

Table des matières

1	Rappel	2
1.1	Concepts fondamentaux de statistique inférentielle	2
1.1.1	Population, variable et échantillon	2
1.1.2	Espérance et variance	2
2	Régression linéaire simple	4
2.1	Modèle de la régression linéaire simple :	4
2.1.1	Écriture matricielle	4
2.2	Hypothèses du modèle	5
2.2.1	Hypothèse de linéarité	5
2.2.2	Hypothèse d'indépendance des erreurs	5
2.2.3	Hypothèse d'homoscédasticité	5
2.2.4	Hypothèse de normalité	6
2.3	Estimation des paramètres méthode des moindres carrés	7
2.3.1	Principe de la méthode des moindres carrés	7
2.3.2	Calcul des estimations des paramètres :	10
2.3.3	Interprétation des estimations	12
2.3.4	Propriétés des estimateurs des MCO :	12
2.3.5	Propriétés de l'erreur :	18
2.4	Evaluation de la qualité de la régression	19
2.4.1	Coefficient de corrélation de Pearson :	19
2.4.2	Coefficient de détermination R^2	19
2.5	Tests des hypothèses	20
2.5.1	Distribution	20
2.5.2	Test de Fisher	22
2.5.3	Test de Student	23
2.6	NumPy : bibliothèque de calcul scientifique	24

Ce chapitre présente les concepts fondamentaux de statistique inférentielle, d'algèbre linéaire et de calcul des probabilités nécessaires à la compréhension des modèles de régression linéaire simple et multiple .

1.1 Concepts fondamentaux de statistique inférentielle

La statistique inférentielle vise à tirer des conclusions sur une population entière à partir de l'étude d'un sous-ensemble de celle-ci, appelé *échantillon*.

1.1.1 Population, variable et échantillon

Définition 1.1.1 (Population). La **population** est l'ensemble complet des individus (sujets, objets, mesures) sur lesquels porte une étude statistique. La définition précise de la population d'intérêt est une étape cruciale avant toute analyse.

Définition 1.1.2 (Variable). Une **variable** est une caractéristique mesurable ou observable qui peut prendre différentes valeurs pour les différents individus de la population (par exemple : âge, taille, revenu, catégorie socio-professionnelle). La nature de la variable (quantitative continue/discrète, qualitative nominale/ordinaire) détermine les outils statistiques appropriés.

Définition 1.1.3 (Distribution). La **distribution** d'une variable décrit la manière dont ses différentes valeurs possibles sont réparties au sein de la population. Elle peut être caractérisée par des mesures de tendance centrale, de dispersion, et par sa forme (par exemple : loi normale).

Définition 1.1.4 (Échantillon). Un **échantillon** est un sous-ensemble d'individus sélectionnés dans la population. Pour que les inférences tirées de l'échantillon soient généralisables à la population, l'échantillon doit être **représentatif**. L'échantillonnage aléatoire simple est une méthode courante pour tenter d'assurer cette représentativité.

Définition 1.1.5 (Modèle d'échantillonnage). Le **modèle d'échantillonnage** décrit les hypothèses faites sur la manière dont l'échantillon a été constitué à partir de la population. Le modèle le plus courant est celui de l'échantillon aléatoire simple, où les observations $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ sont supposées indépendantes et identiquement distribuées (i.i.d.) selon la loi de la variable dans la population.

1.1.2 Espérance et variance

Ces concepts décrivent les caractéristiques centrales d'une variable aléatoire (théoriques) et d'un échantillon (empiriques).

Définitions théoriques

Soit X une variable aléatoire.

Définition 1.1.6 (Espérance mathématique). L'**espérance** de X , notée $\mathbb{E}[X]$ ou μ , est la valeur moyenne théorique de la variable.

— Si X est discrète, prenant les valeurs x_i avec probabilités $\mathbb{P}(X = x_i)$:

$$\mathbb{E}[X] = \sum_i x_i \mathbb{P}(X = x_i)$$

— Si X est continue avec une densité $f(x)$:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

Définition 1.1.7 (Variance). La **variance** de X , notée $\mathbb{V}(X)$ ou σ^2 , mesure la dispersion des valeurs de X autour de son espérance :

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

(Formule de König-Huygens)

Définition 1.1.8 (Écart-type). L'**écart-type** de X , noté σ_X ou simplement σ , est la racine carrée de la variance :

$$\sigma_X = \sqrt{\mathbb{V}(X)}$$

Il s'exprime dans la même unité que la variable X .

Propriété 1.1.1 (Propriétés de l'espérance et de la variance). Soient X, Y des variables aléatoires, et $a, b \in \mathbb{R}$ des constantes :

- $\mathbb{E}[aX + b] = a \mathbb{E}[X] + b$ (linéarité)
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{V}(aX + b) = a^2 \mathbb{V}(X)$
- $\mathbb{V}(X) \geq 0$
- Si X et Y sont indépendantes : $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$

Mesures empiriques (Échantillon)

Soit (X_1, \dots, X_n) un échantillon de taille n .

Définition 1.1.9 (Moyenne empirique). La **moyenne empirique** (ou moyenne de l'échantillon), notée \bar{X} , est définie par :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

C'est un estimateur de l'espérance $\mu = \mathbb{E}[X]$.

Définition 1.1.10 (Variance empirique biaisée). La **variance empirique biaisée** est donnée par :

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

C'est un estimateur sans biais de la variance σ^2 lorsque les X_i sont i.i.d.

Définition 1.1.11 (Écart-type empirique corrigé). L'**écart-type empirique corrigé**, noté S , est la racine carrée de la variance empirique corrigée :

$$S = \sqrt{S^2}$$

Régression linéaire simple

Ce chapitre est une introduction à la modélisation linéaire par le modèle le plus élémentaire, la régression linéaire simple. Le principe de ce modèle est de supposer qu'une variable Y est expliquée, modélisée par une fonction affine d'une seule variable explicative X . Après avoir explicité les hypothèses nécessaires et les termes du modèle, on discute les méthodes d'estimation des paramètres, les lois des estimateurs, les intervalles de confiance puis la signification des tests d'hypothèse et la qualité d'ajustement du modèle dont le but est la prévision.

