

Analyse de données

Régression linéaire

Jamal Atif

jamal.atif@dauphine.fr

Université Paris-Dauphine, Licence MIDO



2014-2015

Ce cours est adapté librement des ressources suivantes :

- ▶ Livre : Régression avec , Pierre-André Cornillon et Eric Matzner- Lober, Springer 2010.

Exemple introductif

Un analyste en webmarketing s'intéresse à la relation de **causalité** entre les frais de publicité et les ventes d'un produit donné. En particulier il cherche à savoir s'il est possible d'**expliquer** le nombre de ventes par les frais de publicité. Il collecte **l'échantillon** des données suivant :

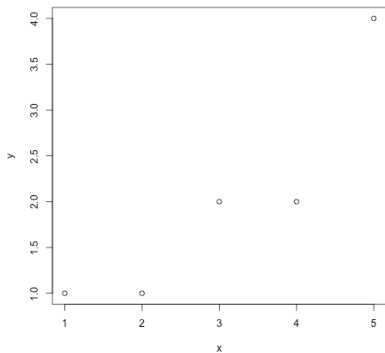
Pub. (€)	Ventes (unité)
1	1
2	1
3	2
4	2
5	4

Il s'agit alors :

- ▶ de trouver un *modèle* permettant d'expliquer *Ventes* en fonction de *Pub*,
- ▶ de *prédire* les valeurs de *Ventes* par de nouvelles valeurs de *Pub*.

⇒ Régression

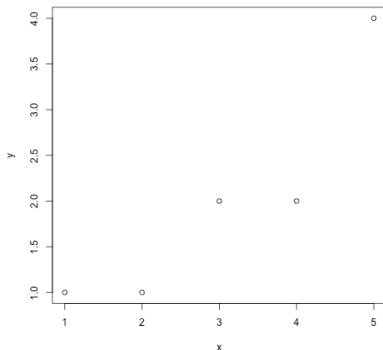
Nuage de points



- ▶ Pub (x) est une variable indépendante, dite variable **explicative** ou encore de **régression**
- ▶ Ventés (y) est dépendante dite **réponse**.

Première démarche

Examiner le nuage de points



Trouver une relation entre la variable x et la variable y , telle que :

$$y_i \approx f(x_i), i = 1, \dots, n$$

Modèle

$$y_i \approx f(x_i)$$

→ Se fixer une famille de fonction \mathcal{F} (ex : fonctions linéaires) et une fonction de coût L telle que :

$$\sum_{i=1}^n L(y - f(x)) \text{ est minimale pour une fonction } f \in \mathcal{F} \text{ donnée,}$$

où n représente le nombre de données disponibles (taille de l'échantillon) et L une fonction de coût ou de perte (Loss).



$$\hat{f} = \operatorname{argmin}_{f \in \mathcal{F}} \sum_{i=1}^n L(y - f(x)), i = 1, \dots, n$$

Exemples de L :

- ▶ $L(x) = |x|$,
- ▶ $L(x) = x^2$,
- ▶ etc.

Modèle de régression linéaire simple

- ▶ \mathcal{F} est une famille de fonctions linéaires (affines) de \mathbb{R} dans \mathbb{R} .
- ▶ On suppose disposer d'un échantillon de n points (x_i, y_i) .

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \forall i = 1, \dots, n$$

- ▶ ε_i modélisent le bruit et sont supposés aléatoires (les points n'étant jamais parfaitement alignés sur une droite).
- ▶ β_1 et β_2 sont les paramètres inconnues du modèles.

