Statistiques à 2 variables

I) Représentations

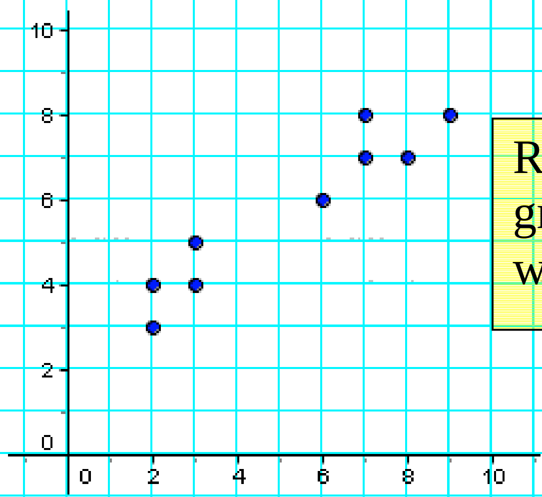
a) Tableau simple

x_i	2	2	3	3	6	7	7	8	9
y_i	3	4	4	5	6	7	8	7	8

b) Tableau pondéré

y_i \ x_i	[2,4[3	[4,6[5	[6,8[7	[8,10[9
[3,5[4	3	0	0	0
[5,7[6	1	0	1	0
[7,9] 8	0	0	2	2

c) Nuage de points



On peut définir , à partir d'un nuage de points , le point G de coordonnées (x_G , y_G). Qui est appelé point moyen d'un nuage

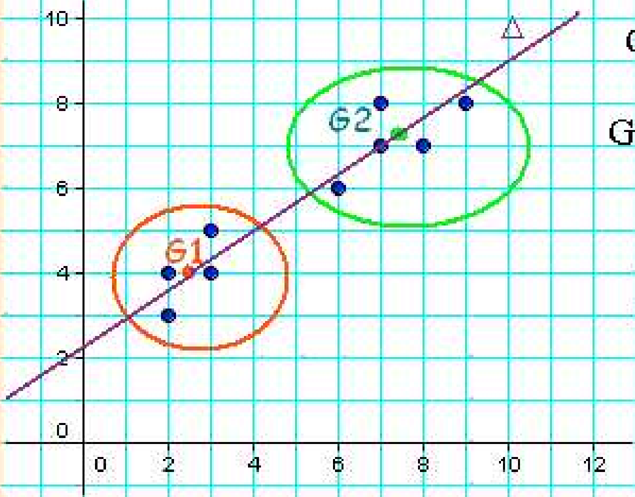
Avec
$$\begin{cases} x_G = \dfrac{1}{n} \sum x_i = \overline{x} \\ y_G = \dfrac{1}{n} \sum y_i = \overline{y} \end{cases}$$

II) Ajustements linéaires

1) Ajustement linéaire par méthode de Mayer (si le nuage "forme" une droite)

On fractionne le nuage de points en 2 sous groupes de taille égale (ou si n est impair , l'un des sous groupes aura 1 point de plus que l'autre) . On prend les points moyens G1 et G2 des nuages partiels.

La droite de Mayer est la droite passant par G1 et G2



$$G1 = \left(\frac{2+2+3+3}{4}, \frac{3+4+4+5}{4}\right) = \left(\frac{10}{4}, \frac{14}{4}\right)$$

$$G2 = \left(\frac{6+7+7+8+9}{5}, \frac{6+7+8+7+8}{5}\right) = \left(\frac{37}{5}, \frac{36}{5}\right)$$

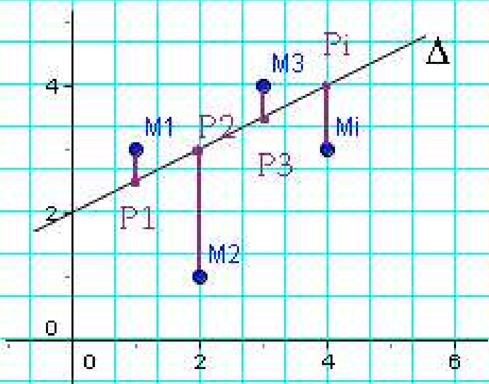
Droite de Mayer :

$$\Delta : y = ax + b$$

$$a = \frac{\Delta y}{\Delta x} = \frac{\frac{36}{5} - \frac{11}{4}}{\frac{37}{5} - \frac{10}{4}}$$

b) Méthode des moindres carrés , droites de régression

* La droite de régression de y en x , Δ , est telle que $\Sigma\, P_i M_i^2$ soit minimale

