

Statistiques à une ou deux variables, cours, terminale STMG

F.Gaudon

8 juillet 2015

Table des matières

1	Statistiques à une variable (rappels)	2
2	Statistiques à deux variables	3
2.1	Vocabulaire	3
2.2	Ajustement d'un nuage de points	4
2.3	Détermination d'une équation de droite d'ajustement affine	4

1 Statistiques à une variable (rappels)

Propriété (couple moyenne/écart type) :

Soient x_i pour i allant de 1 à p où p est un entier les valeurs distinctes d'une série statistique et n_i pour i allant de 1 à p les effectifs correspondants. On note N l'effectif total, somme des n_i pour i allant de 1 à p . On définit :

- La *moyenne pondérée* notée \bar{x} :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n_1 + n_2 + \dots + n_p}$$

- L'*écart type* σ qui mesure la dispersion autour de la moyenne. On utilisera la calculatrice pour déterminer l'écart type.

Exemple :

On a relevé le prix de la baguette de pain dans différentes boulangeries :

Prix en euros (x_i)	0,82	0,83	0,84	0,85	0,86	0,87
Nombre de boulangeries (n_i)	10	42	85	23	18	2

La moyenne est donnée par :

$$\bar{x} = \frac{10 \times 0,82 + 42 \times 0,83 + 85 \times 0,84 + 23 \times 0,85 + 18 \times 0,86 + 2 \times 0,87}{10 + 42 + 85 + 23 + 18 + 2} \approx 0,84 \text{€}$$

L'écart type est 0,01 € environ ce qui signifie qu'il y a un faible écart entre les prix du pain dans les différentes boulangeries.

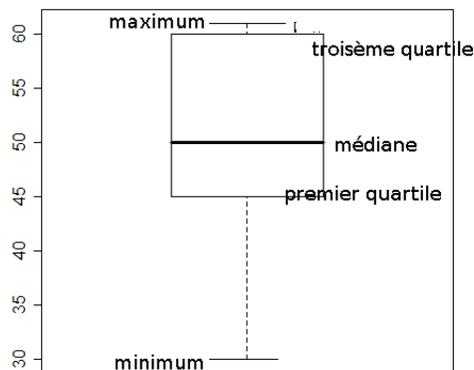
Définition (couple médiane/quartiles) :

- La médiane d'une série statistique est une valeur du caractère telle que la moitié des effectifs lui sont supérieurs ou égaux et la moitié des effectifs lui sont inférieurs ou égaux.
- Le premier quartile est la plus petite valeur pour laquelle au moins 25 % des valeurs lui sont inférieures ou égales ;
- le troisième quartile est la plus petite valeur pour laquelle au moins 75 % des valeurs lui sont inférieures ou égales.

Définition (diagramme en boîte) :

Le diagramme en boîte permet de résumer les différentes caractéristiques statistiques suivantes :

- Le maximum et le minimum ;
- le premier et le troisième quartile ;
- la médiane.



2 Statistiques à deux variables

2.1 Vocabulaire

Définition :

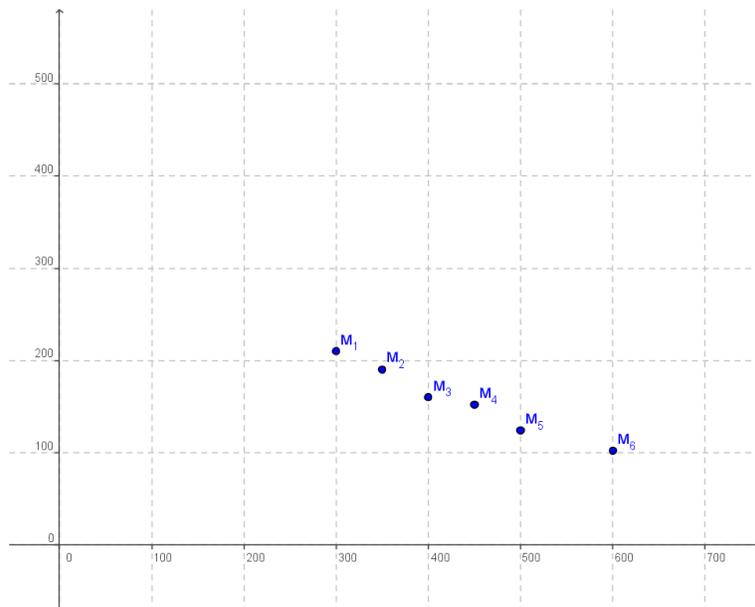
- Soient x et y deux caractères quantitatifs d'une même population. A chaque individu de la population on associe un couple $(x_i; y_i)$ où x_i et y_i pour $i \in \{1; \dots; n\}$ avec n entier naturel sont les valeurs prises respectivement par x et y . L'ensemble de ces couples constitue une *série statistique à deux variables* x et y .
- Dans un repère $(O; \vec{i}; \vec{j})$, l'ensemble des points M_i de coordonnées $(x_i; y_i)$ est appelé *nuage de points* associé à la série statistique.

Exemple :

Un magasin réalise une étude sur l'influence du prix de vente sur le nombre de machines à laver vendues au cours d'une année. Le tableau suivant donne les résultats de cette étude :

Prix x_i en euros	300	350	400	450	500	600
Nombre de machines vendues	210	190	160	152	124	102

Le nuage de points associé à cette série est constitué des points M_i pour i allant de 1 à 6 dont les coordonnées sont $(300; 210)$, $(350; 190)$, ..., $(600; 102)$.



