



Cours Math310

Deuxième partie : statistique descriptive (Suite)

Chapitre 2: Séries statistiques à deux variables (Suite)

M. BEZOUÏ

17 janvier 2012

Plan de travail

- 1 Nuage de points
- 2 La covariance
- 3 Droite de regression
- 4 Coefficient de corrélation
- 5 Droite de Mayer

Un peu de silence S.V.P

Dans le cours précédent

Dans le cours précédent

Tableau de contingence

Distribution conjointe

Distribution marginale

Distribution conditionnelle

Moyennes conditionnelles

Variances conditionnelles

Ecart type

Moyenne des moyenne

Indépendance

Dans le cours précédent

Tableau de contingence

Distribution conjointe

Distribution marginale

Distribution conditionnelle

Moyennes conditionnelles

Variances conditionnelles

Ecart type

Moyenne des moyenne

Indépendance

Dans le cours précédent

Tableau de contingence

Distribution conjointe

Distribution marginale

Distribution conditionnelle

Moyennes conditionnelles

Variances conditionnelles

Ecart type

Moyenne des moyenne

Indépendance

Dans le cours précédent

Tableau de contingence
Distribution conjointe
Distribution marginale
Distribution conditionnelle
Moyennes conditionnelles
Variances conditionnelles
Ecart type
Moyenne des moyenne
Indépendance

Dans le cours précédent

Tableau de contingence
Distribution conjointe
Distribution marginale
Distribution conditionnelle
Moyennes conditionnelles
Variances conditionnelles
Ecart type
Moyenne des moyenne
Indépendance

Dans le cours précédent

Tableau de contingence
Distribution conjointe
Distribution marginale
Distribution conditionnelle
Moyennes conditionnelles
Variances conditionnelles
Ecart type
Moyenne des moyenne
Indépendance

Dans le cours précédent

Tableau de contingence
Distribution conjointe
Distribution marginale
Distribution conditionnelle
Moyennes conditionnelles
Variances conditionnelles
Ecart type
Moyenne des moyenne
Indépendance

Dans le cours précédent

Tableau de contingence
Distribution conjointe
Distribution marginale
Distribution conditionnelle
Moyennes conditionnelles
Variances conditionnelles
Ecart type
Moyenne des moyenne
Indépendance

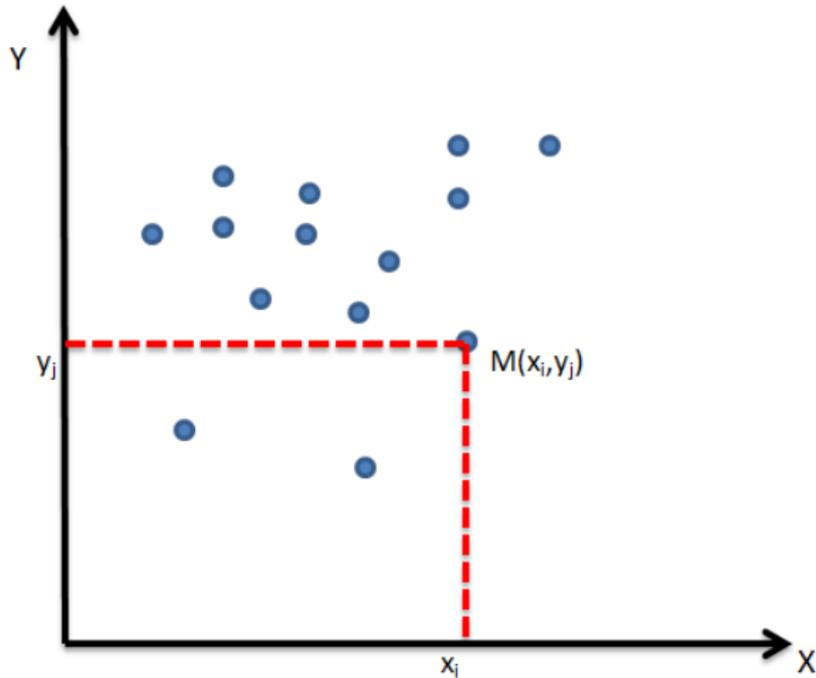
Dans le cours précédent

Tableau de contingence
Distribution conjointe
Distribution marginale
Distribution conditionnelle
Moyennes conditionnelles
Variances conditionnelles
Ecart type
Moyenne des moyenne
Indépendance

