Proportions et pourcentages

I. Proportions et pourcentages

I.1 Généralités

Soit E un ensemble non vide et A une partie de cet ensemble. On note n_{E} le nombre d'éléments (ou d'individus) de E et n_{A} le nombre d'éléments de A.

On appelle **proportion de A dans E** le quotient $p=\frac{n_{\mathrm{A}}}{n_{\mathrm{E}}}$.

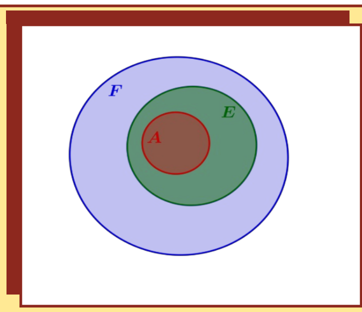
- Pour tout ensemble A contenu dans un ensemble non vide E, on a $\mathbf{0 \leqslant p \leqslant 1}$.
- Pour obtenir un résultat en pourcentage, il suffit de **multiplier la proportion obtenue par 100**.

À partir de la première propriété, il est possible de déterminer chacun des paramètres en fonction des deux autres :

- Si on connaît n_{A} et n_{E}, on peut déterminer p : $p=\frac{n_{\mathrm{A}}}{n_{\mathrm{E}}}$.
- Si on connaît p (où $p\neq 0$) et n_{A}, on peut déterminer n_{E} : $n_{\mathrm{E}}=\frac{n_{\mathrm{A}}}{p}$.
- Si on connaît p et n_{E}, on peut déterminer n_{A} : $n_{\mathrm{A}}=p\times n_{\mathrm{E}}$.

I.2 Pourcentage de pourcentage

Soit F un ensemble non vide, E une partie non vide de F et A une partie non vide de E.
On note p_1 la proportion de A dans E et p_2 la proportion de E dans F.
Alors la **proportion de A dans F** est $p=p_1\times p_2$.



II. Taux d'évolution

II.1 Variation absolue et variation relative

Soit une grandeur ayant pour valeur initiale $\mathrm{V_I}$ et pour valeur finale $\mathrm{V_F}$.

- La **variation absolue $\Delta \mathrm{V}$** est la différence entre $\mathrm{V_F}$ et $\mathrm{V_I}$. On a $\Delta\mathrm{V}=\mathrm{V_F}-\mathrm{V_I}$.
- Le **taux d'évolution (ou variation relative) t** est le quotient de la différence entre $\mathrm{V_F}$ et $\mathrm{V_I}$ par $\mathrm{V_I}$. On a $t=\frac{\mathrm{V_F}-\mathrm{V_I}}{\mathrm{V_I}}$.

Pour obtenir un **pourcentage d'évolution**, il suffit de **multiplier le taux d'évolution obtenu par 100**.

II.2 Coefficient multiplicateur

• Soit t le taux d'évolution qui permet à une quantité de passer de V_I à V_F.
On a alors : $V_F=(1+t)\times V_I$.
• $1+t$ est appelé **coefficient multiplicateur** associé au taux d'évolution t. On le note CM.
Avec ces notations, on a alors $V_F=CM\times V_I$ et $CM=1+t$.

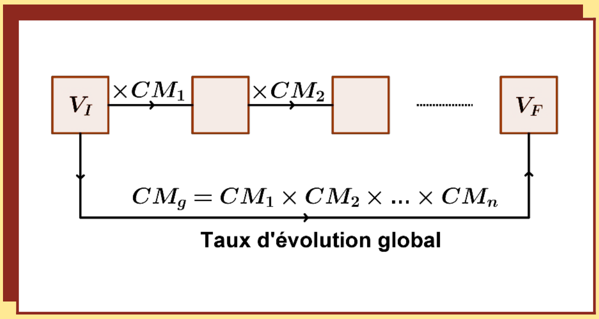
• Dans le cas d'une **augmentation**, t est **positif** et CM est un réel **supérieur à 1**.
• Dans le cas d'une **diminution**, t est **négatif** et CM est **compris entre 0 et 1**.

III. Évolutions successives et réciproques

III.1 Évolutions successives

Lorsqu'une quantité subit des **évolutions successives** t_1, t_2, …, t_n de sa valeur, elle subit alors une **évolution globale** t_g.

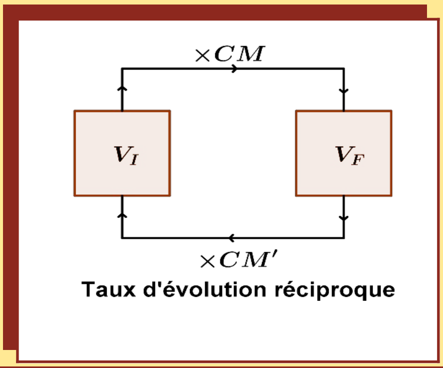
Le **coefficient multiplicateur global** CM_g associé à l'évolution t_g est le produit des coefficients multiplicateurs CM_1, CM_2, …, CM_n associés respectivement aux évolutions t_1, t_2, …, t_n.
On a $CM_g=CM_1\times CM_2\times\ldots\times CM_n$.



