

Statistiques descriptives

Karim Saïd

Ecole Technique, Année scolaire 2020-2021

1 Introduction

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^{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.

2 Définitions

Définition.

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, \dots et ses modalités par la même lettre minuscule affectée d'indices : x_1, x_2, \dots pour la variable X ou y_1, y_2, \dots pour la variable Y .

Exemple. On fait une étude auprès des étudiants de l'ET. 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.

La population considérée est "les étudiants de l'ET". 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	$x_1 = \text{homme}, x_2 = \text{femme}$
Y : âge	$y_1 = 18, y_2 = 19, \dots$
Z : taille	$z_i \in [150; 200]$
U : taux de satisfaction	$u_1 = S, u_2 = I, u_3 = SR$

Définition.

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 *ordinaire*, si ses valeurs ne possèdent pas d'ordre, respectivement si ses valeurs possèdent un certain ordre.
3. On dit d'une variable statistique 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.

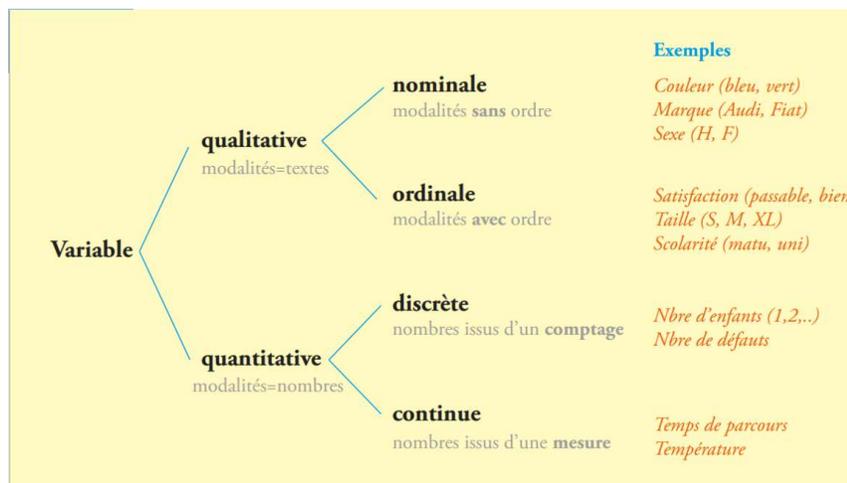


FIGURE 1 – Caractérisation des différentes variables statistiques.

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

Exercice 1. Les voitures particulières neuves immatriculées en 2015 sont réparties dans le tableau ci-dessous selon le type d'énergie consommée.

Type d'énergie	Milliers de voitures
Essence	1'450
Diesel	630
Hybrides	32
Autres	9

- a) Quelle est la population étudiée ?
- b) Quel est le caractère étudié ?

Exercice 2. Dans chacune des situations suivantes, déterminer si il s'agit d'une variable aléatoire qualitative (nominale ou ordinale) ou quantitative (discrète ou continue).

- a) La taille des étudiants de l'Université de Neuchâtel.
- b) Le nombre de pages d'un support de cours.
- c) La nationalité des élèves d'une classe.
- d) Le poids d'un nouveau né.
- e) Le degré de qualification du personnel d'une entreprise.
- f) Le nombre de jours de pluie pendant le mois d'août.
- g) Le degré de satisfaction des étudiants du CPLN au cours de mathématiques.
- h) Le lieu de résidence des enfants d'une classe primaire.
- i) La vitesse du vent.
- j) Le nombre d'erreurs typographiques sur chaque page d'un journal.
- k) Le nombre de personnes vivant dans les ménages du canton de Neuchâtel.
- l) La gravité des blessures des personnes admises aux urgences d'un hôpital.
- m) L'état civil des habitants de la Suisse.
- n) Le nombre de télévisions par famille.
- o) Le temps passé par chaque patient dans un cabinet médical.
- p) Le degré de difficultés des pistes de ski d'une station.

Exercice 3. On a demandé aux employés d'une entreprise pour quel parti politique ils avaient voté lors des dernières élections. Voici les données brutes obtenues :

PS	PLR	PS	PDC	PS	UDC
PS	UDC	PLR	PS	verts	PDC
UDC	PLR	verts	UDC	UDC	UDC
PLR	PS	PLR	PDC	PLR	PDC
UDC	PDC	PS	UDC	UDC	UDC

- a) Identifier la population ainsi que la variable statistique.
- b) Donner l'ensemble des modalités.
- c) De quel type est cette variable statistique ?

Exercice 4. Un professeur de l'Uni a noté le nombre de points obtenus par 80 étudiants lors d'un test de statistiques.

2	3	5	5	4	6	6	5	4	3
7	7	7	6	2	7	7	9	8	10
5	6	6	8	6	6	3	7	3	5
9	7	6	4	7	5	9	9	6	9
6	3	9	8	8	7	5	6	10	6
9	7	7	7	4	7	10	8	7	10
3	5	8	5	8	7	4	8	10	7
4	6	6	8	7	7	7	8	8	9

- a) Identifier la population ainsi que la variable statistique.
- b) Donner l'ensemble des modalités.
- c) De quel type est cette variable statistique?

3 Traitement des données

3.1 Regroupement des données par modalités

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

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.

