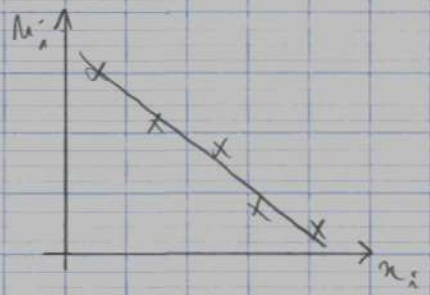
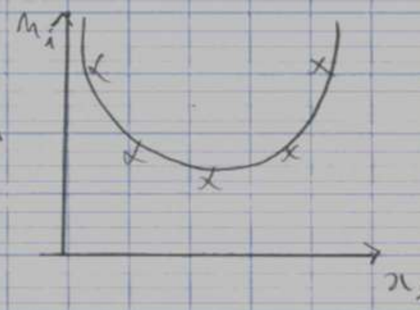
Chapitre III : Ajustement de séries statistiques

On se donne une série stat de valeurs x_i (i variant de 1 à p), d'effectifs n_i.
Cette série peut ê discrète ou continue
On peut représenter ces séries sur des graphes:



ou



Problème: identifier la forme de la courbe d'ajustement: 1/ droite
2/ parabole.

- déterminer les équations de la courbe d'ajustement:

$$n_i = f(x_i) \rightarrow n_i = a x_i$$
$$n_i = a x_i^2 + bx + c$$
$$\ldots$$

Ceci permet | de remplacer les $2p$ valeurs (x_i, n_i) par seulement 2 ou 3 paramètres.
| de calculer des n_i pour d'autres valeurs de x_i

(R) On notera les effectifs par y_i pour revenir à la notation classique
$$y_i = f(x_i)$$

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 (x_i, n_i) pour éventuellement faciliter ensuite, la détermination de la courbe d'ajustement.

a) Méthode des points médians

ex : On a un phénomène économique qui a été observé sur 20 mois consécutifs

méthode : - tracer x_i, y_i

- déterminer les maxima locaux : A, B, C, D, E et F et on construit l'enveloppe sup.
- idem avec les minima.
- tracer un segment vertical pour chaque x_i
- déterminer les points médians entre les points supérieurs et inférieurs.

Les variations de la courbe ainsi obtenue sont plus faibles par comparaison avec la courbe initiale.

(R) aspect subjectif selon la déterminat° des pts max et min.

b) Méthode de la moyenne mobile

méthode : on va remplacer y_i par $y'_i = \frac{y_{i-1} + y_i + y_{i+1}}{3}$

(si on effectue une moyenne sur 3 pts) sauf pour le premier point : $y'_1 = \frac{y_1 + y_2}{2}$

et le dernier point : $y'_p = \frac{y_{p-1} + y_p}{2}$

Sur l'exemple du formulaire correspondant au cours III
Les variations de y'_i sont plus faibles que celles de y_i

(R) • choix subjectif à faire sur le nombre de points considérés (3, 5, 7, ...)
• possibilité de pondérer ces moyennes : $y'_i = \frac{y_{i-1} + 2y_i + y_{i+1}}{4}$

2) Méthodes d'ajustement

a) Méthode graphique

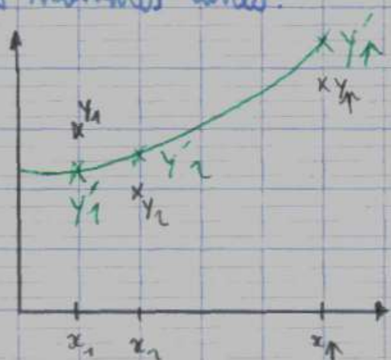
• Essentiellement pour un ajustement linéaire

• méthode : on peut tracer une droite sur un papier calque et positionner cette droite pour avoir autant de point au dessus de la droite qu'en dessous.
Ensuite, si on connait 2 pts de cette droite, on peut en déduire son équation

⊖ méthode subjective

b) Méthode des moindres carrés.

