Semaine Régression linéaire

Sommaire

Introduction

Quels sont les attendus de cette semaine ?

- A travers les heures de cours, je vais vous présenter la régression linéaire ainsi que les sujets connexes comme la sélection de variables, l'analyse de la variance

Introduction

Quels sont les attendus de cette semaine ?

- A travers les heures de cours, je vais vous présenter la régression linéaire ainsi que les sujets connexes comme la sélection de variables, l'analyse de la variance
- Je vous présenterai des cas d'étude avec le logiciel R sans pour autant vous montrer le code

Introduction

Quels sont les attendus de cette semaine ?

- A travers les heures de cours, je vais vous présenter la régression linéaire ainsi que les sujets connexes comme la sélection de variables, l'analyse de la variance
- Je vous présenterai des cas d'étude avec le logiciel R sans pour autant vous montrer le code
- En dehors des heures de cours, vous travaillerez, par groupe de 3, en autonomie sur un ou des jeux de données avec une feuille de route pour vous aider

Introduction

Quels sont les attendus de cette semaine ?

- A travers les heures de cours, je vais vous présenter la régression linéaire ainsi que les sujets connexes comme la sélection de variables, l'analyse de la variance
- Je vous présenterai des cas d'étude avec le logiciel R sans pour autant vous montrer le code
- En dehors des heures de cours, vous travaillerez, par groupe de 3, en autonomie sur un ou des jeux de données avec une feuille de route pour vous aider
- Pendant ces heures, nous resterons à votre disposition afin que vous puissiez nous poser des questions

Introduction

Quels sont les attendus de cette semaine ?

- A travers les heures de cours, je vais vous présenter la régression linéaire ainsi que les sujets connexes comme la sélection de variables, l'analyse de la variance
- Je vous présenterai des cas d'étude avec le logiciel R sans pour autant vous montrer le code
- En dehors des heures de cours, vous travaillerez, par groupe de 3, en autonomie sur un ou des jeux de données avec une feuille de route pour vous aider
- Pendant ces heures, nous resterons à votre disposition afin que vous puissiez nous poser des questions
- Vous devrez être à l'école impérativement

Introduction

Quels sont les attendus de cette semaine ?

- A travers les heures de cours, je vais vous présenter la régression linéaire ainsi que les sujets connexes comme la sélection de variables, l'analyse de la variance
- Je vous présenterai des cas d'étude avec le logiciel R sans pour autant vous montrer le code
- En dehors des heures de cours, vous travaillerez, par groupe de 3, en autonomie sur un ou des jeux de données avec une feuille de route pour vous aider
- Pendant ces heures, nous resterons à votre disposition afin que vous puissiez nous poser des questions
- Vous devrez être à l'école impérativement
- Le lundi 2 juin, votre travail donnera lieu à une soutenance qui comprendra 10 minutes de présentation et 5 minutes de questions. Vous serez répartis par session, avec obligation d'assister aux soutenances de toute votre session

Introduction

Quels sont les attendus de cette semaine ?

- A travers les heures de cours, je vais vous présenter la régression linéaire ainsi que les sujets connexes comme la sélection de variables, l'analyse de la variance
- Je vous présenterai des cas d'étude avec le logiciel R sans pour autant vous montrer le code
- En dehors des heures de cours, vous travaillerez, par groupe de 3, en autonomie sur un ou des jeux de données avec une feuille de route pour vous aider
- Pendant ces heures, nous resterons à votre disposition afin que vous puissiez nous poser des questions
- Vous devrez être à l'école impérativement
- Le lundi 2 juin, votre travail donnera lieu à une soutenance qui comprendra 10 minutes de présentation et 5 minutes de questions. Vous serez répartis par session, avec obligation d'assister aux soutenances de toute votre session
- Par la suite, dans les 15 jours après votre soutenance, vous devrez me rendre un rapport écrit de votre travail qui devra contenir l'intégralité du code R utilisé.

Modèle de régression linéaire

- Le modèle de régression linéaire est très certainement le modèle le plus simple, mais aussi le modèle le plus utilisé en pratique.

Modèle de régression linéaire

- Le modèle de régression linéaire est très certainement le modèle le plus simple, mais aussi le modèle le plus utilisé en pratique.
- le cadre habituel : une variable réponse qui est quantitative, et une ou plusieurs variables explicatives elles aussi quantitatives

Modèle de régression linéaire

- Le modèle de régression linéaire est très certainement le modèle le plus simple, mais aussi le modèle le plus utilisé en pratique.
- le cadre habituel : une variable réponse qui est quantitative, et une ou plusieurs variables explicatives elles aussi quantitatives
- le cadre est plus étendu que ce que bien des personnes pensent, car la régression dite quadratique en partie intégrante de la régression linéaire comme nous le verrons

Modèle de régression linéaire

- Le modèle de régression linéaire est très certainement le modèle le plus simple, mais aussi le modèle le plus utilisé en pratique.
- le cadre habituel : une variable réponse qui est quantitative, et une ou plusieurs variables explicatives elles aussi quantitatives
- le cadre est plus étendu que ce que bien des personnes pensent, car la régression dite quadratique en partie intégrante de la régression linéaire comme nous le verrons
- cadre additionnel : une variable réponse qui est quantitative, et une ou plusieurs variables explicatives qui sont qualitatives.
  On parle dans ce cas, d'analyse de la variance à un ou plusieurs facteurs.

Sommaire

Introduction

Le cadre de la régression linéaire simple est le suivant :

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- une variable explicative qui est quantitative

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- une variable explicative qui est quantitative
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i\in\{1,\dots,n\}}\}$, où :

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- une variable explicative qui est quantitative
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i \in \{1,\dots,n\}}\}$, où :
  - i représente l'individu statistique considéré

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- une variable explicative qui est quantitative
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i \in \{1,\dots,n\}}\}$, où :
  - i représente l'individu statistique considéré
  - $x_1, \dots, x_n$ sont les observations de la variable explicative

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- une variable explicative qui est quantitative
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i \in \{1,\ldots,n\}}\}$, où :
  - i représente l'individu statistique considéré
  - $x_1, \ldots, x_n$ sont les observations de la variable explicative
  - $y_1, \ldots, y_n$ sont les observations de la variable réponse

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- une variable explicative qui est quantitative
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i\in\{1,\ldots,n\}}\}$, où :
  - i représente l'individu statistique considéré
  - $x_1, \ldots, x_n$ sont les observations de la variable explicative
  - $y_1, \ldots, y_n$ sont les observations de la variable réponse
- La question est de déterminer, par l'intermédiaire de ces données, si il existe une relation fonctionnelle entre la variable réponse et la variable explicative, autrement dit l'existence d'une fonction f telle que :

$$\forall i \in \{1, \ldots, n\},\ y_i \approx f(x_i)$$

Introduction

- principe : estimer la fonction f

Introduction

- principe : estimer la fonction f
- critère : le risque quadratique

$$R(g) = \mathbb{E}[(Y - g(X))^2]$$

où Y représente la variable réponse et X celle explicative

Introduction

- principe : estimer la fonction f
- critère : le risque quadratique

$$R(g) = \mathbb{E}[(Y - g(X))^2]$$

  où Y représente la variable réponse et X celle explicative
- estimation de f : minimisation du risque quadratique

$$f^* = \underset{g}{argmin} R(g)$$

Introduction

- principe : estimer la fonction f
- critère : le risque quadratique

$$R(g) = \mathbb{E}[(Y - g(X))^2]$$

  où Y représente la variable réponse et X celle explicative
- estimation de f : minimisation du risque quadratique

$$f^* = \underset{g}{argmin} R(g)$$

- Ce risque est une quantité purement théorique car il faut connaître la loi du couple (X, Y) ce qui est rare et si tel était le cas, nous n'aurions pas besoin d'estimer f car nous aurions accès théoriquement à la meilleure minimisation du risque qui est $\mathbb{E}[Y|X]$

Introduction

- principe : estimer la fonction f
- critère : le risque quadratique

$$R(g) = \mathbb{E}[(Y - g(X))^2]$$

  où Y représente la variable réponse et X celle explicative
- estimation de f : minimisation du risque quadratique

$$f^* = \underset{g}{argmin} R(g)$$

- Ce risque est une quantité purement théorique car il faut connaître la loi du couple (X, Y) ce qui est rare et si tel était le cas, nous n'aurions pas besoin d'estimer f car nous aurions accès théoriquement à la meilleure minimisation du risque qui est $\mathbb{E}[Y|X]$
- risque empirique : il se substitue au risque

$$R_n(g) = \frac{1}{n}.\sum_{k=1} n(Y_i - g(X_i))^2$$

  où :

Introduction

- principe : estimer la fonction f
- critère : le risque quadratique

$$R(g) = \mathbb{E}[(Y - g(X))^2]$$

  où Y représente la variable réponse et X celle explicative
- estimation de f : minimisation du risque quadratique

$$f^* = \underset{g}{argmin} R(g)$$

- Ce risque est une quantité purement théorique car il faut connaître la loi du couple (X, Y) ce qui est rare et si tel était le cas, nous n'aurions pas besoin d'estimer f car nous aurions accès théoriquement à la meilleure minimisation du risque qui est $\mathbb{E}[Y|X]$
- risque empirique : il se substitue au risque

$$R_n(g) = \frac{1}{n}.\sum_{k=1} n(Y_i - g(X_i))^2$$

  où :
  - Y_i est la variable aléatoire associée à l'observation y_i

Introduction

- principe : estimer la fonction f
- critère : le risque quadratique

$$R(g) = \mathbb{E}[(Y - g(X))^2]$$

où Y représente la variable réponse et X celle explicative
- estimation de f : minimisation du risque quadratique

$$f^* = \underset{g}{argmin} R(g)$$

- Ce risque est une quantité purement théorique car il faut connaître la loi du couple (X, Y) ce qui est rare et si tel était le cas, nous n'aurions pas besoin d'estimer f car nous aurions accès théoriquement à la meilleure minimisation du risque qui est $\mathbb{E}[Y|X]$
- risque empirique : il se substitue au risque

$$R_n(g) = \frac{1}{n} . \sum_{k=1} n(Y_i - g(X_i))^2$$

où :
  - Y_i est la variable aléatoire associée à l'observation y_i
  - X_i est la variable aléatoire associée à l'observation x_i

