المدرسة العليا للأساتذة بفاس
ECOLE NORMALE SUPÉRIEURE DE FES

جامعة سيدي محمد بن عبد الله بفاس
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES

Module: Mathématique 1

Chapitre 5: Suites

Prof. Mohammed SRATI

Licence d'Éducation

Spécialité: Enseignement Primaire

02-01-2023

Sommaire

3. Suites

- Suites numériques
- Suites arithmétiques
- Suites géométriques

La statistique est un domaine des mathématiques et de plus en plus, elle fait partie de ce que l'on appelle aujourd'hui la science des données. Plus précisément, C'est une méthode scientifique qui consiste à réunir des données chiffrées sur des ensembles nombreux, puis à analyser, à commenter et à critiquer ces données. Il ne faut pas confondre **la** statistique qui est la science qui vient d'être définie et **une** statistique qui est un ensemble de données chiffrées sur un sujet précis.
Les premières statistiques correctement élaborées ont été celles des recensements démographiques. Ainsi le vocabulaire statistique est essentiellement celui de la démographie. La statistique s'applique à la plupart des disciplines : biologie, agronomie, démographie, économie, sociologie, métrologie, médecine, psychologie, ...

Le calcul statistique a double objectif :

- 1^{er} but : À partir de données brutes en grand nombre, on peut dégager avec un minimum d'effort, un certain nombre de renseignements qualitatifs ou quantitatifs permettant de visualiser cette statistique avec une bonne précision. Cela afin de pouvoir la comparer à d'autres statistiques du même type (*Statistique Descriptive*).
- $2^{\text{ème}}$ but : La statistique ayant été effectuée sur un échantillon, nous permet d'extrapoler les résultats partiels en vue de déduire des précisions globales (*Statistique Inférentielle*).

• ***Population*** : est l'ensemble que l'on observe et qui sera soumis à une analyse statistique. Par exemple : étudiants de l'ENS de Fès, salariés d'une entreprise, ...

• ***Individu*** ou ***unité statistique*** : chaque élément de la population sur lequel porte l'observation. Par exemple : étudiant ; salarié, ...

• ***Échantillon*** : est un sous-ensemble d'individus prélevés de la population. Par exemple : étudiants de moins de 20 ans, jeunes salariés, ...

• ***Caractère*** ou une ***variable statistique*** : désigne une grandeur observable sur un individu et susceptible de varier prenant ainsi différents états appelés ***modalités*** tel que : l'âge des étudiants, le nombre d'heures du travail des salariés, ...

• L'ensemble des valeurs possibles ou des *modalités* est appelé le domaine de la variable.

Plus généralement, on appelle caractère toute application X de la population Ω dans un ensemble E, dont les éléments x sont appelés modalités du caractère X (ou valeurs du caractère).

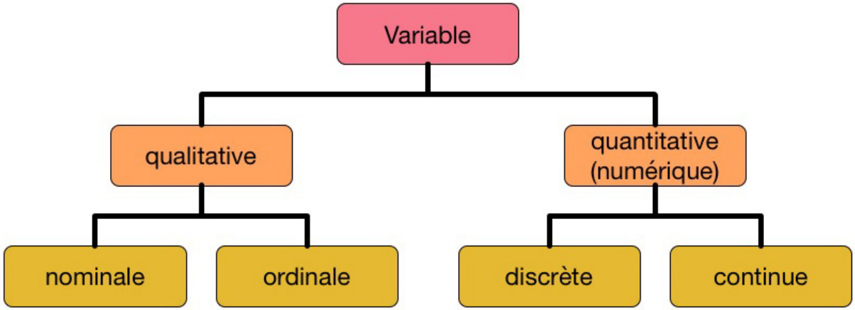
$$X \; : \quad \begin{array}{l} \Omega \longrightarrow E \\ w \longmapsto X(w) := x, \end{array} \qquad \begin{cases} X & : \text{ le caractère,} \\ \Omega & : \text{ la population,} \\ w & : \text{ individu,} \\ E & : \text{ l'ensemble des modalités,} \\ X(w) := x & : \text{ modalité.} \end{cases}$$

Variables quantitatives : se sont des variables numériques et mesurables exprimant une quantité. Les variables quantitatives peuvent être classées en :

- ***Variables quantitatives discrètes*** (ou discontinues) : représentées par un nombre fini de valeurs (de modalités) isolées.
  - Exemple : nombre d'enfant par ménage, nombre d'heures de travail par jour des salariés d'une entreprise,... ($X(\Omega) = \{0, 1, 2, 3, 4\}$).
- ***Variables quantitatives continues*** : un caractère continu peut prendre un nombre infini de valeurs dans son intervalle de définition. Ses valeurs peuvent être regroupées en classes (en pratique un intervalle de $\mathbb{R}$).
  - Exemple : revenu mensuel par ménage, poids ou taille des étudiants de la filière LPEP-S1, ... ($X(\Omega) = [0, 5[\cup[5, 10[\cup[10, 15[\cup[15, 20[$).

Variables qualitatives : se présentent en plusieurs modalités (différentes *catégories* prises par un caractère qualitatif) et ils ne peuvent pas faire l'objet d'une mesure car ils *ne se présentent pas sous forme numérique*. ile peuvent être classées en :

- ***Variables qualitatives nominales*** : dont les modalités ne peuvent pas être ordonnées ou hiérarchisées.
  - Exemple : Pour le caractère "sexe des étudiants", les modalités féminin et masculin ($X(\Omega) = \{$féminin, masculin$\}$) ne peuvent pas être ordonnées.
- ***Variables qualitatives ordinales*** : dont les modalités peuvent être ordonnées ou hiérarchisées.
  - Exemple : pour le caractère "mention du baccalauréat", les modalités sont ordonnées par ordre croissante comme suit :
  $X(\Omega) = \{$Passable, Assez-bien, Bien, Très bien, Excellent$\}$.


Variable
qualitative
quantitative
(numérique)
nominale
ordinale
discrète
continue

Exemple

- *Étude du nombre d'enfants des couples d'un quartier donné.*

Population	:	*Ensembles des couples du quartier,*
Individu	:	*Couple,*
Caractère	:	*Nombre d'enfants,*
Type	:	*Quantitatif discret.*

- *Étude des marques de voitures d'une ville donnée.*

Population	:	*Ensembles des voitures de la ville,*
Individu	:	*Voiture,*
Caractère	:	*Marque,*
Type	:	*Qualitatif nominal.*

Exercice

Spécifier pour chacune des variables suivantes si elle est qualitative (ordinale ou nominale), quantitative (continue ou discrète). Proposer, pour les variables qualitatives, des modalités adéquates.
Niveau de scolarité, Couleur des yeux, Nombre de voiture par famille, Revenu, État mécanique d'une voiture, Profession, Nombre d'heures de travail des salariés, Taille, Grade militaire, Nationalité, Note obtenues par les étudiants, Âge, Masse, Surface.

