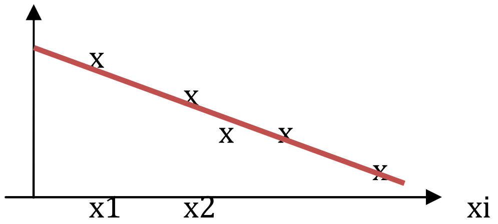
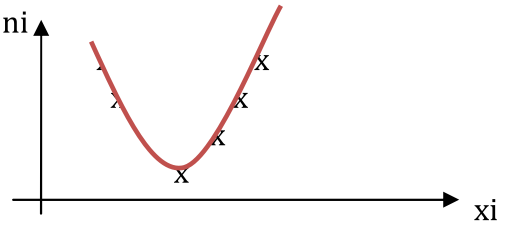
## Chapitre 3 : Ajustement de séries statistiques

On se donne une série statistique de valeurs xi, (i variant de 1 à p), d'effectifs ni.
Cette série peut être discrète ou continue.
On peut représenter ces séries sur des graphes :

Cas 1:
ni(effectif)




ou
Cas 2:
ni




Problème : -identifier la forme de la courbe d'ajustement : droite ou parabole?
-déterminer les équations de la courbe d'ajustement :

$ni = f(xiu) \longrightarrow ni = a.xi + b$
$\longrightarrow ni = a.xi^2 + b.xi + c$

Ceci permet : -de remplacer les 2p valeurs (xi,ni) par seulement 2 ou 3 paramètres
-de calculer des ni par d'autres valeurs de xi

Remarque :
On notera les effectifs par yi pour revenir à la notation classique yi=f(xi)

### 1. Méthode de lissage

Ce n'est pas réellement une méthode d'ajustement car on ne donne pas directement l'équation de la courbe d'ajustement.
Cela permet de réduire les variations de la courbe (xi,ni) pour éventuellement facilité ensuite la détermination de la courbe d'ajustement.
Il existe plusieurs méthodes : nous n'en voyons que deux

#### a. Méthode des points médians

Exemple : On a un phénomène économique qui a été observé sur 20 mois consécutifs.
Méthode : -tracer xi,yi
-déterminer les maximas locaux : A,B,C,D,E et F et on construit l'enveloppe supérieure
-déterminer les minimas locaux : G,H,I,J,K et L et on construit l'enveloppe inférieure
-tracer un segment vertical pour chaque xi
-déterminer les points médians entre les points supérieurs et inférieurs
Les variations de la courbe ainsi obtenues sont plus faibles par comparaison avec la courbe initiale.

Remarque : aspect subjectif selon la détermination des points minimaux et maximaux.

### b. Méthode de la moyenne mobile

Méthode : on va remplace yi par :

$$y'i = \frac{y_{i-1} + yi + y_{i+1}}{3}$$

(Si on effectue une moyenne sur 3 points)
Sauf pour le premier point :

$$y'i = \frac{y1 + y2}{2}$$

et le dernier point :

$$y'p = \frac{y_{p-1} + yp}{2}$$

Sur l'exemple du formulaire correspond au cours III les variations de y'i sont plus faibles que celles de yi.

Remarque : -choix subjectif à faire sur le nombre de points considérés (3,5,7,...)
-possibilité de pondérer ces moyennes :

$$y'i = \frac{1 * y_{i-1} + 2 * yi + 1 * y_{i+1}}{4}$$

## 2. Méthode d'ajustement

### a. Méthode graphique

- Essenciellement pour un ajustement linéaire
- Méthode : on peut tracer une droite sur un papier calque et positionner cette droite pour avoir autant de points au dessus de la droite qu'en dessous. Ensuite, si on connaît deux points de cette droite, on peut en deduire l'équation de cette dernière.

Remarque : méthode subjective.

### b. Méthode des moindres carrés

Méthode :

- Sur la courbe d'ajustement, on note y'i les valeurs ajustées de yi
- L'écart de yi par rapport à cette courbe est dessiné par la quantité suivante:

$$\sum_{i=1}^{p} (yi - y'i)^2$$

- Une fois la forme de la courbe d'ajustement (droite, parabole,exponencielle...), on va chercher les parametres de l'équation considérée qui minimise la quantité $\sum_{i=1}^{p}(yi - y'i)^2$ .

Remarque : - Pas de subjectivité de la méthode (tous les logiciels vont donner le même résultat).

- Notation : $\sum_{i=1}^{p}(yi)$ est notée $\sum_y yi$ et dans certains cas $\sum y$

Rappels : - $\sum_y yi = p\bar{y}$
- $\sum_i (yi - \bar{y})^2 = p\ ox^2 = \sum_i (yi)^2 - \frac{1}{p}\ (\sum_i yi)^2$

On va exposer la méthode sur 4 exemples simples mais la méthode des moindres carrés peut se généraliser à des fonctions plus compliquées.