2.1 Modèle de la régression linéaire simple :

Définition 2.1.1. La régression linéaire simple est une méthode statistique qui modélise la relation entre une variable dépendante Y et une variable indépendante X à l'aide d'une équation linéaire.

Cette relation est généralement exprimée par l'équation :

$$Y = \beta_1 X + \beta_0 + \epsilon$$

où :

- β_0 est l'ordonnée à l'origine, représentant la valeur de Y lorsque $X = 0$.
- β_1 est le coefficient de régression, indiquant la variation de Y pour une unité de variation de X .
- ϵ est le terme d'erreur, reflétant les variations de Y non expliquées par X .

Pour n observations, on peut écrire le modèle de régression linéaire simple sous la forme :

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- ϵ_i est une variable aléatoire, non observée,
- x_i est observée et non aléatoire
- y_i est observée et aléatoire.

Le principe de la régression linéaire simple est de trouver la droite (c'est-à-dire déterminer son équation) qui passe au plus près de l'ensemble des points formés par les couples (x_i, y_i) .

2.1.1 Écriture matricielle

Notons que le modèle de régression linéaire peut encore s'écrire sous forme matricielle, comme suit :

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

telle que :

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}; \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \text{ et } \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

2.2 Hypothèses du modèle

Pour qu'un modèle de régression linéaire puisse être construit et interprété de manière fiable, et pour que les méthodes d'estimation qui en découlent puissent produire des conclusions valides, il est indispensable de poser un certain nombre d'hypothèses sur la nature des données et des erreurs. Ces hypothèses constituent le cadre théorique du modèle : elles sont essentielles pour comprendre ses propriétés, ses limites et les conditions dans lesquelles ses résultats peuvent être généralisés. Elles s'appliquent de manière similaire dans les cas de régression linéaire simple comme multiple.

2.2.1 Hypothèse de linéarité

l'hypothèse de linéarité stipule que la relation entre la variable indépendante X et la variable dépendante Y peut être modélisée par une fonction linéaire. En d'autres termes, un changement unitaire dans la variable explicative doit entraîner un changement constant dans la variable réponse. Cette hypothèse peut être vérifiée graphiquement à l'aide d'un nuage de points, où la tendance linéaire peut être visualisée. En d'autres termes, la moyenne de Y , pour une valeur donnée de X , doit se situer sur une droite définie par l'équation :

$$E(Y|X = x) = \beta_0 + \beta_1 x$$

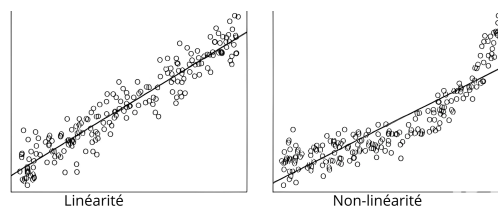


FIGURE 2.1 – Illustration de relation linéaire et non linéaire entre variables

2.2.2 Hypothèse d'indépendance des erreurs

Cette hypothèse stipule que les erreurs (ou résidus) ϵ_i du modèle de régression sont indépendantes les uns des autres. En d'autres termes, il n'y a pas de corrélation entre les erreurs pour des observations différentes. Mathématiquement, cela s'exprime par :

$$Cov(\epsilon_i, \epsilon_j) = 0 \quad \text{pour tout } i \neq j$$

2.2.3 Hypothèse d'homoscédasticité

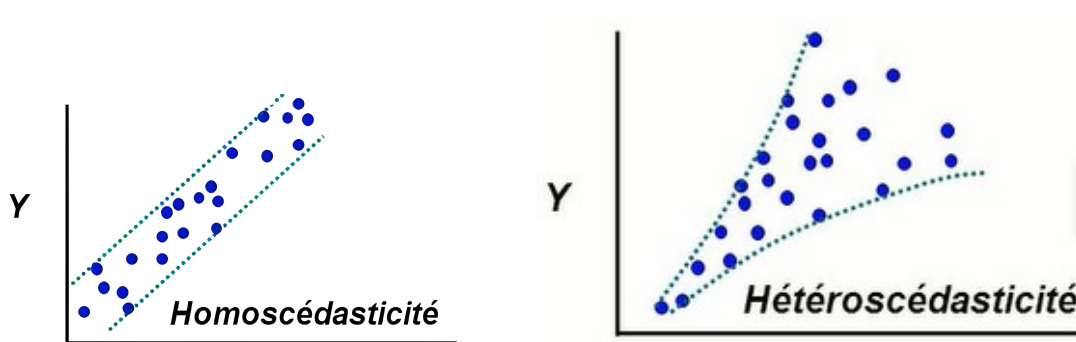
Une autre hypothèse clé est celle de l'homoscédasticité, qui stipule que la variance des résidus doit être constante à travers toutes les valeurs prédites. Cela signifie que l'éparpillement des résidus ne doit pas changer selon les valeurs de la variable dépendante. Si la variance des résidus augmente ou diminue

systématiquement, on parle alors d'hétéroscédasticité, ce qui peut entraîner des estimations biaisées et réduire la précision des prévisions.

$$\mathbb{V}(\varepsilon_i) = \sigma^2, \quad \forall i = 1, \dots, n,$$

Ou de manière équivalente :

$$\mathbb{V}(y_i) = \sigma^2, \quad \forall i = 1, \dots, n.$$



2.2.4 Hypothèse de normalité

Les erreurs doivent être normalement distribuées autour de la droite de régression pour chaque groupe de X. Cela est crucial pour la validité des tests d'hypothèses statistiques et pour assurer que les intervalles de confiance soient fiables.