Plan de travail

- 1 Nuage de points
- 2 La covariance
- 3 Droite de regression
- 4 Coefficient de corrélation
- 5 Droite de Mayer

Nuage de points



Nuage de points

Les couples $\{(x_i, y_j), i = \overline{1, p}, j = \overline{1, q}\}$ peuvent être représentés graphiquement de la manière précédente.

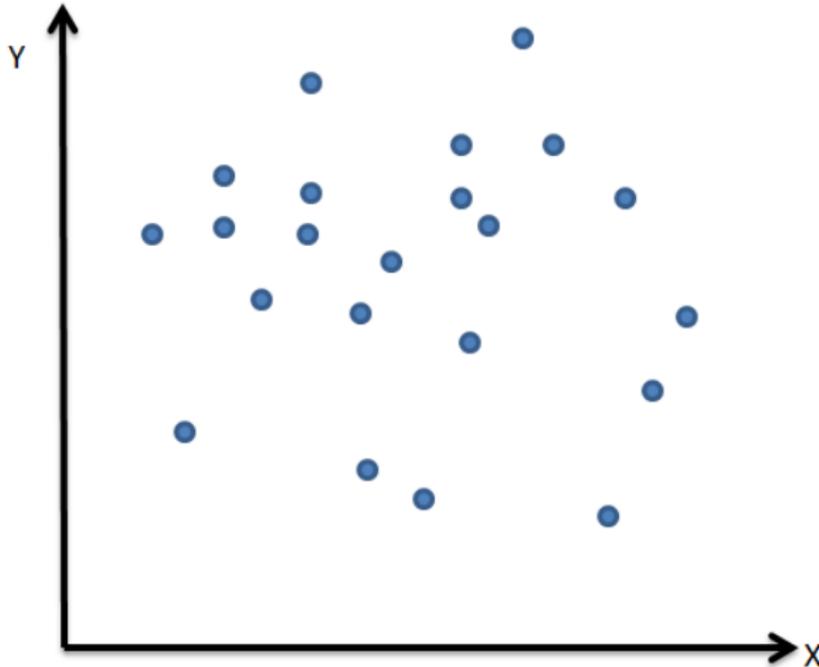
Le nuage de point c'est l'ensemble des points $\{M_{ij}, i = \overline{1, p}, j = \overline{1, q}\}$. Le nuage de points est un bon indicateur sur l'existence d'une liaison entre X et Y .

