COURS 11

3 février 2026

PLAN

- Régression linéaire simple
- Régression linéaire multiple
- Exercices sous R

RÉGRESSION LINÉAIRE SIMPLE

- **Modélisation :** quel est l'effet moyen d'une variation d'une variable X sur une variable Y ?

$$\boldsymbol{Y = \alpha + \beta X + \varepsilon}$$

  - Y: **Variable dépendante**/expliquée/prédite/réponse
  - X: **Variable indépendante**/explicative/prédictive/contrôle
  - α : **Constante**/ordonnée à l'origine
  - β : **Coefficient** de la pente dans la relation entre *Y* et *X*
  - ε : **Terme d'erreur**, perturbations qui affectent *Y* et viennent d'autres facteurs que *X* (inobservables)

ESTIMATEUR DES MOINDRES CARRÉS ORDINAIRES

- **Paramètres inconnus** : α et β
- **Estimateur des Moindres Carrés Ordinaires (MCO) :** sous conditions, le meilleur estimateur statistique de α et β qui minimise la somme des carrés des résidus

$$\boldsymbol{\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta} X_i}$$

pour un échantillon de taille *n* issu de la population (*i* = *1*,…, *n*)

avec le **résidu** $\boldsymbol{\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i}$

ESTIMATEUR DES MOINDRES CARRÉS ORDINAIRES

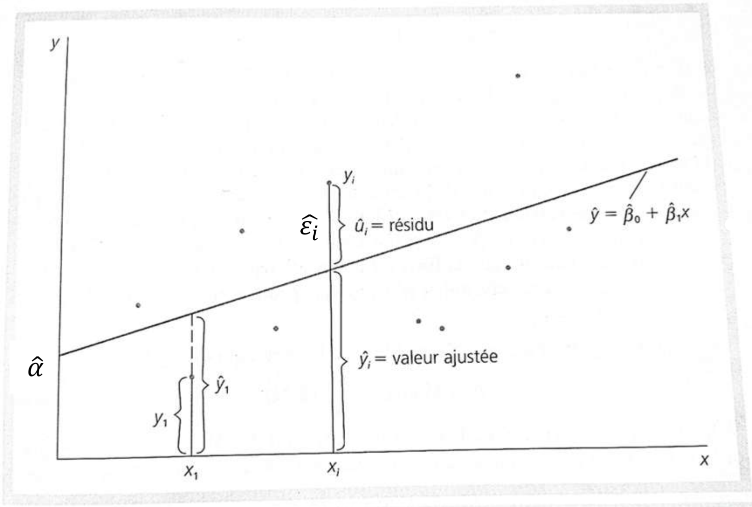


Figure 2.4
Valeurs ajustées et résidus.

© Cengage Learning, 2013

ESTIMATION ET INTERPRÉTATION

- **Estimation** : à partir d'un échantillon donné, calcul des valeurs de $\hat{\alpha}$ et $\hat{\beta}$ (estimateurs des MCO)
- **Interprétation** :
  - **Significativité globale et pouvoir explicatif** de la régression (R^2)
  - **Significativité statistique** de ces coefficients (test d'hypothèse, statistiquement différents de 0 ou non ?)
  - **Valeur moyenne des estimateurs** $\hat{\alpha}$ et $\hat{\beta}$

