Chp. La regression linéaire simple

I/ Introduction

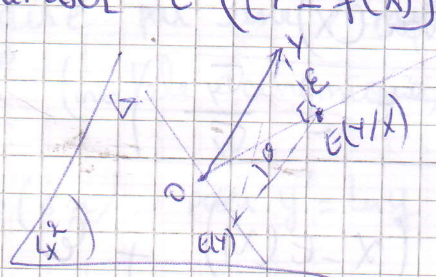
Etant donné un couple de v.a numerique (X, Y). Si X et Y ne sont pas indépendantes, la connaissance de la valeur prise par X change notre incertitude conservant la realisation de Y: elle la diminue generalement, car la variance de la distribution de $Y/X=x$ est en moyenne inferieure à la variance de Y : $E(V(Y/X)) \leq V(\cancel{X})$ (Theorème de la var totale)
Lorsqu'on peut admettre que X peut servire pour ~~expr~~ predir Y on doit alors chercher une formule de predition de Y par X de type de $\hat{Y} = f(Y)$ sans biais $E(\hat{Y} - Y) = 0$, ainsi qu'à evaluer l'ordre de grandeur de l'erreur de prevision qui est mesurée par $Var(\varepsilon) = Var(Y - \hat{Y})$. On cherche bien sûr à minimiser $Var(\varepsilon)$

Dans ce chp nous cherchons la formule de prevision "idéale" au sens des moindres carrés", particulierement le cas où cette formule est linéaire avec un ecart type conditionnelle C^{te}
$\sigma(\varepsilon / X = x) = \sigma < \infty$ dit homocidacité
X est dite variable explicative ou predicteur
Y est dite variable à expliquer ou critère

II) Methode de regression simple

II.1) Approximation conditionnelle

Considerons deux v.a Y et X, pour trouver la meilleure formule qui est le plus proche possible de Y en moyenne quadratique revient à minimiser $E([Y - f(X)]^2)$



L^2_X est l'espace des v.a de type $f(X)$
$E(Y/X)$ est la projection ortho de Y sur L^2_X
Δ des C^{te}

La qualité de l'approximation de Y par $E(Y/X)$ est mesurée par le rapport de

$$\eta^2_{Y/X} = \frac{V_a(E(Y/X))}{V_a(Y)} = \frac{\text{variance expliquée}}{\text{variance totale}} = \cos^2\theta \quad *$$

Pour une valeur x de X ; on associe $E(Y/X=x)$ dite fonction de regression de Y en X son graphe est la courbe de regression de Y en X

on peut donc poser $Y = E(Y/\cancel{X=x}) + \varepsilon$

où ε est un residu ; v.a pas tjs neglegeable.

on a $E(Y) = E(E(Y/X)) \Rightarrow E(\varepsilon) = 0$

De plus ε est non correlee lineairement avec X et avec $E(Y/X)$ car ε est orthogonale à L^2_X.

La variance de ε, ou variance residuelle est alors $V(\varepsilon) =$

$$Var(\varepsilon) = Var(Y) - Var(E(Y/X))$$

et d'après $(*)$ $Var(E(Y/X)) = \eta^2_{Y/X} \cdot Var(Y)$

$$V(\varepsilon) = Var(Y) - \eta^2_{Y/X} \cdot Var(Y) = (1 - \eta^2_{Y/X}) Var(Y)$$

II.2) Cas de la régression linéaire

En générale, en pratique, on utilise le cas où $E(Y/X) = \alpha + \beta X$ (ce qui se produit en particulier si X et Y $\sim$ une loi normale à deux dimension)

$$\text{on a donc } Y = \alpha + \beta X + \varepsilon \quad (1)$$

$$E(Y) = \alpha + \beta E(X) \quad (2)$$

La droite de régression passe donc par le p^t $(E(X); E(Y))$

(1)-(2) on a $Y - E(Y) = \beta(X - E(X)) + \varepsilon$ (3)

$$\Rightarrow (Y - E(Y))(X - E(X)) = \beta(X - E(X))^2 + \varepsilon(X - E(X))$$

et $E\big((Y - E(Y))(X - E(X))\big) = \beta E(X - E(X))^2 + \underbrace{E(\varepsilon)}_{0} \cdot E(X - E(X))$

$$\Leftrightarrow \operatorname{cov}(Y, X) = \beta \, Var(X)$$

$$\Leftrightarrow \boxed{\beta = \frac{\operatorname{cov}(Y,X)}{Var(X)} = \rho \frac{\sigma_Y}{\sigma_X}}$$

