

Statistiques

Table des matières

A Proportion et pourcentage	1
A - I Proportions	1
A - II Pourcentage de pourcentage	2
B Coefficients multiplicateurs, variations successives, variation réciproque	3
B - I Coefficient multiplicateur et taux d'évolution	3
B - II Taux d'évolutions successives	4
B - III Taux d'évolution réciproque	5
C Statistiques	6
C - I Vocabulaire	6
C - II Représentations graphiques	7
II.1 Diagramme en bâtons	7
II.2 Diagramme circulaire	8
II.3 Nuage de points	9
II.4 Histogramme	9
II.5 Polygone d'effectifs ou de fréquences cumulés	10
D Paramètres statistiques	11
D - I Paramètres de position	11
I.1 Mode	11
I.2 Moyenne	11
I.3 Médiane	12
D - II Paramètres de dispersion	12
II.1 Étendue	12
II.2 Écart interquartile	14

Information chiffrée et statistique descriptive

A Proportion et pourcentage

A - I Proportions



Définition

Soit A un ensemble non vide ayant n_A éléments et B une partie de A ayant n_B éléments.

La proportion des éléments de B dans A est le nombre réel $\frac{n_B}{n_A}$.

Exemple . Dans une classe de 35 élèves, 12 suivent un cours d'allemand, 23 suivent un cours d'espagnol et 7 suivent un cours de latin.

La proportion d'élèves de la classe suivant un cours d'allemand est $\frac{12}{35}$, la proportion d'élèves de la classe suivant un cours d'espagnol est $\frac{23}{35}$ et la proportion d'élèves suivant un cours de latin est $\frac{7}{35} = \frac{1}{5}$.

Autre exemple : dans la classe de seconde 6, il y a 34 élèves dont 15 garçons et 19 filles.

La proportion de garçons est donc $\frac{15}{34}$ et celle des filles $\frac{19}{34}$.

Remarque Une proportion est un réel compris entre 0 et 1 car, avec les notations de la définition 1, on a $0 \leq n_B \leq n_A$ et donc, en divisant par $n_A > 0$, $0 \leq \frac{n_B}{n_A} \leq 1$.

Remarque : si x est un nombre, la notation $x\%$ signifie $\frac{x}{100}$.

Pour transformer un nombre en %, on l'écrit comme une fraction avec 100 comme dénominateur.

Exemple : $0,23 = \frac{23}{100} = 23\%$.

Exercice Dans une population, on estime que 5 % des individus possèdent un caractère génétique C. Sachant que cette population compte 2 500 individus, combien de personnes possèdent le caractère C ?

Solution

Notons n_C le nombre de personnes ayant le caractère C dans la population, alors, $\frac{n_C}{2500} = 5\% = \frac{5}{100}$, donc $n_C = \frac{5}{100} \times 2500 = 125$. Il y a donc 125 personnes qui possèdent le caractère C dans cette population.

A - II Pourcentage de pourcentage



Propriété

On considère un ensemble non vide A . On suppose que B est une partie non vide de A et que C est une partie de B .

Si la proportion des éléments de B dans A est p_1 et la proportion des éléments de C dans B est p_2 , alors la proportion des éléments de C dans A est $p_1 \times p_2$.

Démonstration :

Notons n_A , n_B et n_C les nombres d'éléments respectifs de A , B et C .

Alors, par définition, $p_1 = \frac{n_B}{n_A}$ et $p_2 = \frac{n_C}{n_B}$.

On a donc $\frac{n_C}{n_A} = \frac{n_C}{n_B} \times \frac{n_B}{n_A} = p_2 \times p_1 = p_1 \times p_2$.

Exemple :

Sur une année, un musée français a obtenu les chiffres suivants concernant sa fréquentation : 26 % des visiteurs du musée ne sont pas français et 32 % des visiteurs étrangers ne sont pas des européens.

Ici, A est l'ensemble des visiteurs du musée dans l'année, B est l'ensemble des visiteurs étrangers et C est l'ensemble des visiteurs qui ne sont pas européens.

On en déduit donc que $\frac{26}{100} \times \frac{32}{100} = \frac{53}{625} = 8,32\%$ des visiteurs du musée ne sont pas européens.