• méthode :



Sur la courbe d'ajustement, on note y'_i les valeurs ajustées de y_i.
L'écart des y_i par rapport à cette courbe est donné par la quantité suivante :

$$\sum_{i=1}^{n} (y_i - y'_i)^2$$

• Une fois choisie la forme de la courbe d'ajustement (droite, parabole, exp, ...)
On va chercher les paramètres de l'équation considérée et qui minimise la quantité :

$$\sum_{i=1}^{n} (y_i - y'_i)^2$$

⊕ - pas de subjectivité de la méthode (les logiciels donnent le même résultat).
- notation : $\sum_{i=1}^{n} y_i$ est noté $\sum_i y_i$ et les autres ou $\sum y$

Rappels : $\sum_i y_i = n\bar{y}$ $\quad$ $\sum_i (y_i - \bar{y})^2 = n\sigma_y^2 = \sum_i (y_i^2 - 2\bar{y}y_i + \bar{y}^2)$

$$= \sum_i y_i^2 - 2\bar{y}\sum_i y_i + \bar{y}^2 \sum_i 1$$

$$= \sum_i y_i^2 - 2n\bar{y}^2 + n\bar{y}^2 = \sum_i y_i^2 - n\bar{y}^2 = \sum_i y_i^2 - \frac{1}{n}\left(\sum_i y_i\right)^2$$

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

α) ajustement à l'aide d'une droite

pb : · Courbe ajustée est une droite d'équation $y = ax + b$

Ici on notera $y'_i = ax_i + b$

· On va chercher à minimiser la quantité : $\sum_i (y_i - ax_i - b)^2$

(on veut trouver a et b pour que cette quantité soit la + petite possible).

La quantité $\sum_i (y_i - ax_i - b)^2$ dépend de a et b $\Rightarrow f(a,b)$

méthode : $\frac{\partial f}{\partial b} = 0 \Rightarrow -2\sum_i (y_i - ax_i - b) = 0$

$\sum_i y_i = a\sum_i x_i + nb$ ...... (1)

On divise (1) par n : $\quad \bar{y} = a\bar{x} + b$ ...... (2)

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

$\frac{\partial f}{\partial a} = 0 \Rightarrow -2\sum_i (y_i - ax_i - b)x_i = 0 \Rightarrow \sum_i x_i y_i = a\sum x_i^2 + b\sum x_i$ ...... (3)

$(3)\times n - (1)\times\left(\sum_i x_i\right) \Rightarrow \begin{cases} n\sum_i x_i y_i = a n \sum x_i^2 + nb\sum x_i \\ -\sum_i x_i \sum_i y_i = -a\left(\sum_i x_i\right)^2 - nb\sum x_i \end{cases}$

d'où $a = \dfrac{n \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n \sum_i x_i^2 - \left(\sum_i x_i\right)^2}$ .... (4)

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

autres écritures de a : $(4) \times \dfrac{1/n^2}{1/n^2} \Rightarrow a = \dfrac{1/n \sum_i x_i y_i - \frac{\sum x_i}{n} \frac{\sum y_i}{n}}{1/n \sum_i x_i^2 - \left(\frac{\sum x_i}{n}\right)^2}$

$$a = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - (\bar{x})^2} \quad \ldots\ldots (6)$$

OU BIEN : $\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y})$

$$= \sum_i x_i y_i - \overbrace{\bar{y} \sum x_i}^{= n\bar{x}\bar{y}} - \bar{x} \sum_i y_i + \bar{x}\bar{y}\,n$$

$$= \sum_i x_i y_i - \frac{1}{n} \sum_i x_i \sum_i y_i$$

• $\sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - \frac{1}{n} \left(\sum x_i\right)^2 \leftarrow$ vu plus tôt.

donc d'après (4) : $a = \dfrac{n \sum_i (x_i - \bar{x})(y_i - \bar{y})}{n \sum_i (x_i - \bar{x})^2} \Rightarrow a = \dfrac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$ ....(7)

Qualité de l'ajustement.

$$\sum_i (y_i - \bar{y})^2 = \sum_i (y_i - y'_i + y'_i - \bar{y})^2 = \sum_i (y_i - y'_i)^2 + \sum (y'_i - \bar{y})^2 + \overbrace{2 \sum_i (y_i - y'_i)(y'_i - \bar{y})}^{= \Delta}$$

