

# Régression linéaire

David Causeur

*Laboratoire de Mathématiques Appliquées*

*Agrocampus Rennes*

*IRMAR CNRS UMR 6625*

*<http://www.agrocampus-rennes.fr/math/causeur/>*

# Plan de la présentation

## 1 Introduction

## 2 Approche exploratoire

- Graphiques
- Coefficient de corrélation linéaire

## 3 Approche inférentielle

- Le modèle de régression linéaire simple
- Test du lien entre  $x$  et  $Y$
- Estimation des paramètres
- Prédiction

## 4 Ouverture

# Problématique

## Etude du lien entre 2 variables quantitatives

- L'appréciation globale dépend-t'elle de l'amertume perçue ?

# Problématique

## Etude du lien entre 2 variables quantitatives

- L'appréciation globale dépend-t'elle de l'amertume perçue ?
- Peut-on prévoir le chiffre de ventes du mois prochain ?

# Problématique

## Etude du lien entre 2 variables quantitatives

- L'appréciation globale dépend-t'elle de l'amertume perçue ?
- Peut-on prévoir le chiffre de ventes du mois prochain ?
- La note au bac est-elle liée à celle du 3ème trimestre ?
- ...

# Un problème de prédiction

|   | $x = \text{Épaisseur de gras (mm)}$ | $y = \text{Teneur en viande maigre (kg/quintal)}$ |
|---|-------------------------------------|---------------------------------------------------|
| 1 | 12                                  | 56.52                                             |
| 2 | 13                                  | 57.58                                             |
| 3 | 11                                  | 55.89                                             |
| 4 | 10                                  | 61.82                                             |
| 5 | 9                                   | 62.96                                             |
| 6 | 10                                  | 54.58                                             |
| 7 | 7                                   | 67.09                                             |
| 8 | 15                                  | 55.00                                             |
| 9 | 10                                  | 60.49                                             |
| ⋮ | ⋮                                   | ⋮                                                 |

# Plan de la présentation

## 1 Introduction

## 2 Approche exploratoire

- Graphiques
- Coefficient de corrélation linéaire

## 3 Approche inférentielle

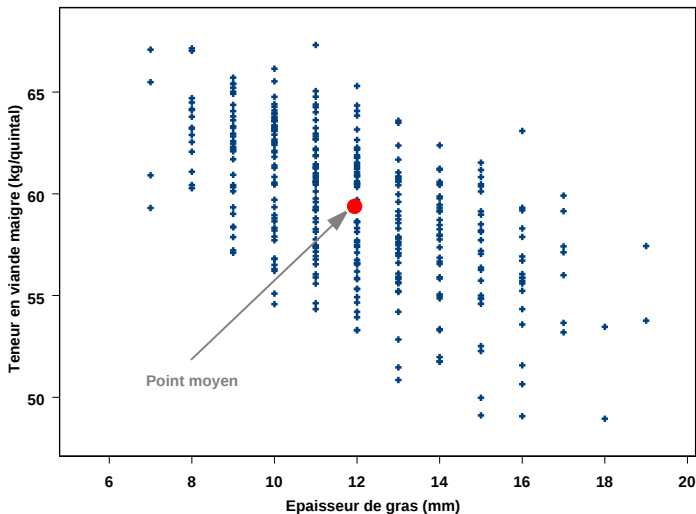
- Le modèle de régression linéaire simple
- Test du lien entre  $x$  et  $Y$
- Estimation des paramètres
- Prédiction

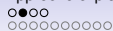
## 4 Ouverture

●○○○  
○○○○○○○○○○

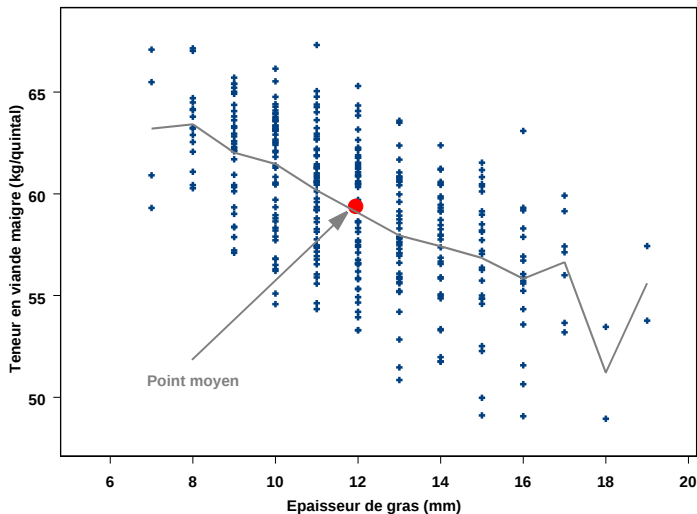
○○  
○○○○○○○○○○  
○○○○  
○○

# Nuage de points





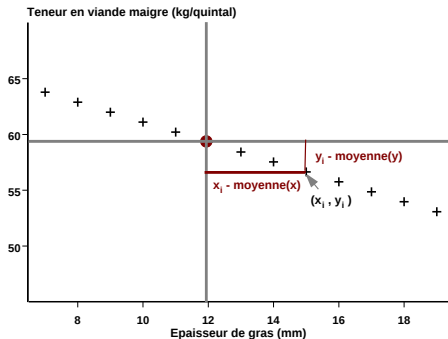
## Forme de la relation



# Lien linéaire parfait

Les variations de  $y$  sont proportionnelles à celles de  $x$

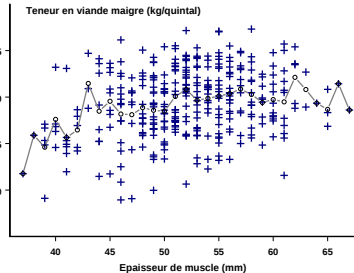
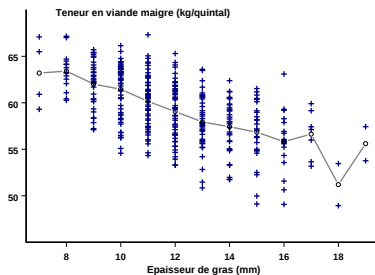
$$(y_i - \bar{y}) = k(x_i - \bar{x})$$



