

COURS 1 : PUISSANCES

Soit b un nombre *strictement positif* et n un entier positif. Le nombre b^n , appelé b à la **puissance** n , est le nombre

$$b^n := \underbrace{b \times b \times \dots \times b}_{n \text{ fois}}$$

IDENTITÉS ET PUISSANCES

On dispose des identités suivantes :

- $a^n \times b^n = (ab)^n$;
- $(b^n)^m = b^{nm}$;
- $b^0 = 1$;
- $b^1 = b$.

Par exemple,

$$\begin{aligned} 10^2 \times 3^2 &= 100 \times 9 = 900 & \text{et} & \quad (10 \times 3)^2 = 30^2 = 900 \\ (10^2)^3 &= 100^3 = 1\,000\,000 & \text{et} & \quad 10^{2 \times 3} = 10^6 = 1\,000\,000 \end{aligned}$$

Attention, les puissances se comportent plus mal avec l'addition (rappelez-vous des identités remarquables!) :

$$(2 + 3)^2 = 5^2 = 25 \text{ mais } 2^2 + 3^2 = 4 + 9 = 13$$

VALEURS DE L'EXPOSANT

On peut généraliser ce qui précède à des valeurs de n négatives, en posant :

$$b^{-n} := \frac{1}{b^n}.$$

On peut aussi prendre des puissances non entières. Par exemple, le nombre $b^{1,3}$ sera bien défini, et compris entre b^1 et b^2 . Tout ce qui précède reste valable.

Un cas particulier est la racine carrée : $b^{0,5} = \sqrt{b}$, car $(b^{0,5})^2 = b^{0,5 \times 2} = b^1 = b$.

Par exemple,

$$\begin{aligned} 2^{-3} &= \frac{1}{2^3} = \frac{1}{8} = 0,125 \\ 4^{0,5} &= \sqrt{4} = 2 \\ 10^{0,3} &\simeq 1,995 \end{aligned}$$

COURS 3 : ÉVOLUTION À TAUX CONSTANT

On s'intéresse à une grandeur qui dépend du temps, comme par exemple :

- la population d'une ville, d'une région, d'un pays ;
- le nombre de clients d'une entreprise ;
- l'argent placé sur un compte dédié...

On se donne une période de référence : une semaine, un mois, une année...

Soit y_0 la valeur de cette grandeur au début d'une période et y_1 celle à la fin. On dit que cette valeur a cru d'un taux t si :

$$y_1 = (1 + t)y_0.$$

Attention à convertir les pourcentages : un taux de 30% correspond à $t = 30/100$.

Soit y_n la valeur après n périodes (par exemple, la population d'une ville n mois après le 1er janvier 2020, ou l'argent sur un compte l'année 2020 + n). On dit que (y_n) **évolue à taux constant** si elle a le même taux de croissance à chaque période.

$$y_0 \xrightarrow{1+t} y_1 = y_0(1+t) \xrightarrow{1+t} \dots \xrightarrow{1+t} y_n = y_0(1+t)^n$$

Si (y_n) suit une évolution à taux constant t , alors :

$$(0.1) \quad y_n = y_0(1+t)^n.$$

On appelle $(1+t)$ le coefficient multiplicateur. Si on connaît deux des trois grandeurs y_0 (valeur initiale), y_n (valeur finale) et t (taux d'évolution), on peut en déduire la troisième.

Attention : le taux dépend de la période choisie ! Un taux de 10% par mois n'est pas la même chose qu'un taux de 10% par an. Dans la formule ci-dessus, il faut rester cohérent : si n est en mois, t est le taux mensuel ; si n est en années, t est le taux annuel.

VALEURS DE n

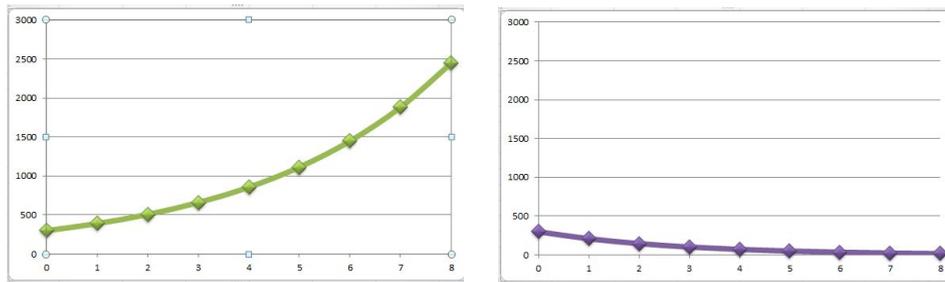
Comme vu dans le cours précédent, la valeur de n peut être décimale. Cela permet de calculer l'évolution après un nombre de périodes non entier. Par exemple, si le taux de croissance annuel est de 30%, alors on peut calculer l'évolution après 1 an et 6 mois = 1,5 ans :

$$y_{1,5} = y_0 \times 1,3^{1,5}.$$

Comme vu dans le cours précédent, la valeur de n peut aussi être négative. Cela permet de retrouver la valeur de y_n dans le passé.