Remarque

Toute statistique qualitative peut se transformer en une statistique quantitative à l'aide d'un codage des valeurs possibles du caractère. Par exemple :

- *1 : masculin et 2 : féminin est le codage usuel du sexe.*
- *Le code postal est le codage de lieux géographiques.*

On appelle *série statistique* la suite des valeurs prises par une variable X sur les unités d'observation.

- Le nombre d'unités d'observation est noté n.
- Les valeurs de la variable X sont notées : $x_1, \ldots, x_i, \ldots, x_n$.

Exemple

On s'intéresse à la variable "état-civil" notée X *et à la série statistique des valeurs prises par* X *sur 20 personnes. La codification est*

$$\begin{array}{ll} \mathrm{C}: & \textit{célibataire,} \\ \mathrm{M}: & \textit{marié(e),} \\ \mathrm{V}: & \textit{veuf(ve),} \\ \mathrm{D}: & \textit{divorcée.} \end{array}$$

Le domaine de la variable X *est* $\{C, M, V, D\}$*. Considérons la série statistique suivante :*

$$\begin{array}{cccccccccc} M & M & D & C & C & M & C & C & C & M \\ C & M & V & M & V & D & C & C & C & M \end{array}$$

Ici, $n = 20$ *et* $x_1 = M, x_2 = M, x_3 = D, x_4 = C, x_5 = C, \ldots, x_{20} = M.$

L'un des objectifs de la statistique descriptive est de résumer les données brutes recueillies sur une population dans des tableaux statistiques afin d'avoir une présentation des données d'une façon lisible.

Exemple

Enquête auprès d'un échantillon de 60 familles de la région Fès-Meknès sur le nombre d'enfant par ménage. Les résultats brutes des nombres d'enfants sont :

$$2, 1, 4, 2, 2, 0, 1, 2, 3, 0, 4, 5, 2, 5, 4, 2, 6, 2, 6, 4, 2, 1, 3, 2, 1, 3, 3, 3, 1, 1,$$

$$1, 3, 2, 3, 3, 2, 4, 2, 5, 2, 3, 3, 1, 5, 1, 5, 2, 6, 2, 5, 2, 3, 1, 2, 2, 0, 1, 4, 3, 1.$$

Remarquons que les données brutes ne sont pas lisibles d'où la nécessité de regrouper ces données dans un tableau pour faciliter leur traitement.

La présentation d'un tableau statistique doit respecter des principes généraux :

- Le tableau doit porter des *intitulés* de lignes et de colonnes clairement définis ainsi que préciser les unités utilisés.
- Le tableau doit porter un *titre* précisant son contenu et la source des informations lorsque les données sont empruntées à une publication ou à un organisme.
- Dans un tableau statistique :
  - La première colonne présent les différentes *modalités* (x_i) prises par le caractère étudié.
  - La deuxième colonne présente les *effectifs* (n_i) nombre d'individus correspondant à chaque modalité (x_i) du caractère.

Considérons une population statistique de n individus décrite selon le caractère x dont les k modalités sont $x_1, x_2, \ldots, x_i, \ldots, x_k$.

Modalité (x_i)	**Effectif** (n_i)
x_1	n_1
x_2	n_2
$\vdots$	$\vdots$
x_i	n_i
$\vdots$	$\vdots$
x_p	n_p
Total	n

• n_i : représente le nombre d' individus, appelé ***effectif partiel*** présentant la modalité x_i

• n : la somme des effectifs partiels n_i est appelé ***effectif total*** de la population.

$$n = \sum_{i=1}^{p} n_i.$$

Maintenant on regroupe les données de l'exemple 1.3 dans un tableau pour faciliter leur traitement.

Nombre d'enfants par ménage (x_i)	Effectif des ménages (x_i)
0	3
1	12
2	18
3	12
4	6
5	6
6	3
Total	60

Titre : Nombre d'enfants observé dans un échantillon des ménages de la région de Fès-Meknès.

Définition (Fréquence)

La fréquence relative ou fréquence notée f_i est la proportion d'individus présentant la même modalité dans la population. Elle est obtenue en divisant chaque effectif n_i par l'effectif total n, c'est-à-dire :

$$f_i = \frac{n_i}{n}.$$

Remarques :

- Il est recommandé d'exprimer la fréquence f_i en pourcentage $\%$.
- La somme des fréquences f_i est égale à 1 et la somme des fréquences exprimées en pourcentage est égale à 100%.

Le tableau statistique de la distribution du caractère étudié se présentera donc sous la forme suivante :

Modalité (x_i)	**Effectif** (n_i)	**Fréquence** (f_i)
x_1	n_1	f_1
x_2	n_2	f_2
$\vdots$	$\vdots$	$\vdots$
x_i	n_i	f_i
$\vdots$	$\vdots$	$\vdots$
x_p	n_p	f_p
Total	n	1

Il en est de même si on considère les fréquences en pourcentage au lieu des fréquences.

Exemple

Le tableau statistique de la distribution de nombre d'enfants par ménage dans la région de Fès-Meknès est présenté comme suit :

Nombre d'enfants par ménage	Effectif des ménages	Fréquence des ménages en %
0	3	5
1	12	20
2	18	30
3	12	20
4	6	10
5	6	10
6	3	5
Total	60	100

Dans l’exemple précédent, s’il est demandé de répondre à certains questions de type :

1. Combien de familles ont **moins de** quatre enfants ?
2. Combien de familles ont **au moins** quatre enfants ?
3. Quelle est la proportion de familles ayant **au plus** quatre enfants ?
4. Quelle est la proportion de familles ayant **plus** de quatre enfants ?

Le calcul des effectifs cumulés, notés N_i, et des fréquences cumulées, notées F_i, nous permet de donner ces valeurs. Ce calcul se fait en **cumulant** (**sommant**) les effectifs et les fréquences relatives dans une colonne du tableau. En effet :

- Pour calculer un *effectif cumulé croissant* d'une valeur d'un caractère, il suffit d'ajouter à l'effectif de cette valeur le ou les effectifs des valeurs précédentes.
- Pour calculer une *fréquence cumulée croissante* d'une valeur d'un caractère, il suffit d'ajouter à la fréquence de cette valeur la ou les fréquences des valeurs précédentes.
- Pour calculer un *effectif cumulé décroissant* d'une valeur d'un caractère, il suffit d'ajouter à l'effectif de cette valeur le ou les effectifs des valeurs suivantes.
- Pour calculer une *fréquence cumulée décroissante* d'une valeur d'un caractère, il suffit d'ajouter à la fréquence de cette valeur la ou les fréquences des valeurs suivantes.

Sommer de haut en bas

$\Downarrow$

Calculer les effectifs cumulés croissants et les fréquences cumulées croissantes

$\Downarrow$

Répondre aux questions «moins de» et «au plus»

Sommer de bas en haut

$\Downarrow$

Calculer les effectifs cumulés décroissants et les fréquences cumulées décroissantes

$\Downarrow$

Répondre aux questions «plus de» et «au moins»

Exemple

Pour répondre aux questions posées au début de ce paragraphe, il suffit de calculer les effectifs cumulés et les fréquences cumulées du caractère représentant le nombre d'enfants par ménage dans la région Fès-Meknès.