○ ○ ○ ●  
○ ○ ○ ○ ○ ○ ○ ○ ○ ○

○ ○  
○ ○ ○ ○ ○ ○ ○ ○ ○ ○  
○ ○ ○ ○  
○ ○ ○ ○  
○ ○

# Intensité du lien linéaire



**Intuition :**

le lien *Teneur en viande maigre* × *Ép. de gras*  
est plus fort que  
le lien *Teneur en viande maigre* × *Ép. de muscle*

# Mesure de la corrélation linéaire

**Objectif** : mesurer l'intensité du lien linéaire

# Mesure de la corrélation linéaire

**Objectif** : mesurer l'intensité du lien linéaire

**Propriété** : mesure indépendante des unités de mesure

# Mesure de la corrélation linéaire

**Objectif** : mesurer l'intensité du lien linéaire

**Propriété** : mesure indépendante des unités de mesure

**Idée ...** : mesurer les variations en nombres d'écarts-types

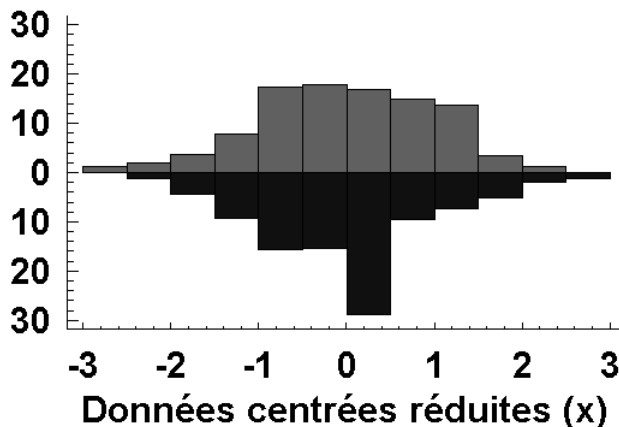
| Variable | Valeur observée | Écart à la moyenne | Écart à la moyenne<br>(en nb d'écarts-types) |
|----------|-----------------|--------------------|----------------------------------------------|
| $x$      | $x_i$           | $x_i - \bar{x}$    | $\frac{x_i - \bar{x}}{S_x}$                  |
| $y$      | $y_i$           | $y_i - \bar{y}$    | $\frac{y_i - \bar{y}}{S_y}$                  |

# Données centrées-réduites

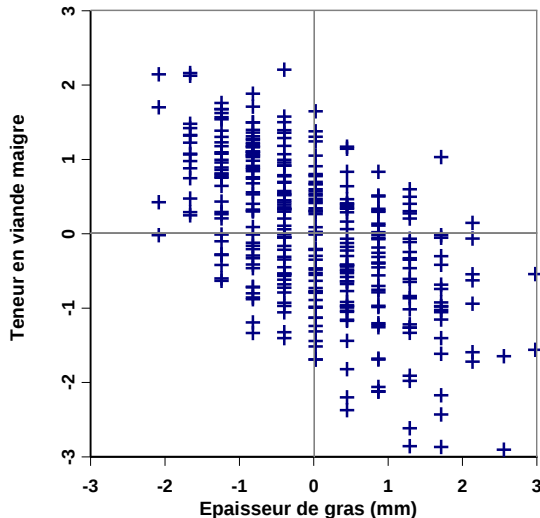
|   | x = Épaisseur de gras |                 | y = Teneur en viande maigre |                 |
|---|-----------------------|-----------------|-----------------------------|-----------------|
|   | Donnée                | Donnée          | Donnée                      | Donnée          |
|   | (mm)                  | centrée-réduite | (kg/quintal)                | centrée-réduite |
| 1 | 12                    | 0.03            | 56.52                       | -0.79           |
| 2 | 13                    | 0.45            | 57.58                       | -0.50           |
| 3 | 11                    | -0.40           | 55.89                       | -0.97           |
| 4 | 10                    | -0.82           | 61.82                       | 0.68            |
| 5 | 9                     | -1.24           | 62.96                       | 1.00            |
| 6 | 10                    | -0.82           | 54.58                       | -1.33           |
| 7 | 7                     | -2.08           | 67.09                       | 2.14            |
| ⋮ | ⋮                     | ⋮               | ⋮                           | ⋮               |

# Répartition des données centrées réduites

## Données centrées réduites (y)



# Nuage des données centrées réduites



# Coefficient de corrélation linéaire $r(x, y)$

$$\begin{aligned}
 r(x, y) &= \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right), \\
 &= \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}}{s_x s_y} = \frac{s_{xy}}{s_x s_y}
 \end{aligned}$$

# Coefficient de corrélation linéaire $r(x, y)$

$$\begin{aligned}
 r(x, y) &= \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{S_x} \right) \left( \frac{y_i - \bar{y}}{S_y} \right), \\
 &= \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}}{S_x S_y} = \frac{S_{xy}}{S_x S_y}
 \end{aligned}$$

## Propriétés

- $-1 \leq r(x, y) \leq 1$

# Coefficient de corrélation linéaire $r(x, y)$

$$\begin{aligned} r(x, y) &= \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{S_x} \right) \left( \frac{y_i - \bar{y}}{S_y} \right), \\ &= \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}}{S_x S_y} = \frac{S_{xy}}{S_x S_y} \end{aligned}$$