REPRÉSENTATION

Enfin, si $t > 0$, alors (y_n) croît (taux de croissance positif), d'autant plus vite que t est élevé ; si $-1 < t < 0$, alors (y_n) décroît (taux de croissance négatif). Dans ce cours, la valeur de t ne sera jamais inférieure à $-1 = -100\%$.



Graphe de (y_n) pour un taux de croissance $t = 30\%$ (à droite) et $t = -30\%$ (à gauche).

COURS 4 : CALCUL ET CONVERSION DE TAUX

On considère une grandeur (y_n) évoluant à taux constant. Comme vu lors du cours précédent,

$$y_n = y_0(1+t)^n,$$

où t est le taux de croissance.

CALCUL DE TAUX

Si l'on connaît y_0 et y_n (pour une certaine valeur de n), alors on peut calculer ce taux. En effet,

$$(1+t)^n = \frac{y_n}{y_0},$$

et donc le coefficient multiplicateur vaut :

$$\left(\frac{y_n}{y_0}\right)^{\frac{1}{n}} - 1 = ((1+t)^n)^{\frac{1}{n}} - 1 = (1+t)^{\frac{n}{n}} - 1 = (1+t)^1 - 1 = t.$$

Exemple : Le nombre d'habitants d'une ville passe de 100 000 au 1er janvier 2019 à 115 000 au 1er janvier 2020. Quel est son taux de croissance *mensuel* ?

Une période d'un an, donc de 12 mois, s'est écoulée. On connaît $y_0 = 100\,000$ et $y_{12} = 115\,000$. On calcule donc :

$$t = \left(\frac{y_{12}}{y_0}\right)^{\frac{1}{12}} - 1 \simeq 1,0117 - 1 = 0,0117 = 1,17\%.$$

CONVERSION DE TAUX

Dans l'exemple précédent, le taux de croissance *annuel* est de 15%. Un taux annuel de 15% correspond donc à un taux mensuel de 1,17%. Ce même type de calcul permet de convertir des taux annuels, mensuel, hebdomadaires... entre eux.

Notons t_a le taux annuel, t_m le taux mensuel, et t_t le taux trimestriel. Comme 1 an = 12 mois, le coefficient multiplicateur sur un an vaut :

$$(1+t_a) = (1+t_m)^{12}.$$

Cette égalité permet de calculer le taux annuel t_a si l'on connaît le taux mensuel t_m . De plus,

$$(1+t_a)^{\frac{1}{12}} = (1+t_m),$$

ce qui permet de calculer t_m si l'on connaît t_a . Les conversions avec le taux trimestriel sont similaires : comme 1 an = 4 trimestres,

$$(1+t_a) = (1+t_t)^4 \quad \Leftrightarrow \quad (1+t_a)^{\frac{1}{4}} = (1+t_t).$$

On peut convertir de même entre taux mensuel et taux trimestriel.

Attention : La conversion de taux n'est pas immédiate. Il n'y a pas de méthode plus simple ; en particulier, le taux annuel n'est **pas** égal à 12 fois le taux mensuel ! Dans l'exemple donné, un taux annuel de 15% correspond à un taux mensuel de 1,17%, mais $12 \times 1,17\% = 14,04\% \neq 15\%$.

COURS 5 : LOGARITHME ET RECHERCHE DE DURÉE

On considère une grandeur (y_n) évoluant à taux constant. Comme vu lors des cours précédents,

$$y_n = y_0(1+t)^n,$$

où t est le taux de croissance. Connaissant le taux t et la valeur initiale y_0 , on veut savoir combien de temps il faut pour atteindre une cible y_n fixée. Autrement dit, connaissant y_0 , y_n et t , on cherche n .

LOGARITHME

Pour cela, il va falloir utiliser la fonction logarithme. Le **logarithme décimal** \log d'un nombre $x > 0$ est la puissance de 10 auquel il correspond :

$$10^{\log(x)} = x.$$

Exemple : $\log(100) = 2$ car $10^2 = 100$; $\log(0,1) = -1$ car $10^{-1} = 0,1$; $\log(2) \simeq 0,301$ car $10^{0,301} \simeq 2$.

La fonction logarithme vérifie les propriétés suivantes (qui se déduisent des propriétés des puissances) :

- $\log(xy) = \log(x) + \log(y)$;
- $\log(b^x) = x \log(b)$;
- $\log(1) = 0$;
- $\log(10) = 1$.

Nous nous servirons essentiellement de la seconde de ces propriétés.

RECHERCHE DE DURÉE

Soient $a > 0$, $b > 0$, n trois nombres tels que $b^n = a$. En prenant le logarithme de chaque côté, on trouve

$$\log(b^n) = n \log(b) = \log(a),$$

et donc :

$$n = \frac{\log(a)}{\log(b)}.$$

Exemple : La population d'une ville est de 100 000 habitants, et augmente de 10% par an. Dans combien de temps aura-t-elle atteint 200 000 habitants ?

