MATHÉMATIQUES – TOUTES SÉRIES
Statistiques

Note liminaire

Programme selon les sections :
- pourcentages : toutes sections
- étude d'une série statistique : S – ES/L – STMG – STL – hôtellerie
- nuage de points : ST2S – STMG – STL – hôtellerie
- ajustement affine : STMG – STL

Prérequis

Série statistique – fréquence – effectif – fréquences cumulées croissantes – effectifs cumulés croissants

Plan du cours

1. Pourcentages

Définition :

- Un **pourcentage** est une fraction dont le dénominateur est 100.

Taux d'évolution :

Le **taux d'évolution** est la valeur en pourcentage d'une augmentation ou d'une réduction.

$$\frac{t}{100} = \frac{V_2 - V_1}{V_1}$$

Coefficient multiplicateur :

Le **coefficient multiplicateur** CM correspond au facteur par lequel il faut multiplier la valeur V_1 pour obtenir V_2, nouvelle valeur réduite ou augmentée de t%.

MATHÉMATIQUES – TOUTES SÉRIES
Statistiques

$\boldsymbol{V_2 = CM \times V_1}$
$\boldsymbol{CM = 1 \pm \frac{t}{100}}$ (+ si augmentation, – si réduction)

Evolution réciproque :

Si $CM = \frac{V_2}{V_1}$ permet de passer de V_1 à V_2, alors CM' permet de passer de V_2 à V_1.

$$\boldsymbol{CM' = \frac{1}{CM}}$$

$CM' = \boldsymbol{1 \pm \frac{t'}{100}}$ d'où $\boldsymbol{t' = (CM' \pm 1) \times 100 = \left(\frac{1}{CM} \pm 1\right) \times 100}$

Evolutions successives :

Si CM_1 permet de passer de V_0 à V_1, CM_2 permet de passer de V_1 à V_2, et ainsi de suite jusqu'à V_n, alors :

$\boldsymbol{V_n = CM_{global} \times V_0}$ **avec** $\boldsymbol{CM_{global} = CM_1 \times CM_2 \times \ldots \times CM_n}$

Indice :

Les indices sont utilisées pour connaître le rapport entre une valeur à une date et la valeur à la date de référence.

$$I_n = \frac{V_n}{V_{ref}} \times 100$$

Ex : Soit le tableau suivant :

Année	2012	2013	2014	2015
Loyer moyen d'un T2 (€)	910	900	930	940

Si 2012 est l'année de référence, les indices correspondant sont :

Année	2012	2013	2014	2015
Indice	100	99	102	103

MATHÉMATIQUES – TOUTES SÉRIES
Statistiques

2. Moyenne et écart-type

Moyenne :

La moyenne d'une série statistique est le **quotient de la somme de ses valeurs et de son effectif total.**

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Si à chaque valeur V_i est associé un effectif n_i on a (**moyenne pondérée**) :

$$\overline{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \cdots + n_p \times x_p}{n_1 + n_2 + \cdots + n_p} = \frac{\sum_{i=1}^{p} n_i \times x_i}{\sum_{i=1}^{p} n_i}$$

Si on connaît les **moyennes partielles** de deux sous-groupes, on peut également calculer la moyenne :

$$\overline{x} = \frac{n_a \times \overline{x_a} + n_b \times \overline{x_b}}{n_a + n_b}$$

Cette formule se généralise lorsqu'il y a plus de deux sous-groupes.

Variance :

Soit une série de valeurs V_i associées à des effectifs n_i. La **variance** est :

$$V = \frac{n_1(V_1 - \overline{x})^2 + n_2(V_2 - \overline{x})^2 + \cdots + n_p\left(V_p - \overline{x}\right)^2}{n_1 + n_2 + \cdots + n_p} = \frac{\sum_{i=1}^{p} n_i(V_i - \overline{x})^2}{\sum_{i=1}^{p} n_i}$$

Soit également :

$V = \frac{\sum_{i=1}^{p} n_i \times {V_i}^2}{\sum_{i=1}^{p} n_i} - \overline{x}^2$ *(« la moyenne des carrés moins le carré de la moyenne »)*

Ecart-type :

L'écart-type est la racine carrée de la variance :

$$\sigma = \sqrt{V}$$

MATHÉMATIQUES – TOUTES SÉRIES
Statistiques

L'écart-type correspond à l'écart moyen entre une valeur et la moyenne.