B Coefficients multiplicateurs, variations successives, variation réciproque

B - I Coefficient multiplicateur et taux d'évolution

Rappel : Si une quantité passe d'une valeur initiale à une valeur finale :

- on appelle variation absolue la quantité valeur finale – valeur initiale
- on appelle variation relative ou taux d'évolution la quantité $\frac{\text{valeur finale} - \text{valeur initiale}}{\text{valeur initiale}}$

Exemple : En 2020, le SMIC (salaire minimum interprofessionnel de croissance) était de 1 539,42 €. En 2021, il est de 1 554,58 €.

Le taux d'évolution est donc $\frac{1554,58 - 1539,42}{1539,42} = \frac{15,16}{1539,42} \approx 0,0098 = \frac{0,98}{100} = 0,98\%$.

'Le taux d'évolution (augmentation) du SMIC a été d'environ 0,98 % entre 2020 et 2021.



Propriété

Appliquer un taux d'évolution t à un nombre revient à le multiplier par $C = 1 + t$, appelé coefficient multiplicateur.

Justification : soit x un nombre.

Le montant de l'évolution est xt , donc ce nombre devient $x + xt = x(1 + t)$.

Remarque ; si t est positif, on a affaire à une augmentation ; si t est négatif, il s'agit d'une diminution.

Exemples :

a) Un objet coûte 30 € ; son prix augmente de 3 %. Le coefficient multiplicateur est $C = 1 + \frac{3}{100} = 1,03$. Le prix devient $30 \times 1,03 = 30,90$ €.

b) La population d'une ville diminue de 1 %. Elle est donc multipliée par $C = 1 - \frac{1}{100} = 0,99$.

c) Un prix passe de 12 € à 12,36 €. Quel est le taux d'augmentation ?

Si C est le coefficient multiplicateur, on a $12 \times C = 12,36$, donc $C = \frac{12,36}{12} = 1,03$.

Si t est le taux d'évolution, $C = 1 + t$ donc $t = 0,03 = \frac{3}{100}$.

Remarque : on peut aussi calculer t en calculant $t = \frac{\text{Valeur finale} - \text{Valeur initiale}}{\text{Valeur initiale}} = \frac{12,36 - 12}{12} = \frac{0,36}{12}$

$0,03 = \frac{3}{100} = 3\%$, mais on remarque que $\frac{12,36 - 12}{12}$
 $= \frac{12,36}{12} - \frac{12}{12} = C - 1$ donc, en fait, c'est le **même calcul!**

d) Une année, la population d'une ville est de 22 000 habitants. L'année suivante, elle n'est plus que de 20 900 habitants.

Le coefficient multiplicateur est $C = \frac{20900}{22000} = 0,95$.

Si t est le taux, $C = 1 + t$ d'où $t = C - 1 = -0,05 = -\frac{5}{100}$. Le taux est de -5% .

La population a baissé de 5 %.

e) Un billet de spectacle est vendu 31,65 €. Le taux de T.V.A. est de 5,5 %. Quel est le prix hors-taxes ?

Le coefficient multiplicateur est $C = 1 + t = 1 + 5,5\% = \boxed{1,055}$.

Notons x le prix H.T. (hors-taxes). $x \times C = 31,65$. Le prix H.T est $\frac{31,65}{1,055} = 30 \text{ €}$.

B - II Taux d'évolutions successives

Exemple : En 2005, un objet avait un prix de 120 €. Son prix a augmenté en un an de 3 %. L'année suivante, son prix augmenté de 2 %.

Quel est alors son prix ?

Solution :

Le premier coefficient multiplicateur est $C_1 = 1 + 3\% = 1 + \frac{3}{100} = 1,03$.

En 2006, son prix vaut : $120 \times C_1 = 120 \times \left(1 + \frac{3}{100}\right) = 120 \times 1,03 = 123,6 \text{ €}$.

Le deuxième coefficient directeur est $C_2 = 1 + 2\% = 1 + \frac{2}{100} = 1,02$.

En 2007, son prix vaut : $123,60 \times C_2 = 120 \times 1,03 \times \left(1 + \frac{2}{100}\right) = 120 \times 1,03 \times 1,02 = 126,072 \approx \boxed{126,7 \text{ €}}$. Cela

revient à calculer $120 \times C$ en posant $\boxed{C = C_1 \times C_2}$

On voit que cela revient à multiplier les coefficients multiplicateurs entre eux.

Le nombre $C = C_1 \times C_2$ est appelé **coefficient multiplicateur global**.

B - III Taux d'évolution réciproque

Exemple

Un objet coûte 20 €. Son prix subit une hausse de 2 %.

1. Quel est son nouveau prix?
2. Quel est le montant de la baisse qu'il doit alors subir pour retrouver sa valeur initiale?