On peut formuler cette hypothèse comme suit :

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

(c) non normalité

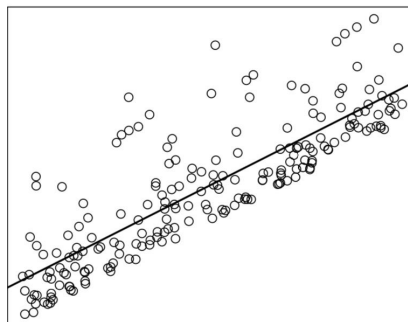


FIGURE 2.2 – Nuage de points lorsque l'hypothèse de normalité n'est pas respectée

2.3 Estimation des paramètres méthode des moindres carrés

Une fois le modèle de régression linéaire défini et ses hypothèses fondamentales posées, l'étape suivante consiste à estimer les valeurs inconnues de ses paramètres, notamment le vecteur des coefficients de régression $\beta = (\beta_0, \beta_1)^T$ ainsi que la variance des erreurs σ^2 . Ces paramètres traduisent respectivement l'influence des variables explicatives sur la variable dépendante, et la dispersion des erreurs autour de la droite de régression. La Méthode des Moindres Carrés (MMC) est la méthode d'estimation la plus utilisée et la plus intuitive. Elle repose sur le principe de la minimisation de l'erreur quadratique. Il s'agit de déterminer les valeurs des coefficients β qui minimisent la somme des carrés des écarts entre les valeurs observées de la variable dépendante Y_i et les valeurs prédites par le modèle \hat{Y}_i . Il convient de souligner que, si la MMC permet d'obtenir directement une estimation analytique des coefficients β , elle ne fournit pas d'estimation directe de la variance des erreurs σ^2 dans le cadre même de sa procédure de minimisation. Cette dernière est généralement obtenue de manière secondaire, à partir des résidus calculés après l'estimation de β .

2.3.1 Principe de la méthode des moindres carrés

La méthode des moindres carrés ordinaires (MCO) vise à estimer les paramètres β_0 (ordonnée à l'origine) et β_1 (pente) du modèle de régression linéaire :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

où ε_i représente l'erreur aléatoire.

La méthode consiste à minimiser une **fonction de coût** (ou critère) qui mesure l'écart global entre les valeurs observées y_i et les valeurs prédites par le modèle :

$$J(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

Cette fonction quadratique est appelée **somme des carrés des résidus** (SCR). Elle présente les propriétés suivantes :

- Toujours positive (somme de carrés)
- Convexe (admet un minimum global unique sous certaines conditions)
- Sensible aux valeurs aberrantes (car les erreurs sont au carré)

Le problème d'optimisation s'écrit formellement :

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0, \beta_1) \in \mathbb{R}^2}{\text{Arg min}} J(\beta_0, \beta_1) = \underset{(\beta_0, \beta_1) \in \mathbb{R}^2}{\text{Arg min}} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$$

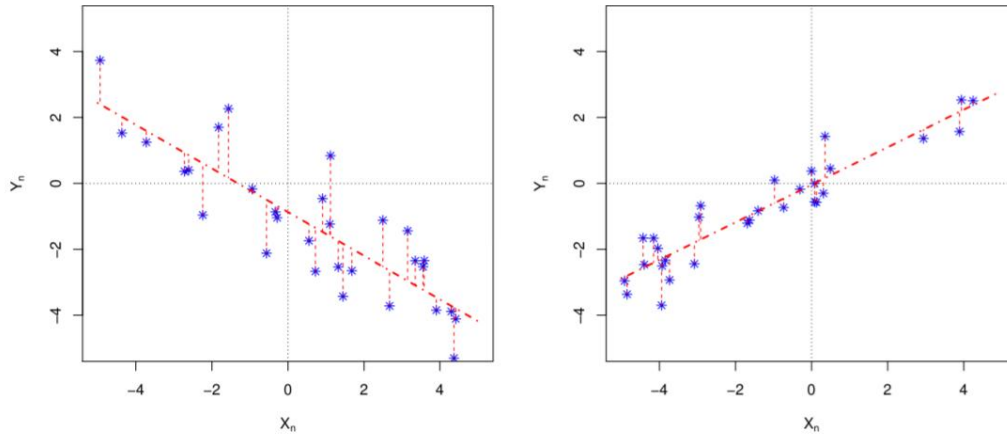
- **Vocabulaire :**

- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$: valeur prédite par le modèle
- $\hat{\varepsilon}_i = y_i - \hat{y}_i$: résidu (erreur observée)
- $\sum \hat{\varepsilon}_i^2$: mesure la qualité d'ajustement du modèle

• Interprétation graphique

Graphiquement, $\hat{\beta}_0$ et $\hat{\beta}_1$ sont construits pour minimiser les distances verticales entre les observations (y_n) et la droite de régression théorique $y = \beta_0 + \beta_1 x$.

Nous avons représenté ces distances sur les figures ci-dessous.



La droite d'équation $y = \hat{\beta}_0 + \hat{\beta}_1 x$ est la droite de régression estimée sur le nuage de points.

Pour la régression linéaire simple, la fonction de coût est donnée par

$$J(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2.$$

Théorème 2.3.1.

J est strictement convexe et admet un unique minimum global.

Preuve 2.3.1.

Pour démontrer l'existence et l'unicité du minimum :

1. Existence du minimum :

- J est continue
- J est coercive vers $+\infty$, car

$$\lim_{\|\beta\| \rightarrow \infty} J(\beta_0, \beta_1) = +\infty$$

- Par le théorème des valeurs extrêmes, J admet un minimum global.

2. Unicité :

Si l'on définit le vecteur des paramètres

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},$$

et la matrice de conception X telle que

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

alors la fonction de coût peut s'écrire en forme matricielle :

$$J(\boldsymbol{\beta}) = (\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta}),$$

où

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Dans le but d'étudier la convexité de J (et ainsi postuler l'existence et l'unicité de l'optimum), on calcule la matrice Hessienne de J . Ici, la Hessienne est donnée par :

$$\nabla^2 J(\boldsymbol{\beta}) = 2X^T X.$$

Calculons $X^T X$:

$$X^T X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}.$$

La matrice Hessienne devient donc :

$$\nabla^2 J(\boldsymbol{\beta}) = 2 \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}.$$

On va prouver que toutes les valeurs propres de la matrice $X^T X$ sont **positives ou nulles**. Pour ce faire, nous démontrons que $X^T X$ est une matrice **semi-définie positive**, c'est-à-dire que pour tout vecteur $\mathbf{v} \in \mathbb{R}^n$, on a :

$$\mathbf{v}^T (X^T X) \mathbf{v} \geq 0.$$

Considerons un vecteur arbitraire ($\mathbf{v} \in \mathbb{R}^n$). Nous évaluons l'expression quadratique associée :

$$\mathbf{v}^T (X^T X) \mathbf{v}.$$

Employons l'associativité du produit matriciel :

$$\mathbf{v}^T (X^T X) \mathbf{v} = (X\mathbf{v})^T (X\mathbf{v}).$$

Remarquons que cette expression est un *produit scalaire*. En effet, pour tout vecteur \mathbf{w} , la norme euclidienne est définie par :

$$\|\mathbf{w}\| = \sqrt{\mathbf{w}^T \mathbf{w}}.$$

Alors :

$$\|X\mathbf{v}\|^2 = (X\mathbf{v})^T (X\mathbf{v}).$$

Or,

$$\|X\mathbf{v}\|^2 \geq 0.$$

donc :

$$\mathbf{v}^T(X^T X)\mathbf{v} = \|X\mathbf{v}\|^2 \geq 0.$$

Alors la matrice $X^T X$ est semi-définie positif.

