Statistiques à une et deux variables, cours, terminale STG

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 $\bar{x}$* :

...

- *Variance V* :

...

- *Écart type σ* :

...

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 :

$\bar{x} =$ ......

L'écart type est ....

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 $i \in \{1; \ldots; n\}$ 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 .... 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 .... et .... 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 :

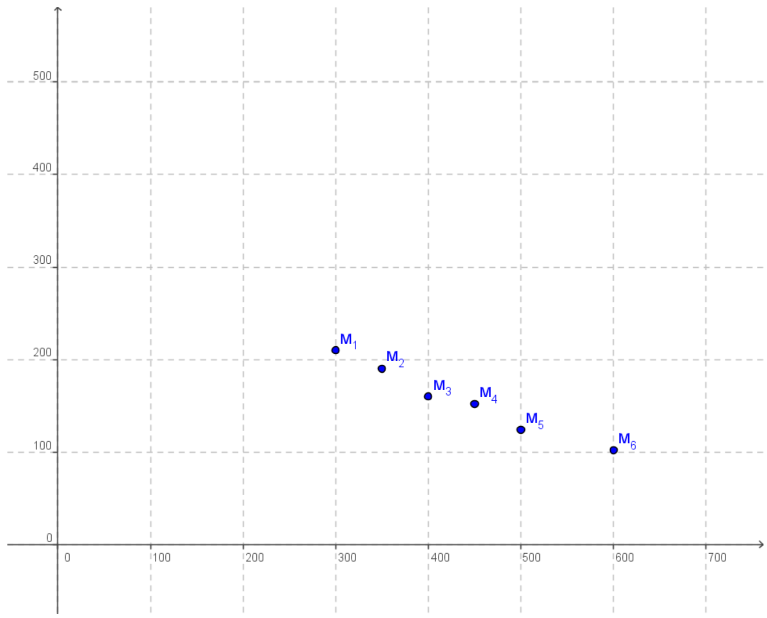
$\bar{x} = \ldots$

$\bar{y} = \ldots$

2.2 Ajustement d'un nuage de points

Définition :

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

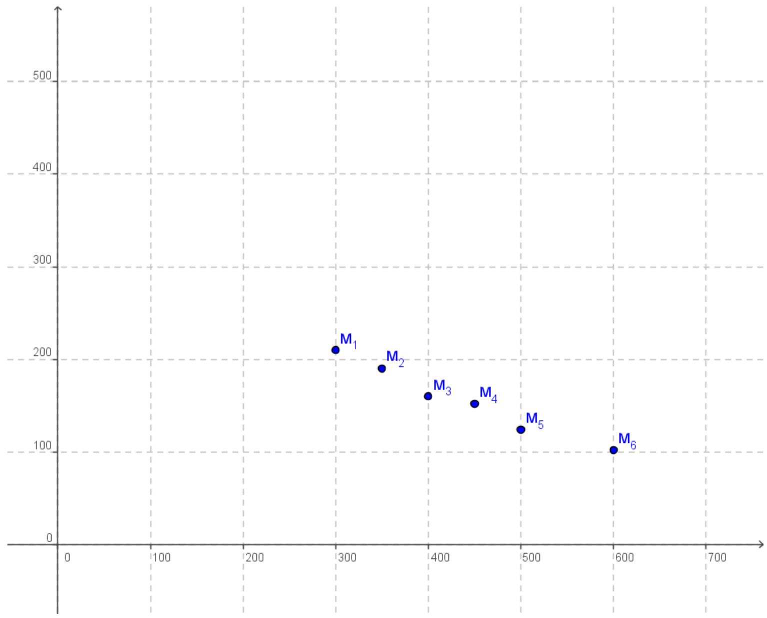


Remarque :
Il existe d'autres types d'ajustement : dans certains cas, on peut observer que visiblement une droite ne convient pas mais que le nuage de points semble être approché par un autre type de courbe, parabole par exemple. En outre, certains nuages peuvent ne pas sembler être approchables par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

$\bar{x_1} = ...$

$\bar{y_1} = ...$

Le point moyen G_2 du deuxième nuage a pour coordonnées :

$\bar{x_2} = ...$

$\bar{y_2} = ...$

La droite de MAYER est alors la droite (G_1G_2).

Méthode des moindres carrés :

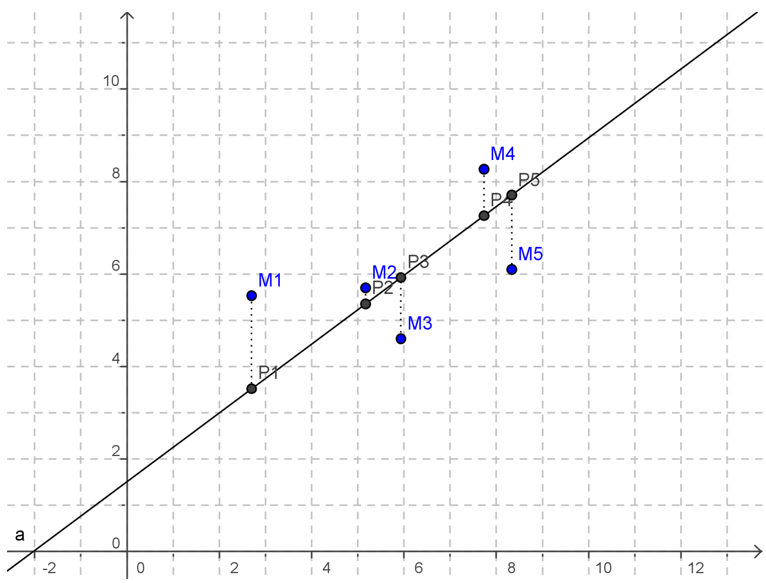
Avec les notations de la figure ci-dessous, étant donné un nuage de n points M_i, il existe une droite passant par le point moyen G et telle que la somme des carrés des écarts (ou *résidus*) $P_1M_1^2 + P_2M_2^2 + \ldots + P_nM_n^2$ soit minimale. Cette droite est appelée *droite de régression de y en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

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

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 $m = ...$

et $p = ...$

- Recherche de l'équation réduite avec la calculatrice :
  * TI 82 et plus :
  Aller dans le menu $\boxed{\text{STAT}}$ puis $\boxed{\text{EDIT}}$. Entrer les valeurs x_i dans la colonne L_1 et les valeurs y_i dans la colonne L_2. Quitter ($\boxed{\text{2nde}}$ $\boxed{\text{QUIT}}$) puis menu $\boxed{\text{STAT}}$ et $\boxed{\text{CALC}}$. Choisir $\boxed{\text{LinReg(ax+b)}}$ puis $\boxed{\text{2nd}}\boxed{\text{L1}}\boxed{,}\boxed{\text{2nd}}\boxed{\text{L2}}$ pour indiquer les deux colonnes à utiliser. Valider ensuite $\boxed{\text{ENTER}}$.
  * CASIO Graph 25 et plus :
  Aller dans le menu $\boxed{\text{STAT}}$ puis entrer les valeurs x_i dans la colonne 1 et les valeurs y_i dans la colonne 2. Choisir ensuite $\boxed{\text{CALC}}$ puis $\boxed{\text{REG}}$ puis $\boxed{\text{X}}$.