III.2 Évolution réciproque

Soit t le taux d'évolution d'une valeur passant de V_I à V_F. Alors son **taux d'évolution réciproque** t' est le taux permettant de passer de V_F à V_I.

Le **coefficient multiplicateur réciproque** CM' associé à l'évolution réciproque t' est l'inverse du coefficient multiplicateur non nul CM associé à l'évolution de départ t.
On a $CM'=\dfrac{1}{CM}$.



Statistiques descriptives

I. Présentation d'une série statistique

• La **population** d'une série statistique est l'ensemble des éléments appelés **individus** sur lesquels porte l'étude statistique.
• Le caractère d'une série statistique est la propriété étudiée sur chaque individu. Il est dit :
- **qualitatif** lorsqu'il ne prend pas que des valeurs numériques.
- **quantitatif discret** lorsqu'il ne peut prendre qu'un nombre fini de valeurs numériques.
- **quantitatif continu** lorsqu'il peut prendre une infinité de valeurs numériques.

• L'**effectif total** est le nombre d'éléments au sein de la population étudiée. On le note N .
• L'**effectif** d'une valeur du caractère est le nombre d'individus de la population prenant cette valeur (nombre de fois où cette valeur apparaît dans la série.). On le note n_i .
• La **fréquence** d'une valeur du caractère est le quotient de l'effectif de cette valeur par l'effectif total. On la note f_i et on a $f_i=\frac{n_i}{\mathrm{N}}$ (on multiplie par 100 si on veut un résultat en pourcentage).
• Le **mode** (ou **classe modale**) de la série est la valeur (ou la classe) du caractère ayant le plus grand effectif.

On note x_i une valeur prise par un caractère quantitatif : ses valeurs sont numériques.
• L'**effectif cumulé croissant** (respectivement **décroissant**) de x_i est la somme des effectifs des valeurs inférieures (respectivement supérieures) ou égales à x_i . On le note ECC .
• La **fréquence cumulée croissante** (respectivement **décroissante**) de x_i est la somme des fréquences des valeurs inférieures (respectivement supérieures) ou égales à x_i . On la note FCC .

La somme de toutes les fréquences est toujours égale à 1.

II. Paramètres d'une série statistique

Variable x_i	x_1	x_2	...	x_p
Effectif n_i	n_1	n_2	...	n_p

II.1 Caractères de position

• La **moyenne** d'une série statistique est le nombre, noté $\overline{x}$, défini par :

$$\overline{x}=\frac{n_1x_1+n_2x_2+\ldots+n_px_p}{\mathrm{N}}=f_1x_1+f_2x_2+\ldots f_nx_n$$

• Soit a et b deux réels.
Si une série statistique a pour moyenne $\overline{x}$, alors la série $\{ax_i+b\}$ a pour moyenne $a\overline{x}+b$.

La **médiane** d'une série statistique est le nombre, noté Me, qui partage la série en deux groupes de même effectif (50% des valeurs en dessous et 50% au dessus).

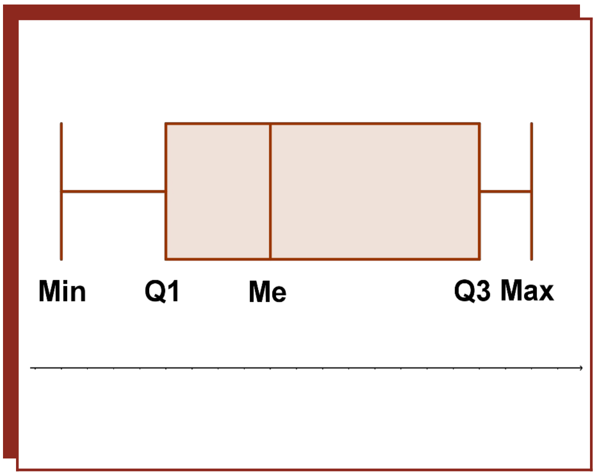
Les **quartiles** d'une série partagent celle-ci en quatre groupes de même effectif :
• Le **premier quartile** Q_1 est la plus petite valeur de la série telle que au moins un quart des valeurs sont inférieures ou égales à Q_1.
• Le **troisième quartile** Q_3 est la plus petite valeur de la série telle que au moins trois quarts des valeurs de la série sont inférieures ou égales à Q_3.