## Propriétés

- $-1 \leq r(x, y) \leq 1$
- $r(x, y) > 0 \Rightarrow$  relation linéaire croissante

# Coefficient de corrélation linéaire $r(x, y)$

$$\begin{aligned}
 r(x, y) &= \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{S_x} \right) \left( \frac{y_i - \bar{y}}{S_y} \right), \\
 &= \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}}{S_x S_y} = \frac{S_{xy}}{S_x S_y}
 \end{aligned}$$

## Propriétés

- $-1 \leq r(x, y) \leq 1$
- $r(x, y) > 0 \Rightarrow$  relation linéaire croissante
- $r(x, y) < 0 \Rightarrow$  relation linéaire décroissante

# Coefficient de corrélation linéaire $r(x, y)$

$$\begin{aligned} r(x, y) &= \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{S_x} \right) \left( \frac{y_i - \bar{y}}{S_y} \right), \\ &= \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}}{S_x S_y} = \frac{S_{xy}}{S_x S_y} \end{aligned}$$

## Propriétés

- $-1 \leq r(x, y) \leq 1$
- $r(x, y) > 0 \Rightarrow$  relation linéaire croissante
- $r(x, y) < 0 \Rightarrow$  relation linéaire décroissante
- $r(x, y) = 0 \Rightarrow$  pas de relation linéaire

# Coefficient de corrélation linéaire $r(x, y)$

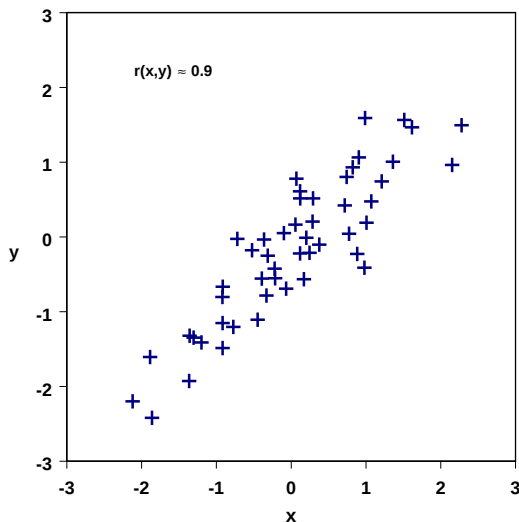
$$\begin{aligned} r(x, y) &= \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{S_x} \right) \left( \frac{y_i - \bar{y}}{S_y} \right), \\ &= \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}}{S_x S_y} = \frac{S_{xy}}{S_x S_y} \end{aligned}$$

## Propriétés

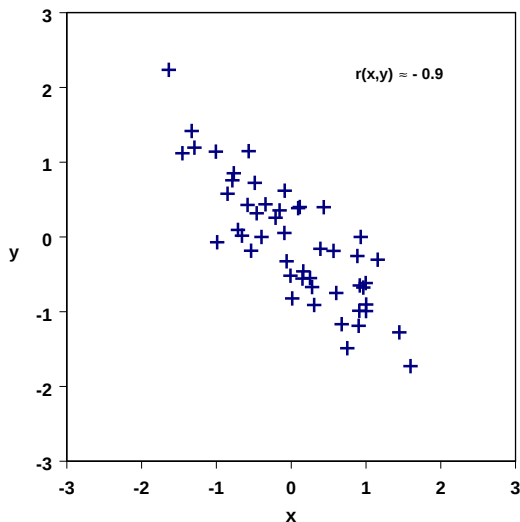
- $-1 \leq r(x, y) \leq 1$
- $r(x, y) > 0 \Rightarrow$  relation linéaire croissante
- $r(x, y) < 0 \Rightarrow$  relation linéaire décroissante
- $r(x, y) = 0 \Rightarrow$  pas de relation linéaire
- $r(x, y) = \pm 1 \Rightarrow y = \beta_0 + \beta_1 x$