(3) devient $Y - E(Y) = \rho \frac{\sigma_Y}{\sigma_X}(X - E(X)) + \varepsilon$

ce qui fait $Y = E(Y) + \rho \frac{\sigma_Y}{\sigma_X}(X - E(X)) + \varepsilon$

$$Var(Y) = Var\left(E(Y) + \rho \cdot \frac{\sigma_Y}{\sigma_X}(X - E(X)) + \varepsilon\right)$$

$$= \rho^2 \frac{\sigma_Y^2}{\sigma_X^2} Var((X - E(X))) + Var(\varepsilon)$$

$$= \rho^2 \frac{\sigma_Y^2}{\sigma_X^2} Var(X) + Var(\varepsilon)$$

$$= \rho^2 \sigma_Y^2 + (1 - \eta^2_{Y/X}) Var(Y)$$

$$(=) \quad 1 = \rho^2 + 1 - \eta^2_{Y/X}$$

$$(=) \quad \boxed{\rho^2 = \eta^2_{Y/X}}$$

II-3 / Ajustement de la regression lineaire par des données experimentales

Soient n couple de la v.a (X, Y) : (x_i, y_i) $i = 1, .., n$ echantillon de (X, Y)

On suppose qu'on $E(Y/X) = \alpha + \beta X$

Le pb est donc d'estimer les coefficients reels α et β ainsi que la variance des residus σ^2

Rq Cette methode s'applique aussi lorsque la v. X n'est pas aleatoire mais contrôlée par l'experimentateur.

comme exp, mesurer la gradeur Y à des instants x_i fixés

Y est aleatoire grâce aux erreurs ε mais X non aleatoire.

on a alors $y_i = \cancel{\alpha x_i}\ a + b x_i + \varepsilon_i$ $i = \overline{1, n}$

où ε_i sont des realisations indépendantes de moyenne nulle et de variance $\sigma^2 < \infty$. On parle alors de modele lineaire plutôt que de régression lineaire.

Rqs : • on parle de correlation entre Y et X que lorsque X est aleatoire

• De nombreux modeles non lineaires se ramene facilement au modeles lineaire par une simple transformation tel que

- $y \# \alpha x^\beta$ (modele economique) en passe Ln $y' = Ln y = Ln \alpha + \beta Ln x = \alpha' + \beta x'$

- $y \simeq \alpha \exp(\beta x)$ poser $y' = Ln y \Rightarrow y' = Ln \alpha + \beta x$. modele croissance expo

$- \; y = \dfrac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$ on pose $y' = \ln \dfrac{y}{1-y}$ modèle logistique

alors $y' = \alpha + \beta x$.

En effet, on a $\dfrac{y}{1-y} = \exp y' \iff \dfrac{\exp(\alpha+\beta x)}{[1+\exp(\alpha+\beta x)]}$

$y = (1-y)\exp(y') \iff y + y\exp(y') = \exp(y')$

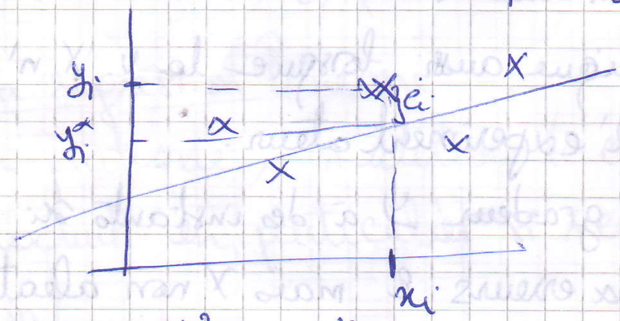
$\iff y = \dfrac{e^{y'}}{1+e^{y'}}$

D'où, $\dfrac{e^{y'}}{1+e^{y'}} = \dfrac{e^{\alpha+\beta x}}{1+\exp(\alpha+\beta x)}$

$y' = \alpha + \beta x$.

a/ Estimation des paramètres α, β et σ^2

On utilise le critère des moindres carrés, sachant que $E(Y/X) = \alpha + \beta x$ est la meilleure approximation de Y par X en moyenne quadratique. On cherche donc à ajuster le nuage (x_i, y_i) par une droite d'équation $y^* = \hat{a} + \hat{b}x$ de sorte que $\sum_{i=1}^{n}(y_i - y_i^*)^2$ soit minimal.



