Rappel :

5. Ajustement, corrélation et régression

a. Ajustement linéaire – Droites de régression

Méthode des moindres carrés:

1er cas :

$$a = \frac{cov(x,y)}{\sigma x^2}$$

et

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

2ème cas :

$$a' = \frac{cov(x,y)}{\sigma x^2}$$

et

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

Angle entre D et D':

D : y = ax + b a: coeff de régression de y en x
D': y = a'y + b' a': coeff de régression de x en y
ou encore :

$$y = \frac{1}{a'}x - \frac{b'}{a'}$$

on a : $\theta = Arctan\left(\frac{1}{a'}\right) - Arctan\ (a)$

$\Rightarrow$

$$\tan\theta = \frac{1 - \mathrm{aa}'}{\mathrm{a} + \mathrm{a}'}$$

On va considérer 2 cas limites :

- Si x et y sont indépendantes : cov(x,y) = 0

donc a = a' $\Rightarrow$ aa'=0

d'où $\tan\theta = +\infty \quad \Rightarrow \quad \theta = 90°$

Donc si x et y indépendantes alors D et D' sont perpendiculaires.

- Si x et y sont liées par un fonction de type linéaire :

alors les points (xi, yi) sont sur la droite qui est donc D et D', dans ce cas là :

$\frac{a}{a'} = a$ car c'est la même pente. $\Rightarrow$ $aa' = 1$

Règle générale :

$$0 \le aa' \le 1$$

Alors aa' sera la coefficient de corrélation linéaire.

b. Coefficient de corrélation linéaire :

On a vu dans le chapitre sur les ajustements :

σy^2 = variance expliquée (VE) + variance résiduelle (VR)

On va partir de la variance de y :

$$\sigma y^2 = \frac{1}{n}\sum_i\sum_j nij\ (yj - \bar{y})^2$$

$$\sigma y^2 = \frac{1}{n}\sum_i\sum_j nij\ ((yj - y'i) + (y'i - \bar{y}))^2$$

$$\sigma y^2 = \underbrace{\frac{1}{n}\sum_i\sum_j nij\ (yj - y'i)}_{VR} + \underbrace{\frac{1}{n}\sum_i\sum_j nij\ (y'i - \bar{y})^2}_{VE} + \underbrace{\frac{2}{n}\sum_i\sum_j nij\ (yj - y'i)(y'i - \bar{y})}_{C}$$

On veut montrer que $\sigma y^2 = VR + VE$
Donc que C=0 ?
On a

$$C = \frac{2}{n}\sum_i \sum_j nij\ (yj - y'i)(y'i - \bar{y})$$

$$C = \frac{2}{n}\sum_i \sum_j nij\ (yj - axi - b)(axi + b - \bar{y})$$

car y'i = axi + b
d'où

$$C = \frac{2a}{n}\sum_i \sum_j nij\ (yj - axi - b)xi + \frac{2}{n}(b - \bar{y})\sum_i \sum_j nij\ (yj - axi - b)$$

$\Rightarrow \quad \frac{d\, f(a,b)}{da} = 0 \qquad\qquad \Rightarrow \quad \frac{d\, f(a,b)}{db} = 0$

donc C = 0
De plus,

$$VE = \frac{1}{n}\sum_i \sum_j nij\ (y'i - \bar{y})^2$$

$$VE = \frac{1}{n}\sum_i \sum_j nij\ (axi + b - a\bar{x} - b)^2$$

$$VE = \frac{1}{n}\sum_i \sum_j nij\ a^2(xi - \bar{x})^2 = \frac{a^2}{n}\sum_i \sum_j nij\ (xi - \bar{x})^2 = a^2\, \sigma^2 x$$

On va poser :

$$R^2 = \frac{VE}{\sigma y^2} = \frac{a^2 \sigma x^2}{\sigma y^2}$$

Or :

$$a = \frac{cov(x,y)}{\sigma x^2}$$

donc :

$$R^2 = \frac{cov^2(x,y)}{\sigma x^4} * \frac{\sigma x^2}{\sigma y^2} = \frac{cov^2(x,y)}{\sigma x^2 \sigma y^2}$$

<u>Remarques :</u>
Si $|R| \geq 0{,}9$, on a alors une corrélation linéaire entre x et y (le nuage de points utilisé dans la représentation de y par rapport à x correspond à une droite)
On a :

$$aa' = \frac{cov(x,y)}{\sigma x^2}$$

$$\frac{cov(x,y)}{\sigma x^2} = \frac{cov^2(x,y)}{\sigma x^2 \sigma y^2} = R^2$$

Extension faible au cas non linéaire après un changement de variable de type log.

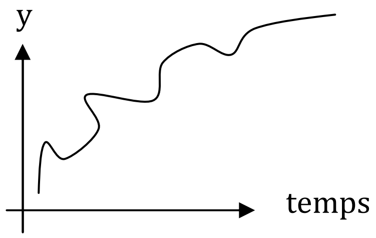
Chapitre 4 : Séries chronologiques

1. Ajustement, corrélation et régression

a. Ajustement linéaire – Droites de régression

Une série chronologique correspond à une suite d'observations d'un variable quantitative ordonnées dans le temps.
Exemple: indices statistiques, prix du pétrole, production d'automobiles, consommation d'électricité ...

Représentation :



Notation : yt ou yi ou y=f(t)
Temps doit correspondre à des intervalles réguliers.

Remarque : Il s'agit d'une série à caractère double (yi , ti) avec une seule valeur de yi par ti

b. Composante d'une série chronologique

cf graphe sur le formulaire

On distingue 4 composantes:

1. Tendance générale (trend):

- c'est la courbe qui lisse la série
- donne l'équation à très long terme des phénomènes observés

2. Cycle :

- variation de type sinusoïdale autour de la tendance générale
- cela traduit la succession de phases sur un temps long : prospérité, crise, dépression, reprise…

On assimile le trend + cycle à des mouvements extra-saisonniers ou conjoncturel (noté Ct)

3. Variations saisonnières : St

- mouvements courts de pics et de creux qui se répètent périodiquement
- dues aux saisons, congés, aux inégalités des différents mois.

4. Variations accidentelles : εt

- dues à des circonstances imprévisibles : grèves, krach financiers
- c'est ce qui n'est pas expliqués par le mouvement conjoncturel

On parle aussi de variations résiduelles.