

Statistiques à une ou deux variables, cours, T STMG

8 juillet 2015

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 . On définit :

- *La moyenne \bar{x}* :

...

- *L'écart type σ* qui mesure

On utilisera la calculatrice pour déterminer l'écart type.

Exemple :

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

| | | | | | | |
|------------------------|------|------|------|------|------|------|
| Prix en euros | 0,82 | 0,83 | 0,84 | 0,85 | 0,86 | 0,87 |
| Nombre de boulangeries | 10 | 42 | 85 | 23 | 8 | 2 |

On a :

$$\bar{x} = \dots$$

et

$$\sigma = \dots$$

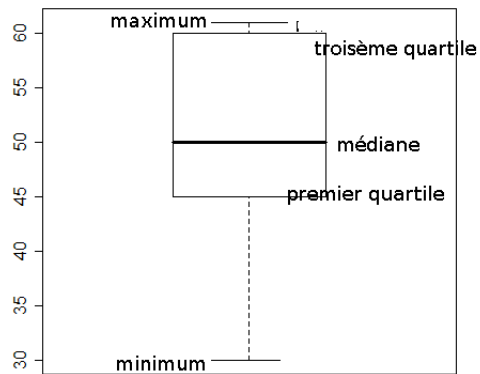
Définition (couple médiane/quartiles) :

- La médiane d'une série statistique est une valeur du caractère telle que
- Le premier quartile est la plus petite valeur pour laquelle
- le troisième quartile est la plus petite valeur pour laquelle

Définition (diagramme en boîte) :

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

-
-
-



2 Statistiques à deux variables

2.1 Vocabulaire

Définition :

- Soient x et y deux caractères quantitatifs d'une même population. À 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 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 :

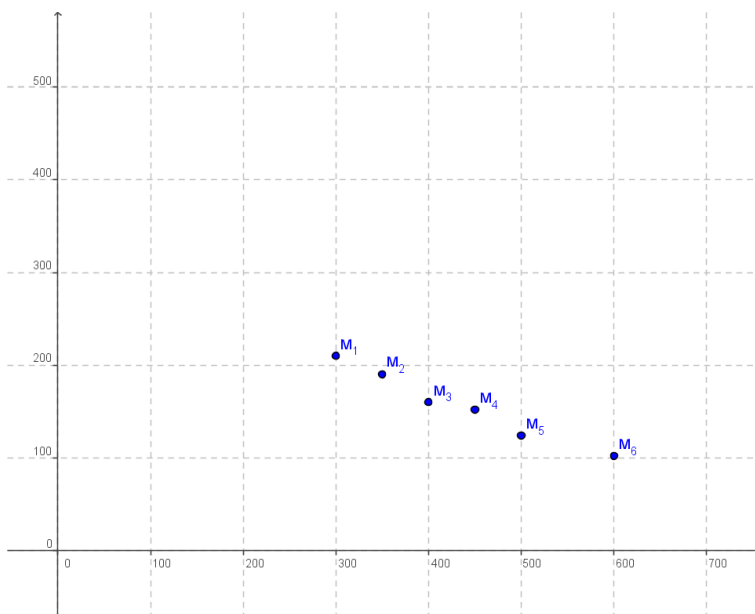
| | | | | | | |
|----------------------------|-----|-----|-----|-----|-----|-----|
| Prix x_i en euros | 300 | 350 | 400 | 448 | 500 | 600 |
| Nombre de machines vendues | 208 | 190 | 160 | 152 | 124 | 102 |

Le nuage de points associé à cette série est constitué des points $M_1 (300; 210)$, $M_2 (350; 190)$, ..., $M_6 (600; 102)$.

2.2 Ajustement d'un nuage de points

Définition :

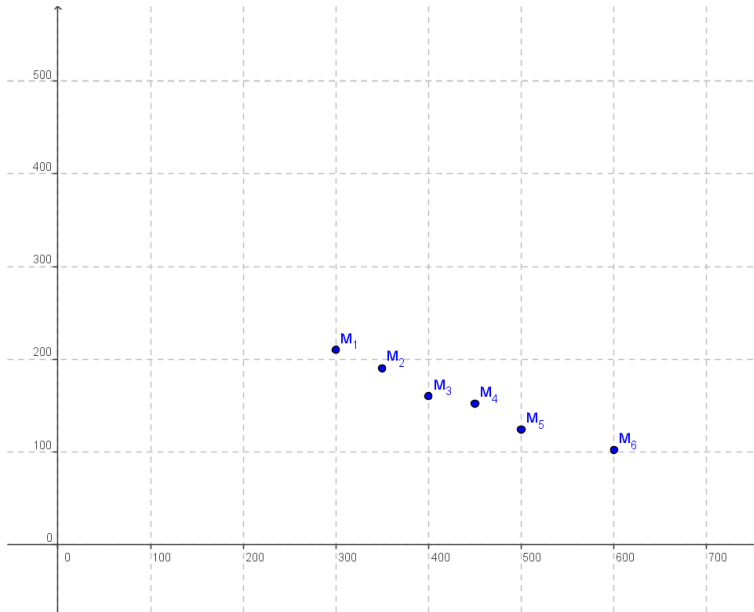
Toute droite "résumant approximativement" le nuage est appelée du nuage de points.



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



Méthode des moindres carrés :

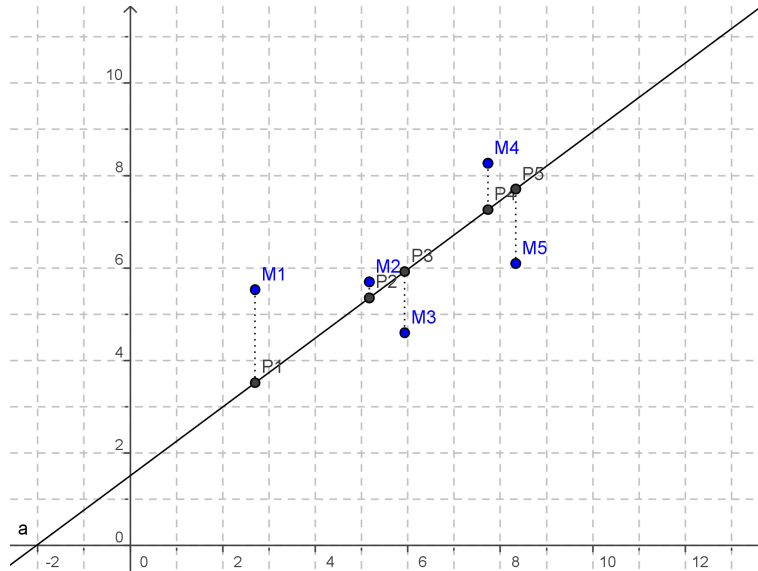
Avec les notations de la figure ci-dessous, étant donné un nuage de n points M_i , il existe une droite passant par le point moyen G et telle que la somme des carrés des écarts (ou *résidus*) $P_1M_1^2 + P_2M_2^2 + \dots + P_nM_n^2$ soit minimale. Cette droite est appelée *droite de régression de y en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

$m = \dots$

et $p = \dots$

- 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`.