STATISTIQUES À DEUX VARIABLES

I Représenter une série statistique à deux variables

Définition n°1. ***Série statistique à deux variables***

Une série statistique à deux variables est une série statistique étudiant simultanément deux caractères sur un même échantillon de n individus extraits d'une population.

Connaissance n°1 ***Représentation sous forme de tableau***

On peut présenter une série statistique à deux variables à l'aide d'un tableau statistique de la forme suivante :

Valeurs du 1er caractère	x_1	x_2	$\dots$	x_n
Valeurs du 2nd caractère	y_1	y_2	$\dots$	y_n

Connaissance n°2 ***Représentation graphique : le nuage de points***

On peut aussi représenter une série statistique à deux variables dans un repère orthogonal à l'aide d'un nuage de points :

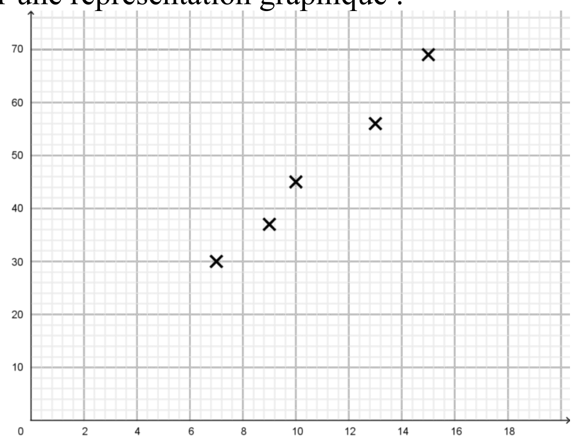
$M_1(x_1\ ;\ y_1)$, $M_2(x_2\ ;\ y_2)$, … , $M_n(x_n\ ;\ y_n)$

Exemple n°1. ***Ma petite entreprise ♪ ♫...***

Le tableau suivant présente l'évolution du nombre de clients et du chiffre d'affaires en milliers d'euros d'une micro-entreprise au cours des 6 mois :

Nombre de clients : x_i	7	9	10	13	15	18
Chiffre d'affaires : y_i	30	37	45	56	69	81

On peut en donner une représentation graphique :



Définition n°2. ***Le point moyen du nuage***

On note $\overline{x} = \dfrac{x_1+x_2+\dots+x_n}{n}$ la moyenne des x_i et

$\overline{y} = \dfrac{y_1+y_2+\dots+y_n}{n}$ la moyenne des y_i .

On appelle point moyen du nuage le point $G(\overline{x}\ ;\ \overline{y})$

Méthode n°1. ***Déterminer le point moyen d'un nuage***

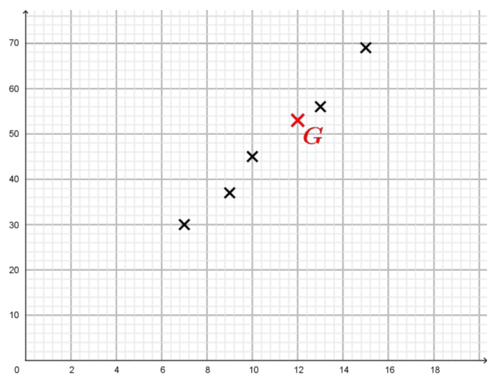
On va calculer le point moyen du nuage de l'exemple n°1

Notons $G(x_G\ ;\ y_G)$ le point moyen du nuage alors :

$$x_G = \frac{7+9+10+13+15+18}{6} = 12 \text{ et}$$

$$y_G = \frac{30+37+45+56+69+81}{6} = 53$$

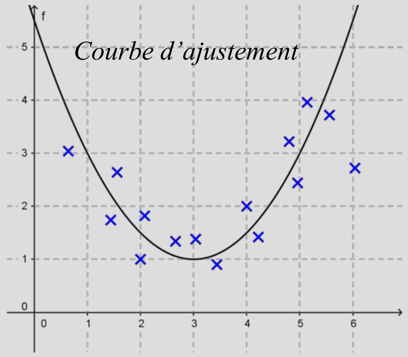
Ainsi $\boxed{G(12\ ;\ 53)}$



II Ajustement affine

II.1 Le principe

Le problème d'ajustement d'un nuage de points consiste à trouver la courbe d'une fonction f qui approche (ajuste) au mieux les points du nuage. On établie alors une relation idéale (théorique) entre y et x de la forme $y=f(x)$. Grâce à cette courbe, on pourra faire des prévisions (approximations) de valeurs non données dans la série.



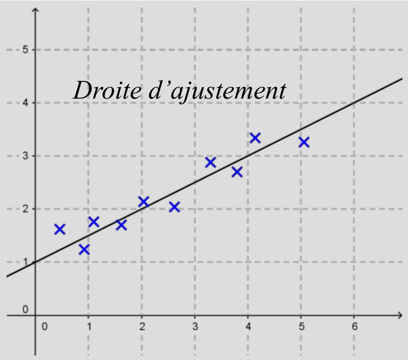
Si ces valeurs sont incluses dans la plage de données du tableau on parle d'interpolation et si elles sont à l'extérieur du tableau on parle d'extrapolation.

Remarque n°1.

Parfois, il n'existe pas de courbe d'ajustement.

II.2 Le cas particulier

L'ajustement affine est l'ajustement qui consiste à trouver une droite qui rende compte de la forme « aligné » d'un nuage en approchant au mieux les points qui le constituent.

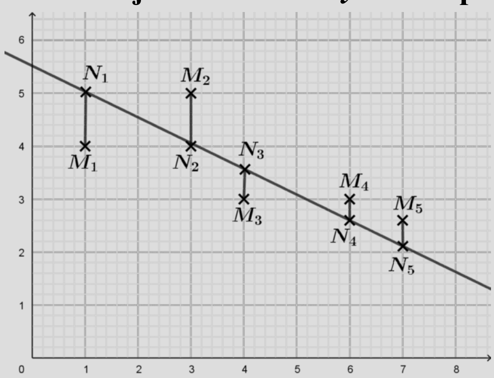


Connaissance n°3 ***La méthode des moindres carrées***

On se donne une série statistique à deux variables que l'on représente par un nuage de points (les M_i) et on cherche à tracer une droite qui minimise la somme des $M_iN_i^2$ (Les N_i étant les points de la droite qui correspondent aux M_i).

On nomme cette droite :

droite d'ajustement de y en x par la méthode des moindres carrés.



Cette droite a pour équation :

$y = mx+p$

avec $m = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$

et $p = \overline{y}-m\overline{x}$

Rassurez-vous c'est la calculatrice qui fera les calculs pour nous et nous remercions Yvan Monka pour son TUTORIEL.

Propriété n°1. ***(admise)***

La droite d'ajustement de y en x par la méthode des moindres carrés passe par le point moyen du nuage de points de la série statistique.