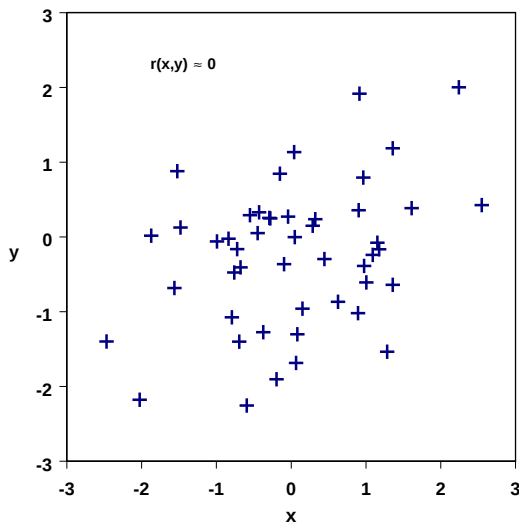
# Relation linéaire croissante



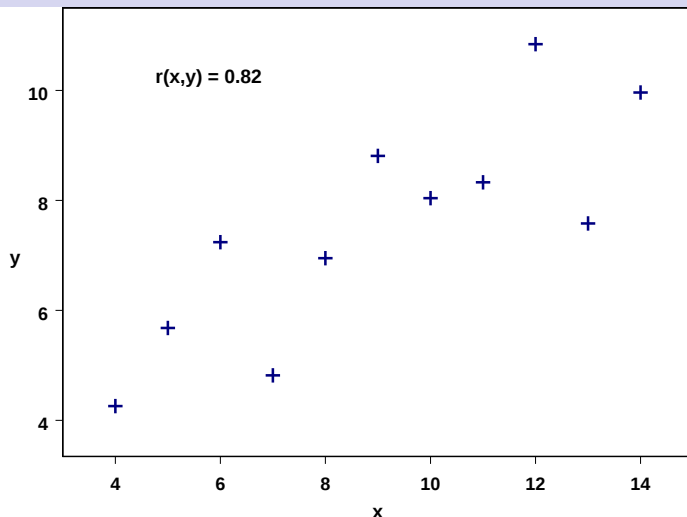
# Relation linéaire décroissante



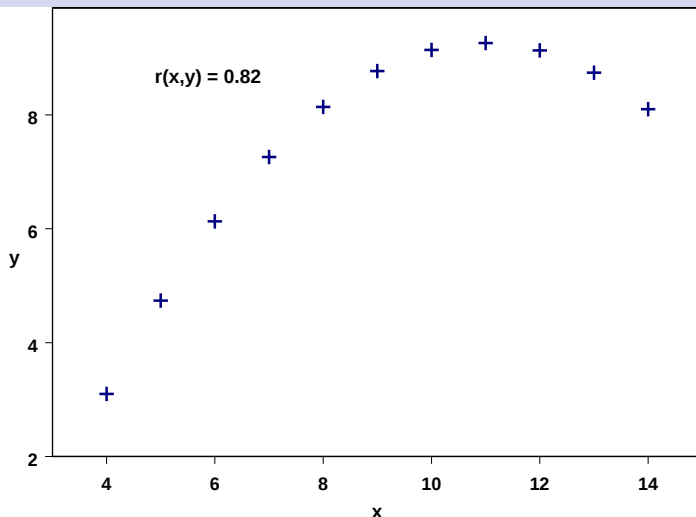
# Pas de relation linéaire

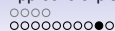


# Attention à l'interprétation !

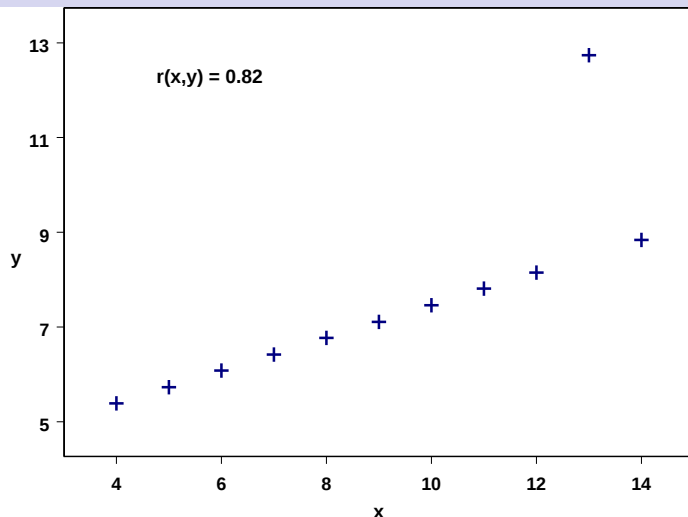


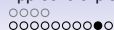
# Attention à l'interprétation !



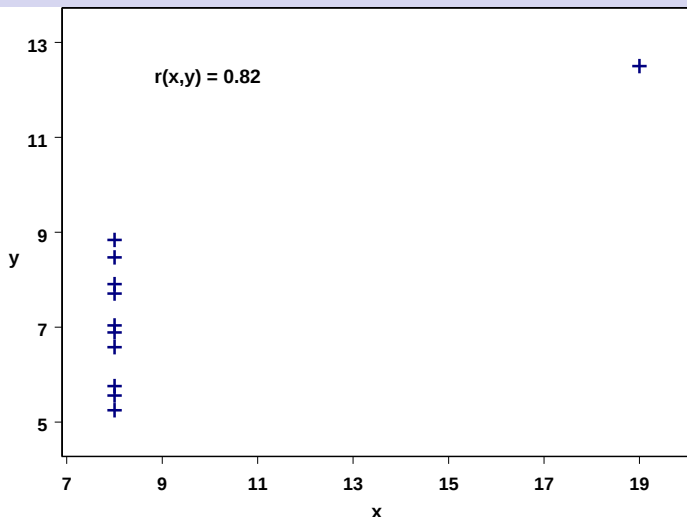


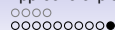
# Attention à l'interprétation !





# Attention à l'interprétation !





# Matrice de corrélation

|                  | Maths  |         |         |      | Philosophie |         |         |      |
|------------------|--------|---------|---------|------|-------------|---------|---------|------|
|                  | 1er t. | 2ème t. | 3ème t. | Bac  | 1er t.      | 2ème t. | 3ème t. | Bac  |
| M <sub>1</sub>   | 1.00   | 0.77    | 0.72    | 0.58 | 0.19        | 0.22    | 0.16    | 0.24 |
| M <sub>2</sub>   | 0.77   | 1.00    | 0.81    | 0.66 | 0.20        | 0.25    | 0.22    | 0.30 |
| M <sub>3</sub>   | 0.72   | 0.81    | 1.00    | 0.68 | 0.22        | 0.26    | 0.24    | 0.30 |
| M <sub>bac</sub> | 0.58   | 0.66    | 0.68    | 1.00 | 0.17        | 0.23    | 0.21    | 0.28 |
| P <sub>1</sub>   | 0.19   | 0.20    | 0.22    | 0.17 | 1.00        | 0.64    | 0.61    | 0.42 |
| P <sub>2</sub>   | 0.22   | 0.25    | 0.26    | 0.23 | 0.64        | 1.00    | 0.69    | 0.43 |
| P <sub>3</sub>   | 0.16   | 0.22    | 0.24    | 0.21 | 0.61        | 0.69    | 1.00    | 0.47 |
| P <sub>bac</sub> | 0.24   | 0.30    | 0.30    | 0.28 | 0.42        | 0.43    | 0.47    | 1.00 |

# Plan de la présentation

## 1 Introduction

## 2 Approche exploratoire

- Graphiques
- Coefficient de corrélation linéaire

## 3 Approche inférentielle

- Le modèle de régression linéaire simple
- Test du lien entre  $x$  et  $Y$
- Estimation des paramètres
- Prédiction

## 4 Ouverture

# Approche inférentielle

## Objectifs

- Tester l'existence d'une relation entre  $x$  et  $y$

# Approche inférentielle

## Objectifs

- Tester l'existence d'une relation entre  $x$  et  $y$
- Construire une équation de prédiction de  $y$  par  $x$

# Approche inférentielle

## Objectifs

- Tester l'existence d'une relation entre  $x$  et  $y$
- Construire une équation de prédiction de  $y$  par  $x$
- Calculer un intervalle de confiance (d'estimation, de prédiction, ...)

