

Chapitre 3

Régression linéaire

Classe Master 1 : I2A + STIC

Plan

- 1. Introduction**
- 2. Rappels d'algèbre (calcul matriciel)**
- 3. Régression linéaire simple**
- 4. Régression linéaire multiple**

1. Introduction

- L'origine du mot *régression* vient de sir Francis Galton en 1885.
- Il travaille sur l'hérédité, il cherche à expliquer la taille des fils en fonction de celle des pères.
- Il constate que lorsque le père était plus grand que la moyenne son fils avait tendance à être plus petit que lui et lorsque le père était plus petit que la moyenne son fils avait tendance à être plus grand que lui
- Ces résultats l'on conduit à considérer sa théorie « *Regression toward mediocrity* » .

1. Introduction

- En statistique, en *data mining* et en apprentissage automatique, un modèle de régression linéaire cherche à établir une *relation linéaire* entre une variable dite à *expliquer* (noté Y) et une variable dite *explicative* (noté X).
 - ✓ Lorsqu'il existe une seule variable d'entrée (X),
→ la méthode est appelée *régression linéaire simple*.
 - ✓ Lorsqu'il y a plusieurs variables d'entrée,
→ la méthode est appelée *régression linéaire multiple*.

1. Introduction

- Les données ont généralement la forme suivante :

Nbr d'observations	X : Variable explicative	Y : Variable à expliquer
1	x_1	y_1
.	.	.
.	.	.
.	.	.
i	x_i	y_i
.	.	.
.	.	.
n	x_n	y_n
Moyenne	$\bar{x}$	$\bar{y}$
Écart-type	δ_x	δ_y

2. Rappels d'algèbre (calcul matriciel)

Définitions & notations

- Une **matrice** de format (n, m) est un tableau rectangulaire de $n * m$ éléments, rangés en n lignes et m colonnes.

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

- Les éléments sont nommés en utilisant deux indices, le premier est *l'indice de ligne*, le second est *l'indice de colonne*. On note alors, par exemple : $\mathbf{A} = [a_{i,j}]$.

2. Rappels d'algèbre (calcul matriciel)

Définitions & notations

- on appelle matrice *identité* d'ordre n la matrice notée I définie par : $I = [\delta_{i,j}]$, avec : $\delta_{i,j} = 0$ si $i \neq j$, $\delta_{i,j} = 1$ si $i = j$.

$$I = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

- La matrice I est l'élément neutre du produit des matrices carrées d'ordre n : pour toute matrice carrée A d'ordre n , $AI = IA = A$.
- Plus généralement, pour toute matrice A de format (n,p) : $IA = A$, et pour toute matrice B de format (m,n) : $BI = B$.

2. Rappels d'algèbre (calcul matriciel)

Définitions & notations

- Un **vecteur** est une matrice à une colonne.

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$$

- Remarque :
 - ✓ Lorsque $m = n$, on dit plutôt : *matrice carrée* d'ordre n .
 - ✓ Si $n = 1$, on parle de *matrice-ligne* d'ordre m ,
 - ✓ si $m = 1$, on parle de *matrice-colonne* d'ordre n .

2. Rappels d'algèbre (calcul matriciel)

Opérations

- La *transposée* d'une matrice $A = [a_{i,j}]$ est la matrice $A^t = [a_{j,i}]$, obtenue en échangeant les lignes et les colonnes de A .
- **Propriétés :**
 - ✓ Si A est de format (n, m) , alors A^t est de format (m, n) . En particulier, si A est carrée d'ordre n , alors A^t a le même format.
 - ✓ La transposée d'une matrice-colonne est une matrice-ligne, et réciproquement.
 - ✓ $(A^t)^t = A$ pour toute matrice A .
 - ✓ $\det(A^t) = \det(A)$

2. Rappels d'algèbre (calcul matriciel)

Opérations

- Le *déterminant* d'une matrice *carrée* es donné comme suit :
 - ✓ on choisit une ligne ou une colonne de la matrice
 - ✓ on multiplie chaque coefficient de cette ligne (ou colonne) par le déterminant de la matrice obtenu en rayant la colonne et la ligne de ce coefficient (la matrice obtenue est une matrice 2×2) ,
 - ✓ chaque résultat obtenu doit être multiplié de plus par **-1** dans le cas ou sa colonne L et sa ligne C sont telles que $L + C$ est impaire.
 - ✓ On ajoute ensuite les 3 résultats.

2. Rappels d'algèbre (calcul matriciel)

Opérations

- Le calcul de *déterminant* d'une matrice *carrée* es donné comme suit :
- exemple :

$$\begin{vmatrix} -1 & 2 & 5 \\ 1 & 2 & 3 \\ -2 & 8 & 10 \end{vmatrix}$$

$$\begin{vmatrix} -1 & 2 & 5 \\ 1 & 2 & 3 \\ -2 & 8 & 10 \end{vmatrix} = -1 \begin{vmatrix} 2 & 3 \\ 8 & 10 \end{vmatrix} - 2 \begin{vmatrix} 1 & 3 \\ -2 & 10 \end{vmatrix} + 5 \begin{vmatrix} 1 & 2 \\ -2 & 8 \end{vmatrix}$$

$$\begin{aligned} &= -1(2 \times 10 - 3 \times 8) - 2(1 \times 10 - 3 \times (-2)) + 5(1 \times 8 - 2 \times (-2)) \\ &= -1(-4) - 2(16) + 5(12) = 4 - 32 + 60 = 32 \end{aligned}$$

2. Rappels d'algèbre (calcul matriciel)

Opérations

- une matrice carrée A est *inversible* si et seulement s'il existe une matrice B (de même format) telle que :
 $AB = BA = I$.
- B est alors appelée *l'inverse de* A , et est notée A^{-1}

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

telle que $\det(A) \neq 0$, où $\det(A) = ad - bc$

Propriétés :

- Si $AB = I$, alors A est inversible et $B = A^{-1}$.
- $(A^{-1})^{-1} = A$, autrement dit, A^{-1} est inversible, d'inverse A .
- Si A et B sont inversibles, alors AB l'est aussi et : $(AB)^{-1} = B^{-1}A^{-1}$.
- Si A est inversible, alors A^t l'est aussi, et : $(A^t)^{-1} = (A^{-1})^t$.