Nuage de points

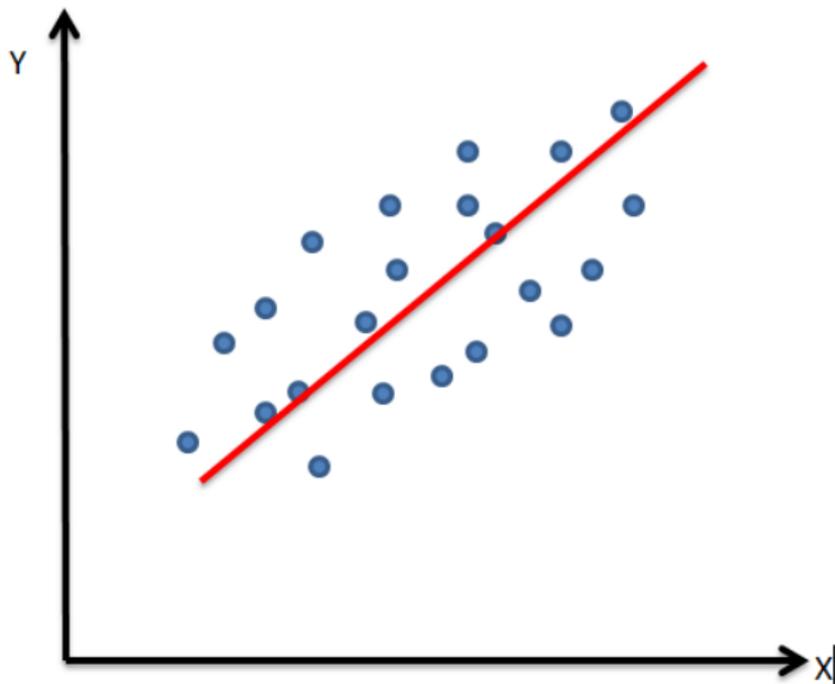
Remarque

Si les points du nuage sont répartis sur une bande régulière et allongées, on peut proposer un ajustement linéaire pour expliquer Y en fonction de X ou inversement

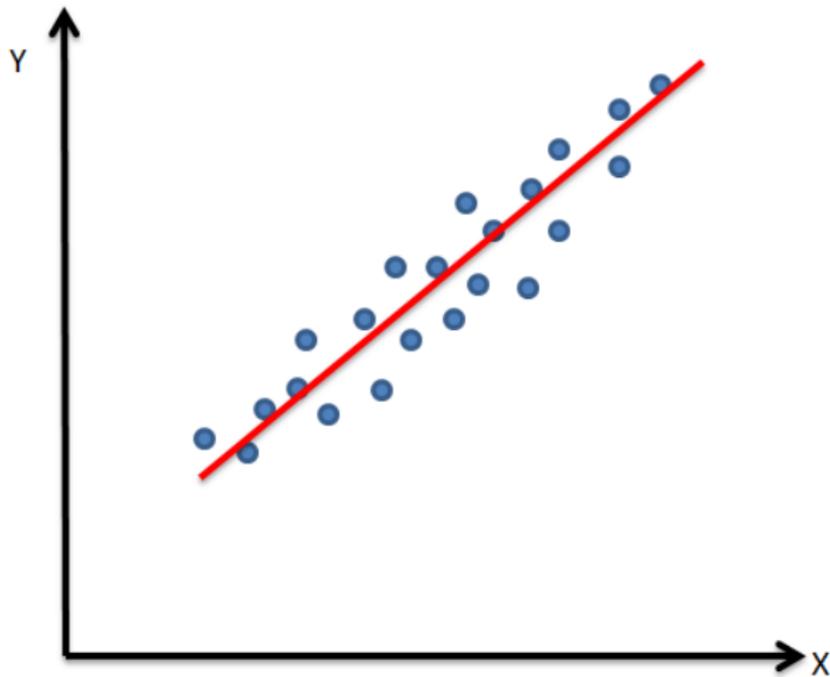
Absence de liaison linéaire



Forte liaison linéaire



Faible liaison linéaire



Plan de travail

- 1 Nuage de points
- 2 La covariance**
- 3 Droite de regression
- 4 Coefficient de corrélation
- 5 Droite de Mayer

La covariance

On appelle covariance entre "X et Y" la quantité :

$$\begin{aligned} \text{cov}(x, y) &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i - \bar{x})(y_j - \bar{y}) \\ &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j (\bar{x} \bar{y}) \end{aligned}$$

Remarque

La variance peut être positive ou négative, elle joue un rôle analogue à celui de la variance.

La covariance

On appelle covariance entre "X et Y" la quantité :

$$\begin{aligned} cov(x, y) &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i - \bar{x})(y_j - \bar{y}) \\ &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j (\bar{x} \bar{y}) \end{aligned}$$