α) Ajustement à l'aide d'une parabole

Problème : Coube ajustée est une droite d'équation y = ax + b
Ici, on notera y'i = axi + b
On va chercher à minimiser la quantité :

$$\sum_i (yi - axi - b)^2$$

( on veut trouver a et b pour que cette quantité soit la plus petite possible.)
La quantité $\sum_i (yi - axi - b)^2$ dépend de a et de b ⇒ fonction de a et de b ⇒ f (a,b) .

Méthode :

$$\frac{df}{db} = 0$$

⇒ $\sum_i (yi - axi - b) = 0$

⇒ $\sum_i yi = a \sum_i xi + \sum_i b$

⇒ $\sum_i yi = a \sum_i xi + pb \qquad (1)$

On divise (1) par p :

(1) ⇒ $\bar{y} = a\bar{x} + b \qquad (2)$

(2) indique que le point $(\bar{x}, \bar{y})$ appartient à la droite ajustée

$$\frac{df}{db} = 0$$

⇒ $\sum_i (yi - axi - b)\, xi = 0$

⇒ $\sum_i xi.yi = a \sum_i xi^{\,2} + b \sum_i xi \qquad (3)$

$$(3) * p - (1) * (\sum_i xi)$$

⇒ $p \sum_i xi.yi = ap\,(\sum_i xi^2) + pb\ \sum_i xi$

$- \sum_i xi * \sum_i yi = -a\,(\sum_i xi)^2 - pb\ \sum_i xi$

d'où

$$a = \frac{p + \sum_i xi.yi - \sum_i xi * \sum_i yi}{p \sum_i xi^2 - (\sum_i xi)^2} \qquad (4)$$

et $b = \bar{y} - a\bar{x} \qquad (5)$

<u>Autres écritures de a :</u>

$$4 * \frac{\frac{1}{p^2}}{\frac{1}{p^2}}$$

⇒

$$a = \frac{\frac{1}{p} + \sum_i xi.yi - \frac{\sum_i xi}{p} * \frac{\sum_i yi}{p}}{\frac{1}{p}\sum_i xi^2 - \left(\frac{\sum_i xi}{p}\right)^2}$$

$$a = \frac{\bar{x}\bar{y} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} \qquad (6)$$

<u>Autres façons d'écrire a :</u>

$$\sum_i (xi - \bar{x})(yi - \bar{y}) = \sum_i (xi.yi - xi.\bar{y} - \bar{x}.yi + \bar{x}\bar{y})$$

$$\sum_i (xi.yi) - \underbrace{\bar{y} \sum_i (xi)}_{[\ \ -p\frac{1}{x}\bar{y}\ \ ]} - \bar{x} \sum_i (yi) + \bar{x}\bar{y}p$$

$$= \sum_i (xi.yi) - \frac{1}{p}\sum_i (xi) - \bar{x} \sum_i (yi) * \sum_i (xi - \bar{x}^2)$$

Au plus haut :

$$= \sum_i (xi^2) - \frac{1}{p}\left(\sum_i (xi)\right)^2$$

alors d'apres (4)

$$a = \frac{p + \sum_i (xi - \bar{x})(yi - \bar{y})}{p \sum_i (xi - \bar{x})^2}$$

$\Rightarrow$

$$a = \frac{\sum_i (xi - \bar{x})(yi - \bar{y})}{\sum_i (xi - \bar{x})^2} \qquad (4)$$

- Qualité de l'ajustement :

$$\sum_i (yi - \bar{y})^2 = \sum_i (yi - y'i + y'i - \bar{y})^2 = \sum_i (yi - y'i)^2 + \sum_i (y'i - \bar{y})^2 + 2\underbrace{\sum_i (yi - y'i)(y'i - \bar{y})}$$

$\Rightarrow$ et on va demontrer que $\Delta=0$

$$\Delta = \sum_i (yi - y'i)(y'i - \bar{y})$$

$$\Delta = \sum_i (yi - axi - b)\big(axi + (b - \bar{y})\big)$$

$$\Delta = a\sum_i (yi - axi - b)xi + (b - \bar{y})\sum_i (yi - axi - b) - \frac{df(a,b)}{db} = 0$$