Introduction

- principe : estimer la fonction f
- critère : le risque quadratique

$$R(g) = \mathbb{E}[(Y - g(X))^2]$$

où Y représente la variable réponse et X celle explicative
- estimation de f : minimisation du risque quadratique

$$f^* = \underset{g}{argmin} R(g)$$

- Ce risque est une quantité purement théorique car il faut connaître la loi du couple (X, Y) ce qui est rare et si tel était le cas, nous n'aurions pas besoin d'estimer f car nous aurions accès théoriquement à la meilleure minimisation du risque qui est $\mathbb{E}[Y|X]$
- risque empirique : il se substitue au risque

$$R_n(g) = \frac{1}{n}.\sum_{k=1} n(Y_i - g(X_i))^2$$

où :
  - Y_i est la variable aléatoire associée à l'observation y_i
  - X_i est la variable aléatoire associée à l'observation x_i
- On considère alors l'estimation f_n^* de f définie par :

$$f_n^* = \underset{g}{argmin} R_n(g)$$

Introduction

- Le problème ainsi posé est très complexe

Introduction

- Le problème ainsi posé est très complexe
- réduction à une famille de fonction : la famille des fonctions linéaires

$$\mathcal{F} = \{g : \mathbb{R} \to \mathbb{R},\ \text{telle que}\ \ g(x) = ax + b,\ \ a, b \in \mathbb{R}\}$$

Introduction

- Le problème ainsi posé est très complexe
- réduction à une famille de fonction : la famille des fonctions linéaires

$$\mathcal{F} = \{g : \mathbb{R} \to \mathbb{R}, \text{ telle que } g(x) = ax + b, \ a, b \in \mathbb{R}\}$$

- ainsi naît le modèle de régression linéaire simple qui consiste à minimiser le risque empirique sur cette famille.
  On considère ainsi l'estimation f_n^* de f définie par :

$$f_n^* = \underset{g \in \mathcal{F}}{argmin} R_n(g)$$

Modèle mathématique

Le cadre mathématique du modèle de régression linéaire simple est le suivant :

$$\forall i \in \{1, \ldots, n\}, \;\; Y_i = a.x_i + b + \varepsilon_i$$

avec :

Modèle mathématique

Le cadre mathématique du modèle de régression linéaire simple est le suivant :

$$\forall i \in \{1, \dots, n\}, \ \ Y_i = a.x_i + b + \varepsilon_i$$

avec :

- $\forall i \in \{1, \dots, n\}, \mathbb{E}[\varepsilon_i] = 0$

Modèle mathématique

Le cadre mathématique du modèle de régression linéaire simple est le suivant :

$$\forall i \in \{1, \dots, n\}, \ \ Y_i = a.x_i + b + \varepsilon_i$$

avec :

- $\forall i \in \{1, \dots, n\}, \mathbb{E}[\varepsilon_i] = 0$
- $\forall i \in \{1, \dots, n\}, \mathbb{V}[\varepsilon_i] = \sigma^2$

Modèle mathématique

Le cadre mathématique du modèle de régression linéaire simple est le suivant :

$$\forall i \in \{1, \ldots, n\},\ \ Y_i = a.x_i + b + \varepsilon_i$$

avec :

- $\forall i \in \{1, \ldots, n\}$, $\mathbb{E}[\varepsilon_i] = 0$
- $\forall i \in \{1, \ldots, n\}$, $\mathbb{V}[\varepsilon_i] = \sigma^2$
- $\forall i, k \in \{1, \ldots, n\}$, avec $i \neq k$, $cov(\varepsilon_i, \varepsilon_k) = 0$

Modèle mathématique

Le cadre mathématique du modèle de régression linéaire simple est le suivant :

$$\forall i \in \{1, \ldots, n\}, \;\; Y_i = a.x_i + b + \varepsilon_i$$

avec :

- $\forall i \in \{1, \ldots, n\}$, $\mathbb{E}[\varepsilon_i] = 0$
- $\forall i \in \{1, \ldots, n\}$, $\mathbb{V}[\varepsilon_i] = \sigma^2$
- $\forall i, k \in \{1, \ldots, n\}$, avec $i \neq k$, $cov(\varepsilon_i, \varepsilon_k) = 0$

Bien souvent, on ajoute une hypothèse supplémentaire qui est la normalité du bruit, à savoir que si l'on considère le vecteur aléatoire

$$\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

alors il s'agit d'un vecteur gaussien d'espérance 0_n et de variance $\sigma^2.I_n$

Modèle mathématique

Remarque

- Dans cette écriture, les x_i ne sont pas des quantités aléatoires (on travaille à fix design).

Modèle mathématique

Remarque

- Dans cette écriture, les x_i ne sont pas des quantités aléatoires (on travaille à fix design).
- Les seules quantités aléatoires sont les Y_i et les ε_i

Modèle mathématique

Remarque

- Dans cette écriture, les x_i ne sont pas des quantités aléatoires (on travaille à fix design).
- Les seules quantités aléatoires sont les Y_i et les ε_i
- Si l'on suppose l'hypothèse de gaussianité du bruit, les variables ε_i sont des variables aléatoires indépendantes et de même loi

Modèle mathématique

Remarque

- Dans cette écriture, les x_i ne sont pas des quantités aléatoires (on travaille à fix design).
- Les seules quantités aléatoires sont les Y_i et les ε_i
- Si l'on suppose l'hypothèse de gaussianité du bruit, les variables ε_i sont des variables aléatoires indépendantes et de même loi
- Si l'on suppose l'hypothèse de gaussianité du bruit, les variables Y_i sont des variables aléatoires indépendantes mais elles ne sont pas de même loi

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

- On dit que R est un vecteur aléatoire si $R_1, \ldots, R_n$ sont des variables aléatoires

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

- On dit que R est un vecteur aléatoire si $R_1, \ldots, R_n$ sont des variables aléatoires
- On dit que R est un vecteur gaussien si R est un vecteur aléatoire tel que toute combinaison linéaire de ses composantes est une variable aléatoire gaussienne, autrement dit si :

$$\forall \lambda_2, \ldots, \lambda_n,\ \ \lambda_1.R_1 + \ldots + \lambda_n.R_n \text{est une variable aléatoire de loi normale.}$$

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

- On dit que R est un vecteur aléatoire si $R_1, \ldots, R_n$ sont des variables aléatoires
- On dit que R est un vecteur gaussien si R est un vecteur aléatoire tel que toute combinaison linéaire de ses composantes est une variable aléatoire gaussienne, autrement dit si :

$$\forall \lambda_2, \ldots, \lambda_n, \;\; \lambda_1.R_1 + \ldots + \lambda_n.R_n \text{est une variable aléatoire de loi normale.}$$

- Automatiquement, toutes les composantes $R_1, \ldots, R_n$ sont des variables aléatoires de loi normale.

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

- Si toutes les variables aléatoires $R_1, \ldots, R_n$ possèdent une espérance, R possède une espérance et :

$$\mathbb{E}[R] = \begin{pmatrix} \mathbb{E}[R_1] \\ \vdots \\ \mathbb{E}[R_n] \end{pmatrix}$$

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

- Si toutes les variables aléatoires $R_1, \ldots, R_n$ possèdent une espérance, R possède une espérance et :

$$\mathbb{E}[R] = \begin{pmatrix} \mathbb{E}[R_1] \\ \vdots \\ \mathbb{E}[R_n] \end{pmatrix}$$

- Si toutes les variables aléatoires $R_1, \ldots, R_n$ possèdent une variance, R possède une variance et :

$$\mathbb{V}[R] = \mathbb{E}[(R - \mathbb{E}[R]).(R - \mathbb{E}[R])']$$

Il s'agit d'une matrice carrée et symétrique *nxn*

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

- Si toutes les variables aléatoires $R_1, \ldots, R_n$ possèdent une espérance, R possède une espérance et :

$$\mathbb{E}[R] = \begin{pmatrix} \mathbb{E}[R_1] \\ \vdots \\ \mathbb{E}[R_n] \end{pmatrix}$$

- Si toutes les variables aléatoires $R_1, \ldots, R_n$ possèdent une variance, R possède une variance et :

$$\mathbb{V}[R] = \mathbb{E}[(R - \mathbb{E}[R]).(R - \mathbb{E}[R])']$$

  Il s'agit d'une matrice carrée et symétrique nxn

- $\mathbb{V}[R]$ a pour terme général $cov(R_i, R_j)$.

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

Proposition

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

Proposition

- Soit A une matrice non aléatoire. Si $A.R$ possède une espérance, alors

$$\mathbb{E}[A.R] = A.\mathbb{E}[R]$$

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

Proposition

- Soit A une matrice non aléatoire. Si $A.R$ possède une espérance, alors

$$\mathbb{E}[A.R] = A.\mathbb{E}[R]$$

- Soit A une matrice non aléatoire. Si $A.R$ possède une variance, alors

$$\mathbb{V}[A.R] = A.\mathbb{V}[R].A'$$

Vecteur gaussien

Soit $R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$.

Proposition

- Soit A une matrice non aléatoire. Si $A.R$ possède une espérance, alors

$$\mathbb{E}[A.R] = A.\mathbb{E}[R]$$

- Soit A une matrice non aléatoire. Si $A.R$ possède une variance, alors

$$\mathbb{V}[A.R] = A.\mathbb{V}[R].A'$$

- Soit A une matrice non aléatoire. Si on suppose que R est un vecteur gaussien, alors $A.R$ est lui-même un vecteur gaussien.