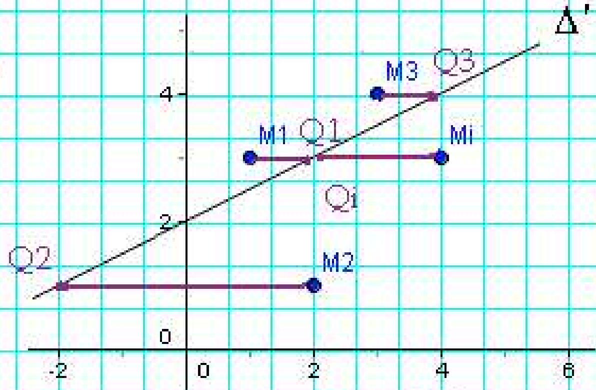


Equation de Δ

$$\boxed{y = m\,(x - \bar{x}) + \bar{y}}$$

Ces 2 droites passent par le point G

* La droite de régression de x en y , Δ', est telle que $\Sigma\, P_i Q_i^2$ soit minimale



Equation de Δ'

$$\boxed{y = m'\,(x - \bar{x}) + \bar{y}}$$

$$\text{Avec } m = \frac{\operatorname{cov}(x,y)}{V(x)} = \frac{\Sigma\,(x_i - \bar{x})(y_i - \bar{y})}{\Sigma\,(x_i - \bar{x})^2} = \frac{\sigma_{xy}}{\sigma_x^2}$$

$$\text{Avec } \frac{1}{m'} = \frac{\operatorname{cov}(x,y)}{V(y)} = \frac{\Sigma\,(x_i - \bar{x})(y_i - \bar{y})}{\Sigma\,(y_i - \bar{y})^2} = \frac{\sigma_{xy}}{\sigma_y^2}$$

On appelle coefficient de corrélation linéaire des variables x et y la valeur :

$$\boxed{r = \sqrt{\frac{m}{m'}} = \sqrt{m * \frac{1}{m'}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}}$$

si $|r| = 1$ Corrélation linéaire parfaite

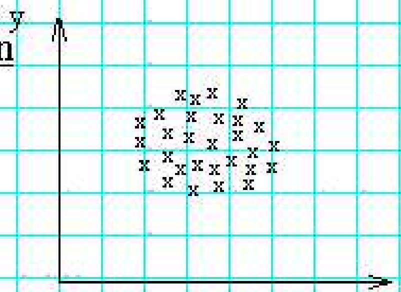
$0.5 < |r| < 1$ Corrélation linéaire bonne

$|r| < 0.5$ Corrélation linéaire mauvaise -> il faut trouver une autre corrélation