# Le modèle de régression linéaire simple

$Y$  : variable à expliquer (à prédire, endogène, réponse, dépendante) quantitative

$x$  : variable explicative (prédictrice, exogène, covariable, indépendante) quantitative

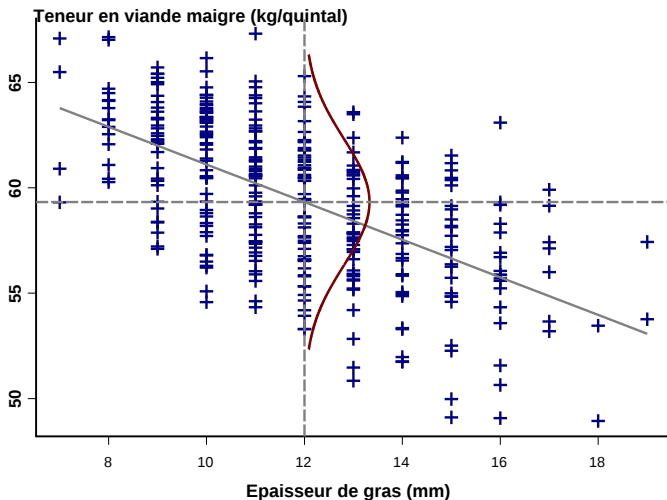
# Le modèle de régression linéaire simple

$Y$  : variable à expliquer (à prédire, endogène, réponse, dépendante) quantitative

$x$  : variable explicative (prédictrice, exogène, covariable, indépendante) quantitative

Pour un  $x$  fixé,  $Y$  est une variable aléatoire

# Le modèle de régression linéaire simple



# Le modèle de régression linéaire simple

$Y$  : variable à expliquer (à prédire, endogène, réponse, dépendante) quantitative

$x$  : variable explicative (prédictrice, exogène, covariable, indépendante) quantitative

Pour un  $x$  fixé,  $Y$  est une variable aléatoire

$$Y = \underbrace{\beta_0 + \beta_1 x}_{\text{Variations imputables à } x} + \underbrace{\varepsilon}_{\text{Variations non-imputables à } x}$$

$$\varepsilon \sim \mathcal{N}(0; \sigma)$$

# Paramètres du modèle

- Coefficients du modèle :
  - $\beta_0$  ordonnée à l'origine
  - $\beta_1$  pente

# Paramètres du modèle

- Coefficients du modèle :
  - $\beta_0$  ordonnée à l'origine
  - $\beta_1$  pente
- Écart-type résiduel :  $\sigma$

# Existence d'un lien entre $x$ et $Y$

$H_0$  : il n'y a pas de lien entre  $x$  et  $Y$

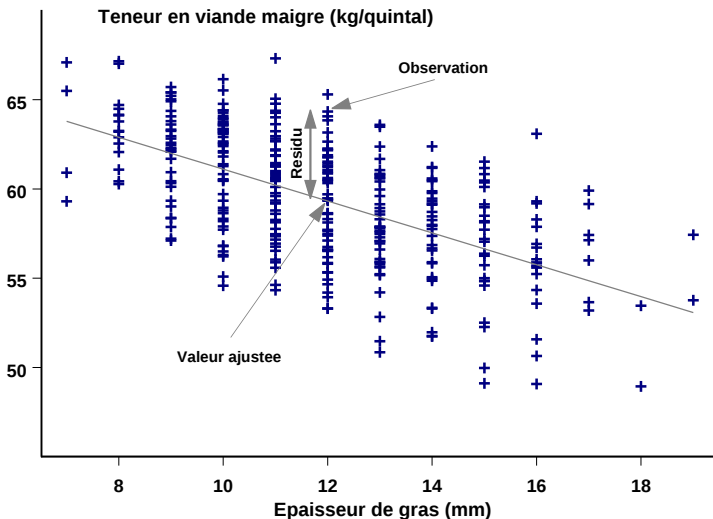
$H_1$  : il y a un lien entre  $x$  et  $Y$

ou encore ...

$H_0$  :  $\beta_1 = 0$  ,

$H_1$  :  $\beta_1 \neq 0$

# Décomposition de la variabilité de $Y$



# Décomposition de la variabilité de $Y$

## Equation d'«analyse de la variance»

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{\text{SCET}} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{\text{SCEM}} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{\text{SCER}}$$

# Décomposition de la variabilité de $Y$

## Equation d'«analyse de la variance»

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{\text{SCET}} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{\text{SCEM}} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{\text{SCER}}$$