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
La résolution de ce problème de minimisation a pour solution :

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
La résolution de ce problème de minimisation a pour solution :

Théorème

La solution du problème des moindres carrés est :

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
La résolution de ce problème de minimisation a pour solution :

Théorème

La solution du problème des moindres carrés est :

- $\hat{a}_n = \frac{\sum_{i=1}^{n} x_i . Y_i - n . \bar{x}_n . \bar{Y}_n}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}$

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
La résolution de ce problème de minimisation a pour solution :

Théorème

La solution du problème des moindres carrés est :

- $\hat{a}_n = \frac{\sum_{i=1}^{n} x_i . Y_i - n.\bar{x}_n . \bar{Y}_n}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}$
- $\hat{b}_n = \bar{Y}_n - \hat{a}_n . \bar{x}_n$

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
La résolution de ce problème de minimisation a pour solution :

Théorème

La solution du problème des moindres carrés est :

- $\hat{a}_n = \frac{\sum_{i=1}^{n} x_i . Y_i - n.\bar{x}_n.\bar{Y}_n}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}$
- $\hat{b}_n = \bar{Y}_n - \hat{a}_n.\bar{x}_n$
- avec :

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
La résolution de ce problème de minimisation a pour solution :

Théorème

La solution du problème des moindres carrés est :

- $\hat{a}_n = \frac{\sum_{i=1}^{n} x_i . Y_i - n.\bar{x}_n.\bar{Y}_n}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$
- $\hat{b}_n = \bar{Y}_n - \hat{a}_n.\bar{x}_n$
- avec :
  - $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
La résolution de ce problème de minimisation a pour solution :

Théorème

La solution du problème des moindres carrés est :

- $\hat{a}_n = \frac{\sum_{i=1}^{n} x_i.Y_i - n.\bar{x}_n.\bar{Y}_n}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$
- $\hat{b}_n = \bar{Y}_n - \hat{a}_n.\bar{x}_n$
- avec :
  - $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$

Visualisation

- $x \to \hat{a}_n.x + \hat{b}_n$ est l'équation de la droite de régression

Visualisation

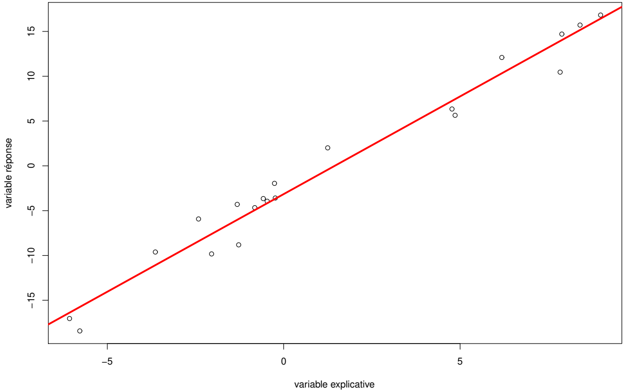
- $x \to \hat{a}_n.x + \hat{b}_n$ est l'équation de la droite de régression
- $\forall i \in \{1, \ldots, n\}$, $\hat{Y}_i = \hat{a}_n.x_i + \hat{b}_n$ est la prédiction de Y_i

Visualisation

- $x \to \hat{a}_n.x + \hat{b}_n$ est l'équation de la droite de régression
- $\forall i \in \{1, \ldots, n\}$, $\hat{Y}_i = \hat{a}_n.x_i + \hat{b}_n$ est la prédiction de Y_i
- $\forall i \in \{1, \ldots, n\}$, $\hat{\varepsilon}_i = Y_{-}\hat{Y}_i$ est le résidu pour l'individu i

Visualisation

- $x \to \hat{a}_n.x + \hat{b}_n$ est l'équation de la droite de régression
- $\forall i \in \{1, \ldots, n\}$, $\hat{Y}_i = \hat{a}_n.x_i + \hat{b}_n$ est la prédiction de Y_i
- $\forall i \in \{1, \ldots, n\}$, $\hat{\varepsilon}_i = Y_- \hat{Y}_i$ est le résidu pour l'individu i
- visualisation



Logiciel R

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.4716 -1.1260  0.3359  1.5306  2.4861

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.1521     0.4302  -7.327 8.37e-07 ***
X             2.1801     0.0898  24.277 3.32e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.856 on 18 degrees of freedom
Multiple R-squared:  0.9704,
  Adjusted R-squared:  0.9687
F-statistic: 589.4 on 1 and 18 DF,  p-value: 3.316e-15
```

Propriétés

- $\hat{a}_n$ et $\hat{b}_n$ sont des estimateurs respectivement des paramètres a et b

Propriétés

- $\hat{a}_n$ et $\hat{b}_n$ sont des estimateurs respectivement des paramètres a et b
- $\hat{a}_n$ et $\hat{b}_n$ sont des estimateurs sans biais respectivement des paramètres a et b

Propriétés

- $\hat{a}_n$ et $\hat{b}_n$ sont des estimateurs respectivement des paramètres a et b
- $\hat{a}_n$ et $\hat{b}_n$ sont des estimateurs sans biais respectivement des paramètres a et b
- On a :

$$\mathbb{V}[\hat{b}_n] = \frac{\sigma^2 . \sum_{i=1}^n x_i^2}{n . \sum_{i=1}^n (x_i - \bar{x}_n)^2}, \quad \mathbb{V}[\hat{a}_n] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2},$$

$$cov(\hat{a}_n, \hat{b}_n) = \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}$$

Propriétés

- $\hat{a}_n$ et $\hat{b}_n$ sont des estimateurs respectivement des paramètres a et b
- $\hat{a}_n$ et $\hat{b}_n$ sont des estimateurs sans biais respectivement des paramètres a et b
- On a :

$$\mathbb{V}[\hat{b}_n] = \frac{\sigma^2 . \sum_{i=1}^n x_i^2}{n . \sum_{i=1}^n (x_i - \bar{x}_n)^2}, \quad \mathbb{V}[\hat{a}_n] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2},$$

$$cov(\hat{a}_n, \hat{b}_n) = \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}$$

- Ces quantités sont là aussi théoriques car elles dépendent de σ^2 qui est dans la pratique un paramètre inconnu.

Propriétés

- $\hat{a}_n$ et $\hat{b}_n$ sont des estimateurs respectivement des paramètres a et b
- $\hat{a}_n$ et $\hat{b}_n$ sont des estimateurs sans biais respectivement des paramètres a et b
- On a :

$$\mathbb{V}[\hat{b}_n] = \frac{\sigma^2 . \sum_{i=1}^n x_i^2}{n . \sum_{i=1}^n (x_i - \bar{x}_n)^2}, \quad \mathbb{V}[\hat{a}_n] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2},$$

$$cov(\hat{a}_n, \hat{b}_n) = \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}$$

- Ces quantités sont là aussi théoriques car elles dépendent de σ^2 qui est dans la pratique un paramètre inconnu.
- Un estimateur sans biais de σ^2 est :

$$\hat{\sigma}_n^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

avec $\hat{Y}_i = \hat{a}_n . x_i + \hat{b}_n$ (prédiction de Y_i par le modèle de régression linéaire simple)

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n = \begin{pmatrix} \hat{b}_n \\ \hat{a}_n \end{pmatrix}$ est un vecteur gaussien d'espérance $\begin{pmatrix} b \\ a \end{pmatrix}$ et de variance Σ avec :

$$\Sigma = \begin{pmatrix} \frac{\sigma^2 . \sum_{i=1}^n x_i^2}{n . \sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \\ \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \end{pmatrix}$$

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n = \begin{pmatrix} \hat{b}_n \\ \hat{a}_n \end{pmatrix}$ est un vecteur gaussien d'espérance $\begin{pmatrix} b \\ a \end{pmatrix}$ et de variance Σ avec :

$$\Sigma = \begin{pmatrix} \frac{\sigma^2 . \sum_{i=1}^n x_i^2}{n . \sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \\ \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \end{pmatrix}$$

- $\hat{a}_n$ et $\hat{b}_n$ sont des variables aléatoires de loi normale d'espérance respective a et b et de variance respective $\mathbb{V}[\hat{a}_n]$ et $\mathbb{V}[\hat{b}_n]$

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n = \begin{pmatrix} \hat{b}_n \\ \hat{a}_n \end{pmatrix}$ est un vecteur gaussien d'espérance $\begin{pmatrix} b \\ a \end{pmatrix}$ et de variance Σ avec :

$$\Sigma = \begin{pmatrix} \frac{\sigma^2 . \sum_{i=1}^n x_i^2}{n . \sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \\ \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \end{pmatrix}$$

- $\hat{a}_n$ et $\hat{b}_n$ sont des variables aléatoires de loi normale d'espérance respective a et b et de variance respective $\mathbb{V}[\hat{a}_n]$ et $\mathbb{V}[\hat{b}_n]$
- $(n-2)\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-2)$

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n = \begin{pmatrix} \hat{b}_n \\ \hat{a}_n \end{pmatrix}$ est un vecteur gaussien d'espérance $\begin{pmatrix} b \\ a \end{pmatrix}$ et de variance Σ avec :

$$\Sigma = \begin{pmatrix} \frac{\sigma^2 . \sum_{i=1}^n x_i^2}{n . \sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \\ \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \end{pmatrix}$$

- $\hat{a}_n$ et $\hat{b}_n$ sont des variables aléatoires de loi normale d'espérance respective a et b et de variance respective $\mathbb{V}[\hat{a}_n]$ et $\mathbb{V}[\hat{b}_n]$
- $(n-2)\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-2)$
- $\hat{\sigma}_n^2$ et $\hat{\beta}_n$ sont indépendants

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n = \begin{pmatrix} \hat{b}_n \\ \hat{a}_n \end{pmatrix}$ est un vecteur gaussien d'espérance $\begin{pmatrix} b \\ a \end{pmatrix}$ et de variance Σ avec :
$$\Sigma = \begin{pmatrix} \frac{\sigma^2 . \sum_{i=1}^n x_i^2}{n. \sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \\ \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \end{pmatrix}$$
- $\hat{a}_n$ et $\hat{b}_n$ sont des variables aléatoires de loi normale d'espérance respective a et b et de variance respective $\mathbb{V}[\hat{a}_n]$ et $\mathbb{V}[\hat{b}_n]$
- $(n-2)\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-2)$
- $\hat{\sigma}_n^2$ et $\hat{\beta}_n$ sont indépendants
- Soit
$$T_b = \frac{\hat{b}_n - b}{\hat{\sigma}_n . \sqrt{\frac{\sum_{i=1}^n x_i^2}{n. \sum_{i=1}^n (x_i - \bar{x}_n)^2}}}, \quad T_a = \frac{\hat{a}_n - a}{\hat{\sigma}_n . \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}}}$$

On a :

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n = \begin{pmatrix} \hat{b}_n \\ \hat{a}_n \end{pmatrix}$ est un vecteur gaussien d'espérance $\begin{pmatrix} b \\ a \end{pmatrix}$ et de variance Σ avec :

$$\Sigma = \begin{pmatrix} \frac{\sigma^2 . \sum_{i=1}^n x_i^2}{n. \sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \\ \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \end{pmatrix}$$

- $\hat{a}_n$ et $\hat{b}_n$ sont des variables aléatoires de loi normale d'espérance respective a et b et de variance respective $\mathbb{V}[\hat{a}_n]$ et $\mathbb{V}[\hat{b}_n]$
- $(n-2)\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-2)$
- $\hat{\sigma}_n^2$ et $\hat{\beta}_n$ sont indépendants
- Soit

$$T_b = \frac{\hat{b}_n - b}{\hat{\sigma}_n . \sqrt{\frac{\sum_{i=1}^n x_i^2}{n. \sum_{i=1}^n (x_i - \bar{x}_n)^2}}}, \quad T_a = \frac{\hat{a}_n - a}{\hat{\sigma}_n . \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}}}$$

On a :

- $T_b \sim T(n-2)$

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n = \begin{pmatrix} \hat{b}_n \\ \hat{a}_n \end{pmatrix}$ est un vecteur gaussien d'espérance $\begin{pmatrix} b \\ a \end{pmatrix}$ et de variance Σ avec :

$$\Sigma = \begin{pmatrix} \frac{\sigma^2 . \sum_{i=1}^n x_i^2}{n. \sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \\ \frac{-\sigma^2 . \bar{x}_n}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \end{pmatrix}$$

- $\hat{a}_n$ et $\hat{b}_n$ sont des variables aléatoires de loi normale d'espérance respective a et b et de variance respective $\mathbb{V}[\hat{a}_n]$ et $\mathbb{V}[\hat{b}_n]$
- $(n-2)\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-2)$
- $\hat{\sigma}_n^2$ et $\hat{\beta}_n$ sont indépendants
- Soit

$$T_b = \frac{\hat{b}_n - b}{\hat{\sigma}_n . \sqrt{\frac{\sum_{i=1}^n x_i^2}{n. \sum_{i=1}^n (x_i - \bar{x}_n)^2}}}, \quad T_a = \frac{\hat{a}_n - a}{\hat{\sigma}_n . \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}}}$$

On a :

- $T_b \sim T(n-2)$
- $T_a \sim T(n-2)$

Aspects pratiques

Il est toujours possible de réaliser une régression linéaire, mais plusieurs questions se posent :

Aspects pratiques

Il est toujours possible de réaliser une régression linéaire, mais plusieurs questions se posent :

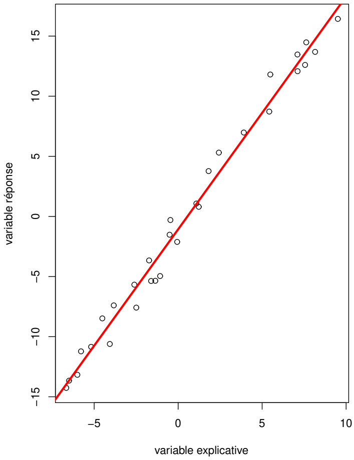
- Validité de la régression linéaire

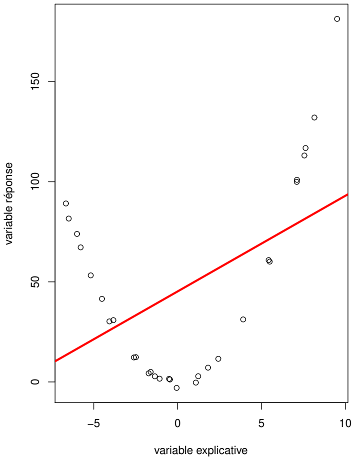
Aspects pratiques

Il est toujours possible de réaliser une régression linéaire, mais plusieurs questions se posent :

- Validité de la régression linéaire
- Qualité de la prédiction pour une nouvelle observation

Validité de la régression linéaire





Validité de la régression linéaire

Comment évaluer la validité de la régression linéaire :

Validité de la régression linéaire

Comment évaluer la validité de la régression linéaire :

- le coefficient de détermination : R^2 (multiple R-squared)

Validité de la régression linéaire

Comment évaluer la validité de la régression linéaire :

- le coefficient de détermination : R^2 (multiple R-squared)
- test $\mathcal{H}_0 : a = 0$ contre $\mathcal{H}_1 : a \neq 0$

R^2

Définition

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y}_n)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2} = 1 - \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}$$

R^2

Définition

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y}_n)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2} = 1 - \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}$$

Proposition

- $0 \leq R^2 \leq 1$

R^2

Définition

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y}_n)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2} = 1 - \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}$$

Proposition

- $0 \leq R^2 \leq 1$
- Plus le coefficient R^2 est proche de 1 et plus la régression linéaire est une bonne modélisation

R^2

Définition

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y}_n)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2} = 1 - \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}$$

Proposition

- $0 \leq R^2 \leq 1$
- Plus le coefficient R^2 est proche de 1 et plus la régression linéaire est une bonne modélisation
- Sur les deux exemples précédents, voici les valeurs respectives du coefficient R^2

