d'individus possédant une valeur du caractère supérieure à la médiane.
La *classe médiane* d'une variable continue est la première classe où la fréquence cumulée atteint ou dépasse 50%.

Exemple

1. *L'ensemble des nombres 3, 4, 4, 5, 6, 8, 8, 8, 10 a pour médiane* $M_e = 6$*.*

Definition

La *médiane*, notée M_e, est la valeur du caractère qui partage en deux l'effectif total.C'est la valeur du caractère qui correspond à une fréquence cumulée égale à 50%.Dans une population, il y a ainsi autant d'individus possédant une valeur du caractère inférieure au caractère médian que d'individus possédant une valeur du caractère supérieure à la médiane.
La *classe médiane* d'une variable continue est la première classe où la fréquence cumulée atteint ou dépasse 50%.

Exemple

1. *L'ensemble des nombres 3, 4, 4, 5, 6, 8, 8, 8, 10 a pour médiane* $M_e = 6$.
2. *L'effectif de l'ensemble 5, 5, 7, 9, 11, 12, 15, 18 étant pair, ce dernier a pour médiane* $M_e = \dfrac{9+11}{2} = 10$.

Remarque

On constate que la médiane correspond à la valeur du caractère de l'individu occupant le rang

$$m = \frac{N+1}{2}.$$

Remarque

On constate que la médiane correspond à la valeur du caractère de l'individu occupant le rang

$$m = \frac{N+1}{2}.$$

- *Si N est impair, il s'agit d'un individu réel occupant le rang entier m.*

Remarque

On constate que la médiane correspond à la valeur du caractère de l'individu occupant le rang

$$m = \frac{N+1}{2}.$$

- *Si N est impair, il s'agit d'un individu réel occupant le rang entier m.*
- *Si N est pair, il s'agit d'un individu virtuel placé entre les rangs $N/2$ et $N/2+1$.*

Exemple

Reprenons notre exemple du nombre de personnes par ménage. Pour déterminer la valeur de la médiane, il convient de calculer les effectifs cumulés croissants.

Exemple

Reprenons notre exemple du nombre de personnes par ménage. Pour déterminer la valeur de la médiane, il convient de calculer les effectifs cumulés croissants.

Modalités	Effectifs	Effectifs cumulés
1	5	5
2	9	14
3	15	29
4	10	39
5	6	45
6	3	48
8	2	50

Exemple

Reprenons notre exemple du nombre de personnes par ménage. Pour déterminer la valeur de la médiane, il convient de calculer les effectifs cumulés croissants.

Modalités	Effectifs	Effectifs cumulés
1	5	5
2	9	14
3	15	29
4	10	39
5	6	45
6	3	48
8	2	50

La médiane est comprise entre les rangs 25 et 26. Donc, $M_e = 3$.

Exemple

Reprenons notre exemple des exploitations agricoles.

Classes	*Effectifs*	*Effectifs cumulés croissants*
$]0;10]$	*48*	*48*
$]10;15]$	*62*	*110*
$]15;20]$	*107*	*217*
$]20;25]$	*133*	*350*
$]25;30]$	*84*	*434*
$]30;40]$	*66*	*500*

Exemple

La superficie médiane est comprise entre celles des $250^{ème}$ *et* $251^{ème}$ *individus.*

Exemple

La superficie médiane est comprise entre celles des $250^{ème}$ *et* $251^{ème}$ *individus.Ces deux exploitations appartiennent à la classe* $]20;25]$ *d'effectif 133.*

Exemple

La superficie médiane est comprise entre celles des $250^{ème}$ *et* $251^{ème}$ *individus.Ces deux exploitations appartiennent à la classe* $]20;25]$ *d'effectif 133.Comme ces derniers occupent respectivement les* $33^{ème}$ $(= 250 - 217)$ *et* $34^{ème}$ $(= 251 - 217)$ *rangs,*

Exemple

La superficie médiane est comprise entre celles des $250^{ème}$ *et* $251^{ème}$ *individus.Ces deux exploitations appartiennent à la classe* $]20;25]$ *d'effectif 133.Comme ces derniers occupent respectivement les* $33^{ème}$ $(= 250 - 217)$ *et* $34^{ème}$ $(= 251 - 217)$ *rangs,la médiane sera donc égale à*

$$M_e = 20 + \frac{33,5}{133} \cdot 5$$

Exemple

La superficie médiane est comprise entre celles des $250^{ème}$ *et* $251^{ème}$ *individus.Ces deux exploitations appartiennent à la classe* $]20; 25]$ *d'effectif 133.Comme ces derniers occupent respectivement les* $33^{ème}$ $(= 250 - 217)$ *et* $34^{ème}$ $(= 251 - 217)$ *rangs,la médiane sera donc égale à*

$$M_e = 20 + \frac{33,5}{133} \cdot 5 \cong 21,26 \ ha.$$

Exemple

La superficie médiane est comprise entre celles des $250^{ème}$ *et* $251^{ème}$ *individus.Ces deux exploitations appartiennent à la classe* $]20; 25]$ *d'effectif 133.Comme ces derniers occupent respectivement les* $33^{ème}$ $(= 250 - 217)$ *et* $34^{ème}$ $(= 251 - 217)$ *rangs,la médiane sera donc égale à*

$$M_e = 20 + \frac{33,5}{133} \cdot 5 \cong 21,26 \ ha.$$

Ce calcul repose sur l'hypothèse d'une répartition uniforme des 133 exploitations à l'intérieur de leur classe $]20; 25]$.

Exemple

La superficie médiane est comprise entre celles des $250^{ème}$ *et* $251^{ème}$ *individus.Ces deux exploitations appartiennent à la classe* $]20;25]$ *d'effectif 133.Comme ces derniers occupent respectivement les* $33^{ème}$ $(= 250 - 217)$ *et* $34^{ème}$ $(= 251 - 217)$ *rangs,la médiane sera donc égale à*

$$M_e = 20 + \frac{33,5}{133} \cdot 5 \cong 21,26 \; ha.$$

Ce calcul repose sur l'hypothèse d'une répartition uniforme des 133 exploitations à l'intérieur de leur classe $]20;25]$.
Graphiquement, la médiane est la première coordonnée du point d'intersection des polygones des effectifs cumulés croissants et cumulés décroissants. Dans notre exemple le résultat graphique concorde bien avec le calcul précédent.

Comparaison entre les valeurs centrales

Nous pouvons maintenant faire quelques comparaisons sommaires entre les trois mesures de tendance centrale.

1. Elle est sans doute la mesure de tendance centrale la plus familière.

1. Elle est sans doute la mesure de tendance centrale la plus familière.
2. Elle demande plus de calculs, mais s'exprime algébriquement d'une manière simple.

1. Elle est sans doute la mesure de tendance centrale la plus familière.
2. Elle demande plus de calculs, mais s'exprime algébriquement d'une manière simple.
3. Elle tient compte de toutes les données et est donc influencée par les données extrêmes de la distribution. Dans le cas où une distribution est fortement dissymétrique, ceci devient un inconvénient qui justifie l'usage de la médiane au lieu de la moyenne.

1. Elle est sans doute la mesure de tendance centrale la plus familière.
2. Elle demande plus de calculs, mais s'exprime algébriquement d'une manière simple.
3. Elle tient compte de toutes les données et est donc influencée par les données extrêmes de la distribution. Dans le cas où une distribution est fortement dissymétrique, ceci devient un inconvénient qui justifie l'usage de la médiane au lieu de la moyenne.
4. Elle est peu influencée par le choix des classes, mais ne peut cependant pas être calculée s'il y a une classe ouverte (par exemple une classe du type "80 ans et plus").

1. Elle est sans doute la mesure de tendance centrale la plus familière.
2. Elle demande plus de calculs, mais s'exprime algébriquement d'une manière simple.
3. Elle tient compte de toutes les données et est donc influencée par les données extrêmes de la distribution. Dans le cas où une distribution est fortement dissymétrique, ceci devient un inconvénient qui justifie l'usage de la médiane au lieu de la moyenne.
4. Elle est peu influencée par le choix des classes, mais ne peut cependant pas être calculée s'il y a une classe ouverte (par exemple une classe du type "80 ans et plus").
5. Elle se prête facilement aux manipulations algébriques à cause de son expression mathématique simple.

1. Elle est sans doute la mesure de tendance centrale la plus familière.
2. Elle demande plus de calculs, mais s'exprime algébriquement d'une manière simple.
3. Elle tient compte de toutes les données et est donc influencée par les données extrêmes de la distribution. Dans le cas où une distribution est fortement dissymétrique, ceci devient un inconvénient qui justifie l'usage de la médiane au lieu de la moyenne.
4. Elle est peu influencée par le choix des classes, mais ne peut cependant pas être calculée s'il y a une classe ouverte (par exemple une classe du type "80 ans et plus").
5. Elle se prête facilement aux manipulations algébriques à cause de son expression mathématique simple.
6. Sa valeur est stable, c'est-à-dire qu'elle varie peu d'un échantillon à l'autre, du fait qu'elle tient compte de toutes les données. C'est sa plus grande qualité pour faire de l'inférence statistique.

1. Elle est sans doute la mesure de tendance centrale la plus familière.
2. Elle demande plus de calculs, mais s'exprime algébriquement d'une manière simple.
3. Elle tient compte de toutes les données et est donc influencée par les données extrêmes de la distribution. Dans le cas où une distribution est fortement dissymétrique, ceci devient un inconvénient qui justifie l'usage de la médiane au lieu de la moyenne.
4. Elle est peu influencée par le choix des classes, mais ne peut cependant pas être calculée s'il y a une classe ouverte (par exemple une classe du type "80 ans et plus").
5. Elle se prête facilement aux manipulations algébriques à cause de son expression mathématique simple.
6. Sa valeur est stable, c'est-à-dire qu'elle varie peu d'un échantillon à l'autre, du fait qu'elle tient compte de toutes les données. C'est sa plus grande qualité pour faire de l'inférence statistique.
7. Il s'agit de la mesure de tendance centrale la plus utilisée.

1. Il peut y en avoir plusieurs dans une distribution. Le cas échéant, la présence de plusieurs modes peut être une indication que la population étudiée se compose de sous-groupes distincts. Selon l'étude désirée, cela pourrait inviter à scinder la population.

1. Il peut y en avoir plusieurs dans une distribution. Le cas échéant, la présence de plusieurs modes peut être une indication que la population étudiée se compose de sous-groupes distincts. Selon l'étude désirée, cela pourrait inviter à scinder la population.
2. Il est facile à déterminer.

1. Il peut y en avoir plusieurs dans une distribution. Le cas échéant, la présence de plusieurs modes peut être une indication que la population étudiée se compose de sous-groupes distincts. Selon l'étude désirée, cela pourrait inviter à scinder la population.
2. Il est facile à déterminer.
3. Il ne tient pas compte de toutes les données. Par contre, il n'est pas influencé par les données extrêmes de la distribution.

1. Il peut y en avoir plusieurs dans une distribution. Le cas échéant, la présence de plusieurs modes peut être une indication que la population étudiée se compose de sous-groupes distincts. Selon l'étude désirée, cela pourrait inviter à scinder la population.
2. Il est facile à déterminer.
3. Il ne tient pas compte de toutes les données. Par contre, il n'est pas influencé par les données extrêmes de la distribution.
4. Il peut être grandement influencé par le choix des classes.

1. Il peut y en avoir plusieurs dans une distribution. Le cas échéant, la présence de plusieurs modes peut être une indication que la population étudiée se compose de sous-groupes distincts. Selon l'étude désirée, cela pourrait inviter à scinder la population.
2. Il est facile à déterminer.
3. Il ne tient pas compte de toutes les données. Par contre, il n'est pas influencé par les données extrêmes de la distribution.
4. Il peut être grandement influencé par le choix des classes.
5. Il n'est vraiment significatif que si l'effectif correspondant est nettement supérieur aux autres.