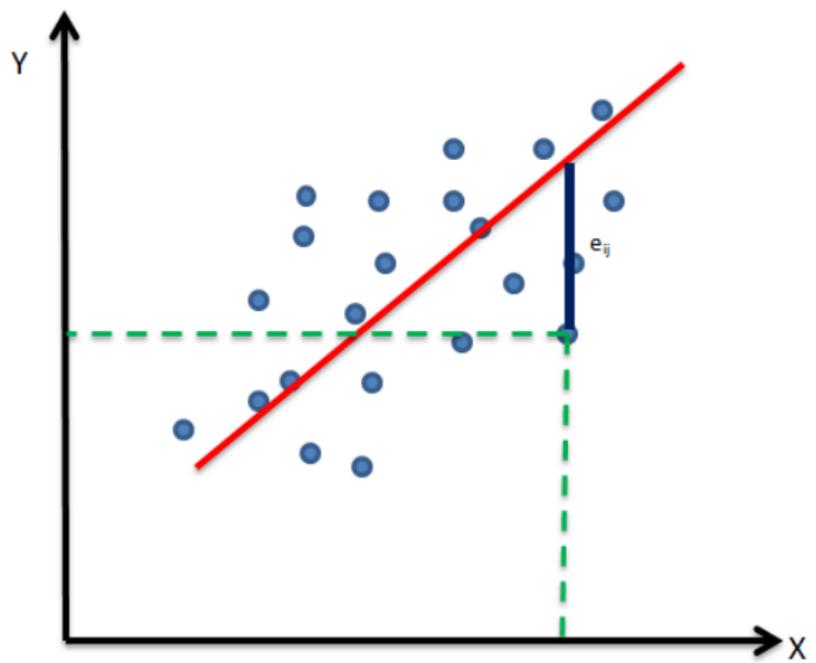
Remarque

La variance peut être positive ou négative, elle joue un rôle analogue à celui de la variance.

Plan de travail

- 1 Nuage de points
- 2 La covariance
- 3 Droite de regression**
- 4 Coefficient de corrélation
- 5 Droite de Mayer

Droite de Y en X



Droite de regression (Ou droite des moindres carrés)

Soit, (Δ) , la droite de regression de Y en X :

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

Soit le point $M = (x_i, y_j)$,

$$y'_j = ax_i + b$$

y_j : Valeur observée, et y'_j : valeur ajustée.

Soit, l'erreur entre y_j et y'_j , $e_{ij} = y_j - y'_j = y_j - (ax_i + b)$

Chercher la droite (Δ) revient à chercher les paramètres a et b ?

Pour cela, on utilise le critère de minimisation de la somme des erreurs e_{ij} au carré, on minimise le critère :

$$S(a, b) = \sum_{i=1}^p \sum_{j=1}^q n_{ij} e_{ij}^2 = \sum_{i=1}^p \sum_{j=1}^q n_{ij} (y_j - (ax_i + b))^2$$

Droite de regression

Théorème

Soit $M_{ij}, i = \overline{1, p}, j = \overline{1, q}$, un nuage de point de coordonnées (x_i, y_j) .

L'équation de la droite de regression de Y en X est écrite sous la forme suivante $y = ax + b$ est définie par :

$$a = \frac{\text{cov}(X, Y)}{V(X)}$$

$$b = \bar{y} - a\bar{x}$$

Droite de X en Y

Soit (Δ') : $x = \alpha y + \beta$

Avec la même procédure appliquée pour trouver la droite de Y en X , on trouve que la droite de X en Y s'écrit sous la forme $x = \alpha y + \beta$ tels que :

$$a = \frac{\text{cov}(X, Y)}{V(Y)}$$

$$b = \bar{x} - a\bar{y}$$

- \bar{x} : moyenne marginale de X ,
- \bar{y} : moyenne marginale de Y ,
- $V(X)$: variance marginale de X ,
- $V(Y)$: variance marginale de Y ,

Droite de regression

Remarque

Le point (\bar{x}, \bar{y}) est le point d'intersection des deux droites de regression (Δ) et (Δ') . Il est applé "**Centre de gravité du nuage de point**" ou "**Point moyen**"

Plan de travail

- 1 Nuage de points
- 2 La covariance
- 3 Droite de regression
- 4 Coeffecient de corrélation**
- 5 Droite de Mayer

Coefficient de corrélation

Pour mesurer l'intensité de la liaison linéaire entre X et Y , on calcule le coefficient de corrélation noté $R(\rho, R_{xy})$, ρ_{xy} définie par