Donc Ses valeurs propres sont non-négatives (la matrice Hessienne est semi-définie positif), ce qui assure la convexité de J .

Maintenant on va calculer Déterminant de la Hessienne :

Le déterminant de $\nabla^2 J$ est :

$$\det(\nabla^2 J) = 4 \left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) = 4n^2 \text{Var}(X).$$

On a les x_i ne sont pas tous égaux d'où $\text{Var}(X) > 0$, alors $\det(\nabla^2 J) > 0$ et la matrice Hessienne est strictement définie positive .

Puisque la fonction J admet un minimum. De plus, la matrice Hessienne étant strictement définie positive implique que J est strictement convexe, garantissant ainsi l'unicité de ce minimum.

■

2.3.2 Calcul des estimations des paramètres :

On appelle estimateurs des moindres carrés ordinaires (MCO) de β_0 et β_1 les estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$ tels que

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg \min_{(\beta_0, \beta_1) \in \mathbb{R}^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$$

la méthode des moindres carrés consiste à chercher les valeurs pour β_0 et β_1 de telle sorte que la somme des carrés des erreurs $\sum_{i=1}^n \varepsilon_i^2$ soit la plus petite possible, c'est-à-dire que la droite passe le plus près possible de l'ensemble des points.

la résolution de ce problème de minimisation mène aux estimations $\hat{\beta}_1$ et $\hat{\beta}_0$ des paramètres β_1 et β_0 suivantes :

$$\begin{cases} \hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\text{Cov}(x, y)}{\text{Var}(x)} \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \end{cases}$$

Preuve 2.3.2.

On a $J(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$.

La fonction J est une fonction de deux variables réelles, ses points critiques sont obtenus par la résolution du système :

$$\begin{cases} \frac{\partial J}{\partial \beta_0}(\beta_0, \beta_1) = 0 \\ \frac{\partial J}{\partial \beta_1}(\beta_0, \beta_1) = 0 \end{cases} \quad (2.1)$$

donc :

$$\begin{cases} -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0, \\ -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0. \end{cases} \quad (2.2)$$

Ainsi, les valeurs estimées de β_0 et β_1 sont données par :

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \quad (*)$$

$$\sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) = 0 \quad (**)$$

En développant les équations (*) et (**), nous obtenons :

$$\begin{aligned} \sum_{i=1}^n y_i - n\beta_0 - \beta_1 \sum_{i=1}^n x_i &= 0 \\ \sum_{i=1}^n x_i y_i - \beta_0 \sum_{i=1}^n x_i - \beta_1 \sum_{i=1}^n x_i^2 &= 0 \end{aligned}$$

C'est-à-dire :

$$\begin{cases} \sum_{i=1}^n y_i = n\beta_0 + \beta_1 \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i = \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 \end{cases} \quad (\text{Ces équations sont appelées } \textit{équations normales}).$$

Sachant que :

$$\begin{aligned} \bar{X} &= \frac{x_1 + \cdots + x_n}{n} = \sum_{i=1}^n x_i/n \\ \bar{Y} &= \frac{y_1 + \cdots + y_n}{n} = \sum_{i=1}^n y_i/n \end{aligned}$$

nous trouvons les solutions suivantes :

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

et

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)/n}{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2/n} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{Cov(x, y)}{Var(x)}$$

On obtient le point critique $(\hat{\beta}_0, \hat{\beta}_1) = \left(\bar{y} - \hat{\beta}_1 \bar{x}, \frac{Cov(x, y)}{Var(x)} \right)$

Alors nous avons démontré que le couple (β_0, β_1) constituait la solution optimale du problème d'optimisation .



2.3.3 Interprétation des estimations

Le signe des coefficients β_0 et β_1 est essentiel pour estimer le comportement de la regression

- Influence du signe de β_1 (la pente) :
 - **Si** $\beta_1 > 0$: La droite est **croissante** → relation positive entre X et Y . *Exemple* : plus X croît (comme les heures de révision), plus Y croît (la note).
 - **Si** $\beta_1 < 0$: La droite est **décroissante** → relation négative entre X et Y . *Exemple* : plus X augmente (comme le stress), plus Y diminue (la performance).
 - **Si** $\beta_1 = 0$: La droite est **horizontale** → absence de relation linéaire entre X et Y .
- Effet du signe de β_0 (l'ordonnée à l'origine) :
 - **Si** $\beta_0 > 0$: Lorsque $X = 0$, la valeur prédite de Y est positive. *Exemple* : si $\beta_0 = 10$, même sans heures de révision ($X = 0$), la note prévue (Y) est 10.
 - **Si** $\beta_0 < 0$: Lorsque $X = 0$, Y commence en dessous de zéro. *Remarque* : Ce cas peut être irréaliste selon le contexte (exemple : une note négative n'a pas de sens).

2.3.4 Propriétés des estimateurs des MCO :

Théorème 2.3.2.

Les coefficients $\hat{\beta}_0$ et $\hat{\beta}_1$ sont des estimateurs sans biais de β_0 et β_1 .

Preuve 2.3.3.

