Mise à niveau

Modèle linéaire

A. Godichon-Baggioni

I. Modèle linéaire

Modèle de régression

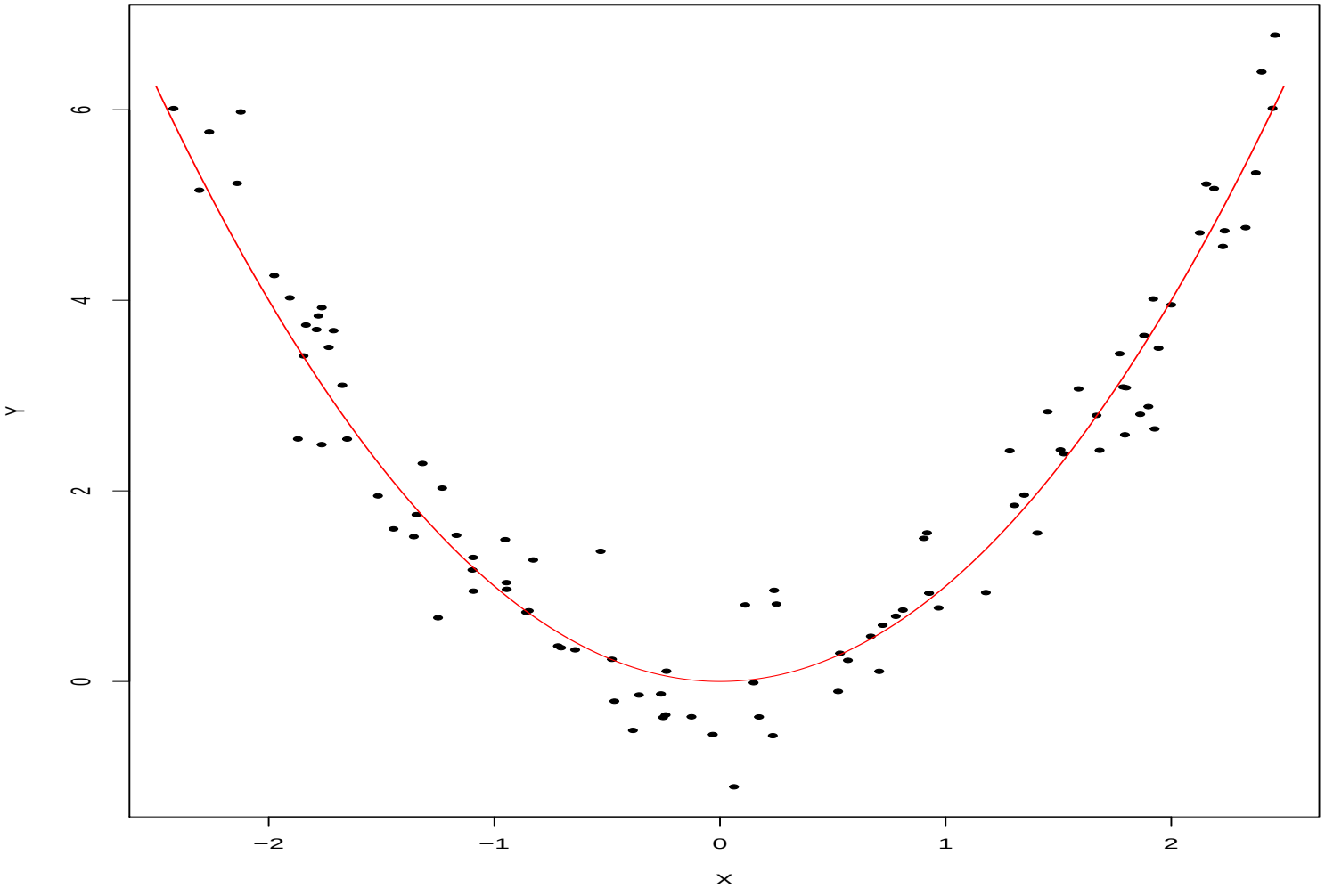
L'objectif est d'expliquer une variable Y en fonction de variables explicatives X. Plus précisément, un modèle de regression est un modèle où on considère :