$\Delta(a,b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - y_i^*)^2 = \sum_{i=1}^{n}(y_i - a - bx_i)^2$

$\min \Delta(a,b) \Rightarrow \begin{cases} \dfrac{\partial \Delta(a,b)}{\partial a} = 0 \\ \dfrac{\partial \Delta(a,b)}{\partial b} = 0 \end{cases}$

La solution

$\hat{b} = \dfrac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = r\dfrac{S_y}{S_x} = \dfrac{\operatorname{cov}(x,y)}{S_x^2}$

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

$$\boxed{y^* = \bar{y} + r\frac{S_y}{S_x}(x - \bar{x})}$$

Rq : La droite des moindres carrés passe par le centre de gravité du nuage $G = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$ et sa pente est l'analogue "valeur" empirique de la pente de droite de regression $\rho \frac{\sigma_y}{\sigma_x}$

Théorème 1

$\hat{a}$, $\hat{b}$, et y^* sont des estimateurs sans biais de α, β et $E(Y/X=x) = \alpha + \beta x$

$\hat{b} = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ est la realisation de la v.a :

$$B = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

$$E(B/X=x_i) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})E(Y_i-\bar{Y}/X_i=x_i)}{\sum_{i=1}^{n}(x_i-\bar{X})^2}$$

comme $E(Y/X=x_i) = \alpha + \beta x_i$
$E(\bar{Y}/X=x_i) = \alpha + \beta\bar{x}$ $\Rightarrow E(Y_i - \bar{Y}/X_i = x_i) = \beta(x_i - \bar{x})$

Alors

$$E(B/X=x_i) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\beta(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \beta\frac{\sum(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2} = \beta$$

D'où $E(E(B/X_i=x_i)) = E(B) = \beta$

comme $a = \bar{y} - b\bar{x}$ donc a est la realisation de la v.a $A = \bar{Y} - B\bar{X}$ et par le même procédé

$$E(A/X_i=x_i) = E(\bar{Y}/X=x_i) - E(B\bar{X}/X_i=x_i)$$
$$= \alpha + \beta\bar{x} - \bar{x}E(B/X=x_i)$$
$$= \alpha + \beta\bar{x} - \bar{x}\beta = \alpha.$$

Et $E(E(A/X_i=x_i)) = E(A) = \alpha$

comme $E(Y/X=x) = \alpha + \beta x$ et $y^* = \hat{a} + \hat{b}x$

~~$E(Y) = E(\alpha+\beta x) = \alpha + \beta x$~~ $E(y^*) = E(\hat{a}) + E(\hat{b}).x = \alpha + \beta x$

D'où $\alpha + \beta x$ estimateur sans biais de y^*

Proposition

B non corelé avec $\bar{Y}$ en effet,

$$\hat{b} = \frac{\sum_1^n(x_i-\bar{x})(y_i-\bar{y})}{\sum_1^n(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} - \frac{\bar{y}\overbrace{\sum_{i=1}^{n}(x_i-\bar{x})}^{0}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\mathrm{cov}(B,\bar{Y}) = \mathrm{cov}\left(\frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}, \bar{Y}\right)$$

$$= \frac{1}{\sum_1^n(x_i-\bar{x})^2}\sum_{i=1}^n(x_i-\bar{x})\,\mathrm{cov}(y_i,\bar{y})$$

$$= \frac{1}{\sum(x_i-\bar{x})^2}\sum_{i=1}^n(x_i-\bar{x})\,\mathrm{cov}\left(y_i ; \frac{1}{n}\sum_{j=1}^n y_j\right)$$

$(y_1, \ldots, y_n$ indep^{ts} alors $\mathrm{cov}(y_i,y_j)=0$ sauf pour $i=j$

$$\mathrm{cov}(B,\bar{Y}) = \frac{1}{\sum_1^n(x_i-\bar{x})^2}\sum_{i=1}^n(x_i-\bar{x})\frac{1}{n}\mathrm{cov}(y_i,y_i)$$

$$= \frac{\sigma^2}{n\sum_{i=1}^n(x_i-\bar{x})^2}\underbrace{\sum_{i=1}^n(x_i-\bar{x})}_{0} = 0$$

D'où B et $\bar{Y}$ non corrélés.