```
0.9861887 0.2338926
```

Test du paramètre a

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit

Test du paramètre a

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit
- un estimateur de a est $\hat{a}_n$, estimateur pour lequel on sait que $T_a = \frac{\hat{a}_n - a}{\hat{\sigma}_n . \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}} \sim T(n-2)$

Test du paramètre a

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit
- un estimateur de a est $\hat{a}_n$, estimateur pour lequel on sait que $T_a = \frac{\hat{a}_n - a}{\hat{\sigma}_n \cdot \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}} \sim T(n-2)$
- Région de rejet de $\mathcal{H}_0$: $\left\{ \left| \frac{\hat{a}_n}{\hat{\sigma}_n \cdot \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}} \right| > t_{1-\alpha/2;n-2} \right\}$ où $t_{1-\alpha/2;n-2}$ est défini par $P(T \leq t_{1-\alpha/2;n-2}) = 1 - \alpha/2$ avec T une variable aléatoire de loi $T(n-2)$.

Test du paramètre a

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit
- un estimateur de a est $\hat{a}_n$, estimateur pour lequel on sait que $T_a = \frac{\hat{a}_n - a}{\hat{\sigma}_n \cdot \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}}} \sim T(n-2)$
- Région de rejet de $\mathcal{H}_0$: $\left\{ \left| \frac{\hat{a}_n}{\hat{\sigma}_n \cdot \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}}} \right| > t_{1-\alpha/2;n-2} \right\}$ où $t_{1-\alpha/2;n-2}$ est défini par $P(T \leq t_{1-\alpha/2;n-2}) = 1 - \alpha/2$ avec T une variable aléatoire de loi $T(n-2)$.
- pratique : on ne calcule pas le seuil théorique mais une p-valeur

```
              Estimate Std. Error  t value     Pr(>|t|)
(Intercept) -1.069084  0.2116328 -5.05160 2.408415e-05
X            1.936450  0.0433076 44.71387 1.382205e-27
```

Test du paramètre a

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit
- un estimateur de a est $\hat{a}_n$, estimateur pour lequel on sait que $T_a = \frac{\hat{a}_n - a}{\hat{\sigma}_n \cdot \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}} \sim T(n-2)$
- Région de rejet de $\mathcal{H}_0$: $\left\{ \left| \frac{\hat{a}_n}{\hat{\sigma}_n \cdot \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}} \right| > t_{1-\alpha/2;n-2} \right\}$ où $t_{1-\alpha/2;n-2}$ est défini par $P(T \leq t_{1-\alpha/2;n-2}) = 1 - \alpha/2$ avec T une variable aléatoire de loi $T(n-2)$.
- pratique : on ne calcule pas le seuil théorique mais une p-valeur

```
             Estimate Std. Error  t value     Pr(>|t|)
(Intercept) -1.069084  0.2116328 -5.05160 2.408415e-05
X            1.936450  0.0433076 44.71387 1.382205e-27
```

- Règle de décision :

Test du paramètre a

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit
- un estimateur de a est $\hat{a}_n$, estimateur pour lequel on sait que $T_a = \frac{\hat{a}_n - a}{\hat{\sigma}_n \cdot \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}}} \sim T(n-2)$
- Région de rejet de $\mathcal{H}_0$: $\left\{ \left| \frac{\hat{a}_n}{\hat{\sigma}_n \cdot \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}}} \right| > t_{1-\alpha/2;n-2} \right\}$ où $t_{1-\alpha/2;n-2}$ est défini par $P(T \leq t_{1-\alpha/2;n-2}) = 1 - \alpha/2$ avec T une variable aléatoire de loi $T(n-2)$.
- pratique : on ne calcule pas le seuil théorique mais une p-valeur

```
              Estimate Std. Error  t value     Pr(>|t|)
(Intercept) -1.069084  0.2116328 -5.05160 2.408415e-05
X            1.936450  0.0433076 44.71387 1.382205e-27
```

- Règle de décision :
    - si la p-valeur est plus petite que α, on décide $\mathcal{H}_1$

Test du paramètre a

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit
- un estimateur de a est $\hat{a}_n$, estimateur pour lequel on sait que $T_a = \frac{\hat{a}_n - a}{\hat{\sigma}_n \cdot \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}} \sim T(n-2)$
- Région de rejet de $\mathcal{H}_0$: $\left\{ \left| \frac{\hat{a}_n}{\hat{\sigma}_n \cdot \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}} \right| > t_{1-\alpha/2;n-2} \right\}$ où $t_{1-\alpha/2;n-2}$ est défini par $P(T \leq t_{1-\alpha/2;n-2}) = 1 - \alpha/2$ avec T une variable aléatoire de loi $T(n-2)$.
- pratique : on ne calcule pas le seuil théorique mais une p-valeur

