

# Statistiques à une ou deux variables, cours, terminale STG

F.Gaudon

15 novembre 2009

## Table des matières

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

# 1 Statistiques à une variable (rappels)

## Propriété :

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  $p$ .

- *Moyenne  $\bar{x}$*  :

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

- *Variance  $V$*  :

$$V(x) = \frac{n_1x_1^2 + n_2x_2^2 + \dots + n_px_p^2}{n_1 + n_2 + \dots + n_p} - \bar{x}^2$$

- *Écart type  $\sigma$*  :

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

## 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.

# 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.
- Soit une série statistique à deux variables  $x$  et  $y$  de moyennes  $\bar{x}$  et  $\bar{y}$ . Le point  $G$  de coordonnées  $(\bar{x}; \bar{y})$  avec  $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$  et  $\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}$  est appelé le *point moyen* du 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 :

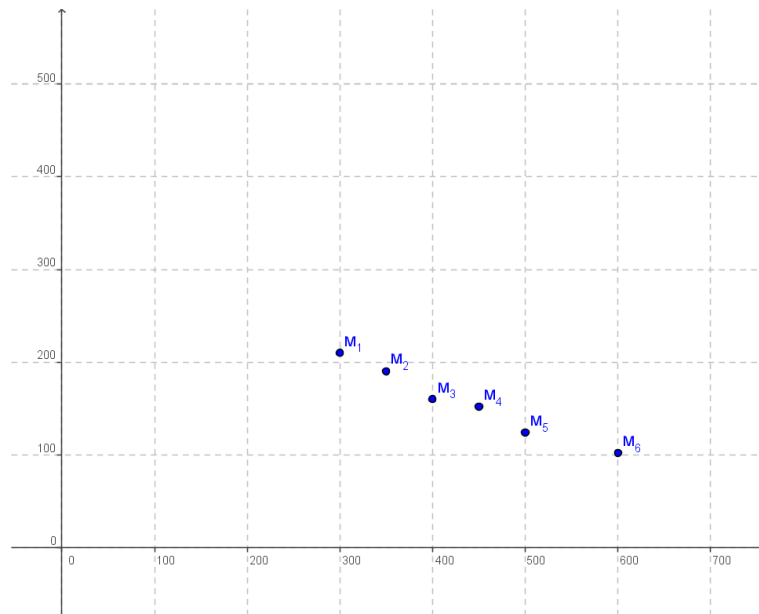
Prix $x_i$ en euros	300	350	400	450	500	600
Nombre de machines vendues	210	190	160	152	124	102

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  $(300; 210)$ ,  $(350; 190)$ , ...,  $(600; 102)$ .

Le point moyen associé à ce nuage de points est le point  $G$  de coordonnées  $(\bar{x}; \bar{y})$  données par :

$$\bar{x} = \frac{300+350+\dots+600}{6} = \frac{2285}{6} \approx 380,3$$

$$\bar{y} = \frac{210+190+\dots+102}{6} = \frac{938}{6} \approx 156,33$$



## 2.2 Ajustement d'un nuage de points

### Définition :

Toute droite passant par le point moyen du nuage et "résumant approximativement" le nuage est appelée *droite d'ajustement affine* du nuage de points.