<u>Théorème de Gauss - Markov</u>

A et B sont des estimateurs sans biais de a et b de variance minimale.

Preuve ultérieurement.

. Calcul de la variance conditionnelle de B pour X fixé

$$V(B/X=x_i) = V\left(\frac{\sum_1^n(x_i-\bar{x})Y_i}{\sum(x_i-\bar{x})^2}\right) = \frac{\sum V((x_i-\bar{x})y_i)_{X=x_i}}{\left[\sum_{i=1}^n(x_i-\bar{x})^2\right]^2} = \frac{\sum_{i=1}^n(x_i-\bar{x})^2}{\left[\sum_{i=1}^n(x_i-\bar{x})^2\right]^2}V(Y_i)_{X=x_i}$$

$$\boxed{\overset{x_i}{V}(B) = \frac{\sigma^2}{\sum(x_i-\bar{x})^2}}$$

car pour X fixé $V(Y_i/X=x_i) = V(a+bx_i+\varepsilon_i) = V(\varepsilon_i) = \sigma^2$

D'autre part $A = \bar{Y} - \bar{X}B$

$V(A) = V(\bar{Y}-\bar{X}B)$ et $V(A/X=x_i) = V(\bar{Y}) + \bar{x}^2V(B) = \frac{\sigma^2}{n} + \bar{x}^2\frac{\sigma^2}{\sum(x_i-\bar{x})^2}$

$$\boxed{\overset{x_i}{V}(A) = \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum_1^n(x_i-\bar{x})^2}\right)}$$

Pour estimer la variance σ^2 des ~~residus~~ ε, il est naturel de penser à estimer la variance des residus $e_i = y_i - y_i^*$

Théorème

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - y_i^*)^2}{(n-2)}$$ est un estimateur sans biais de σ^2

Proprieté des ecarts residuels

$e_i = y_i - y_i^*$ ecart residuel.

1) Théorème.

~~$E(e_i)$ zero~~ e_i de moyenne nulle.

Preuve.

$$y_i^* = \bar{y} + b(x_i - \bar{x})$$

$$\sum e_i = \sum_{i=1}^{n} (y_i^* - y_i) = \sum_{i=1}^{n} (\bar{y} - y_i) + b \sum_{1}^{n} (x_i - \bar{x}) = 0$$

ceci prouve que les residus e_i ne sont pas des realisations independants d'une v.a

Variance empirique des résidus

$$\frac{1}{n} \sum_{i=1}^{n} e_i^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y} - b(x_i - \bar{x}))^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 + \frac{b^2}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 - \frac{2b}{n} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})$$

$$= S_y^2 + b^2 S_x^2 - 2b \operatorname{cov}(x, y)$$

$$\cancel{= S_y + \frac{b^2}{n} S_x}$$

avec $b = r \frac{S_y}{S_x} = \frac{\operatorname{cov}(x,y)}{S_x^2}$

$$\frac{1}{n} \sum_{i=1}^{n} e_i^2 = S_y^2 + r^2 \frac{S_y^2}{S_x^2} \cdot S_x^2 - 2r \frac{S_y}{S_x} \cdot r \frac{S_y}{S_x} \cdot S_x$$

$$= S_y^2 + r^2 S_y^2 - 2r^2 S_y^2 = \boxed{(1 - r^2) S_y^2}$$

III / Cas ou les ε suit une loi normale.