Modèle de régression linéaire simple

Hypothèses :

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \forall i = 1, \dots, n$$

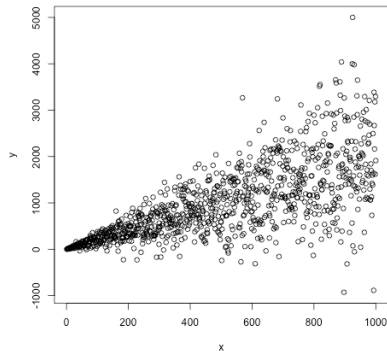
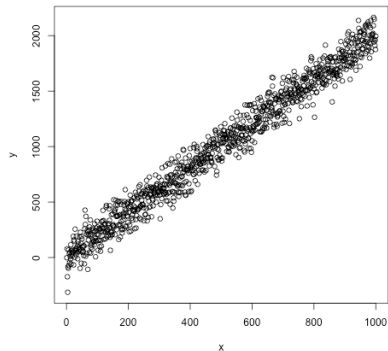
1. le bruit est une variable aléatoire d'espérance nulle et de variance inconnue fixe (homoscédacité) : $\mathbb{E}(\varepsilon_i) = 0$ et $Var(\varepsilon_i) = \sigma^2$,
2. ε_i et ε_j sont décorrélés pour tout $i \neq j$: $cov(\varepsilon_i, \varepsilon_j) = 0$
3. ε_i une v.a distribuée selon une loi normale de moyenne nulle et de variance σ^2 : $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Modèle de régression linéaire simple

Hypothèses :

Homoscédacité vs Hétéroscédacité

$$\text{Ex : } y_i = 2x_i + 1 + \varepsilon_i$$

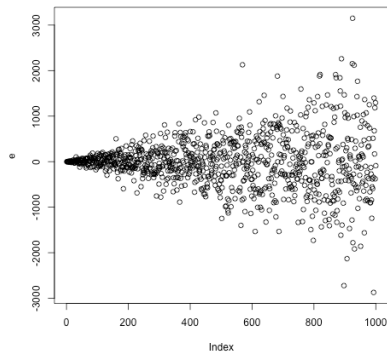
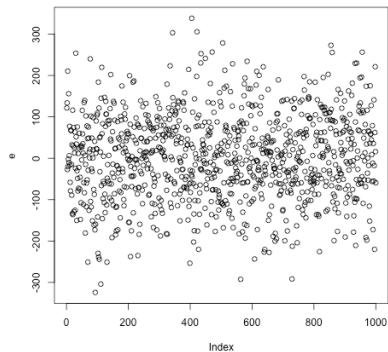


Nuage de points (x_i, y_i) .

Modèle de régression linéaire simple

Hypothèses :

Homoscédasticité vs Hétéroscédasticité



Nuage de points (x_i, ε_i) .

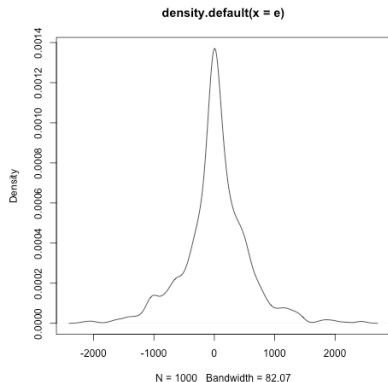
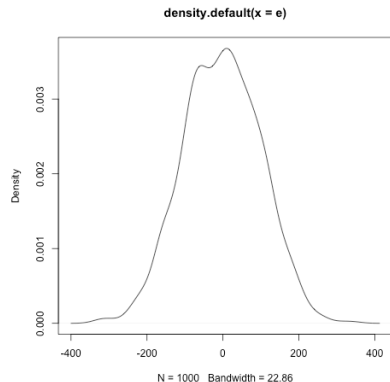
$$\varepsilon_i \sim \mathcal{N}(0, x_i)$$

Modèle de régression linéaire simple

Hypothèses :

Homoscédasticité vs Hétéroscédasticité

$$\text{Ex : } y_i = 2x_i + 1 + \varepsilon_i$$



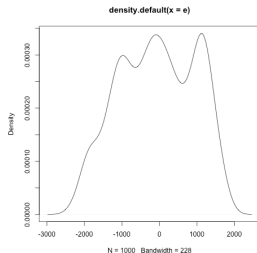
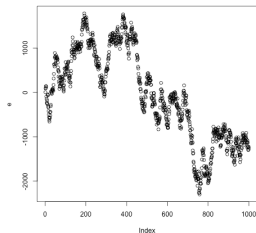
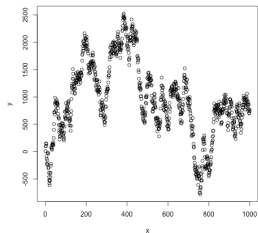
Distribution des erreurs.