Nombre d'enfants par ménage	Effectifs des Ménages	Fréquences des Ménage en %	Effectifs cumulés croissants	Effectifs cumulés décroissants	Fréquences cumulées croissantes en %	Fréquences cumulées décroissante s en %
0	3	5	3	60	5	100
1	12	20	15	57	25	95
2	18	30	33	45	55	75
3	12	20	45	27	75	45
4	6	10	51	15	85	25
5	6	10	57	9	95	15
6	3	5	60	3	100	5
Total	60	100				

D'après le tableau :

- *45 ménages ont moins de 4 enfants.*
- *15 ménages ont au moins 4 enfants.*
- 85% *des ménages ont au plus de 4 enfants.*
- 15% *des ménages ont plus de 4 enfants.*

Une variable quantitative continue peut prendre une infinité de valeurs possibles. Le domaine de la variable est alors $\mathbb{R}$ ou un intervalle de $\mathbb{R}$. En pratique, une mesure est limitée en précision. La taille peut être mesurée en centimètres, voire en millimètres. On peut alors traiter les variables continues comme des variables discrètes. Cependant, pour construire le tableau statistique et faire des représentations graphiques, il faut procéder à des regroupements en classes. Le tableau regroupé en classe est souvent appelé distribution groupée. Si $\left[c_j^- ; c_j^+\right[$ désigne la classe j, on note, de manière générale :

1. c_j^- : la *borne inférieure* de la classe j,
2. c_j^+ : la *borne supérieure* de la classe j,
3. $c_j = \dfrac{\left(c_j^+ + c_j^-\right)}{2}$: le *centre* de la classe j,
4. $a_j = c_j^+ - c_j^-$: l'*amplitude* de la classe j,
5. n_j : l'*effectif* de la classe j
6. N_j : l'*effectif cumulé* de la classe j,
7. f_j : la *fréquence* de la classe j,
8. F_j : la *fréquence cumulée* de la classe j.

La répartition en classes des données nécessite de définir a priori le *nombre de classes* J et donc l'amplitude de chaque classe. En règle générale, on choisit au moins cinq classes de même amplitude. Cependant, il existent des formules qui nous permettent d'établir le nombre de classes et la longueur de classe (l'amplitude) pour une série statistique de n observations.

- ***La règle de Sturge*** : $J = 1 + \big(3.3 \times \log_{10}(n)\big)$.
- ***La règle de Yule*** : $J = 2.5 \times \sqrt[4]{n}$.

L'intervalle de classe est obtenue ensuite de la manière suivante :

$$\text{Longueur de l'intervalle (l'amplitude)} = \frac{(x_{\max} - x_{\min})}{J},$$

où $x_{\max}$ (resp. $x_{\min}$) désigne la plus grande (resp. la plus petite) valeur observée.

Remarque

1. *Il faut arrondir le nombre de classe J à l'entier le plus proche. Par commodité, on peut aussi arrondir la valeur obtenue de l'intervalle de classe.*
2. *A partir de la plus petite valeur observée, on obtient les bornes de classes en additionnant successivement la longueur de classe (l'amplitude).*

Exemple

On mesure la taille en centimètres de 50 élèves d'une classe :

152	152	152	153	153	154	154	154	155	155
156	156	156	156	156	157	157	157	158	158
159	159	160	160	160	161	160	160	161	162
162	162	163	164	164	164	164	165	166	167
168	168	168	169	169	170	171	171	171	171

On a les classes de tailles définies comme il suit :

$$[151,5;155,5[,\qquad [155,5;159,5[,\qquad [159,5;163,5[,\qquad [163,5;167,5[,\qquad [167,5;171,5[.$$

Alors, on construit le tableau statistique la série statistique des tailles des élèves.

$\left[c_j^-, c_j^+\right]$	Effectif n_j	Effectif cumulé N_j	Fréquence f_j	Fréquence cumulée F_j
$[151,5;155,5[$	10	10	0.20	0.20
$[155,5;159,5[$	12	22	0.24	0.44
$[159,5;163,5[$	11	33	0.22	0.66
$[163,5;167,5[$	7	40	0.14	0.80
$[167,5;171,5[$	10	50	0.20	1.00
Total	50		1.00	

Les graphiques permettent de donner une synthèse visuelle de la distribution d'une variable et mettre en évidence certaines informations données par le tableau.
Les représentations graphiques sont spécifiques à chaque type de variables ou de caractères (qualitatif, quantitatif discret ou quantitatif continu). Les variables qualitatives peuvent être représentées graphiquement de différentes manières. Les diagrammes les plus utilisés sont le **diagramme à bandes** (ou diagramme en tuyaux d'orgues) et le **diagramme à secteurs circulaires**.
Un diagramme en bâtons est constitué d'une suite de bâtons (verticaux ou horizontaux). À chaque modalité du caractère, on associe un «bâton» de longueur proportionnelle à l'effectif ou à la fréquence de cette modalité.

Exemple A

Répartition des salariés de l'entreprise X selon le contrat de travail (contrat de sécurisation professionnelle : CSP).

CSP : x_i	Effectif n_i	Fréquence f_i	Fréquence f_i en %
Cadres	10	0,125	12.5
Contremaîtres	5	0,0625	6.25
Employés	20	0,25	25
Ouvriers	40	0,50	50
Autres	5	0,0625	6.25
Total	80	1	100