Réponses :

1. Le coefficient multiplicateur associé à une hausse de $t=2\%$ est $C = 1 + 2\% = 1,02$.

Le nouveau prix est : $20 \times 1,02 = 20,4$.

2. Soit t' le taux de baisse ; le coefficient multiplicateur est alors : $C' = 1 + t'$.

On doit donc avoir : $20 \times C \times C' = 20$ d'où, après simplification par 20 :

$$CC' = 1 \text{ et donc : } C' = \frac{1}{C}.$$

$$\text{Par conséquent : } 1 + t' = \frac{1}{1 + t}.$$

$$\text{On en déduit : } t' = \frac{1}{1 + t} - 1 = \frac{1}{1,02} - 1 \approx -0,01960.$$

$$t' \approx -\frac{1,96}{100} = -1,96\%.$$

On dit que le **taux d'évolution réciproque** de 2 % est de $-1,96\%$.



Définition et propriété

Soit t le taux d'évolution subi par un nombre. On appelle taux d'évolution réciproque le taux t' qu'il faut alors appliquer pour retrouver le nombre de départ.

$$\text{On a : } t' = \frac{1}{1 + t} - 1$$

Démonstration :

Soit x un nombre, qui subit un taux d'évolution égal à t .

Le coefficient multiplicateur est $1 + t$, donc la nouvelle valeur est $x(1 + t)$.

On cherche alors le montant du taux d'évolution t' qui permet de retrouver la valeur x initiale.

Le coefficient multiplicateur associé est $1 + t'$.

On doit donc avoir : $[x(1 + t)] \times (1 + t') = x$.

En simplifiant par x , on obtient : $(1 + t)(1 + t') = 1$.

$$\text{On en déduit : } 1 + t' = \frac{1}{1 + t} \text{ d'où : } t' = \frac{1}{1 + t} - 1.$$

t' est le **taux d'évolution réciproque** du taux t

Exemples :

1. Pour un taux $t = 3\%$, on obtient $t' = \frac{1}{1 + 0,03} - 1 \approx -0,029 \approx -2,9\%$.

Le taux d'évolution réciproque de 2 % est de $-2,9\%$.

2. Pour un taux $t = -10\%$, on obtient $t' = \frac{1}{1-0,1} - 1 \approx 0,111 \approx 11,1\%$.

Le taux dévolution réciproque d'une baisse de 10% est d'environ $11,1\%$.

C Statistiques

C - I Vocabulaire

Une étude statistique commence par un recueil de données. On utilise le vocabulaire suivant pour décrire cette étude :

- **Série statistique** : Ensemble des valeurs collectées.
- **Population** : Ensemble sur lequel porte l'étude statistique.
- **Individus** : Éléments qui composent la population.
- **échantillon** : Partie de la population.
- **Caractère étudié** : Propriété que l'on observe sur les individus. Les différentes valeurs obtenues sont appelées **valeurs du caractère** ou **modalités**, souvent notées x_1, x_2, \dots, x_p . On distingue deux types de caractères.
 - ◊ Un caractère peut être **qualitatif** (situation de famille, sexe, couleur des yeux, type d'habitation...).
 - ◊ Un caractère peut être **quantitatif**. Il est dit **discret** lorsqu'il ne prend que des valeurs isolées (nombre d'enfants, notes dans une classe...).

Il est dit **continu** lorsqu'il peut prendre théoriquement toutes les valeurs d'un intervalle (taille, temps d'écoute...); dans ce cas, les valeurs sont regroupées en intervalles appelés des **classes**.

- **Effectif** : Pour une valeur du caractère (modalité ou classe), on appelle effectif le nombre d'individus de la population ayant cette valeur. On note souvent n_1, n_2, \dots, n_p les effectifs respectifs des modalités x_1, x_2, \dots, x_p .
- **Effectif total** : Nombre total d'individus de la population (ou de l'échantillon). Il est égal à $n_1 + n_2 + \dots + n_p$, souvent noté N .
- **Fréquence** : Pour une valeur du caractère (modalité ou classe), on appelle fréquence le quotient de l'effectif de cette valeur par l'effectif total. On note souvent f_1, f_2, \dots, f_p les fréquences respectives des modalités x_1, x_2, \dots, x_p , donc :

$$f_1 = \frac{n_1}{N}, f_2 = \frac{n_2}{N}, \dots, f_p = \frac{n_p}{N}.$$

On en déduit que $0 \leq f_1 \leq 1, 0 \leq f_2 \leq 1, \dots, 0 \leq f_p \leq 1$, et $f_1 + f_2 + \dots + f_p = 1$.