- des variables aléatoires à expliquer $Y_1, \ldots, Y_n$
- des vecteurs $X_1, \ldots, X_n \in \mathbb{R}^p$ (variables explicatives)
- une fonction de régression $g : \mathbb{R}^p \longrightarrow \mathbb{R}$
- des variables aléatoires indépendantes et centrées $\epsilon_1, \ldots, \epsilon_n$

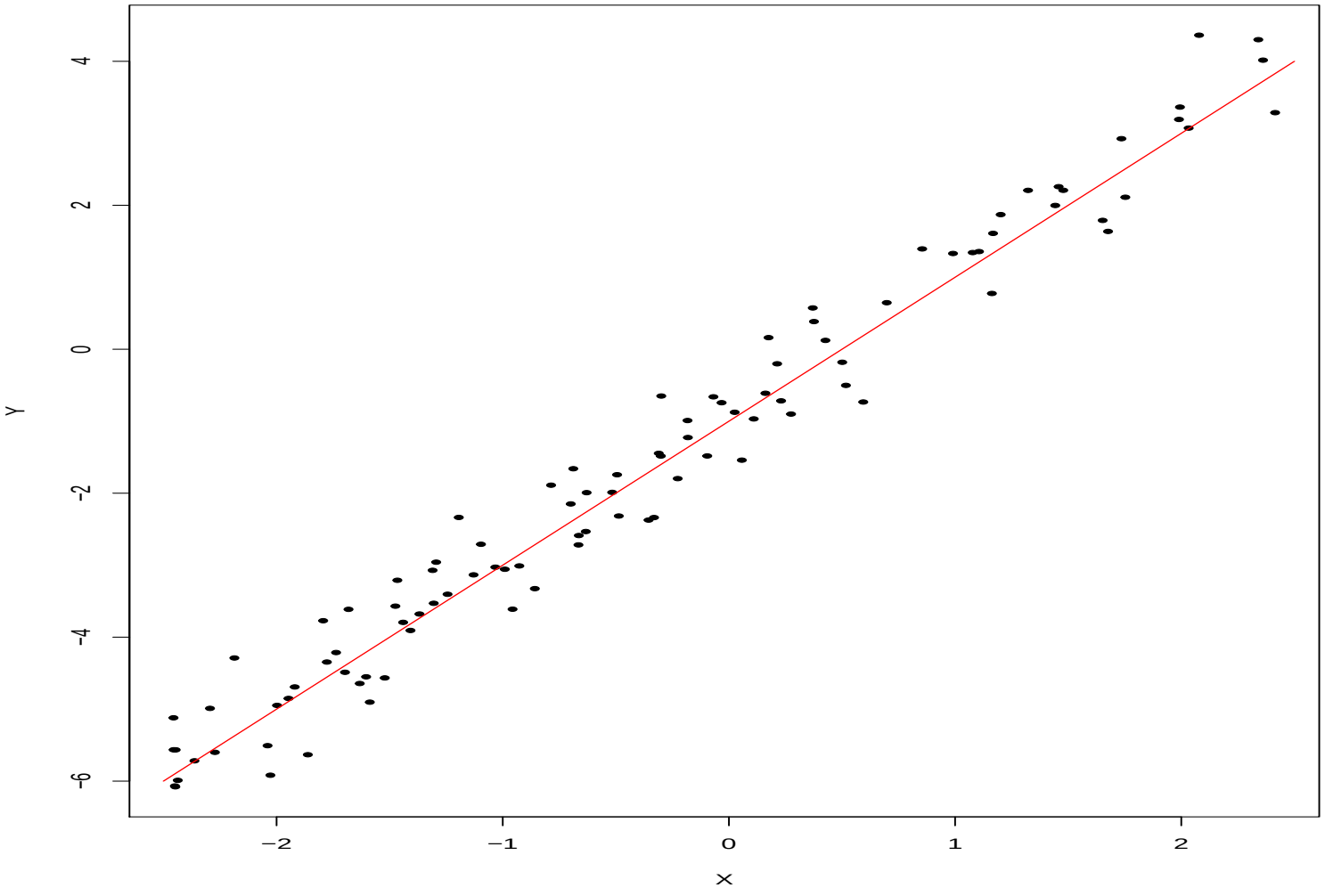
Le modèle de régression est défini comme

$$Y_i = g(X_i) + \epsilon_i.$$

Exemple 1 : $Y_i = ax_i^2 + bx_i + c + \epsilon_i$



Exemple 2 : $Y_i = ax_i + b + \epsilon_i$



Modèle linéaire

Définition

Lorsque la fonction de régréssion est linéaire, i.e de la forme

$$g(X) = X^T \beta$$

avec $\beta \in \mathbb{R}^p$, le modèle de régression associé est dit linéaire.