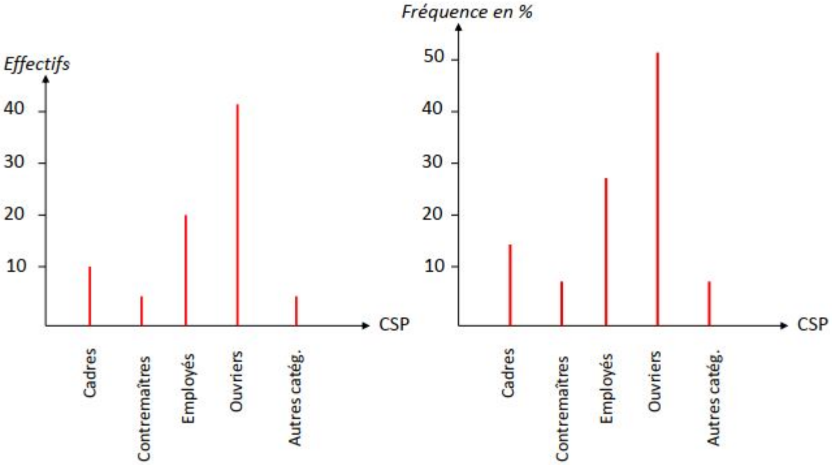


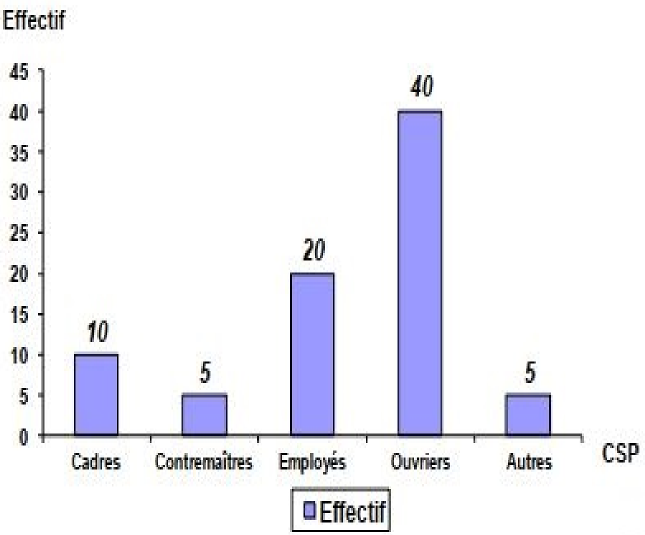
FIGURE : *

Répartition des salariés de l'entreprise selon la CSP.

Dans un diagramme à bandes, on associe une bande verticale à chaque modalité. La largeur de chacune de bande est la même et sa hauteur est proportionnelle à l'effectif ou la fréquence de la modalité correspondante. La distance entre les bandes est constante. Au dessus de chaque bande on note des étiquettes permettant connaître l'effectif ou la fréquence de la modalité associée.

Exemple

Reprenons l'exemple A, alors le diagramme à bandes correspondant aux effectifs des salariés de l'entreprise X est présenté comme suit :



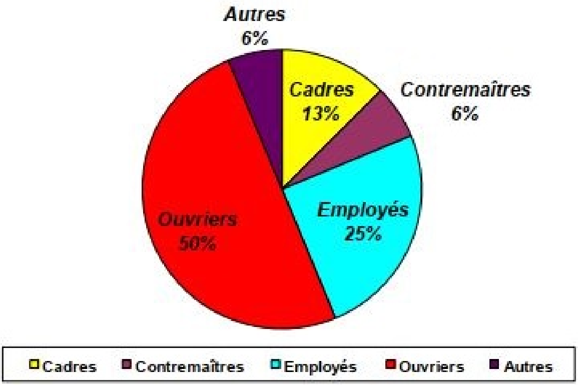
Répartition des salariés de l'entreprise selon la CSP.

C'est un disque divisé en secteurs angulaires représentant l'ensemble de la population. Les différentes modalités du caractère sont représentées par des angles aux centres proportionnelles aux effectifs ou fréquences de leurs modalités respectives. L'angle de chaque secteur α_i est proportionnel à la fréquence f_i : $\alpha_i = 360 \times f_i$.

Exemple

Répartition des salariés de l'entreprise X selon la CSP.

CSP : x_i	Effectif n_i	Fréquence f_i	Angle α_i
Cadres	10	0,125	45
Contremaîtres	5	0,0625	22,5
Employés	20	0,25	90
Ouvriers	40	0,50	180
Autres	5	0,0625	22,5
Total	80	1	360



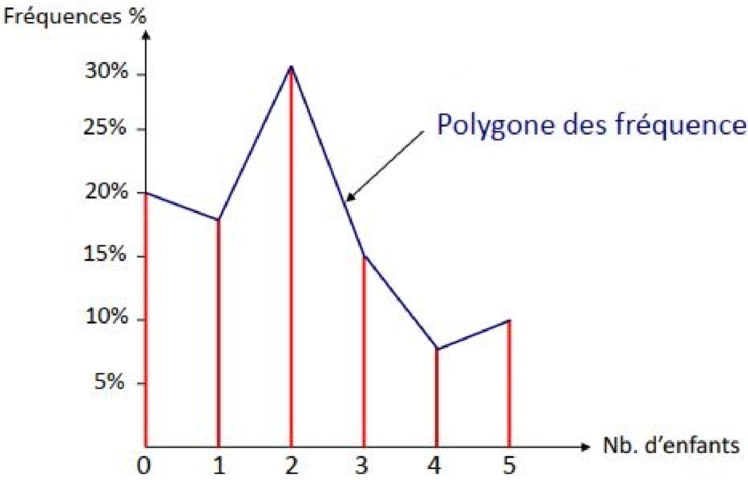
Répartition par secteurs des salariés de l'entreprise selon la CSP.

Représentation d'une distribution de fréquences (ou effectifs). Quand la variable est discrète, les effectifs sont représentés par un diagramme en bâtons.

Exemple

Nombre d'enfants de 40 salariés d'une entreprise :

Nbr. d'enfants	Effectifs	Fréquences en %
0	8	20
1	7	17,5
2	12	30
3	6	15
4	3	7,5
5	4	10
Total	40	100



Distribution des fréquences des salariés selon leur nombre d'enfants.

Le polygone de fréquence est obtenu en joignant les sommets des bâtons par une ligne brisée.

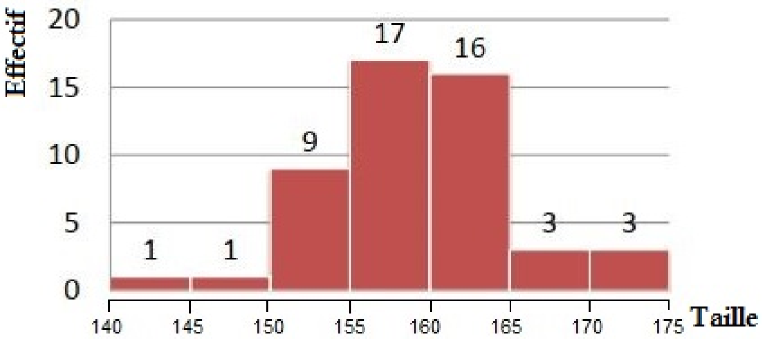
Un histogramme est constitué d'une suite de rectangles, dont les bases coïncident avec les classes divisant le domaine de variation de la variable et dont les hauteurs soient telles que les effectifs (ou les fréquences) sont traduits par les surfaces des rectangles.

Exemple B

Dans le cadre de l'étude de la population des adolescents d'un cartier populaire, les valeurs de leurs tailles peuvent être réparties de la façon suivante :

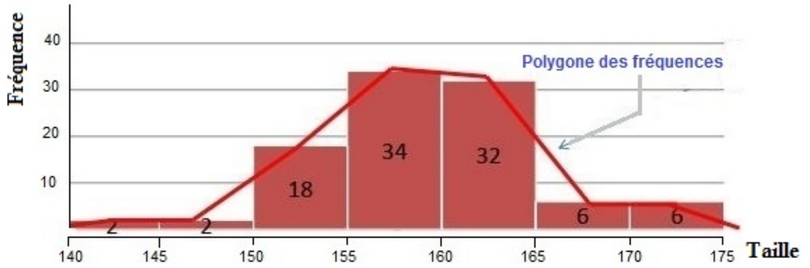
Taille	effectif	Fréquence en %
$[140, 145[$	1	2
$[145, 150[$	1	2
$[150, 155[$	9	18
$[155, 160[$	17	34
$[160, 165[$	16	32
$[165, 170[$	3	6
$[170, 175[$	3	6

L'histogramme des effectifs de cette série est présenté par le graphique suivant :



Effectifs des adolescents selon leurs tailles.