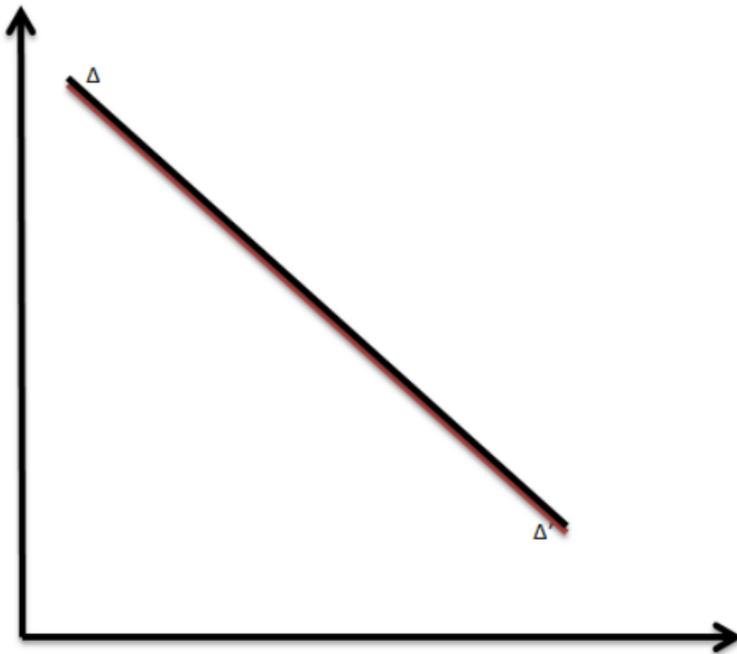
$$R = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \text{ où,}$$

$$\sigma_x = \sqrt{V(x)}, \sigma_y = \sqrt{V(y)}.$$

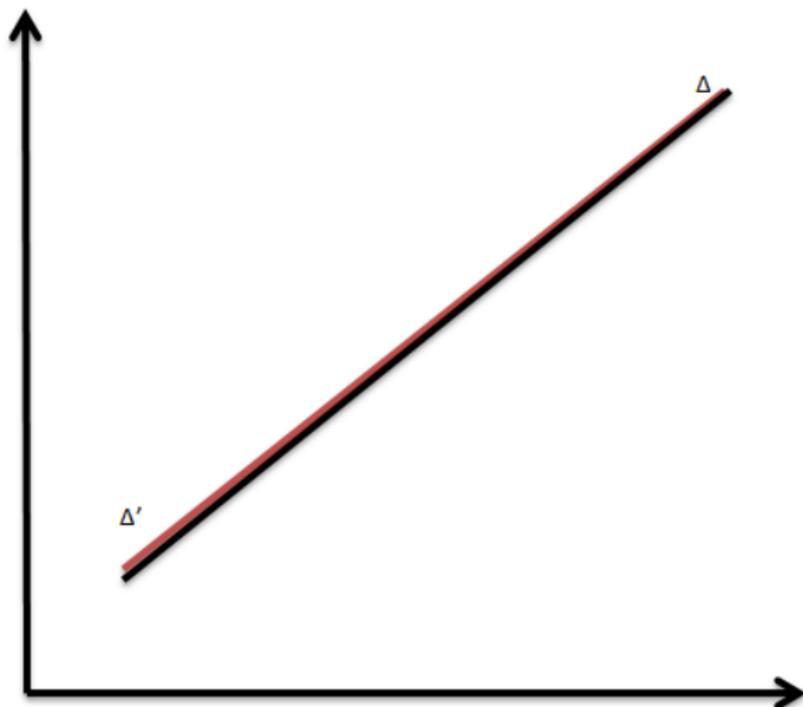
Propriétés

- $R^2 = a\alpha$
- $R^2 \leq 1 \Leftrightarrow -1 \leq R^2 \leq 1$
- $R^2 \Rightarrow$ les points sont alignés et les droites Δ et Δ' sont confondues