$\Rightarrow \Delta=0 \qquad \frac{df(a,b)}{db} = 0$

donc

| $\sum_i (yi - \bar{y})^2$ | $=$ | $\sum_i (yi - y'i)^2$ | $+$ | $\sum_i (y'i - \bar{y})^2$ |
|---|---|---|---|---|
| Cela represente l'écart total entre yi et $\bar{y}$ | | Ecart résiduel de yi par rapport à la valeur estimée yi | | Ecart expliqué entre yi=axi+b et $\bar{y}$.Il ne dé- -pend que de xi,pas des yi |

d'où le coefficient de corrélation:

$$R^2 = \frac{\text{écart expliqué}}{\text{écart total}} = \frac{\sum_i (y'i - \bar{y})^2}{\sum_i (yi - \bar{y})^2}$$

et $R^2$ varie de 0 à 1 .
Si ajustement idéal : écart résiduel = $\sum_i yi - y'i = 0$
donc $\sum_i (yi - \bar{y})^2 = \sum_i (y'i - \bar{y})^2$
donc $R^2 = 1$

<u>Autres expressions de R :</u>

Ecart expliqué :

$$\sum_i (y'i - \bar{y})^2 = \sum_i (axi + b - a\bar{x} - b)^2 = a^2 \sum_i (xi - \bar{x})^2$$

donc

$$R^2 = \frac{a\sum_i (xi - \bar{x})^2}{\sum_i (yi - \bar{y})^2} = R$$

$$R = |a|\sqrt{\frac{\sum_i (xi - \bar{x})^2}{\sum_i (yi - \bar{y})^2}} = |a|\frac{\sigma x}{\sigma y}$$

Exemple :
Ajustement à l'aide d'une droite.

| | xi | yi | $xi^2$ | xiyi | $yi^2$ |
|---|---|---|---|---|---|
| Données experimentales | 2<br>3<br>.<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. |
| Somme : | 25 | 26 | 151 | 163 | 182 |

Formule pour calculer a, b et R

$$a=\frac{p\sum xy-\sum x\sum y}{p\sum x^2-(\sum x)^2}=\frac{5.163-25.26}{5.151-25^2}=\frac{33}{26}=1{,}2692$$

$$b=\frac{\sum y}{p}-a\frac{\sum y}{p}=\frac{26}{5}-1{,}2692*\frac{25}{5}=-\frac{149}{130}=-1{,}1462$$

Valeurs obtenues sont identiques à celles obtenues avec un logiciel

$$R=\frac{\sum xy-\frac{1}{p}\sum x\sum y}{\sqrt{\sum x^2-\frac{1}{p}(\sum x)^2\left(\sum y^2-\frac{1}{p}(\sum y^2)\right)}}=\frac{\left(163-\frac{1}{5}*25*26\right)}{\sqrt{\left(151-\frac{1}{5}*25^2\right)\left(182-\frac{1}{5}*26^2\right)}}=\frac{11\sqrt{5}}{26}$$

$$=0{,}9460$$

β) <u>Ajustement à l'aide d'une parabole</u>

Courbe ajustée est une parabole d'équation, $y = ax^2 + bx + c$
Avec notre notation, on aura:
$y'i = axi^2 + bxi + c$
On va chercher le minimum de $\sum_i (yi-y'i)^2$
Ici, $\sum_i (yi-y'i)^2 = \sum_i (yi-axi^2-bxi+c)^2 = f(a,b,c)$

Méthode:

$$\frac{df}{dc}=0 \;\; => \; -2\sum_i (yi-axi^2-bxi+c)^2=0$$

$$=> \sum_i yi = a\sum_i xi^2 + b\sum_i xi + \sum_i c \qquad (1)$$

$$=> \bar{y} = a\overline{x^2} + b\bar{x} + c$$

$$\sum_i yi = a\sum_i xi^2 + b\sum_i xi + pc \qquad (1)$$

$$\frac{df}{dc}=0 \;\; => \; -2\sum_i (yi-axi^2-bxi+c)^2=0$$

$$=> \sum_i xiyi = a\sum_i xi^3 + b\sum_i xi^2 + c\sum_i xi \qquad (2)$$

$$\frac{df}{dc}=0 \;\; => \; -2\sum_i (yi-axi^2-bxi+c)xi^2=0$$

$$=> \sum_i xi^2yi = a\sum_i xi^4 + b\sum_i xi^3 + c\sum_i xi^2 \qquad (3)$$

On reprend le formulaire:

| | xi | yi | $xi^2$ | $xi^3$ | $xi^4$ | xiyi | $xi^2yi$ | $yi^2$ |
|---|---|---|---|---|---|---|---|---|
| Données experimentales | .<br>.<br>.<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. |
| Somme : | 18 | 25 | 88 | 504 | 3124 | 100 | 560 | 155 |

d'où :

$$\left.\begin{array}{l}(1): 25 = 88a + 18b + 5c \\ (2): 100 = 504a + 88b + 18c \\ (3): 560 = 3124a + 504b + 88c\end{array}\right\}$$

avec (1) et (2) $\Rightarrow 936a + 116b = 50$
(1) et (3) $\Rightarrow 7876a + 936b = 600$