Soit n le nombre d'années avant que la population atteigne 200 000 habitants. Comme on a une évolution à taux constant,

$$200\,000 = 100\,000(1+0,1)^n,$$

et donc $2 = 1,1^n$. En appliquant la formule précédente, on trouve $n = \log(2)/\log(1,1) \simeq 7,27$ ans $\simeq 7$ ans et 3 mois. Remarquons au passage que si la population augmente de 10% par an, le temps de doublement est très inférieur à 10 ans !

COURS 6 : TAUX D'INTÉRÊTS ET ACTUALISATION

Les intérêts, qu'ils soient sur un capital déposé (livret A...) ou emprunté, sont en général composés : d'un mois sur l'autre, le capital est multiplié par un coefficient $(1 + t)$, où t est le taux d'intérêt mensuel. Si le taux d'intérêt peut varier, dans de nombreuses situations, il est constant ou presque constant. Dans ce cas, le capital suit une **évolution à taux constant** :

$$C_n = C_0(1 + t)^n.$$

Toutes les méthodes des cours précédents s'appliquent donc :

- Si l'on connaît le capital initial C_0 et le taux t , on peut calculer le capital en tout temps.
- Si l'on connaît le capital à deux instants donnés, on peut calculer le taux d'intérêt t .
- On peut convertir des taux d'intérêts (annuels, mensuels...).
- Si l'on connaît le capital initial C_0 et le taux t , on peut calculer le temps nécessaire pour atteindre un certain capital cible.

ACTUALISATION

Lorsqu'un compte rapporte des intérêts, on ne peut pas ajouter des capitaux correspondant à des dates différentes. Par exemple, si je dépose 100€ au 1er janvier 2019 et 100€ au 1er janvier 2020, je n'ai pas 200€ : il me faut rajouter les intérêts perçus. Il y a principalement deux méthodes pour le faire :

- Actualiser à chaque nouvelle opération.
- Traiter séparément chaque somme déposée.

La première méthode est de loin la plus intuitive et la plus flexible, mais la seconde est utile dans le contexte des prêts.

Exemple : Je dépose sur un compte 1 000€ au 1er janvier 2010, puis 500€ au 1er janvier 2012 et 2 000€ au 1er janvier 2015. Le taux d'intérêts annuel est de 5%. Combien ai-je au 1er janvier 2020 ?

Date	Capital (€)
01/01/2010	1 000
Méthode 1 : 01/01/2012	$1\,000 \times 1,05^2 + 500 = 1\,602,5$
01/01/2015	$1\,602,5 \times 1,05^3 + 2\,000 = 3\,855,09$
01/01/2020	$3\,855,09 \times 1,05^5 = 4\,920,19$

Montant (€)	1 000	500	2 000
Méthode 2 : Durée d'intérêts (ans)	10	8	5
Valeur (€) au 01/01/2020	$1\,000 \times 1,05^{10} = 1\,628,89$	$500 \times 1,05^8 = 738,73$	$2\,000 \times 1,05^5 = 2\,552,56$

Soit un total final de 4920,18€.

Ces deux méthodes s'appliquent aussi bien à des montants négatifs (retraits).

COURS 7 : EMPRUNTS ET MENSUALITÉS/ANNUITÉS

On emprunte un capital C à taux constant t . On rembourse pendant n périodes un montant constant p . Ces quantités sont reliées par la relation :

$$(0.1) \quad C = p \left(\frac{1 - (1 + t)^{-n}}{t} \right).$$

Attention : Encore une fois, il faut faire attention aux unités. Si la période de référence est le mois, alors t est le taux mensuel, les remboursements p sont mensuels (*mensualités*), et la durée du prêt n est en mois. Si la période de référence est l'année, alors t est le taux annuel, les remboursements p sont annuels (*annuités*), et la durée du prêt n est en années. Si les unités ne concordent pas (taux annuels mais remboursements mensuels, par exemple), il faut faire des conversions (par exemple, calculer le taux d'intérêt mensuel).

Attention : Faites attention aux signes (en particulier dans l'exposant) !

Cette relation permet entre autres :

- Si l'on connaît le montant des remboursements p , la durée du prêt n et le taux du prêt t , de calculer le capital emprunté C .
- Si l'on connaît le capital à emprunter C , la durée du prêt n et le taux du prêt t , de calculer le montant des remboursements p .
- Si l'on connaît le capital à emprunter C , le montant des remboursements p et le taux du prêt t , de calculer la durée du prêt n .

COURS 8 : VOCABULAIRE DE STATISTIQUES DESCRIPTIVES

Une **population** statistique est l'ensemble sur lequel on effectue des observations. Les éléments de cet ensemble sont appelés les **individus**.

Ce qui est observé ou mesuré sur les individus d'une population statistique est une **variable (ou caractère) statistique**. Les valeurs prises par une variable statistique sont appelées ses **modalités**.