Forme matricielle

On considère le modèle linéaire

$$Y_i = g(X_i) + \epsilon_i = X_i^T \beta + \epsilon_i.$$

On note $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ et

$$\mathbf{X} = (X_1, \ldots, X_n)^T = \begin{pmatrix} X_{1,1} & \ldots & X_{1,p} \\ \vdots & \ddots & \vdots \\ X_{n,1} & \ldots & X_{n,p} \end{pmatrix}.$$

La forme matricielle du modèle linéaire s'écrit alors

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon.$$

EXEMPLE 1 : MODÈLE LINÉAIRE SIMPLE

On a $x_1, \ldots, x_n \in \mathbb{R}$ et pour tout i

$$y_i = a + bx_i + \epsilon_i.$$

Le modèle s'écrit sous la forme

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon.$$

avec

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad \text{et} \qquad \beta = \begin{pmatrix} a \\ b \end{pmatrix}.$$

La fonction $x \longmapsto a + bx$ est dite droite de régression.

EXEMPLE 2 : RÉGRESSION POLYNOMIALE

On a $x_1, \ldots, x_n \in \mathbb{R}$ et pour tout i

$$y_i = a_0 + a_1 x_i + \ldots + a_p x_i^p + \epsilon_i.$$

Le modèle s'écrit sous la forme

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon.$$

avec

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & \ldots & x_1^p \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \ldots & x_n^p \end{pmatrix}, \qquad \text{et} \qquad \beta = \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix}.$$

Comment estimer β ?

II. Méthode des moindres carrés

ESTIMATEUR DES MOINDRES CARRÉS

La méthode des moindres carrés consiste à chercher $\hat{\beta} \in \mathbb{R}^p$ qui minimise la quantité suivante :

$$\|\mathbf{Y} - \mathbf{X}\beta'\|^2 = \sum_{i=1}^{n} \left(Y_i - X_i^T \beta'\right)^2 .$$

$\hat{\beta}$ est appelé estimateur des moindres carrés.

- Existence d'une solution ?
- Unicité ?

Existence et unicité

Théorème

Si $\mathbf{X}$ *est de rang* p*, alors l'estimateur des moindres carrés est unique et est défini par*

$$\hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}$$

Proposition

Proposition

La matrice $\mathbf{X}^T\mathbf{X}$ est semi-définie positive. De plus, elle est positive si et seulement si $\mathbf{X}$ est de rang p.

Remarque : La matrice $\mathbf{X}$ ne peut être de rang p que si $n \geq p$.

MOINDRES CARRÉS ET PROJECTION

On dote $D \subset \mathbb{R}^p$ le sous espace vectoriel engendré par les colonnes de $\mathbf{X}$. Le problème de minimization peut s'écrire comme

$$\min_{\beta' \in \mathbb{R}^p} \|\mathbf{Y} - \mathbf{X}\beta'\|^2 = \min_{h \in D} \|\mathbf{Y} - h\|^2$$

et chercher le minimum revient donc à cherche la projection orthogonale de $\mathbf{Y}$ sur D.

Remarque : $\dim(D) = \text{rang}(\mathbf{X})$.

MOINDRES CARRÉS ET PROJECTION

Théorème

Si $rang(\mathbf{X}) = p$*, i.e si* $\mathbf{X}^T\mathbf{X}$ *est inversible, alors la matrice*

$$P_D = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$$

est la matrice de projection orthogonale sur D *et* $rang(H) = p$.