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

```

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.

3.2 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 ;
 - $c_i = \frac{b_{i-1} + b_i}{2}$ 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.

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.

Superficie 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

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^{ème}) est affecté de la valeur 25, borne supérieure de l'intervalle.

Exercice 5. Quelle est la superficie de la 52^{ème} exploitation de la classe $]20; 25]$ qui en compte 133 ? Même question pour la 78^{ème} exploitation de la classe $]15; 20]$ qui en compte 107.

Exercice 6. Des chimistes viennent de composer une nouvelle fibre synthétique qui devrait se caractériser par sa résistance. Afin de vérifier sa capacité de tension, on prélève de la production, au hasard, un échantillon de 60 fibres qu'on soumet à des essais de résistance. Les résultats (en kg) sont les suivants :

35	65	71	75	77	80	81	82	84	86	87	89	91	97	100
48	69	72	75	78	80	81	83	85	86	88	89	94	97	103
53	69	73	76	79	80	81	83	85	87	88	89	95	99	104
63	71	74	77	79	81	82	84	86	87	89	91	97	99	114

Regrouper les données en 6 classes d'amplitude 15 avec 30 comme valeur minimale. Pour chaque classe, donner le centre, l'effectif correspondant et les fréquences.

4 Représentations graphiques

4.1 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.

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.

Superficie 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 2 – Données avec angles.

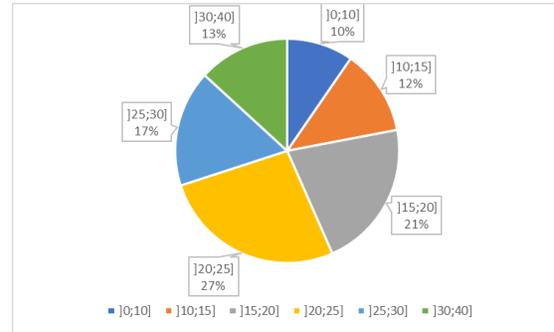


FIGURE 3 – Diagramme en secteurs.

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.

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 4 – Données avec angles.

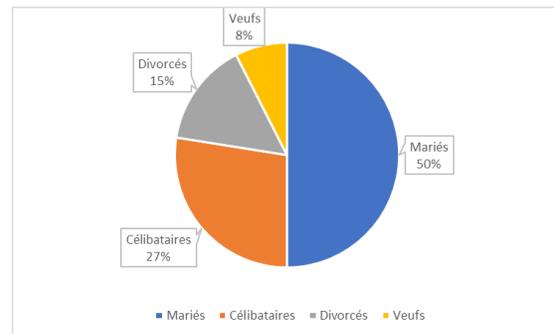
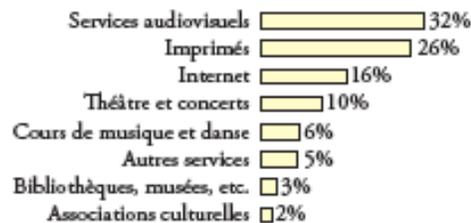


FIGURE 5 – Diagramme en secteurs.

Exercice 7. De 2009 à 2011, les dépenses culturelles ont représenté 5% de la dépense totale de consommation des ménages suisses. Leur répartition est donnée par le graphique ci-dessous.



Déterminer si les affirmations sont vraies ou fausses et justifier.

- La population étudiée est "les ménages suisses".
- Le caractère étudié est "les catégories de dépenses culturelles".
- La nature du caractère est qualitatif discret.
- Ce diagramme est faux, car la somme des pourcentages devrait être de 5%.

Exercice 8. Un sondage effectué sur 1400 adolescentes ayant assisté à un concert du groupe *One Direction* a été réalisé. La question posée était "Quel est votre chanteur favori du groupe ?" Les réponses ont été rassemblées au sein du tableau ci-dessous.

Chanteur	Effectifs	Fréquences (en %)	Angle (en degrés)
Liam Payne	350		
Niall Horan	210		
Zayn Malik	140		
Harry Styles	560		
Louis Tomlinson	140		

- Compléter le tableau ci-dessus.
- Construire le diagramme en secteurs à la main.

Exercice 9. Lors d'un sondage, avant une votation populaire, on a obtenu les indications suivantes :

Avis	Nombre
Oui	1'240
Plutôt oui	350
Indécis	780
Plutôt non	410
Non	1'120

- Déterminer en % les fréquences.
- Dessiner un diagramme circulaire.

4.2 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.

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

FIGURE 6 – Données.

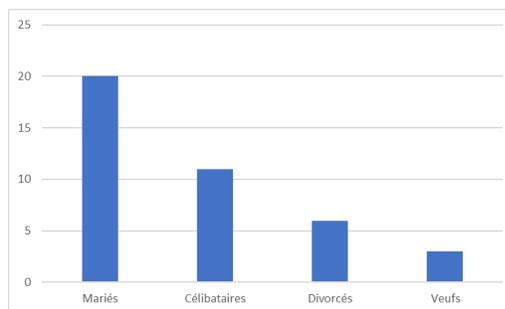


FIGURE 7 – Diagramme en bâtons.

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 8 – Données.

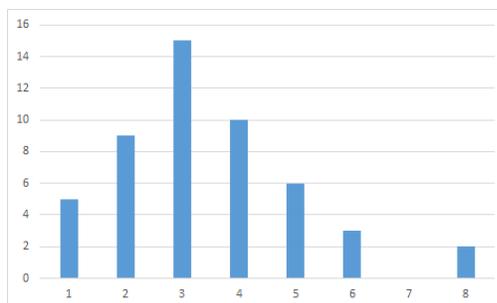


FIGURE 9 – Diagramme en bâtons.

Exercice 10. Un magasin de vêtements s'intéresse à la couleur des robes de sa collection. En notant la couleur de toutes ses robes, on a recueilli les données brutes ci-dessous.

jaune	verte	jaune	noire	rose	noire
bleue	bleue	blanche	orange	verte	beige
rouge	rouge	blanche	rouge	orange	bleue
bleue	rose	beige	verte	blanche	jaune
verte	verte	beige	blanche	rouge	rose
noire	jaune	rose	rose	bleue	beige
blanche	orange	verte	grise	beige	verte
noire	beige	grise	grise	noire	bleue
bleue	noire	rouge	noire	rouge	rouge
verte	grise	noire	orange	verte	noire

- Identifier la population et la variable statistique.
- Donner l'ensemble des modalités.
- Donner le tableau de distribution des effectifs et des fréquences.
- Représenter les données par un diagramme en bâtons.

Exercice 11. On étudie l'état civil des 30 employés (numérotés de 1 à 30) d'une petite entreprise.

1	Marié	11	Marié	21	Célibataire
2	Mariée	12	Célibataire	22	Marié
3	Célibataire	13	Marié	23	Veuf
4	Divorcé	14	Veuve	24	Célibataire
5	Marié	15	Marié	25	Divorcée
6	Célibataire	16	Divorcé	26	Divorcé
7	Célibataire	17	Célibataire	27	Marié
8	Mariée	18	Mariée	28	Marié
9	Mariée	19	Marié	29	Marié
10	Divorcée	20	Marié	30	Marié

- a) Identifier la population.
- b) Caractériser la variable statistique.
- c) Donner l'ensemble des modalités.
- d) Construire le tableau des distributions des effectifs et des fréquences.
- e) Proposer le diagramme en bâtons des effectifs de cette variable statistique.
- f) Proposer le diagramme en bâtons des fréquences de cette variable statistique.
- g) Comparer ces deux représentations graphiques.

4.3 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. 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.

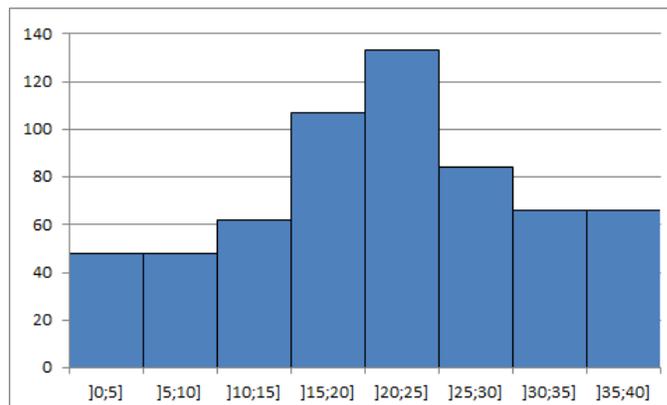


FIGURE 10 – Histogramme trompeur.

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.

Superficie en ha	Nombre d'exploitations
]0; 5]	24
]5; 10]	24
]10; 15]	62
]15; 20]	107
]20; 25]	133
]25; 30]	84
]30; 35]	33
]35; 40]	33

FIGURE 11 – Effectifs corrigés.

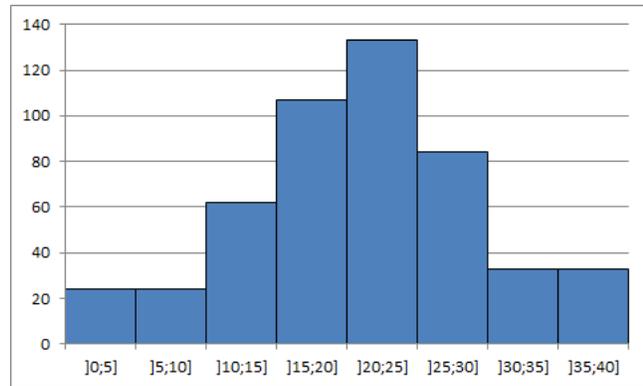


FIGURE 12 – Histogramme correct.

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 $c = x \cdot l = 4,8 \cdot 5 = 24$.

Exercice 12. Un recensement agricole en 2015 a permis de classer les exploitations agricoles selon la surface agricole utilisée. Les résultats sont présentés ci-dessous

Taille de l'exploitation en ha	Effectifs
[0; 20[125'000
[20; 50[44'000
[50; 100[62'000
[100; 200[35'000
[200; 1000[12'000

- a) Déterminer la population ainsi que le caractère étudié.
- b) Comment appelle-t-on les intervalles de la première colonne ?
- c) Pourquoi a-t-on fait un regroupement ?
- d) Quel est le type de cette variable statistique ?
- e) Par quel type de diagramme peut-on représenter ce tableau ?

Exercice 13. Construire correctement l'histogramme correspondant à partir des classes et effectifs suivants.

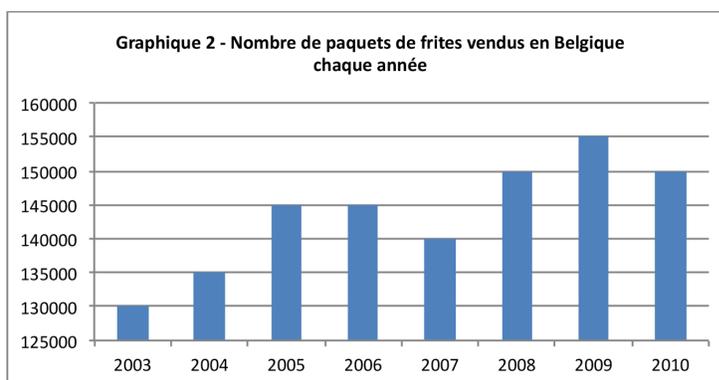
Classes	Effectifs
[50; 100[3
[100; 125[5
[125; 150[4
[150; 175[6
[175; 200[5
[200; 300[2

4.4 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.