2. Rappels d'algèbre (calcul matriciel)

Opérations

- Exemple: calcul d'*inverse* d'une matrice 3 * 3:

$$A = \begin{pmatrix} -1 & 2 & 5 \\ 1 & 2 & 3 \\ -2 & 8 & 10 \end{pmatrix} \quad \det A = 32$$
$$A^{-1} = \frac{1}{32} \begin{pmatrix} \begin{vmatrix} 2 & 3 \\ 8 & 10 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ -2 & 10 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ -2 & 8 \end{vmatrix} \\ -\begin{vmatrix} 2 & 5 \\ 8 & 10 \end{vmatrix} & \begin{vmatrix} -1 & 5 \\ -2 & 10 \end{vmatrix} & -\begin{vmatrix} -1 & 2 \\ -2 & 8 \end{vmatrix} \\ \begin{vmatrix} 2 & 5 \\ 2 & 3 \end{vmatrix} & -\begin{vmatrix} -1 & 5 \\ 1 & 3 \end{vmatrix} & \begin{vmatrix} -1 & 2 \\ 1 & 2 \end{vmatrix} \end{pmatrix} =$$
$$\frac{1}{32} \begin{pmatrix} -4 & -16 & 12 \\ 20 & 0 & 4 \\ -4 & 8 & -4 \end{pmatrix} = \frac{1}{32} \begin{pmatrix} -4 & 20 & -4 \\ -16 & 0 & 8 \\ 12 & 4 & -4 \end{pmatrix} = \begin{pmatrix} -\frac{1}{8} & \frac{5}{8} & \frac{-1}{8} \\ \frac{-1}{2} & 0 & \frac{1}{4} \\ \frac{3}{8} & \frac{1}{8} & \frac{-1}{8} \end{pmatrix}$$

2. Rappels d'algèbre (calcul matriciel)

Résolution d'une système d'équations

Exemple: soit le système de 2 équations:
$$\begin{cases} x + 3y = 1 \\ 2x + 5y = 3 \end{cases}$$

On peut réécrire ce système sous forme matricielle comme suit:

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$$

D'où: $\mathbf{Ax} = \mathbf{b}$

$\mathbf{A}$ est inversible, alors: $\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \Rightarrow \boxed{\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}}$

3. Régression linéaire simple

Le modèle

- **Hypothèse** : Supposons que les données proviennent d'un phénomène suivants le modèle :

$$Y = f(X) = \alpha + \beta X$$

- Ou $X = (x_1, x_2, \dots, x_n)$
- Cette équation est appelée *mdèle de régression linéaire simple*
- α, β Son appelées des *paramètres* ou des *poids* de modèle.
- La fonction $Y = f(x)$ est une droite. On appelle cette droite *la droite de regression*.

3. Régression linéaire simple

Le modèle

- **Exemple1** : Considérons un exemple de prédiction du prix d'un logement sachant que nous disposons la base d'apprentissage suivante :

X ↓	Surface	Prix (10 Millions)	Y ↓
$x_1 =$	2104	400	$= y_1$
$x_2 =$	1600	330	$= y_2$
$x_3 =$	2000	369	$= y_3$
$x_4 =$	1416	232	$= y_4$
$x_5 =$	3000	540	$= y_5$

- Le model qui décrit ces données : $Y = f(X) = \alpha + \beta X$

3. Régression linéaire simple

Le modèle

- Pour déterminer le modèle : $Y = f(X) = \alpha + \beta X$
- Nous pouvons écrire:

$$y_i = \alpha + \beta x_i + \xi_i, \quad i = 1, \dots, n$$

Ou :

- ✓ les x_i sont des valeurs non aléatoires *connus*.
 - ✓ Les paramètres α, β du modèle sont *inconnus*.
 - ✓ Les ε_i sont les réalisation *inconnus* d'une variable aléatoire.
 - ✓ Les y_i sont les observations *connus*.
- Les hypothèses à ce modèle sont les suivantes :

$$E(\xi_i) = 0, \quad \text{var}(\xi_i) = \sigma^2. \quad \text{cov}(\xi_i, \xi_j) = 0, \quad \forall i \neq j.$$

3. Régression linéaire simple

Estimation des paramètres

- Quelle droite choisir ? Comment prendre les paramètres α et β ?
- (α, β) doit rendre $f(x)$ très proche des valeurs des Y .
- Celle qui « colle » le mieux aux données, selon un certain *critère*.
- On doit alors définir un *critère*, une *fonction d'erreur* ou *de perte* pour mesurer combien notre prédiction est loin des « vraies » valeurs.
- On prend (α, β) qui minimise la fonction d'erreur.