Tout les resultats établis precedamment supposent uniquement que ~~de~~ * $E(Y/X) = a + bX$

Supposant de plus que $\varepsilon \sim N(0, \sigma)$

On a :

a) $Y/[X=x] \sim N(a+bx\,;\,\sigma)$

b) B, A et Y^* suivent des lois pour les x_i fixés, gaussiennes car ils sont combinaisons linéaires des lois de Gauss.

$$B \sim N\left(b, \frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}}\right)$$

$$A \sim N\left(a, \sigma\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}\right)$$

$$Y^* \sim N\left(a+bx\,;\,\sigma\sqrt{\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2}}\right)$$

c) A, B et $\hat{\sigma}^2$ sont des estimateurs de variance minimale de a, b, σ^2

d) $\dfrac{(n-2)\hat{\sigma}^2}{\sigma^2} = \dfrac{\sum(y_i-y_i^*)^2}{\sigma^2} = \dfrac{n S^2_{Y/x}}{\sigma^2} \sim \chi^2_{n-2}$ indépendante de $\bar{Y}$, B et de A.

Rq :

A et B ne sont pas indépendants.

de b) on a :

$$\frac{B-b}{\frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}\,S_x}} \sim N(0,1) \quad \text{et} \quad \frac{n S^2_{Y/x}}{\sigma^2} \sim \chi^2_{n-2}$$

alors :

$$\frac{(B-b)S_x}{S_{Y/x}}\sqrt{n-2} \sim T_{n-2}$$

Ce qui permet de construire des intervalles de confiance pour b

De même

$$\frac{A-a}{\sigma\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}} \sim N(0,1)$$

$$\frac{A-a}{\sigma\sqrt{\frac{1}{n}+\frac{\bar{x}}{\sum(x_i-\bar{x})^2}}} \cdot \frac{\sigma}{\sqrt{n}\,S_{Y/x}}\sqrt{n-2} \sim T_{n-2}.$$

IV/ Intervalles de confiance

On a

1) $$T = \frac{(B - b) S_x}{S_{y/x}} \sqrt{n-2} \sim T_{n-2}$$

$$u \leq T \leq v \iff \frac{S_{y/x}}{S_x \sqrt{n-2}} u \leq B - b \leq \frac{S_{y/x}}{S_x \sqrt{n-2}} v$$

$$\iff B - \frac{S_{y/x}}{S_x \sqrt{n-2}} v \leq b \leq -\frac{S_{y/x}}{S_x \sqrt{n-2}} + B$$

tq $P(u \leq T \leq v) = 1 - \alpha$; u et v donnés par la table T_{n-2}

et $$B = \frac{\sum_{i=1}^{n} (y_i - \bar{Y})(x_i - \bar{X})}{\sum_{i=1}^{n} (x_i - \bar{X})^2}$$

2) On a

$$U = \frac{A - a}{\sqrt{\frac{1}{n} + \frac{\bar{x}}{\sum (x_i - \bar{x})^2}}} \frac{\sqrt{n-2}}{\sqrt{n}\, S_{y/x}} \sim T_{n-2}$$

$$u \leq U \leq v \iff u \sqrt{\frac{1}{n} + \frac{\bar{x}}{\sum (x_i - \bar{x})^2}} \frac{\sqrt{n} S_{y/x}}{\sqrt{n-2}} \leq A - a \leq v \sqrt{\frac{1}{n} + \frac{\bar{x}}{\sum (x_i - \bar{x})^2}} \cdot \frac{\sqrt{n}\, S_{y/x}}{\sqrt{n-2}}$$

$$\text{et } A - v \sqrt{\frac{1}{n} + \frac{\bar{x}}{\sum (x_i - \bar{x})^2}} \frac{\sqrt{n} S_{y/x}}{\sqrt{n-2}} < a < A - u \sqrt{\frac{1}{n} + \frac{\bar{x}}{\sum (x_i - \bar{x})^2}} \cdot \frac{\sqrt{n}\, S_{y/x}}{\sqrt{n-2}}$$

u et v donnés par la table de student T_{n-2}

V/ Erreurs d'estimation

Soit $\hat{y}_i = \hat{a} + \hat{b} x_i$

Les résidus quadratiques est la quantité moyen donnée par :