Sur le même graphique des effectifs (resp. des fréquences), on présente le polygone des effectifs (resp. des fréquences). Ce polygone permet de représenter la distribution sous la forme d'une courbe en joignant les milieux des bases supérieures de chaque rectangle de l'histogramme par des segments de droite.



Fréquences des adolescents selon leurs tailles.

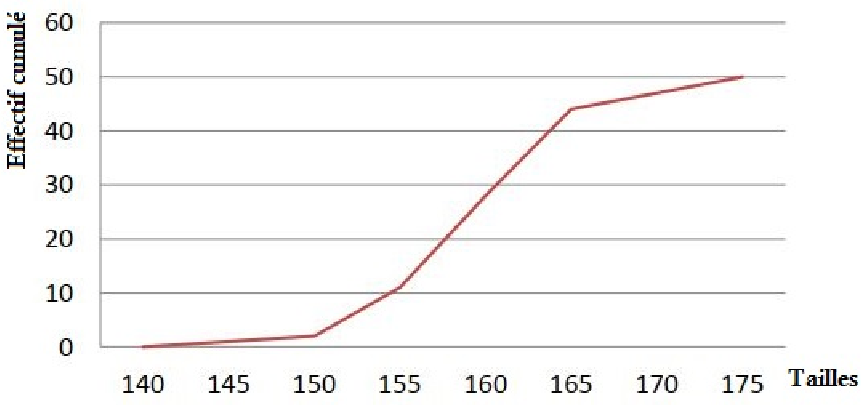
On construit dans un repère cartésien orthogonal les points dont les abscisses sont égales aux bornes supérieurs des classes (sauf pour le premier point), et dont les ordonnées sont les effectifs cumulés croissant correspondants. En joignant ces points par des segments de droites nous obtenons le polygone cumulatif croissant de la distribution donnée.

Exemple

Reprenons l'exemple B. On calcule les effectifs cumulés croissant :

Taille	effectif	Effectif cumulé
$[140,145[$	1	1
$[145,150[$	1	2
$[150,155[$	9	11
$[155,160[$	17	28
$[160,165[$	16	44
$[165,170[$	3	47
$[170,175[$	3	50

Le Polygone des effectifs cumulés des adolescents selon leurs tailles est présenté par le graphique suivant :



Polygone des effectifs cumulés des adolescents selon leurs tailles

On entend par statistique univariée l'étude d'une seule variable, que celle-ci soit quantitative ou qualitative. La statistique univariée fait partie de la statistique descriptive.   Appelés également caractéristiques de position ou de tendance centrale, les paramètres de position tentent de donner une information sur la valeur de la modalité "autour de laquelle se situent les autres modalités" (d'où le terme de tendance "centrale"). Cette notion suppose donc que les modalités constituent un ensemble ordonné, ou pour le moins que la notion de "distance" entre deux modalités ait un sens. Il est bien évident que seuls les caractères quantitatifs permettent de donner un sens précis à ces notions intuitives.

Définition (Le mode)

On appelle mode ou valeur dominante d'une série statistique la valeur observée de la variable ayant le ***plus grand effectif*** *(ou la* ***fréquence la plus élevée****). On note généralement le mode M_o.*

Exemple

Soit la série : $\{8, 4, 4, 3, 4, 3, 8, 2, 5\}$. La valeur la plus fréquente de cette série est 4. Le mode est $M_o = 4$.

Remarques

- *Le mode peut être calculé pour tous les types de variable, quantitative et qualitative.*
- *Le mode n'est pas nécessairement unique, c'est-à-dire, Une série peut avoir plusieurs modes : Soit la série $S = \{4, 0, 1, 1, 2, 2, 2, \underline{3}, \underline{3}, 4, 2, 3, 4, 5, 2, 1, \underline{3}, \underline{3}, 4, 5\}$. "2" et "3" sont les valeurs qui reviennent le plus souvent : 5 fois chacune. Cette série a deux modes : 2 et 3. On peut avoir des séries avec $3, 4, \ldots$ modes. Ce sont alors des séries multimodales.*

Remarques

- *Le mode n'existe pas forcément : C'est le cas lorsque toutes les valeurs ont le même effectif comme dans l'exemple suivant :* $\{8, 6, 5, 7, 3, 1\}$*. Dans ce cas, on peut aussi dire que toutes les valeurs sont modales.*
- *Le mode n'est pas la valeur la plus élevée : Il ne faut pas confondre le mode, qui est la valeur la plus fréquente, avec la valeur la plus élevée de la série. Dans la série* $\{8, 6, 5, 7, 3, 1\}$*, il n'y a pas de mode, mais la valeur la plus élevée est* 8*. Il peut arriver que le mode soit aussi la valeur la plus élevée, mais ce n'est alors qu'une coïncidence.*
- *Quand une variable continue est découpée en classes, on peut définir une classe modale (classe correspondant à l'effectif le plus élevé).*

Le mode dans le cas d'une variable qualitative : Le mode est la modalité correspondante à l'effectif le plus important.

Exemple

Répartition des salariés de l'entreprise X selon le contrat de travail :

CSP	*Cadres supérieurs*	*Contremaîtres*	*Employés*	*Ouvriers spécialisés*	*Autres catégories*
Effectif des salariés	*10*	*5*	*20*	***40***	*5*

Le mode de ce caractère est la modalité «ouvriers spécialisés».

Le mode dans le cas d'une variable quantitative discrète : Le mode est la valeur correspondante à l'effectif le plus important

Exemple

Nombre d'enfants de 40 salariés d'une entreprise

Nbr. d'enfants	Effectifs	Fréquences en %
0	8	20
1	7	17,5
2	**12**	**30**
3	6	15
4	3	7,5
5	4	10
Total	40	100

Le mode de ce caractère est égal à 2 *enfants, c'est-à-dire :* $M_o = 2$.

La classe modale dans le cas d'une variable quantitative continue :

- Si les classes sont toutes de *même amplitude*, la classe modale est la classe d'*effectif le plus élevé* ou de *fréquence la plus élevée*.
- Si les classes *ne sont pas toutes de même amplitude*, on définit l'*effectif corrigé* $\widetilde{n_i}$ (resp. la *fréquence corrigée*) par

$$\widetilde{n_i} = \frac{n_i}{a_i}, \quad \left(\text{resp. } \widetilde{f_i} = \frac{f_i}{a_i}\right),$$

où $a_i = c_i^+ - c_i^-$ l'*amplitude* de la classe i : $[c_i^-, c_i^+[$. Alors la classe modale est la classe dont l'effectif corrigé ou la fréquence corrigée est *maximale*.