Exemple :

Voici les notes correspondant à un groupe d'élèves :

notes	7	9	10	12	14	18
effectif	2	2	2	1	4	1

Moyenne :
$$\bar{x} = \frac{2 \times 7 + 2 \times 9 + 2 \times 10 + 1 \times 12 + 4 \times 14 + 1 \times 18}{12}$$
$$\bar{x} = 11{,}5$$

Ecart-type :
$$\sigma = \sqrt{V} = \sqrt{\frac{2\times(7-11{,}5)^2+2\times(9-11{,}5)^2+2\times(10-11{,}5)^2+1\times(12-11{,}5)^2+4\times(14-11{,}5)^2+1\times(18-11{,}5)^2}{12}}$$
$$\sigma \approx 3{,}23$$

3. Médiane et écart interquartile

Etendue :

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

Médiane :

La **médiane** est la valeur qui **sépare la série en deux groupes de même effectif** : l'un ayant toutes ses valeurs en-dessous de la médiane, l'autre ayant toutes ses valeurs au-dessus.

- **S'il y a $2n+1$ valeurs** dans la série (effectif total impair) : la médiane est x_{n+1} **($n+1^{\text{ème}}$ valeur)**
- **S'il y a $2n$ valeurs** dans la série (effectif total pair) : la médiane est la **moyenne des $n^{\text{ème}}$ et $n+1^{\text{ème}}$ valeurs soit $\frac{x_n+x_{n+1}}{2}$**

MATHÉMATIQUES – TOUTES SÉRIES
Statistiques

<u>Premier quartile :</u>

Le **premier quartile** Q_1 est la **plus petite valeur de la série** telle qu'**au moins** $\frac{1}{4}$ des valeurs de la série (25%) sont **inférieures ou égales à** $\boldsymbol{Q_1}$.

<u>Troisième quartile :</u>

Le **troisième quartile** Q_3 est la **plus petite valeur de la série** telle qu'**au moins** $\frac{3}{4}$ des valeurs de la série (75%) sont **inférieures ou égales à** $\boldsymbol{Q_3}$.

<u>Ecart interquartile :</u>

L'écart interquartile est la différence entre les premier et troisième quartiles.

$$\boldsymbol{EI = Q_3 - Q_1}$$

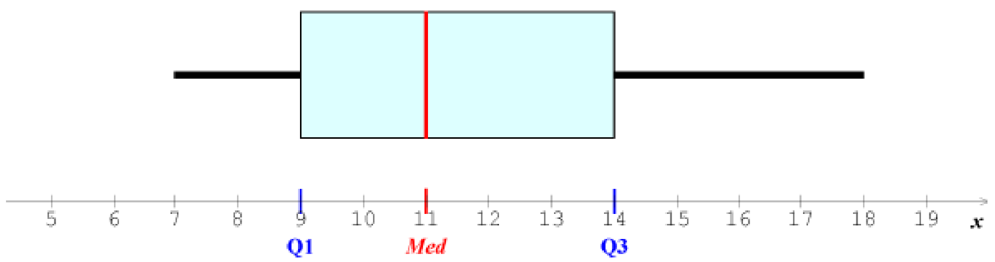
<u>Diagramme en boîtes :</u>

Un **diagramme en boîtes** (ou « boîte à moustaches ») est un diagramme permettant de visualiser **médiane, quartiles, écart interquartile et étendue de la série.**
Les valeurs sont reportées sur un axe horizontal, les quartiles sont représentés par des rectangles.