L'**effectif** d'une modalité i est le nombre d'individus n_i ayant cette modalité, et sa **fréquence** f est la proportion d'individus ayant cette modalité, c'est-à-dire

$$f = \frac{n_i}{n},$$

où n est l'effectif total.

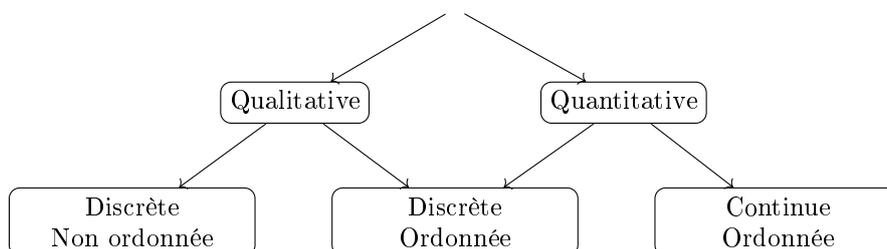
Exemple : Les clients d'une entreprise forment une population. On peut observer leur âge, leurs dépenses annuelles, etc. : ce sont des modalités.

TAXONOMIE

Nous allons faire trois distinctions importantes :

- Une variable est dite **quantitative** si ses modalités représentent une quantité. On peut calculer la moyenne et les autres paramètres statistiques. Une variable qui n'est pas quantitative est dite **qualitative**.
- Une variable est dite **ordonnée** si ses modalités ont un ordre naturel, et **non ordonnée** sinon.
- Une variable est dite **discrète** si elle ne peut prendre que des valeurs isolées, généralement entières. Une variable est dite **continue** si elle peut prendre n'importe quelle valeur dans un intervalle, en particulier des nombres décimaux.

Toute variable qualitative est discrète, et toute variable quantitative est ordonnée. La détermination de caractères d'une variable peut donc se faire à l'aide du diagramme suivant :



Attention : Une variable représentée par un numéro n'est pas quantitative ! Par exemple, un numéro de téléphone, un numéro de département ou une adresse sont des variables qualitatives : faire une moyenne de numéros de téléphones n'a aucun sens.

REPRÉSENTATION

De nombreux types de représentations graphiques existent. Dans les grandes lignes (des exceptions existent toujours!), nous recommandons :

- variables **discrètes, non ordonnées** : diagramme circulaire (“camembert”);
- variables **discrètes, ordonnées** : diagramme en bâtons;
- variables **continues** : histogrammes.

Cette recommandation n’est pas rigide, et on pourra par exemple utiliser un diagramme en bâtons au lieu d’un diagramme circulaire s’il y a trop de modalités (plus de 4 ou 5), ou un histogramme au lieu d’un diagramme en bâtons pour une variable quantitative discrète ayant beaucoup de modalités.

Bien entendu, cela n’aborde pas d’autres types de représentations (cartes, notamment).

COURS 9 : HISTOGRAMMES ET FRÉQUENCES CUMULÉES

On étudie une variable continue dans une population donnée, par exemple la taille en centimètres.

HISTOGRAMMES

Afin de représenter graphiquement la distribution de cette variable, on va utiliser un **histogramme**. La première étape consiste à trier les modalités dans des **classes**; idéalement, ces classes ont des extrémités qui sont des nombres “ronds”, et il faut choisir un juste équilibre (trop peu de classes et on perd beaucoup d’informations; trop de classes et il y aura peu d’individus dans chaque classe). Par exemple, disons que l’on obtient :

Taille (en cm)	[120,150[[150,160[[160,170[[170,180[[180,190[[190, 210[
Effectif	3	22	65	62	16	4

L’**amplitude** d’une classe est la longueur de l’intervalle correspondant. Par exemple, la classe [120, 150[a une amplitude de $150 - 120 = 30$ centimètres.

La première étape consiste à calculer la fréquence de chaque classe. Pour cela, on calcule l’effectif total $N = 3 + 22 + 65 + 62 + 16 + 4 = 172$, et on divise chaque effectif par l’effectif total. Par exemple, la fréquence de la classe [120, 150[est $3/172 \simeq 1,7\%$. On obtient :

Taille (en cm)	[120,150[[150,160[[160,170[[170,180[[180,190[[190, 210[
Fréquence (en %)	1,7	12,8	37,8	36,0	9,3	2,3