Exemple

• On considère la distribution statistique d'une population d'étudiants selon leur taille :

Taille (cm)	$[150, 160[$	$[160, 170[$	$[170, 180[$	$[180, 190[$	$[190, 200[$	Total
Effectif	6	7	**8**	2	1	24
Fréquences en $\%$	25	29,1	**33,3**	8,3	4,3	100

L'effectif ou la fréquence les plus élevés montrent que le classe modale est $[170; 180[$. Soit la distribution d'une population des étudiants répartis suivant leur poids (en Kg) :

Poids (en Kg)	Effectif n_i	Fréquence f_i en $\%$	Amplitude a_i	Fréq. corrigée $\widetilde{f_i}$
$[50; 55[$	2	8,33	5	1,67
$[55; 60[$	3	12,5	5	2,5
$[60; 70[$	4	16,67	10	1,67
$[70; 75[$	5	20,83	5	**4,17**
$[75; 85[$	6	25	10	2,5
$[85; 95[$	4	16,67	10	1,67
Total	24	100		

La classe modale, à laquelle est associée la fréquence corrigée la plus grande, est la classe $[70; 75[$.

Calcul du mode dans le cas d'une variable quantitative continue :

• Cas des effectifs groupés par classes d'amplitudes ***égales*** : Dans ce cas, la classe modale est la classe d'effectif n_i le plus élevé, soit $\left[c_i^-, c_i^+\right[$. L'effectif de la classe qui précède la classe modale est n_{i-1} et celui de la classe qui suit la classe modale est n_{i+1} alors

(1) $M_o = c_i^- + a_i\left(\dfrac{m_1}{m_1+m_2}\right)$

avec

$$m_1 = n_i - n_{i-1},$$
$$m_2 = n_i - n_{i+1}.$$

Exemple

Dans le tableau ci-dessous, les valeurs d'une variable X *ont été groupées par classes de valeurs d'amplitudes égales.*

Classes	n_i	N_i
$[0,5[$	2	2
$[5,10[$	7	9
$[10,15[$	**18**	27
$[15,20[$	3	30

Titre : Valeurs groupées par classes de valeurs d'amplitudes égales.

La classe modale est $[10,15[$. *Donc par application de la formule* (1) *on trouve que le mode de ce caractère est :*

$$M_o = c_i^- + a_i\left(\frac{m_1}{m_1+m_2}\right) = 10+5\times\left(\frac{18-7}{(18-7)+(18-3)}\right) = 10+5\left(\frac{11}{11+15}\right) = 12,115.$$

• Cas des effectifs groupés par classes d'amplitudes ***inégales*** : Dans ce cas, pour calculer le mode, il faut appliquer la formule (1) mais la définition de m_1 et de m_2 change, car il faut remplacer les effectifs n_i par les effectifs corrigées $\widetilde{n}_i = \dfrac{n_i}{a_i}$, c'est-à-dire :

$$\widetilde{m}_1 = \widetilde{n}_i - \widetilde{n}_{i-1}, \quad \text{et} \quad \widetilde{m}_2 = \widetilde{n}_i - \widetilde{n}_{i+1}.$$

Exemple

Soit le tableau suivant où les données sont présentées par classes d'amplitudes inégales.

x_i	n_i	a_i	$\widetilde{n}_i = \frac{n_i}{a_i}$
$[0, 10[$	9	10	0,9
$[10, 12[$	9	2	**4,5**
$[12, 20[$	12	8	1,5

D'après le tableau, la classe modale est $[10, 12[$*, On a donc*

$$\widetilde{n}_{i-1} = \frac{n_{i-1}}{a_{i-1}} = \frac{9}{10} = 0,9; \quad \widetilde{n}_i = \frac{n_i}{a_i} = \frac{9}{2} = 4,5; \quad \widetilde{n}_{i+1} = \frac{n_{i+1}}{a_{i+1}} = \frac{12}{8} = 1,5.$$

$$\Longrightarrow \quad \widetilde{m}_1 = \widetilde{n}_i - \widetilde{n}_{i-1} = 4,5 - 0,9 = 3,6 \quad \text{et} \quad \widetilde{m}_2 = \widetilde{n}_i - \widetilde{n}_{i+1} = 4,5 - 1,5 = 3.$$

On trouve donc

$$M_o = c_i^- + a_i \left(\frac{\widetilde{m}_1}{\widetilde{m}_1 + \widetilde{m}_2} \right) = 10 + 2 \times \left(\frac{3,6}{3,6 + 3} \right) = 11,09.$$

Définition (La médiane)

La médiane M_e est telle que l'effectif des observations dont les modalités sont inférieures à M_e est égal à l'effectif des observations dont les modalités sont supérieures à M_e. Elle partage la série statistique en deux parties d'égal effectif.

Détermination pratique : Il ya deux méthodes.

- On classe les données par ordre croissant, la médiane est la *valeur centrale* qui sépare la série en deux parties égales.
- On la détermine à partir des *effectifs cumulés* ou du *diagramme cumulatif*.

Cas d'une variable discrète : - *À partir de la série statistique :* On classe les données par ordre croissant :

$$x_1 < x_2 < \ldots < x_n.$$

- Si n est *paire* $(n = 2k)$, la médiane est égale à

  $$M_e = \frac{x_{k+1} + x_k}{2}.$$

  Exemple : $S = \{0; 0; 1; 1; 2; 2; 3; 4\}$. La médiane est égale à $M_e = \dfrac{1+2}{2} = 1,5.$
- Si n est *impaire* $(n = 2k+1)$, la médiane est égale à la $(k+1)^{\text{ème}}$ valeur de la série :

  $$M_e = x_{k+1}.$$

  Exemple : $S = \{0; 1; 1; 2; 2; 3; 4\}$. La médiane est égale à $M_e = 2$.

- *À partir du tableau statistique :* On la détermine à partir des *fréquence cumulées*.

Exemple

Nombre d'enfants (x_i) observés dans un échantillon de $n = 55$ familles (n_i) :

x_i	n_i	N_i	f_i en %	F_i en %
0	3	3	5	5
1	4	7	7	12
2	8	15	15	27
3	7	22	13	40
4	14	**36**	25	**65**
5	9	45	16	81
6	6	51	11	92
7	2	53	4	96
8	1	54	2	98
9	1	55	2	100
Total	55		100	

$\Longrightarrow$ *D'après le tableau on a :* $M_e = 4$.