- Le coefficient $\hat{\beta}_1$ est un estimateur sans biais de β_1 . En effet :
On commence par la formule de l'estimateur de la pente $\hat{\beta}_1$:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\text{Cov}(x, y)}{\text{V}(x)} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}\end{aligned}$$

Nous savons que le modèle linéaire réel est $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.
En prenant la moyenne des deux côtés :

$$\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{\varepsilon}$$

Ainsi, la déviation $y_i - \bar{y}$ peut être exprimée comme :

$$y_i - \bar{y} = \beta_1 (x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon})$$

Substituons cette expression dans la formule de $\hat{\beta}_1$:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x}) [\beta_1 (x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon})]}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon})}{\sum_{i=1}^n (x_i - \bar{x})^2}\end{aligned}$$

Le terme $\sum_{i=1}^n (x_i - \bar{x})\bar{\varepsilon}$ vaut 0 car :

$$\sum_{i=1}^n (x_i - \bar{x}) = n\bar{x} - n\bar{x} = 0$$

Donc :

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

En prenant l'espérance (conditionnellement aux x_i) :

$$\begin{aligned} \mathbb{E}(\hat{\beta}_1) &= \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})\mathbb{E}(\varepsilon_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta_1 \quad (\text{car } \mathbb{E}(\varepsilon_i) = 0 \text{ par hypothèse}) \end{aligned}$$

- Le coefficient $\hat{\beta}_0$ est un estimateur sans biais de β_0 , en effet :
L'estimateur de l'ordonnée à l'origine est :

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

En substituant $\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{\varepsilon}$:

$$\begin{aligned} \mathbb{E}(\hat{\beta}_0) &= \mathbb{E}(\beta_0 + \beta_1 \bar{x} + \bar{\varepsilon} - \hat{\beta}_1 \bar{x}) \\ &= \beta_0 + \beta_1 \bar{x} - \bar{x} \mathbb{E}(\hat{\beta}_1) \quad (\text{car } \mathbb{E}(\bar{\varepsilon}) = 0) \\ &= \beta_0 \quad (\text{car } \mathbb{E}(\hat{\beta}_1) = \beta_1) \end{aligned}$$

■

On peut également exprimer la variances et la covariance de ces estimateurs

Théorème 2.3.3.

Les variances des estimateurs sont :

$$\mathbb{V}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \quad (2.3)$$

et

$$\mathbb{V}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.4)$$

tandis que leur covariance vaut :

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.5)$$

Preuve 2.3.4.

- Variance de $\hat{\beta}_1$

Nous avons précédemment démontré que l'estimateur de la pente $\hat{\beta}_1$ peut s'écrire sous la forme :

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Calculons sa variance :

$$\begin{aligned}
 V(\hat{\beta}_1) &= V \left[\beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) \epsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \\
 &= V(\beta_1) + V \left[\frac{\sum_{i=1}^n (x_i - \bar{x}) \epsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \quad (\text{car } \beta_1 \text{ est constant}) \\
 &= 0 + V \left[\frac{\sum_{i=1}^n (x_i - \bar{x}) \epsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \\
 &= \frac{1}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} V \left(\sum_{i=1}^n (x_i - \bar{x}) \epsilon_i \right) \\
 &= \frac{1}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} \sum_{i=1}^n (x_i - \bar{x})^2 V(\epsilon_i) \quad (\text{par indépendance des } \epsilon_i) \\
 &= \frac{1}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} \sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2 \quad (\text{car } V(\epsilon_i) = \sigma^2) \\
 &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}
 \end{aligned}$$

- Variance de $\hat{\beta}_0$

L'estimateur de l'ordonnée à l'origine $\hat{\beta}_0$ est défini par :

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

Nous savons que $\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{\epsilon}$. En substituant \bar{y} dans l'expression de $\hat{\beta}_0$:

$$\begin{aligned}
 \hat{\beta}_0 &= (\beta_0 + \beta_1 \bar{x} + \bar{\epsilon}) - \hat{\beta}_1 \bar{x} \\
 &= \beta_0 + \bar{x}(\beta_1 - \hat{\beta}_1) + \bar{\epsilon}
 \end{aligned}$$

D'où, la déviation de $\hat{\beta}_0$ par rapport à sa vraie valeur β_0 est :

$$\hat{\beta}_0 - \beta_0 = -\bar{x}(\hat{\beta}_1 - \beta_1) + \bar{\epsilon}$$

Calculons maintenant la variance de $\hat{\beta}_0$:

$$\mathbb{V}(\hat{\beta}_0) = \mathbb{V} \left[-\bar{x}(\hat{\beta}_1 - \beta_1) + \bar{\epsilon} \right]$$

Nous avons montré précédemment que $\mathbb{E}[(\hat{\beta}_1 - \beta_1)\bar{\epsilon}] = 0$ (voir la preuve de la covariance ci-dessous), ce qui implique que $\hat{\beta}_1$ et $\bar{\epsilon}$ sont non corrélés. Par conséquent, $\mathbb{V}(A + B) = \mathbb{V}(A) + \mathbb{V}(B)$ si A et B sont non corrélés.

$$\begin{aligned}
 \mathbb{V}(\hat{\beta}_0) &= \mathbb{V} \left[-\bar{x}(\hat{\beta}_1 - \beta_1) \right] + \mathbb{V}(\bar{\epsilon}) \\
 &= (-\bar{x})^2 \mathbb{V}(\hat{\beta}_1 - \beta_1) + \mathbb{V} \left(\frac{1}{n} \sum_{i=1}^n \epsilon_i \right)
 \end{aligned}$$

Puisque β_1 est une constante, $\mathbb{V}(\hat{\beta}_1 - \beta_1) = \mathbb{V}(\hat{\beta}_1)$. De plus, par indépendance et homoscédasticité des ε_i , $\mathbb{V}\left(\frac{1}{n} \sum \varepsilon_i\right) = \frac{1}{n^2} \sum \mathbb{V}(\varepsilon_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$.

$$= \bar{x}^2 \mathbb{V}(\hat{\beta}_1) + \frac{\sigma^2}{n}$$

En substituant l'expression de $\mathbb{V}(\hat{\beta}_1)$ que nous avons déjà trouvée :

$$\begin{aligned} &= \bar{x}^2 \cdot \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{\sigma^2}{n} \\ &= \sigma^2 \left(\frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{1}{n} \right) \\ &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \end{aligned}$$

Ainsi, la variance de $\hat{\beta}_0$ est $\mathbb{V}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$.