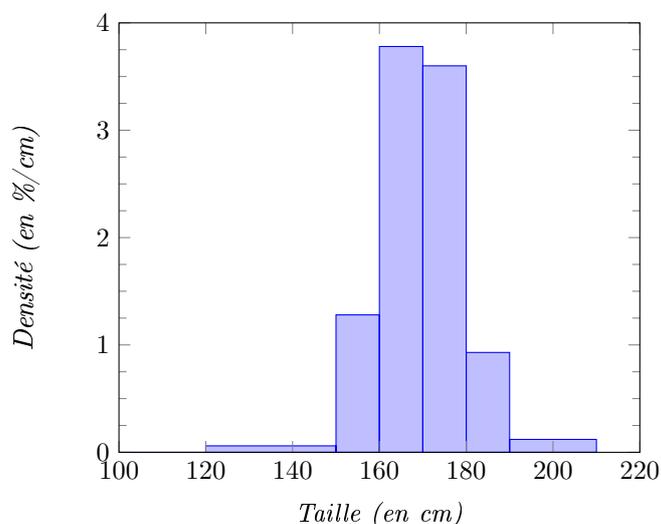
Quand on trace un histogramme, on met en abscisse la variable observée (ici, la taille en centimètres), que l’on divise selon les classes. La population de chaque classe est représentée par un rectangle. Cependant, pour le graphe obtenu ne soit pas trompeur, on ne met pas en ordonnée la fréquence, mais la **densité** :

$$\text{densité} = \frac{\text{fréquence}}{\text{amplitude}}.$$

Le tableau suivant résume cela :

Taille (en cm)	[120,150[[150,160[[160,170[[170,180[[180,190[[190, 210[
Fréquence (en %)	1,7	12,8	37,8	36,0	9,3	2,3
Amplitude (en cm)	30	10	10	10	10	20
Densité (en %/cm)	0,06	1,28	3,78	3,60	0,93	0,12

Nous pouvons maintenant tracer l'histogramme :



Dans un histogramme, la fréquence d'une classe est égale à l'aire du rectangle correspondant. Par exemple, le rectangle correspondant à la classe $[160, 170[$ a une largeur de 10 centimètres (c'est l'amplitude de la classe) et une hauteur de 3,78% par centimètre (c'est la densité). Son aire est donc de :

$$\text{amplitude} \times \text{densité} = (10 \text{ cm}) \times (3,78\%/ \text{cm}) = 37,8\%,$$

ce qui est bien la fréquence de cette classe.

Attention : La densité a une unité étrange (ici en % par centimètre; elle serait, par exemple, en % par an si on mesurait des âges en années, ou en % par M€ si on mesurait des chiffres d'affaires en M€).

FRÉQUENCES CUMULÉES

Étant donnée une variable quantitative, la fréquence cumulée à la valeur x est la fréquence des modalités inférieures ou égale à x . On la notera F_x . Afin de montrer cette construction dans le cas de variables continues triées par classes, reprenons l'exemple ci-dessus :

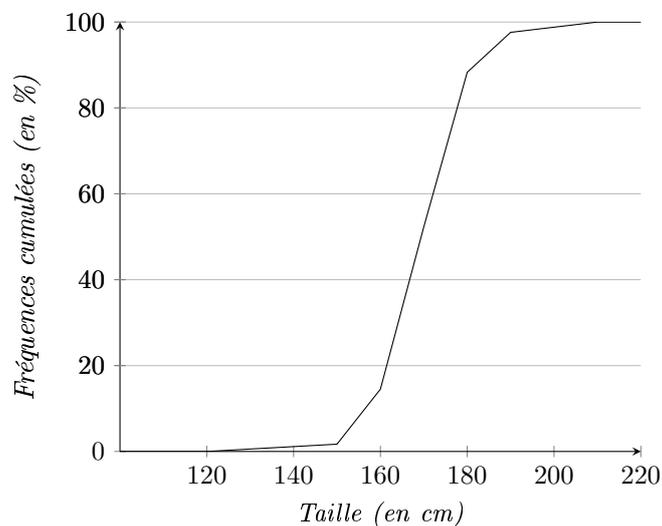
- Personne n'a une taille inférieure ou égale à 120 cm, donc $F_{120} = 0$.
- 1,7% de la population a une taille inférieure ou égale à 150 cm, donc $F_{150} = 1,7\%$.
- $1,7 + 12,8 = 14,5\%$ de la population a une taille inférieure ou égale à 160 cm, donc $F_{160} = 14,5\%$.
- $14,5 + 37,8 = 52,3\%$ de la population a une taille inférieure ou égale à 170 cm, donc $F_{170} = 52,3\%$.

On peut ainsi calculer de proche en proche les fréquences cumulées pour les valeurs de x qui sont les extrémités des classes.

Taille (en cm)	120	150	160	170	180	190	210
Fréquences cumulées (en %)	0	1,7	14,5	52,3	88,3	97,6	100

Attention : La fréquence cumulée au plus bas échelon doit valoir 0% (personne n'est plus petit), et à l'échelon le plus élevée doit valoir 100% (tout le monde est plus petit), éventuellement avec un erreur d'arrondi.

Pour tracer la courbe des fréquences cumulées, on place en abscisse la taille en centimètres, et en ordonnées les fréquences cumulées. On place d'abord les points dont les coordonnées sont données par le tableau ci-dessus (par exemple, le point d'abscisse 150 cm et d'ordonnée 1,7%), puis on relie les points par des segments.



Remarque : La pente d'un segment est donnée par la densité de la classe, donc par la hauteur de l'histogramme : plus un rectangle est haut dans l'histogramme, plus la pente est raide dans la courbe des fréquences cumulées.