<u>Exemple :</u>

notes	7	9	10	12	14	18
effectif	2	2	2	1	4	1

L'étendue est 11.
L'effectif total étant 12, la médiane correspond à la moyenne entre la 6ème et la 7ème valeur.
$M = \frac{10+12}{2} = 11$ $\quad Q_1 = 9$ $\quad Q_3 = 14$ $\quad EI = 14 - 9 = 5$



MATHÉMATIQUES – TOUTES SÉRIES
Statistiques

La largeur de la boîte n'a pas d'importance.

Remarques :

- **Moyenne et médiane** sont des **paramètres de position**. Elles permettent de situer la série vers une certaine valeur.

- **Ecart-type et écart interquartile** sont des **paramètres de dispersion**. Ils permettent d'évaluer si l'ensemble des valeurs de la série est proche ou non de la valeur moyenne ou médiane. La comparaison des diagrammes en boîtes de deux séries permet de comparer à vue d'œil leur dispersion autour de la valeur médiane.

MATHÉMATIQUES – TOUTES SÉRIES
Statistiques

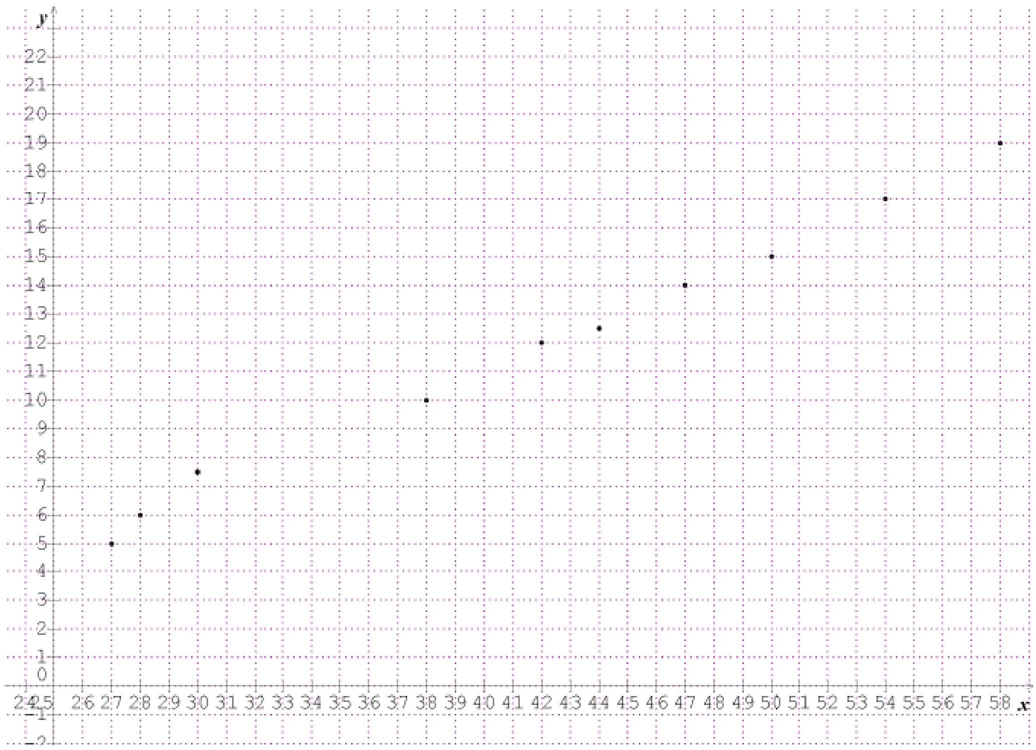
4. Ajustement affine

Nuage de points :

Un **nuage de points** est la **représentation graphique d'une série statistique à deux variables.** On met en abscisse l'une des variables, et en ordonnée l'autre variable. On utilise cette représentation quand on veut montrer la corrélation entre deux variables, en particulier la proportionnalité.