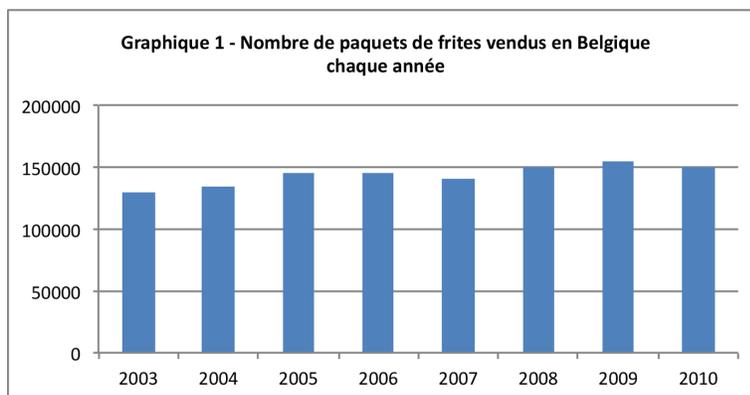
Exemple.

1. 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.
 - a) En dépit des chiffres de 2007, le diagramme en bâtons ci-dessous semble indiquer une augmentation des ventes du nombre de paquets de frites.



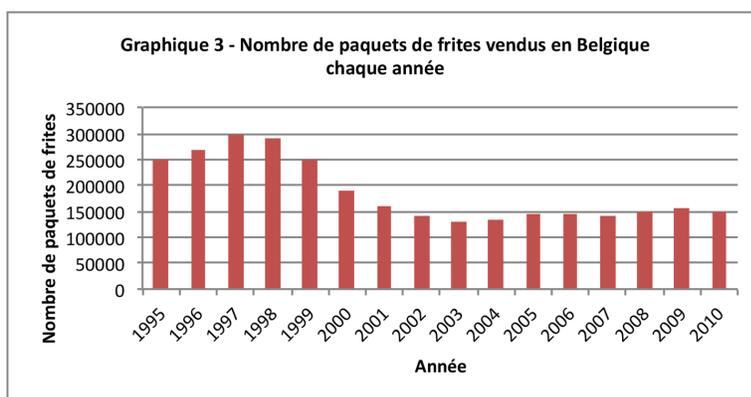
Cependant, on y regardant de plus près, on observe que l'axe des y ne part pas à 0, mais à 125'000 ! Sur ce diagramme, on peut y lire que 130'000 paquets de frites ont été vendus en 2003, contre 135'000 en 2004, ce qui correspond à une augmentation de 5'000 paquets en une année, soit d'environ 3,85%. Or, l'effet visuel du diagramme laisse supposer au premier abord que les ventes ont doublé durant cette période, c'est-à-dire qu'elles ont augmenté de 100% ! Remarquons enfin que le diagramme donne l'impression que les ventes ont été multipliées par 6 entre 2003 et 2009, alors qu'elles sont passées de 130'000 à 155'000, ce qui fait 25'000 de plus, soit une augmentation de presque 20% seulement !

- b) 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.



- c) 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 précédant 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 !



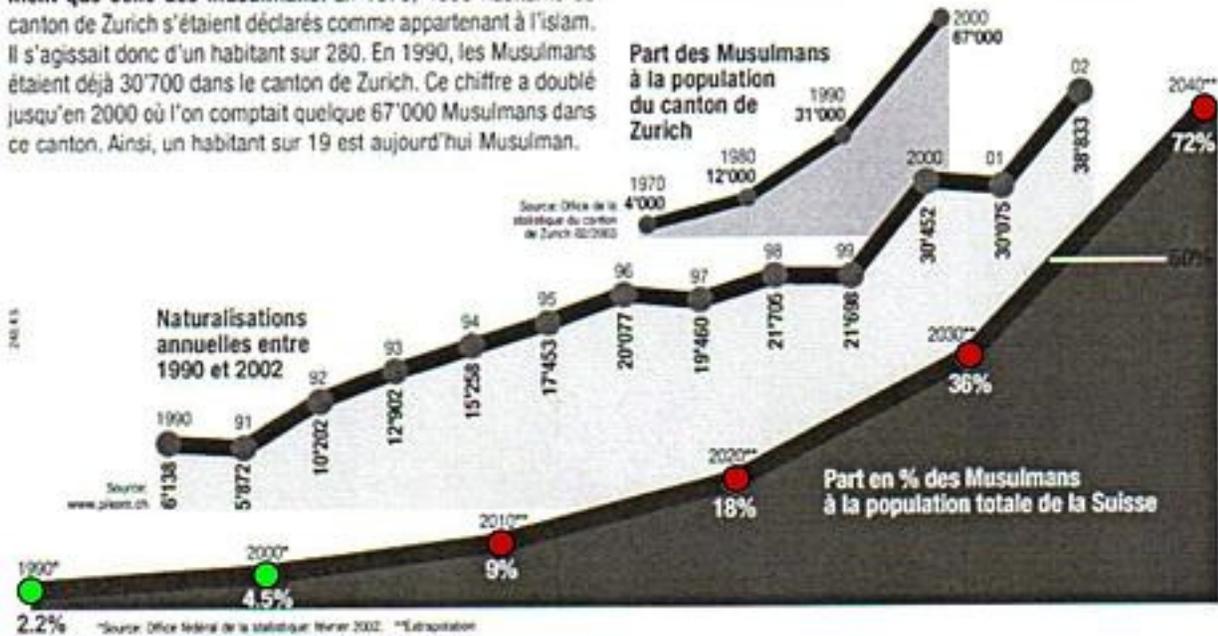
Avec une même étude, il est donc possible de faire passer trois messages complètement différents selon la manière dont on présente l'information.

2. 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



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 zurichoïses, 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 !

3. 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](#).



4. 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 mêmes données de façon correcte, et donne une impression différente.

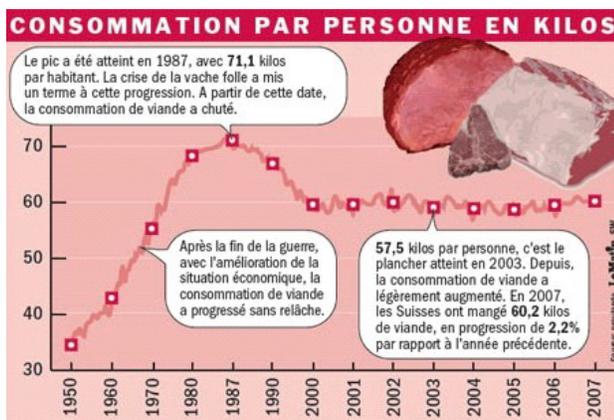


FIGURE 13 – Graphe faux.

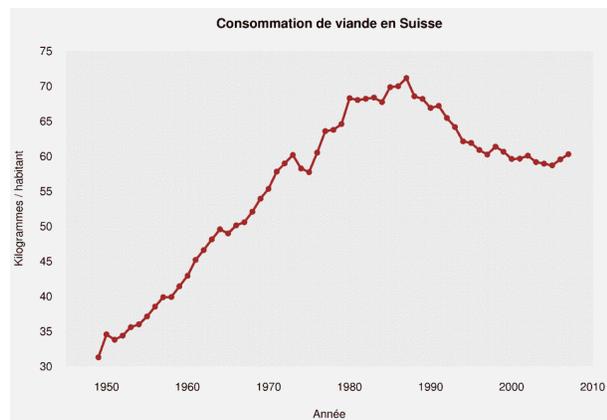
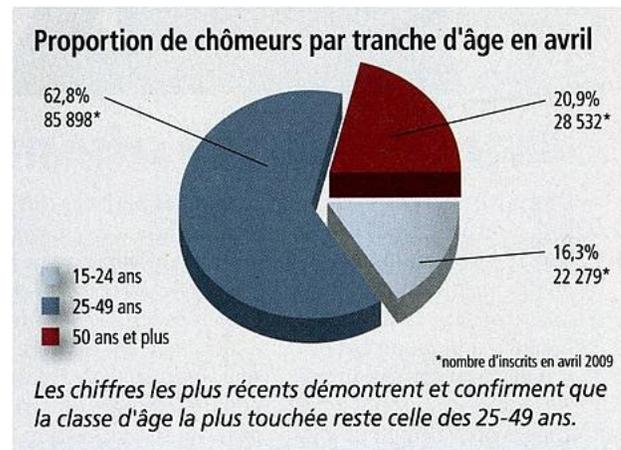


FIGURE 14 – Graphe correct.

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.

5. 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.

6. 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 15 – Diagramme erroné.

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.

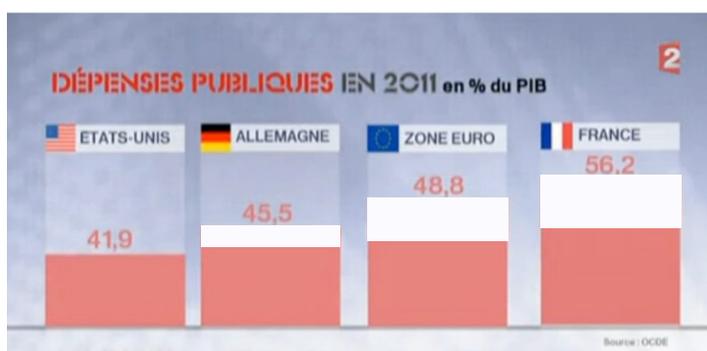


FIGURE 16 – Diagramme correct.

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

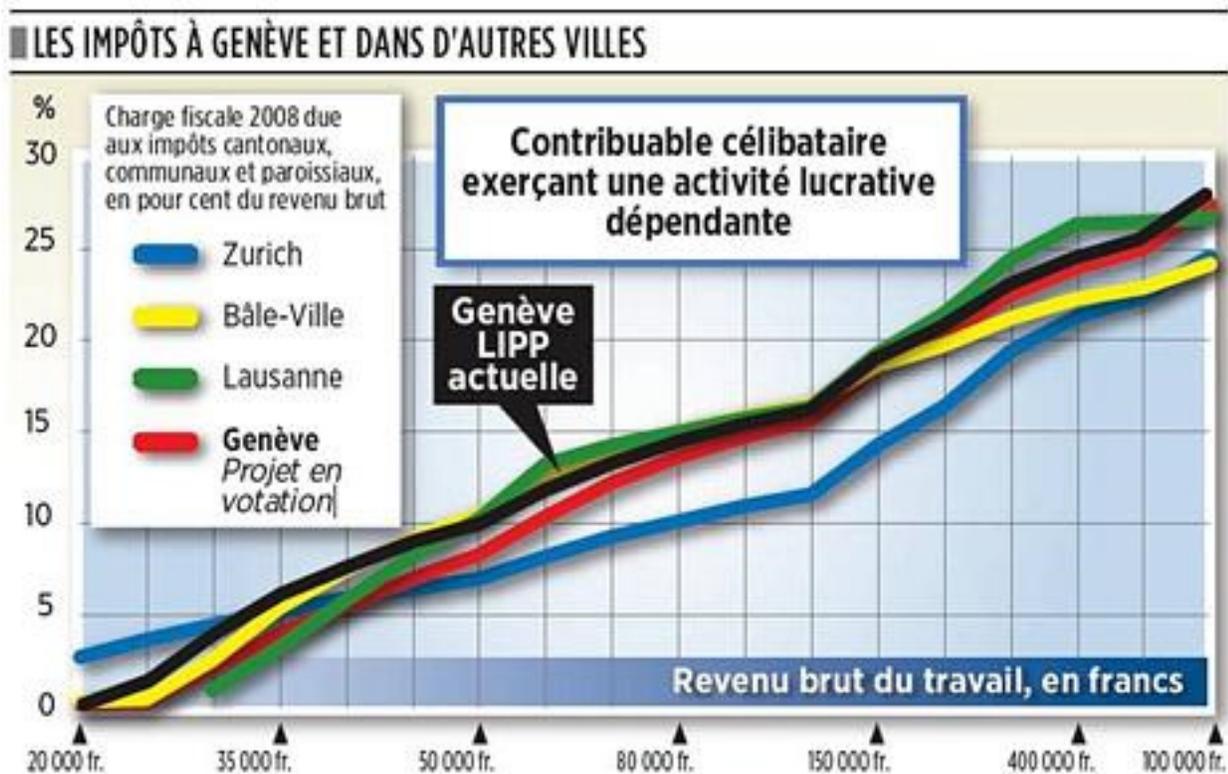


Le graphiste a voulu associer cette augmentation au diamètre des montres. On peut vérifier qu'ils augmentent bien de 40%. Le lecteur voit l'augmentation de la surface des horloges qui, elle, n'est pas de 40%, mais proche de 100% ! Enfin, les aiguilles ont été ajoutées dans un but purement esthétique,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

8. 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%.

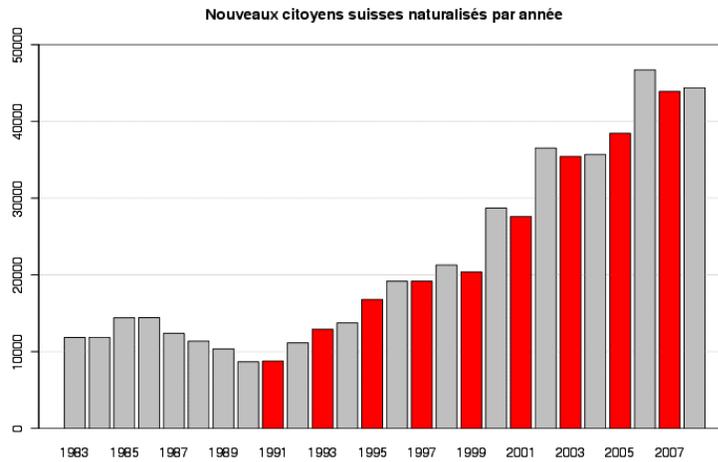


Exercice 14. La figure ci-dessous est parue dans un quotidien genevois.

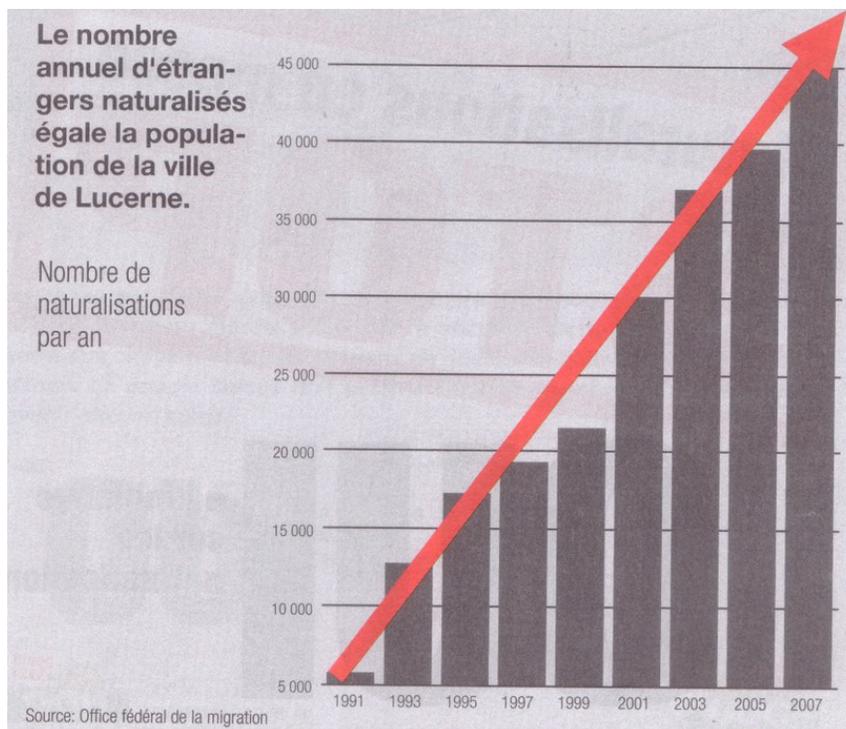


Expliquer en quoi ce graphe est trompeur.

Exercice 15. Le diagramme ci-dessous rend compte du nombre de citoyens suisses naturalisés par année de 1983 à 2008.



Donner deux techniques utilisées pour construire le diagramme suivant.



Exercice 16. La figure ci-dessous datant de 2010 explique comment compter correctement la proportion d'étrangers en Suisse.



Les chiffres ci-dessus reposent sur une estimation de la population à 7,7 millions et à 550'000 naturalisés entre 1985 et 2010.

- Déterminer, en %, le taux d'étrangers en Suisse en 2010, en tenant également compte des sans-papiers, des frontaliers et des requérants d'asile.
- Comment a été obtenu le chiffre de 27,2% ?
- Expliquer comment a été obtenu le chiffre de 34,3% et porter un jugement sur sa pertinence.

4.5 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.

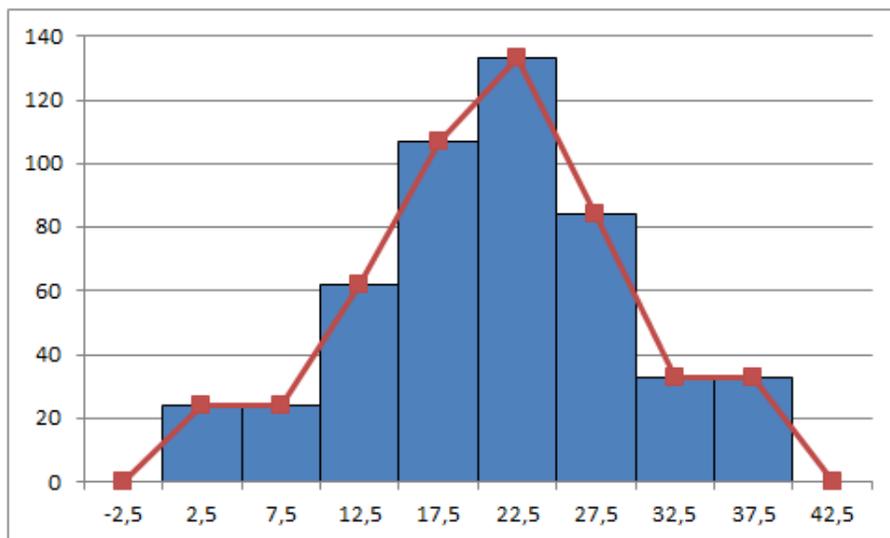


FIGURE 17 – Polygone des effectifs.

Exercice 17. Une entreprise a enregistré le salaire de tous ses vendeurs pour l'année dernière. Voici les données rangées :

Classes (salaires)	Centres	Effectifs	Fréquences
[10000; 15000[12500	2	
[15000; 20000[8	10%
[20000; 25000[22500	14	
[25000; 30000[27500	21	26,25%
[30000; 35000[20%
[35000; 40000[37500	12	15%
[40000; 45000[42500	5	6,25%
[45000; 50000[47500		2,5%
Total		80	100%

- Compléter le tableau des distributions des effectifs et des fréquences.
- Construire l'histogramme correspondant.
- Construire le polygone des effectifs.

4.6 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 : la *polygone des effectifs cumulés croissants* et la *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.

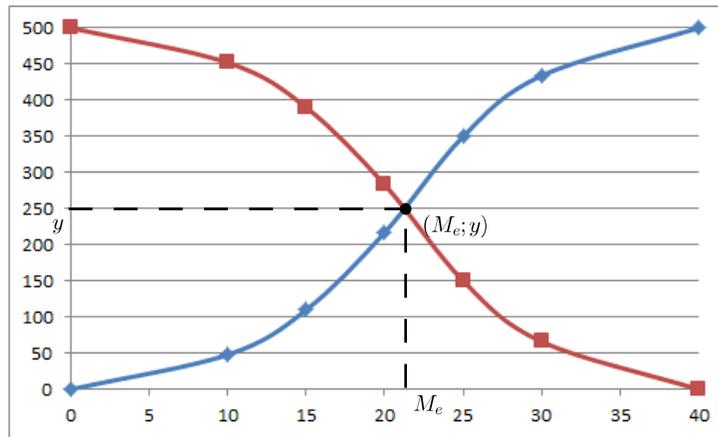


FIGURE 18 – Polygones des effectifs cumulés.

Remarque. L'intersection de ces deux polygones est un point $(M_e; y)$, dont la première coordonnée est appelée *mediane* 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 inférieure à M_e ; le reste, c'est-à-dire $500 - y$, ont une superficie supérieure à M_e . Ce point étant également sur le polygone des effectifs cumulés décroissants, y décrit le nombre d'exploitations ayant une superficie supérieure à M_e . On en déduit 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 .

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 .

Exercice 18. La gendarmerie de Fribourg a relevé sur l'autoroute les vitesses suivantes un samedi soir.

Vitesses (en km/h)	Effectifs
]80; 100]	4
]100; 120]	34
]120; 140]	84
]140; 160]	58
]160; 180]	20
Total	200

- Quel est le type de cette variable statistique ?
- Construire l'histogramme des fréquences, ainsi que le polygone des fréquences.
- Construire le polygone des fréquences cumulées.
- Combien d'automobilistes roulaient trop vite ?
- Combien d'automobilistes seront amendés, sachant qu'une marge de 4% est déduite de la vitesse réelle observée ?

5 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*.

5.1 Moyenne arithmétique

5.1.1 Cas discret

Définition. La *moyenne arithmétique* \bar{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

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_Nx_N}{N}$$

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%

la moyenne arithmétique \bar{x} est donnée par

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

Exercice 19. Calculer la moyenne arithmétique de l'ensemble des nombres

$$E = \{2, 3, 3, 3, 3, 4, 5, 7, 9, 10, 11, 11, 13\}.$$

5.1.2 Cas continu

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

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

Exercice 20. Calculer la moyenne arithmétique de chacune des populations suivantes.

Âges	Effectifs
]0; 20]	72
]20; 65]	180