Modèle de régression linéaire simple

Hypothèses :

Dépendance vs indépendance du bruit

$$\text{Ex : } y_i = 2x_i + 1 + \varepsilon_i$$



A gauche (x_i, y_i) , au milieu (x_i, ε_i) et à droite la distribution de ε .

$$\varepsilon_{i+1} \sim \mathcal{N}(\varepsilon_i, 100)$$

Estimateur des Moindres carrés Ordinaires

$$\hat{f} = \operatorname{argmin}_{f \in \mathcal{F}} \sum_{i=1}^n L(y_i - f(x_i)), \quad i = 1, \dots, n$$

On fixe $L(x) = x^2$ et $f(x) = \beta_1 + \beta_2 x$

$$(\hat{\beta}_1, \hat{\beta}_2) = \operatorname{argmin}_{\beta_1, \beta_2} \sum_{i=1}^n (y_i - \beta_1 - \beta_2 x_i)^2$$

Estimateur des Moindres carrés Ordinaires

Cela revient à minimiser le carré du bruit ε_i pour chaque i :

$$\varepsilon_i = y_i - \beta_1 + \beta_2 x_i = y_i - \tilde{y}_i$$

y_i : le point observé, et \tilde{y}_i le point de la droite *théorique*.

$$\begin{aligned}(\hat{\beta}_1, \hat{\beta}_2) &= \operatorname{argmin}_{\beta_1, \beta_2} \sum_{i=1}^n (y_i - \beta_1 - \beta_2 x_i)^2, \\ &= \operatorname{argmin}_{\beta_1, \beta_2} \sum_{i=1}^n (y_i - \tilde{y}_i)^2, \\ &= \operatorname{argmin}_{\beta_1, \beta_2} \sum_{i=1}^n \varepsilon_i^2, \\ &= \operatorname{argmin}_{\beta_1, \beta_2} \|\varepsilon\|^2\end{aligned}$$

Illustration

Au tableau !

Calcul des estimateurs de β_1 et β_2

- ▶ On notera $S(\beta_1, \beta_2) = \sum_{i=1}^n (y_i - \beta_1 - \beta_2 x_i)^2$
- ▶ $S(\beta_1, \beta_2)$ est quadratique donc convexe et différentiable \Rightarrow admet un minimum unique en $(\hat{\beta}_1, \hat{\beta}_2)$.
- ▶ On calcule les points pour lesquelles les dérivées partielles de S en β_1 et β_2 s'annulent. On obtient les équations normales suivantes :

$$\begin{cases} \frac{\partial S}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \\ \frac{\partial S}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^n x_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \end{cases} \quad (1)$$

Calcul des estimateurs de β_1 et β_2

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0$$

$$\Rightarrow \hat{\beta}_1 n + \hat{\beta}_2 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$\Rightarrow \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}.$$

Calcul des estimateurs de β_1 et β_2

$$\frac{\partial \mathcal{S}}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^n x_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0$$

$$\Rightarrow \hat{\beta}_1 \sum_{i=1}^n x_i + \hat{\beta}_2 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

$$\Rightarrow \hat{\beta}_2 = \frac{\sum x_i y_i - \sum x_i \bar{y}}{\sum x_i^2 - \sum x_i \bar{x}} = \frac{\sum x_i (y_i - \bar{y})}{\sum x_i (x_i - \bar{x})} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})(x_i - \bar{x})}$$

Exercice de TD!