3. Régression linéaire simple

Estimation des paramètres

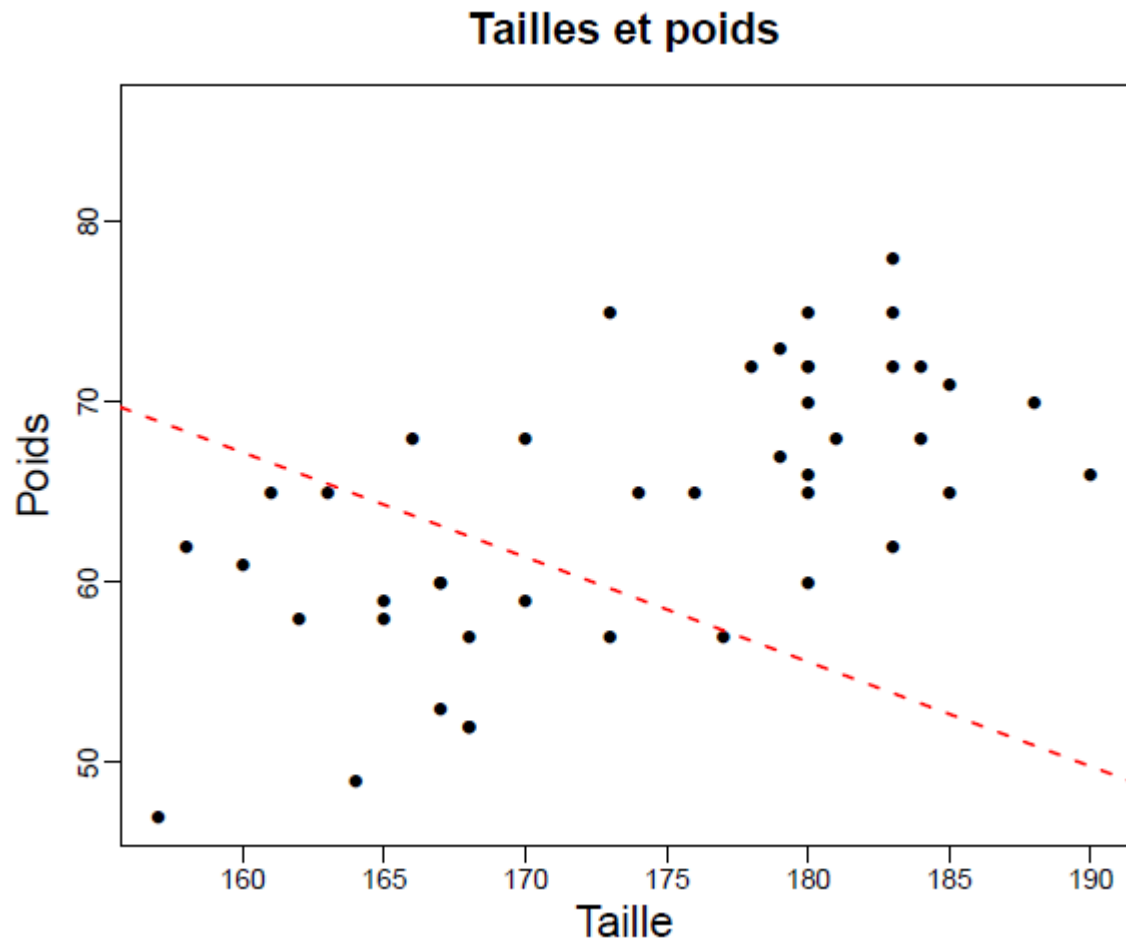
- Quelle droite choisir ?
- **Exemple 2:**
- Critère généralement utilise: **critère des moindres carres.**
! Choisir la droite qui minimise la somme des carres des distances entre la droite et les observations $\sum (\varepsilon_i)^2$.