- Covariance entre $\hat{\beta}_1$ et $\hat{\beta}_0$

Nous voulons montrer que $\text{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\bar{x} \cdot \mathbb{V}(\hat{\beta}_1)$.

Par définition, la covariance entre deux variables aléatoires X et Y est

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

En utilisant le fait que $\mathbb{E}(\hat{\beta}_1) = \beta_1$ et $\mathbb{E}(\hat{\beta}_0) = \beta_0$ (démontré précédemment) :

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_0) = \mathbb{E}[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_0 - \beta_0)]$$

Nous avons établi que $\hat{\beta}_0 - \beta_0 = -\bar{x}(\hat{\beta}_1 - \beta_1) + \bar{\varepsilon}$.

Substituons cette expression dans la covariance :

$$\begin{aligned} \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) &= \mathbb{E}[(\hat{\beta}_1 - \beta_1) \left(-\bar{x}(\hat{\beta}_1 - \beta_1) + \bar{\varepsilon} \right)] \\ &= \mathbb{E} \left[-\bar{x}(\hat{\beta}_1 - \beta_1)^2 + (\hat{\beta}_1 - \beta_1)\bar{\varepsilon} \right] \end{aligned}$$

Par linéarité de l'espérance :

$$= -\bar{x} \mathbb{E}[(\hat{\beta}_1 - \beta_1)^2] + \mathbb{E}[(\hat{\beta}_1 - \beta_1)\bar{\varepsilon}]$$

Puisque $\mathbb{E}[(\hat{\beta}_1 - \mathbb{E}(\hat{\beta}_1))^2] = \mathbb{V}(\hat{\beta}_1)$, le premier terme est $-\bar{x} \mathbb{V}(\hat{\beta}_1)$.

$$= -\bar{x} \mathbb{V}(\hat{\beta}_1) + \mathbb{E}[(\hat{\beta}_1 - \beta_1)\bar{\varepsilon}]$$

Il nous reste à démontrer que le deuxième terme $\mathbb{E}[(\hat{\beta}_1 - \beta_1)\bar{\varepsilon}]$ est nul.

Nous savons que $\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$ et $\bar{\varepsilon} = \frac{1}{n} \sum_{j=1}^n \varepsilon_j$.

$$\begin{aligned} \mathbb{E}[(\hat{\beta}_1 - \beta_1)\bar{\varepsilon}] &= \mathbb{E} \left[\frac{\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \frac{1}{n} \sum_{j=1}^n \varepsilon_j \right] \\ &= \frac{1}{n \sum_{i=1}^n (x_i - \bar{x})^2} \mathbb{E} \left[\left(\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i \right) \left(\sum_{j=1}^n \varepsilon_j \right) \right] \end{aligned}$$

Soit $\omega_i = (x_i - \bar{x})$. L'expression dans l'espérance devient $\left(\sum_{i=1}^n \omega_i \varepsilon_i \right) \left(\sum_{j=1}^n \varepsilon_j \right)$.

En développant ce produit, on obtient des termes de la forme $\omega_i \varepsilon_i^2$ et $\omega_i \varepsilon_i \varepsilon_j$ pour $i \neq j$.

$$\mathbb{E}[(\hat{\beta}_1 - \beta_1)\bar{\varepsilon}] = \frac{1}{n \sum \omega_i^2} \mathbb{E} \left[\sum_{i=1}^n \omega_i \varepsilon_i^2 + \sum_{i \neq j} \omega_i \varepsilon_i \varepsilon_j \right]$$

Par linéarité de l'espérance :

$$= \frac{1}{n \sum \omega_i^2} \left(\sum_{i=1}^n \omega_i \mathbb{E}[\varepsilon_i^2] + \sum_{i \neq j} \omega_i \mathbb{E}[\varepsilon_i \varepsilon_j] \right)$$

En vertu des hypothèses du MMC :

- $\mathbb{E}(\varepsilon_i^2) = \mathbb{V}(\varepsilon_i) + [\mathbb{E}(\varepsilon_i)]^2 = \sigma^2 + 0^2 = \sigma^2$.
- $\mathbb{E}(\varepsilon_i \varepsilon_j) = \text{Cov}(\varepsilon_i, \varepsilon_j) + \mathbb{E}(\varepsilon_i)\mathbb{E}(\varepsilon_j) = 0 + 0 \cdot 0 = 0$ pour $i \neq j$ (par non-autocorrélation et moyenne nulle).

En substituant ces résultats :

$$\begin{aligned} \mathbb{E}[(\hat{\beta}_1 - \beta_1)\bar{\varepsilon}] &= \frac{1}{n \sum \omega_i^2} \left(\sum_{i=1}^n \omega_i \sigma^2 + \sum_{i \neq j} \omega_i \cdot 0 \right) \\ &= \frac{1}{n \sum \omega_i^2} \left(\sigma^2 \sum_{i=1}^n \omega_i \right) \\ &= \frac{\sigma^2}{n \sum (x_i - \bar{x})^2} \sum_{i=1}^n (x_i - \bar{x}) \end{aligned}$$

Puisque $\sum_{i=1}^n (x_i - \bar{x}) = 0$ (somme des écarts à la moyenne est nulle) :

$$\begin{aligned} &= \frac{\sigma^2}{n \sum (x_i - \bar{x})^2} \cdot 0 \\ &= 0 \end{aligned}$$

Finalement, en substituant ce résultat dans l'expression de la covariance :

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\bar{x}\mathbb{V}(\hat{\beta}_1) + 0$$

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\bar{x}\mathbb{V}(\hat{\beta}_1)$$

Ceci conclut la démonstration de la covariance. ■

Théorème 2.3.4.

les estimateurs de MCO $\hat{\beta}_0$ et $\hat{\beta}_1$ sont convergents. Cela se traduit par le fait que la variance de l'estimateur tend vers zéro :

$$\lim_{n \rightarrow \infty} \mathbb{V}(\hat{\beta}_0) = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \mathbb{V}(\hat{\beta}_1) = 0.$$

Preuve 2.3.5.