```
              Estimate Std. Error  t value     Pr(>|t|)
(Intercept) -1.069084  0.2116328 -5.05160 2.408415e-05
X            1.936450  0.0433076 44.71387 1.382205e-27
```

- Règle de décision :
  - si la p-valeur est plus petite que α, on décide $\mathcal{H}_1$
  - si la p-valeur est plus grande que α, on ne rejette pas $\mathcal{H}_0$

Prédiction

L'objectif principal de la modélisation est de pouvoir se servir du modèle afin de faire de la prévision, à savoir avoir une idée de la valeur de la variable réponse quand on dispose d'une nouvelle donnée pour la variable explicative.

Prédiction

L'objectif principal de la modélisation est de pouvoir se servir du modèle afin de faire de la prévision, à savoir avoir une idée de la valeur de la variable réponse quand on dispose d'une nouvelle donnée pour la variable explicative.

- Si on note x_{new} une nouvelle observation de la variable explicatice, une prévision de la variable réponse est donnée par :

$$\hat{y}_{new} = \hat{a}_n.x_{new} + \hat{b}_n$$

Prédiction

L'objectif principal de la modélisation est de pouvoir se servir du modèle afin de faire de la prévision, à savoir avoir une idée de la valeur de la variable réponse quand on dispose d'une nouvelle donnée pour la variable explicative.

- Si on note x_{new} une nouvelle observation de la variable explicatice, une prévision de la variable réponse est donnée par :

$$\hat{y}_{new} = \hat{a}_n.x_{new} + \hat{b}_n$$

- cependant, nous savons qu'il y a des fluctuations d'échantillonage et que l'observation de $\hat{a}_n$ et de $\hat{b}_n$ sont dépendante du jeu de donnée d'apprentissage

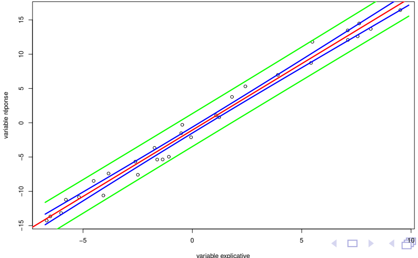
Prédiction

L'objectif principal de la modélisation est de pouvoir se servir du modèle afin de faire de la prévision, à savoir avoir une idée de la valeur de la variable réponse quand on dispose d'une nouvelle donnée pour la variable explicative.

- Si on note x_{new} une nouvelle observation de la variable explicatice, une prévision de la variable réponse est donnée par :

$$\hat{y}_{new} = \hat{a}_n.x_{new} + \hat{b}_n$$

- cependant, nous savons qu'il y a des fluctuations d'échantillonage et que l'observation de $\hat{a}_n$ et de $\hat{b}_n$ sont dépendante du jeu de donnée d'apprentissage
- visualisation : comment ternir compte de ce phénomène ?



Prédiction

Théorème

Un intervalle de confiance pour la prédiction de Y pour une nouvelle valeur x_{new} de la variable explicative, intervalle de niveau de confiance $100.(1-\alpha)\%$ est :

$$\hat{a}_n.x_{new} + \hat{b}_n \pm \hat{\sigma}_n.\sqrt{1 + \frac{1}{n} + \frac{(x_{new} - \bar{x}_n)^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}}.t_{1-\alpha/2;n-2}$$

Validité des hypothèses sur le bruit

- Un certain nombre de résultats supposent d'avoir notamment l'hypothèse de gaussianité du bruit, et-elle toujours satisfaite ?

Validité des hypothèses sur le bruit

- Un certain nombre de résultats supposent d'avoir notamment l'hypothèse de gaussianité du bruit, et-elle toujours satisfaite ?
- difficulté : le bruit est non observable, donc comment vérifier?
  On va regarder les résidus car si le bruit est gaussien, alors les résidus aussi.
  Mais pas si simple !

Validité des hypothèses sur le bruit

- Un certain nombre de résultats supposent d'avoir notamment l'hypothèse de gaussianité du bruit, et-elle toujours satisfaite ?
- difficulté : le bruit est non observable, donc comment vérifier?
  On va regarder les résidus car si le bruit est gaussien, alors les résidus aussi. Mais pas si simple !
- difficulté numéro 1 : les résidus ne sont pas tous de même loi (donc oublier l'idée de faire un histogramme des résidus)

Validité des hypothèses sur le bruit

- Un certain nombre de résultats supposent d'avoir notamment l'hypothèse de gaussianité du bruit, et-elle toujours satisfaite ?
- difficulté : le bruit est non observable, donc comment vérifier?
  On va regarder les résidus car si le bruit est gaussien, alors les résidus aussi. Mais pas si simple !
- difficulté numéro 1 : les résidus ne sont pas tous de même loi (donc oublier l'idée de faire un histogramme des résidus)
- introduction des résidus standardisés qui seront tous de même loi, mais avec simplement une loi connue asymptotiquement

$$\hat{\varepsilon}_{i,sd} = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_n.\sqrt{1-h_i}}$$

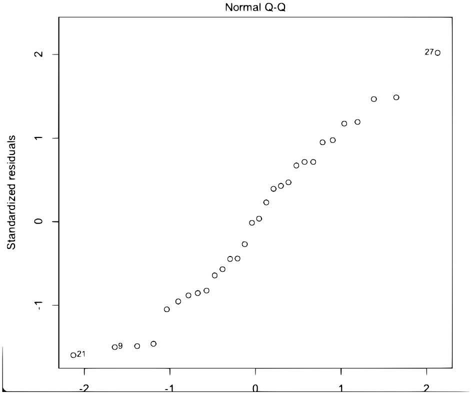
avec h_i qui est le i-ème terme diagonal de la matrice $\mathbb{X}.(\mathbb{X}'.\mathbb{X})^{-1}.\mathbb{X}'$ où :

$$\mathbb{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$\hat{\varepsilon}_{i,sd} \underset{(\mathcal{L})}{\rightsquigarrow} \mathcal{N}(0,1)$$

Validité des hypothèses sur le bruit

Une manière de valider ainsi l'hypothèse de gaussianité du bruit consiste à vérifier si les résidus standardisés suivent approximativement une loi normale standard.



Validité des hypothèses sur le bruit

- Mais, nous ne nous comparons qu'à une loi asymptotique.

Validité des hypothèses sur le bruit

- Mais, nous ne nous comparons qu'à une loi asymptotique.
- introduction des résidus studentisés

$$\hat{\varepsilon}_{i,st} = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_{n,i}.\sqrt{1 - h_i}}$$

avec $\hat{\sigma}^2_{n,i}$ un estimateur de σ^2 obtenu en considérant le modèle de régression linéaire construit à partir des données d'apprentissage auxquelles l'individu i a été retiré.

Validité des hypothèses sur le bruit

- Mais, nous ne nous comparons qu'à une loi asymptotique.
- introduction des résidus studentisés

$$\hat{\varepsilon}_{i,st} = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_{n,i}.\sqrt{1 - h_i}}$$

  avec $\hat{\sigma}^2_{n,i}$ un estimateur de σ^2 obtenu en considérant le modèle de régression linéaire construit à partir des données d'apprentissage auxquelles l'individu i a été retiré.
- Le résultat important est :

$$\hat{\varepsilon}_{i,st} \sim T(n-3)$$

Validité des hypothèses sur le bruit

- Mais, nous ne nous comparons qu'à une loi asymptotique.
- introduction des résidus studentisés

$$\hat{\varepsilon}_{i,st} = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_{n,i}.\sqrt{1 - h_i}}$$

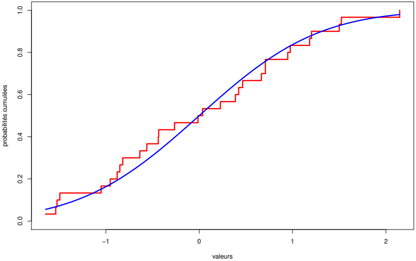
  avec $\hat{\sigma}^2_{n,i}$ un estimateur de σ^2 obtenu en considérant le modèle de régression linéaire construit à partir des données d'apprentissage auxquelles l'individu i a été retiré.
- Le résultat important est :

$$\hat{\varepsilon}_{i,st} \sim T(n-3)$$

- On peut donc vérifier l'hypothèse d egaussianité du bruit en tester le fait que les résidus studentisés suivent ou non une loi de Student à (n-3) degré de liberté.
  On peut ainsi réaliser un test d'adéquation de Kolmogorov.

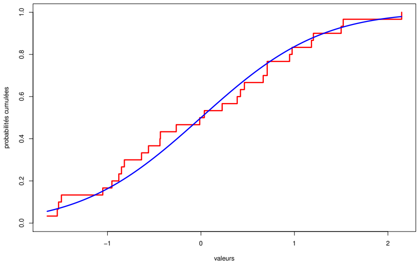
Test de Kolmogorov

Principe :



Test de Kolmogorov

Principe :



Dans notre cas :