]65; 100]	43

Salaire (en francs)	Nombre de collaborateurs
]0 ;80]	32
]80 ;100]	48
]100 ;260]	20

5.2 Mode

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*.

5.2.1 Cas discret

Définition. 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

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

Exercice 21. Indiquer le mode pour les deux séries d'observations suivantes :

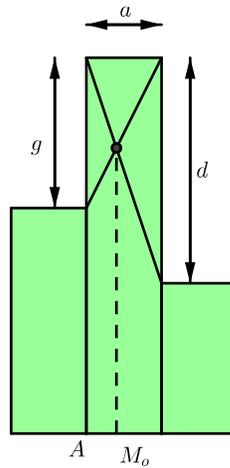
- a) 7, 8, 9, 3, 3, 7, 6, 7, 8, 7, 3, 9, 6, 7
 b) 4, 4, 7, 6, 8, 12, 6, 4, 8, 7, 8, 13, 4, 8, 6, 5

5.2.2 Cas continu

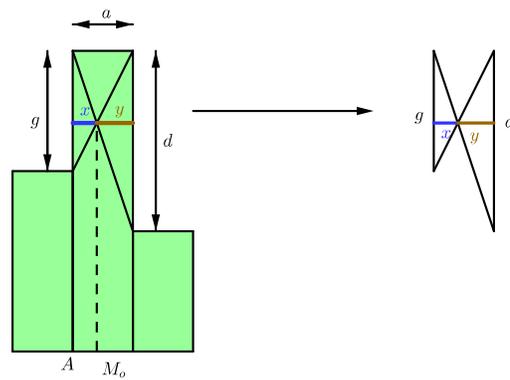
Pour des séries de données groupées par classes, la détermination du mode s'effectue comme suit :

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

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



Preuve de la formule du mode :

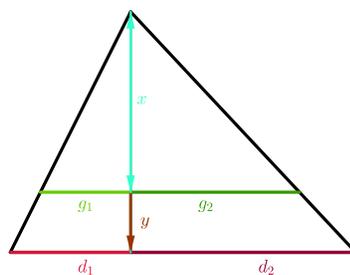


Il est clair que

$$- x = M_o - A.$$

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

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 :

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

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

On a alors

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

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

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

Exemple. Dans notre exemple des exploitations 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

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

Exercice 22. Déterminer le mode de la variable statistique continue suivante.

Classes	Effectifs
[0; 2[3
[2; 4[8
[4; 6[15
[6; 8[14
[8; 10[6
[10; 12[2

Exercice 23. Le tableau ci-dessous rend compte du diamètre de pièces en mm.

Classes	Effectifs
]20; 25]	9
]25; 30]	27
]30; 35]	36
]35; 45]	45
]45; 55]	18
]55; 60]	9
]60; 65]	3
]65; 70]	3

Calculer le mode de cette distribution.

Exercice 24. Déterminer le mode de chacune des populations suivantes.

Ages	Effectifs
]0; 20]	72
]20; 65]	180
]65; 100]	43

Salaire (en francs)	Nombre de collaborateurs
]0 ;80]	32
]80 ;100]	48
]100 ;260]	20

Exercice 25. Calculer les différents modes de cette distribution.

Classes	Effectifs
[18; 20[107
[20; 22[110
[22; 24[91
[24; 28[220

5.3 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. 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 !

5.3.1 Cas discret

Définition. 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
$$M_e = \frac{9 + 11}{2} = 10.$$

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

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

- 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.

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, $M_e = 3$.

Exercice 26. Les valeurs ci-dessous indiquent le nombre de langues parlées par chaque individu sur la base d'un échantillon de 30 salariés d'une compagnie d'assurance.

3 5 2 4 3 4 5 4 3 4
 2 1 2 1 3 4 1 2 5 4
 1 1 1 1 1 2 4 4 2 2

Calculer la médiane de cette distribution.

Exercice 27. On a enregistré au guichet d'une poste (arrondi au kg) de 20 colis envoyés en une heure

3 3 2 3 1 2 1 2 4 4
 4 1 1 1 3 1 4 2 4 1

- Regrouper ces données au sein d'un tableau.
- Déterminer le poids moyen.
- Déterminer le mode.
- Déterminer le poids médian.

Exercice 28. Calculer la moyenne, la médiane et le mode de la variable statistique suivante.

Modalités	Effectifs
10	2
11	3
12	7
13	9
14	14
15	8
16	3
17	1

Exercice 29. Une PME comprend sept employés ainsi que son patron. Le tableau ci-contre indique les salaires mensuels en francs de ces huit personnes.

Jean B.	12'750
Nicole C.	4'950
Alphonse G.	4'200
Jennifer Z.	8'720
Christian A.	4'800
Adrien A.	5'080
Kim D.	4'500
Natacha T.	4'750

- a) Calculer la moyenne des salaires.
- b) Calculer la médiane des salaires.
- c) Un employé supplémentaire est engagé.
 - (a) Quel est son salaire si la moyenne des 9 salaires est de 6'100 francs ? Quelle est alors la médiane ?
 - (b) Quel est son salaire si la médiane est de 4'800 francs ?

Exercice 30. Donner une série statistique dont l'effectif total est 3, la médiane 8, la moyenne arithmétique 7 et dont l'une des valeurs est 4.

Exercice 31. Lors d'un loto, on a demandé à 9 grand-mamans le nombre de petits enfants qu'elles avaient. Voici les réponses obtenues : 10, 5, 6, 8, 9, 10, 6 et 12. Malheureusement, la réponse de la dernière s'est perdue.

- a) Qu'a-t-elle répondu si la moyenne du nombre de petits-enfants est de 8 ?
- b) Qu'a-t-elle pu répondre si la médiane du nombre de petits-enfants est de 8 ?

5.3.2 Cas continu

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

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

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

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.

Remarque. Dans le cas d'une population très nombreuse, il est loisible de convertir tous les effectifs en fréquences et d'attribuer à la médiane la valeur du caractère correspondant à la fréquence de 50% des effectifs cumulés. La valeur obtenue est une approximation d'autant meilleure que l'effectif est important. Dans notre cas, cette convention conduit aux calculs suivants.

Classes	Effectifs	Effectifs cumulés croissants	Fréquence en %	Fréquence cumulés croissants en %
]0; 10]	48	48	9,6	9,6
]10; 15]	62	110	12,4	22
]15; 20]	107	217	21,4	43,4
]20; 25]	133	350	26,6	70
]25; 30]	84	434	16,8	86,8
]30; 40]	66	500	13,2	100

Desquels, il ressort que la médiane est donnée par

$$M_e = 20 + \frac{50 - 43,4}{26,6} \cdot 5 \cong 21,24 \text{ ha.}$$

Approximation très proche de la valeur exacte 21,26 ha.

Exercice 32. Quelle est la médiane de la variable statistique continue suivante ?

Classes	Effectifs
[0; 2[3
[2; 4[8
[4; 6[15
[6; 8[14
[8; 10[6
[10; 12[2

Exercice 33. Une entreprise organise un grand tournoi de quilles. Voici le tableau de distribution des scores :

Classes	Effectifs
[120; 140[1
[140; 160[9
[160; 180[22
[180; 200[51
[200; 220[12
[220; 240[5
Total	100

Calculer la moyenne, puis la médiane.

5.4 Comparaison entre les valeurs centrales

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

La moyenne arithmétique

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.

Le mode

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.

La médiane

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.

Exercice 34. Un contrôle radar a permis de mesurer la vitesse de 10 automobilistes sur une route limitée à 60 km/h.

70 60 80 60 60 60 90 200 70 60

Quelle mesure de tendance centrale décrit le mieux cette situation ?

Exercice 35. Soient les échantillons ci-dessous.

Echantillon A	2	9	3	8	3	9	3	2	10	3
Echantillon B	9	2	2	8	3	10	3	9	3	8

