

## 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  $p = \frac{n_A}{n_E}$ .

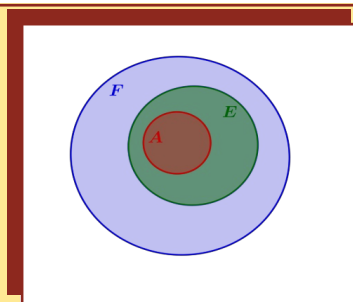
- Pour tout ensemble  $A$  contenu dans un ensemble non vide  $E$ , on a  $0 \leq p \leq 1$ .
- Pour obtenir un résultat en pourcentage, il suffit de **multiplier la proportion obtenue par 100**.

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  $p$  :  $p = \frac{n_A}{n_E}$ .
- Si on connaît  $p$  (où  $p \neq 0$ ) et  $n_A$ , on peut déterminer  $n_E$  :  $n_E = \frac{n_A}{p}$ .
- Si on connaît  $p$  et  $n_E$ , on peut déterminer  $n_A$  :  $n_A = p \times n_E$ .

#### 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  $V_I$  et pour valeur finale  $V_F$ .

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

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  $CM = 1 + t$ .

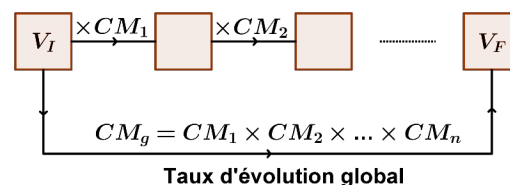
- Dans le cas d'une **augmentation**,  $t$  est **positif** et  $CM$  est un réel **supérieur à 1**.
- Dans le cas d'une **diminution**,  $t$  est **négatif** et  $CM$  est **compris entre 0 et 1**.

## III. Évolutions successives et réciproques

### III.1 Évolutions successives

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

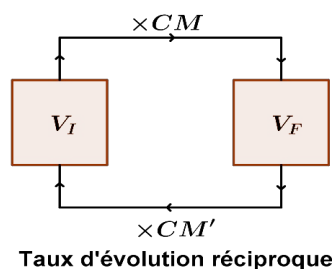
Le **coefficient multiplicateur global**  $CM_g$  associé à l'évolution  $t_g$  est le produit des coefficients multiplicateurs  $CM_1, CM_2, \dots, CM_n$  associés respectivement aux évolutions  $t_1, t_2, \dots, t_n$ .  
On a  $CM_g = CM_1 \times CM_2 \times \dots \times CM_n$ .



### 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  $CM' = \frac{1}{CM}$ .



## Statistiques à deux variables quantitatives

### I. Statistiques à une variable

- 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  $f_i = \frac{n_i}{N}$  (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.

#### Remarques

- Une fréquence est toujours un nombre compris entre 0 et 1 (entre 0 et 100% si on donne les fréquences en pourcentage).
- Une série statistique apparaîtra ainsi généralement sous la forme d'un tableau donnant les effectifs pour chacune des valeurs possibles du caractère, celles-ci étant classées par ordre croissant.

|                                  |       |       |     |       |
|----------------------------------|-------|-------|-----|-------|
| <b>Variable <math>x_i</math></b> | $x_1$ | $x_2$ | ... | $x_p$ |
| <b>Effectif <math>n_i</math></b> | $n_1$ | $n_2$ | ... | $n_p$ |

Effectif total :  $N = n_1 + n_2 + \dots + n_p$ .

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

### II. Croisement de deux variables catégorielles

Lorsque l'on s'intéresse à un sujet, il est fréquent de recueillir plusieurs données ou caractères. Il est parfois intéressant de ne pas regarder chaque caractère indépendamment, mais d'essayer d'établir une éventuelle corrélation entre ces deux caractères. C'est le rôle des statistiques.

#### II.1 Tableau croisé d'effectifs

Lorsqu'une série statistique étudie simultanément deux caractères  $X$  et  $Y$ , on dit qu'il s'agit d'une série statistique à deux variables (ou double) que l'on note  $(X; Y)$ .

Les valeurs prises par le caractère  $X$  sont notées  $(x_1; x_2; \dots; x_n)$ .

Les valeurs prises par le caractère  $Y$  sont notées  $(y_1; y_2; \dots; y_n)$ .