COURS 10 : PARAMÈTRES DE POSITION

Étant donnée une variable statistique quantitative, nous définissons deux **paramètres de position**, c'est-à-dire des valeurs autour desquelles la variable est centrée :

- la moyenne ;
- la médiane.

MOYENNE

Supposons que l'on dispose d'une série statistique brute (x_1, \dots, x_N) , où l'effectif total est N . La **moyenne** est définie comme :

$$\bar{x} := \frac{x_1 + x_2 + \dots + x_N}{N}.$$

En général, avec des variables discrètes, on ne travaille pas avec la série brute mais avec la série triée par classes : pour chaque modalité i , on calcule l'effectif n_i ou la fréquence $f_i = n_i/N$. La moyenne est alors :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N} = f_1x_1 + f_2x_2 + \dots + f_px_p.$$

Dans le cas de variables continues triées par classes, on ne dispose pas des valeurs exactes x_i . La méthode la plus simple consiste à utiliser pour chaque classe la valeur centrale de la classe, c'est-à-dire la moyenne de ses extrémités.

Exemple : considérons l'exemple, utilisé la semaine passée, de la taille dans une promotion d'étudiants :

Taille (en cm)	[120,150[[150,160[[160,170[[170,180[[180,190[[190, 210[
Centre (en cm)	135	155	165	175	185	200
Fréquence (en %)	1,7	12,8	37,8	36,0	9,3	2,3

La taille moyenne vaut alors (attention à convertir les pourcentages en nombres décimaux!) :

$$0,017 \times 135 + 0,128 \times 155 + 0,378 \times 165 + 0,360 \times 175 + 0,093 \times 185 + 0,023 \times 200 = 169,31.$$

La taille moyenne de cette population est donc de 169 centimètres.

MÉDIANE

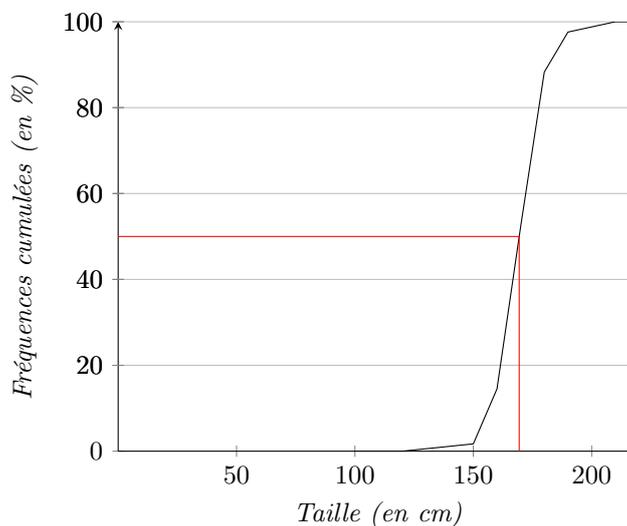
Supposons que l'on dispose d'une série statistique brute (x_1, \dots, x_N) , classée en ordre croissant, où l'effectif total est N . La **médiane** est la valeur centrale¹ de la série.

En général, avec des variables discrètes, on ne travaille pas avec la série brute mais avec la série triée par classes. Il suffit alors de calculer les fréquences cumulées, et de choisir la première valeur x dont la fréquence cumulée associée F_x est supérieure ou égale à 50%.

1. Il y a des conventions – peu importantes – si l'effectif total est pair.

Dans le cas de variables continues triées par classes, on ne dispose pas des valeurs exactes x_i . La méthode la plus simple consiste à chercher la valeur de x dont la fréquence cumulée associée est 50%. Cela se détermine aisément graphiquement².

Exemple : reprenons l'exemple précédent :



Graphiquement, on trouve une médiane d'environ 170 centimètres (une valeur plus précise, accessible par le calcul, serait de 169,39 centimètres).

Remarque : la moyenne peut être fortement affectée par des valeurs extrêmes, là où la médiane y est insensible. Si la distribution est à peu près symétrique, comme dans l'exemple utilisé ici, la moyenne sera très proche de la médiane. Cependant, dans le cas de distributions concentrées vers le bas ou vers le haut, la différence entre les deux peut être non négligeable. C'est le cas par exemple des revenus : une minorité de revenus élevés tire la moyenne vers le haut. Par exemple, en France en 2019, le salaire médian est de 1789€, alors que le salaire moyen est de 2238€ – une différence de 20% ! Ce phénomène se retrouve aussi pour les dépenses des clients dans certains secteurs.

2. Au vu de l'imprécision associée au tri par classes, cette détermination graphique est suffisante.

COURS 11 : FRACTILES

Les **fractiles** sont une généralisation de la médiane pour des variables statistiques quantitatives. Rappelons que la médiane d'une série statistique est la valeur du milieu, c'est-à-dire la plus petite valeur x telle que $F_x \geq 50\%$, où F_x est la fréquence cumulée.