1. Il peut y en avoir plusieurs dans une distribution. Le cas échéant, la présence de plusieurs modes peut être une indication que la population étudiée se compose de sous-groupes distincts. Selon l'étude désirée, cela pourrait inviter à scinder la population.
2. Il est facile à déterminer.
3. Il ne tient pas compte de toutes les données. Par contre, il n'est pas influencé par les données extrêmes de la distribution.
4. Il peut être grandement influencé par le choix des classes.
5. Il n'est vraiment significatif que si l'effectif correspondant est nettement supérieur aux autres.
6. Sa valeur n'est pas stable, c'est-à-dire qu'elle varie beaucoup d'un échantillon à l'autre choisi dans la même population.

1. Il peut y en avoir plusieurs dans une distribution. Le cas échéant, la présence de plusieurs modes peut être une indication que la population étudiée se compose de sous-groupes distincts. Selon l'étude désirée, cela pourrait inviter à scinder la population.
2. Il est facile à déterminer.
3. Il ne tient pas compte de toutes les données. Par contre, il n'est pas influencé par les données extrêmes de la distribution.
4. Il peut être grandement influencé par le choix des classes.
5. Il n'est vraiment significatif que si l'effectif correspondant est nettement supérieur aux autres.
6. Sa valeur n'est pas stable, c'est-à-dire qu'elle varie beaucoup d'un échantillon à l'autre choisi dans la même population.
7. Il s'agit de la mesure de tendance centrale la moins utilisée.

1. Elle provient d'une conception simple de la notion de centre.

1. Elle provient d'une conception simple de la notion de centre.
2. Elle n'est pas très difficile à calculer, mais elle est plus difficile à exprimer algébriquement que la moyenne arithmétique.

1. Elle provient d'une conception simple de la notion de centre.
2. Elle n'est pas très difficile à calculer, mais elle est plus difficile à exprimer algébriquement que la moyenne arithmétique.
3. Elle ne tient pas compte de toutes les données, mais uniquement de la position des données. Elle n'est donc pas influencée par les données extrêmes de la distribution.

1. Elle provient d'une conception simple de la notion de centre.
2. Elle n'est pas très difficile à calculer, mais elle est plus difficile à exprimer algébriquement que la moyenne arithmétique.
3. Elle ne tient pas compte de toutes les données, mais uniquement de la position des données. Elle n'est donc pas influencée par les données extrêmes de la distribution.
4. Elle peut être influencée par le choix des classes.

1. Elle provient d'une conception simple de la notion de centre.
2. Elle n'est pas très difficile à calculer, mais elle est plus difficile à exprimer algébriquement que la moyenne arithmétique.
3. Elle ne tient pas compte de toutes les données, mais uniquement de la position des données. Elle n'est donc pas influencée par les données extrêmes de la distribution.
4. Elle peut être influencée par le choix des classes.
5. Elle est surtout utilisée lorsque la distribution des effectifs est fortement dissymétrique ou lorsqu'elle contient des classes ouvertes.

1. Elle provient d'une conception simple de la notion de centre.
2. Elle n'est pas très difficile à calculer, mais elle est plus difficile à exprimer algébriquement que la moyenne arithmétique.
3. Elle ne tient pas compte de toutes les données, mais uniquement de la position des données. Elle n'est donc pas influencée par les données extrêmes de la distribution.
4. Elle peut être influencée par le choix des classes.
5. Elle est surtout utilisée lorsque la distribution des effectifs est fortement dissymétrique ou lorsqu'elle contient des classes ouvertes.
6. Sa valeur est moins stable que celle de la moyenne. Ceci s'explique par le fait que cette valeur ne dépend que de quelques données parmi celles choisies dans un échantillon.

1. Elle provient d'une conception simple de la notion de centre.
2. Elle n'est pas très difficile à calculer, mais elle est plus difficile à exprimer algébriquement que la moyenne arithmétique.
3. Elle ne tient pas compte de toutes les données, mais uniquement de la position des données. Elle n'est donc pas influencée par les données extrêmes de la distribution.
4. Elle peut être influencée par le choix des classes.
5. Elle est surtout utilisée lorsque la distribution des effectifs est fortement dissymétrique ou lorsqu'elle contient des classes ouvertes.
6. Sa valeur est moins stable que celle de la moyenne. Ceci s'explique par le fait que cette valeur ne dépend que de quelques données parmi celles choisies dans un échantillon.
7. Elle est plus utilisée que le mode, mais moins que la moyenne arithmétique.

Definition

On appelle *quartiles* les valeurs du caractère qui partagent l'effectif total de la série en 4 groupes d'effectifs égaux. On les note Q_1, Q_2 et Q_3.

Definition

On appelle *quartiles* les valeurs du caractère qui partagent l'effectif total de la série en 4 groupes d'effectifs égaux. On les note Q_1, Q_2 et Q_3.Un quart de l'effectif total possède donc un caractère inférieur à Q_1.

Definition

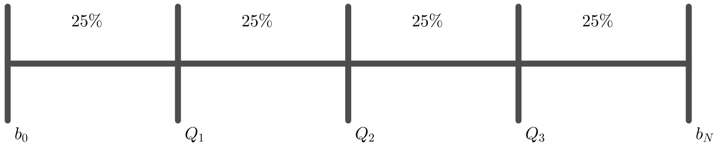
On appelle *quartiles* les valeurs du caractère qui partagent l'effectif total de la série en 4 groupes d'effectifs égaux. On les note Q_1, Q_2 et Q_3.Un quart de l'effectif total possède donc un caractère inférieur à Q_1.Le deuxième quartile $Q_2 = M_e$ n'est autre que la médiane.

Definition

On appelle *quartiles* les valeurs du caractère qui partagent l'effectif total de la série en 4 groupes d'effectifs égaux. On les note Q_1, Q_2 et Q_3.Un quart de l'effectif total possède donc un caractère inférieur à Q_1.Le deuxième quartile $Q_2 = M_e$ n'est autre que la médiane.Enfin, les trois quarts de la population se trouvent en dessous de la valeur définie par le troisième quartile Q_3.

Definition

On appelle *quartiles* les valeurs du caractère qui partagent l'effectif total de la série en 4 groupes d'effectifs égaux. On les note Q_1, Q_2 et Q_3.Un quart de l'effectif total possède donc un caractère inférieur à Q_1.Le deuxième quartile $Q_2 = M_e$ n'est autre que la médiane.Enfin, les trois quarts de la population se trouvent en dessous de la valeur définie par le troisième quartile Q_3.



Remarque

Il faut être attentif au fait qu'il existe de nombreuses méthodes différentes pour déterminer les quartiles, qui ne conduisent pas aux mêmes résultats. Dans ce cours, nous calculerons les quartiles selon la méthode établie par John Tukey en 1983.

Remarque

Il faut être attentif au fait qu'il existe de nombreuses méthodes différentes pour déterminer les quartiles, qui ne conduisent pas aux mêmes résultats. Dans ce cours, nous calculerons les quartiles selon la méthode établie par John Tukey en 1983.

On classe les N données dans l'ordre croissant et on les coupe en deux ensembles sur lesquels on calcule la médiane.

Remarque

Il faut être attentif au fait qu'il existe de nombreuses méthodes différentes pour déterminer les quartiles, qui ne conduisent pas aux mêmes résultats. Dans ce cours, nous calculerons les quartiles selon la méthode établie par John Tukey en 1983.

On classe les N données dans l'ordre croissant et on les coupe en deux ensembles sur lesquels on calcule la médiane.

- Si N est impair, il y a une valeur centrale (la médiane), et on coupe les données en deux sous-ensembles en mettant la médiane dans chacun des deux ensembles. Le quartile Q_1 est alors la médiane du premier sous-ensemble ; le quartile Q_3 est la médiane du deuxième sous-ensemble.

Remarque

Il faut être attentif au fait qu'il existe de nombreuses méthodes différentes pour déterminer les quartiles, qui ne conduisent pas aux mêmes résultats. Dans ce cours, nous calculerons les quartiles selon la méthode établie par John Tukey en 1983.

On classe les N données dans l'ordre croissant et on les coupe en deux ensembles sur lesquels on calcule la médiane.

- Si N est impair, il y a une valeur centrale (la médiane), et on coupe les données en deux sous-ensembles en mettant la médiane dans chacun des deux ensembles. Le quartile Q_1 est alors la médiane du premier sous-ensemble ; le quartile Q_3 est la médiane du deuxième sous-ensemble.
- Si N est pair, il y a deux valeurs centrales (la médiane est la moyenne arithmétique de ces deux valeurs), et on coupe en deux sous-ensembles en mettant dans chaque sous-ensemble la valeur centrale correspondante. Le quartile Q_1 est alors la médiane du premier sous-ensemble ; le quartile Q_3 est la médiane du deuxième sous-ensemble.

Exemple

N	*Mesures*	*Sous-ensembles*	Q_1	Q_3	*Rang* Q_1	*Rang* Q_3
4	*1 3 4 5*	$\{1;3\}$ *et* $\{4;5\}$	2	4,5	1,5	3,5
5	*1 3 5 5 7*	$\{1;3;5\}$ *et* $\{5;5;7\}$	3	5	2	4
6	*1 3 4 6 7 9*	$\{1;3;4\}$ *et* $\{6;7;9\}$	3	7	2	5
7	*1 3 5 6 7 9 15*	$\{1;3;5;6\}$ *et* $\{6;7;9;15\}$	4	8	2,5	5,5

Théorème

- *Si N est pair, le rang du quartile Q_1 est donné par* $\dfrac{N+2}{4}$ *et celui de Q_3 par* $\dfrac{3N+2}{4}$.

Exemple

N	Mesures	Sous-ensembles	Q_1	Q_3	Rang Q_1	Rang Q_3
4	1 3 4 5	$\{1;3\}$ et $\{4;5\}$	2	4,5	1,5	3,5
5	1 3 5 5 7	$\{1;3;5\}$ et $\{5;5;7\}$	3	5	2	4
6	1 3 4 6 7 9	$\{1;3;4\}$ et $\{6;7;9\}$	3	7	2	5
7	1 3 5 6 7 9 15	$\{1;3;5;6\}$ et $\{6;7;9;15\}$	4	8	2,5	5,5

Théorème

- *Si N est pair, le rang du quartile Q_1 est donné par $\dfrac{N+2}{4}$ et celui de Q_3 par $\dfrac{3N+2}{4}$.*
- *Si N est impair, le rang du quartile Q_1 est donné par $\dfrac{N+3}{4}$ et celui de Q_3 par $\dfrac{3N+1}{4}$.*

Exemple

Reprenons notre exemple du nombre de personnes par ménage.

Exemple

Reprenons notre exemple du nombre de personnes par ménage.

Modalités	Effectifs	Effectifs cumulés	Fréquences (en %)	Fréquences cumulées
1	5	5	10	10
2	9	14	18	28
3	15	29	30	58
4	10	39	20	78
5	6	45	12	90
6	3	48	6	96
8	2	50	4	100

Exemple

Reprenons notre exemple du nombre de personnes par ménage.

Modalités	Effectifs	Effectifs cumulés	Fréquences (en %)	Fréquences cumulées
1	5	5	10	10
2	9	14	18	28
3	15	29	30	58
4	10	39	20	78
5	6	45	12	90
6	3	48	6	96
8	2	50	4	100

On connaît déjà $Q_2 = M_e = 3$. *Le quartile* Q_1 *est la valeur de l'observation de rang* $\dfrac{50+2}{4} = 13$. *Donc* $Q_1 = 2$. *Quant au quartile* Q_3, *il est égal à la valeur de l'observation de rang* $\dfrac{3 \cdot 50 + 2}{4} = 38$. *Donc* $Q_3 = 4$.