3. Régression linéaire simple

Estimation des paramètres

- Quelle droite choisir ?
- **Exemple 2:**

$$\sum (\varepsilon_i)^2$$



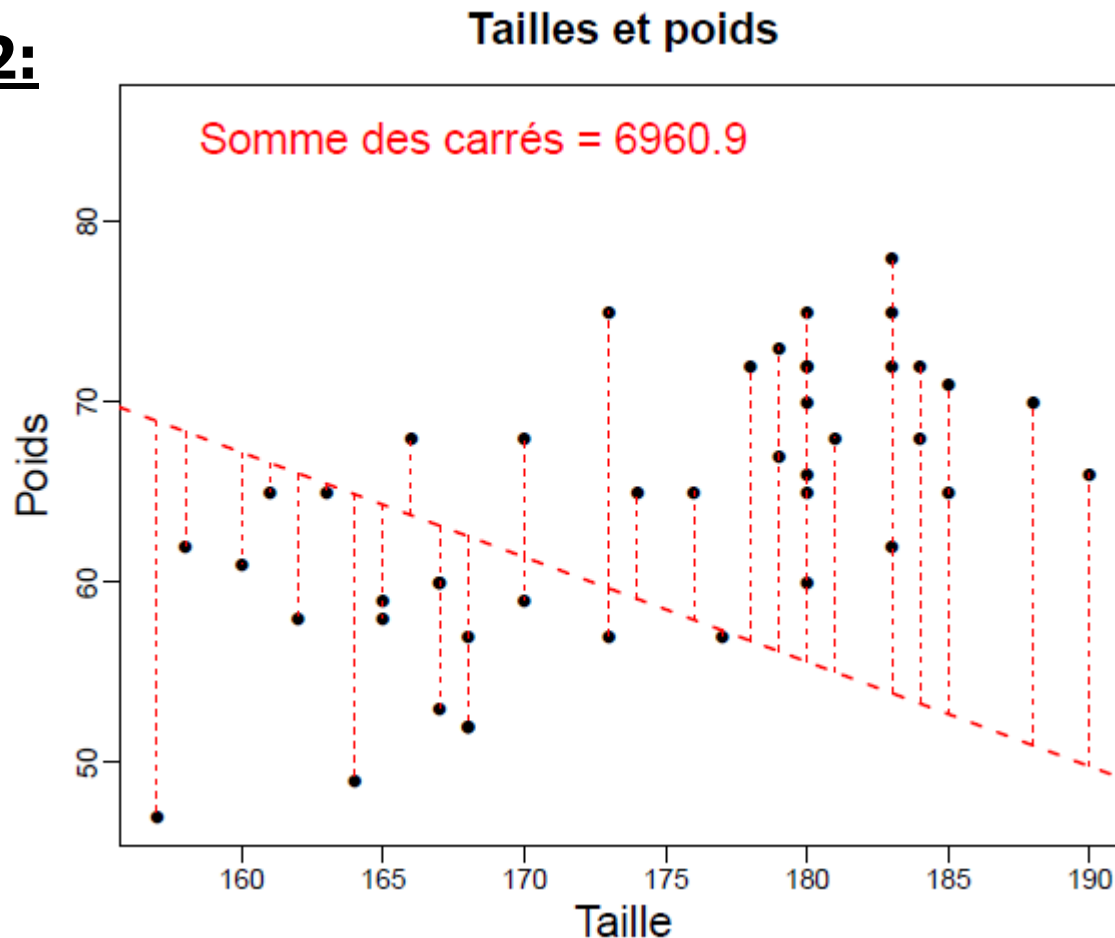
3. Régression linéaire simple

Estimation des paramètres

- Quelle droite choisir ?

- **Exemple 2:**

$$\sum (\varepsilon_i)^2$$



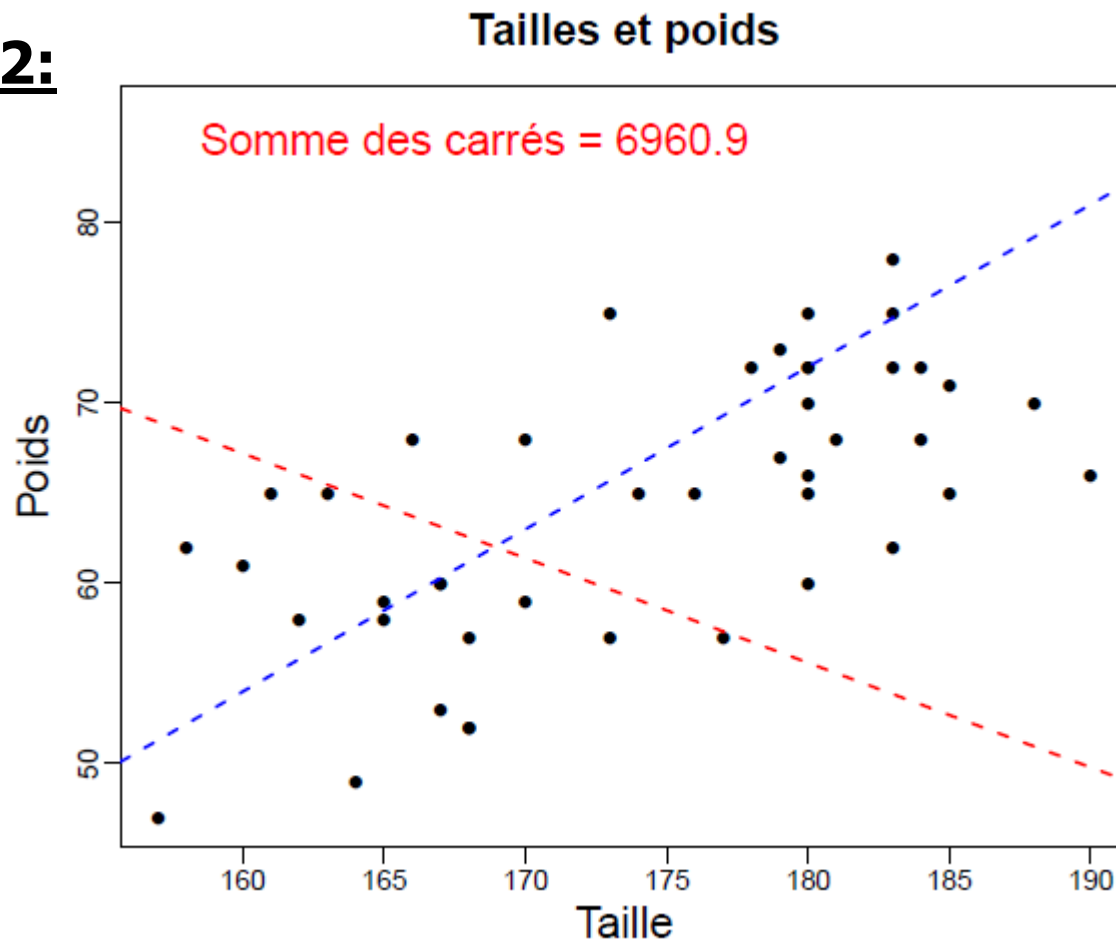
3. Régression linéaire simple

Estimation des paramètres

- Quelle droite choisir ?