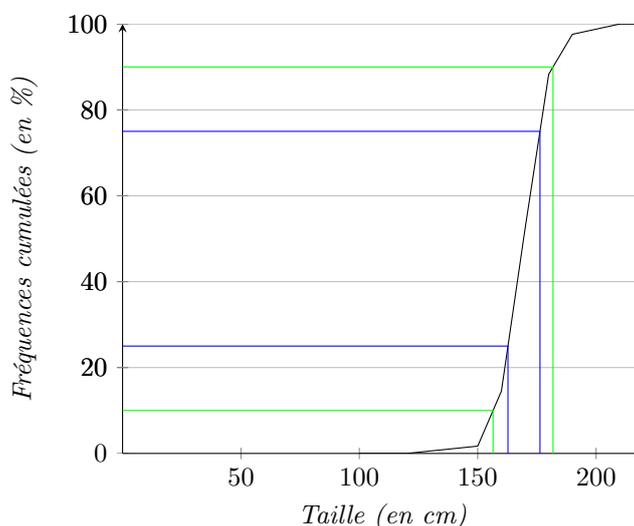
QUARTILES, DÉCILES, CENTILES

Les premier, deuxième et troisième **quartiles** sont valeurs à $1/4$, à $1/2$ et à $3/4$ de l'effectif. Par exemple, le premier quartile est la plus petite valeur x telle que $F_x \geq 25\%$. On notera le premier quartile **Q1** et le troisième **Q3**; le second quartile n'est rien d'autre que la médiane.

On peut de même définir les **déciles D1, D2, ... , D9**, qui sont les valeurs à 10%, 20%, ... , 90% de l'effectif. On parle parfois aussi de **centiles C1, C2, ... , C99**.

Comme pour la médiane, il est possible de déterminer les fractiles graphiquement à l'aide de la courbe des fréquences cumulées. On place le seuil voulu (par exemple, 25% pour le premier quartile) sur l'axe des ordonnées, on se déplace horizontalement jusqu'à la courbe des fréquences cumulées, puis on descend verticalement pour repérer l'abscisse.

Exemple : reprenons l'exemple de la taille, présenté dans le résumé de cours précédent :



Les quartiles sont donnés en bleu, et les premier et neuvième déciles en vert. Graphiquement, on trouve $Q1 \simeq 160$ cm, $Q3 \simeq 175$ cm, $D1 \simeq 155$ cm et $D9 \simeq 180$ cm (des valeurs plus précises, accessibles par le calcul, seraient de 162,78 et 176,31 centimètres pour les quartiles, et de 156,48 et 181,83 centimètres pour les déciles).

DIAGRAMME EN BOÎTE

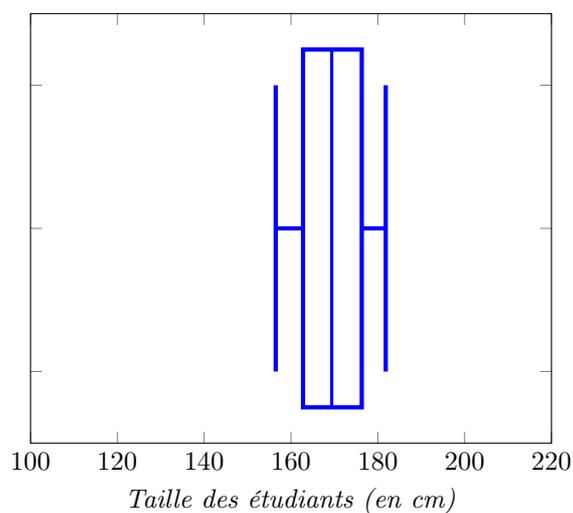
Le **diagramme en boîte**, aussi appelé **boîte à moustache**, est une représentation graphique qui ne garde que peu d'information, mais permet de visualiser et comparer rapidement des médianes, quartiles et les premier et neuvièmes déciles.

On dessine :

- Une boîte centrale, de hauteur arbitraire, mais dont les extrémités sont $Q1$ et $Q3$.
- Une barre verticale à la médiane.
- Deux “moustaches” reliant $Q1$ à $D1$ et $Q3$ à $D9$ respectivement (premier quartile et premier décile, dernier quartile et dernier décile).

Remarques : On peut aussi dessiner des boîtes à moustaches verticalement. Dans certains cas, les moustaches s'arrêtent non pas aux déciles extrêmes, mais aux centiles extrêmes, voire aux valeurs extrêmes.

Exemple : toujours pour l'exemple de la taille :



COURS 12 : PARAMÈTRES DE DISPERSION

Étant donnée une variable statistique quantitative, nous définissons deux **paramètres de dispersion**, c'est-à-dire des valeurs mesurant l'étalement de la distribution de la variable (ou l'écart entre les valeurs prises par cette variable et la moyenne) :

- l'écart-type ;
- l'écart inter-quartiles.

ÉCART-TYPE

Supposons que l'on dispose d'une série statistique brute (x_1, \dots, x_N) , où l'effectif total est N . La **variance** σ^2 est définie comme la **moyenne des carrés des écarts à la moyenne** :

$$\sigma^2 := \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N},$$

où \bar{x} est la moyenne.