Remarque

La plupart du temps, lorsqu'il s'agit par exemple de définir les quartiles, il n'est pas possible de trouver des rangs qui divisent la population en quatre classes d'effectifs égaux. Dans ce cas, on convertit les effectifs en fréquences et on définit les quartiles Q_1, M_e *et* Q_3 *par les valeurs du caractère associées aux fréquences cumulées* 25%, 50% *et* 75%.

Remarque

La plupart du temps, lorsqu'il s'agit par exemple de définir les quartiles, il n'est pas possible de trouver des rangs qui divisent la population en quatre classes d'effectifs égaux. Dans ce cas, on convertit les effectifs en fréquences et on définit les quartiles Q_1*,* M_e *et* Q_3 *par les valeurs du caractère associées aux fréquences cumulées* 25%, 50% *et* 75%.

Remarque

Il est possible de généraliser la notion de quartile à celle de quantile d'ordre n. Les autres quantiles les plus souvent utilisés sont :

Remarque

La plupart du temps, lorsqu'il s'agit par exemple de définir les quartiles, il n'est pas possible de trouver des rangs qui divisent la population en quatre classes d'effectifs égaux. Dans ce cas, on convertit les effectifs en fréquences et on définit les quartiles Q_1, M_e *et* Q_3 *par les valeurs du caractère associées aux fréquences cumulées* 25%, 50% *et* 75%.

Remarque

Il est possible de généraliser la notion de quartile à celle de quantile d'ordre n. Les autres quantiles les plus souvent utilisés sont :

- *Les déciles* D_1, D_2, ..., D_9 *partagent l'effectif total en dix groupes égaux. Un dixième de la population a un caractère inférieur à* D_1, *et neuf dixièmes ont un caractère supérieur à* $D_1, \ldots$, *et ainsi de suite. Le décile* D_5 *est égal à la médiane ;*

Remarque

La plupart du temps, lorsqu'il s'agit par exemple de définir les quartiles, il n'est pas possible de trouver des rangs qui divisent la population en quatre classes d'effectifs égaux. Dans ce cas, on convertit les effectifs en fréquences et on définit les quartiles Q_1, M_e *et* Q_3 *par les valeurs du caractère associées aux fréquences cumulées* 25%, 50% *et* 75%.

Remarque

Il est possible de généraliser la notion de quartile à celle de quantile d'ordre n. Les autres quantiles les plus souvent utilisés sont :

- *Les déciles* D_1, D_2, ..., D_9 *partagent l'effectif total en dix groupes égaux. Un dixième de la population a un caractère inférieur à* D_1, *et neuf dixièmes ont un caractère supérieur à* D_1, . . ., *et ainsi de suite. Le décile* D_5 *est égal à la médiane ;*
- *Les centiles* C_1, C_2,. . . , C_{99} *partagent la population en 100 groupes d'effectifs égaux.*

Exemple

Calculons les quartiles pour notre exemple des exploitations agricoles.

Classes	Effectifs	Effectifs cumulés croissants
]0; 10]	48	48
]10; 15]	62	110
]15; 20]	107	217
]20; 25]	133	350
]25; 30]	84	434
]30; 40]	66	500

Exemple

On connait déjà la mediane $Q_2 = M_e \cong 21,26$.

Exemple

On connait déjà la mediane $Q_2 = M_e \cong 21,26$.
Le premier quartile Q_1 *est la valeur de la superficie de l'exploitation de rang* $\dfrac{500+2}{4} = 125,5$.

Exemple

On connait déjà la mediane $Q_2 = M_e \cong 21,26$.
Le premier quartile Q_1 *est la valeur de la superficie de l'exploitation de rang* $\dfrac{500+2}{4} = 125,5$. *Comme elle se trouve dans la troisieme classe, d'effectif 107, on a*

Exemple

On connait déjà la mediane $Q_2 = M_e \cong 21,26$.
Le premier quartile Q_1 *est la valeur de la superficie de l'exploitation de rang* $\dfrac{500+2}{4} = 125,5$.*Comme elle se trouve dans la troisieme classe, d'effectif 107, on a*

$$Q_1 = 15 + 5 \cdot \frac{125,5 - 110}{107}$$

Exemple

On connait déjà la mediane $Q_2 = M_e \cong 21,26$.
Le premier quartile Q_1 *est la valeur de la superficie de l'exploitation de rang* $\dfrac{500+2}{4} = 125,5$.*Comme elle se trouve dans la troisieme classe, d'effectif 107, on a*

$$Q_1 = 15 + 5 \cdot \frac{125,5 - 110}{107} \cong 15,72 \; ha.$$

Le troisième quartile Q_3 *est défini par est la valeur de la superficie de l'exploitation de rang* $\dfrac{3 \cdot 500 + 2}{4} = 375,5$.

Exemple

On connait déjà la mediane $Q_2 = M_e \cong 21,26$.
Le premier quartile Q_1 *est la valeur de la superficie de l'exploitation de rang* $\dfrac{500+2}{4} = 125,5$.*Comme elle se trouve dans la troisieme classe, d'effectif 107, on a*

$$Q_1 = 15 + 5 \cdot \frac{125,5 - 110}{107} \cong 15,72 \ ha.$$

Le troisième quartile Q_3 *est défini par est la valeur de la superficie de l'exploitation de rang* $\dfrac{3 \cdot 500 + 2}{4} = 375,5$.*Celle-ci se trouvant en position 25,5 dans la classe* $]25;30]$, *d'effectif 84,*

Exemple

On connait déjà la mediane $Q_2 = M_e \cong 21,26$*.*
Le premier quartile Q_1 *est la valeur de la superficie de l'exploitation de rang* $\dfrac{500+2}{4} = 125,5$*.Comme elle se trouve dans la troisieme classe, d'effectif 107, on a*

$$Q_1 = 15 + 5 \cdot \frac{125,5 - 110}{107} \cong 15,72 \; ha.$$

Le troisième quartile Q_3 *est défini par est la valeur de la superficie de l'exploitation de rang* $\dfrac{3 \cdot 500 + 2}{4} = 375,5$*.Celle-ci se trouvant en position 25,5 dans la classe* $]25;30]$*, d'effectif 84,on a*

$$Q_3 = 25 + 5 \cdot \frac{25,5}{84}$$

Exemple

On connait déjà la mediane $Q_2 = M_e \cong 21,26$.
Le premier quartile Q_1 *est la valeur de la superficie de l'exploitation de rang* $\dfrac{500+2}{4} = 125,5$.*Comme elle se trouve dans la troisieme classe, d'effectif 107, on a*

$$Q_1 = 15 + 5 \cdot \frac{125,5 - 110}{107} \cong 15,72 \; ha.$$

Le troisième quartile Q_3 *est défini par est la valeur de la superficie de l'exploitation de rang* $\dfrac{3 \cdot 500 + 2}{4} = 375,5$.*Celle-ci se trouvant en position 25,5 dans la classe* $]25;30]$, *d'effectif 84,on a*

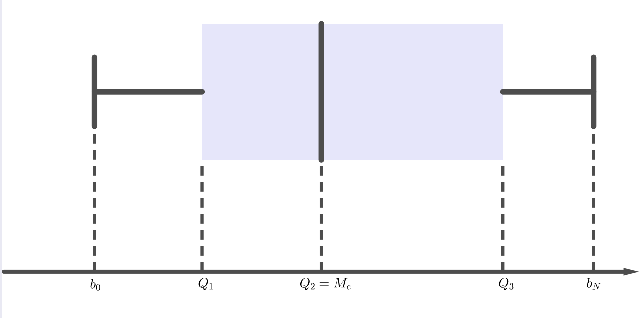
$$Q_3 = 25 + 5 \cdot \frac{25,5}{84} \cong 26,52 \; ha.$$

Definition

Le *diagramme de Tukey*, plus communément appelé *boîte à moustaches* ou *box plot*, est une représentation codifiée des quantiles Q_1, M_e, Q_3 et des valeurs extrêmes b_0 et b_N de la distribution qui donne une information graphique concernant la symétrie de la distribution.

Definition

Le *diagramme de Tukey*, plus communément appelé *boîte à moustaches* ou *box plot*, est une représentation codifiée des quantiles Q_1, M_e, Q_3 et des valeurs extrêmes b_0 et b_N de la distribution qui donne une information graphique concernant la symétrie de la distribution.

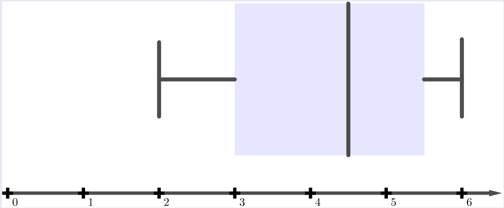


Exemple

Les notes d'une classe ont été représentées à l'aide de la boîte à moustache ci-dessous

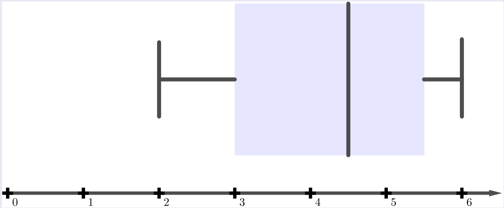
Exemple

Les notes d'une classe ont été représentées à l'aide de la boîte à moustache ci-dessous



Exemple

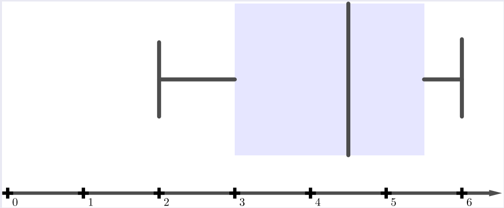
Les notes d'une classe ont été représentées à l'aide de la boîte à moustache ci-dessous



Cette boîte à moustaches fournit les informations suivantes :

Exemple

Les notes d'une classe ont été représentées à l'aide de la boîte à moustache ci-dessous

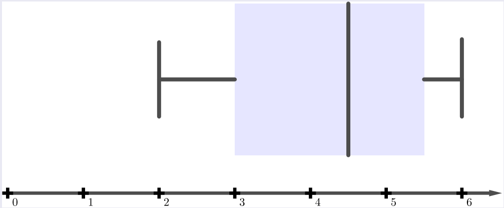


Cette boîte à moustaches fournit les informations suivantes :

- *La moins bonne note est 2 et la meilleure 6.*

Exemple

Les notes d'une classe ont été représentées à l'aide de la boîte à moustache ci-dessous

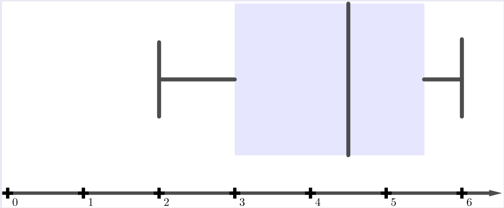


Cette boîte à moustaches fournit les informations suivantes :

- *La moins bonne note est 2 et la meilleure 6.*
- 25% *des élèves ont fait une note égale ou inférieure à 3.*

Exemple

Les notes d'une classe ont été représentées à l'aide de la boîte à moustache ci-dessous

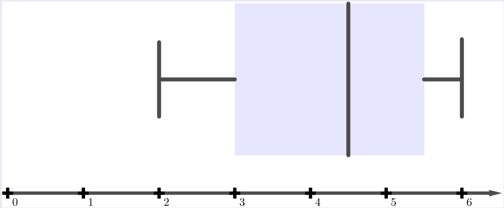


Cette boîte à moustaches fournit les informations suivantes :

- *La moins bonne note est 2 et la meilleure 6.*
- *25% des élèves ont fait une note égale ou inférieure à 3.*
- *La moitié des élèves ont fait* $4,5$ *au moins (et au plus !).*

Exemple

Les notes d'une classe ont été représentées à l'aide de la boîte à moustache ci-dessous

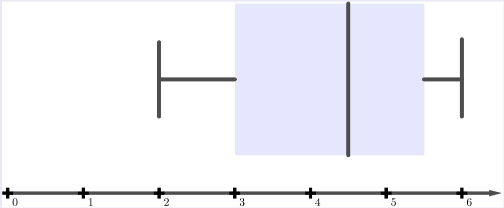


Cette boîte à moustaches fournit les informations suivantes :

- *La moins bonne note est 2 et la meilleure 6.*
- *25% des élèves ont fait une note égale ou inférieure à 3.*
- *La moitié des élèves ont fait* $4,5$ *au moins (et au plus !).*
- *75% des élèves ont fait une note inférieure ou égale à* $5,5$.