- **Exemple 2:**

$$\sum (\varepsilon_i)^2$$



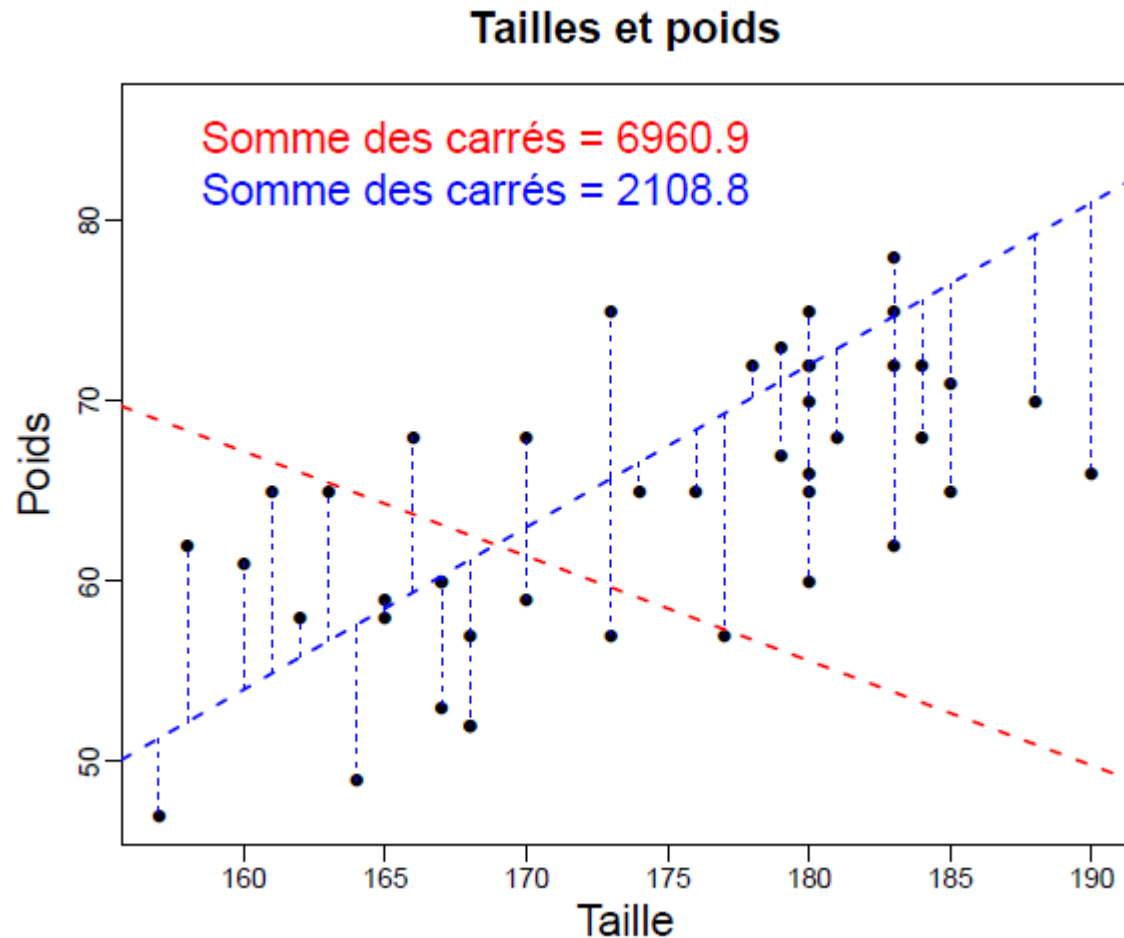
3. Régression linéaire simple

Estimation des paramètres

- Quelle droite choisir ?