III) Les formes de nuages de points

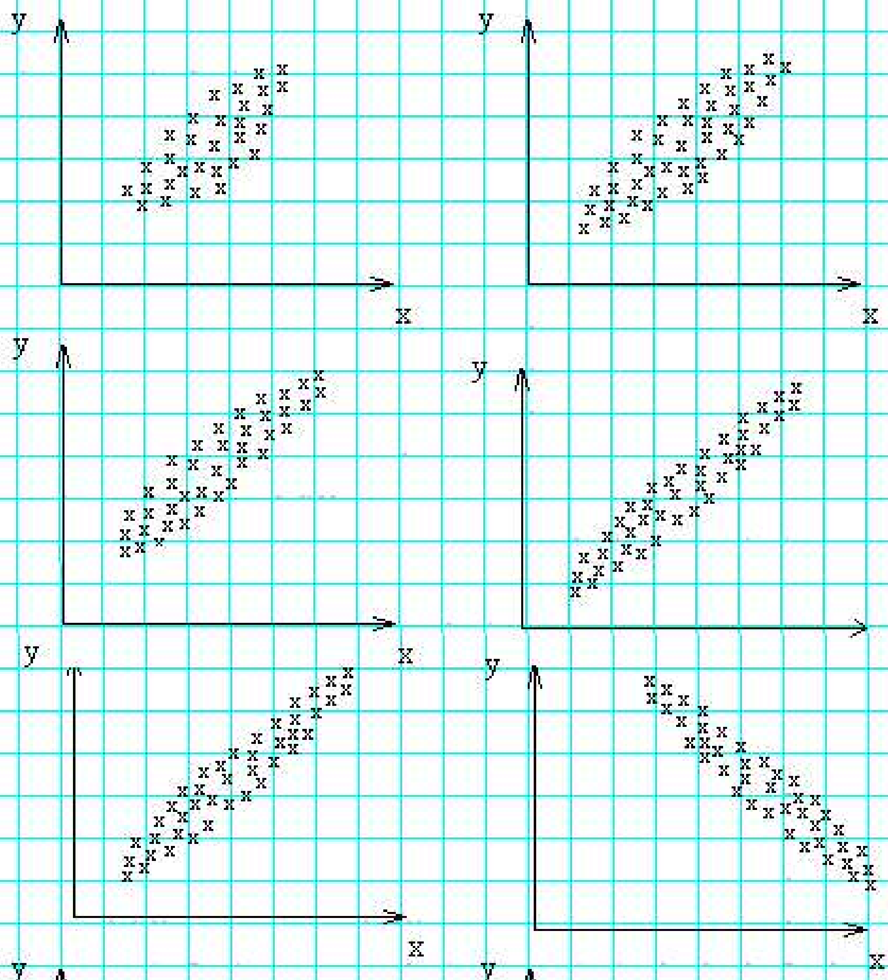
Absence de liaison

$r \sim 0$



Liaison linéaire

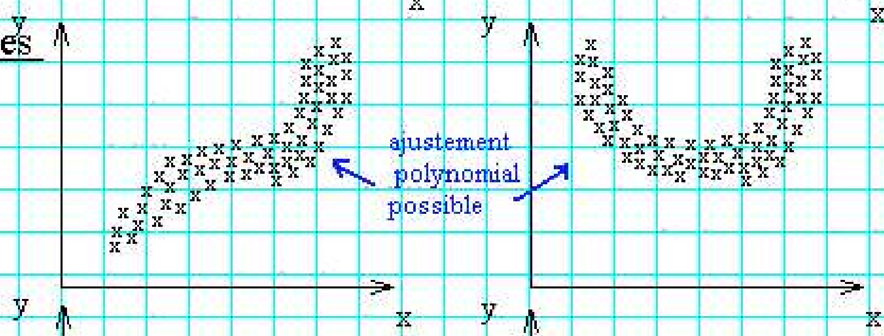
-> Ajustement linéaire

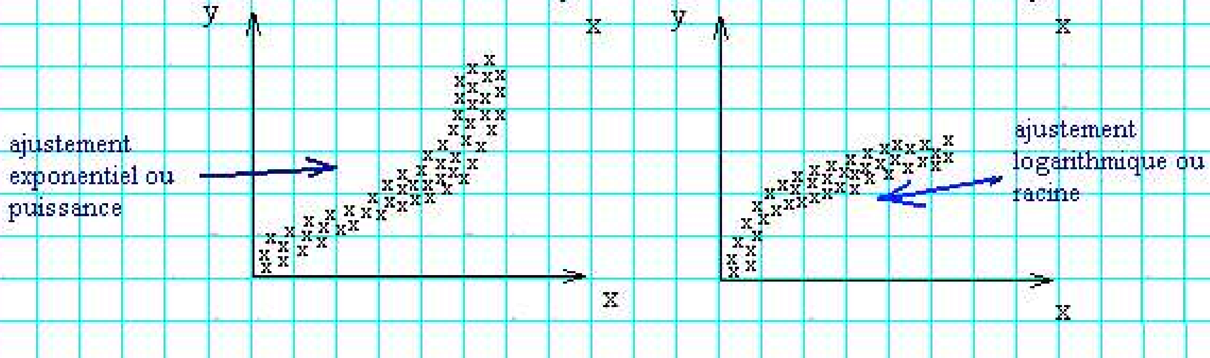


Liaisons non linéaires

$r \sim 0$

-> Ajustements non linéaires





La forme du nuage de points peut conduire à rechercher une fonction autre qu'une fonction affine.Dans ce cas , on recherche un ajustement autre que linéaire (rappel : un ajustement consiste à trouver l'équation d'une courbe la plus proche de la forme du nuage).

IV) Ajustements non linéaires

Type du nuage	Equation "type" du nuage	Chgt variable X	Chgt variable Y	Equation de la droite
Logarithmique	$y = a \ln x + b$	$X = \ln x$	$Y = y$	$Y = a X + b$
Exponentiel	$y = a\, e^{bx}$	$X = x$	$Y = \ln y$	$Y = a X + \ln b$
Puissance	$y = a\, x^{n} + b$	$X = x^{n}$	$Y = y$	$Y = a X + b$