Ex : Relevés sur un groupe de 10 personnes de l'âge des membres du groupe et de leur distance minimale de lecture

Âge	27	28	30	38	42	44	47	50	54	58
Distance minimale de lecture (cm)	5	6	7,5	10	12	12,5	14	15	17	19

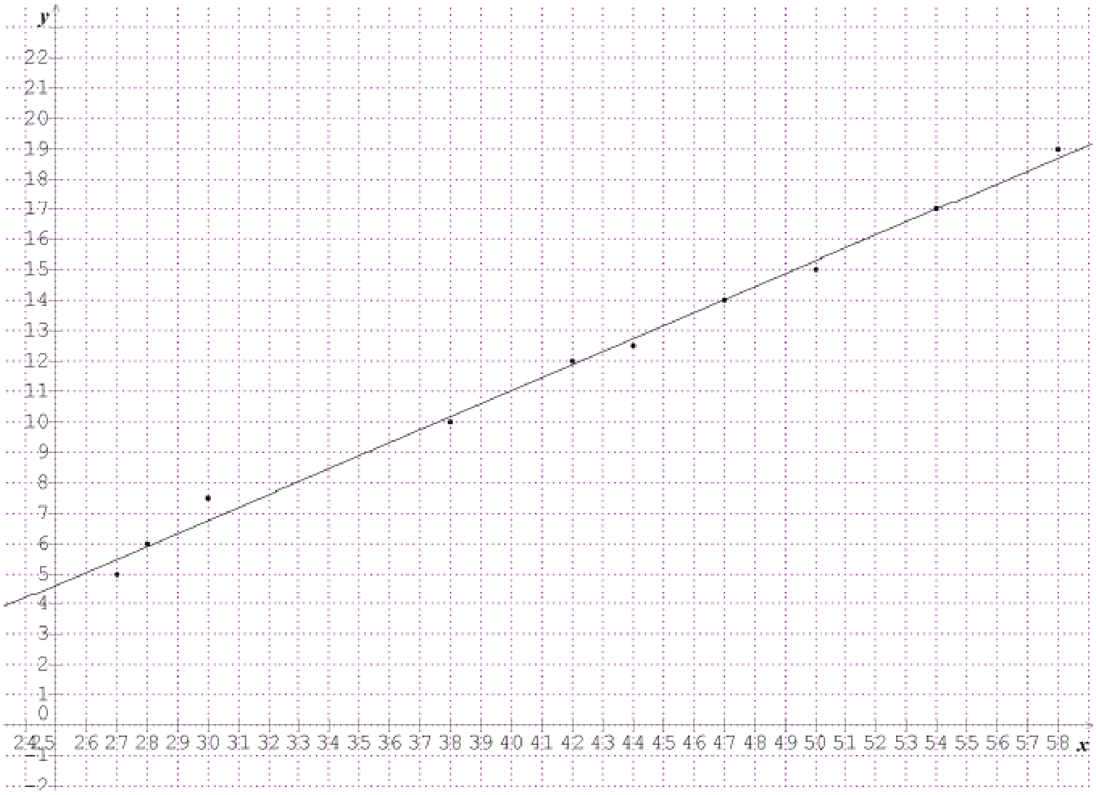


MATHÉMATIQUES – TOUTES SÉRIES
Statistiques

Ajustement affine :

L'ajustement affine consiste à tracer dans le nuage de points **une droite au plus près des différents points.**

Ex :



S'il est possible de tracer une droite, il y a alors une relation de proportionnalité entre les deux variables.

MATHÉMATIQUES – TOUTES SÉRIES
Statistiques

Méthode des moindres carrés :

On peut tracer cette droite à l'œil nu, ou appliquer une méthode de calcul pour l'ajustement affine. **La méthode des moindres carrés** en est une.

La droite tracée a pour équation $y = ax + b$

$a = \dfrac{cov(x, y)}{V(x)}$

avec $cov(x, y) = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x} \times \bar{y}$ et $V(x)$ variance de x

$b = \bar{y} - a\bar{x}$

soit : $\boldsymbol{y = \frac{cov(x,y)}{V(x)} (x - \bar{x}) + \bar{y}}$

Dans l'exemple, précédent, on a :

$y = 0{,}43x - 6{,}02$