Propriétés de l'estimateur des moindres carrés

Proposition

Soit $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ *avec* $\mathbb{E}[\epsilon] = 0_{\mathbb{R}^n}$ *et* $Var[\epsilon] = \sigma^2 I_n$, $\sigma^2 > 0$. *On suppose de plus que* $rang(\mathbf{X}) = p$. *Alors*

$$\mathbb{E}\left[\hat{\beta}\right] = \beta \qquad et \qquad Var\left[\hat{\beta}\right] = \sigma^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1}.$$

DÉFINITIONS

Définition

1. *On appelle*
$$\hat{\epsilon} = \mathbf{Y} - \mathbf{X}\hat{\beta}$$
le vecteur des résidus.
2. *On appelle*
$$SCR = \|\hat{\epsilon}\|^2 = \left\|\mathbf{Y} - \mathbf{X}\hat{\beta}\right\|^2$$
la somme des carrés des résidus.

ESTIMATION DE σ^2

Théorème

Soit $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ *avec* $\mathbb{E}[\epsilon] = 0_{\mathbb{R}^n}$ *et* $Var[\epsilon] = \sigma^2 I_n$, $\sigma^2 > 0$. *On suppose de plus que* $rang(\mathbf{X}) = p$. *Alors*

$$\hat{\sigma}^2 = \frac{\left\| \mathbf{Y} - \mathbf{X}\hat{\beta} \right\|^2}{n-p} = \frac{1}{n-p} \sum_{i=1}^{n} \left(Y_i - X_i^T \hat{\beta} \right)^2$$

est un estimateur sans biais de σ^2.

III. Modèle linéaire gaussien

Modèle linéaire gaussien

Définition

Le modèle de régression linéaire

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

est dit modèle linéaire gaussien si ϵ est un vecteur gaussien de loi $\mathcal{N}\left(0, \sigma^2 I_n\right)$ avec $\sigma^2 > 0$.

Proposition

Dans le cadre du modèle linéaire gaussien, l'estimateur des moindres carrés coïncide avec l'estimateur du maximum de vraisemblance.

PROPRIÉTÉS DES ESTIMATEURS

Théorème

Soit $\epsilon \sim \mathcal{N}\left(0, \sigma^2 I_n\right)$, $\sigma^2 > 0$. *On suppose de plus que* $rang(\mathbf{X}) = p$. *Alors*

1. $\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)$.
2. $\hat{\beta}$ *et* $\hat{\sigma}^2$ *sont indépendantes.*
3. $\frac{(n-p)}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{n-p}$.

Intervalle de confiance

Théorème

Soit $x_0 \in \mathbb{R}^p$. Un intervalle de confiance de niveau $1-\alpha$ pour $x_0^T\beta$ est donné par

$$\left[x_0^T\hat{\beta} - \hat{\sigma}\sqrt{v_0}t_{n-p,1-\alpha/2}; x_0^T\hat{\beta} + \hat{\sigma}\sqrt{v_0}t_{n-p,1-\alpha/2}\right]$$

avec $v_0 = x_0^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}x_0$ et $t_{n-p,1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi de Student à $n-p$ degrés de liberté.

INTERVALLES DE PRÉDICTION

Théorème

Soit $x_0 \in \mathbb{R}^p$. Un intervalle de prédiction de niveau $1-\alpha$ de y_0 est donné par

$$\left[x_0^T\hat{\beta} - \hat{\sigma}\sqrt{1+v_0}t_{n-p,1-\alpha/2}; x_0^T\hat{\beta} + \hat{\sigma}\sqrt{1+v_0}t_{n-p,1-\alpha/2}\right]$$

avec $v_0 = x_0^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}x_0$ et $t_{n-p,1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi de Student à $n-p$ degrés de liberté.

Remarque : L'intervalle de prédiction est plus grand que l'intervalle de confiance car il doit également prendre en compte la variance de ϵ_0.

TEST DE SIGNIFICATIVITÉ D'UN PARAMÈTRE

Soit $k \in \{1, ..., p\}$, on souhaite tester

$$H0 : "\beta_k = 0" \qquad \text{contre} \qquad H1 : "\beta_k \neq 0".$$

Proposition

Dans le cadre du modèle linéaire gaussien, un test de significativité du k-ème coefficient de niveau α est donné par la zone de rejet

$$ZR_{\alpha,k} = \left\{ 0 \notin \left[\hat{\beta}_k \pm \hat{\sigma} \sqrt{v_k} t_{n-p,1-\alpha/2} \right] \right\},$$

où $v_k = \left(\left(\mathbf{X}^T \mathbf{X} \right)^{-1} \right)_{k,k}$ est le k-ème coefficient diagonal de $\left(\mathbf{X}^T \mathbf{X} \right)^{-1}$ et $t_{n-p,1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi de Student à $n - p$ degrés de liberté.