Exemple

Nombre d'enfants par foyer :

x_i	n_i	N_i	f_i en %	F_i en %
0	22	22	27,5	27,5
1	18	40	22,5	**50**
2	16	56	20	**70**
3	10	66	12,5	82,5
4	7	73	8,75	91,25
5	5	78	6,25	97,5
6	2	80	2,5	100
Total	80		100	

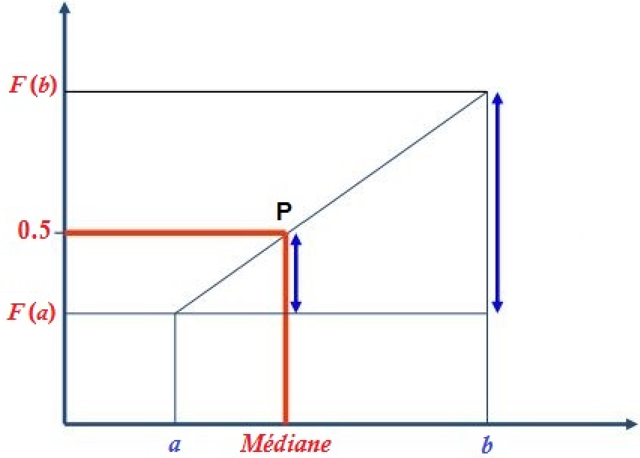
$$\Longrightarrow \quad M_e = \frac{x_{40} + x_{41}}{2} = \frac{1+2}{2} = 1,5.$$

Cas d'une variable continue : Il n'y a aucune différence de calcul pour la médiane selon que les classes sont d 'amplitudes constantes ou variables. Le calcul de la médiane dans le cas de variable continue passe, d'abord, par la détermination de la classe médiane. Ensuite, par interpolation linéaire, on peut calculer la valeur précise de la médiane à l'intérieur de la classe médiane.

• Pour déterminer la médiane M_e, on recherche la classe qui la contient. C'est la classe :

$$\left[c_i^-, c_i^+\right[= [a, b[\text{ telle que } F(a) \leqslant 0,5 < F(b,)$$

en notant F la fonction "*fréquences relatives cumulées*".
Alors, la médiane M_e est l'abscisse du point P d'ordonnée $0,5$ ou 50% dans la représentation graphique des fréquences cumulés croissantes (voir la figure ci-dessous). Son calcul est le suivant :

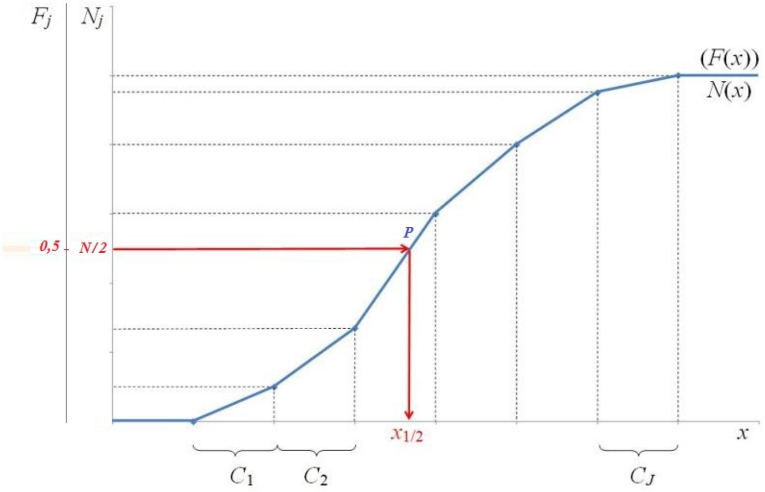

F (b)
P
0.5
F (a)
a
Médiane
b

$$\frac{0.5 - F(a)}{F(b) - F(a)} = \frac{M_e - a}{b - a}$$

$$\Downarrow$$

$$M_e = a + \left[\frac{0.5 - F(a)}{F(b) - F(a)}\right](b - a)\,.$$

Détermination de la médiane par interpolation linéaire.



Exemple

La distribution statistique d'une population d'étudiants selon leur taille (en cm) :

Taille (cm)	$[150;160[$	$[160;170[$	$[170;180[$	$[180;190[$	$[190;200[$	*Total*
Effectif	*6*	*7*	*8*	*2*	*1*	*24*
Fréquences en %	*25*	*29,1*	*33,3*	*8,3*	*4,3*	*100*
Fréquences cumulées croissantes en %	*25*	*54,1*	*87,4*	*95,7*	*100*	

La médiane de cette distribution est :

$$M_e = 160 + \frac{50-25}{54,1-25}(170-160) \approx 168,6.$$

Définition (Moyenne Arithmétique)

La moyenne est la somme des valeurs observées divisée par leur nombre.

La moyenne ne peut être définie que sur une variable quantitative.
La moyenne d'une variable discrète $\{(x_i; n_i)\}$ est donnée par :

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{n_1 + n_2 + \ldots + n_p} = \frac{1}{n}\sum_{i=1}^{p} n_ix_i.$$

Exemple C

L'étude de 20 familles a conduit à la distribution des nombres d'enfants dans chaque famille :

Nombre d'enfants x_i	**0**	**1**	**2**	**3**	**4**	**5**
Nombre de familles n_i	5	3	6	1	3	2

Le nombre moyen d'enfants par famille est :

$$\overline{x} = \frac{(0 \times 5) + (1 \times 3) + (2 \times 6) + (3 \times 1) + (4 \times 3) + (5 \times 2)}{5 + 3 + 6 + 1 + 3 + 2} = \frac{40}{20} = 2.$$

La moyenne d'une variable continue : La variable continue est connue par ses classes $[c_i^-; c_i^+[$ et les effectifs n_i associées à chaque classe. Si on note c_i le centre de la classe $[c_i^-; c_i^+[$, alors $\overline{x} = \dfrac{n_1c_1 + n_2c_2 + \ldots + n_pc_p}{n_1 + n_2 + \ldots + n_p} = \dfrac{1}{n}\sum_{i=1}^{p} n_ic_i.$

Exemple

La distribution des tailles de 40 étudiants est donnée par le tableau ci-après.

Taille en cm x_i	$[150;160[$	$[160;165[$	$[165;170[$	$[170;175[$	$[175;185]$
Nombre d'étudiants n_i	*4*	*8*	*10*	*16*	*2*

La moyenne des tailles des étudiants :

$$\begin{aligned}\overline{x} &= \frac{(155 \times 4) + (162,5 \times 8) + (167,5 \times 10) + (172,5 \times 16) + (180 \times 2)}{4+8+10+16+2} \\ &= \frac{620 + 1300 + 1675 + 2760 + 360}{40} = 167,875.\end{aligned}$$

Dans le paragraphe précédent, nous avons quantifié l'aspect "moyen" d'une série statistique. Il est bien évident que des séries ayant la même moyenne et la même médiane peuvent être trés différemment étalées. Les paramètres de dispersion que nous allons maintenant définir permettent d'apprécier l'importance de la dispersion d'une série statistique autour de sa moyenne ou de sa médiane.

Définition

*On appelle **étendue** d'une distribution statistique quantitative* $X = (x_i, n_i)_{i=1}^{p}$ *la différence entre la plus grande et la plus petite valeur observée, notée* E *:*

$$E = x_{\max} - x_{\min}.$$

L'étendue est simple et facile à calculer. Toutefois, il est très sensible aux valeurs extrêmes "aberrantes". Si la série statistique $(x_i, n_i)_{i=1}^{p}$ est classée dans l'ordre croissant $x_1 < x_2 < \cdots < x_p$, alors

$$E = x_p - x_1.$$

Exemple

On considère la distribution statistique suivante :

x_i	7	8	9	10	11	12	13
n_i	2	3	7	10	6	2	1
N_i	2	5	12	22	28	30	31

l'étendue est :

$$E = x_7 - x_1 = 13 - 7 = 6.$$

En statistique, la *variance* est une mesure arbitraire servant à caractériser la dispersion d'une distribution ou d'un échantillon.