```
        One-sample Kolmogorov-Smirnov test

data:  rt
D = 0.099391, p-value = 0.9
alternative hypothesis: two-sided
```

Sommaire

Introduction

La régression linéaire simple peut assez facilement se généraliser au cadre de plusieurs variables expllicatives.
Cependant, il ne s'agit pas de la seule généralisation possible.
L'objet ici est d'évoquer :

Introduction

La régression linéaire simple peut assez facilement se généraliser au cadre de plusieurs variables expllicatives.
Cependant, il ne s'agit pas de la seule généralisation possible.
L'objet ici est d'évoquer :

- cadre de plusieurs variables

Introduction

La régression linéaire simple peut assez facilement se généraliser au cadre de plusieurs variables expllicatives.
Cependant, il ne s'agit pas de la seule généralisation possible.
L'objet ici est d'évoquer :

- cadre de plusieurs variables
- régression quadratique

Introduction

La régression linéaire simple peut assez facilement se généraliser au cadre de plusieurs variables expllicatives.
Cependant, il ne s'agit pas de la seule généralisation possible.
L'objet ici est d'évoquer :

- cadre de plusieurs variables
- régression quadratique
- analyse de la variance

Introduction

Le cadre de la régression linéaire simple est le suivant :

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- p variables explicatives qui sont quantitatives

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- p variables explicatives qui sont quantitatives
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i\in\{1,\dots,n\}}\}$, où :

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- p variables explicatives qui sont quantitatives
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i\in\{1,\dots,n\}}\}$, où :
    - i représente l'individu statistique considéré

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- p variables explicatives qui sont quantitatives
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i \in \{1,\ldots,n\}}\}$, où :
  - i représente l'individu statistique considéré
  - $x_i = (x_i^{(1)}, \ldots, x_i^{(p)})$

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- p variables explicatives qui sont quantitatives
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i\in\{1,\ldots,n\}}\}$, où :
  - i représente l'individu statistique considéré
  - $x_i = (x_i^{(1)}, \ldots, x_i^{(p)})$
  - $x_1^{(k)}, \ldots, x_n^{(k)}$ sont les observations de la variable explicative numéro k

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- p variables explicatives qui sont quantitatives
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i \in \{1,\ldots,n\}}\}$, où :
  - i représente l'individu statistique considéré
  - $x_i = (x_i^{(1)}, \ldots, x_i^{(p)})$
  - $x_1^{(k)}, \ldots, x_n^{(k)}$ sont les observations de la variable explicative numéro k
  - $y_1, \ldots, y_n$ sont les observations de la variable réponse

Introduction

Le cadre de la régression linéaire simple est le suivant :

- une variable réponse qui est quantitative
- p variables explicatives qui sont quantitatives
- on dispose d'observations pour ces deux variables $\mathcal{L} = \{(x_i, y_i)_{i\in\{1,\dots,n\}}\}$, où :
  - i représente l'individu statistique considéré
  - $x_i = (x_i^{(1)}, \dots, x_i^{(p)})$
  - $x_1^{(k)}, \dots, x_n^{(k)}$ sont les observations de la variable explicative numéro k
  - $y_1, \dots, y_n$ sont les observations de la variable réponse
- La question est de déterminer, par l'intermédiaire de ces données, si il existe une relation fonctionnelle entre la variable réponse et la variable explicative, autrement dit l'existence d'une fonction f telle que :

$$\forall i \in \{1, \dots, n\},\ \ y_i \approx f(x_i)$$

Modèle mathématique

Le cadre mathématique du modèle de régression linéaire multiple est le suivant :

$$\forall i \in \{1, \ldots, n\}, \;\; Y_i = \beta_0 + \beta_1.x_i^{(1)} + \beta_2.x_i^{(2)} + \ldots + \beta_p.x_i^{(p)} + \varepsilon_i$$

avec :

Modèle mathématique

Le cadre mathématique du modèle de régression linéaire multiple est le suivant :

$$\forall i \in \{1, \ldots, n\}, \;\; Y_i = \beta_0 + \beta_1.x_i^{(1)} + \beta_2.x_i^{(2)} + \ldots + \beta_p.x_i^{(p)} + \varepsilon_i$$

avec :

- $\forall i \in \{1, \ldots, n\}$, $\mathbb{E}[\varepsilon_i] = 0$

Modèle mathématique

Le cadre mathématique du modèle de régression linéaire multiple est le suivant :

$$\forall i \in \{1, \ldots, n\}, \;\; Y_i = \beta_0 + \beta_1 . x_i^{(1)} + \beta_2 . x_i^{(2)} + \ldots + \beta_p . x_i^{(p)} + \varepsilon_i$$

avec :

- $\forall i \in \{1, \ldots, n\}$, $\mathbb{E}[\varepsilon_i] = 0$
- $\forall i \in \{1, \ldots, n\}$, $\mathbb{V}[\varepsilon_i] = \sigma^2$

Modèle mathématique

Le cadre mathématique du modèle de régression linéaire multiple est le suivant :

$$\forall i \in \{1, \ldots, n\},\ \ Y_i = \beta_0 + \beta_1.x_i^{(1)} + \beta_2.x_i^{(2)} + \ldots + \beta_p.x_i^{(p)} + \varepsilon_i$$

avec :

- $\forall i \in \{1, \ldots, n\}$, $\mathbb{E}[\varepsilon_i] = 0$
- $\forall i \in \{1, \ldots, n\}$, $\mathbb{V}[\varepsilon_i] = \sigma^2$
- $\forall i, k \in \{1, \ldots, n\}$, avec $i \neq k$, $cov(\varepsilon_i, \varepsilon_k) = 0$

Modèle mathématique

Le cadre mathématique du modèle de régression linéaire multiple est le suivant :

$$\forall i \in \{1, \ldots, n\},\ \ Y_i = \beta_0 + \beta_1.x_i^{(1)} + \beta_2.x_i^{(2)} + \ldots + \beta_p.x_i^{(p)} + \varepsilon_i$$

avec :

- $\forall i \in \{1, \ldots, n\}$, $\mathbb{E}[\varepsilon_i] = 0$
- $\forall i \in \{1, \ldots, n\}$, $\mathbb{V}[\varepsilon_i] = \sigma^2$
- $\forall i, k \in \{1, \ldots, n\}$, avec $i \neq k$, $cov(\varepsilon_i, \varepsilon_k) = 0$

Bien souvent, on ajoute une hypothèse supplémentaire qui est la normalité du bruit, à savoir que si l'on considère le vecteur aléatoire

$$\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

alors il s'agit d'un vecteur gaussien d'espérance 0_n et de variance $\sigma^2.I_n$

Modèle mathématique

Remarque

- Dans cette écriture, les x_i ne sont pas des quantités aléatoires (on travaille à fix design).

Modèle mathématique

Remarque

- Dans cette écriture, les x_i ne sont pas des quantités aléatoires (on travaille à fix design).
- Les seules quantités aléatoires sont les Y_i et les ε_i

Modèle mathématique

Remarque

- Dans cette écriture, les x_i ne sont pas des quantités aléatoires (on travaille à fix design).
- Les seules quantités aléatoires sont les Y_i et les ε_i
- Si l'on suppose l'hypothèse de gaussianité du bruit, les variables ε_i sont des variables aléatoires indépendantes et de même loi

Modèle mathématique

Remarque

- Dans cette écriture, les x_i ne sont pas des quantités aléatoires (on travaille à fix design).
- Les seules quantités aléatoires sont les Y_i et les ε_i
- Si l'on suppose l'hypothèse de gaussianité du bruit, les variables ε_i sont des variables aléatoires indépendantes et de même loi
- Si l'on suppose l'hypothèse de gaussianité du bruit, les variables Y_i sont des variables aléatoires indépendantes mais elles ne sont pas de même loi

Modèle mathématique

Remarque

- Si l'écriture sous forme d'un système d'équations semble adapté dans le cadre de la régression linéaire simple, tel ne semble plus être le cas dans le cadre multiple.
  En effet, ce sont $p+1$ paramètres inconnus qu'il faut à présent estimer.

Modèle mathématique

Remarque

- Si l'écriture sous forme d'un système d'équations semble adapté dans le cadre de la régression linéaire simple, tel ne semble plus être le cas dans le cadre multiple.
  En effet, ce sont $p+1$ paramètres inconnus qu'il faut à présent estimer.
- L'écriture du modèle se fait davantage sous forme matricielle dans le cadre de la régression linéaire multiple

Modèle mathématique

Remarque

- Si l'écriture sous forme d'un système d'équations semble adapté dans le cadre de la régression linéaire simple, tel ne semble plus être le cas dans le cadre multiple.
  En effet, ce sont $p+1$ paramètres inconnus qu'il faut à présent estimer.
- L'écriture du modèle se fait davantage sous forme matricielle dans le cadre de la régression linéaire multiple
- Notations :

$$\mathbb{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \ \mathbb{U} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}$$

Et enfin :

$$\mathbb{X} = \begin{pmatrix} 1 & x_1^{(1)} & \dots & x_1^{(p)} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_n^{(1)} & \dots & x_n^{(p)} \end{pmatrix}$$

Modèle mathématique

Remarque

- Si l'écriture sous forme d'un système d'équations semble adapté dans le cadre de la régression linéaire simple, tel ne semble plus être le cas dans le cadre multiple.
  En effet, ce sont $p+1$ paramètres inconnus qu'il faut à présent estimer.
- L'écriture du modèle se fait davantage sous forme matricielle dans le cadre de la régression linéaire multiple
- Notations :

$$\mathbb{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \; \mathbb{U} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \; \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}$$

  Et enfin :

$$\mathbb{X} = \begin{pmatrix} 1 & x_1^{(1)} & \dots & x_1^{(p)} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_n^{(1)} & \dots & x_n^{(p)} \end{pmatrix}$$

- Le modèle s'écrit alors :

$$\mathbb{Y} = \mathbb{X}.\beta + \mathbb{U}$$

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
Comment trouver la solution dans ce contexte ?

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
Comment trouver la solution dans ce contexte ?

- $\forall a = \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix}$, on a :

$$\sum_{i=1}^{n}(y_i - a_0 - a_1.x_i^{(1)} - \ldots - a_p.x_i^{(p)})^2 = \|\mathbb{Y} - \mathbb{X}.a\|^2$$

$$= (\mathbb{Y} - \mathbb{X}.a)'.(\mathbb{Y} - \mathbb{X}.a) = \mathbb{Y}'.\mathbb{Y} - 2.\mathbb{Y}'.\mathbb{X}.a + a'.\mathbb{X}'.\mathbb{X}.a$$

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
Comment trouver la solution dans ce contexte ?

- $\forall a = \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix}$, on a :

$$\sum_{i=1}^{n}(y_i - a_0 - a_1.x_i^{(1)} - \ldots - a_p.x_i^{(p)})^2 = \|\mathbb{Y} - \mathbb{X}.a\|^2$$

$$= (\mathbb{Y} - \mathbb{X}.a)'.(\mathbb{Y} - \mathbb{X}.a) = \mathbb{Y}'.\mathbb{Y} - 2.\mathbb{Y}'.\mathbb{X}.a + a'.\mathbb{X}'.\mathbb{X}.a$$

- Minimiser le critère des moindres carrés en a revient donc à minimiser $S(a) = -2.\mathbb{Y}'.\mathbb{X}.a + a'.\mathbb{X}'.\mathbb{X}.a$ en a

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
Comment trouver la solution dans ce contexte ?

- $\forall a = \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix}$, on a :

$$\sum_{i=1}^{n}(y_i - a_0 - a_1.x_i^{(1)} - \ldots - a_p.x_i^{(p)})^2 = \|\mathbb{Y} - \mathbb{X}.a\|^2$$

$$= (\mathbb{Y} - \mathbb{X}.a)'.(\mathbb{Y} - \mathbb{X}.a) = \mathbb{Y}'.\mathbb{Y} - 2.\mathbb{Y}'.\mathbb{X}.a + a'.\mathbb{X}'.\mathbb{X}.a$$

- Minimiser le critère des moindres carrés en a revient donc à minimiser $S(a) = -2.\mathbb{Y}'.\mathbb{X}.a + a'.\mathbb{X}'.\mathbb{X}.a$ en a
- Ce second critère se décompose en une forme linéaire et une forme quadratique en a.

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
Comment trouver la solution dans ce contexte ?