- **Valeurs extrêmes** : Valeurs minimales et maximales d'un caractère quantitatif.
- **Effectif cumulé** : Pour une valeur x d'une série statistique quantitative, l'effectif cumulé croissant (respectivement décroissant) de x est la somme des effectifs des valeurs inférieures (respectivement supérieures) ou égales à x .
- **Fréquence cumulée** : Pour une valeur x d'une série statistique quantitative, la fréquence cumulée croissante (respectivement décroissante) de x est la somme des fréquences des valeurs inférieures (respectivement supérieures) ou égales à x .

Exemple avec des notes :

Dans le tableau suivant sont regroupées les notes obtenues par les élèves d'une seconde lors du contrôle n° 1 (éventuellement arrondies pour simplifier l'étude) :

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

Dans cet exemple :

- La série statistique est l'ensemble des notes collectées.
- La population est l'ensemble des élèves de seconde .
- Les individus sont chacun des élèves de seconde Turner.
- Le caractère étudié est le résultat obtenu au contrôle n° 1.
- Les modalités sont les valeurs chiffrées des notes obtenues au contrôle n° 1.
- L'effectif total est le nombre d'élèves de la classe, à savoir 34.
- Les valeurs extrêmes sont 4 et 19.

Exemple avec des notes : Pour une meilleure lisibilité et pour simplifier l'étude, on peut commencer par compter le nombre d'individus ayant obtenu chaque note :

Note	4	5	6	8	9	10	11	12	13	14	16	17	19
Effectif	1	1	3	2	1	1	3	4	2	2	7	4	3
Fréquence à 10^{-2} près	0,03	0,03	0,09	0,06	0,03	0,03	0,09	0,12	0,06	0,06	0,21	0,12	0,09

Remarque

Dans le tableau précédent, la somme des fréquences est supérieure à 1 à cause des arrondis.

Exemple : série continue On a interrogé en 2008 un échantillon de 4812 Français concernant la durée hebdomadaire d'écoute de la télévision (en heures) :

Durée	[0 ; 10[[10 ; 15[[15 ; 20[20 ; 30[30 ; 50]
Effectif	972	924	826	1069	1021

Le caractère étudié, à savoir la durée d'écoute, est quantitatif continu : il peut prendre théoriquement toutes les valeurs de l'intervalle [0 ; 50]. Les données sont regroupées en classes [0 ; 10], [10 ; 15[, [15 ; 20[, [20 ; 30[et [30 ; 50].

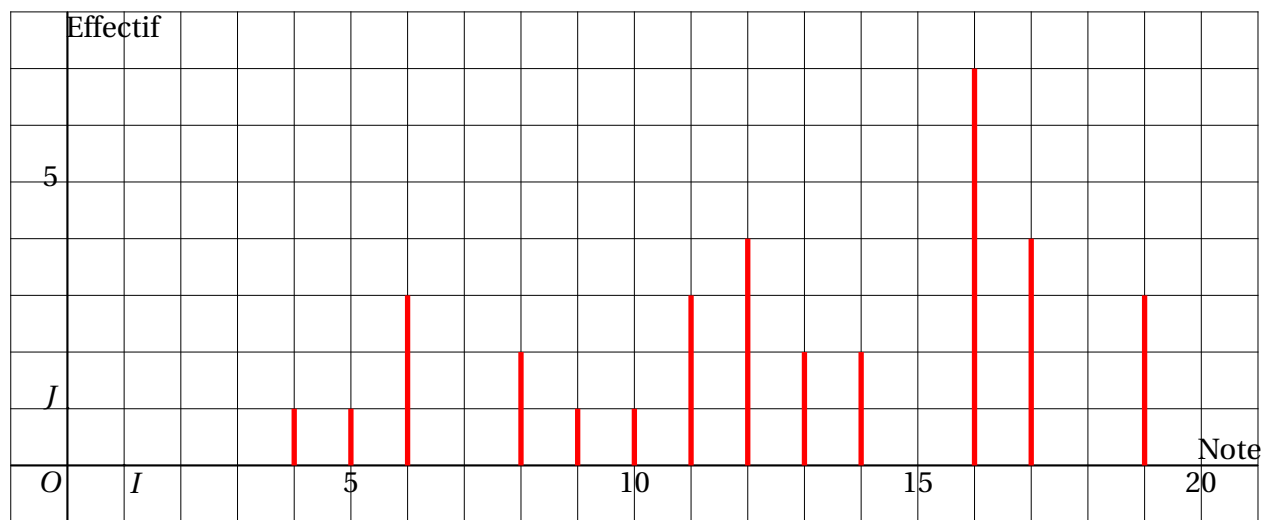
C - II Représentations graphiques

Séries à caractère quantitatif discret

II.1 Diagramme en bâtons

Dans un **diagramme en bâtons**, on représente une série statistique discrète par des segments dont la hauteur est proportionnelle à l'effectif de la valeur qu'ils représentent.