Quelques remarques

- ▶ La relation $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$ montre que la droite des moindres carrés passe par le centre de gravité des nuages (\bar{x}, \bar{y}) .
- ▶ Les expressions obtenues pour $\hat{\beta}_1$ et $\hat{\beta}_2$ montrent que ces deux estimateurs sont linéaires par rapport au vecteur y .
- ▶ L'estimateur $\hat{\beta}_2$ peut s'écrire (exercice de TD) :

$$\hat{\beta}_2 = \beta_2 + \frac{\sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x})^2}$$

→ La variation de $\hat{\beta}_1$ et $\hat{\beta}_2$ vient seulement de ε

Quelques propriétés de $\hat{\beta}_1$ et $\hat{\beta}_2$

Sous les hypothèses 1 et 2 (centrage, décorrélation et homoscedasticité) β_1 et β_2 sont des estimateurs sans biais de β_1 et β_2 .

Nous savons que :

$$\hat{\beta}_2 = \beta_2 + \frac{\sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x})^2}$$

Dans cette expression, seuls les bruits ε_i sont aléatoires d'espérance nulle. Nous avons donc :

$$\mathbb{E}(\hat{\beta}_2) = \beta_2$$

Pour $\hat{\beta}_1$, on part de l'expression :

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x},$$

d'où l'on tire :

$$\mathbb{E}(\hat{\beta}_1) = \mathbb{E}(\bar{y}) - \bar{x} \mathbb{E}(\hat{\beta}_2) = \beta_1 + \bar{x} \beta_2 - \bar{x} \beta_2 = \beta_1$$

Quelques propriétés de $\hat{\beta}_1$ et $\hat{\beta}_2$

Les variances des estimateurs sont :

$$\text{var}(\hat{\beta}_1) = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right)$$

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

Et la covariance vaut :

$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\frac{\sigma^2 \bar{x}}{\sum (x_i - \bar{x})^2}$$

Preuve au tableau.

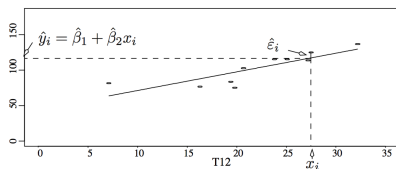
Résultat fondamental

Théorème de Gauss-Markov :

Parmi les estimateurs sans biais linéaires de y , les estimateurs $\hat{\beta}_i$ sont de variances minimales.

Nous omettrons la preuve.

Calcul des résidus et de la variance résiduelle



$$\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i = (y_i - \bar{y}) - \hat{\beta}_2 (x_i - \bar{x})$$

Par construction, nous avons :

$$\sum_i \hat{\epsilon}_i = \sum_i (y_i - \bar{y}) - \hat{\beta}_2 \sum_i (x_i - \bar{x}) = 0.$$

Estimateur non biaisé de σ^2

La statistique $\hat{\sigma}^2 = \sum \hat{\epsilon}_i^2 / (n - 2)$ est un estimateur sans biais de σ^2 .

Détails au tableau.

Prévision

Rappel : un des buts de la régression est la prévision/prédiction.

Soit x_{n+1} une nouvelle valeur pour laquelle nous voulons prédire y_{n+1} . Le modèle s'écrit :

$$y_{n+1} = \beta_1 + \beta_2 x_{n+1} + \varepsilon_{n+1},$$

avec $\mathbb{E}(\varepsilon_{n+1}) = 0$, $\text{var}(\varepsilon_{n+1}) = \sigma^2$ et $\text{cov}(\varepsilon_{n+1}, \varepsilon_n) = 0$ pour tout $i = 1, \dots, n$.

La valeur y_{n+1} peut être prédite comme suit :

$$\hat{y}_{n+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{n+1}$$

Erreur de prévision

$$\hat{y}_{n+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{n+1}$$

L'erreur de prévision $\hat{\varepsilon}_{n+1} = (y_{n+1} - \hat{y}_{n+1})$ satisfait les propriétés suivantes :

$$\begin{cases} \mathbb{E}(\hat{\varepsilon}_{n+1}) &= 0 \\ \text{var}(\hat{\varepsilon}_{n+1}) &= \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \end{cases} \quad (2)$$