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

On a :

$$y_i - \hat{y}_i = y_i - \hat{a} - \hat{b} x_i = y_i - \left(\bar{y} - \frac{S_{xy}}{S_x^2} \bar{x}\right) - \left(\frac{S_{xy}}{S_x^2}\right) x_i$$

$$= y_i - \bar{y} + \frac{S_{xy}}{S_x^2} (\bar{x} - x_i)$$

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

$$= \frac{1}{n}\sum_{i=1}^{n}\left[(y_i-\bar{y}) - \frac{S_{xy}}{S_x^2}(x_i-\bar{x})\right]^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{S_{xy}^2}{S_x^4}(x_i-\bar{x})^2 - 2\frac{S_{xy}}{S_x^2}(y_i-\bar{y})(x_i-\bar{x})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{S_{xy}^2}{S_x^4}\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 - 2\frac{S_{xy}}{S_x^2}\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{S_{xy}^2}{S_x^4}S_x^{\cancel{2}} - 2\frac{S_{xy}}{S_x^2}\cdot S_{xy}$$

$$\eta^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 - \frac{S_{xy}^2}{S_x^2}$$

$$\boxed{\eta^2 = S_y^2 - \frac{S_{xy}^2}{S_x^2}}$$

On pose $r = \frac{S_{xy}}{S_x S_y}$ alors $\eta^2 = S_y^2(1-r^2)$

r est appelé coefficient de correlation empirique entre X et Y

Rq : Si $\eta^2 = 0 \iff y_i = \hat{y}_i$ i.e les points (x_i, y_i) sont alignés

Interpretation du coefficient de correlation emperique

Proposition

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

Preuve

$$(y_i-\bar{y})^2 = (y_i-\hat{y}_i+\hat{y}_i-\bar{y})^2$$

$$= (y_i-\hat{y}_i)^2 + (\hat{y}_i-\bar{y})^2 + 2(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$$

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

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) = \sum_{i=1}^{n}(y_i-\hat{y}_i)\hat{y}_i - \bar{y}\sum(y_i-\hat{y}_i)$$

$$= \sum_{i=1}^{n}(y_i-\hat{y}_i)\left(\bar{y}-\frac{S_{xy}}{S_x^2}\bar{x}+\frac{S_{xy}}{S_x^2}x_i\right) - \bar{y}\sum_{i=1}^{n}(y_i-\hat{y}_i)$$

$$= \cancel{\bar{y}\sum_{i=1}^{n}(y_i-\hat{y}_i)} - \frac{S_{xy}}{S_x^2}\bar{x}\sum_{i=1}^{n}(y_i-\hat{y}_i) + \frac{S_{xy}}{S_x^2}\sum_{i=1}^{n}x_i(y_i-\hat{y}_i) - \cancel{\bar{y}\sum_{i=1}^{n}(y_i-\hat{y}_i)}$$

$$= \frac{S_{xy}}{S_x^2}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\hat{y}_i)$$

$$= \frac{S_{xy}}{S_x^2}\sum_{i=1}^{n}(x_i-\bar{x})\left(y_i-\bar{y}+\frac{S_{xy}}{S_x^2}\bar{x}-\frac{S_{xy}}{S_x^2}x_i\right)$$

$$= \frac{S_{xy}}{S_x^2}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) - \left(\frac{S_{xy}}{S_x}\right)^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$= n\frac{S_{xy}^2}{S_x^2} - n\frac{S_{xy}^2}{S_x^2} = 0$$

cqfd.

Comme $\eta^2 = S_y^2(1-r^2)$

$$\Leftrightarrow 1-r^2 = \frac{\eta^2}{S_y^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

Alors $r^2 = 1 - \dfrac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}$

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

$$r^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 - \sum_{1}^{n}(\hat{y}_i-\bar{y})^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2} = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$r^2 = \frac{\text{Ecart expliqué}}{\text{Ecart total}}.$$

r^2 est une mesure de la quantité d'ajustement des moindres carrés par la droite de régression

qd $r^2 \longrightarrow 1 \Leftrightarrow \hat{y}_i$ est très proche de $y_i \Rightarrow$ la droite est un bon ajustement du nuage des p^ts^ $(x_i ; y_i)$

Si $r^2 \longrightarrow 0 \Leftrightarrow \hat{y}_i$ loin de $y_i \Rightarrow$ la droite n'est pas la bonne modelisation