- **Exemple 2:**

$$\sum (\varepsilon_i)^2$$



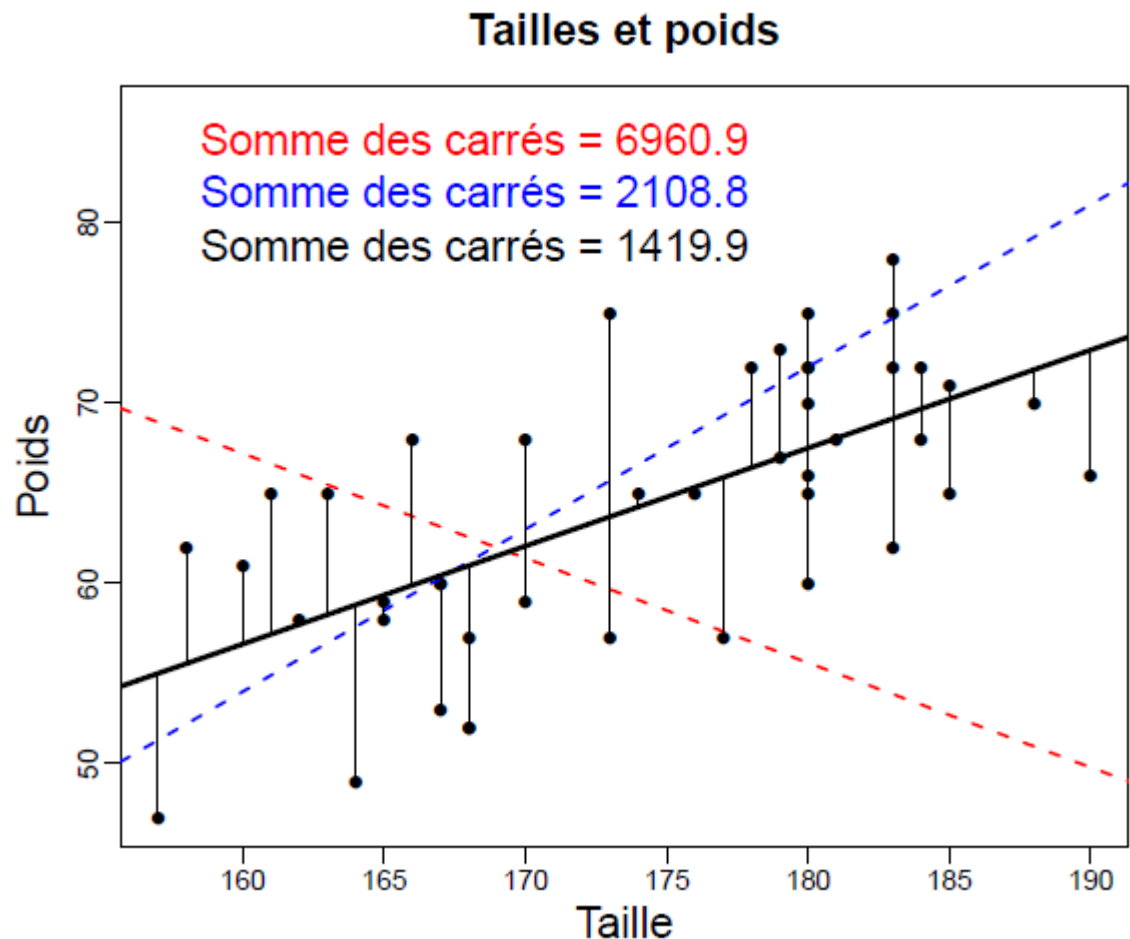
3. Régression linéaire simple

Estimation des paramètres

- Quelle droite choisir ?

- **Exemple 2:**

$$\sum (\varepsilon_i)^2$$



3. Régression linéaire simple

Estimation des paramètres

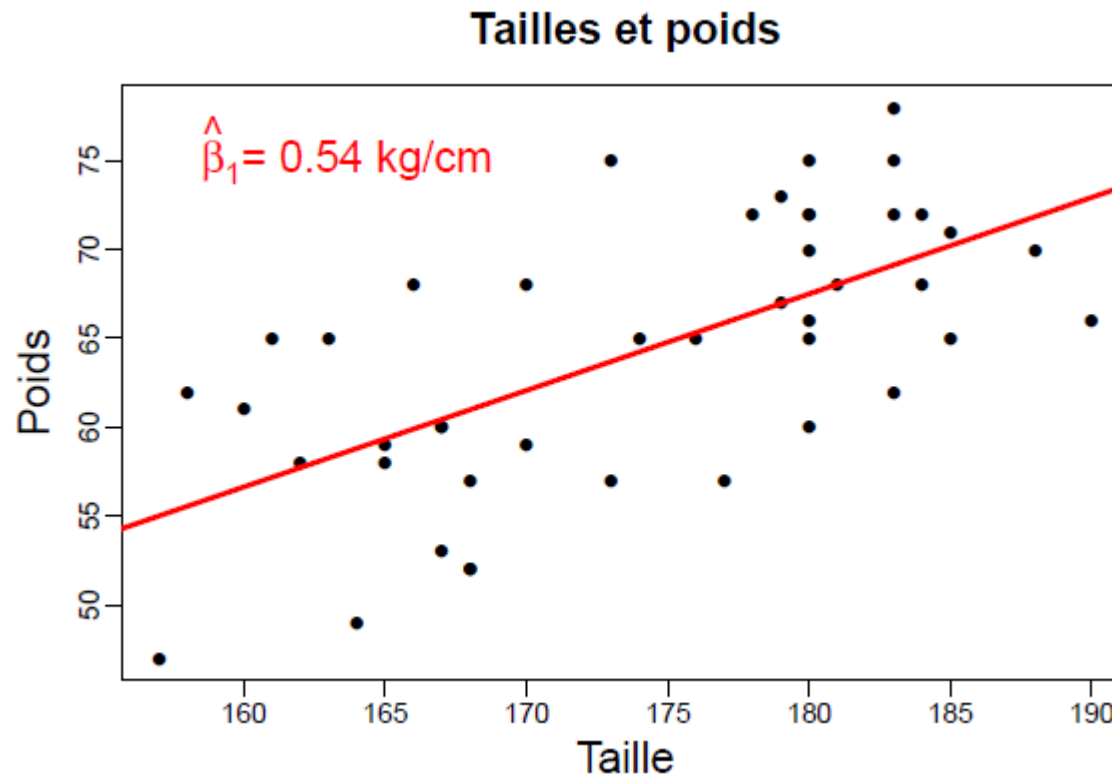
- Choisir la droite qui *minimise* la somme des carrés des *distances* entre la droite et les observations y_i .
- Différentes Critères et techniques permettent de préparer (ou d'entraîner) l'équation de régression linéaire à partir de données, les plus courantes:
 - ✓ *Méthode des Moindres Carrés Ordinaires (MMCO):*
Ordinary Least Squares Method (OLSM)
 - ✓ *Méthode du Maximum de Vraisemblance (MMV):*
Maximum Likelihood Method (MLM)
- On utilise souvent le « $\hat{}$ »(chapeau) pour indiquer qu'une variable est une estimation d'un parametre.

3. Régression linéaire simple

Estimation des paramètres

- Quelle droite choisir ?