Interprétation

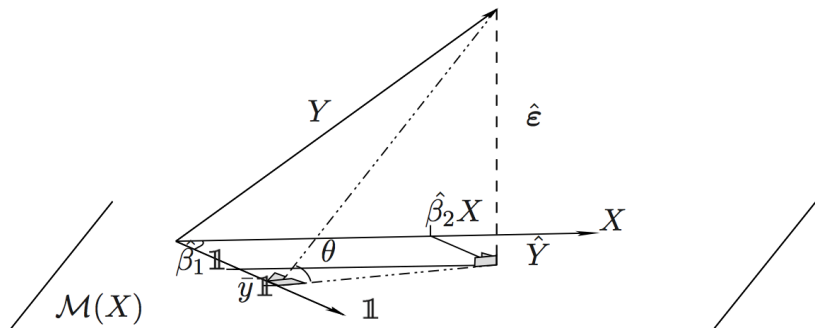
La variance augmente lorsque x_{n+1} s'éloigne du centre de gravité. Autrement dit, faire la prévision lorsque x_{n+1} est «loin» de \bar{x} est périlleux, puisque la variance de l'erreur de prévision peut être très grande.

Interprétation géométrique

Le problème de régression peut prendre la forme matricielle :

$y = Ab + \varepsilon$, avec

$$A = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, b = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$



Décomposition de la variance

$$\hat{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}}$$

Par le théorème de Pythagore, nous avons :

$$\|\mathbf{y} - \bar{y}\mathbf{1}\|^2 = \|\hat{\mathbf{y}} - \bar{y}\mathbf{1}\|^2 + \|\hat{\varepsilon}\|^2 \quad (3)$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n \hat{\varepsilon}_i^2, \quad (4)$$

$$\text{SCT} = \text{SCE} + \text{SCR} \quad (5)$$

- ▶ SCT : *Somme des Carrés des écarts Totale* ; elle possède $(n - 1)$ degrés de liberté
- ▶ SCE : *Somme des Carrés des écarts Expliquée* ; elle possède (1) degrés de liberté
- ▶ SCR : *Somme des Carrés des écarts Résiduelle* ; elle possède $(n - 2)$ degrés de liberté

Rappels sur les degré de liberté

Somme des carrés à la moyenne, SC

$$SC = \sum_{i=1}^n (x_i - \bar{x})^2$$

Nombre de degrés de liberté (ddl)

- ▶ ddl = nombre total des valeurs – nombre des valeurs estimées.
- ▶ Pour la somme précédente (SC), on a estimé la moyenne, donc $ddl = n - 1$.

Variance estimée

$$\text{var}(x) = \frac{SC}{ddl} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Décomposition de la variance

$$SCT = SCE + SCR$$

Les degrés de liberté (ddl) s'additionnent, tout comme les sommes carrés d'écart :

$$(n - 1) = 1 + (n - 2)$$

Les variances s'obtient en divisant chaque somme de carrés d'écart par le nombre de ddl correspondants :

$$V_t = \frac{SCT}{n - 1}, \quad V_e = SCE, \quad V_r = \frac{SCE}{n - 2}$$

Ce sont respectivement les variances totale, expliquée, et résiduelle. (Ces variances ne s'additionnent pas !)

Quelques quantités

Le coefficient de détermination R^2

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = \frac{\|\hat{y} - \bar{y}\mathbf{1}\|^2}{\|y - \bar{y}\mathbf{1}\|^2} = 1 - \frac{\|\hat{\epsilon}\|^2}{\|y - \bar{y}\mathbf{1}\|^2} = 1 - \frac{\text{SCR}}{\text{SCT}} = \rho_{xy}^2$$

- ▶ Si $R^2 = 1$, le modèle explique tout, l'angle θ vaut zéro et y est dans $\mathcal{M}(X)$, c'est-à-dire que $y_i = \beta_1 + \beta_2 x_i$ pour tout i : les points de l'échantillon sont parfaitement alignés sur la droite des moindres carrés ;
- ▶ Si $R^2 = 0$, cela veut dire que $(\hat{y}_i - \bar{y})^2 = 0$, donc $\hat{y}_i = \bar{y}$ pour tout i . Le modèle de régression linéaire est inadapté puisqu'on ne modélise rien de mieux que la moyenne ;
- ▶ Si R^2 est proche de zéro, cela veut dire que Y est quasiment dans l'orthogonal de $\mathcal{M}(X)$, le modèle de régression linéaire est inadapté, la variable x n'explique pas bien la variable réponse y (du moins pas de façon affine).