On va montrer que $\Delta = 0$:

$$\Delta = \sum_i (y_i - y'_i)(y'_i - \bar{y}) = \sum (y_i - a x_i - b)(a x_i + (b - \bar{y})) = a \sum_i (y_i - a x_i - b) x_i$$

$$\hookrightarrow + \sum (b - \bar{y})(y_i - a x_i - b)$$

$$\frac{\partial f(a,b)}{\partial a} = 0 \qquad \frac{\partial f(a,b)}{\partial b} = 0$$

$$\Delta = a \sum_i (y_i - a x_i - b) x_i + (b - \bar{y}) \sum_i (y_i - a x_i - b) = 0$$

donc $\boxed{\sum_i (y_i - \bar{y})^2} = \boxed{\sum_i (y_i - y'_i)^2} + \boxed{\sum_i (y'_i - \bar{y})^2}$

cela représente l'écart total entre y_i et $\bar{y}$

écart résiduel de y_i par rapport à la valeur estimée y'_i

écart expliqué entre $y'_i = a x_i + b$ et $\bar{y}$. Il ne dépend que de x_i et pas des y_i

d'où le coefficient de corrélation :

$$R^2 = \frac{\text{écart expliqué}}{\text{écart total}} = \frac{\sum_i (y'_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$ et R^2 de 0 à 1

L'ajustement est idéal si l'écart résiduel est nul et donc que $R^2 = 1$

autre expression de R^2 :

écart expliqué : $\sum_i (y'_i - \bar{y})^2 = \sum_i (a x_i + \cancel{b} - a \bar{x} - \cancel{b})^2 = a^2 \sum_i (x_i - \bar{x})^2$

donc $R^2 = \dfrac{a^2 \sum_i (x_i - \bar{x})^2}{\sum_i (y_i - \bar{y})^2} \Rightarrow R = |a| \sqrt{\dfrac{\sum_i (x_i - \bar{x})^2}{\sum_i (y_i - \bar{y})^2}} = |a| \dfrac{\sigma_x}{\sigma_y}$

OU ENCORE : à partir de (7) : $R = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$

Exemple : Ajustement à l'aide d'une droite (voir formulaire cours 4)

Formule pour calculer a, b, k

$$a = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{5 \times 163 - 25 \times 26}{5 \times 151 - (25)^2} = \frac{33}{26} = 1{,}2692$$

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

Les valeurs obtenues sont identiques à celles obtenues avec un logiciel.

$$R = \frac{\sum xy - \frac{1}{n}\sum x_i \sum y}{\sqrt{\left(\sum x^2 - \frac{1}{n}(\sum x)^2\right)\left(\sum y - \frac{1}{n}(\sum y)^2\right)}}$$

$$= \frac{163 - \frac{1}{5} \times 25 \times 26}{\left[\left(151 - \frac{1}{5} \times 25^2\right)\left(182 - \frac{1}{5} \times 26^2\right)\right]^{1/2}} = \frac{11\sqrt{5}}{26} = 0{,}94603$$

β) ajustement à l'aide d'une parabole

· La courbe ajustée est une parabole d'équation, $y = ax^2 + bx + c$
Avec notre notation, on aura : $y'_i = ax_i^2 + bx_i + c$

On va rechercher le minimum de $\sum_i (y_i - y'_i)^2$

Ici, $\sum_i (y_i - y'_i)^2 = \sum_i (y_i - ax_i^2 - bx_i - c)^2 = f(a, b, c)$

méthode : $\frac{\partial f}{\partial c} = 0 \Rightarrow -2\sum_i (y_i - ax_i^2 - bx_i - c) = 0$

$$\sum_i y_i = a \sum_i x_i^2 + b \sum_i x_i + \sum_i c \quad \ldots\ldots (1)$$

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

$$\frac{df}{db} = 0 \Rightarrow -2 \sum_i (y_i - a x_i^2 - b x_i - c)\, x_i = 0$$

$$\boxed{\sum_i x_i y_i = a \sum_i x_i^3 + b \sum_i x_i^2 + c \sum_i x_i} \quad \ldots (2)$$

$$\frac{df}{da} = 0 \Rightarrow -2 \sum_i (y_i - a x_i^2 - b x_i - c)\, x_i^2 = 0$$

$$\boxed{\sum_i x_i^2 y_i = a \sum_i x_i^4 + b \sum_i x_i^3 + c \sum_i x_i^2} \quad \ldots (3)$$

exemple: fontaine:

(1) $25 = 88a + 18b + 5c$

(2) $100 = 504a + 88b + 18c$

(3) $560 = 3124 + 504b + 88c$

avec (1) et (2) $936a + 116b = 50$

(1) et (3) $7876a + 936b = 600$

d'où $a = \frac{285}{469} = 0,6077$ $\quad b = -\frac{4195}{938} = -4,4723$ et $c = \frac{4880}{469} = 10,405$

$$R^2 = \frac{\sum (y'_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{N}{D} \qquad \text{On a } D = \sum_i y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2$$

On a $N = \sum_i (y'_i - \bar{y})^2 = \sum_i (a x_i^2 + b x_i + c - a\overline{x^2} - b\bar{x} - c)^2$

$$\Rightarrow N = a^2 \sum_i (x_i^2 - \overline{x^2})^2 + b^2 \sum (x_i - \bar{x})^2 + 2ab \sum (x_i^2 - \overline{x^2})(x_i - \bar{x})$$

$$N = a^2\left[\sum x_i^4 - \frac{1}{n}\left(\sum x_i^2\right)^2\right] + b^2\left[\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2\right] + 2ab\left[\sum x_i^3 - \frac{1}{n}\sum x_i \left(\sum x_i\right.\right.$$

Dans l'exemple on aura $N = 28{,}203$ et $D = 155 - \frac{1}{5} \times 25^2 = 30$

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

γ) Ajustement à l'aide d'une exponentielle : $y = ba^x$

ici : $y = ba^x \Rightarrow \log y = \log b + x \log a = x \log a + \log b$

on tire $Y = Ax + B$ où $Y = \log y$ $B = \log b$ $A = \log a$

On aura : $A = \dfrac{n \sum x_i Y_i - \sum x_i \sum Y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}$ et $B = \bar{Y} - A\bar{x}$

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

$y = bx^a$ $\log y = a \log x + \log b$

$Y = aX + B$ où $Y = \log y$ $X = \log x$ $B = \log |b|$

Chapitre IV : Indices statistiques

Les indices statistiques servent à suivre l'évolut° d'une grandeur au cours du temps.

- indice des prix à la consommation (IPC)
- indice de la product° industrielle (IPI)
- " du coût de la construct° (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 part d'un exemple:
Si 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'↗ en % $\frac{54-50}{50} = 8\%$
- le rapport des prix : $\frac{54}{50} = 1{,}08$

L'intérêt de ce rapport: calcul facile pour estimer le prix final ($50 \times 1{,}08 = 54$)

- d'où la définition de l'indice simple $\frac{54}{50} \times 100 = 108$

déf: Soit p_0 la valeur à une date t_0 (de base ou référence) d'une grandeur.
Soit p_1 " " t_1

Alors l'indice simple de la grandeur à la date t_1 calculé sur la base 100 à la date t_0 est $\boxed{I_{1/0} = \frac{p_1}{p_0} \times 100}$

R - On multiplie par 100 pour éviter les décimales
- On note $i_{1/0} = \frac{p_1}{p_0}$ avec i en minuscule.

propriétés : un indice simple est réversible : $i_{1/0} \times i_{0/1} = 1$

on écrit $\frac{I_{0/1}}{100} = \frac{1}{\frac{I_{1/0}}{100}}$

- indice transférable : $\forall t_0, t_1, t_2$: $i_{2/1} \times i_{1/0} = i_{2/0}$

R 2 hausses de suite de 8% donne une hausse de 16,64% et pas de 16%.
car $1,08 \times 1,08 = 1,1664$