Exemple

Les notes d'une classe ont été représentées à l'aide de la boîte à moustache ci-dessous



Cette boîte à moustaches fournit les informations suivantes :

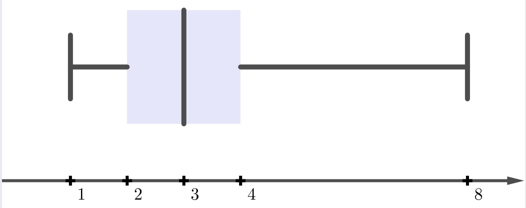
- *La moins bonne note est 2 et la meilleure 6.*
- 25% *des élèves ont fait une note égale ou inférieure à 3.*
- *La moitié des élèves ont fait* $4,5$ *au moins (et au plus !).*
- 75% *des élèves ont fait une note inférieure ou égale à* $5,5$.
- 50% *se tiennent dans un écart de* $2,5$.

Exemple

Dans notre exemple du nombre de personnes par ménage, la boîte à moustaches est donnée par

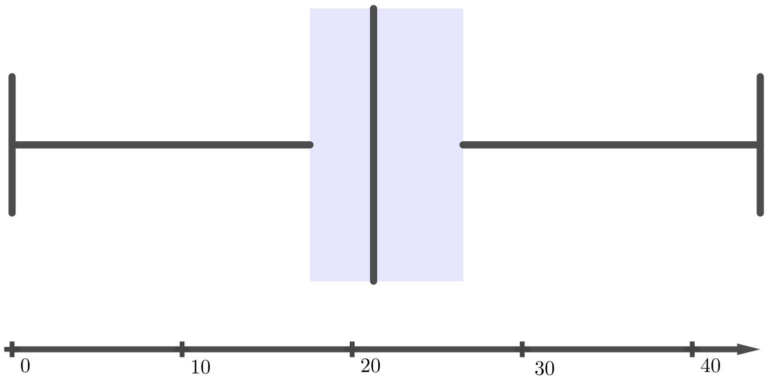
Exemple

Dans notre exemple du nombre de personnes par ménage, la boîte à moustaches est donnée par



Exemple

Dans notre exemple des exploitations agricoles, la distribution étant presque symétrique, la boîte à moustaches s'étale symétriquement sur l'intervalle $[0; 40]$.



Si les valeurs centrales sont généralement nécessaires pour caractériser une série, elles ne sont toutefois pas suffisantes.

Si les valeurs centrales sont généralement nécessaires pour caractériser une série, elles ne sont toutefois pas suffisantes.Deux populations différentes peuvent avoir les mêmes valeurs centrales et différer notablement quant à la dispersion des individus autour de ces valeurs centrales.

Si les valeurs centrales sont généralement nécessaires pour caractériser une série, elles ne sont toutefois pas suffisantes.Deux populations différentes peuvent avoir les mêmes valeurs centrales et différer notablement quant à la dispersion des individus autour de ces valeurs centrales.
Les deux ensembles $A = \{6; 8; 10; 12; 14\}$ et $B = \{2; 6; 10; 14; 18\}$ ont, par exemple, la même moyenne arithmétique et la même médiane, à savoir 10.

Si les valeurs centrales sont généralement nécessaires pour caractériser une série, elles ne sont toutefois pas suffisantes.Deux populations différentes peuvent avoir les mêmes valeurs centrales et différer notablement quant à la dispersion des individus autour de ces valeurs centrales.
Les deux ensembles $A = \{6; 8; 10; 12; 14\}$ et $B = \{2; 6; 10; 14; 18\}$ ont, par exemple, la même moyenne arithmétique et la même médiane, à savoir 10.Pourtant, les individus qui les composent ne sont pas répartis de la même manière autour de cette valeur centrale.

Si les valeurs centrales sont généralement nécessaires pour caractériser une série, elles ne sont toutefois pas suffisantes.Deux populations différentes peuvent avoir les mêmes valeurs centrales et différer notablement quant à la dispersion des individus autour de ces valeurs centrales.
Les deux ensembles $A = \{6; 8; 10; 12; 14\}$ et $B = \{2; 6; 10; 14; 18\}$ ont, par exemple, la même moyenne arithmétique et la même médiane, à savoir 10.Pourtant, les individus qui les composent ne sont pas répartis de la même manière autour de cette valeur centrale.L'ensemble B est moins régulier ou plus dispersé que l'ensemble A.

Si les valeurs centrales sont généralement nécessaires pour caractériser une série, elles ne sont toutefois pas suffisantes.Deux populations différentes peuvent avoir les mêmes valeurs centrales et différer notablement quant à la dispersion des individus autour de ces valeurs centrales.

Les deux ensembles $A = \{6; 8; 10; 12; 14\}$ et $B = \{2; 6; 10; 14; 18\}$ ont, par exemple, la même moyenne arithmétique et la même médiane, à savoir 10.Pourtant, les individus qui les composent ne sont pas répartis de la même manière autour de cette valeur centrale.L'ensemble B est moins régulier ou plus dispersé que l'ensemble A.On dit que A et B n'ont pas la même dispersion.

Pour comparer deux populations, on considère, outre leurs valeurs centrales, des mesures de leur dispersion.

Si les valeurs centrales sont généralement nécessaires pour caractériser une série, elles ne sont toutefois pas suffisantes.Deux populations différentes peuvent avoir les mêmes valeurs centrales et différer notablement quant à la dispersion des individus autour de ces valeurs centrales.

Les deux ensembles $A = \{6; 8; 10; 12; 14\}$ et $B = \{2; 6; 10; 14; 18\}$ ont, par exemple, la même moyenne arithmétique et la même médiane, à savoir 10.Pourtant, les individus qui les composent ne sont pas répartis de la même manière autour de cette valeur centrale.L'ensemble B est moins régulier ou plus dispersé que l'ensemble A.On dit que A et B n'ont pas la même dispersion.

Pour comparer deux populations, on considère, outre leurs valeurs centrales, des mesures de leur dispersion.Les mesures classiques de dispersion sont les suivantes :

Si les valeurs centrales sont généralement nécessaires pour caractériser une série, elles ne sont toutefois pas suffisantes.Deux populations différentes peuvent avoir les mêmes valeurs centrales et différer notablement quant à la dispersion des individus autour de ces valeurs centrales.
Les deux ensembles $A = \{6; 8; 10; 12; 14\}$ et $B = \{2; 6; 10; 14; 18\}$ ont, par exemple, la même moyenne arithmétique et la même médiane, à savoir 10.Pourtant, les individus qui les composent ne sont pas répartis de la même manière autour de cette valeur centrale.L'ensemble B est moins régulier ou plus dispersé que l'ensemble A.On dit que A et B n'ont pas la même dispersion.
Pour comparer deux populations, on considère, outre leurs valeurs centrales, des mesures de leur dispersion.Les mesures classiques de dispersion sont les suivantes :*l'étendue, la variance et l'écart-type*.

Etendue de la série

C’est la valeur de dispersion la plus simple.

Etendue de la série

C'est la valeur de dispersion la plus simple.

Definition

Aussi appelée *intervalle de variation*, *amplitude de la série* ou *intervalle maximal*, *l'étendue E* est la différence des valeurs extrêmes de la série.

C'est la valeur de dispersion la plus simple.

Definition

Aussi appelée *intervalle de variation*, *amplitude de la série* ou *intervalle maximal*, *l'étendue E* est la différence des valeurs extrêmes de la série.

Exemple

Dans notre exemple des exploitations agricoles, l'étendue vaut $E = 40 - 0 = 40$ *ha.*

Etendue de la série

C'est la valeur de dispersion la plus simple.

Definition

Aussi appelée *intervalle de variation*, *amplitude de la série* ou *intervalle maximal*, *l'étendue E* est la différence des valeurs extrêmes de la série.

Exemple

Dans notre exemple des exploitations agricoles, l'étendue vaut $E = 40 - 0 = 40$ *ha.*

Remarque

Simple à calculer, cette mesure de dispersion n'est pas très fiable puisqu'elle ne tient compte que de deux observations marginales et néglige toutes les autres.

Ecart interquartile

Definition

L'écart interquartile I_Q est défini par la différence des quartiles extrêmes.

Definition

L'écart interquartile I_Q est défini par la différence des quartiles extrêmes.Autrement dit, on

Definition

L'écart interquartile I_Q est défini par la différence des quartiles extrêmes.Autrement dit, ona

$$\boxed{I_Q = Q_3 - Q_1.}$$

Ecart interquartile

Definition

L'écart interquartile I_Q est défini par la différence des quartiles extrêmes.Autrement dit, ona

$$\boxed{I_Q = Q_3 - Q_1.}$$

Remarque

Cette mesure est plus fiable que l'étendue puisqu'elle exclut les 50% *des valeurs marginales inférieures et supérieures.*

Ecart interquartile

Definition

L'écart interquartile I_Q est défini par la différence des quartiles extrêmes.Autrement dit, ona

$$\boxed{I_Q = Q_3 - Q_1 .}$$

Remarque

Cette mesure est plus fiable que l'étendue puisqu'elle exclut les 50% *des valeurs marginales inférieures et supérieures.*

Definition

L'écart semi-interquartile Q est défini par la moyenne arithmétique des écarts entre les quartiles et la médiane.

Definition

L'écart interquartile I_Q est défini par la différence des quartiles extrêmes.Autrement dit, ona

$$\boxed{I_Q = Q_3 - Q_1.}$$

Remarque

Cette mesure est plus fiable que l'étendue puisqu'elle exclut les 50% *des valeurs marginales inférieures et supérieures.*

Definition

L'écart semi-interquartile Q est défini par la moyenne arithmétique des écarts entre les quartiles et la médiane.Autrement dit, on a

$$Q$$

Ecart interquartile

Definition

L'écart interquartile I_Q est défini par la différence des quartiles extrêmes.Autrement dit, ona

$$\boxed{I_Q = Q_3 - Q_1 .}$$

Remarque

Cette mesure est plus fiable que l'étendue puisqu'elle exclut les 50% *des valeurs marginales inférieures et supérieures.*

Definition

L'écart semi-interquartile Q est défini par la moyenne arithmétique des écarts entre les quartiles et la médiane.Autrement dit, on a

$$Q = \frac{(Q_3 - M_e) + (M_e - Q_1)}{2}$$

Definition

L'écart interquartile I_Q est défini par la différence des quartiles extrêmes.Autrement dit, ona

$$\boxed{I_Q = Q_3 - Q_1.}$$

Remarque

Cette mesure est plus fiable que l'étendue puisqu'elle exclut les 50% *des valeurs marginales inférieures et supérieures.*

Definition

L'écart semi-interquartile Q est défini par la moyenne arithmétique des écarts entre les quartiles et la médiane.Autrement dit, on a

$$Q = \frac{(Q_3 - M_e) + (M_e - Q_1)}{2} = \frac{Q_3 - Q_1}{2}$$

Definition

L'écart interquartile I_Q est défini par la différence des quartiles extrêmes.Autrement dit, ona

$$\boxed{I_Q = Q_3 - Q_1.}$$

Remarque

Cette mesure est plus fiable que l'étendue puisqu'elle exclut les 50% *des valeurs marginales inférieures et supérieures.*

Definition

L'écart semi-interquartile Q est défini par la moyenne arithmétique des écarts entre les quartiles et la médiane.Autrement dit, on a

$$\boxed{Q = \frac{(Q_3 - M_e) + (M_e - Q_1)}{2} = \frac{Q_3 - Q_1}{2} = \frac{I_Q}{2}.}$$

Definition

L'écart interquartile I_Q est défini par la différence des quartiles extrêmes.Autrement dit, ona

$$\boxed{I_Q = Q_3 - Q_1\,.}$$

Remarque

Cette mesure est plus fiable que l'étendue puisqu'elle exclut les 50% *des valeurs marginales inférieures et supérieures.*

Definition

L'écart semi-interquartile Q est défini par la moyenne arithmétique des écarts entre les quartiles et la médiane.Autrement dit, on a

$$\boxed{Q = \frac{(Q_3 - M_e) + (M_e - Q_1)}{2} = \frac{Q_3 - Q_1}{2} = \frac{I_Q}{2}.}$$

Exemple

Dans notre exemple du nombre de personnes par ménage, on a

Exemple

Dans notre exemple du nombre de personnes par ménage, on a

$$I_Q = 4 - 2 = 2$$

Exemple

Dans notre exemple du nombre de personnes par ménage, on a

$$I_Q = 4 - 2 = 2$$

et

$$Q = \frac{4-2}{2} = 1.$$

Remarque

Il est également possible de définir l'écart interdécile I_D par

$$\boxed{I_D = D_9 - D_1.}$$

Cela définit un intervalle comprenant les 80% de la population.