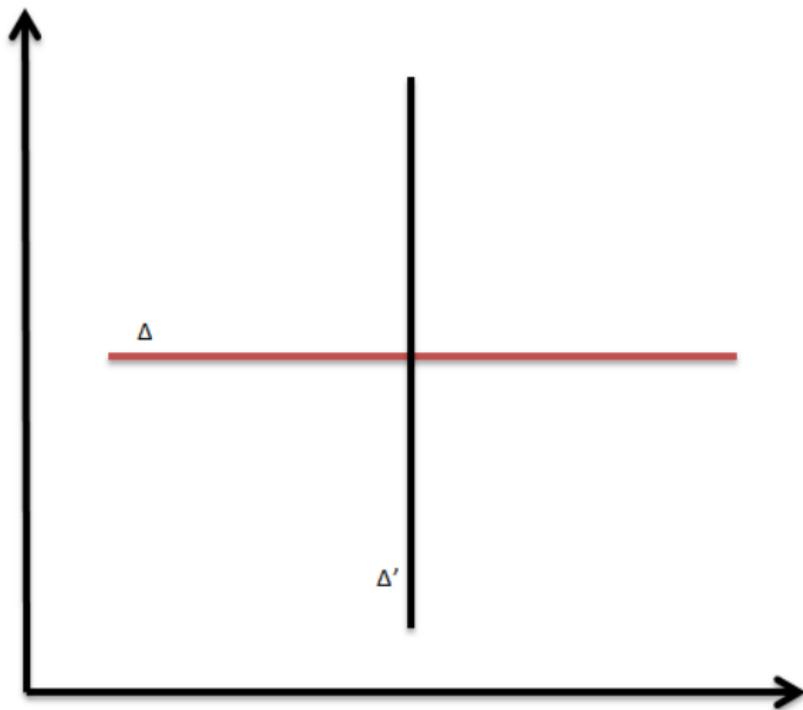
$R=-1$



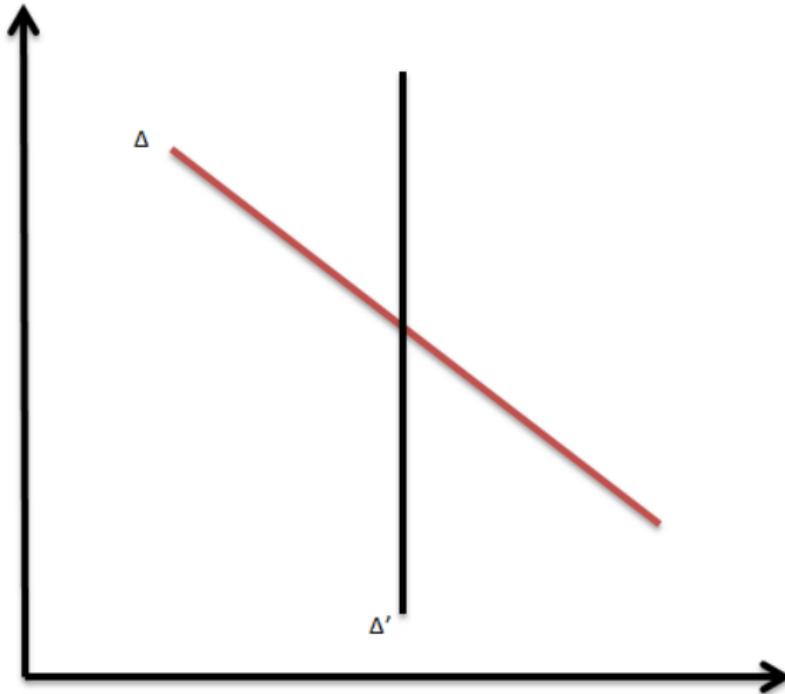
$R=+1$



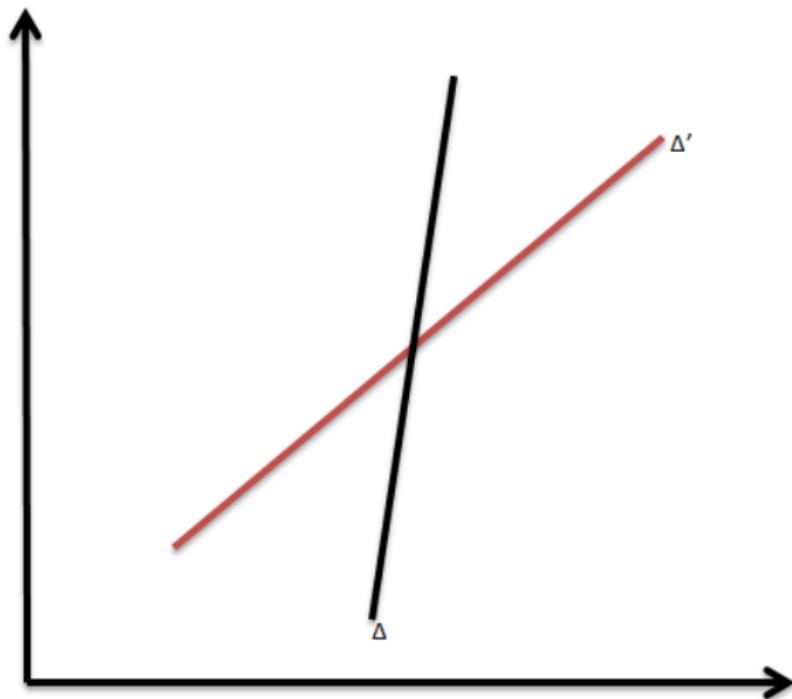
$R=0$



$$R < 0$$



$R > 0$



Remarques

$|R| = 1$: forte liaison linéaire

$|R| \simeq 0$ absence de liaison linéaire

Plus (Δ) et (Δ') se rapproche plus la corrélation est forte.

X, Y deux variables statistiques continues x_i, y_j représentent les centres de classes

Dans le cas de données brutes $(x_i, y_j), i = \overline{1, p}, j = \overline{1, q}$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}$$

Comment juger de la qualité de l'ajustement linéaire entre X et Y

Si $R \simeq 0$ Il faut rejeter l'ajustement linéaire.