Quelques quantités

Le coefficient de corrélation R : c'est la racine du coefficient de détermination, affecté du signe de la pente $\hat{\beta}_2$. Il est toujours compris entre -1 et 1 .

L'écart-type résiduel $\hat{\sigma}$: c'est la racine carrée de la variance résiduelle ($\hat{\sigma} = \sqrt{V_r}$). C'est une estimation de l'erreur faite sur la mesure de la variable dépendante y . Une valeur de 0 indiquerait un ajustement parfait.

Le rapport de variance F : c'est le rapport de la variance expliquée à la variance résiduelle ($F = \frac{V_e}{V_r}$).

Régions de confiance et tests d'hypothèses

Mieux que les expressions des estimateurs et celles de leurs variances, on aimerait connaître leurs lois : ceci permettrait par exemple d'obtenir des régions de confiance et d'effectuer des tests d'hypothèses. Dans cette optique, il faut bien entendu faire une hypothèse plus forte sur notre modèle, à savoir préciser la loi des erreurs.

Cas d'erreurs gaussiennes

Hypothèses supplémentaires sur le modèle :

1. $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$
2. ε_i sont mutuellement indépendants

Il s'en suit :

$$\forall i \in \{1, \dots, n\} \quad y_i \sim \mathcal{N}(\beta_1 + \beta_2 x_i, \sigma^2)$$

Estimateurs du maximum de vraisemblance

La vraisemblance vaut

$$\begin{aligned}\mathcal{L}(y; \beta_1, \beta_2, \sigma^2) &= \prod_{i=1}^n \mathcal{N}(\beta_1 + \beta_2 x_i, \sigma^2) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_1 - \beta_2 x_i)^2 \right] \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left[-\frac{1}{2\sigma^2} S(\beta_1, \beta_2) \right]\end{aligned}$$

⇒ Trouver $(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2)$ qui maximisent la vraisemblance.

⇒ Pour simplifier le calcul on prend la log-vraisemblance : $\log(\mathcal{L})$.

$$\log \mathcal{L}(y; \beta_1, \beta_2, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} S(\beta_1, \beta_2)$$

Estimateurs du maximum de vraisemblance

$$\hat{\beta}_{1mv}, \hat{\beta}_{2mv}$$

$$\log \mathcal{L}(y; \beta_1, \beta_2, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} S(\beta_1, \beta_2)$$

- ▶ Maximiser par rapport à (β_1, β_2) revient à minimiser $-S(\beta_1, \beta_2)$.
- ⇒ Les estimateurs du maximum de vraisemblance de β_1 et β_2 sont égaux aux estimateurs des moindres carrés.

$$\hat{\beta}_{1mv} = \hat{\beta}_1, \quad \hat{\beta}_{2mv} = \hat{\beta}_2$$

Estimateurs du maximum de vraisemblance

 $\hat{\sigma}_{mv}^2$

Il reste à maximiser $\log \mathcal{L}(y; \hat{\beta}_1, \hat{\beta}_2, \sigma^2)$ par rapport à σ^2 .

$$\begin{aligned}\frac{\partial \mathcal{L}(y; \hat{\beta}_1, \hat{\beta}_2, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} S(\hat{\beta}_1, \hat{\beta}_2) \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2\end{aligned}$$

$$\frac{\partial \mathcal{L}(y; \hat{\beta}_1, \hat{\beta}_2, \sigma^2)}{\partial \hat{\sigma}_{mv}^2} = 0 \Rightarrow \hat{\sigma}_{mv}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2$$

- ▶ L'estimateur du maximum de vraisemblance de σ^2 est différent de l'estimateur MCO $\hat{\sigma}^2$.
 - ▶ L'estimateur du maximum de vraisemblance de σ^2 est donc biaisé. En effet $\mathbb{E}(\hat{\sigma}_{mv}^2) = \frac{1}{n} \sum \mathbb{E}(\hat{\varepsilon}_i^2) = \frac{n-2}{n} \sigma^2$
- ⇒ Ce biais est d'autant plus négligeable que le nombre d'observations est grand.