- Quelle est la moyenne arithmétique et la médiane des deux échantillons ?
- Laquelle de ces deux mesures est la mieux adaptée pour décrire la situation ?

Exercice 36. Dans notre classe, nous sommes 10 étudiants. Au cours de 3 examens différents je n'ai obtenu que 8 points sur 20, ce qui est assez lamentable. À l'aide des indicateurs de tendance centrale, comment vais-je arriver à présenter ces résultats à mon employeur pour que mes résultats ne paraissent pas si mauvais que ça ?

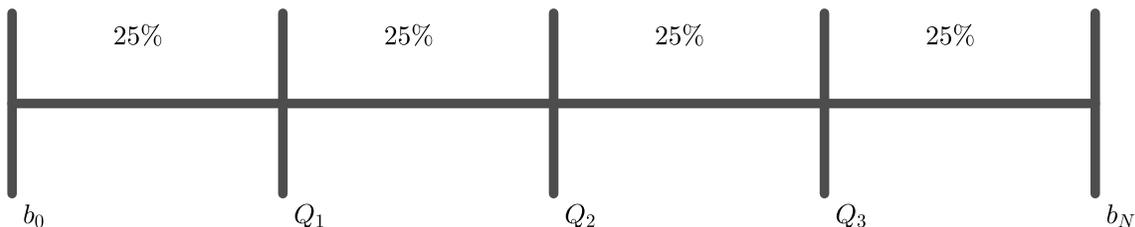
Examen	Moi	Marc	Lili	Jojo	Bob	Fred	Karl	Léa	Luc	Jo
1 ^{er}	8	2	2	2	9	9	9	9	10	19
2 ^{er}	8	2	3	4	5	7	9	9	18	19
3 ^{er}	8	2	7	7	7	10	11	12	18	19

6 Quartiles et boîte à moustaches

6.1 Quartiles

6.1.1 Cas discret

Définition. 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.

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

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 $\frac{50+2}{4} = 13$. Donc $Q_1 = 2$. Quant au quartile Q_3 , il est égal à la valeur de l'observation de rang $\frac{3 \cdot 50 + 2}{4} = 38$. 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%.

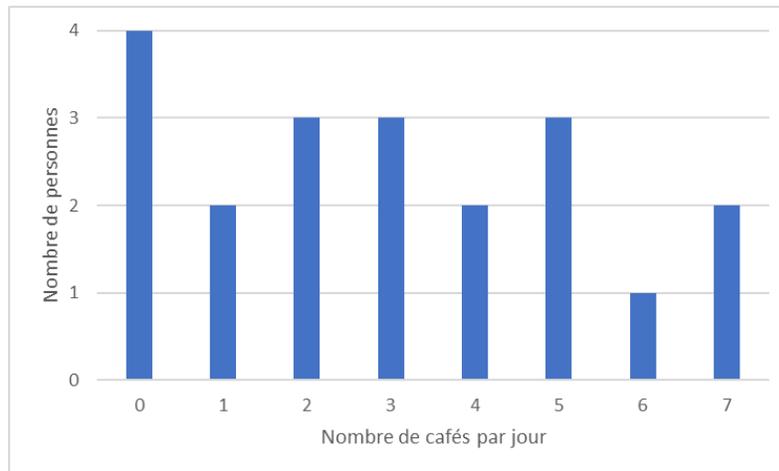
Remarque. Il est possible de généraliser la notion de quartile à celle de *quantile d'ordre n* . Les autres quantiles les plus souvent utilisés sont :

- Les *déciles* D_1, D_2, \dots, 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, \dots , et ainsi de suite. Le décile D_5 est égal à la médiane ;
- Les *centiles* C_1, C_2, \dots, C_{99} partagent la population en 100 groupes d'effectifs égaux.

Exercice 37. Calculer les trois quartiles de la distribution ci-dessous.

1	1	4	8	8	10	10	10	15	20
1	1	4	8	8	10	10	10	15	20
1	4	8	8	8	10	10	10	15	20
1	4	8	8	8	10	10	10	15	20
1	4	8	8	8	10	10	10	20	20

Exercice 38. Ci-dessous figure un histogramme représentant la consommation quotidienne de cafés de toutes les personnes travaillant dans une PME.



- a) Combien d'employés y a-t-il dans cette PME ?
- b) Combien de cafés chaque employé boit-il en moyenne ?
- c) Donner la médiane, le premier et le troisième quartiles.
- d) Déterminer le mode.

Exercice 39. Ajouter un nombre à la liste de nombres

$$-12; -10; -6; -3; 0; 3; 5; 8; 9; 11; 13$$

de manière à avoir la médiane à 4 et le troisième quartile à 9.

Exercice 40. A un concours de saut en longueur, on a mesuré les performances ci-dessous exprimées en mètres :

3,45 3,60 3,70 2,95 3,00 4,10 3,10 3,90
 3,10 3,20 4,50 3,15 3,20 3,05 3,60 3,55
 4,00 3,35 3,55 2,70 2,95 3,05 3,85 3,40

- a) Donner la médiane, le maximum, le minimum et les quartiles.
 b) Les énoncés suivants sont-ils corrects ? Justifier.
 (a) Environ 75% ont fait un saut de plus de 3,60 m.
 (b) Environ 50% ont sauté au plus 3,35 m.
 (c) Environ 50% ont sauté au moins 3,35 m.

6.1.2 Cas continu

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

On connaît déjà la médiane $Q_2 = M_e \cong 21,26$.

Le premier quartile Q_1 est la valeur de la superficie de l'exploitation de rang $\frac{500+2}{4} = 125,5$. Comme elle se trouve dans la troisième classe, d'effectif 107, on a

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

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

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

Exercice 41. Le tableau ci-dessous recense la taille en centimètres des élèves d'une école.

Taille des élèves	Nombre d'élèves
[120; 130[6
[130; 140[21
[140; 150[45
[150; 160[55
[160; 170[26
[170; 180[7

Calculer la valeur des quartiles Q_1 et Q_3 .

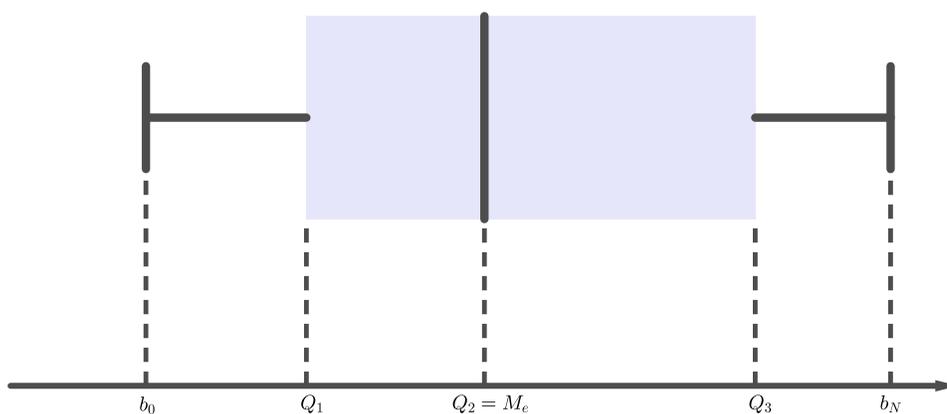
Exercice 42. Un échantillon de 200 sportifs est à la base du tableau ci-dessous.

Taille (en cm)	Effectifs	Fréquences cumulées
[150; 160[0,2
[160; 170[60	
[170; 180[0,6
[180; 190[30	
[190; 200[

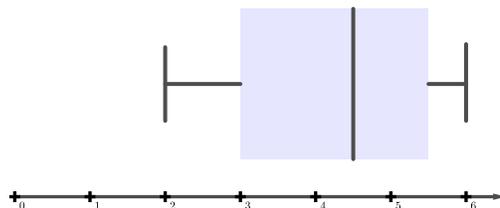
- Compléter le tableau ci-dessus.
- Calculer la valeur du mode, de la médiane et des quartiles Q_1 et Q_3 .

6.2 Boîte à moustaches

Définition. 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.



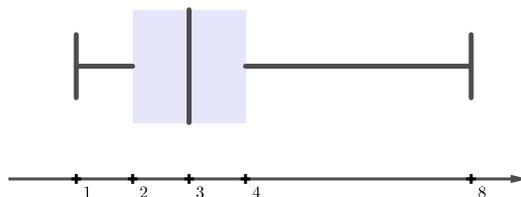
Exemple. Les notes d'une classe ont été représentées à l'aide de la boîte à moustache ci-dessous



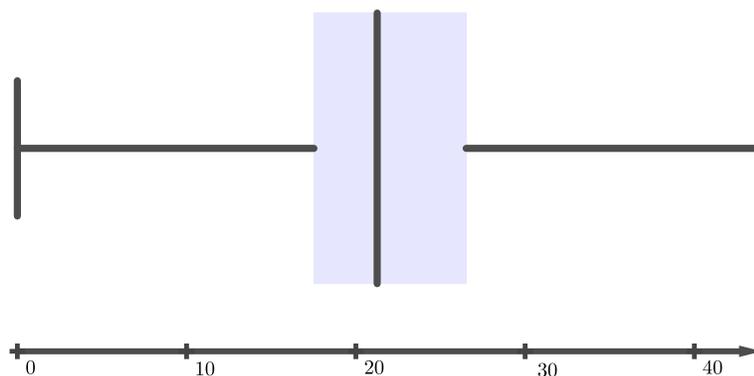
Cette boîte à moustaches fournit les informations suivantes :

- 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.

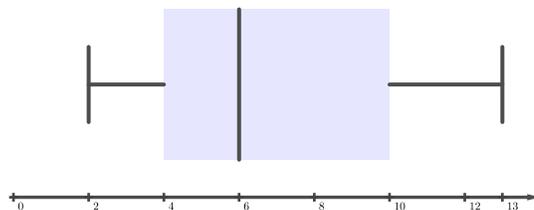
Exemple. Dans notre exemple du nombre de personnes par ménage, la boîte à moustaches est donnée par



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

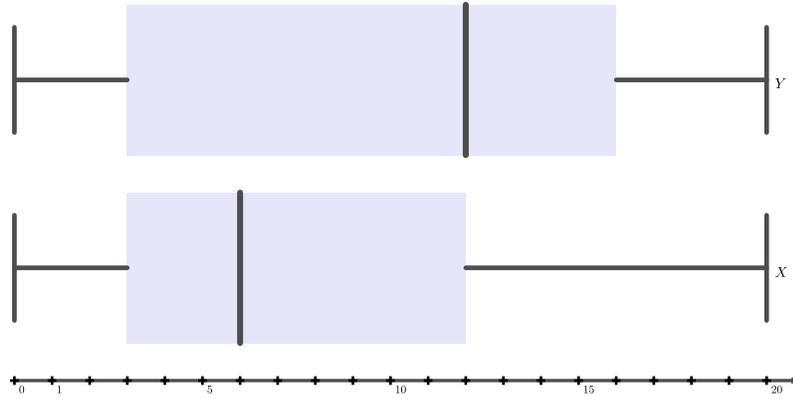


Exercice 43. Dans une entreprise, on interroge un certain nombre d'employés pour savoir combien de mails ils ont envoyé durant une journée. Les résultats sont synthétisés par la boîte à moustaches ci-dessous.



- Déterminer la valeur des trois quartiles.
- Entre combien et combien de mails les employés ont-ils envoyé ?
- La moitié des employés interrogés ont envoyé plus de combien de mails ?
- Quel pourcentage d'employés interrogés a envoyé entre 4 et 10 mails ?

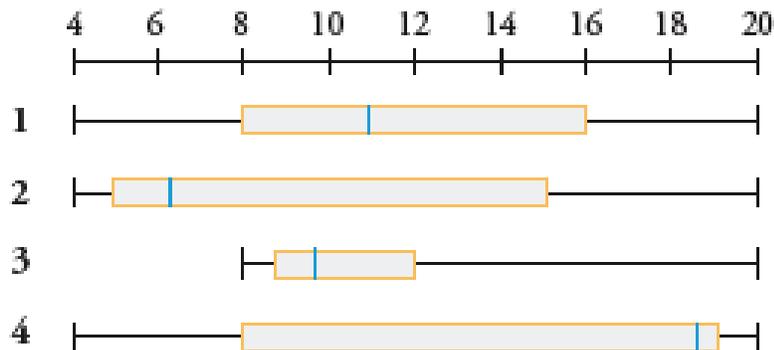
Exercice 44. Dans deux villes X et Y , on a sélectionné un échantillon de 1000 personnes à qui on a demandé combien de cigarettes ils fumaient quotidiennement. Les résultats sont présentés au sein des boîtes à moustache ci-dessous.



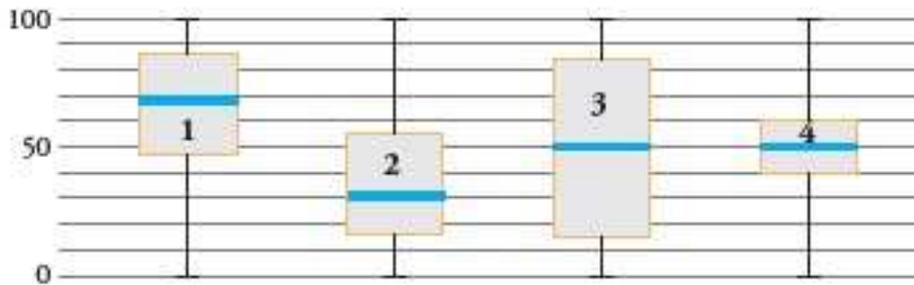
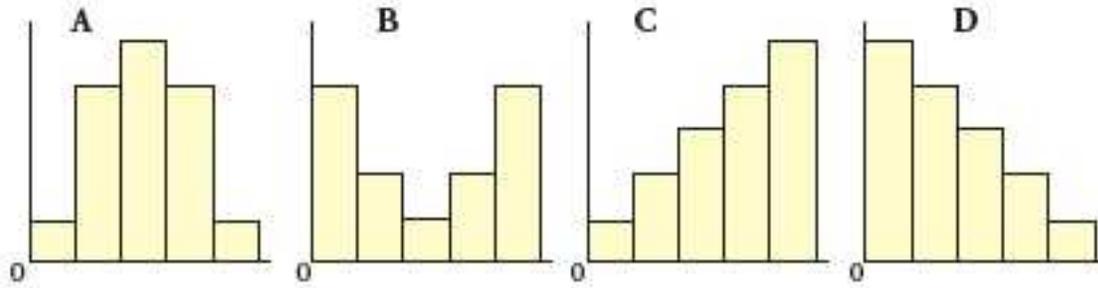
- a) Quelle est la valeur des quartiles Q_1 , Q_2 et Q_3 pour les deux villes ?
- b) Quelle est la ville la plus consommatrice de cigarettes ?
- c) Peut-on affirmer qu'un quart de la population de la ville X fume plus de 3 cigarettes par jour ?
- d) Peut-on affirmer que la moitié des habitants de la ville Y fume moins de 11 cigarettes par jour ?

Exercice 45. Associer les quatre distributions suivantes au boîtes à moustaches ci-dessous.

X	4	5	8	9	18	19	19	19	19	20
Y	8	9	9	9	9	10	11	12	18	20
Z	4	4	5	5	6	7	10	15	18	20
U	4	6	8	8	10	12	16	16	16	20



Exercice 46. Associer chacun des quatre histogrammes ci-dessous aux boîtes à moustaches correspondantes.



Exercice 47. A une station météo, la température extérieure a été mesurée et enregistrée, chaque jour à midi, durant tout le mois de février. Il en résulte le tableau des températures exprimées en °C.

Jour	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Température	7	-2	0	1	-5	-9	-6	-5	-2	4	7	11	12	10

Jour	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Température	10	10	7	4	3	-2	-3	-6	-1	4	7	11	12	10

Représenter la boîte à moustaches correspondant à ces mesures.

Exercice 48. 3 classes de 34 élèves ont obtenu les points suivants au dernier test évalué sur 20 points.

Classe 1 :