- $\forall a = \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix}$, on a :

$$\sum_{i=1}^{n}(y_i - a_0 - a_1.x_i^{(1)} - \ldots - a_p.x_i^{(p)})^2 = \|\mathbb{Y} - \mathbb{X}.a\|^2$$

$$= (\mathbb{Y} - \mathbb{X}.a)'.(\mathbb{Y} - \mathbb{X}.a) = \mathbb{Y}'.\mathbb{Y} - 2.\mathbb{Y}'.\mathbb{X}.a + a'.\mathbb{X}'.\mathbb{X}.a$$

- Minimiser le critère des moindres carrés en a revient donc à minimiser $S(a) = -2.\mathbb{Y}'.\mathbb{X}.a + a'.\mathbb{X}'.\mathbb{X}.a$ en a
- Ce second critère se décompose en une forme linéaire et une forme quadratique en a.
- Il faut donc calculer le gradient, annuler le gradient et vérifier que les points critiques éventuels sont des minimum ou pas

Moindres carrés

Dans le cadre de la régression, la minimisation du risque empirique porte un autre nom, le principe des moindres carrés.
Comment trouver la solution dans ce contexte ?

- $\forall a = \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix}$, on a :

$$\sum_{i=1}^{n} (y_i - a_0 - a_1.x_i^{(1)} - \ldots - a_p.x_i^{(p)})^2 = \|\mathbb{Y} - \mathbb{X}.a\|^2$$

$$= (\mathbb{Y} - \mathbb{X}.a)'.(\mathbb{Y} - \mathbb{X}.a) = \mathbb{Y}'.\mathbb{Y} - 2.\mathbb{Y}'.\mathbb{X}.a + a'.\mathbb{X}'.\mathbb{X}.a$$

- Minimiser le critère des moindres carrés en a revient donc à minimiser $S(a) = -2.\mathbb{Y}'.\mathbb{X}.a + a'.\mathbb{X}'.\mathbb{X}.a$ en a
- Ce second critère se décompose en une forme linéaire et une forme quadratique en a.
- Il faut donc calculer le gradient, annuler le gradient et vérifier que les points critiques éventuels sont des minimum ou pas
- On a :

$$\nabla S(a) = -2.\mathbb{X}'.\mathbb{Y} + 2.\mathbb{X}'.\mathbb{X}.a$$

Moindres carrés

La solution du problème de minimisation est donnée par :

Théorème

Un estimateur de β est :

$$\hat{\beta}_n = (\mathbb{X}'.\mathbb{X})^{-1}.\mathbb{X}'.\mathbb{Y}$$

si $\mathbb{X}'.\mathbb{X}$ est une matrice inversible

Moindres carrés

La solution du problème de minimisation est donnée par :

Théorème

Un estimateur de β est :

$$\hat{\beta}_n = (\mathbb{X}'.\mathbb{X})^{-1}.\mathbb{X}'.\mathbb{Y}$$

si $\mathbb{X}'.\mathbb{X}$ est une matrice inversible

Remarque

- On se place dans le cadre où $p << n$

Moindres carrés

La solution du problème de minimisation est donnée par :

Théorème

Un estimateur de β est :

$$\hat{\beta}_n = (\mathbb{X}'.\mathbb{X})^{-1}.\mathbb{X}'.\mathbb{Y}$$

si $\mathbb{X}'.\mathbb{X}$ est une matrice inversible

Remarque

- On se place dans le cadre où $p << n$
- $\mathbb{X}'.\mathbb{X}$ est une matrice inversible si $\mathbb{X}$ est une matrice de rang plein, autrement dit si les colonnes de $\mathbb{X}$ constituent une famille libre. Dans ce cas $rang(\mathbb{X}) = p + 1$.

Moindres carrés

La solution du problème de minimisation est donnée par :

Théorème

Un estimateur de β est :

$$\hat{\beta}_n = (\mathbb{X}'.\mathbb{X})^{-1}.\mathbb{X}'.\mathbb{Y}$$

si $\mathbb{X}'.\mathbb{X}$ est une matrice inversible

Remarque

- On se place dans le cadre où $p << n$
- $\mathbb{X}'.\mathbb{X}$ est une matrice inversible si $\mathbb{X}$ est une matrice de rang plein, autrement dit si les colonnes de $\mathbb{X}$ constituent une famille libre.
  Dans ce cas $rang(\mathbb{X}) = p + 1$.
- Si $\mathbb{X}'.\mathbb{X}$ n'est pas une matrice inversible, cela signifie que les colonne de $\mathbb{X}$ constituent une famille liée et que donc certaines sont des combinaisons linéaires des autres.
  On extrait alors la famille libre la plus grande possible, famille libre contenant la première colonne de $\mathbb{X}$, on construit une nouvelle matrice $\mathbb{X}$ associée (ayant pour colonnes que les colonnes associées aux éléments de la famille libre) et un nouveau vecteur β (vecteur dans lequel ont été retirés les coefficients des variables qui ne sont pas dans la famille libre).

Propriétés

- $\hat{\beta}_n$ est un estimateur sans biais de β

Propriétés

- $\hat{\beta}_n$ est un estimateur sans biais de β
- $\hat{\beta}_n$ a pour variance $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}$

Propriétés

- $\hat{\beta}_n$ est un estimateur sans biais de β
- $\hat{\beta}_n$ a pour variance $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}$
- Un estimateur sans biais de σ^2 est :

$$\hat{\sigma}_n^2 = \frac{1}{n - rang(\mathbb{X})} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

avec $\hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \mathbb{X}.\hat{\beta}_n$

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n$ est un vecteur gaussien d'espérance β et de variance $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}$

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n$ est un vecteur gaussien d'espérance β et de variance $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}$
- les $\hat{\beta}_k$ sont des variables aléatoires de loi normale d'espérance respective β_k et de variance respective $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}_{k+1,k+1}$

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n$ est un vecteur gaussien d'espérance β et de variance $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}$
- les $\hat{\beta}_k$ sont des variables aléatoires de loi normale d'espérance respective β_k et de variance respective $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}_{k+1,k+1}$
- $(n - rang(\mathbb{X}))\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n - rang(\mathbb{X}))$

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n$ est un vecteur gaussien d'espérance β et de variance $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}$
- les $\hat{\beta}_k$ sont des variables aléatoires de loi normale d'espérance respective β_k et de variance respective $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}_{k+1,k+1}$
- $(n - rang(\mathbb{X}))\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n - rang(\mathbb{X}))$
- $\hat{\sigma}_n^2$ et $\hat{\beta}_n$ sont indépendants

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n$ est un vecteur gaussien d'espérance β et de variance $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}$
- les $\hat{\beta}_k$ sont des variables aléatoires de loi normale d'espérance respective β_k et de variance respective $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}_{k+1,k+1}$
- $(n - rang(\mathbb{X}))\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n - rang(\mathbb{X}))$
- $\hat{\sigma}_n^2$ et $\hat{\beta}_n$ sont indépendants
- Soit

$$T_k = \frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_n.\sqrt{(\mathbb{X}'.\mathbb{X})^{-1}_{k+1,k+1}}}$$

On a :

Propriétés

Si on suppose l'hypothèse de gaussianité du bruit

- $\hat{\beta}_n$ est un vecteur gaussien d'espérance β et de variance $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}$
- les $\hat{\beta}_k$ sont des variables aléatoires de loi normale d'espérance respective β_k et de variance respective $\sigma^2.(\mathbb{X}'.\mathbb{X})^{-1}_{k+1,k+1}$
- $(n - rang(\mathbb{X}))\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n - rang(\mathbb{X}))$
- $\hat{\sigma}_n^2$ et $\hat{\beta}_n$ sont indépendants
- Soit

$$T_k = \frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_n.\sqrt{(\mathbb{X}'.\mathbb{X})^{-1}_{k+1,k+1}}}$$

  On a :

  - $T_k \sim T(n - rang(\mathbb{X}))$

Aspects pratiques

Il est toujours possible de réaliser une régression linéaire, mais plusieurs questions se posent :

Aspects pratiques

Il est toujours possible de réaliser une régression linéaire, mais plusieurs questions se posent :

- Validité de la régression linéaire

Aspects pratiques

Il est toujours possible de réaliser une régression linéaire, mais plusieurs questions se posent :

- Validité de la régression linéaire
- Qualité de la prédiction pour une nouvelle observation

Validité de la régression linéaire

Comment évaluer la validité de la régression linéaire :

Validité de la régression linéaire

Comment évaluer la validité de la régression linéaire :

- le coefficient de détermination ajusté : R^2 (adjusted R-squared)

Validité de la régression linéaire

Comment évaluer la validité de la régression linéaire :

- le coefficient de détermination ajusté : R^2 (adjusted R-squared)
- test $\mathcal{H}_0$: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$ contre $\mathcal{H}_1$: ce n'est pas le cas

R^2 ajusté

Définition

$$R_a^2 = 1 - \frac{n-1}{n - rang(\mathbb{X})}(1 - R^2)$$

R^2 ajusté

Définition

$$R_a^2 = 1 - \frac{n-1}{n - rang(\mathbb{X})}(1 - R^2)$$

Proposition

- $R_a^2 \leq 1$

R^2 ajusté

Définition

$$R_a^2 = 1 - \frac{n-1}{n - rang(\mathbb{X})}(1 - R^2)$$

Proposition

- $R_a^2 \leq 1$
- Plus le coefficient R_a^2 est grand et plus la régression linéaire est une bonne modélisation

Test de significativité du modèle

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit

Test de significativité du modèle

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit
- Sous l'hypothèse $\mathcal{H}_0$, le modèle sécrit simplement $\mathbb{Y} = \beta_0 . \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \mathbb{U}$.

  L'idée est de ce dire que si les vecteurs $\hat{Y}$ et $\hat{Y}_0$ (prédcition dans le modèle $\mathcal{H}_0$) sont proche, autant conserver le modèle impliquant le moins de variables explicatives.

Test de significativité du modèle

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit
- Sous l'hypothèse $\mathcal{H}_0$, le modèle sécrit simplement $\mathbb{Y} = \beta_0.\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \mathbb{U}$.

  L'idée est de ce dire que si les vecteurs $\hat{Y}$ et $\hat{Y}_0$ (prédcition dans le modèle $\mathcal{H}_0$) sont proche, autant conserver le modèle impliquant le moins de variables explicatives.
- On crée un distance entre les deux modèle :

$$F = \frac{\|\hat{Y} - \bar{Y}_n\|^2/(rang(\mathbb{X}) - 1)}{\|Y - \bar{Y}_n\|^2/(n - rang(\mathbb{X}))} \underset{\mathcal{H}_0}{\sim} \mathcal{F}(rang(\mathbb{X}) - 1, n - rang(\mathbb{X}))$$

Test de significativité du modèle

- A partir du moment où l'on parle de test, il faut la notion de loi, donc l'hypothèse de gaussianité du bruit
- Sous l'hypothèse $\mathcal{H}_0$, le modèle sécrit simplement $\mathbb{Y} = \beta_0 . \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \mathbb{U}$.

  L'idée est de ce dire que si les vecteurs $\hat{Y}$ et $\hat{Y}_0$ (prédcition dans le modèle $\mathcal{H}_0$) sont proche, autant conserver le modèle impliquant le moins de variables explicatives.
- On crée un distance entre les deux modèle :

$$F = \frac{\|\hat{Y} - \bar{Y}_n\|^2/(rang(\mathbb{X}) - 1)}{\|Y - \bar{Y}_n\|^2/(n - rang(\mathbb{X}))} \underset{\mathcal{H}_0}{\sim} \mathcal{F}(rang(\mathbb{X}) - 1, n - rang(\mathbb{X}))$$

- Région de rejet de $\mathcal{H}_0$: $\left\{F > f_{1-\alpha;rang(\mathbb{X})-1,n-rang(\mathbb{X})}\right\}$ où $f_{1-\alpha;rang(\mathbb{X})-1,n-rang(\mathbb{X})}$ est défini par $P(R \leq f_{1-\alpha;rang(\mathbb{X})-1,n-rang(\mathbb{X})}) = 1 - \alpha$ avec R une variable aléatoire de loi $\mathcal{F}(rang(\mathbb{X}) - 1, n - rang(\mathbb{X}))$.

Test de significativité du modèle

- pratique : on ne calcule pas le seuil théorique mais une p-valeur