SIGNIFICATIVITÉ GLOBALE ET POUVOIR EXPLICATIF

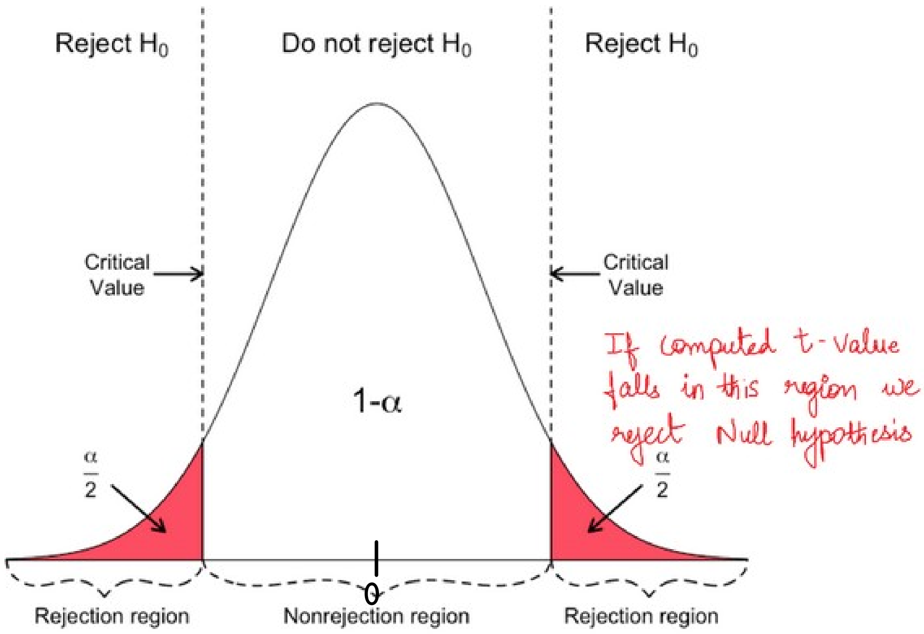
- **Qualité d'ajustement** de la droite de régression MCO aux données
- **Coefficient de détermination (R^2) :** entre 0 et 1 (1 = ajustement parfait)
- **Interprétation :** $\mathrm{R}^2 \times 100$ est le pourcentage de la variation de Y qui est expliquée par X

TEST DE SIGNIFICATIVITÉ DES COEFFICIENTS

- **Test de Student** (t-test) : exemple de test de significativité des coefficients estimés d'une régression
- **Hypothèses** $\begin{cases} \boxed{H_0 : \hat{\beta} = 0} \\ H_1 : \hat{\beta} \neq 0 \end{cases}$ ⟵ **Hypothèse nulle** : absence d'effet de X sur Y
- **Statistique de Student :** $T = \frac{\hat{\beta}}{\sigma_{\hat{\beta}}}$ avec $\sigma_{\hat{\beta}}$ l'écart-type de $\hat{\beta}$ (standard errors)
- **Conclusion du test :** on rejette H0 au seuil de significativité de 5 % si |T| est > 1.96 ou valeur-p < 0,05.
- **Interprétation :** statistiquement (non-)significative au seuil de 5%.

TEST DE SIGNIFICATIVITÉ DES COEFFICIENTS

α est le seuil de rejet de H_0



VALEUR MOYENNE DES ESTIMATEURS

- $\widehat{\boldsymbol{\alpha}}$ **:** valeur estimée de Y lorsque $X = 0$
- $\widehat{\boldsymbol{\beta}}$ **:** variation de $\widehat{Y}_i$ suite à une variation d'une unité de X
- **Attention :** interprétation du coefficient de régression pour un modèle <u>linéaire</u> (niveau-niveau)

INTRODUIRE DE LA NON-LINÉARITÉ

- **Linéaire :** l'effet d'une variation de X sur Y est identique pour toute valeur initiale de X
- **Exemple :** chaque année d'éducation supplémentaire augmente le salaire estimé du même montant
- **Problème :** rendement d'échelle croissant de l'éducation (chaque année d'étude supplémentaire a un effet croissant sur les salaires)
- **Modèle :** log (logarithme naturel) – niveau
- **Autre modèle non-linéaire :** log – log (pour les élasticités)