$\Rightarrow a = \frac{285}{469} = 0{,}6077$
$\Rightarrow b = \frac{4195}{938} = -4{,}4723$
$\Rightarrow c = \frac{4880}{469} = 10{,}4051$

Calcul du coefficient de correlation dans le cas d'une parabole :
On a vu que $R^2 = \frac{\sum_i (y'i - \bar{y})^2}{\sum_i (yi - \bar{y})^2} = \frac{N}{D}$
On a : $D = \sum_i yi^2 - \frac{1}{p}(\sum yi)^2$ Déjà vu
On a

$$N = \sum_i (y'i - \bar{\bar{y}})^2 = \sum_i \left(ax + bxi + c - a\,\overline{x^2} - b\bar{x} - c\right)^2$$
$$= a^2 \sum_i \left(xi^2 - \overline{x^2}\right)^2 + b^2 \sum_i (xi - \bar{x})^2 + 2ab \sum_i \left(xi^2 - \overline{x^2}\right)\ (xi - \bar{x})$$

Dans l'exemple retenu :
On aura:

$$N = (0{,}6077)^2 \left[3124 - \frac{1}{5} * 88\right] + (4{,}4723)^2 \left[88 - \frac{1}{5} * 18^2\right] - 2 * 0{,}6077 * 4{,}4723 \left[504 - \frac{1}{5} * 18 * 88\right] = \cdots = 28{,}203$$

et

$$D = 155 - \frac{1}{5} * 25^2 = 30$$

donc

$$R^2 = \frac{N}{D} = 0{,}9401 \ \ => \ \ R = 0{,}9695$$

<u>δ) Ajustement à l'aide d'unee exponentielle : $y=ba^x$</u>

Si $y=b\ ba^x \Rightarrow \log y = \log b + x \log a = x \log a + \log b$
ou encore $y=Ax + B$
où : $Y=\log(y)$
$A=\log(a)$
$B=\log(b)$
On aura:

$$A = p \sum xiYi - \frac{\sum xi \sum Yi}{p \sum_i xi^2\ (\sum_i xi)^2}$$

et B=$\bar{Y} - A\bar{x}$

δ) Ajustement à l'aide d'une fonction puissance y= bx$^a$

donc log y = a log x + log b
⇒ Y=aX + B
où : Y=log (y)
A=log (a)
B=log (b)
et on aura:

$$a = \frac{p\sum_i XiYi - \sum_i Xi \sum_i Yi}{p\sum_i Xi^2 - (\sum_i Xi)^2}$$

et B=$\bar{Y} - a\bar{X}$

# Chapitre 4 : Indices statistiques

Ces indices statistiques servent à suivre l'évolution d'une grandeur au cours du temps
-indice des prix à la consommation (IPC)
-indice de la production (IPI)
-indice ddu coût de la construction (ICC)
On distingue deux notions d'indice:
-indice simple si une seule grandeur est étudiée
-indice synthétique si plusieurs grandeurs sont étudiées

## 1. Indice simple (ou élémentaire)

On va partir d'un exemple :
Si un matériel coûte 50€ en 2004 et 54€ en 2005, on peut comparer ces prix avec :
-l'augmentation : 54-50 = 4€ mais cela ne donne pas l'importance de l'augmentation
-l'augmentation en pourcent : $\frac{54-50}{100} = 0{,}08 = 8\%$
-le rapport des prix : $\frac{54}{50} = 1{,}08$

Interêt de ce rapport : calcul facile pour estimer le prix final (50 * 1,08 = 54)
D'où la définition de l'indice simple

$$\frac{54}{50} * 100 = 108$$

Définition :
Soit p0 la valeur a une date t0 (de base ou reférence)d'un grandeur.
Soit p1 la valeur a une date t1 d'un grandeur.
Alors l'indice simple de la grandeur à la date t1 calculé sur la base 100 à la date t0 est :

$$I_{Y0} = \frac{p1}{p0} * 100$$

Remarques :
-on multiplie par 100 pour éviter les décimales
-on note $i_{1/0} = \frac{p1}{p0}$ avec i en minuscule

Propriétés :
- un indice simple est reversible :

on a $i_{0/t} * i_{t/0} = 1$
ou encore :

$$\frac{i_{0/t}}{100} = \frac{1}{\frac{i_{t/0}}{100}}$$

-indice transfèrable:
qlq soit t0,t1,t2 : $i_{2/1} * i_{1/0} = i_{2/0}$
ou

$$\frac{i_{2/1}}{100} * \frac{i_{1/0}}{100} = \frac{i_{2/0}}{100}$$

Remarque:
2 hausses de suite de 8% donnent une hausse de 16,64% et par de 16% car :
1,08 * 1,08 = 1,1664.