On va montrer que

$$\lim_{n \rightarrow \infty} \mathbb{V}(\hat{\beta}_1) = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \mathbb{V}(\hat{\beta}_0) = 0$$

On sait que

$$\mathbb{V}(\hat{\beta}_1) = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{et} \quad \mathbb{V}(\hat{\beta}_0) = \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

Avec $\sigma_\varepsilon^2 = \mathbb{V}(\varepsilon)$ Alors

$$\lim_{n \rightarrow \infty} \mathbb{V}(\hat{\beta}_1) = \lim_{n \rightarrow \infty} \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma_\varepsilon^2}{+\infty} = 0$$

Donc $\hat{\beta}_1$ est un estimateur convergent.

$$\lim_{n \rightarrow \infty} \mathbb{V}(\hat{\beta}_0) = \lim_{n \rightarrow \infty} \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) = \sigma_\varepsilon^2 \left(\frac{1}{+\infty} + 0 \right) = 0$$

D'où $\hat{\beta}_0$ est un estimateur convergent. ■

Maintenant, proposons un estimateur sans biais de σ^2

Théorème 2.3.5.

La statistique $\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{(n-2)}$ est un estimateur sans biais de σ^2 .

Preuve 2.3.6.

Réécrivons les résidus en constatant que $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ et $\beta_0 = \bar{y} - \beta_1 \bar{x}$,

ce qui donne :

$$\begin{aligned} \hat{\varepsilon}_i &= \beta_0 + \beta_1 x_i + \varepsilon_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \\ &= \bar{y} - \beta_1 \bar{x} + \beta_1 x_i + \varepsilon_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i \\ &= (\beta_1 - \hat{\beta}_1)(x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon}) \end{aligned}$$

En développant et en nous servant de l'écriture vue plus haut :

$$\hat{\beta}_1 = \beta_1 + \frac{\sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x})^2}$$

nous avons :

$$\begin{aligned} \sum \hat{\varepsilon}_i^2 &= (\beta_1 - \hat{\beta}_1)^2 \sum (x_i - \bar{x})^2 + \sum (\varepsilon_i - \bar{\varepsilon})^2 + 2(\beta_1 - \hat{\beta}_1) \sum (x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon}) \\ &= (\beta_1 - \hat{\beta}_1)^2 \sum (x_i - \bar{x})^2 + \sum (\varepsilon_i - \bar{\varepsilon})^2 - 2(\beta_1 - \hat{\beta}_1)^2 \sum (x_i - \bar{x})^2 \end{aligned}$$

Prenons l'espérance :

$$\mathbb{E} \left(\sum \hat{\varepsilon}_i^2 \right) = \mathbb{E} \left(\sum (\varepsilon_i - \bar{\varepsilon})^2 \right) - \sum (x_i - \bar{x})^2 \text{V}(\hat{\beta}_1) = (n-2)\sigma^2$$

Bien sûr, lorsque n est grand, cet estimateur diffère très peu de l'estimateur empirique de la variance des résidus, à savoir $\sum_{i=1}^n \hat{\varepsilon}_i^2 / n$.

■

2.3.5 Propriétés de l'erreur :

- ε_i est l'erreur inconnue introduite dans la spécification du modèle :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (2.6)$$

- L'estimation des paramètres du modèle nous a permis de déduire la valeur estimée de l'endogène Y pour l'individu i :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (2.7)$$

- Ainsi, on déduit l'erreur observée $\hat{\varepsilon}_i$, appelée « résidu », à partir de (4.6)-(4.7) :

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

Remarque 2.3.1.

- La relation $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ montre que la droite des MCO passe par le centre de gravité du nuage (\bar{x}, \bar{y}) .
- Les expressions obtenues pour $\hat{\beta}_0$ et $\hat{\beta}_1$ montrent que ces deux estimateurs sont linéaires par rapport au vecteur $Y = [y_1, \dots, y_n]'$.
- La somme des résidus est nulle

$$\begin{aligned} \sum_{i=1}^n \hat{\varepsilon}_i &= \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\ &= \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_0 - \sum_{i=1}^n \hat{\beta}_1 x_i \\ &= n \frac{1}{n} \sum_{i=1}^n y_i - n \hat{\beta}_0 - n \frac{1}{n} \sum_{i=1}^n \hat{\beta}_1 x_i \\ &= n \bar{y} - n \hat{\beta}_0 - n \hat{\beta}_1 \bar{x} \\ &= n \hat{\beta}_0 - n \underbrace{(\bar{y} - \hat{\beta}_1 \bar{x})}_{\hat{\beta}_0} \end{aligned}$$

Donc :

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

- Minimiser les critères des moindres carrés revient donc à définir la droite de régression comme étant la droite pour laquelle les erreurs de régression ont une moyenne nulle et une variance minimale.

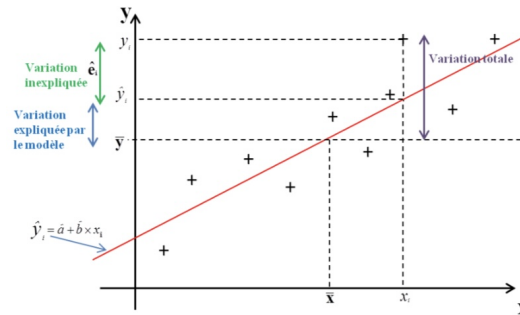


FIGURE 2.3 – Droite obtenue par régression linéaire simple

2.4 Evaluation de la qualité de la régression

Une fois les coefficients du modèle de régression estimés, que ce soit par la méthode des moindres carrés ordinaires, la descente de gradient, ou le maximum de vraisemblance, l'étape suivante consiste à évaluer la pertinence et la qualité de cet ajustement.

Cette évaluation permet de déterminer si le modèle est non seulement statistiquement valide, mais aussi s'il offre une bonne capacité explicative et prédictive des phénomènes étudiés.

Elle s'appuie sur des indicateurs clés tels que le coefficient de détermination (R^2) et le coefficient de corrélation r .

2.4.1 Coefficient de corrélation de Pearson :

Le **coefficient de corrélation de Pearson**, noté r , mesure la force et la direction de la relation linéaire entre deux variables quantitatives. Sa valeur est comprise entre -1 et $+1$:

$$r = \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \in [-1; 1]$$

- $r = +1$: corrélation linéaire positive parfaite.
- $r = -1$: corrélation linéaire négative parfaite.
- $r = 0$: absence de corrélation linéaire.