5 10 15 12 13 14 12
 5 8 9 6 12 14 18
 16 15 19 14 12 10 11
 12 16 12 15 4 12 8
 9 14 5 3 18 17

Classe 2 :

8	8	9	10	8	9	10
11	8	9	12	8	12	11
10	12	11	10	12	10	11
12	9	8	7	13	10	11
12	11	10	8	10	11	

Classe 3 :

2	18	5	15	3	17	6
14	0	20	1	19	8	12
9	11	10	10	12	8	11
9	16	4	15	5	12	8
9	11	9	11	19	1	

- Proposer le tableau d'effectif pour les 3 classes.
- Réaliser le diagramme en boîte à moustaches.
- Quelles sont les moyennes pour chaque classe ?
- À partir des informations obtenues, indiquer la ou les classes répondant le mieux aux critères suivants :

Critère *A* : "Les élèves ont des résultats proches. La classe est homogène".

Critère *B* : "Les élèves ont des résultats très différents. La classe est hétérogène".

Critère *C* : "Cette classe possède les meilleurs résultats".

Critère *D* : "50% des élèves au moins ont une note comprise entre 8 et 12".

Critère *E* : "25% des élèves au plus ont une note comprise entre 8 et 12".

7 Mesures de 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 : *l'étendue*, *la variance* et *l'écart-type*.

7.1 Etendue de la série

C'est la valeur de dispersion la plus simple.

Définition. 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.

Exercice 49. La série d'observations suivantes porte sur le relevé des températures durant le mois de février dans une ville.

-10	-9	-5	-12	-8	-5	-2
5	2	3	5	6	2	0
1	-1	-2	0	-6	-10	-11
-9	-6	-4	-3	0	0	1

Calculer l'étendue de la série.

7.2 Ecart interquartile

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

$$I_Q = Q_3 - Q_1.$$

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

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

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

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

$$I_D = D_9 - D_1.$$

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

Exercice 50. Le tableau ci-dessous donne la capacité des stades des équipes de la saison 2019-2020 de Premier League de football.

Club	Stade	Club	Stade
AFC Bournemouth	11'464	Everton FC	39'571
Watford FC	20'877	Chelsea FC	41'798
Burnley FC	21'401	Aston Villa	42'788
Crystal Palace	26'255	Newcastle United	52'405
Norwich City	27'244	Liverpool FC	54'074
Brighton & Hove Albion	30'750	Manchester City	55'097
Wolverhampton Wanderers	31'700	West Ham United	60'000
Southampton FC	32'505	Arsenal FC	60'704
Sheffield United	32'702	Tottenham Hotspur	62'062
Leicester City	36'262	Manchester United	75'635

Donner la valeur des quartiles Q_1 , Q_2 et Q_3 , puis en déduire l'écart interquartile et l'écart semi-interquartile.

Exercice 51. Une compagnie a révélé les chiffres des absences de ses employés pour le mois dernier.

Nombre de jours d'absence	Nombre d'employés
0	36
1	42
2	20
3	11
4	3
5	2
12	1

- Calculer l'étendue et l'écart semi-interquartile.
- Calculer la proportion des employés ayant manqué plus de deux jours de travail.

7.3 Variance écart-type

7.3.1 Cas discret

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

TABLE 1 – 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

TABLE 2 – 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 ?

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

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

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

Ces deux moyennes \bar{x} et \bar{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.

On obtient :

$x_i - \bar{x}$	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

$y_i - \bar{y}$	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.

On obtient alors les distributions suivantes :

$(x_i - \bar{x})^2$	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

$(y_i - \bar{y})^2$	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

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

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

$$\begin{aligned} \overline{(y_i - \bar{y})^2} &= \frac{7,5625 \cdot 1 + 3,0625 \cdot 2 + 1,5625 \cdot 4 + 0,0625 \cdot 3 + 0,5625 \cdot 6 + 1,5625 \cdot 2 + 3,0625 \cdot 2}{20} \\ &= 1,6375. \end{aligned}$$

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.

Définition. 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

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

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

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

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 .

La moyenne arithmétique de ces nombres vaut

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

Du tableau suivant

	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
	-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. On rappelle que la moyenne arithmétique valait 3,44.

Modalités x_i	Effectifs n_i	Écarts $x_i - \bar{x}$	Carrés des écarts $(x_i - \bar{x})^2$	Produits $n_i \cdot (x_i - \bar{x})^2$
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

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

Exercice 52. Calculer la variance et l'écart-type des nombres suivants : $-9, -4, 1, 7, 10, 21$.

Exercice 53. Un casino a demandé à son croupier de noter durant 60 jours consécutifs le nombre de fois par jour où le 0 sort au jeu de la roulette. Il a obtenu les données suivantes.

Nombre de 0 par jour	Nombre de jours
7	1
8	3
9	6
10	9
11	14
12	11
13	7
14	6
15	2
16	1

- Calculer les mesures de tendance centrale.
- Calculer la variance et l'écart-type.

7.3.2 Cas continu

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

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 $(c_i - \bar{x})^2$	Produits $n_i \cdot (c_i - \bar{x})^2$
]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 $V = \frac{34862,5}{500} = 69,725 \text{ ha}^2$ et l'écart-type est donné par $\sigma = \sqrt{69,725 \text{ ha}^2} \cong 8,35 \text{ ha}$.

Exercice 54. Déterminer la variance et l'écart-type de chacune des populations suivantes.

Ages	Effectifs
]0; 20]	72
]20; 65]	180
]65; 100]	43

Salaire (en francs)	Nombre de collaborateurs
]0 ;80]	32
]80 ;100]	48
]100 ;260]	20

7.3.3 Autre méthode de calcul

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 \bar{x}^2 . Ainsi, on a

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

Preuve. On a

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

□

Remarque. Cette seconde formulation sera préférée à la première chaque fois que les termes $x_i - \bar{x}$ 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 $7,2$.

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 $V = \frac{516}{5} - 7,2^2 = 51,36$ 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.

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 $\overline{x^2} = \frac{255362,5}{500} = 510,725$. Comme $\bar{x} = 21$, $\bar{x}^2 = 441$, il suit que