Un tableau croisé d'effectifs est un tableau dans lequel les valeurs d'un caractère sont présentées en ligne et celles de l'autre caractère sont présentées en colonne :

- On note  $n_{ij}$  l'effectif correspondant au couple  $(x_i; y_j)$  et  $N$  l'effectif total.
- À l'intersection d'une ligne et d'une colonne, on indique le nombre d'individus présentant simultanément la valeur du premier caractère correspondant à la ligne et la valeur du deuxième caractère correspondant à la colonne.
- L'**effectif marginal** d'une valeur d'un caractère est le nombre d'individus de la population présentant cette valeur du caractère.

| X \ Y        | $y_1$             | ... | $y_j$             | ... | $y_m$             | Total             |
|--------------|-------------------|-----|-------------------|-----|-------------------|-------------------|
| $x_1$        | $n_{11}$          |     | $n_{1j}$          |     | $n_{1m}$          | Effectif de $x_1$ |
| ...          |                   |     |                   |     |                   |                   |
| $x_i$        | $n_{i1}$          |     | $n_{ij}$          |     | $n_{im}$          | Effectif de $x_i$ |
| ...          |                   |     |                   |     |                   |                   |
| $x_k$        | $n_{k1}$          |     | $n_{ki}$          |     | $n_{km}$          | Effectif de $x_k$ |
| <b>Total</b> | Effectif de $y_1$ |     | Effectif de $y_i$ |     | Effectif de $y_m$ | $N$               |

## II.2 Tableau des fréquences par rapport à l'effectif global

Un tableau des fréquences par rapport à l'effectif global se présente de la même manière qu'un tableau croisé d'effectifs. Il permet d'étudier simultanément les fréquences de deux caractères avec l'un des caractères présenté en ligne et l'autre en colonne.

- On note  $f_{ij}$  la fréquence correspondant au couple  $(x_i; y_j)$  et on a  $f_{ij} = \frac{n_{ij}}{N}$ .
- La **fréquence marginale** d'une valeur d'un caractère est le quotient de l'effectif marginal de cette valeur par l'effectif total de la population.

| X \ Y        | $y_1$                       | ... | $y_j$                       | ... | $y_m$                       | Total              |
|--------------|-----------------------------|-----|-----------------------------|-----|-----------------------------|--------------------|
| $x_1$        | $f_{11} = \frac{n_{11}}{N}$ |     | $f_{1j} = \frac{n_{1j}}{N}$ |     | $f_{1m} = \frac{n_{1m}}{N}$ | Fréquence de $x_1$ |
| ...          |                             |     |                             |     |                             |                    |
| $x_i$        | $f_{i1} = \frac{n_{i1}}{N}$ |     | $f_{ij} = \frac{n_{ij}}{N}$ |     | $f_{im} = \frac{n_{im}}{N}$ | Fréquence de $x_i$ |
| ...          |                             |     |                             |     |                             |                    |
| $x_k$        | $f_{k1} = \frac{n_{k1}}{N}$ |     | $f_{kj} = \frac{n_{kj}}{N}$ |     | $f_{km} = \frac{n_{km}}{N}$ | Fréquence de $x_k$ |
| <b>Total</b> | Fréquence de $y_1$          |     | Fréquence de $y_i$          |     | Fréquence de $y_m$          | 1                  |

## II.3 Tableau des fréquences conditionnelles

Si on fixe une valeur de l'un des deux caractères, on obtient une série à une variable appelée série conditionnelle. On peut donc obtenir un tableau des effectifs de cette série à une variable correspondant à une colonne du tableau croisé d'effectifs.

Une **fréquence conditionnelle** se calcule donc par rapport au caractère fixé, c'est-à-dire suivant la ligne ou la colonne du tableau croisé d'effectifs.

On repart du tableau croisé d'effectif puis on fixe  $x_i$ . On obtient donc le tableau des effectifs de la variable  $Y$ .

| $Y$               | $y_1$    | ... | $y_j$    | ... | $y_m$    | Total                                          |
|-------------------|----------|-----|----------|-----|----------|------------------------------------------------|
| Effectifs des $Y$ | $n_{il}$ |     | $n_{ij}$ |     | $n_{im}$ | $n = \text{effectif des } Y \text{ pour } x_i$ |

On peut donc construire le tableau des fréquences conditionnelles du caractère  $Y$  par rapport à  $x_i$ .

| $Y$                | $y_1$    | ... | $y_j$    | ... | $y_m$    | Total |
|--------------------|----------|-----|----------|-----|----------|-------|
| Fréquences des $Y$ | $f_{il}$ |     | $f_{ij}$ |     | $f_{im}$ | 1     |

## III. Point moyen et forme d'un nuage de points

### III.1 Point moyen

Dans cette partie, on représentera une série statistique à deux variables à l'aide d'un tableau de la forme :

|                |       |       |     |       |
|----------------|-------|-------|-----|-------|
| Variable $x_i$ | $x_1$ | $x_2$ | ... | $x_n$ |
| Variable $y_i$ | $y_1$ | $y_2$ | ... | $y_n$ |

L'ensemble des points  $M_i(x_i; y_i)$ , pour  $i$  allant de 1 à  $n$ , est appelé le nuage de points de cette série statistique dans un repère orthogonal.