Exemple

Deux classes de 20 élèves ont effectué un travail écrit de mathématiques, dont les résultats de ces travaux écrits sont présentés dans les tableaux ci-dessous.

Exemple

Deux classes de 20 élèves ont effectué un travail écrit de mathématiques, dont les résultats de ces travaux écrits sont présentés dans les tableaux ci-dessous.

Note x_i	Nombre d'élèves n_i
1	0
1, 5	0
2	0
2, 5	0
3	0
3, 5	3
4	7
4, 5	8
5	1
5, 5	1
6	0

FIGURE – *Notes de la première classe.*

Note y_i	Nombre d'élèves n_i
1	0
1, 5	1
2	0
2, 5	2
3	4
3, 5	0
4	0
4, 5	3
5	6
5, 5	2
6	2

FIGURE – *Notes de la deuxième classe.*

Exemple

Deux classes de 20 élèves ont effectué un travail écrit de mathématiques, dont les résultats de ces travaux écrits sont présentés dans les tableaux ci-dessous.

Note x_i	Nombre d'élèves n_i
1	0
1,5	0
2	0
2,5	0
3	0
3,5	3
4	7
4,5	8
5	1
5,5	1
6	0

FIGURE – *Notes de la première classe.*

Note y_i	Nombre d'élèves n_i
1	0
1,5	1
2	0
2,5	2
3	4
3,5	0
4	0
4,5	3
5	6
5,5	2
6	2

FIGURE – *Notes de la deuxième classe.*

Au vu des résultats ci-dessus, il est naturel de se poser la question suivante.

Exemple

Deux classes de 20 élèves ont effectué un travail écrit de mathématiques, dont les résultats de ces travaux écrits sont présentés dans les tableaux ci-dessous.

Note x_i	Nombre d'élèves n_i
1	0
1,5	0
2	0
2,5	0
3	0
3,5	3
4	7
4,5	8
5	1
5,5	1
6	0

FIGURE – *Notes de la première classe.*

Note y_i	Nombre d'élèves n_i
1	0
1,5	1
2	0
2,5	2
3	4
3,5	0
4	0
4,5	3
5	6
5,5	2
6	2

FIGURE – *Notes de la deuxième classe.*

Au vu des résultats ci-dessus, il est naturel de se poser la question suivante.
Question : *Laquelle de ces deux classes a été la plus performante lors de ce travail écrit ?*

Exemple

Un moyen de répondre à cette question consiste à calculer la moyenne arithmétique de chacune des deux classes :

Exemple

Un moyen de répondre à cette question consiste à calculer la moyenne arithmétique de chacune des deux classes :

$$\overline{x} = \frac{3,5 \cdot 3 + 4 \cdot 7 + 4,5 \cdot 8 + 5 \cdot 1 + 5,5 \cdot 1}{20} = 4,25$$

Exemple

Un moyen de répondre à cette question consiste à calculer la moyenne arithmétique de chacune des deux classes :

$$\bar{x} = \frac{3,5 \cdot 3 + 4 \cdot 7 + 4,5 \cdot 8 + 5 \cdot 1 + 5,5 \cdot 1}{20} = 4,25$$

$$\bar{y} = \frac{1,5 \cdot 1 + 2,5 \cdot 2 + 3 \cdot 4 + 4,5 \cdot 3 + 5 \cdot 6 + 5,5 \cdot 2 + 6 \cdot 2}{20} = 4,25.$$

Exemple

Un moyen de répondre à cette question consiste à calculer la moyenne arithmétique de chacune des deux classes :

$$\begin{aligned}\overline{x} &= \frac{3,5\cdot 3+4\cdot 7+4,5\cdot 8+5\cdot 1+5,5\cdot 1}{20}=4,25\\ \overline{y} &= \frac{1,5\cdot 1+2,5\cdot 2+3\cdot 4+4,5\cdot 3+5\cdot 6+5,5\cdot 2+6\cdot 2}{20}=4,25.\end{aligned}$$

Ces deux moyennes $\overline{x}$ et $\overline{y}$ sont égales alors que les résultats sont très différents !

Exemple

Un moyen de répondre à cette question consiste à calculer la moyenne arithmétique de chacune des deux classes :

$$\overline{x} = \frac{3,5 \cdot 3 + 4 \cdot 7 + 4,5 \cdot 8 + 5 \cdot 1 + 5,5 \cdot 1}{20} = 4,25$$

$$\overline{y} = \frac{1,5 \cdot 1 + 2,5 \cdot 2 + 3 \cdot 4 + 4,5 \cdot 3 + 5 \cdot 6 + 5,5 \cdot 2 + 6 \cdot 2}{20} = 4,25.$$

Ces deux moyennes $\overline{x}$ et $\overline{y}$ sont égales alors que les résultats sont très différents !
La moyenne arithmétique ne donne pas d'informations sur la dispersion des résultats autour de la moyenne. Pour l'estimer, on essaie de quantifier la manière dont les notes sont réparties autour de la moyenne.

Exemple

On obtient :

Exemple

On obtient :

$x_i - \overline{x}$	n_i
$-3,25$	*0*
$-2,75$	*0*
$-2,25$	*0*
$-1,75$	*0*
$-1,25$	*0*
$-0,75$	3
$-0,25$	7
$0,25$	8
$0,75$	1
$1,25$	1
$1,75$	*0*

$y_i - \overline{y}$	n_i
$-3,25$	*0*
$-2,75$	1
$-2,25$	*0*
$-1,75$	2
$-1,25$	4
$-0,75$	*0*
$-0,25$	*0*
$0,25$	3
$0,75$	6
$1,25$	2
$1,75$	2

Exemple

On obtient :

$x_i - \overline{x}$	n_i
$-3,25$	*0*
$-2,75$	*0*
$-2,25$	*0*
$-1,75$	*0*
$-1,25$	*0*
$-0,75$	3
$-0,25$	7
$0,25$	8
$0,75$	1
$1,25$	1
$1,75$	*0*

$y_i - \overline{y}$	n_i
$-3,25$	*0*
$-2,75$	1
$-2,25$	*0*
$-1,75$	2
$-1,25$	4
$-0,75$	*0*
$-0,25$	*0*
$0,25$	3
$0,75$	6
$1,25$	2
$1,75$	2

Le calcul de la moyenne de ces écarts est nul, car les écarts négatifs sont exactement compensés par les écarts positifs, ce qui n'amène aucun renseignement sur la dispersion. On choisit alors de calculer le carré des écarts à la moyenne.

Exemple

On obtient alors les distributions suivantes :

Exemple

On obtient alors les distributions suivantes :

$(x_i - \overline{x})^2$	n_i
10,5625	0
7,5625	0
5,0625	0
3,0625	0
1,5625	0
0,5625	3
0,0625	7
0,0625	8
0,5625	1
1,5625	1
3,0625	0

$(y_i - \overline{y})^2$	n_i
10,5625	0
7,5625	1
5,0625	0
3,0625	2
1,5625	4
0,5625	0
0,0625	0
0,0625	3
0,5625	6
1,5625	2
3,0625	2

Exemple

Calculons alors la moyenne arithmétique de $(\overline{x} - x_i)^2$ *et* $(\overline{y} - y_i)^2$ *:*

Exemple

Calculons alors la moyenne arithmétique de $(\overline{x} - x_i)^2$ *et* $(\overline{y} - y_i)^2$ *:*

$$\begin{aligned}
\overline{(x_i - \overline{x})^2} &= \frac{0,5625 \cdot 3 + 0,0625 \cdot 7 + 0,0625 \cdot 8 + 0,5625 \cdot 1 + 1,5625 \cdot 1}{20} \\
&= 0{,}2375 \\
\overline{(y_i - \overline{y})^2} &= \frac{7,5625 \cdot 1 + 3,0625 \cdot 2 + \ldots + 3,0625 \cdot 2}{20} \\
&= 1{,}6375.
\end{aligned}$$

Exemple

Calculons alors la moyenne arithmétique de $(\overline{x} - x_i)^2$ et $(\overline{y} - y_i)^2$:

$$\begin{aligned}
\overline{(x_i - \overline{x})^2} &= \frac{0,5625 \cdot 3 + 0,0625 \cdot 7 + 0,0625 \cdot 8 + 0,5625 \cdot 1 + 1,5625 \cdot 1}{20} \\
&= 0{,}2375 \\
\overline{(y_i - \overline{y})^2} &= \frac{7,5625 \cdot 1 + 3,0625 \cdot 2 + \ldots + 3,0625 \cdot 2}{20} \\
&= 1{,}6375.
\end{aligned}$$

Ces nombres ainsi trouvés sont une mesure de la dispersion des notes autour de la moyenne arithmétique. On voit ainsi que les notes de la première classe sont plus proches de la moyenne que celles de la deuxième classe.

Definition

On appelle *variance* V d'une série statistique la moyenne des carrés des écarts entre toutes les données et leur moyenne arithmétique. On a ainsi

Definition

On appelle *variance V* d'une série statistique la moyenne des carrés des écarts entre toutes les données et leur moyenne arithmétique. On a ainsi

$$\boxed{V = \overline{(x_i - \overline{x})^2} = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{N}.}$$

Definition

On appelle *variance V* d'une série statistique la moyenne des carrés des écarts entre toutes les données et leur moyenne arithmétique. On a ainsi

$$\boxed{V = \overline{(x_i - \overline{x})^2} = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{N}.}$$

Definition

On appelle *ecart-type* σ, la racine carrée de la variance. Autrement dit, on a

$$\boxed{\sigma = \sqrt{V}.}$$

Definition

On appelle *variance V* d'une série statistique la moyenne des carrés des écarts entre toutes les données et leur moyenne arithmétique. On a ainsi

$$\boxed{V = \overline{(x_i - \overline{x})^2} = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{N}.}$$

Definition

On appelle *ecart-type* σ, la racine carrée de la variance. Autrement dit, on a

$$\boxed{\sigma = \sqrt{V}.}$$

Remarque

L'écart-type est une mesure de la dispersion plus significative que la variance. En effet, si les données x_i représentent une distance exprimée en mètres, V est en m^2 tandis que l'écart-type est exprimé en mètres.

Exemple

Soient les nombres -4, 3, 9, 11 et 17.