Si $R \simeq 1$ Il faut compléter par un examen graphique.

Dans le cas d'absence de renseignements complémentaires On admet que $|R| \geq 0.75$ justifie l'ajustement linéaire.

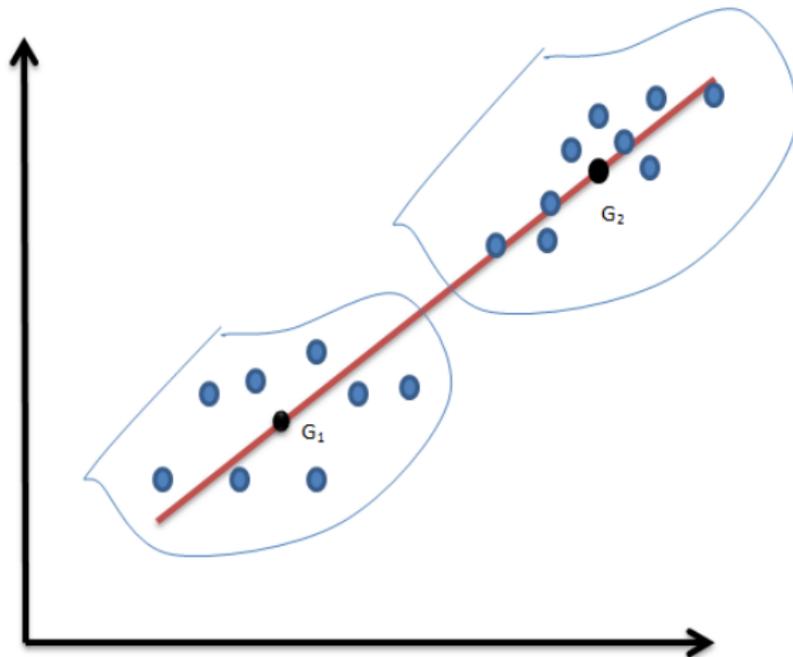
Plan de travail

- 1 Nuage de points
- 2 La covariance
- 3 Droite de regression
- 4 Coefficient de corrélation
- 5 Droite de Mayer**

Droite de Mayer

On partage l'ensemble des points (x, y) en deux sous ensemble E_1 et E_2 ayant presque le même nombre de points, on cherche les centres de gravité G_1 et G_2 , des sous ensembles E_1 et E_2 . La droite de Mayer est celle qui lie les deux points G_1, G_2

Droite de Mayer



Merci de votre
attention !