Exemple 2: cas des poids et Tailles



3. Régression linéaire simple

Méthode des Moindres Carrés

- On calcule *la distance* entre chaque *point* des données et la *ligne* de régression.
- Cette opération engendre des *résidus* ε_i par rapport à Y , Les valeurs de x restent inchangées
- nous calculons la distance et la somme de toutes les erreurs au carré: $\sum(\varepsilon_i)^2$
- C'est la quantité que les *moindres carrés ordinaires* cherchent à minimiser. **Min** $(\sum(\varepsilon_i)^2)$,

3. Régression linéaire simple

Méthode des Moindres Carrés

- On appelle estimateurs des moindres carrés de (α, β) , les estimateurs $(\hat{\alpha}, \hat{\beta})$ obtenus par la minimisation de la quantité :

$$\delta(\alpha, \beta) = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 = \|Y - \alpha \mathbf{1} - \beta X\|^2$$

- Où $\mathbf{1}$ est le vecteur de R^n dont tous les coefficients valent 1.
- Les estimateurs peuvent également s'écrire sous la forme suivante :

$$(\hat{\alpha}, \hat{\beta}) = \operatorname{argmin} \delta(\alpha, \beta)$$

3. Régression linéaire simple

Méthode des Moindres Carrés

- Le Calcul des estimateurs est donné par:

$$\hat{\alpha} = \bar{y} - \beta \bar{x}$$

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

$$\hat{\beta} = \text{cov}(X, Y) / \text{var}(X)$$

3. Régression linéaire simple

Méthode des Moindres Carrés

- **Exemple** : cas de l'exemple 1

X	Y
2104	400
1600	330
2000	369
1416	232
3000	540

Moyenne

2024

374,2

Cov(X,Y)=

53841,6

$\hat{\beta} = 0,17840254$

var(X)=

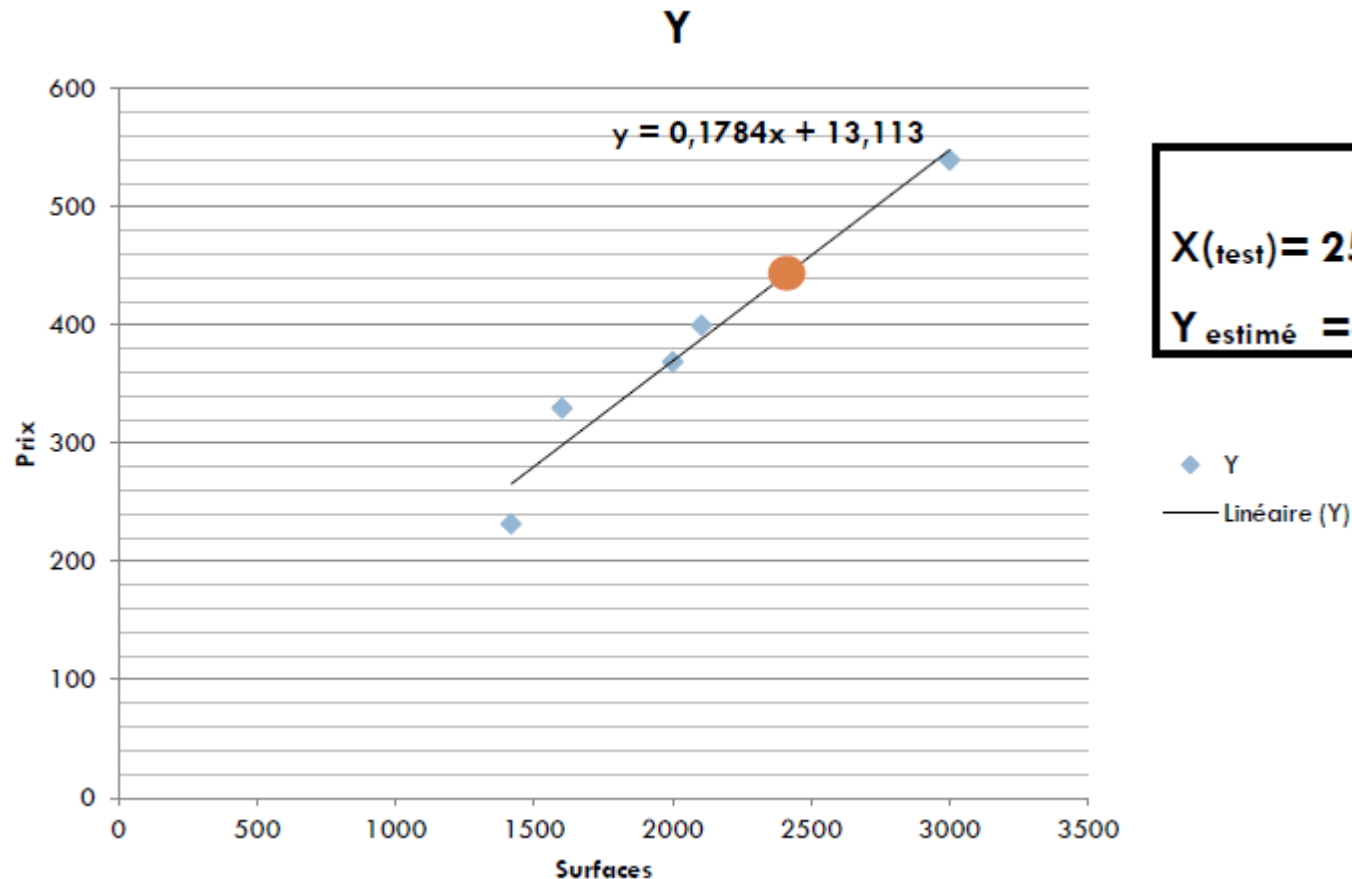
301798,4

$\hat{\alpha} = 13,1132666$

3. Régression linéaire simple

Méthode des Moindres Carrés

- **Exemple** : cas de l'exemple 1



$X_{\text{test}} = 2500$

$Y_{\text{estimé}} = 459,12$

3. Régression linéaire simple

Notation matricielle

- Supposons que les données suivants le modèle :

$$Y = f(X) = \alpha + \beta X$$

- Nous pouvons écrire:

$$y_i = \alpha + \beta x_i + \xi_i, \quad i = 1, \dots, n$$

- Exemple:

Données

i	Y	X	ε
1	1,4	0	0,4
2	6.8	3	0,2
.	.	.	
.	.	.	