Exemple

Soient les nombres -4, 3, 9, 11 et 17.
La moyenne arithmétique de ces nombres vaut

$$\overline{x} = \frac{-4+3+9+11+17}{5} = 7,2.$$

Exemple

Soient les nombres -4*, 3, 9, 11 et 17.*
La moyenne arithmétique de ces nombres vaut

$$\overline{x} = \frac{-4+3+9+11+17}{5} = 7,2.$$

Du tableau suivant

Exemple

Soient les nombres -4, 3, 9, 11 et 17.
La moyenne arithmétique de ces nombres vaut

$$\overline{x} = \frac{-4+3+9+11+17}{5} = 7,2.$$

Du tableau suivant

	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
	-4	$-11,2$	$125,44$
	3	$-4,2$	$17,64$
	9	$1,8$	$3,24$
	11	$3,8$	$14,44$
	17	$9,8$	$96,04$
Total	36	0	$256,8$

Exemple

Soient les nombres -4, 3, 9, 11 et 17.
La moyenne arithmétique de ces nombres vaut

$$\overline{x} = \frac{-4+3+9+11+17}{5} = 7,2.$$

Du tableau suivant

	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
	-4	$-11,2$	$125,44$
	3	$-4,2$	$17,64$
	9	$1,8$	$3,24$
	11	$3,8$	$14,44$
	17	$9,8$	$96,04$
Total	36	0	$256,8$

on en tire la variance

$$V = \frac{256,8}{5} = 51,36$$

Exemple

Soient les nombres -4*, 3, 9, 11 et 17.*
La moyenne arithmétique de ces nombres vaut

$$\overline{x} = \frac{-4+3+9+11+17}{5} = 7,2.$$

Du tableau suivant

	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
	-4	$-11,2$	$125,44$
	3	$-4,2$	$17,64$
	9	$1,8$	$3,24$
	11	$3,8$	$14,44$
	17	$9,8$	$96,04$
Total	36	0	$256,8$

on en tire la variance

$$V = \frac{256,8}{5} = 51,36$$

et l'écart-type

$$\sigma = \sqrt{51,36} \cong 7,167.$$

Exemple

Calculons la variance et l'écart-type de notre exemple du nombre de personnes par ménage. A cet effet, il convient de dresser le tableau ci-dessous.

Exemple

Calculons la variance et l'écart-type de notre exemple du nombre de personnes par ménage. A cet effet, il convient de dresser le tableau ci-dessous.On rappelle que la moyenne arithmétique valait 3,44.

Exemple

Calculons la variance et l'écart-type de notre exemple du nombre de personnes par ménage. A cet effet, il convient de dresser le tableau ci-dessous.On rappelle que la moyenne arithmétique valait 3,44.

Modalités x_i	*Effectifs* n_i	*Ecarts* $x_i - \overline{x}$	*Carrés des écarts* $(x_i - \overline{x})^2$	*Produits* $n_i \cdot (x_i - \overline{x})^2$
1	*5*	$-2,44$	$5,9536$	$29,768$
2	*9*	$-1,44$	$2,0736$	$18,6624$
3	*15*	$-0,44$	$0,1936$	$2,904$
4	*10*	$0,56$	$0,3136$	$3,136$
5	*6*	$1,56$	$2,4336$	$14,6016$
6	*3*	$2,56$	$6,5536$	$19,6608$
8	*2*	$4,56$	$20,7936$	$41,5872$
Total	*50*			*130,32*

Exemple

Calculons la variance et l'écart-type de notre exemple du nombre de personnes par ménage. A cet effet, il convient de dresser le tableau ci-dessous.On rappelle que la moyenne arithmétique valait 3,44.

Modalités x_i	Effectifs n_i	Ecarts $x_i - \overline{x}$	Carrés des écarts $(x_i - \overline{x})^2$	Produits $n_i \cdot (x_i - \overline{x})^2$
1	5	$-2,44$	$5,9536$	$29,768$
2	9	$-1,44$	$2,0736$	$18,6624$
3	15	$-0,44$	$0,1936$	$2,904$
4	10	$0,56$	$0,3136$	$3,136$
5	6	$1,56$	$2,4336$	$14,6016$
6	3	$2,56$	$6,5536$	$19,6608$
8	2	$4,56$	$20,7936$	$41,5872$
Total	50			130,32

On en tire la variance

Exemple

Calculons la variance et l'écart-type de notre exemple du nombre de personnes par ménage. A cet effet, il convient de dresser le tableau ci-dessous.On rappelle que la moyenne arithmétique valait 3,44.

Modalités x_i	*Effectifs* n_i	*Ecarts* $x_i - \overline{x}$	*Carrés des écarts* $(x_i - \overline{x})^2$	*Produits* $n_i \cdot (x_i - \overline{x})^2$
1	*5*	$-2,44$	$5,9536$	$29,768$
2	*9*	$-1,44$	$2,0736$	$18,6624$
3	*15*	$-0,44$	$0,1936$	$2,904$
4	*10*	$0,56$	$0,3136$	$3,136$
5	*6*	$1,56$	$2,4336$	$14,6016$
6	*3*	$2,56$	$6,5536$	$19,6608$
8	*2*	$4,56$	$20,7936$	$41,5872$
Total	*50*			*130,32*

On en tire la variance

$$V = \frac{130,32}{50} = 20,6064$$

Exemple

Calculons la variance et l'écart-type de notre exemple du nombre de personnes par ménage. A cet effet, il convient de dresser le tableau ci-dessous.On rappelle que la moyenne arithmétique valait 3,44.

Modalités x_i	*Effectifs* n_i	*Ecarts* $x_i - \overline{x}$	*Carrés des écarts* $(x_i - \overline{x})^2$	*Produits* $n_i \cdot (x_i - \overline{x})^2$
1	*5*	$-2,44$	$5,9536$	$29,768$
2	*9*	$-1,44$	$2,0736$	$18,6624$
3	*15*	$-0,44$	$0,1936$	$2,904$
4	*10*	$0,56$	$0,3136$	$3,136$
5	*6*	$1,56$	$2,4336$	$14,6016$
6	*3*	$2,56$	$6,5536$	$19,6608$
8	*2*	$4,56$	$20,7936$	$41,5872$
Total	*50*			*130,32*

On en tire la variance

$$V = \frac{130,32}{50} = 20,6064$$

et l'écart-type

Exemple

Calculons la variance et l'écart-type de notre exemple du nombre de personnes par ménage. A cet effet, il convient de dresser le tableau ci-dessous.On rappelle que la moyenne arithmétique valait 3,44.

Modalités x_i	*Effectifs* n_i	*Ecarts* $x_i - \overline{x}$	*Carrés des écarts* $(x_i - \overline{x})^2$	*Produits* $n_i \cdot (x_i - \overline{x})^2$
1	*5*	$-2,44$	$5,9536$	$29,768$
2	*9*	$-1,44$	$2,0736$	$18,6624$
3	*15*	$-0,44$	$0,1936$	$2,904$
4	*10*	$0,56$	$0,3136$	$3,136$
5	*6*	$1,56$	$2,4336$	$14,6016$
6	*3*	$2,56$	$6,5536$	$19,6608$
8	*2*	$4,56$	$20,7936$	$41,5872$
Total	*50*			*130,32*

On en tire la variance

$$V = \frac{130,32}{50} = 20,6064$$

et l'écart-type

$$\sigma = \sqrt{V} \cong 1,614.$$

Comme pour le calcul de la moyenne arithmétique, on affecte à tous les individus d'une classe $]b_{i-1}; b_i]$ la valeur centrale $c = \dfrac{b_{i-1} + b_i}{2}$.

Comme pour le calcul de la moyenne arithmétique, on affecte à tous les individus d'une classe $]b_{i-1}; b_i]$ la valeur centrale $c = \dfrac{b_{i-1} + b_i}{2}$.

Exemple

Dans notre exemple des exploitations agricoles, cet écart se calcule à l'aide du tableau suivant. La moyenne arithmétique valait 21.

Comme pour le calcul de la moyenne arithmétique, on affecte à tous les individus d'une classe $]b_{i-1}; b_i]$ la valeur centrale $c = \dfrac{b_{i-1} + b_i}{2}$.

Exemple

Dans notre exemple des exploitations agricoles, cet écart se calcule à l'aide du tableau suivant. La moyenne arithmétique valait 21.

Classes x_i	*Centres* c_i	*Effectifs* n_i	*Carrés des écarts* $(c_i - \overline{x})^2$	*Produits* $n_i \cdot (c_i - \overline{x})^2$
$]0; 10]$	*5*	*48*	*256*	*12288*
$]10; 15]$	$12,5$	*62*	$72,25$	$4479,5$
$]15; 20]$	$17,5$	*107*	$12,25$	$1310,75$
$]20; 25]$	$22,5$	*133*	$2,25$	$299,25$
$]25; 30]$	$27,5$	*84*	$42,25$	*3549*
$]30; 40]$	*35*	*66*	*196*	*12936*
Total		*500*	*196*	$34862,5$

Comme pour le calcul de la moyenne arithmétique, on affecte à tous les individus d'une classe $]b_{i-1}; b_i]$ la valeur centrale $c = \dfrac{b_{i-1} + b_i}{2}$.

Exemple

Dans notre exemple des exploitations agricoles, cet écart se calcule à l'aide du tableau suivant. La moyenne arithmétique valait 21.

Classes x_i	*Centres* c_i	*Effectifs* n_i	*Carrés des écarts* $(c_i - \overline{x})^2$	*Produits* $n_i \cdot (c_i - \overline{x})^2$
$]0; 10]$	*5*	*48*	*256*	*12288*
$]10; 15]$	$12,5$	*62*	$72,25$	$4479,5$
$]15; 20]$	$17,5$	*107*	$12,25$	$1310,75$
$]20; 25]$	$22,5$	*133*	$2,25$	$299,25$
$]25; 30]$	$27,5$	*84*	$42,25$	*3549*
$]30; 40]$	*35*	*66*	*196*	*12936*
Total		*500*	*196*	$34862,5$

La variance est donc égale à $V = \dfrac{34862,5}{500} = 69,725\ ha^2$ *et l'écart-type est donné par*

Comme pour le calcul de la moyenne arithmétique, on affecte à tous les individus d'une classe $]b_{i-1}; b_i]$ la valeur centrale $c = \dfrac{b_{i-1} + b_i}{2}$.

Exemple

Dans notre exemple des exploitations agricoles, cet écart se calcule à l'aide du tableau suivant. La moyenne arithmétique valait 21.

Classes x_i	Centres c_i	Effectifs n_i	Carrés des écarts $(c_i - \overline{x})^2$	Produits $n_i \cdot (c_i - \overline{x})^2$
$]0; 10]$	5	48	256	12288
$]10; 15]$	$12,5$	62	$72,25$	$4479,5$
$]15; 20]$	$17,5$	107	$12,25$	$1310,75$
$]20; 25]$	$22,5$	133	$2,25$	$299,25$
$]25; 30]$	$27,5$	84	$42,25$	3549
$]30; 40]$	35	66	196	12936
Total		500	196	$34862,5$

La variance est donc égale à $V = \dfrac{34862,5}{500} = 69,725\ ha^2$ *et l'écart-type est donné par* $\sigma = \sqrt{69,725\ ha^2} \cong 8,35\ ha.$

Le calcul de la variance (et donc de l'écart-type) n'est pas toujours commode. En particulier lorsque la moyenne est un nombre dont on ne donne qu'une approximation avec un développement décimal limité.

Le calcul de la variance (et donc de l'écart-type) n'est pas toujours commode. En particulier lorsque la moyenne est un nombre dont on ne donne qu'une approximation avec un développement décimal limité.Les calculs peuvent toutefois être simplifiés de la manière suivante.

Théorème

La variance V peut être obtenue en calculant la différence entre la moyenne $\overline{x^2}$ des carrés des données x_i et le carré de leur moyenne $\overline{x}^2$.

Le calcul de la variance (et donc de l'écart-type) n'est pas toujours commode. En particulier lorsque la moyenne est un nombre dont on ne donne qu'une approximation avec un développement décimal limité.Les calculs peuvent toutefois être simplifiés de la manière suivante.

Théorème

La variance V peut être obtenue en calculant la différence entre la moyenne $\overline{x^2}$ des carrés des données x_i et le carré de leur moyenne $\overline{x}^2$. Ainsi, on a

$$\boxed{V = \overline{x^2} - \overline{x}^2.}$$

Preuve.