Définition

La variance est la somme des carrés des écarts à la moyenne divisée par le nombre d'observations :

$$V_{ar} = \frac{1}{n} \sum_{i=1}^{p} (x_i - \bar{x})^2 .$$

• La variance d'une variable quantitatif discrète $(x_i, n_i)_{i=1}^{p}$ est exprimée par :

$$V_{ar} = \frac{1}{n} \sum_{i=1}^{p} n_i (x_i - \bar{x})^2 .$$

• La variance d'une variable quantitative continue $\left([c_i^-; c_i^+[, n_i\right)_{i=1}^{p}$ est donnée par :

$$V_{ar} = \frac{1}{n} \sum_{i=1}^{p} n_i (c_i - \bar{x})^2 ,$$

où c_i est le centre de la classe $[c_i^-; c_i^+[$.

Remarques

- *On peut dire que plus la variance est élevée, plus la dispersion autour de la moyenne est élevée. Mais comme les écarts à la moyenne ont été élevés au carré, le chiffre obtenu est assez élevé. C'est pourquoi, on utilise surtout la variance comme calcul intermédiaire pour obtenir l'écart-type et le coefficient de variation.*
- *La variance peut s'écrire encore de la forme suivante :*
  - *Pour une variable quantitative discrète :*

$$V_{av} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i^2 - (\bar{x})^2.$$

  - *Pour une variable quantitative continue :*

$$V_{av} = \frac{1}{n}\sum_{i=1}^{p} n_i c_i^2 - (\bar{x})^2.$$

Exemple D

Reprenons l'exemple C, qui présent la distribution des nombres d'enfants dans 20 familles selon le tableau suivant :

Nombre d'enfants x_i	$\mathbf{0}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$
Nombre de familles n_i	5	3	6	1	3	2

On a déjà calculé le nombre moyen d'enfants par famille : $\overline{x} = 2$.
Pour calculer la variance de cette série statistique, on calcule d'abord les valeurs $n_i x_i^2$.

x_i	**0**	**1**	**2**	**3**	**4**	**5**
n_i	5	3	6	1	3	2
x_i^2	0	1	4	9	16	25
$n_i x_i^2$	0	3	24	9	48	50

On a donc :

$$V_{ar} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i^2 - (\bar{x})^2 = \frac{1}{20}\sum_{i=1}^{6} n_i x_i^2 - (\bar{x})^2 = \frac{1}{20}(0+3+24+9+48+50) - 2^2 = 2,7.$$

Exemple E

Reprenons l'exemple qui présent la distribution des tailles de 40 étudiants est donnée par le tableau ci-après.

Taille en cm $[c_i^-; c_i^+[$	$[150; 160[$	$[160; 165[$	$[165; 170[$	$[170; 175[$	$[175; 185]$
Nombre d'étudiants n_i	4	8	10	16	2

On a déjà calculé la moyenne des tailles des étudiants : $\bar{x} = 167,875$.
Pour calculer la variance de cette série statistique, on calcule d'abord les valeurs $c_i x_i^2$.

$[c_i^-; c_i^+[$	$[150; 160[$	$[160; 165[$	$[165; 170[$	$[170; 175[$	$[175; 185]$	Total
n_i	4	8	10	16	2	40
$\boldsymbol{c_i}$	155	162,5	167,5	172,5	180	
$\boldsymbol{x_i^2}$	24025	26406	28056	29756	32400	
$\boldsymbol{n_i x_i^2}$	96100	211250	280562,5	476100	64800	1128813

On a donc

$$V_{av} = \frac{1}{n}\sum_{i=1}^{p} n_i c_i^2 - (\bar{x})^2 = \frac{1}{40}\sum_{i=1}^{5} n_i c_i^2 - (167,875)^2 = \frac{1}{40}(1128813) - (167,875)^2 \simeq 38,3.$$

L'écart type est la mesure la plus courante de la dispersion ou de la répartition des données sur la moyenne.

Définition

L'écart-type *d'une variable, noté par σ, est la racine carrée de la variance :* $\sigma = \sqrt{V_{ar}}$.

Remarques

- *L'écart-type est la mesure de la dispersion la plus couramment utilisée.*
- *Si l'écart-type est petit, cela signifie que les valeurs sont assez concentrées autour de la moyenne. Autrement dit, les valeurs observées sont* `peu dispersées` *(*`homogène`*).*
- *Si l'écart-type est élevé, cela veut dire au contraire que les valeurs sont* `plus dispersées` *autour de la moyenne. Autrement dit, les valeurs observées sont* `hétérogènes`*.*
- *Quelque soit la distribution statistique étudiée, un intervalle dont les extrémités sont «$\overline{x} - 2\sigma$» et «$\overline{x} + 2\sigma$» (i.e. $[\overline{x} - 2\sigma, \overline{x} + 2\sigma]$) contient toujours au moins 75% des unités constituant la population étudiée.*

Exemple

Reprenons les exemples D et E. Alors :

• *L'écart type de la série statistique étudiée dans l'exemple D est :*

$$\sigma = \sqrt{V_{ar}} = \sqrt{2.7} = 1,64.$$

• *L'écart type de la série statistique étudiée dans l'exemple E est :*

$$\sigma = \sqrt{V_{ar}} = \sqrt{38,3} = 6,19.$$

`Le coefficient de variation`, noté par C_{v}, est une mesure de la répartition qui décrit la variation des données par rapport à la moyenne : $C_{\mathsf{v}} = \dfrac{\sigma}{\overline{x}}$. Le coefficient de variation est ajusté de façon à ce que les valeurs soient sur une échelle `sans unités`. C'est pourquoi vous pouvez utiliser le coefficient de variation à la place de l'écart type pour comparer la variation des données ayant des unités ou des moyennes très différentes.
Plus le coefficient de variation est élevé, plus les données sont dispersées.

Exercice

Sur une population bien portante issues d'une certaine région, on a dosé le taux de cholestérol, exprimé en cg/ℓ *et on a obtenu les résultats suivants :*

Classes	$[84, 96[$	$[96, 108[$	$[108, 120[$	$[120, 132[$	$[132, 144[$	$[144, 156[$	$[156, 168[$
Effectif n_i	2	8	16	19	2	2	1

1. *Quel est le nombre d'individus ?*
2. *Quel est le caractère étudié ?*
3. *Déterminer le mode.*
4. *Déterminer la médiane.*
5. *Déterminer la moyenne arithmétique.*
6. *Déterminer l'écart-type.*
7. *Calculer le coefficient de variation, puis conclure.*