Exemple On continue à travailler avec les données de l'exemple sur les notes. Voici le diagramme en bâtons de cette série :



II.2 Diagramme circulaire

Exemple :

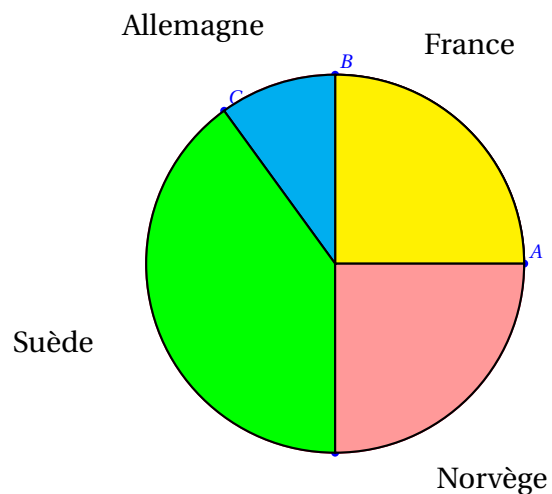
Dans une compétition d'athlétisme, quatre pays s'affrontent : la France, l'Allemagne, la Suède et la Norvège. On note le pourcentage de médailles obtenues par chacun des pays :

Pays	France	Allemagne	Suède	Norvège
Pourcentage de médailles	25 %	10 %	40 %	25 %

Représenter le diagramme circulaire associé à cette série statistique :

Pays	Total	France	Allemagne	Suède	Norvège
Pourcentage de médailles	100 %	25 %	10 %	40 %	25 %
Angle en °	360	90	36	144	90

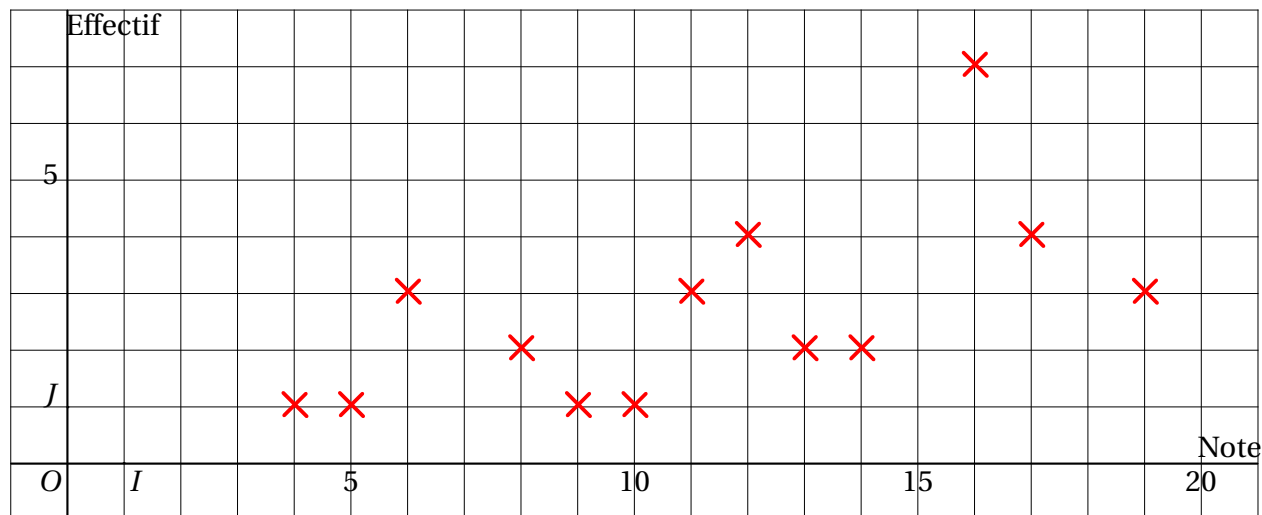
Pour cela, nous avons besoin des angles ; nous les calculons par proportionnalité, sachant que 100 % correspondent à 360°.