On a

$$V$$

Preuve.

On a

$$V = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_N - \overline{x})^2}{N}$$

Preuve.

On a

$$
\begin{aligned}
V &= \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\ldots+(x_N-\overline{x})^2}{N}\\
&= \frac{(x_1^2-2x_1\overline{x}+\overline{x}^2)+(x_2^2-2x_2\overline{x}+\overline{x}^2)+\ldots+(x_N^2-2x_N\overline{x}+\overline{x}^2)}{N}
\end{aligned}
$$

Preuve.

On a

$$
\begin{aligned}
V &= \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\ldots+(x_N-\overline{x})^2}{N} \\
&= \frac{(x_1^2-2x_1\overline{x}+\overline{x}^2)+(x_2^2-2x_2\overline{x}+\overline{x}^2)+\ldots+(x_N^2-2x_N\overline{x}+\overline{x}^2)}{N} \\
&= \frac{x_1^2+x_2^2+\ldots+x_N^2}{N}-\frac{2\overline{x}(x_1+x_2+\ldots+x_N)}{N}+\frac{\overline{x}^2+\overline{x}^2+\ldots+\overline{x}^2}{N}
\end{aligned}
$$

Preuve.

On a

$$\begin{aligned}
V &= \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+...+(x_N-\overline{x})^2}{N}\\
&= \frac{(x_1^2-2x_1\overline{x}+\overline{x}^2)+(x_2^2-2x_2\overline{x}+\overline{x}^2)+...+(x_N^2-2x_N\overline{x}+\overline{x}^2)}{N}\\
&= \frac{x_1^2+x_2^2+...+x_N^2}{N}-\frac{2\overline{x}(x_1+x_2+...+x_N)}{N}+\frac{\overline{x}^2+\overline{x}^2+...+\overline{x}^2}{N}\\
&= \overline{x^2}-2\overline{x}\cdot\overline{x}+\overline{x}^2
\end{aligned}$$

Preuve.

On a

$$\begin{aligned}
V &= \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+...+(x_N-\overline{x})^2}{N} \\
&= \frac{(x_1^2-2x_1\overline{x}+\overline{x}^2)+(x_2^2-2x_2\overline{x}+\overline{x}^2)+...+(x_N^2-2x_N\overline{x}+\overline{x}^2)}{N} \\
&= \frac{x_1^2+x_2^2+...+x_N^2}{N}-\frac{2\overline{x}(x_1+x_2+...+x_N)}{N}+\frac{\overline{x}^2+\overline{x}^2+...+\overline{x}^2}{N} \\
&= \overline{x^2}-2\overline{x}\cdot\overline{x}+\overline{x}^2 \\
&= \overline{x^2}-2\overline{x}^2+\overline{x}^2
\end{aligned}$$

Preuve.

On a

$$
\begin{aligned}
V &= \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\ldots+(x_N-\overline{x})^2}{N}\\
&= \frac{(x_1^2-2x_1\overline{x}+\overline{x}^2)+(x_2^2-2x_2\overline{x}+\overline{x}^2)+\ldots+(x_N^2-2x_N\overline{x}+\overline{x}^2)}{N}\\
&= \frac{x_1^2+x_2^2+\ldots+x_N^2}{N}-\frac{2\overline{x}(x_1+x_2+\ldots+x_N)}{N}+\frac{\overline{x}^2+\overline{x}^2+\ldots+\overline{x}^2}{N}\\
&= \overline{x^2}-2\overline{x}\cdot\overline{x}+\overline{x}^2\\
&= \overline{x^2}-2\overline{x}^2+\overline{x}^2\\
&= \overline{x^2}-\overline{x}^2.
\end{aligned}
$$

Preuve.

On a

$$
\begin{aligned}
V &= \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\dots+(x_N-\overline{x})^2}{N}\\
&= \frac{(x_1^2-2x_1\overline{x}+\overline{x}^2)+(x_2^2-2x_2\overline{x}+\overline{x}^2)+\dots+(x_N^2-2x_N\overline{x}+\overline{x}^2)}{N}\\
&= \frac{x_1^2+x_2^2+\dots+x_N^2}{N}-\frac{2\overline{x}(x_1+x_2+\dots+x_N)}{N}+\frac{\overline{x}^2+\overline{x}^2+\dots+\overline{x}^2}{N}\\
&= \overline{x^2}-2\overline{x}\cdot\overline{x}+\overline{x}^2\\
&= \overline{x^2}-2\overline{x}^2+\overline{x}^2\\
&= \overline{x^2}-\overline{x}^2.
\end{aligned}
$$

□

Remarque

Cette seconde formulation sera préférée à la première chaque fois que les termes $x_i - \overline{x}$ sont plus compliqués que les termes x_i. Ce cas se présente fréquemment lorsque la moyenne n'est pas un nombre entier.

Exemple

Reprenons l'exemple des nombres -4*, 3, 9, 11 et 17 de moyenne arithmétique* $7,2$.

Exemple

Reprenons l'exemple des nombres -4*, 3, 9, 11 et 17 de moyenne arithmétique* $7,2$*.*
Du tableau suivant

Exemple

Reprenons l'exemple des nombres -4*, 3, 9, 11 et 17 de moyenne arithmétique* $7,2$.
Du tableau suivant

	x_i	x_i^2
	-4	16
	3	9
	9	81
	11	121
	17	289
Total	*36*	516

Exemple

Reprenons l'exemple des nombres -4*, 3, 9, 11 et 17 de moyenne arithmétique* $7,2$*. Du tableau suivant*

	x_i	x_i^2
	-4	16
	3	9
	9	81
	11	121
	17	289
Total	*36*	516

on en tire la variance $V = \dfrac{516}{5} - 7,2^2 = 51,36$ *et l'écart-type* $\sigma = \sqrt{51,36} \cong 7,167$*.*
On retrouve bien les résultats obtenus plus haut.

Exemple

Reprenons l'exemple des exploitations agricoles.

Exemple

Reprenons l'exemple des exploitations agricoles.

Classes x_i	Centres c_i	Effectifs n_i	Carrés des centres c_i^2	Produits $n_i \cdot c_i^2$
$]0;10]$	5	48	25	1200
$]10;15]$	$12,5$	62	$156,25$	$9687,5$
$]15;20]$	$17,5$	107	$306,25$	$32768,75$
$]20;25]$	$22,5$	133	$506,25$	$67331,25$
$]25;30]$	$27,5$	84	$756,25$	63525
$]30;40]$	35	66	1225	80850
Total		500		$255362,5$

Exemple

Reprenons l'exemple des exploitations agricoles.

Classes x_i	Centres c_i	Effectifs n_i	Carrés des centres c_i^2	Produits $n_i \cdot c_i^2$
$]0;10]$	5	48	25	1200
$]10;15]$	12,5	62	156,25	9687,5
$]15;20]$	17,5	107	306,25	32768,75
$]20;25]$	22,5	133	506,25	67331,25
$]25;30]$	27,5	84	756,25	63525
$]30;40]$	35	66	1225	80850
Total		500		255362,5

On en déduit que $\overline{x^2} = \dfrac{255362,5}{500} = 510,725$. *Comme* $\overline{x} = 21$, $\overline{x}^2 = 441$, *il suit que*

Exemple

Reprenons l'exemple des exploitations agricoles.

Classes x_i	Centres c_i	Effectifs n_i	Carrés des centres c_i^2	Produits $n_i \cdot c_i^2$
$]0;10]$	5	48	25	1200
$]10;15]$	$12,5$	62	$156,25$	$9687,5$
$]15;20]$	$17,5$	107	$306,25$	$32768,75$
$]20;25]$	$22,5$	133	$506,25$	$67331,25$
$]25;30]$	$27,5$	84	$756,25$	63525
$]30;40]$	35	66	1225	80850
Total		500		$255362,5$

On en déduit que $\overline{x^2} = \dfrac{255362,5}{500} = 510,725$. *Comme* $\overline{x} = 21$, $\overline{x}^2 = 441$, *il suit que*

$$V =$$

Exemple

Reprenons l'exemple des exploitations agricoles.

Classes x_i	Centres c_i	Effectifs n_i	Carrés des centres c_i^2	Produits $n_i \cdot c_i^2$
$]0;10]$	5	48	25	1200
$]10;15]$	$12,5$	62	$156,25$	$9687,5$
$]15;20]$	$17,5$	107	$306,25$	$32768,75$
$]20;25]$	$22,5$	133	$506,25$	$67331,25$
$]25;30]$	$27,5$	84	$756,25$	63525
$]30;40]$	35	66	1225	80850
Total		500		$255362,5$

On en déduit que $\overline{x^2} = \dfrac{255362,5}{500} = 510,725$. *Comme* $\overline{x} = 21$, $\overline{x}^2 = 441$, *il suit que*

$$V = \overline{x^2} - \overline{x}^2 = 510,725 - 441$$

Exemple

Reprenons l'exemple des exploitations agricoles.

Classes x_i	*Centres* c_i	*Effectifs* n_i	*Carrés des centres* c_i^2	*Produits* $n_i \cdot c_i^2$
$]0;10]$	*5*	*48*	*25*	*1200*
$]10;15]$	$12,5$	*62*	$156,25$	$9687,5$
$]15;20]$	$17,5$	*107*	$306,25$	$32768,75$
$]20;25]$	$22,5$	*133*	$506,25$	$67331,25$
$]25;30]$	$27,5$	*84*	$756,25$	*63525*
$]30;40]$	*35*	*66*	*1225*	*80850*
Total		*500*		$255362,5$

On en déduit que $\overline{x^2} = \dfrac{255362,5}{500} = 510,725$. *Comme* $\overline{x} = 21$, $\overline{x}^2 = 441$, *il suit que*

$$V = \overline{x^2} - \overline{x}^2 = 510,725 - 441 = 69,725\ ha^2$$

et l'écart-type est donc donné par

Exemple

Reprenons l'exemple des exploitations agricoles.

Classes x_i	Centres c_i	Effectifs n_i	Carrés des centres c_i^2	Produits $n_i \cdot c_i^2$
$]0;10]$	5	48	25	1200
$]10;15]$	$12,5$	62	$156,25$	$9687,5$
$]15;20]$	$17,5$	107	$306,25$	$32768,75$
$]20;25]$	$22,5$	133	$506,25$	$67331,25$
$]25;30]$	$27,5$	84	$756,25$	63525
$]30;40]$	35	66	1225	80850
Total		500		$255362,5$

On en déduit que $\overline{x^2} = \dfrac{255362,5}{500} = 510,725$. *Comme* $\overline{x} = 21$, $\overline{x}^2 = 441$, *il suit que*

$$V = \overline{x^2} - \overline{x}^2 = 510,725 - 441 = 69,725 \ ha^2$$

et l'écart-type est donc donné par $\sigma =$

Exemple

Reprenons l'exemple des exploitations agricoles.

Classes x_i	Centres c_i	Effectifs n_i	Carrés des centres c_i^2	Produits $n_i \cdot c_i^2$
$]0;10]$	5	48	25	1200
$]10;15]$	$12,5$	62	$156,25$	$9687,5$
$]15;20]$	$17,5$	107	$306,25$	$32768,75$
$]20;25]$	$22,5$	133	$506,25$	$67331,25$
$]25;30]$	$27,5$	84	$756,25$	63525
$]30;40]$	35	66	1225	80850
Total		500		$255362,5$

On en déduit que $\overline{x^2} = \dfrac{255362,5}{500} = 510,725$. *Comme* $\overline{x} = 21$, $\overline{x}^2 = 441$, *il suit que*

$$V = \overline{x^2} - \overline{x}^2 = 510,725 - 441 = 69,725\ ha^2$$

et l'écart-type est donc donné par $\sigma = \sqrt{69,725\ ha^2} \cong 8,35\ ha$.

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.