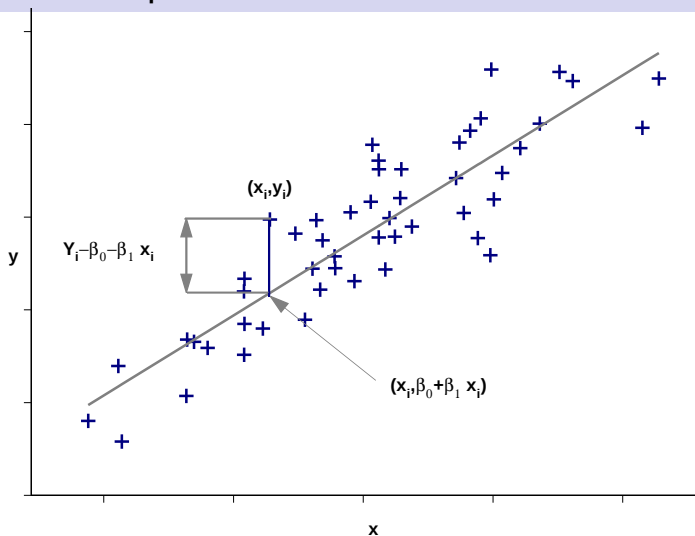
**Problème** : ajustement du modèle ( $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ )

# Ajustement par la méthode des moindres carrés

**Une mesure de l'adéquation entre modèle et données : le critère des moindres carrés  $SC(\beta_0, \beta_1)$**

$$SC(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$$

# Ajustement par la méthode des moindres carrés



# Ajustement par la méthode des moindres carrés

**Une mesure de l'adéquation entre modèle et données : le critère des moindres carrés  $SC(\beta_0, \beta_1)$**

$$SC(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$$

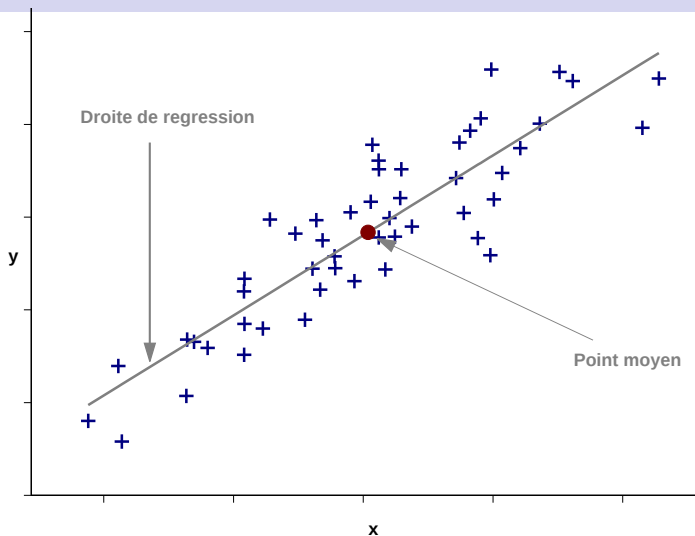
Les estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$  minimisent  $SC(\beta_0, \beta_1)$

$x \mapsto \hat{\beta}_0 + \hat{\beta}_1 x$  droite de régression

# Minimisation du critère des moindres carrés

$$\begin{aligned}\frac{\partial SC}{\partial \beta_0}(\hat{\beta}_0, \hat{\beta}_1) &= 0 \\ -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\ \bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} &= 0 \\ \bar{Y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x}\end{aligned}$$

# Minimisation du critère des moindres carrés



# Minimisation du critère des moindres carrés

$$\begin{aligned}\frac{\partial SC}{\partial \beta_0}(\hat{\beta}_0, \hat{\beta}_1) &= 0 \\ -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\ \bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} &= 0 \\ \bar{Y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x}\end{aligned}$$

Le point moyen  $(\bar{x}, \bar{Y})$  est sur la droite de régression

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$$

## Estimation de la pente

$$SC(\hat{\beta}_0, \beta_1) = \sum_{i=1}^n \left( [Y_i - \bar{Y}] - \hat{\beta}_1 [x_i - \bar{x}] \right)^2$$

D'où :

# Estimation de la pente

$$SC(\hat{\beta}_0, \beta_1) = \sum_{i=1}^n \left( [Y_i - \bar{Y}] - \hat{\beta}_1 [x_i - \bar{x}] \right)^2$$

D'où :

$$\frac{\partial SC}{\partial \beta_1}(\hat{\beta}_0, \hat{\beta}_1) = 0$$

# Estimation de la pente

$$SC(\hat{\beta}_0, \beta_1) = \sum_{i=1}^n \left( [Y_i - \bar{Y}] - \hat{\beta}_1 [x_i - \bar{x}] \right)^2$$

D'où :

$$\frac{\partial SC}{\partial \beta_1}(\hat{\beta}_0, \hat{\beta}_1) = 0$$

$$-2 \sum_{i=1}^n [x_i - \bar{x}] \left( [Y_i - \bar{Y}] - \hat{\beta}_1 [x_i - \bar{x}] \right) = 0$$

# Estimation de la pente

$$SC(\hat{\beta}_0, \beta_1) = \sum_{i=1}^n \left( [Y_i - \bar{Y}] - \hat{\beta}_1 [x_i - \bar{x}] \right)^2$$

D'où :

$$\frac{\partial SC}{\partial \beta_1}(\hat{\beta}_0, \hat{\beta}_1) = 0$$

$$-2 \sum_{i=1}^n [x_i - \bar{x}] \left( [Y_i - \bar{Y}] - \hat{\beta}_1 [x_i - \bar{x}] \right) = 0$$

$$S_{xy} - \hat{\beta}_1 S_x^2 = 0$$

## Estimation de la pente

$$SC(\hat{\beta}_0, \beta_1) = \sum_{i=1}^n \left( [Y_i - \bar{Y}] - \hat{\beta}_1 [x_i - \bar{x}] \right)^2$$

D'où :

$$\frac{\partial SC}{\partial \beta_1}(\hat{\beta}_0, \hat{\beta}_1) = 0$$

$$-2 \sum_{i=1}^n [x_i - \bar{x}] \left( [Y_i - \bar{Y}] - \hat{\beta}_1 [x_i - \bar{x}] \right) = 0$$