Modèle

$$Y = 1 + 2 * X$$



$$\begin{aligned} Y_1 &= 1 + 2 * 0 + 0.4 \\ Y_2 &= 1 + 2 * 3 - 0.2 \end{aligned}$$

3. Régression linéaire simple

Notation matricielle

- Supposons que les données suivants le modèle :

$$Y = f(X) = \alpha + \beta X$$

- Nous pouvons écrire:

$$y_i = \alpha + \beta x_i + \xi_i, \quad i = 1, \dots, n$$

$$\begin{aligned}\alpha &= B_0 \\ \beta &= B_1\end{aligned}$$

- La notation matricielle est :

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & \vdots \\ 1 & \vdots \\ 1 & \vdots \\ 1 & x_n \end{pmatrix} * \begin{pmatrix} B_0 \\ B_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

3. Régression linéaire simple

Notation matricielle

$$\begin{matrix} \textcolor{red}{Y} \\ \left(\begin{array}{c} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ \textcolor{red}{y_n} \end{array} \right) \end{matrix} = \begin{matrix} \textcolor{red}{X} \\ \left(\begin{array}{cc} 1 & x_1 \\ 1 & x_2 \\ 1 & \cdot \\ 1 & \cdot \\ 1 & \cdot \\ 1 & \textcolor{red}{x_n} \end{array} \right) \end{matrix} * \begin{matrix} \textcolor{red}{\beta} \\ \left(\begin{array}{c} B_0 \\ B_1 \end{array} \right) \end{matrix} + \begin{matrix} \textcolor{red}{\varepsilon} \\ \left(\begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{array} \right) \end{matrix}$$

- Calcul des estimateurs : (les paramètres du modèle)

$$\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \longrightarrow \hat{\beta} = (X'X)^{-1}X'Y$$

3. Régression linéaire multiple

Principe

- La régression linéaire multiple est une méthode étendant la régression linéaire simple.
- Elle permet d'écrire les variations d'une variable endogène (à expliquer) Y associée aux plusieurs variables exogènes (expliquatives)

3. Régression linéaire multiple

Le modèle

- Un modèle de **régression linéaire multiple** est défini par l'équation :

$$Y_{(n*1)} = X_{(n*p)}\beta_{(p*1)} + \xi_{(n*1)}$$

- Les hypothèses concernant le modèle sont :

$$(H1) : rg(X) = p$$

$$(H2) : E(\xi) = O_n, \quad var(\xi) = \sigma^2 I_n$$

- (H2) $\leftrightarrow$ Les erreurs sont centrés de même variance et non corrélées entre elles.

3. Régression linéaire multiple

Notation matricielle

Exemple : Données (cas 3 variables)

i	$Y(y_{1i})$	$X(x_{1i})$	$Z(x_{2i})$
1	5	1	3
2	8	2	4
.	.	.	.
.	.	.	.

3. Régression linéaire multiple

Notation matricielle

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_{n-1} \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{1,n-1} & x_{2,n-1} \\ 1 & x_{1,n} & x_{2,n} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_{n-1} \\ \varepsilon_n \end{pmatrix}$$

3. Régression linéaire multiple

Notation matricielle

Exemple : Données (cas p variables)

i	Y(y_i)	X(x_{1i})	Z (x_{2i})	...	W (x_{pi})
1	15	1	3	...	8
2	18	2	4	...	10
.	.	.	.	...	.
.	.	.	.		.

3. Régression linéaire multiple

Notation matricielle

Exemple : Données (cas p variables)

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} * \begin{pmatrix} B_0 \\ B_1 \\ B_2 \\ \vdots \\ B_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

3. Régression linéaire multiple

Notation matricielle

- L'écriture sous forme d'un système d'équations :

$$\left\{ \begin{array}{l} Y_1 = B_0 + B_1 x_{11} + B_2 x_{12} + \dots + B_p x_{1p} + \varepsilon_1 \\ Y_2 = B_0 + B_1 x_{21} + B_2 x_{22} + \dots + B_p x_{2p} + \varepsilon_2 \\ \vdots \\ Y_n = B_0 + B_1 x_{n1} + B_2 x_{n2} + \dots + B_p x_{np} + \varepsilon_n \end{array} \right.$$

3. Régression linéaire multiple

Estimateurs des MCO

- L'estimateur MCO est défini comme suit :

$$\hat{\beta} = \underset{\beta \in R^p}{\operatorname{argmin}} \sum_{i=1}^n (y_i - \sum \beta_j x_i^j)^2,$$

$$= \underset{\beta \in R^p}{\operatorname{argmin}} \|Y - X\beta\|^2,$$

- L'estimateur des MCO est :

$$\mathbf{B} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ \vdots \\ B_p \end{bmatrix} \quad \longrightarrow \quad \hat{\beta} = (X'X)^{-1}X'Y$$

3. Régression linéaire multiple

Propriétés de la droite de régression

- Le modèle de régression linéaire simple ou multiple estimé par la méthode des moindres carrés possède les propriétés suivantes :
 - ✓ La somme des résidus est nulle.
 - ✓ La somme des carrés des résidus est minimale.
 - ✓ La droite de régression passe toujours par le point $(\bar{X}, \bar{Y})$