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

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

Exercice 55. Appliquer cette méthode simplificatrice aux exercices 52 et 54.

Exercice 56. On donne la série suivante :

2 3 5 5 4 4 5 2 2 4

Calculer b_0 , b_{10} , Q_1 , M_e , Q_3 , \bar{x} , $\overline{x^2}$, V et σ .

Exercice 57. Des mesures de vitesse d'automobiles ont donné les résultats suivants.

Vitesse	Nombre
]30; 35]	15
]35; 40]	39
]40; 45]	65
]45; 50]	87
]50; 55]	70
]55; 60]	22
]60; 65]	12

Calculer la moyenne arithmétique et l'écart-type.

Exercice 58. Une étude des salaires annuels des employés d'une grande compagnie a donné les résultats suivants.

Classes	Effectifs
[20'000; 22'000[80
[22'000; 24'000[130
[24'000; 26'000[340
[26'000; 28'000[210
[28'000; 30'000[120
[30'000; 32'000[40

- a) Calculer la moyenne arithmétique.
- b) Calculer la variance et l'écart-type.

7.4 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-interquartile. 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*.

Définition. 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 :

$$C = \frac{\sigma}{\bar{x}}$$

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

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

Ainsi, la dispersion des données est élevée.

Exercice 59. Au cours du mois de février, sur le marché boursier américain de l'industrie électronique, la moyenne des prix quotidiens à la fermeture a été de 1'500 \$ avec un écart-type de 50 \$ pour la catégorie A d'actions, tandis que pour la catégorie B , durant la même période, la moyenne a été de 500 \$ et l'écart-type de 30 \$.

- Déterminer le coefficient de variation pour le titre A .
- Idem pour le titre B .
- Comparer les résultats et les interpréter.

Exercice 60. Les nombres suivants indiquent la masse en grammes de 20 poussins :

67; 73; 76; 82; 60; 62; 60; 62; 55; 64; 64; 55; 75; 66; 61; 69; 72; 73; 54; 59.

Calculer la moyenne arithmétique, la médiane, l'écart-type et le coefficient de variation.

Exercice 61. Soit la population ci-dessous.

Somme dépensée en francs	Nombre de clients
[0; 20[141
[20; 40[239
[40; 60[451
[60; 80[578
[80; 100[321
[100; 200[294

Calculer l'écart-type et le coefficient de variation.

7.5 Comparaison des mesures de dispersion

L'étendue

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.

L'écart semi-interquartile

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.

L'écart interquartile

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

L'écart-type

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.

La variance

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 :

mode	↔	étendue
médiane	↔	écart semi-interquartile
moyenne	↔	écart-type

8 Solutions

Exercice 1.

- a) La population étudiée est "les voitures particulières neuves immatriculées en 2015".
- b) Le caractère étudié est "type d'énergie consommée par ces voitures".

Exercice 2.

- a) Quantitative continue.
- b) Quantitative discrète.
- c) Qualitative nominale.
- d) Quantitative continue.
- e) Qualitative ordinale.
- f) Quantitative discrète.
- g) Qualitative ordinale.
- h) Qualitative nominale.
- i) Quantitative continue.
- j) Quantitative discrète.
- k) Quantitative discrète.
- l) Qualitative ordinale.
- m) Qualitative nominale.
- n) Quantitative discrète.
- o) Quantitative continue.
- p) Qualitative ordinale.

Exercice 3.

- a) La population : les employés d'une entreprise. La variable statistique : le parti politique pour lequel ils ont voté.
- b) Les modalités : {PS; PLR; PDC; UDC; verts}.
- c) Cette variable statistique est qualitative ordinale.

Exercice 4.

- a) La population : les 80 étudiants du professeur de l'Uni. La variable statistique : le nombre de points obtenus lors d'un test statistique.
- b) Les modalités : {2; 3; ...; 10}.
- c) Elle est du type quantitative discrète.

Exercice 5. La 52^{ème} exploitation de la classe]20; 25] a une superficie de 21,95 ha.
La 78^{ème} exploitation de la classe]15; 20] a une superficie de 18,64 ha.

Exercice 6.

Classes	Centres	Effectifs	Fréquences
[30; 45[37,5	1	1, $\bar{6}$ %
[45; 60[52,5	2	3, $\bar{3}$ %
[60; 75[67,5	9	15%
[75; 90[82,5	35	58, $\bar{3}$ %
[90; 105[97,5	12	20%
[105; 120[112,5	1	1, $\bar{6}$ %
Total		60	100%

Exercice 7.

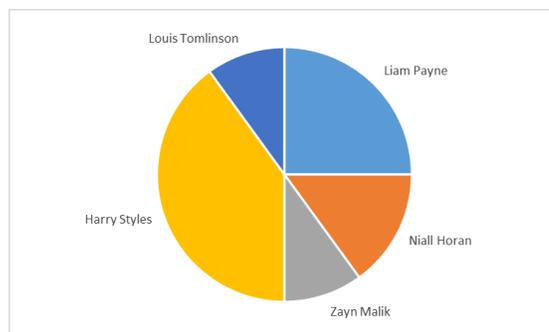
- a) Faux, ce sont les "dépenses culturelles".
- b) Vrai.
- c) Faux. Cela n'a pas de sens.
- d) Faux. Le total doit être de 100%.

Exercice 8.

a)

Chanteur	Effectifs	Fréquences (en %)	Angle (en degrés)
Liam Payne	350	25	90°
Niall Horan	210	15	54°
Zayn Malik	140	10	36°
Harry Styles	560	40	144°
Louis Tomlinson	140	10	36°

b)

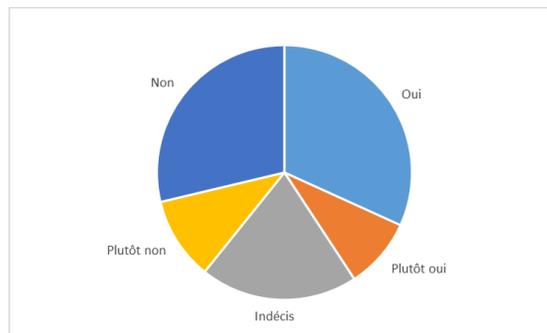


Exercice 9.

a)

Avis	Fréquences
Oui	31,79%
Plutôt oui	8,97%
Indécis	20%
Plutôt non	10,51%
Non	28,72%

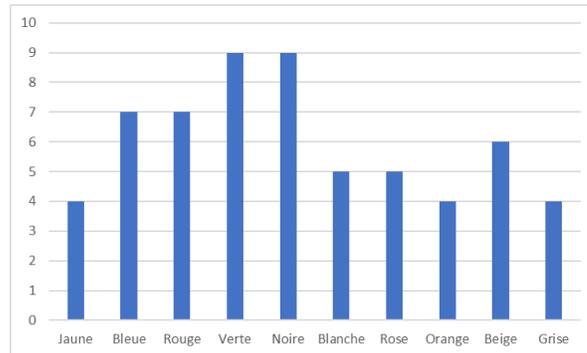
b)

**Exercice 10.**

- a) La population est "les robes du magasin de vêtements" et la variable statistique est "la couleur des robes".
- b) Les modalités : {jaune; bleue; rouge; verte; noire; blanche; rose; orange; beige; grise}.
- c)

Couleurs	Effectifs	Fréquences
Jaune	4	6,6%
Bleue	7	11,6%
Rouge	7	11,6%
Verte	9	15%
Noire	9	15%
Blanche	5	8,3%
Rose	5	8,3%
Orange	4	6,6%
Beige	6	10%
Grise	4	6,6%
Total	60	100%

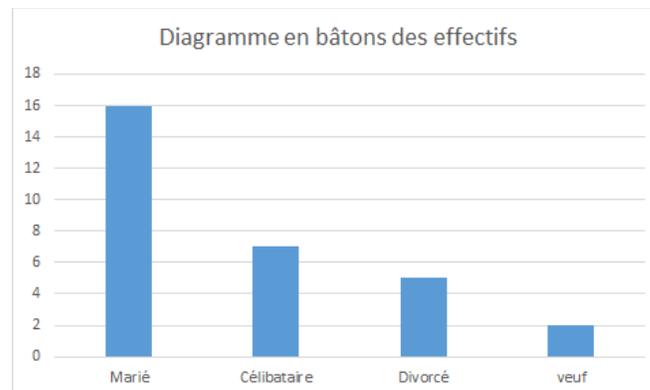
d)

**Exercice 11.**

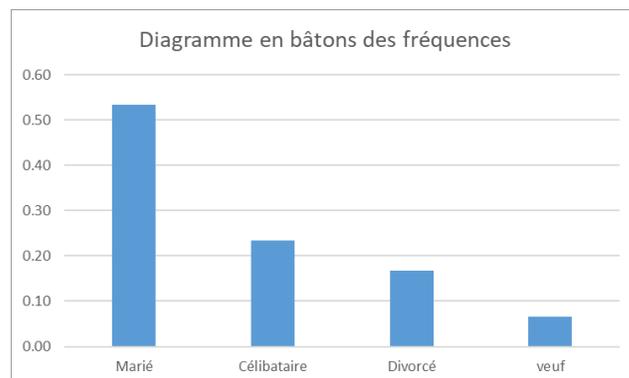
- a) 30 employés d'une petite entreprise.
 b) Leur état civil
 c) {Marié, Célibataire, Divorcé, Veuf}.
 d)

Modalités	Effectifs	Fréquences
Marié	16	53,3%
Célibataire	7	23,3%
Divorcé	5	16,6%
Veuf	2	6,3%
Total	30	100%

e)



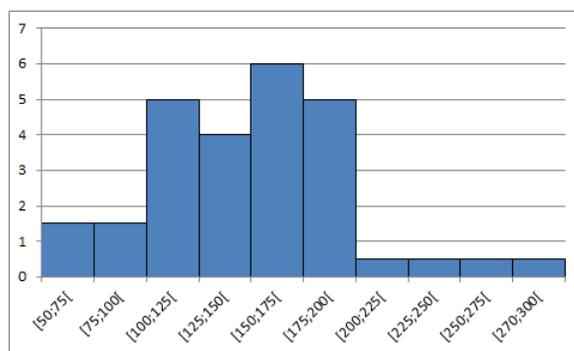
f)



g) Ils sont semblables.

Exercice 12.

- La population est "les exploitations agricoles en 2015" et le caractère étudié est "leur surface".
- Des classes.
- Car il y a beaucoup de valeurs différentes du caractère. Cela facilite la lecture.
- Quantitative continue.
- Par exemple un histogramme.

Exercice 13.

Exercice 14. Ce graphe va donner une image tronquée de la réalité dans la mesure où l'échelle horizontale n'est pas linéaire.

Exercice 15.

- L'origine de l'axe vertical n'est pas à 0, mais juste au dessus de 5000, donnant l'impression d'une multiplication par 6 des naturalisations entre 1991 et 1993 (au lieu d'une multiplication par 2,5).
- L'année 1990 a été choisie comme point de départ parce qu'elle représente l'année avec le plus faible nombre de naturalisations depuis.
- Notons enfin que les chiffres sont absolus, et ne tiennent pas compte de l'augmentation de la population d'une année à l'autre.