INTRODUIRE DE LA NON-LINÉARITÉ

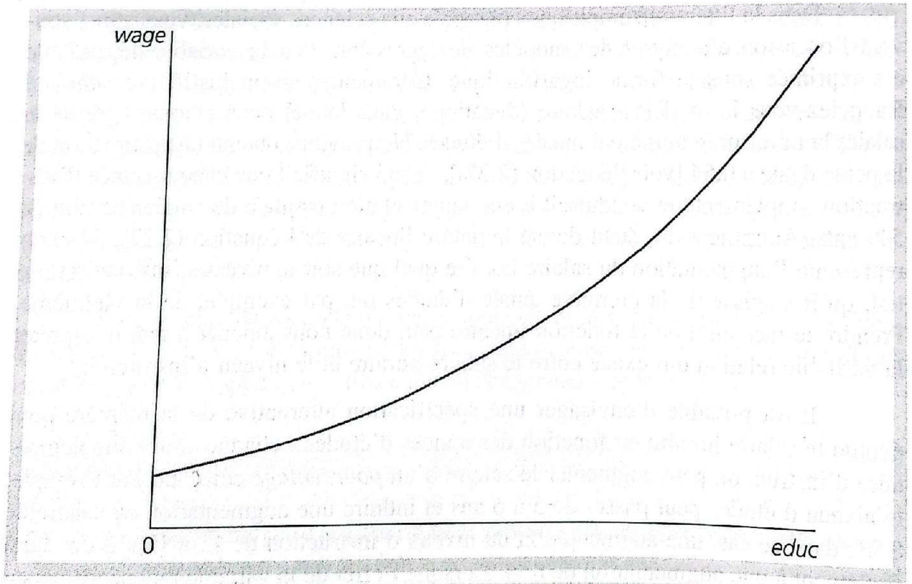


Figure 2.6
wage = exp($\beta_0 + \beta_1$ *educ*), avec $\beta_1 > 0$.

© Cengage Learning, 2013

INTRODUIRE DE LA NON-LINÉARITÉ

Tableau 2.3
Synthèse des formes fonctionnelles ayant recours au log naturel

Modèle	Variable dépendante	Variable indépendante	Interprétation de β_1
Niveau-niveau	y	x	$\Delta y = \beta_1 \Delta x$
Niveau-log	y	$\log(x)$	$\Delta y = (\beta_1 / 100) \%\Delta x$
Log-niveau	$\log(y)$	x	$\%\Delta y = (100\, \beta_1)\, \Delta x$
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

© Cengage Learning, 2013

CS Scanné avec CamScanner

- **Niveau – niveau** : variation de β unité de $\widehat{Y}_i$ suite à une variation d'une unité de X
- **Niveau – log** : variation de $\beta/100$ unités de $\widehat{Y}_i$ suite à une variation d'1% de X
- **Log – niveau** : variation de $100 \times \beta$ % de $\widehat{Y}_i$ suite à une variation d'une unité de X
- **Log – log** : variation de β% de $\widehat{Y}_i$ suite à une variation d'1% de X

RÉGRESSION LINÉAIRE DANS R

- **Régression linéaire** dans R : **lm(variable à expliquer ~ variable(s) explicative(s), ...)**
- **Exemples :**

  model <- lm(nomvar1 ~ nomvar2, data = mesdata)

  model <- lm(mesdata$nomvar1 ~ mesdata$nomvar2)
- **Données manquantes : na.omit** ou **na.fail**
- **Exemples :**

  model <- lm(nomvar1 ~ nomvar2, na.action = na.omit) *par défaut*

  model <- lm(nomvar1 ~ nomvar2, na.action = na.fail) *produit un avertissement si NA(s)*
- **Opérateur Retard : lag(variable, nombre de retard)**

RÉGRESSION LINÉAIRE DANS R

- **Visualiser les estimations : summary()**
- **Exemple :**

  monmodel <- lm(nomvar1 ~ nomvar2, data = mesdata)

  summary(monmodel) *ensemble des informations de la régression (dont tests, significativité, etc.)*

  monmodel$coef *ou* coef(monmodel) *pour les coefficients estimés*

  monmodel$residuals *ou* residuals(monmodel) *pour les résidus*

EXERCICE 2 - RÉGRESSION LINAIRE SIMPLE

QUEL EST L'IMPACT DE LA VITESSE SUR LA DISTANCE DE FREINAGE D'UNE VOITURE ?

RÉGRESSION LINÉAIRE MULTIPLE

- **Estimation :**

$$\hat{Y} = \hat{\alpha} + \widehat{\beta_1} X_1 + \widehat{\beta_2} X_2 + \widehat{\varepsilon_i}$$

- $\boldsymbol{\hat{\alpha}}$ est la valeur estimée de Y quand X_1 = 0 et X_2 = 0
- $\boldsymbol{\widehat{\beta_1}}$ et $\boldsymbol{\widehat{\beta_2}}$ sont les **effets marginaux** (ceteris paribus) :
  - Variation estimée de Y suite aux variations de X_1 quand X_2 est maintenu constant
  - Variation estimée de Y suite aux variations de X_2 quand X_1 est fixé

➢ **Biais :** inclusion de variables non-pertinentes (**sur-spécification** du modèle), variable omise (**sous-spécification**)

EXERCICE 3 - RÉGRESSION LINÉAIRE MULTIPLE

TEMPS RECORD SUR DIFFÉRENTES COURSES À PIED EN ECOSSE