II.3 Nuage de points

Dans un **nuage de points**, on représente une série statistique discrète par des points dont les abscisses sont les valeurs du caractère, et les ordonnées sont les effectifs correspondants, parfois reliés par des segments.

Exemple On travaille toujours avec les données de l'exemple sur les notes. Voici le nuage de points de cette série :



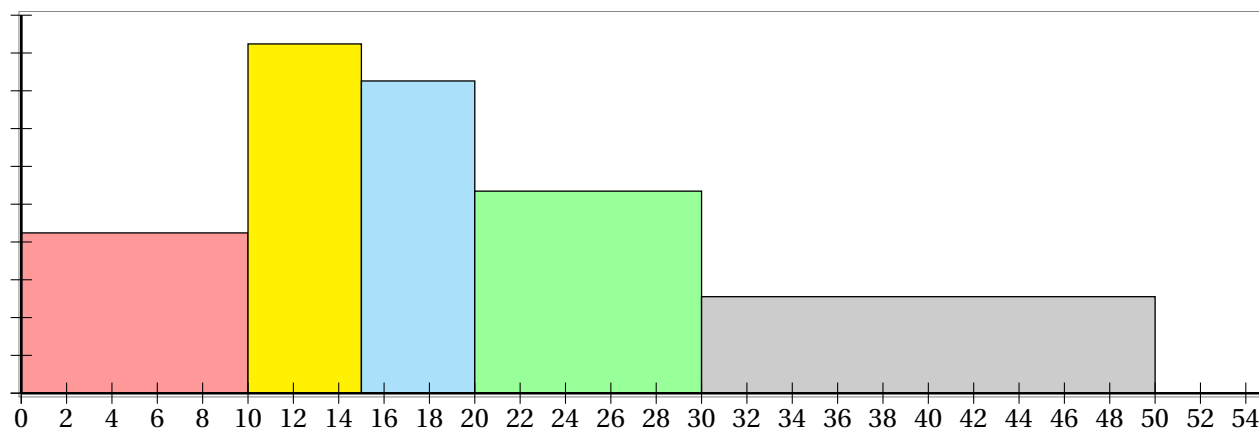
Séries à caractère quantitatif continu

II.4 Histogramme

Dans un **histogramme**, on représente une série statistique continue par des rectangles dont la largeur correspond à l'amplitude de chaque classe et dont l'aire est proportionnelle à l'effectif de la classe.

Exemple

On travaille avec les données de l'exemple sur la durée d'écoute de la télévision. Voici l'histogramme de cette série :



Remarque

Lorsque les classes ont toutes la même amplitude, la hauteur de chaque rectangle est proportionnelle à l'effectif de la classe qu'il représente. On dit alors que l'histogramme est à **pas constant**.

II.5 Polygone d'effectifs ou de fréquences cumulés

- Le **polygone des effectifs cumulés croissants** (respectivement **décroissants**) d'une série statistique continue est la ligne brisée qui joint les points du plan dont les abscisses sont les bornes de chaque classe et dont les ordonnées sont les effectifs cumulés croissants (respectivement décroissants) de ces valeurs.
- Le **polygone des fréquences cumulées croissantes** (respectivement **décroissantes**) d'une série statistique continue est la ligne brisée qui joint les points du plan dont les abscisses sont les bornes de chaque classe et dont les ordonnées sont les fréquences cumulées croissantes (respectivement décroissantes) de ces valeurs.