Coefficient de variation

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.Cependant, dans la grande majorité des cas, on décrit une distribution par sa moyenne et son écart-type.

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.Cependant, dans la grande majorité des cas, on décrit une distribution par sa moyenne et son écart-type.La moyenne indique autour de quelle valeur sont situées les données, alors que l'écart-type donne une idée de la dispersion.

Coefficient de variation

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.Cependant, dans la grande majorité des cas, on décrit une distribution par sa moyenne et son écart-type.La moyenne indique autour de quelle valeur sont situées les données, alors que l'écart-type donne une idée de la dispersion.Cette idée de dispersion doit cependant être située dans son contexte.

Coefficient de variation

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.Cependant, dans la grande majorité des cas, on décrit une distribution par sa moyenne et son écart-type.La moyenne indique autour de quelle valeur sont situées les données, alors que l'écart-type donne une idée de la dispersion.Cette idée de dispersion doit cependant être située dans son contexte.Si l'écart-type d'une distribution est égal à 10, peut-on dire que cette distribution est très dispersée ?

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.Cependant, dans la grande majorité des cas, on décrit une distribution par sa moyenne et son écart-type.La moyenne indique autour de quelle valeur sont situées les données, alors que l'écart-type donne une idée de la dispersion.Cette idée de dispersion doit cependant être située dans son contexte.Si l'écart-type d'une distribution est égal à 10, peut-on dire que cette distribution est très dispersée ?Bien sûr, cela dépend de l'ordre de grandeur des données.

Coefficient de variation

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.Cependant, dans la grande majorité des cas, on décrit une distribution par sa moyenne et son écart-type.La moyenne indique autour de quelle valeur sont situées les données, alors que l'écart-type donne une idée de la dispersion.Cette idée de dispersion doit cependant être située dans son contexte.Si l'écart-type d'une distribution est égal à 10, peut-on dire que cette distribution est très dispersée ?Bien sûr, cela dépend de l'ordre de grandeur des données.En effet, si les données traitées sont de l'ordre de 2000 par exemple, cet écart-type est vraiment petit et les données sont sûrement très concentrées.

Coefficient de variation

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.Cependant, dans la grande majorité des cas, on décrit une distribution par sa moyenne et son écart-type.La moyenne indique autour de quelle valeur sont situées les données, alors que l'écart-type donne une idée de la dispersion.Cette idée de dispersion doit cependant être située dans son contexte.Si l'écart-type d'une distribution est égal à 10, peut-on dire que cette distribution est très dispersée ?Bien sûr, cela dépend de l'ordre de grandeur des données.En effet, si les données traitées sont de l'ordre de 2000 par exemple, cet écart-type est vraiment petit et les données sont sûrement très concentrées.Par contre, si les données sont de l'ordre de 12, par exemple, l'écart-type est grand et les données sont relativement dispersées.

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.Cependant, dans la grande majorité des cas, on décrit une distribution par sa moyenne et son écart-type.La moyenne indique autour de quelle valeur sont situées les données, alors que l'écart-type donne une idée de la dispersion.Cette idée de dispersion doit cependant être située dans son contexte.Si l'écart-type d'une distribution est égal à 10, peut-on dire que cette distribution est très dispersée ?Bien sûr, cela dépend de l'ordre de grandeur des données.En effet, si les données traitées sont de l'ordre de 2000 par exemple, cet écart-type est vraiment petit et les données sont sûrement très concentrées.Par contre, si les données sont de l'ordre de 12, par exemple, l'écart-type est grand et les données sont relativement dispersées.Il est donc utile de mesurer la *dispersion relative*.

Coefficient de variation

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.Cependant, dans la grande majorité des cas, on décrit une distribution par sa moyenne et son écart-type.La moyenne indique autour de quelle valeur sont situées les données, alors que l'écart-type donne une idée de la dispersion.Cette idée de dispersion doit cependant être située dans son contexte.Si l'écart-type d'une distribution est égal à 10, peut-on dire que cette distribution est très dispersée ?Bien sûr, cela dépend de l'ordre de grandeur des données.En effet, si les données traitées sont de l'ordre de 2000 par exemple, cet écart-type est vraiment petit et les données sont sûrement très concentrées.Par contre, si les données sont de l'ordre de 12, par exemple, l'écart-type est grand et les données sont relativement dispersées.Il est donc utile de mesurer la *dispersion relative*.

Definition

Le *coefficient de variation C* d'une variable statistique est le rapport entre l'écart-type et la moyenne exprimé sous la forme d'un pourcentage :

Coefficient de variation

Pour caractériser une distribution, on utilise généralement une mesure de tendance centrale et une mesure de dispersion.Par exemple, on peut donner sa médiane et son intervalle semi-interquartille.Cependant, dans la grande majorité des cas, on décrit une distribution par sa moyenne et son écart-type.La moyenne indique autour de quelle valeur sont situées les données, alors que l'écart-type donne une idée de la dispersion.Cette idée de dispersion doit cependant être située dans son contexte.Si l'écart-type d'une distribution est égal à 10, peut-on dire que cette distribution est très dispersée ?Bien sûr, cela dépend de l'ordre de grandeur des données.En effet, si les données traitées sont de l'ordre de 2000 par exemple, cet écart-type est vraiment petit et les données sont sûrement très concentrées.Par contre, si les données sont de l'ordre de 12, par exemple, l'écart-type est grand et les données sont relativement dispersées.Il est donc utile de mesurer la *dispersion relative*.

Definition

Le *coefficient de variation C* d'une variable statistique est le rapport entre l'écart-type et la moyenne exprimé sous la forme d'un pourcentage :

$$\boxed{C = \frac{\sigma}{\overline{x}}}$$

Remarque

Si l'on souhaite porter un jugement sur la dispersion d'une série, la qualification suivante est généralement admise :

Coefficient de variation

Remarque

Si l'on souhaite porter un jugement sur la dispersion d'une série, la qualification suivante est généralement admise :

Coefficient de variation	*Dispersion*
0 à 10%	*Faible*
10 à 20%	*Moyenne*
Plus de 20%	*Elevée*

Exemple

Dans notre exemple des exploitations agricoles, ce coefficient vaut

Remarque

Si l'on souhaite porter un jugement sur la dispersion d'une série, la qualification suivante est généralement admise :

Coefficient de variation	*Dispersion*
0 à 10%	*Faible*
10 à 20%	*Moyenne*
Plus de 20%	*Elevée*

Exemple

Dans notre exemple des exploitations agricoles, ce coefficient vaut

$$C = \frac{\sigma}{\overline{x}} \cong \frac{8,35}{21} \cong 0,398 = 39,8\%.$$

Remarque

Si l'on souhaite porter un jugement sur la dispersion d'une série, la qualification suivante est généralement admise :

Coefficient de variation	*Dispersion*
0 à 10%	*Faible*
10 à 20%	*Moyenne*
Plus de 20%	*Elevée*

Exemple

Dans notre exemple des exploitations agricoles, ce coefficient vaut

$$C = \frac{\sigma}{\overline{x}} \cong \frac{8,35}{21} \cong 0,398 = 39,8\%.$$

Ainsi, la dispersion des données est élevée.

1. Elle est très simple à calculer et à interpréter.

1. Elle est très simple à calculer et à interpréter.
2. Elle ne tient pas compte de toutes les données et n'implique que les valeurs extrêmes.

1. Elle est très simple à calculer et à interpréter.
2. Elle ne tient pas compte de toutes les données et n'implique que les valeurs extrêmes.
3. Elle est utilisée pour donner une idée sommaire et rapide de la dispersion et pour déterminer les largeurs de classes lorsqu'on fait un regroupement en classes.

1. Elle est très simple à calculer et à interpréter.
2. Elle ne tient pas compte de toutes les données et n'implique que les valeurs extrêmes.
3. Elle est utilisée pour donner une idée sommaire et rapide de la dispersion et pour déterminer les largeurs de classes lorsqu'on fait un regroupement en classes.
4. Sa valeur n'est pas stable, c'est-à-dire qu'elle varie beaucoup d'un échantillon à l'autre choisi dans une même population.

1. Elle est très simple à calculer et à interpréter.
2. Elle ne tient pas compte de toutes les données et n'implique que les valeurs extrêmes.
3. Elle est utilisée pour donner une idée sommaire et rapide de la dispersion et pour déterminer les largeurs de classes lorsqu'on fait un regroupement en classes.
4. Sa valeur n'est pas stable, c'est-à-dire qu'elle varie beaucoup d'un échantillon à l'autre choisi dans une même population.
5. Elle est très peu utilisée.

1. Il est simple à calculer et à interpréter.

1. Il est simple à calculer et à interpréter.
2. Il ne tient pas compte de toutes les données et n'est donc pas influencé par les données extrêmes.

1. Il est simple à calculer et à interpréter.
2. Il ne tient pas compte de toutes les données et n'est donc pas influencé par les données extrêmes.
3. Il est utilisé lorsque la distribution des effectifs est fortement dissymétrique. Dans ce cas, on utilise la médiane comme mesure de tendance centrale.

1. Il est simple à calculer et à interpréter.
2. Il ne tient pas compte de toutes les données et n'est donc pas influencé par les données extrêmes.
3. Il est utilisé lorsque la distribution des effectifs est fortement dissymétrique. Dans ce cas, on utilise la médiane comme mesure de tendance centrale.
4. Sa valeur est moins stable que celle de la variance ou de l'écart-type.

1. Il est simple à calculer et à interpréter.
2. Il ne tient pas compte de toutes les données et n'est donc pas influencé par les données extrêmes.
3. Il est utilisé lorsque la distribution des effectifs est fortement dissymétrique. Dans ce cas, on utilise la médiane comme mesure de tendance centrale.
4. Sa valeur est moins stable que celle de la variance ou de l'écart-type.
5. Il est peu utilisé en général.

1. Il présente les mêmes caractéristiques que l'écart semi-interquartile.

1. Son calcul est plus long et son interprétation est moins immédiate.

1. Son calcul est plus long et son interprétation est moins immédiate.
2. Il tient compte de toutes les données.

1. Son calcul est plus long et son interprétation est moins immédiate.
2. Il tient compte de toutes les données.
3. Il se prête assez bien aux manipulations algébriques. On le retrouve ainsi dans plusieurs calculs en statistiques inférentielles.

1. Son calcul est plus long et son interprétation est moins immédiate.
2. Il tient compte de toutes les données.
3. Il se prête assez bien aux manipulations algébriques. On le retrouve ainsi dans plusieurs calculs en statistiques inférentielles.
4. Sa valeur est stable d'un échantillon à l'autre.

1. Son calcul est plus long et son interprétation est moins immédiate.
2. Il tient compte de toutes les données.
3. Il se prête assez bien aux manipulations algébriques. On le retrouve ainsi dans plusieurs calculs en statistiques inférentielles.
4. Sa valeur est stable d'un échantillon à l'autre.
5. Il est, avec la variance, la mesure de dispersion la plus utilisée.

1. La variance a les mêmes caractéristiques que l'écart-type.

1. La variance a les mêmes caractéristiques que l'écart-type.
2. La présence de carrés accorde plus de poids aux grands écarts. Elle est ainsi fortement influencée par les données extrêmes.

Remarque

Le choix de la mesure de tendance centrale implique le choix de la mesure de dispersion :

Remarque

Le choix de la mesure de tendance centrale implique le choix de la mesure de dispersion :

$$\textit{mode} \qquad \leftrightarrow \qquad \textit{étendue}$$

Remarque

Le choix de la mesure de tendance centrale implique le choix de la mesure de dispersion :

mode	$\leftrightarrow$	*étendue*
médiane	$\leftrightarrow$	*écart semi-interquartile*

Remarque

Le choix de la mesure de tendance centrale implique le choix de la mesure de dispersion :

mode	$\leftrightarrow$	*étendue*
médiane	$\leftrightarrow$	*écart semi-interquartile*
moyenne	$\leftrightarrow$	*écart-type*