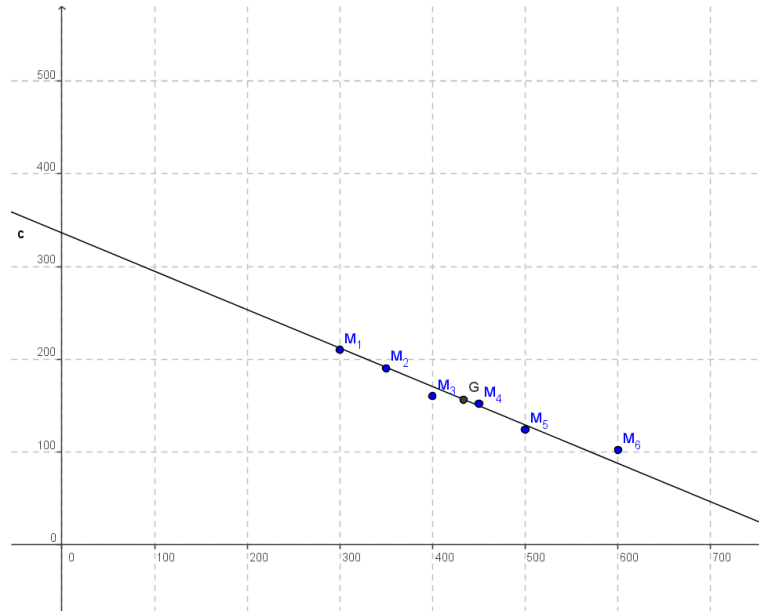
### Remarque :

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 passant par le point moyen du nuage qui « semble résumer » le nuage de points. C'est une méthode simple mais qui dépend de la droite tracée.



Méthode de MAYER :

On sépare le nuage en deux sous nuages et on calcule les coordonnées des points moyens des deux sous nuages. La droite de MAYER est la droite passant par ces deux points. On peut montrer qu'elle passe aussi par le point moyen du nuage.

**Exemple :**

Dans l'exemple précédent, on définit les deux sous nuages constitués des points \$M\_1, M\_2\$ et \$M\_3\$ pour le premier et des points \$M\_4, M\_5\$ et \$M\_6\$ pour le second nuage.

Le point moyen \$G\_1\$ du premier nuage a donc pour coordonnées :

$$\bar{x}_1 = \frac{300+350+400}{3} = \frac{1050}{3} = 350$$

$$\bar{y}_1 = \frac{210+190+160}{3} = \frac{560}{3} \approx 186,67$$

Le point moyen \$G\_2\$ du deuxième nuage a pour coordonnées :

$$\bar{x}_2 = \frac{450+500+600}{3} = \frac{1550}{3} \approx 516,67$$

$$\bar{y}_2 = \frac{152+124+102}{3} = \frac{378}{3} = 126$$

La droite de MAYER est alors la droite \$(G\_1G\_2)\$.

### Méthode des moindres carrés :

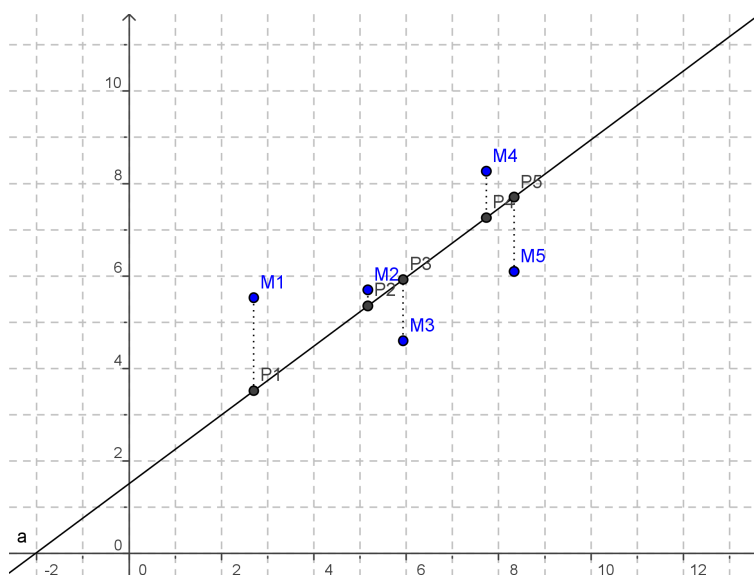
Avec les notations de la figure ci-dessous, étant donné un nuage de  $n$  points  $M_i$ , il existe une droite passant par le point moyen  $G$  et telle que la somme des carrés des écarts (ou *résidus*)  $P_1M_1^2 + P_2M_2^2 + \dots + P_nM_n^2$  soit minimale. Cette droite est appelée *droite de régression 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`.