Soit  $\bar{x}$  la moyenne des valeurs prises par le 1er caractère et  $\bar{y}$  celle des valeurs prises par le 2e caractère. On appelle **point moyen du nuage de points** de la série statistique le point **G** de coordonnées  $(\bar{x}; \bar{y})$  tel que  $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$  et  $\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}$ .

### III.2 Forme d'un nuage de points

La forme d'un nuage de points peut indiquer ou non une liaison entre les deux variables  $X$  et  $Y$ . L'allure du nuage de points représentant la série statistique permet d'avoir une idée de la réponse. En effet, la disposition des points peut parfois suggérer la courbe d'une fonction. Cela signifie alors que le lien approximatif entre  $X$  et  $Y$  est de la forme  $y = f(x)$ .

Parmi les tendances les plus communes, on trouvera :

- les points du nuage sont dispersés, ils n'indiquent pas de tendance particulière ;
- les points du nuage sont sensiblement alignés, alors il existe une **liaison affine** entre les deux variables  $X$  et  $Y$  ;
- les points du nuage sont sensiblement sur une courbe exponentielle, alors il existe une **liaison exponentielle** entre les variables  $X$  et  $Y$ .

#### IV. Ajustement affine, estimations et prévisions

Lorsqu'un nuage de points  $M_i(x_i; y_i)$  met en évidence une liaison affine entre les deux variables étudiées, on peut la représenter par une droite d'équation  $y = ax + b$ . Cette droite est appelée **droite d'ajustement de  $y$  en  $x$**  de la série statistique ou **droite de régression linéaire de  $y$  en  $x$** . Cependant, si la forme d'un nuage de points  $M_i(x_i; y_i)$  suggère un ajustement affine, on peut trouver un grand nombre de droites ajustant ce nuage.

##### IV.1 Ajustement affine graphique

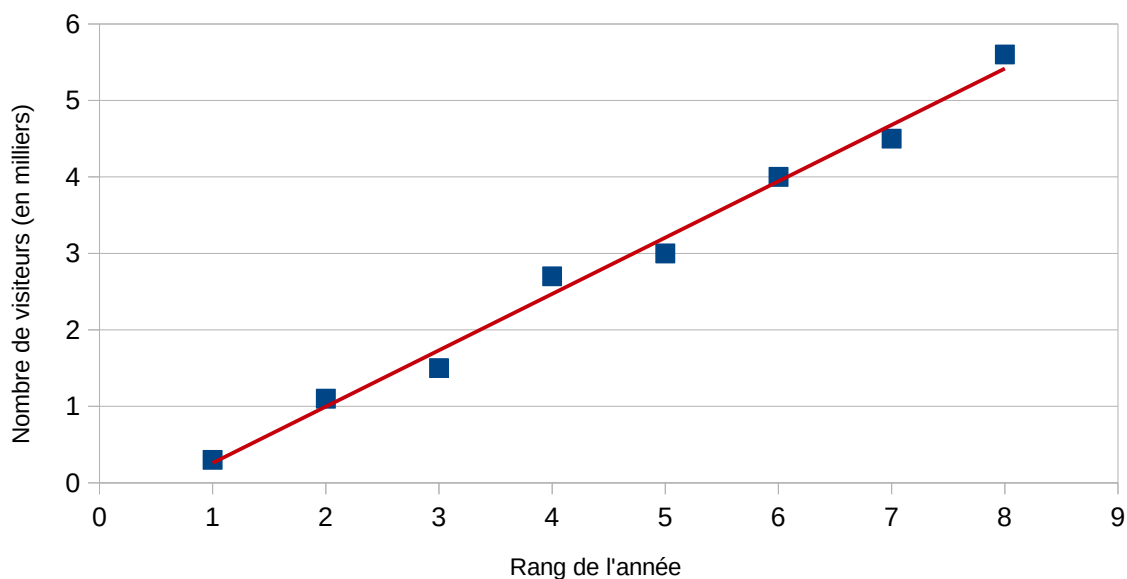
Dans le plan rapporté à un repère, on trace une droite passant au plus près de tous les points du nuage en suivant les indications fournies par l'énoncé.