En général, avec des variables discrètes, on ne travaille pas avec la série brute mais avec la série triée par classes : pour chaque modalité i , on calcule l'effectif n_i ou la fréquence $f_i = n_i/N$. La variance est alors :

$$\sigma^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N} = f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_p(x_p - \bar{x})^2.$$

Dans le cas de variables continues triées par classes, on ne dispose pas des valeurs exactes x_i . La méthode la plus simple consiste à utiliser pour chaque classe la valeur centrale de la classe, c'est-à-dire la moyenne de ses extrémités¹.

Exemple : considérons l'exemple, utilisé les semaines passées, de la taille dans une promotion d'étudiants :

Taille (en cm)	[120,150[[150,160[[160,170[[170,180[[180,190[[190, 210[
Centre (en cm)	135	155	165	175	185	200
Fréquence (en %)	1,7	12,8	37,8	36,0	9,3	2,3

Comme calculé en semaine 10, la taille moyenne vaut alors 169,31 centimètres. Calculons la variance de la taille :

Taille (en cm)	[120,150[[150,160[[160,170[[170,180[[180,190[[190, 210[
Centre (en cm)	135	155	165	175	185	200
Écart à la moyenne (en cm)	135 – 169,31 = –34,31	–14,31	–4,31	5,69	15,69	30,69
Carré de l'écart à la moyenne (en cm^2)	$(-34,31)^2 = 1177,18$	204,78	18,58	32,38	246,18	941,88
Fréquence (en %)	1,7	12,8	37,8	36,0	9,3	2,3

1. Il faudrait en toute rigueur rajouter un terme correctif pour prendre en compte la dispersion à l'intérieur de chaque classe (voir le polycopié de cours), mais s'il y a suffisamment de classes, ce terme correctif sera petit. Pour la suite du cours, nous ignorerons ce terme supplémentaire.

La variance est la moyenne des carrés des écarts à la moyenne :

$$\begin{aligned}\sigma^2 &= 0,017 \times 1177,18 + 0,128 \times 204,78 + 0,378 \times 18,58 + 0,36 \times 32,38 + 0,093 \times 246,18 + 0,023 \times 941,88 \\ &= 109,46 \text{ centimètres carrés.}\end{aligned}$$

Remarque : Si les observations ont une unité u , alors la variance a l'unité u^2 . Par exemple, la variance est en centimètres carrés si on mesure des tailles en centimètres, en années carrées si l'on mesure des âges... La variance n'a pas d'interprétation simple, et n'est pas directement comparable aux données et aux autres paramètres (moyenne, médiane, etc.).

L'écart-type σ est définie comme la **racine carrée de la variance** :

$$\sigma = \sqrt{\sigma^2}.$$

L'écart-type, et non la variance, est un paramètre de dispersion que l'on peut comparer aux données.

Exemple : en gardant l'exemple de la taille dans une promotion d'étudiants : l'écart-type vaut ² :

$$\sigma = \sqrt{109,46} = 10,5 \text{ centimètres.}$$

Les tailles des étudiants sont donc typiquement à une distance d'environ 10 centimètres de la taille moyenne.

Remarque : La définition de l'écart-type (racine carrée de la moyenne des carrés des écarts à la moyenne) peut paraître inutilement compliquée. Elle est en fait assez naturelle, et a de bonnes propriétés mathématiques. On peut la comparer à la distance euclidienne usuelle : si l'on considère deux points A et B dans l'espace, de coordonnées $(1, 2, 3)$ et $(2, 2, 2)$ respectivement dans un repère orthonormé, alors la distance entre A et B vaut, d'après le théorème de Pythagore,

$$\sqrt{(1-2)^2 + (2-2)^2 + (3-2)^2} = \sqrt{1+0+1} = \sqrt{2},$$

une formule que l'on pourrait décrire comme la racine carrée de la somme des carrés des écarts des coordonnées.

ÉCART INTER-QUARTILES

L'écart inter-quartiles est simplement la différence entre les troisième et premier quartiles $Q_3 - Q_1$.

Exemple : en gardant l'exemple de la taille dans une promotion d'étudiants : le premier quartile est à 160 centimètres et le troisième quartile à 175 centimètres. L'écart inter-quartile est donc de $175 - 160 = 15$ centimètres.

Remarque : L'écart-type et l'écart-interquartiles ne sont pas directement comparables³. Il est plus pertinent de comparer les écarts-types (ou les écarts inter-quartiles) d'observations de la même variable dans des populations différentes.

2. Le terme correctif précédemment évoqué ferait passer la variance de 109,46 centimètres carrés à 119,50 centimètres carrés, et donc l'écart-type de 10,5 centimètres à 10,9 centimètres – une différence très faible.

3. Assez souvent, et en particulier pour les distributions gaussiennes, l'écart inter-quartile vaut environ 1,5 fois l'écart-type. C'est le cas de l'exemple.