II.2 Caractères de dispersion

• L'**étendue** d'une série statistique est la différence entre la valeur maximale et la valeur minimale.
• L'**écart interquartile** est le nombre $Q_3 - Q_1$ et mesure la dispersion de la série par rapport à la médiane.
• L'**intervalle interquartile** $[Q_1 ; Q_3]$ contient 50% des valeurs de la série.
• L'**écart-type** d'une série statistique est le nombre réel positif, noté σ, défini par :

$$\sigma = \sqrt{\frac{n_1(x_1-\overline{x})^2 + n_2(x_2-\overline{x})^2 + \ldots + n_p(x_p-\overline{x})^2}{N}}$$

Représentation graphique d'une série
Pour résumer une série statistique sous forme de schéma, on utilisera ce que l'on appelle un diagramme en boîte (ou boîte à moustaches).
Un diagramme en boite représente graphiquement une série en faisant apparaître les principaux indicateurs de la série. On aura ainsi environ 25% des valeurs de la série entre Min et Q_1, environ 25% entre Q_1 et Me, environ 25% entre Me et Q_3 (donc environ 50% des valeurs dans la boite) et environ 25% des valeurs entre Q_3 et Max.



Méthode : Détermination de la médiane et des quartiles d'une série
Pour déterminer la médiane et les quartiles d'une série, on range tout d'abord les N valeurs de la série par ordre croissant (si ce n'est pas déjà fait).
Pour la médiane :
• Si N est un nombre impair, alors la médiane Me est la valeur située au rang $\frac{N+1}{2}$.
• Si N est pair, alors la médiane est la moyenne des valeurs de rangs $\frac{N}{2}$ et $\frac{N}{2}+1$.
Pour les quartiles :
• Q_1 est la valeur de la série de rang $\frac{N}{4}$ (ou l'entier directement supérieur si $\frac{N}{4} \notin \mathbb{N}$)
• Q_3 est la valeur de la série de rang $\frac{3N}{4}$ (ou l'entier directement supérieur si $\frac{3N}{4} \notin \mathbb{N}$)

Échantillonnage et estimation

I. Échantillonnage

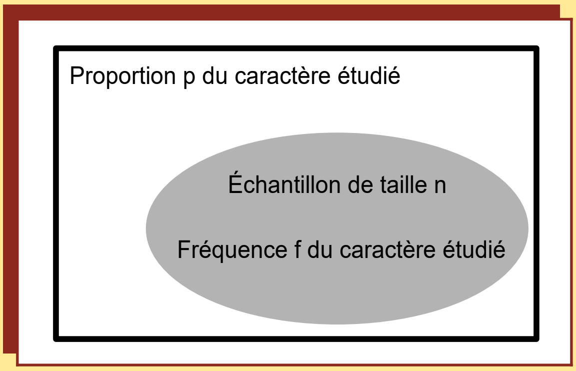
• Un échantillon de taille n est une sélection de n individus choisis au hasard dans une population.
• Sur plusieurs échantillons de même taille, la fréquence d'un caractère observée varie d'un échantillon à l'autre. C'est ce qu'on appelle la **fluctuation d'échantillonnage**.

II. Intervalle de fluctuation

Au sein d'une population, on connaît la proportion p d'un certain caractère.
On prélève un échantillon de taille n et on note f la fréquence du caractère dans l'échantillon.
Si $0{,}2 \leqslant p \leqslant 0{,}8$ et $n \geqslant 25$ alors, dans au moins 95% des cas, f appartient à l'intervalle :

$$\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]$$

Cette intervalle est appelé **intervalle de fluctuation au seuil de 95%**.



III. Intervalle de confiance

Au sein d'une population, la proportion p d'un certain caractère est inconnue.
On prélève un échantillon de taille n et on note f la fréquence du caractère dans l'échantillon.
Si $0{,}2 \leqslant f \leqslant 0{,}8$ et $n \geqslant 25$ alors, dans au moins 95% des cas, p appartient à l'intervalle :

$$\left[f-\frac{1}{\sqrt{n}}\,;f+\frac{1}{\sqrt{n}}\right]$$

Cet intervalle est appelé **intervalle de confiance**.

Méthode : utilisation d'un intervalle de confiance pour un sondage
Avant les élections, le candidat A commande un sondage effectué sur 250 personnes. 138 personnes interrogées déclarent avoir l’intention de voter pour le candidat A.
Le candidat A peut-il espérer être élu ?
Soit p la proportion théorique d’électeurs pour le candidat A.
La fréquence observée est égale à $f=\frac{138}{250}=0{,}552$.
L’intervalle de confiance de p au seuil de $0{,}95$ est $\left[0{,}552-\frac{1}{\sqrt{250}}\,;0{,}552+\frac{1}{\sqrt{250}}\right]$, soit de façon approchée $[0{,}48\,;0{,}62]$.
On a donc $0{,}48< p< 0{,}62$. Or, pour être sûr que le candidat A soit élu, il faudrait que $p> 0{,}5$.
Il est donc possible que le candidat A ne soit pas élu.