Test de significativité de plusieurs coéfficients

On souhaite tester la significativité de plusieurs coefficients $\beta_{i_1}, \ldots, \beta_{i_k}$. Pour simplifier les notations, on souhaite tester

$H0 : "\beta_{p_0+1}, \ldots, \beta_p = 0"$ contre $H1 : "\exists i \in \{p_0 + 1, \ldots, p\}, \beta_i \neq 0"$.

Test de significativité de plusieurs coéfficients

On note

- $\mathbf{X}_0 = \left[X_1, \ldots, X_{p_0}\right] \in \mathbb{R}^{n \times (p_0)}$
- $\mathbf{X}_1 = \left[X_{p_0+1}, \ldots, X_p\right] \in \mathbb{R}^{n \times (p-p_0)}$
- $\mathbf{X} = [\mathbf{X}_0, \mathbf{X}_1] \in \mathbb{R}^{n \times p}$.

Le test de significativité revient à faire la comparaison de modèles suivante :

$$H0 : "\mathbf{Y} = \mathbf{X}_0\beta_0 + \epsilon" \qquad \text{contre} \qquad "H1 : \mathbf{Y} = \mathbf{X}\beta + \epsilon".$$

Test de significativité de plusieurs coéfficients

Théorème

On considère la statistique de test

$$F = \frac{\left\| \hat{Y} - \hat{Y}_0 \right\|^2}{(p - p_0)\,\hat{\sigma}^2} =: \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2},$$

avec $\hat{Y}_0 = \mathbf{X}_0 \hat{\beta}_0$*, et* $\hat{\beta}_0$ *est l'estimateur des moindres carrés dans le modèle H0. Alors, sous H0,*

$$F \sim \mathcal{F}\left(p - p_0, n - p\right).$$

Corollaire

Tester la significativité des coefficients au risque α *revient à considérer la zone de rejet*

$$ZR_\alpha = \left\{ F > f_{1-\alpha, p-p_0, n-p} \right\}$$

IV. Modèles emboités

MODÈLES EMBOITÉS

Le test de Fisher précédent est un cas particulier du test entre modèles emboités. On considère :

- $\mathbf{Y} \sim \mathcal{N}\left(\mu, \sigma^2 I_n\right)$ avec $\sigma^2 > 0$.
- $\mathcal{M}_0 \subset \mathcal{M}_1 \subset \mathbb{R}^n$ deux sous espaces vectoriels de $\mathbb{R}^n$

On souhaite tester

$$H0 : \mathbb{E}[\mathbf{Y}] = \mu \in \mathcal{M}_0 \qquad \text{contre} \qquad H1 : \mathbb{E}[\mathbf{Y}] = \mu \in \mathcal{M}_1.$$

CONSTRUCTION DE LA STATISTIQUE DE TEST

On considère $P_{\mathcal{M}_0}, P_{\mathcal{M}_1}$ les projections orthogonales sur $\mathcal{M}_0$ et $\mathcal{M}_1$. On considère la statistique

$$\mathcal{F} = \frac{\dim\left(\mathcal{M}_1^{\perp}\right) \left\|P_{\mathcal{M}_0}\mathbf{Y} - P_{\mathcal{M}_1}\mathbf{Y}\right\|^2}{\left(\dim\left(\mathcal{M}_1\right) - \dim\left(\mathcal{M}_0\right)\right) \left\|P_{\mathcal{M}_1^{\perp}}\mathbf{Y}\right\|^2}$$

Proposition

Sous H0,

$$\mathcal{F} \sim F\left(dim\left(\mathcal{M}_1\right) - dim\left(\mathcal{M}_0\right), dim\left(\mathcal{M}_1^{\perp}\right)\right)$$

TEST DES MODÈLES EMBOITÉS

Corollaire

Faire le test des modèles emboités revient donc à considérer la zone de rejet

$$ZR_\alpha = \left\{ \mathcal{F} > f_{1-\alpha, dim(\mathcal{M}_1)-dim(\mathcal{M}_0), dim\left(\mathcal{M}_1^\perp\right)} \right\}$$