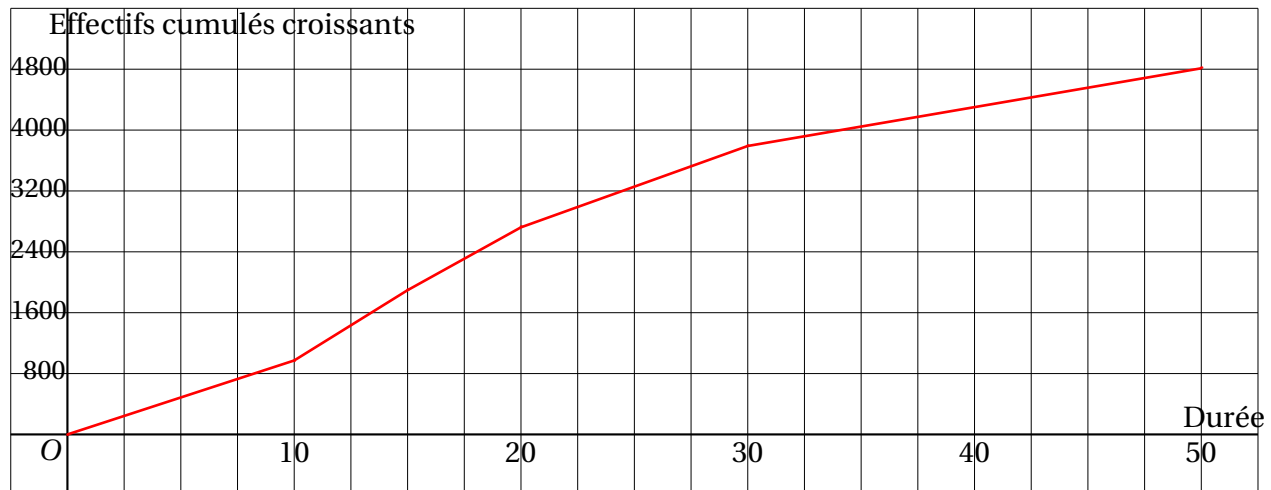
Ces représentations donnent l'allure de la répartition des valeurs de la série.

Exemple

La situation est toujours celle de l'exemple C - I. Le tableau des effectifs cumulés croissants est le suivant :

Durée <	0	10	15	20	30	50
Effectif	0	972	1896	2722	3791	4812

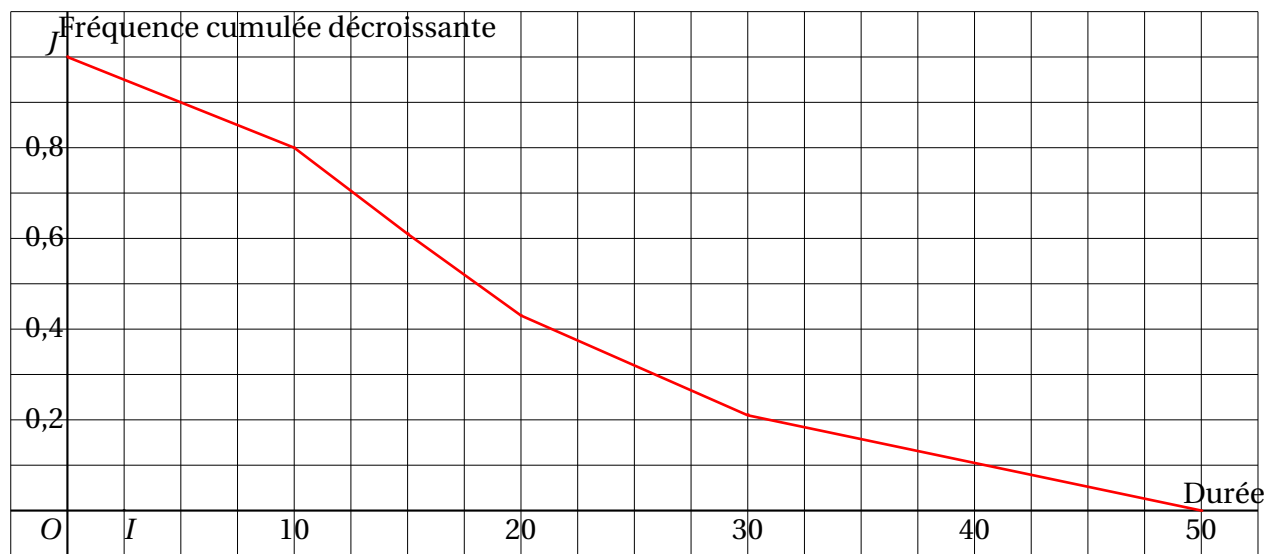
D'où le polygone des effectifs cumulés croissants :



Traisons à présent le cas des fréquences cumulées décroissantes :

Durée \geq	0	10	15	20	30	50
Effectif	4812	3840	2916	2090	1021	0
Fréquence	1	0,8	0,61	0,43	0,21	0

D'où le polygone des fréquences cumulées décroissantes :



D Paramètres statistiques

D - I Paramètres de position

I.1 Mode

Définition :

Un **mode** d'une série statistique est une valeur de la série dont l'effectif est strictement supérieur à celui des autres valeurs.

Remarque

Dans une série statistique, il peut y avoir plusieurs modes.

Exemple

Dans l'exemple sur les notes, le mode est 16.