Lois des estimateurs et régions de confiance

Rappels sur les lois usuelles

Loi du χ^2

Soit X_1, \dots, X_n des variables aléatoires i.i.d. suivant une loi normale centrée réduite. La loi de la variable $X = \sum_{i=1}^n X_i^2$ est appelée loi du χ^2 à n degrés de liberté (ddl), noté $X \sim \chi_n^2$.

Loi de Student

Soit Z une variable aléatoire suivant une loi normale centrée réduite et X une variable suivant une loi du χ^2 à n degrés de liberté, avec Z et X indépendantes. La loi de la variable $T = \frac{Z}{\sqrt{X/n}}$ est appelée loi de Student à n degrés de liberté et on note $T \sim \mathcal{T}_n$.

Loi de Fisher

Soit U_1 une variable aléatoire suivant une loi du χ^2 à n_1 degrés de liberté et U_2 une variable aléatoire suivant une loi du χ^2 à n_2 degrés de liberté, avec U_1 et U_2 indépendantes. La loi de la variable $F = \frac{U_1/n_1}{U_2/n_2}$ est appelée loi de Fisher à (n_1, n_2) degrés de liberté et on note $F \sim F_{n_1}^{n_2}$.

Lois des estimateurs et régions de confiance

Quelques notations préalables :

$$\begin{aligned}c &= \frac{-\sigma^2 \bar{x}}{\sum (x_i - \bar{x})^2} & \hat{\sigma}^2 &= \frac{1}{n-2} \sum \hat{\varepsilon}_i^2 \\ \sigma_1^2 &= \sigma^2 \left(\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} \right) & \hat{\sigma}_1 &= \hat{\sigma}^2 \left(\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} \right) \\ \sigma_2^2 &= \frac{\sigma^2}{\sum (x_i - \bar{x})^2} & \hat{\sigma}_2 &= \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}\end{aligned}$$

- ▶ σ_1^2 , σ_2^2 et c sont les variances et covariance des estimateurs des moindres carrés ordinaires.
- ▶ $\hat{\sigma}_1^2$ et $\hat{\sigma}_2^2$ correspondent quant à elles aux estimateurs des variances de $\hat{\beta}_1$ et $\hat{\beta}_2$.

Lois des estimateurs avec variance connue

Les lois des estimateurs des MCO avec variance σ^2 connue sont :

- ▶ $\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \sim \mathcal{N}(\beta, \sigma^2 V)$ où $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ et
$$V = \frac{1}{\sum (x_i - \bar{x})^2} \begin{bmatrix} \sum x_i^2 / n & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \sigma_1^2 & c \\ c & \sigma_2^2 \end{bmatrix}$$
- ▶ $\frac{(n-2)}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-2}^2$, loi du χ^2 à $(n-2)$ ddl.
- ▶ $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendants.

Lois des estimateurs avec variance estimée

Les lois des estimateurs des MCO avec variance σ^2 estimée sont :

- ▶ $\frac{\hat{\beta}_1 - \beta_1}{\sigma_1} \sim \mathcal{T}_{n-2}$ où \mathcal{T}_{n-2} est une loi de Student à $(n - 2)$ degrés de liberté.
- ▶ $\frac{\hat{\beta}_2 - \beta_2}{\sigma_2} \sim \mathcal{T}_{n-2}$.
- ▶ $\frac{1}{2\hat{\sigma}^2} (\hat{\beta} - \beta)' V^{-1} ((\hat{\beta} - \beta)) \sim \mathcal{F}_{n-2}^2$, loi de Fisher de paramètres $(2, n - 2)$.

Ces dernières propriétés nous permettent de donner des intervalles de confiance (IC) ou des régions de confiance (RC) des estimateurs.