Exercice 16.

- a) Le taux d'étrangers en Suisse en 2010 se montait à

$$\frac{2'092'309}{7'700'000 + 200'000 + 212'566 + 40'797} \cong 25,7\%.$$

- b) Aux 1'638'949 étrangers faisant partie des 7,7 millions, on en a ajouté

$$200'000 + 212'566 + 40'797 = 453'363,$$

qu'il convient également de rajouter au total de 7,7 millions. Autrement dit, ce chiffre a été obtenu en calculant

$$\frac{2'092'308}{7'700'000}$$

- c) En plus des 2'092'309 étrangers obtenus ci-dessus, on a ajouté les 550'000 naturalisés entre 1985 et 2010, ce qui donne un total de 2'642'309. Il a ensuite été calculé le rapport

$$\frac{2'642'309}{7'700'000} \cong 34,3\%.$$

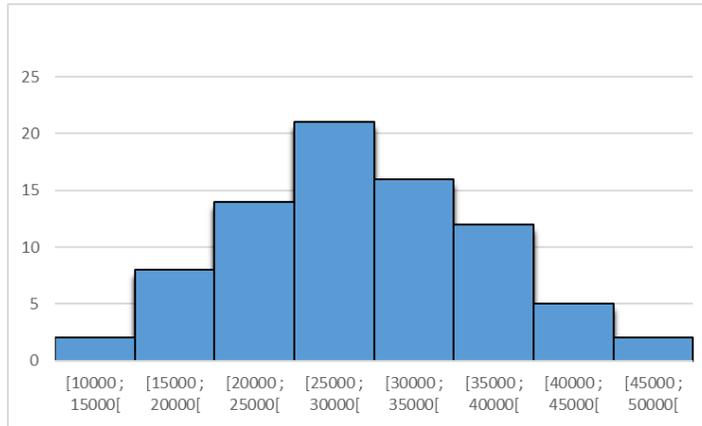
Ce calcul pose deux problèmes. D'une part, le total n'est pas correct, comme au point précédent. D'autre part, ce raisonnement repose sur l'hypothèse selon laquelle les 550'000 naturalisés depuis 25 ans sont tous en vie et résident toujours en Suisse!

Exercice 17.

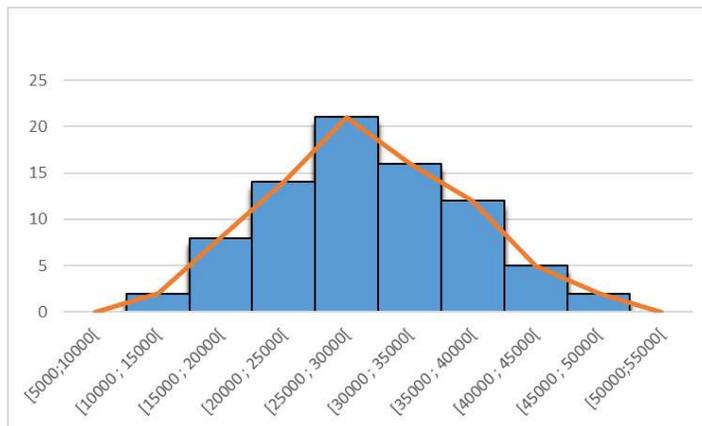
- a)

Classes (salaires)	Centres	Effectifs	Fréquences
[10'000; 15'000[12'500	2	2,5%
[15'000; 20'000[17'500	8	10%
[20'000; 25'000[22'500	14	17,5%
[25'000; 30'000[27'500	21	26,25%
[30'000; 35'000[32'500	16	20%
[35'000; 40'000[37'500	12	15%
[40'000; 45'000[42'500	5	6,25%
[45'000; 50'000[47'500	2	2,5%
Total		80	100%

b)



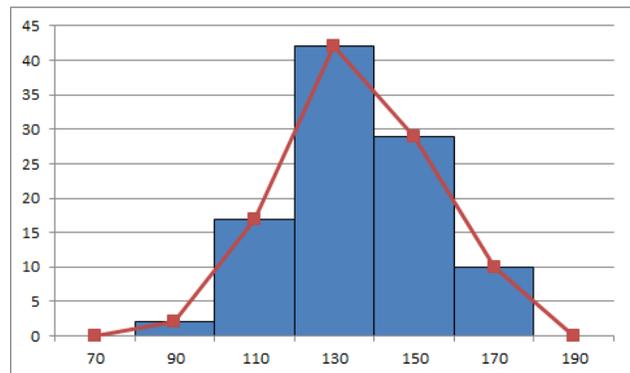
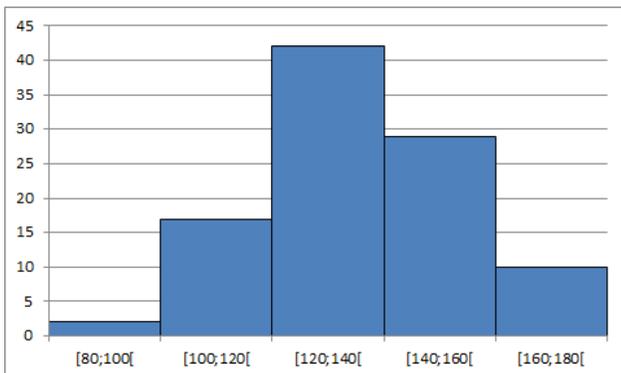
c)



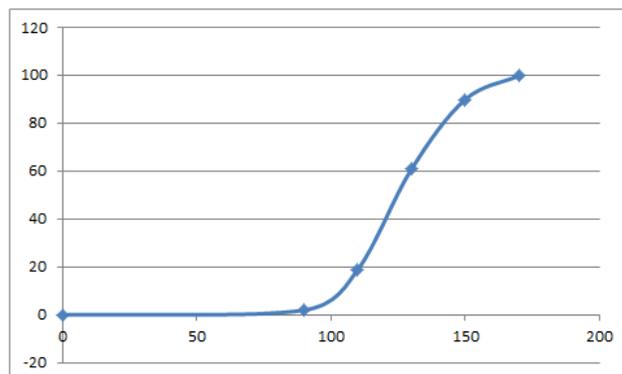
Exercice 18.

a) Quantitative continue.

b)



c)



d) 162.

e) 141.

Exercice 19. $\bar{x} \cong 6,46$.**Exercice 20.** $\bar{x} = 40,4$, $\bar{x} = 92$.**Exercice 21.**a) $M_o = 7$.b) $M_o = 4$ et $M_o = 8$.**Exercice 22.** $M_o = 5,75$.**Exercice 23.** $M_o = 37,5$.**Exercice 24.** $M_o \cong 29,837$ et $M_o = 87,27$.**Exercice 25.** Les modes sont $M_o = 20,27$ et $M_o \cong 24,589$.**Exercice 26.** $M_e = 2,5$.**Exercice 27.**

a)

Poids	Effectifs	Effectifs cumulés
1	7	7
2	4	11
3	4	15
4	5	20

b) $\bar{x} = 2,35$ kg.c) $M_o = 1$ kg.d) $M_e = 2$ kg.

Exercice 28. $\bar{x} \cong 13,51$, $M_e = 14$ et $M_o = 14$.

Exercice 29.

- a) 6'218,75 francs.
- b) 4'875 francs.
- c) (a) 5'150 francs et la médiane devient 4'950 francs.
(b) Il sera compris entre 0 et 4'800 francs.

Exercice 30. 4, 8 et 9.

Exercice 31.

- a) 6 petits-enfants.
- b) Elle a pu répondre n'importe quel nombre inférieur ou égal à 8.

Exercice 32. $M_e = 5,8$.

Exercice 33. $\bar{x} = 185,8$ et $M_e \cong 187,255$.

Exercice 34. La médiane, éventuellement le mode, mais surtout pas la moyenne.

Exercice 35.

- a) Echantillon A : $\bar{x} = 5,2$ et $M_e = 3$
Echantillon B : $\bar{x} = 5,7$ et $M_e = 5,5$.
- b) La moyenne arithmétique.

Exercice 36. Au premier examen, je suis en dessus de la moyenne. Au deuxième, j'ai fait mieux que la moitié des étudiants. Au troisième, j'ai obtenu plus que le nombre de points le plus fréquent de 7.

Exercice 37. $Q_1 = 8$, $Q_2 = 9$ et $Q_3 = 10$.

Exercice 38.

- a) 20.
- b) 3.
- c) $M_e = 3$, $Q_1 = 1$ et $Q_3 = 5$.
- d) $M_o = 0$.

Exercice 39. $Q_3 = 9$.

Exercice 40.

- a) $M_e = 3,375$ m, Maximum = 4,5 m, Minimum = 2,7 m, $Q_1 = 3,075$ m et $Q_3 = 3,65$ m.
- b) (a) Faux.
(b) Vrai.
(c) Vrai.

Exercice 41. $Q_1 = 143$, $Q_3 = 158,8\bar{1}$.

Exercice 42.

a)

Taille (en cm)	Effectifs	Fréquences cumulées
[150; 160[40	0,2
[160; 170[60	0,5
[170; 180[20	0,6
[180; 190[30	0,75
[190; 200[50	1

b) $M_o = 163, \bar{3}$, $M_e = 170$, $Q_1 = 161,75$ et $Q_3 = 190$.**Exercice 43.**a) $Q_1 = 4$, $Q_2 = M_e = 6$ et $Q_3 = 10$.

b) Entre 2 et 13 mails.

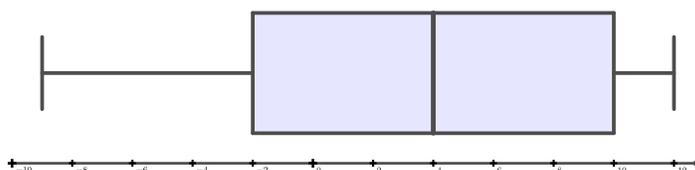
c) Plus de 6 mails.

d) 50%.

Exercice 44.a) Pour X : $Q_1 = 3$, $Q_2 = 6$ et $Q_3 = 12$.Pour Y : $Q_1 = 3$, $Q_2 = 12$ et $Q_3 = 16$.b) La ville Y .

c) Non.

d) Non.

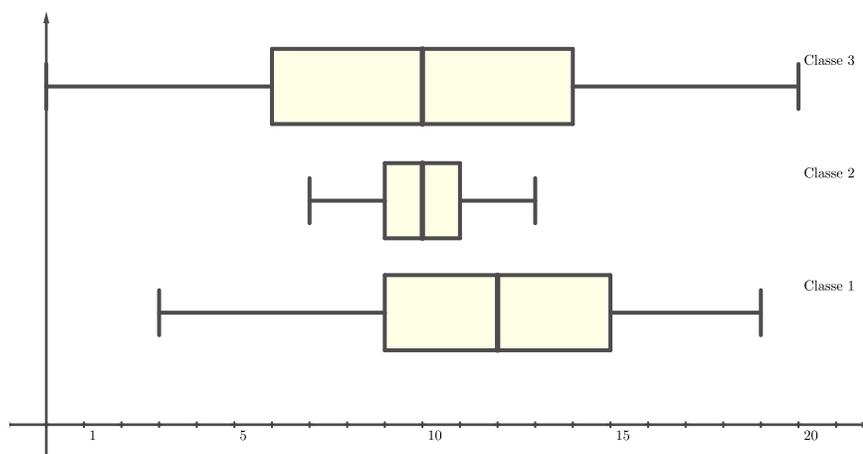
Exercice 45.1 \leftrightarrow U 2 \leftrightarrow Z 3 \leftrightarrow Y 4 \leftrightarrow X **Exercice 46.** $A \leftrightarrow 4$. $B \leftrightarrow 3$. $C \leftrightarrow 1$. $D \leftrightarrow 2$.**Exercice 47.**

Exercice 48.

a)

Points	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Classe 1	0	0	0	1	1	3	1	0	2	2	2	1	7	1	4	3	2	1	2	1	0
Classe 2	0	0	0	0	0	0	0	1	7	4	8	7	6	1	0	0	0	0	0	0	0
Classe 3	1	2	1	1	1	2	1	0	3	4	2	4	3	0	1	2	1	1	1	2	1

b)

c) $\bar{x}_1 \cong 11,618$, $\bar{x}_2 \cong 10,029$ et $\bar{x}_3 = 10$.

d)

Critère A : Classe 2.

Critère B : Classe 3.

Critère C : Classe 1.

Critère D : Classe 2.

Critère E : Aucune.

Exercice 49. $E = 18$.**Exercice 50.** $Q_1 = 28'997$, $Q_2 = 37'916,5$, $Q_3 = 54'585,5$, $I_Q = 25'588,5$ et $Q = 12'794,25$.**Exercice 51.**a) $E = 12$ et $Q = 1$.

b) 14,78%.

Exercice 52. $V = 95,8$ et $\sigma \cong 9,79$.**Exercice 53.**a) $\bar{x} = 11,36$, $M_o = 11$ et $M_e = 11$.b) $V = 3,632$ et $\sigma \cong 1,906$.

Exercice 54.

- a) $V \cong 486,6$ et $\sigma \cong 22,059$.
- b) $V = 2416$ et $\sigma \cong 49,153$.

Exercice 55.

Exercice 56. $b_0 = 2$, $b_{10} = 5$, $Q_1 = 2$, $M_e = 4$, $Q_3 = 5$, $\bar{x} = 3,6$, $\overline{x^2} = 14,4$, $V = 1,44$ et $\sigma = 1,2$.

Exercice 57. $\bar{x} = 46,887$ km/h et $\sigma \cong 7,056$ km/h.

Exercice 58.

- a) $\bar{x} \cong 25'608,696$.
- b) $V \cong 6'151'228,733$ et $\sigma \cong 2'480,167$.

Exercice 59.

- a) $C_A = 3,3\%$.
- b) $C_B = 6\%$.
- c) Dans les deux cas, la dispersion est faible, bien qu'elle le soit davantage pour le titre A , malgré que son écart-type soit inférieur à celui de B .

Exercice 60. $\bar{x} = 65,45$ g, $M_e = 64$ g, $\sigma \cong 7,57$ g et $C \cong 11,57\%$.

Exercice 61. $\sigma \cong 38,99$ et $C \cong 54,58\%$.

Table des matières

1	Introduction	1
2	Définitions	1
3	Traitement des données	4
3.1	Regroupement des données par modalités	4
3.2	Représentation des données à l'intérieur des classes	6
4	Représentations graphiques	7
4.1	Diagramme circulaire	7
4.2	Diagramme en bâtons	9
4.3	Histogramme	11
4.4	Diagrammes trompeurs ou faux	13
4.5	Polygone des effectifs	21
4.6	Polygone des effectifs cumulés	22
5	Valeurs centrales	24
5.1	Moyenne arithmétique	24
5.1.1	Cas discret	24
5.1.2	Cas continu	25
5.2	Mode	25
5.2.1	Cas discret	26
5.2.2	Cas continu	26
5.3	Médiane	30
5.3.1	Cas discret	30
5.3.2	Cas continu	32
5.4	Comparaison entre les valeurs centrales	34
6	Quartiles et boîte à moustaches	35
6.1	Quartiles	35
6.1.1	Cas discret	35
6.1.2	Cas continu	38
6.2	Boîte à moustaches	39
7	Mesures de dispersion	43
7.1	Etendue de la série	44
7.2	Ecart interquartile	44
7.3	Variance écart-type	46
7.3.1	Cas discret	46
7.3.2	Cas continu	49
7.3.3	Autre méthode de calcul	50
7.4	Coefficient de variation	52
7.5	Comparaison des mesures de dispersion	53
8	Solutions	55