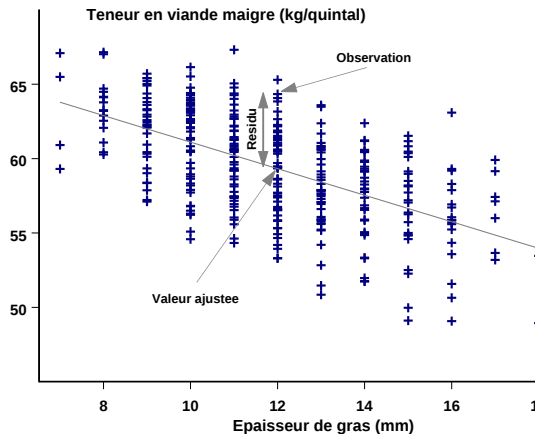
$$S_{xy} - \hat{\beta}_1 S_x^2 = 0$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_x^2}$$

# Valeurs ajustées et résidus d'ajustement

## ● Valeurs ajustées

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$



# Valeurs ajustées et résidus d'ajustement

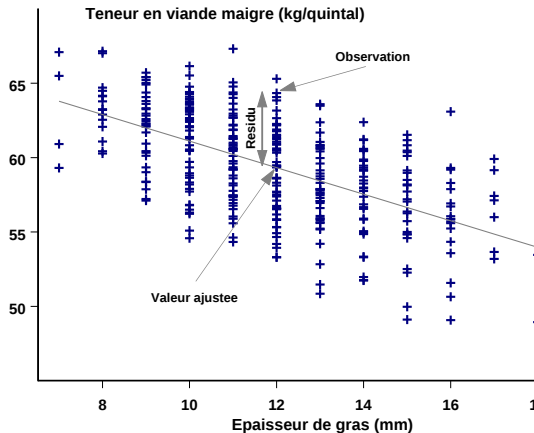
## ● Valeurs ajustées

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

## ● Résidus

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$$

$$\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i = 0$$



# Valeurs ajustées et résidus d'ajustement

## ● Valeurs ajustées

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

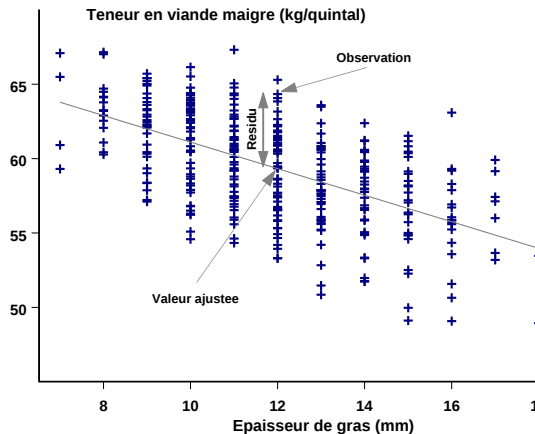
## ● Résidus

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$$

$$\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i = 0$$

## ● Décomposition

$$Y_i = \hat{Y}_i + \hat{\varepsilon}_i$$



# Part de variabilité de $Y$ imputable à $x$

## Equation d'«analyse de la variance»

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{\text{SCET}} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{\text{SCEM}} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{\text{SCER}}$$

## Part de variabilité de $Y$ imputable à $x$

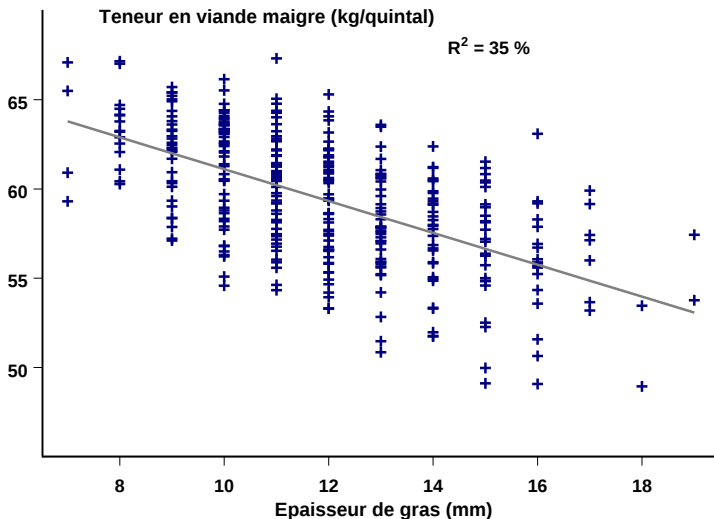
### Equation d'«analyse de la variance»

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{\text{SCET}} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{\text{SCEM}} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{\text{SCER}}$$

### Part de variabilité expliquée par le modèle

$$\begin{aligned} R^2 &= \frac{\text{SCEM}}{\text{SCET}} = 1 - \frac{\text{SCER}}{\text{SCET}} \\ &= [r(x, y)]^2 \end{aligned}$$

# Part de variabilité de $Y$ imputable à $x$



# Test du lien entre $x$ et $Y$

## Equation d'«analyse de la variance»

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{\text{SCET}} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{\text{SCEM}} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{\text{SCER}}$$

# Test du lien entre $x$ et $Y$

## Equation d'«analyse de la variance»

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{\text{SCET}} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{\text{SCEM}} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{\text{SCER}}$$

## Sous l'hypothèse d'absence de lien

$$F = \frac{\frac{\text{SCEM}}{2-1}}{\frac{\text{SCER}}{n-2}} = \frac{\text{CMF}}{\text{CMR}} \sim \mathcal{F}_{1,n-2}$$