I.2 Moyenne

On considère une série statistique donnée par le tableau suivant :

Valeur	x_1	x_2	x_3	...	x_{p-1}	x_p
Effectif	n_1	n_2	n_3	...	n_{p-1}	n_p
Fréquence	f_1	f_2	f_3	...	f_{p-1}	f_p



Définition

La **moyenne** de cette série statistique est le réel noté \bar{x} défini par

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{n_1 + n_2 + \dots + n_p} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$$

en notant $N = n_1 + n_2 + \dots + n_p$ l'effectif total de la série.

Propriété

On peut également calculer la moyenne à l'aide des fréquences :

$$\bar{x} = x_1 f_1 + x_2 f_2 + \dots + x_p f_p.$$

Exemple

Avec les données de l'exemple C - I, la moyenne de la classe est :

$$\bar{x} = \frac{1 \times 4 + 1 \times 5 + 3 \times 6 + 2 \times 8 + 1 \times 9 + 1 \times 10 + 3 \times 11 + 4 \times 12 + 2 \times 13 + 2 \times 14 + 7 \times 16 + 4 \times 17 + 3 \times 19}{34} = \frac{434}{34}$$

Une valeur approchée de \bar{x} à 10^{-2} près est 12,76.

I.3 Médiane

Définition

La **médiane** M d'une série statistique est un réel qui partage cette série en deux parties telles que :

- Au moins 50 % des valeurs sont inférieures ou égales à la médiane ;
- Au moins 50 % des valeurs sont supérieures ou égales à la médiane.

Propriété

En pratique, on adopte la démarche suivante pour déterminer la médiane M d'une série statistiques d'effectif total N :

- On range d'abord les N valeurs du caractère par ordre croissant.
- Si N est pair, M est la moyenne des deux valeurs « centrales » de la série.
- Si N est impair, M est la valeur centrale de la série.

Exemple

Dans l'exemple C - I, l'effectif total est 34, c'est-à-dire pair. La médiane est donc la moyenne des deux valeurs centrales de la série, à savoir les 17^e et 18^e valeurs. Donc $M = \frac{13 + 13}{2} = 13$, ce qui signifie qu'au moins la moitié des notes est inférieure ou égale à 12 (en réalité 18 notes), et qu'au moins la moitié des notes est supérieure ou égale à 12 (en réalité 18 notes également).

Exemple

En France, en 2005, dans le secteur privé et semi-public (SNCF, Poste, Caisse d'épargne...), le salaire net mensuel médian est de 1528 €, alors que le salaire net mensuel moyen est de 1903,5 €.

Source : .

D - II Paramètres de dispersion

II.1 Étendue



Définition

| L'**étendue** d'une série statistique est la différence entre sa plus grande et sa plus petite valeur.

Exemple

Dans l'exemple sur les notes, l'étendue est égale à $19 - 4 = 15$.

Quartiles



Définition

On considère une série statistique.

- Le premier **quartile** Q_1 est la plus petite valeur de la série telle qu'au moins 25 % des données soient inférieures ou égales à Q_1 .
- Le troisième **quartile** Q_3 est la plus petite valeur de la série telle qu'au moins 75 % des données soient inférieures ou égales à Q_3 .

Exemple

On considère toujours les données de l'exemple C - I.

- $34 \times \frac{25}{100} = 8,5$ donc le premier quartile Q_1 de la série est la 9^e valeur, d'où $Q_1 = 10$, ce qui signifie qu'au moins un quart des notes sont inférieures ou égales à 10 (en réalité 9 notes, soit environ 26 %).
- $34 \times \frac{75}{100} = 25,5$ donc le troisième quartile Q_3 de la série est la 26^e valeur, d'où $Q_3 = 16$, ce qui signifie qu'au moins trois quarts des notes sont inférieures ou égales à 16 (en réalité 27 notes, soit environ 79 %).

Remarque

Le fait que le partage théorique en 25 %, 50 % et 75 % de la série statistique à l'aide des indicateurs Q_1 , M et Q_3 ne soit pas tout à fait exact provient du fait que la série comporte des valeurs identiques. Ce phénomène a tendance à s'amoin-drir lors d'une étude sur une population plus importante avec un caractère dont les modalités sont plus disparates.

II.2 Écart interquartile



Définition

On considère une série statistique de premier quartile Q_1 et de troisième quartile Q_3 .

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

Exemple

Dans l'exemple C - I, l'intervalle interquartile est $[10 ; 16]$ et l'écart interquartile est $Q_3 - Q_1 = 16 - 10 = 6$.