Intervalles et régions de confiance

- ▶ $IC(\hat{\beta}_1) \pm t_{n-2}(1 - \alpha/2)\hat{\sigma}_1$ où $t_{n-2}(1 - \alpha/2)$ est le quantile de niveau $(1 - \alpha/2)$ d'une loi de Student \mathcal{T}_{n-2} .
- ▶ $IC(\hat{\beta}_2) \pm t_{n-2}(1 - \alpha/2)\hat{\sigma}_2$
- ▶ $RC(\beta)$: Une région de confiance simultanée pour β_1 et β_2 au niveau $(1 - \alpha)$ est

$$\frac{1}{2\hat{\sigma}^2} \left(n(\hat{\beta}_1 - \beta_1)^2 + 2n\bar{x}(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) + \sum x_i^2(\hat{\beta}_2 - \beta_2)^2 \right) \leq f_{n-2}^2(1 - \alpha),$$

où $f_{n-2}^2(1 - \alpha)$ est le quantile de niveau $(1 - \alpha)$ d'une loi \mathcal{F}_{n-2}^2 .

- ▶ Un intervalle de confiance de σ^2 est donné par :

$$\left[\frac{(n-2)\hat{\sigma}^2}{c_{n-2}(1 - \alpha/2)}, \frac{(n-2)\hat{\sigma}^2}{c_{n-2}(\alpha/2)} \right]$$

où $c_{n-2}(1 - \alpha/2)$ est le quantile de niveau $1 - \alpha/2$ d'une loi χ_{n-2}^2 .

Loi et intervalle de confiance pour la prédiction

On a :

$$y_{n+1} - \hat{y}_{n+1} \sim \mathcal{N} \left(0, \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \right)$$

Avec les notations et hypothèses précédentes, on obtient :

$$\frac{y_{n+1} - \hat{y}_{n+1}}{\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim \mathcal{T}_{n-2}$$

d'où l'on déduit l'intervalle de confiance suivant pour y_{n+1} :

$$\left[\hat{y}_{n+1} \pm t_{n-2}(1 - \alpha/2) \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right]$$

Tests d'hypothèse

Test Bilatéral

$$H_0 : \beta_1 = 0, H_a : \beta_2 \neq 0$$

Sous H_0 , nous avons la v.a :

$$t = \frac{\hat{\beta}_2}{\hat{\sigma}_2} \sim t_{n-2}$$

Considérons un niveau de test α (usuellement 0,05 ou 0,01), On calcule t^* :

- ▶ si $|t^*| \leq t_{n-2}(1 - \alpha/2)$, alors H_0 est acceptée,
- ▶ si $|t^*| > t_{n-2}(1 - \alpha/2)$, alors H_0 est rejetée.

Tests d'hypothèse

Test unilatéral

Exemple : vérifier si β_2 est positive.

$$H_0 : \beta_2 \geq 0, H_a : \beta_2 < 0$$

Sous H_0 , nous avons :

$$t = \frac{\hat{\beta}_2}{\hat{\sigma}_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_2} + \frac{\beta_2}{\hat{\sigma}_2} \sim t_{n-2} + \text{terme positif}$$

Considérons un niveau de test α (usuellement 0,05 ou 0,01), On calcule t^* :

- ▶ si $|t^*| \geq t_{n-2}(\alpha)$, alors H_0 est acceptée,
- ▶ si $|t^*| < t_{n-2}(\alpha)$, alors H_0 est rejetée.

Tests d'hypothèse

Test de passage par l'origine β_1

Test si $\beta_1 = 0$:

$$H_0 : \beta_1 = 0, H_1 : \beta_1 \neq 0$$

Sous H_0 , nous avons la v.a :

$$t = \frac{\hat{\beta}_1}{\hat{\sigma}_1} \sim t_{n-2}$$

Considérons un niveau de test α (usuellement 0,05 ou 0,01), On calcule t^* :

- ▶ si $|t^*| \leq t_{n-2}(1 - \alpha/2)$, alors H_0 est acceptée,
- ▶ si $|t^*| > t_{n-2}(1 - \alpha/2)$, alors H_0 est rejetée.