# Illustration

## Table d'analyse de la variance

|                   | Degrés de liberté | SC      | CM      | F      | Probabilité critique |
|-------------------|-------------------|---------|---------|--------|----------------------|
| Épaisseur de gras | 1                 | 1535.53 | 1535.53 | 181.04 | 0.00                 |
| Résiduelle        | 342               | 2900.69 | 8.48    |        |                      |

# Estimation des coefficients

- $\hat{\beta}_0, \hat{\beta}_1$  sont sans biais

# Estimation des coefficients

- $\hat{\beta}_0, \hat{\beta}_1$  sont sans biais
- **Variance d'estimation :**

$$\text{Var}(\hat{\beta}_1) = \frac{1}{n} \frac{\sigma^2}{S_x^2}$$

# Estimation des coefficients

- $\hat{\beta}_0, \hat{\beta}_1$  sont sans biais
- **Variance d'estimation :**

$$\text{Var}(\hat{\beta}_1) = \frac{1}{n} \frac{\sigma^2}{S_x^2}$$

- **Estimation de la variance d'estimation :**

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{1}{n} \frac{\hat{\sigma}^2}{S_x^2}$$

# Estimation de $\sigma$

$\sigma^2$ , **variance résiduelle**

$$\hat{\sigma}^2 = \frac{1}{n-2} \underbrace{\sum_{i=1}^n [Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i]^2}_{\text{SCER}}$$

**Nombre de degrés de liberté de SCER**

$$\text{ddl}_r = n - 2$$

# Exemple

|              | Coefficients | Écart-type | Statistique t | Probabilité critique |
|--------------|--------------|------------|---------------|----------------------|
| Constante    | 70,03        | 0,81       | 86,75         | 6,02E-235            |
| Variable X 1 | -0,89        | 0,07       | -13,46        | 2,05E-33             |

|              | Limite inférieure<br>(seuil de confiance = 95%) | Limite supérieure |
|--------------|-------------------------------------------------|-------------------|
| Constante    | 68,45                                           | 71,62             |
| Variable X 1 | -1,02                                           | -0,76             |

**Écart-type résiduel : 2.91**

# Tous différents de 0 !

|                  | Maths  |         |         |      | Philosophie |         |         |      |
|------------------|--------|---------|---------|------|-------------|---------|---------|------|
|                  | 1er t. | 2ème t. | 3ème t. | Bac  | 1er t.      | 2ème t. | 3ème t. | Bac  |
| M <sub>1</sub>   | 1.00   | 0.77    | 0.72    | 0.58 | 0.19        | 0.22    | 0.16    | 0.24 |
| M <sub>2</sub>   | 0.77   | 1.00    | 0.81    | 0.66 | 0.20        | 0.25    | 0.22    | 0.30 |
| M <sub>3</sub>   | 0.72   | 0.81    | 1.00    | 0.68 | 0.22        | 0.26    | 0.24    | 0.30 |
| M <sub>bac</sub> | 0.58   | 0.66    | 0.68    | 1.00 | 0.17        | 0.23    | 0.21    | 0.28 |
| P <sub>1</sub>   | 0.19   | 0.20    | 0.22    | 0.17 | 1.00        | 0.64    | 0.61    | 0.42 |
| P <sub>2</sub>   | 0.22   | 0.25    | 0.26    | 0.23 | 0.64        | 1.00    | 0.69    | 0.43 |
| P <sub>3</sub>   | 0.16   | 0.22    | 0.24    | 0.21 | 0.61        | 0.69    | 1.00    | 0.47 |
| P <sub>bac</sub> | 0.24   | 0.30    | 0.30    | 0.28 | 0.42        | 0.43    | 0.47    | 1.00 |

# Problématique

**On connaît  $x_0$  mais pas  $Y_0$**

**Prédicteur**

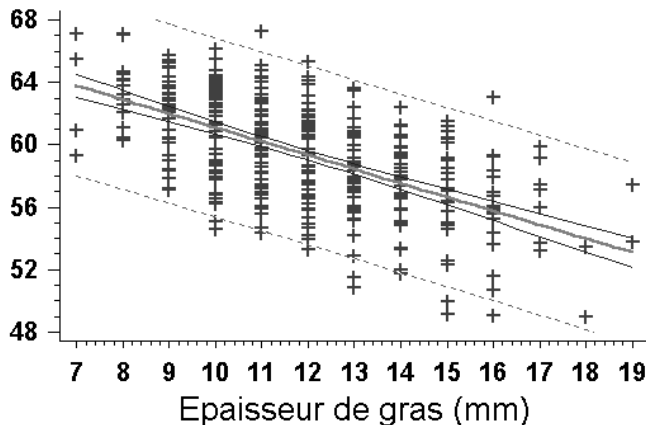
$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

**Erreur de prédiction**

$$\begin{aligned}\hat{\varepsilon}_0 &= Y_0 - \hat{Y}_0 \\ \mathbb{E}(\hat{\varepsilon}_0) &= 0\end{aligned}$$

## Variance de prédiction

Teneur en viande maigre (kg/quintal)



# Plan de la présentation

## 1 Introduction

## 2 Approche exploratoire

- Graphiques
- Coefficient de corrélation linéaire

## 3 Approche inférentielle

- Le modèle de régression linéaire simple
- Test du lien entre  $x$  et  $Y$
- Estimation des paramètres
- Prédiction

## 4 Ouverture

# Ouverture

- Extension à plusieurs variables explicatives (Module suivant)

# Ouverture

- Extension à plusieurs variables explicatives (Module suivant)
- Extensions non-linéaires (Microbiologie, Economie, ...)

# Ouverture

- Extension à plusieurs variables explicatives (Module suivant)
- Extensions non-linéaires (Microbiologie, Economie, ...)
- Prise en compte de variables explicatives quantitatives et qualitatives (Analyse de la covariance)