2.2 Ajustement d'un nuage de points

Définition :

Toute droite "résumant approximativement" le nuage est appelée *droite d'ajustement affine* du nuage de points.

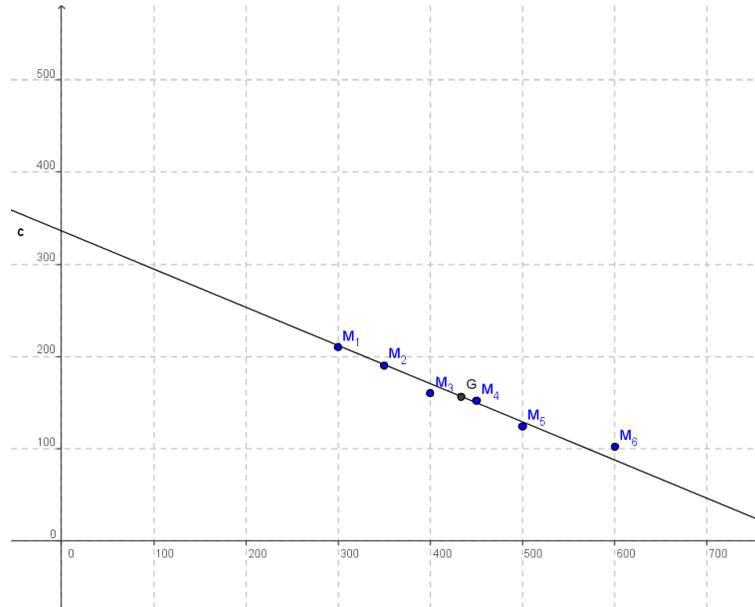
Remarque :

Il existe d'autres types d'ajustement : dans certains cas, on peut observer que visiblement une droite ne convient pas mais que le nuage de points semble être approché par un autre type de courbe, parabole par exemple. En outre, certains nuages peuvent ne pas sembler être approchables par une quelconque courbe auquel cas les deux variables ne sont pas reliées entre elles.

2.3 Détermination d'une équation de droite d'ajustement affine

Méthode graphique au jugé :

On trace « au jugé » une droite qui « semble résumer » le nuage de points. C'est une méthode simple mais qui dépend de la droite tracée.



Méthode des moindres carrés :

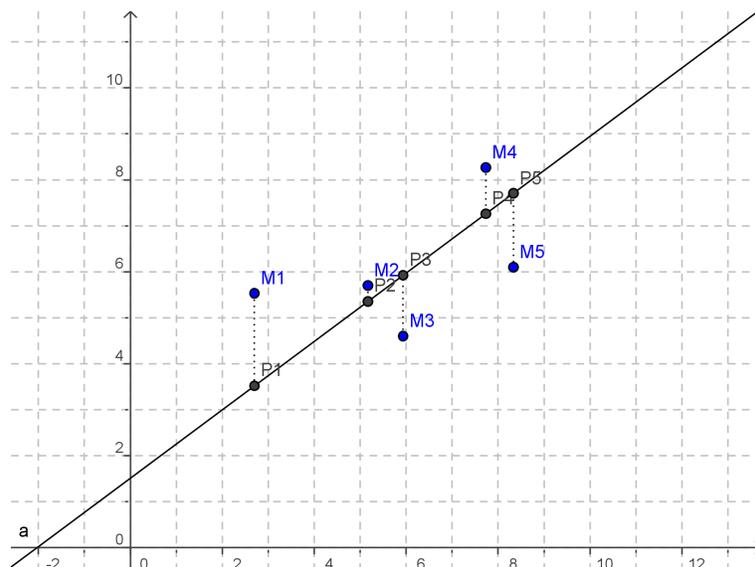
Avec les notations de la figure ci-dessous, étant donné un nuage de n points M_i , il existe une droite passant par le point moyen G et telle que la somme des carrés des écarts (ou *résidus*) $P_1M_1^2 + P_2M_2^2 + \dots + P_nM_n^2$ soit minimale. Cette droite est appelée *droite de régression de y en x* . On peut montrer que son équation réduite est $y = mx + p$ avec :

$$m = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_p - \bar{x})(y_p - \bar{y})}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_p - \bar{x})^2}$$

et

$$p = \bar{y} - m\bar{x}$$

En pratique, on utilisera la calculatrice pour l'obtenir.



Exemple :

On reprend l'exemple précédent.

- Recherche de l'équation réduite à l'aide des formules :

$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
	Total		

D'où $m \approx -0,37$

et $p \approx 315,06$

- Recherche de l'équation réduite avec la calculatrice :

* TI 82 et plus :

Aller dans le menu **STAT** puis **EDIT**. Entrer les valeurs x_i dans la colonne L_1 et les valeurs y_i dans la colonne L_2 . Quitter (**2nde** **QUIT**) puis menu **STAT** et **CALC**. Choisir **LinReg(ax+b)** puis **2nd** **L1** **,** **2nd** **L2** pour indiquer les deux colonnes à utiliser. Valider ensuite **ENTER**.

* CASIO Graph 25 et plus :

Aller dans le menu **STAT** puis entrer les valeurs x_i dans la colonne 1 et les valeurs y_i dans la colonne 2. Choisir ensuite **CALC** puis **REG** puis **X**.