Série statistique à deux variables :

On considère deux variables statistiques numériques x et y observées sur une même population de n individus.
On note $x_1 ; x_2 ; \dots ; x_n$ les valeurs relevées pour la première variable et $y_1 ; y_2 ; \dots ; y_n$ les valeurs relevées pour la deuxième variable.
Les couples $(x_1 ; y_1) ; (x_2 ; y_2) ; \dots ; (x_n ; y_n)$ forment une **série statistique à deux variables**.

Nuage de points :

Dans le plan rapporté à un repère orthogonal, on appelle **nuage de points** associé à cette série statistique à deux variables, l'ensemble des points $M_1(x_1 ; y_1)$; $M_2(x_2 ; y_2)$;...; $M_n(x_n ; y_n)$.

Point moyen :

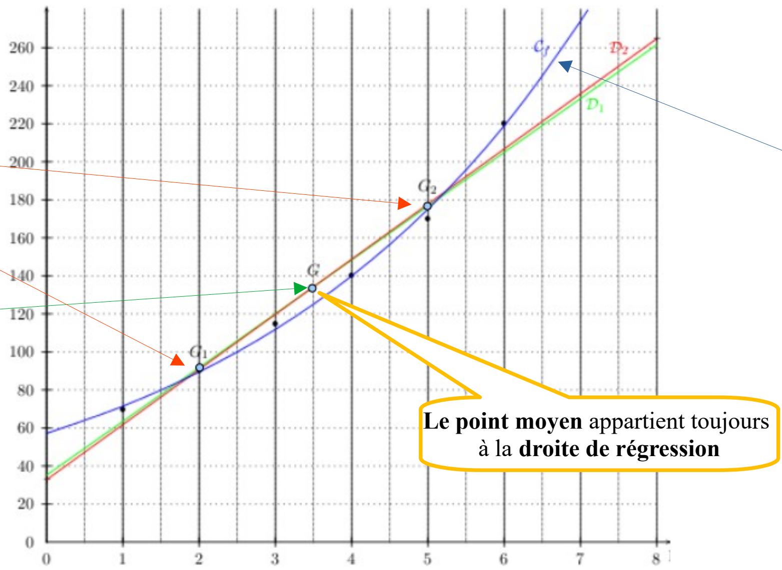
On appelle **point moyen** de cette série le point G de coordonnées $(\overline{x} ; \overline{y})$ où $\overline{x}$ et $\overline{y}$ sont les moyennes respectives des séries $x_1 ; x_2 ; \dots x_n$ et $y_1 ; y_2 ; \dots y_n$.

Ajustement :

Méthode de Mayer :
Cet ajustement consiste à déterminer la droite passant par deux points moyens du nuage de points.

Méthode des moindres carrés :

Très facile à la calculatrice
Pour une Tinspire :
>Menu
>Statistiques
>Calculs statistiques
>Régression linéaire (mx+b)



Ajustement exponentiel

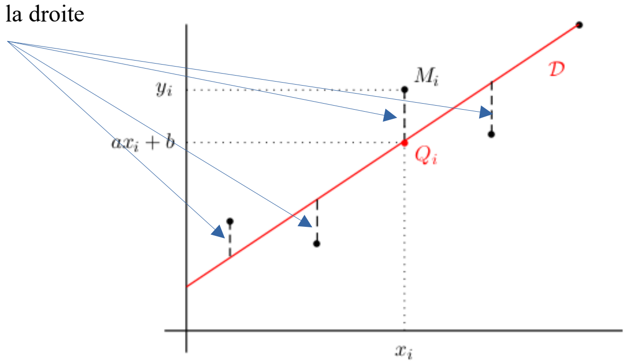
On parle d'**interpolation** pour des valeurs à l'intérieur de la plage des valeurs observées et d'**extrapolation** pour des valeurs à l'extérieur de cette plage.

Les formules de la méthode des moindres carrés :

Dans la pratique, on utilise directement les résultats fournis par la calculatrice.

La droite de régression de y en x est la droite qui minimise la somme des carrés des distances des points aux points de même abscisse de la droite.

$$\sum_{i=1}^{n}(M_i Q_i)^2=\sum_{i=1}^{n}(y_i-(a x_i+b))^2$$



Covariance :

$$\operatorname{cov}(x ; y)=\sigma_{xy}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\frac{1}{n}\sum_{i=1}^{n}x_i y_i-\overline{x}\,\overline{y}$$

On aura aussi besoin de : $\operatorname{cov}(x ; x)=\sigma_{x^2}=\mathrm{V}(x)=(\sigma(x))^2$ (On retrouve **la variance** de x)

Droite de régression :

La droite de régression d de y en x a pour équation $y=ax+b$ où :

$$a=\frac{\sigma_{xy}}{\sigma_{x^2}}=\frac{\sigma_{xy}}{\mathrm{V}(x)} \quad \text{et } b \text{ vérifie } \overline{y}=a\,\overline{x}+b$$

Le point moyen G du nuage appartient toujours à la droite de régression de y en x.

Coefficient de corrélation linéaire :

Le **coefficient de corrélation linéaire** d'une série statistique de variables x et y est le nombre r défini par :

$$r=\frac{\sigma_{xy}}{\sigma(x)\times\sigma(y)}$$

Ce coefficient sert à mesurer la qualité d'un ajustement affine :

Le coefficient de corrélation linéaire r vérifie $-1\leqslant r\leqslant 1$.

Plus le coefficient de corrélation linéaire est **proche de 1 en valeur absolue, meilleur est l'ajustement affine**.

Si $r\leqslant -0{,}75$ ou $r\geqslant 0{,}75$, on dit que la corrélation linéaire entre les séries x et y est **forte**.

Lorsque $r=\pm 1$, la droite de régression passe par tous les points du nuage, qui sont donc alignés.

Si le coefficient de corrélation est **proche de 0**, cela signifie que le nuage de points ne peut pas être ajusté au mieux par une droite. Il faut dans ce cas essayer un autre type de courbe (exponentiel …)