```
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9195 -0.9234 -0.0335  0.7887  3.8104

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.21820    0.51294  -4.324 0.000231 ***
X1           2.05376    0.11177  18.375 1.21e-15 ***
X2          -0.10527    0.09692  -1.086 0.288201
X3          -4.53257    0.85854  -5.279 2.05e-05 ***
X4           0.36748    0.21200   1.733 0.095864 .
X5           0.44280    0.28510   1.553 0.133480
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.672 on 24 degrees of freedom
Multiple R-squared:  0.9587,
  Adjusted R-squared:  0.9501
F-statistic: 111.5 on 5 and 24 DF,  p-value: 8.363e-16
```

Test de significativité du modèle

- pratique : on ne calcule pas le seuil théorique mais une p-valeur

```
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9195 -0.9234 -0.0335  0.7887  3.8104

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.21820    0.51294  -4.324 0.000231 ***
X1           2.05376    0.11177  18.375 1.21e-15 ***
X2          -0.10527    0.09692  -1.086 0.288201
X3          -4.53257    0.85854  -5.279 2.05e-05 ***
X4           0.36748    0.21200   1.733 0.095864 .
X5           0.44280    0.28510   1.553 0.133480
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.672 on 24 degrees of freedom
Multiple R-squared:  0.9587,
  Adjusted R-squared:  0.9501
F-statistic: 111.5 on 5 and 24 DF,  p-value: 8.363e-16
```

- Règle de décision :

Test de significativité du modèle

- pratique : on ne calcule pas le seuil théorique mais une p-valeur

```
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.9195 -0.9234 -0.0335  0.7887  3.8104 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.21820    0.51294  -4.324 0.000231 ***
X1           2.05376    0.11177  18.375 1.21e-15 ***
X2          -0.10527    0.09692  -1.086 0.288201    
X3          -4.53257    0.85854  -5.279 2.05e-05 ***
X4           0.36748    0.21200   1.733 0.095864 .  
X5           0.44280    0.28510   1.553 0.133480    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 1.672 on 24 degrees of freedom
Multiple R-squared:  0.9587,
  Adjusted R-squared:  0.9501
F-statistic: 111.5 on 5 and 24 DF,  p-value: 8.363e-16
```

- Règle de décision :
  - si la p-valeur est plus petite que α, on décide $\mathcal{H}_1$

Test de significativité du modèle

- pratique : on ne calcule pas le seuil théorique mais une p-valeur

```
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9195 -0.9234 -0.0335  0.7887  3.8104

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.21820    0.51294  -4.324 0.000231 ***
X1           2.05376    0.11177  18.375 1.21e-15 ***
X2          -0.10527    0.09692  -1.086 0.288201
X3          -4.53257    0.85854  -5.279 2.05e-05 ***
X4           0.36748    0.21200   1.733 0.095864 .
X5           0.44280    0.28510   1.553 0.133480
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.672 on 24 degrees of freedom
Multiple R-squared:  0.9587,
  Adjusted R-squared:  0.9501
F-statistic: 111.5 on 5 and 24 DF,  p-value: 8.363e-16
```

- Règle de décision :
  - si la p-valeur est plus petite que α, on décide $\mathcal{H}_1$
  - si la p-valeur est plus grande que α, on ne rejette pas $\mathcal{H}_0$

Prédiction

L'objectif principal de la modélisation est de pouvoir se servir du modèle afin de faire de la prévision, à savoir avoir une idée de la valeur de la variable réponse quand on dispose d'une nouvelle donnée pour la variable explicative.

Prédiction

L'objectif principal de la modélisation est de pouvoir se servir du modèle afin de faire de la prévision, à savoir avoir une idée de la valeur de la variable réponse quand on dispose d'une nouvelle donnée pour la variable explicative.

- Si on note x_{new} une nouvelle observation de la variable explicatice, une prévision de la variable réponse est donnée par :

$$\hat{y}_{new} = (1 \ x_{new}).\hat{\beta}_n$$

Prédiction

L'objectif principal de la modélisation est de pouvoir se servir du modèle afin de faire de la prévision, à savoir avoir une idée de la valeur de la variable réponse quand on dispose d'une nouvelle donnée pour la variable explicative.

- Si on note x_{new} une nouvelle observation de la variable explicatice, une prévision de la variable réponse est donnée par :

$$\hat{y}_{new} = (1 \ x_{new}).\hat{\beta}_n$$

- cependant, nous savons qu'il y a des fluctuations d'échantillonage et que les observations des $\hat{\beta}_k$ sont dépendantes du jeu de donnée d'apprentissage

Prédiction

Théorème

Un intervalle de confiance pour la prédiction de Y pour une nouvelle valeur x_{new} de la variable explicative, intervalle de niveau de confiance $100.(1-\alpha)\%$ est :

$$(1\ x_{new}).\hat{\beta}_n \pm \hat{\sigma}_n.\sqrt{(1\ x_{new}).(\mathbb{X}'.\mathbb{X})^{-1}.(1\ x_{new})'}.t_{1-\alpha/2,n-rang(\mathbb{X})}$$

Validité des hypothèses sur le bruit

- Un certain nombre de résultats supposent d'avoir notamment l'hypothèse de gaussianité du bruit, et-elle toujours satisfaite ?

Validité des hypothèses sur le bruit

- Un certain nombre de résultats supposent d'avoir notamment l'hypothèse de gaussianité du bruit, et-elle toujours satisfaite ?
- difficulté : le bruit est non observable, donc comment vérifier?
  On va regarder les résidus car si le bruit est gaussien, alors les résidus aussi.
  Mais pas si simple !

Validité des hypothèses sur le bruit

- Un certain nombre de résultats supposent d'avoir notamment l'hypothèse de gaussianité du bruit, et-elle toujours satisfaite ?
- difficulté : le bruit est non observable, donc comment vérifier?
  On va regarder les résidus car si le bruit est gaussien, alors les résidus aussi. Mais pas si simple !
- difficulté numéro 1 : les résidus ne sont pas tous de même loi (donc oublier l'idée de faire un histogramme des résidus)

Validité des hypothèses sur le bruit

- Un certain nombre de résultats supposent d'avoir notamment l'hypothèse de gaussianité du bruit, et-elle toujours satisfaite ?
- difficulté : le bruit est non observable, donc comment vérifier?
  On va regarder les résidus car si le bruit est gaussien, alors les résidus aussi. Mais pas si simple !
- difficulté numéro 1 : les résidus ne sont pas tous de même loi (donc oublier l'idée de faire un histogramme des résidus)
- introduction des résidus standardisés qui seront tous de même loi, mais avec simplement une loi connue asymptotiquement

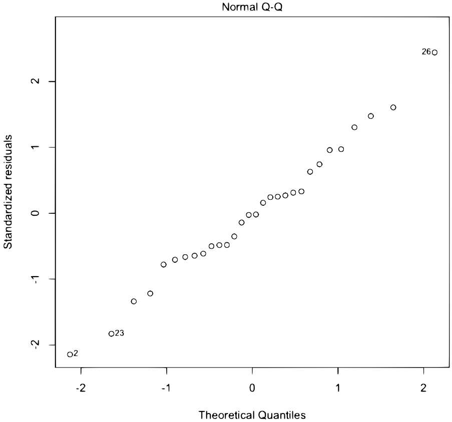
$$\hat{\varepsilon}_{i,sd} = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_n.\sqrt{1-h_i}}$$

avec h_i qui est le i-ème terme diagonal de la matrice $\mathbb{X}.(\mathbb{X}'.\mathbb{X})^{-1}.\mathbb{X}'$

$$\hat{\varepsilon}_{i,sd} \underset{(\mathcal{L})}{\rightsquigarrow} \mathcal{N}(0,1)$$

Validité des hypothèses sur le bruit

Une manière de valider ainsi l'hypothèse de gaussianité du bruit consiste à vérifier si les résidus standardisés suivent approximativement une loi normale standard.



Validité des hypothèses sur le bruit

- Mais, nous ne nous comparons qu'à une loi asymptotique.

Validité des hypothèses sur le bruit

- Mais, nous ne nous comparons qu'à une loi asymptotique.
- introduction des résidus studentisés

$$\hat{\varepsilon}_{i,st} = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_{n,i}.\sqrt{1 - h_i}}$$

avec $\hat{\sigma}^2_{n,i}$ un estimateur de σ^2 obtenu en considérant le modèle de régression linéaire construit à partir des données d'apprentissage auxquelles l'individu i a été retiré.

Validité des hypothèses sur le bruit

- Mais, nous ne nous comparons qu'à une loi asymptotique.
- introduction des résidus studentisés

$$\hat{\varepsilon}_{i,st} = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_{n,i}.\sqrt{1 - h_i}}$$

  avec $\hat{\sigma}^2_{n,i}$ un estimateur de σ^2 obtenu en considérant le modèle de régression linéaire construit à partir des données d'apprentissage auxquelles l'individu i a été retiré.

- Le résultat important est :

$$\hat{\varepsilon}_{i,st} \sim T(n - rang(\mathbb{X} - 1)$$

Validité des hypothèses sur le bruit

- Mais, nous ne nous comparons qu'à une loi asymptotique.
- introduction des résidus studentisés

$$\hat{\varepsilon}_{i,st} = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_{n,i}.\sqrt{1-h_i}}$$

  avec $\hat{\sigma}^2_{n,i}$ un estimateur de σ^2 obtenu en considérant le modèle de régression linéaire construit à partir des données d'apprentissage auxquelles l'individu i a été retiré.
- Le résultat important est :

$$\hat{\varepsilon}_{i,st} \sim T(n - rang(\mathbb{X} - 1)$$

- On peut donc vérifier l'hypothèse d egaussianité du bruit en tester le fait que les résidus studentisés suivent ou non une loi de Student à $(n - rang(\mathbb{X} - 1)$ degré de liberté.
  On peut ainsi réaliser un test d'adéquation de Kolmogorov.

Analyse de la variance à un facteur

Sommaire

Introduction

- R^2 ajusté

Introduction

- R^2 ajusté
- sélection pas à pas

Introduction

- R^2 ajusté
- sélection pas à pas
- Ridge et Lasso