##### Exemple

On a relevé, pour un site internet, le nombre de visiteurs (en milliers) par an pendant les huit premières années de fonctionnement :

| Année                                      | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|--------------------------------------------|------|------|------|------|------|------|------|------|
| Rang $x_i$                                 | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    |
| Nombre de visiteurs $y_i$<br>(en milliers) | 0,3  | 1,1  | 1,5  | 2,7  | 3    | 4    | 4,5  | 5,6  |

Dans cet exemple, les points du nuage sont sensiblement alignés. Traçons une droite d'ajustement :

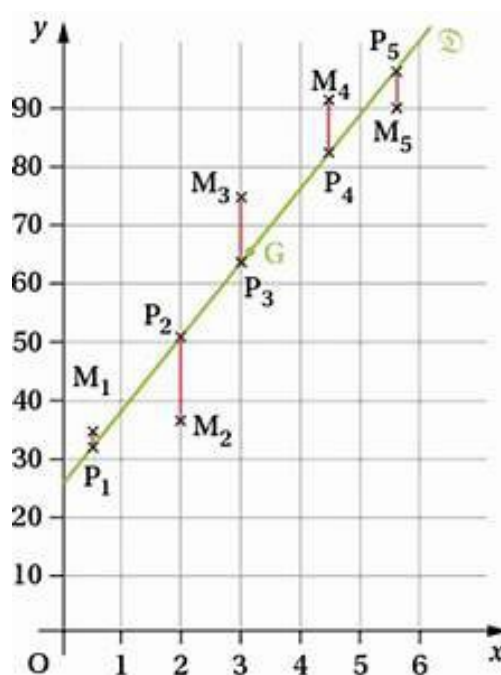


### Remarque

Par cette méthode, on effectue un tracé « au jugé ». Nous n'obtiendrons donc pas à chaque fois exactement la même droite.

### IV.2 Ajustement affine par la méthode des moindres carrés

Cette méthode consiste à déterminer une droite d'ajustement  $D$ , d'équation  $y = ax + b$ , telle que la somme des carrés des écarts entre les points de même abscisse  $M_i(x_i; y_i)$  du nuage de points et  $P_i(x_i; y'_i)$  de la droite soient minimale.



**Remarque**

Pour obtenir l'équation de la droite d'ajustement affine, on utilisera la calculatrice ou un tableur. Dans l'exemple précédent, on obtiendrait la droite d'équation  $y = 0,74x - 0,48$  (en arrondissant à  $10^{-2}$  près).

La droite d'ajustement affine de  $y$  en  $x$  par la méthode des moindres carrés **passé par le point moyen** du nuage de points de la série statistique.

**IV.3 Estimations et prévisions**

En supposant que la tendance observée se maintienne, la droite d'ajustement permet de faire des estimations. Ces estimations s'effectuent soit par lecture graphique (à partir du tracé de la droite d'ajustement), soit par le calcul (en utilisant l'équation de la droite d'ajustement).

Lorsque la variable  $x$  est une mesure du temps, la série est appelée série chronologique. Il s'agit alors de se projeter dans le futur et de faire des prévisions.

**Exemple**

Reprenons l'exemple précédent.

Supposons que les statistiques de l'année numéro 5 aient été perdues. Nous pouvons estimer alors la fréquentation du site grâce à l'ajustement affine que nous avons fait.

On a  $y = 0,74x - 0,48$  avec  $x = 5$ . On obtient alors  $y = 5 \times 0,74 - 0,48 = 3,22$ . Nous estimons le nombre de visiteurs à 3 220 la cinquième année. La valeur cherchée était à l'intérieur du nuage de points, on dit qu'on a fait une **interpolation**. Il s'agit ici d'une **estimation**.

Toujours dans cet exemple, nous voulons estimer la fréquentation du site au bout de 20 ans de fonctionnement, en considérant que l'évolution reste la même.

On a  $y = 0,74x - 0,48$  avec  $x = 20$ . On obtient alors  $y = 20 \times 0,74 - 0,48 = 14,32$ . Nous estimons le nombre de visiteurs à 14 320 la vingtième année. La valeur cherchée était au-delà du nuage de points, on dit qu'on a fait une **extrapolation**. Il s'agit ici d'une **prévision**.