L'interprétation de la valeur absolue de r est donnée par le tableau suivant :

Valeur de $ r $	Interprétation
$0.0 \leq r < 0.2$	Corrélation très faible ou négligeable
$0.2 \leq r < 0.4$	Corrélation faible
$0.4 \leq r < 0.6$	Corrélation modérée
$0.6 \leq r < 0.8$	Corrélation forte
$0.8 \leq r \leq 1.0$	Corrélation très forte

Ainsi, si $r \geq 0.8$, cela indique une **corrélation très forte et positive** entre les deux variables. À l'inverse, si $r \leq -0.8$, on parle d'une **corrélation très forte et négative**.

2.4.2 Coefficient de détermination R^2

Le **coefficient de détermination**, noté R^2 , est une mesure statistique qui évalue la proportion de la variance de la variable dépendante y expliquée par la ou les variables indépendantes x dans un modèle de régression. Il est défini par la formule suivante :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\text{variabilité expliquée (SCE)}}{\text{variabilité totale (SCT)}} = 1 - \frac{\text{SCR}}{\text{SCT}}$$

- $R^2 = 1$: le modèle explique parfaitement la variance de y ;
- $R^2 = 0$: le modèle n'explique aucune variance de y ;
- $0 < R^2 < 1$: le modèle explique partiellement la variance de y .

Dans le cas d'une régression linéaire simple (une seule variable explicative), le coefficient de détermination est égal au carré du coefficient de corrélation de Pearson r entre x et y :

$$R^2 = r^2$$

Cette relation indique que R^2 mesure la proportion de la variance de y expliquée par x . Par exemple, si $r = 0.9$, alors $R^2 = 0.81$, ce qui signifie que 81% de la variance de y est expliquée par x

Remarque 2.4.1. Un R^2 élevé indique une bonne qualité d'ajustement du modèle, mais ne garantit pas la validité du modèle. Il est essentiel de compléter cette analyse par l'examen des résidus et d'autres tests diagnostiques pour évaluer la pertinence du modèle.

2.5 Tests des hypothèses

Lorsqu'un modèle de régression linéaire est ajusté et que ses paramètres sont estimés, il reste une étape critique à accomplir avant de pouvoir parler de relations découvertes. Trouver l'équation de la droite ou de l'hyperplan qui semble s'adapter le mieux selon le principe des moindres carrés ne suffit pas ; il est impératif de vérifier que des relations fur découvertes, et non simplement que les caractéristiques rencontrées durant l'étude d'une certaine relation étaient accidentellement vraies dans l'échantillon. C'est à ce point es test d'hypothèse et de signification statistique entre en scène. Cette section traite des techniques statistiques qui nous apprennent à décider quelles des variables explicatives ne convient pas dans ce modèle et quelle est la confiance exprimée dans l'acceptation du modèle tout entier.

Avant de procéder aux tests d'hypothèses, il est essentiel de caractériser la distribution des estimateurs des coefficients

2.5.1 Distribution

Pour effectuer des tests d'hypothèses sur les coefficients de régression β_0 et β_1 , il est essentiel de connaître la distribution d'échantillonnage de leurs estimateurs par Moindres Carrés Ordinaires (MCO), $\hat{\beta}_0$ et $\hat{\beta}_1$. Cette distribution dépend cruciallement des hypothèses du modèle de régression linéaire classique, en particulier l'hypothèse de normalité des erreurs.

On considère le modèle suivant : $\forall i \in 0, \dots, n$

$$y_i = \beta_1 x_i + \beta_0 + \epsilon_i$$

On sait que $\hat{\beta}_0$ a une espérance :

$$E(\hat{\beta}_0) = \beta_0$$

et une variance :

$$V(\hat{\beta}_0) = \sigma_{\hat{\beta}_0}^2$$

c'est à dire :

$$\hat{\beta}_0 \sim N(\beta_0, \sigma_{\hat{\beta}_0}^2)$$

Donc la forme standardisée :

$$\frac{\hat{\beta}_0 - \beta_0}{\sigma_{\hat{\beta}_0}} \sim N(0, 1)$$

or :

$$\sigma_{\hat{\beta}_0}^2 = \sigma_\epsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

Ainsi :

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \hat{\sigma}_\epsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

D'après hypothèse de normalité des Erreurs

$$\epsilon_i \sim N(0, \sigma_\epsilon^2)$$

Donc :

$$\frac{\epsilon_i}{\sigma_\epsilon} \sim N(0, 1)$$

Comme $\hat{\epsilon}_i$ est une réalisation de ϵ_i :

$$\frac{\hat{\epsilon}_i}{\sigma_\epsilon} \sim N(0, 1)$$

Passant au carré et sommant par rapport à i :

$$\sum_{i=1}^n \left(\frac{\hat{\epsilon}_i}{\sigma_\epsilon} \right)^2 = \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{\sigma_\epsilon^2} \sim \chi_{(n-2)}^2$$

On a l'estimateur de la variance :

$$\frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{n-2} = \hat{\sigma}_\epsilon^2$$

donc :

$$\frac{(n-2)\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2} \sim \chi_{(n-2)}^2$$

Ainsi :

$$\frac{\hat{\sigma}_{\hat{\beta}_0}^2}{\sigma_{\hat{\beta}_0}^2} = \frac{\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2} \sim \frac{\chi_{(n-2)}^2}{n-2}$$

Par conséquent :

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \sim \frac{N(0, 1)}{\sqrt{\chi_{(n-2)}^2 / (n-2)}} = t_{(n-2)}$$

On trouve de même :

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim \frac{N(0, 1)}{\sqrt{\chi^2_{(n-2)}/(n-2)}} = t_{(n-2)}$$

2.5.2 Test de Fisher

Le test de Fisher permet de vérifier si le modèle de régression linéaire simple est globalement significatif, c'est-à-dire si la variable explicative X a une influence réelle sur la variable dépendante Y .

le test de Fisher est se basé sur la démarche suivante :

1. Énoncé des hypothèses :
 - $H_0 : \beta_1 = 0$ (aucune influence de X sur Y :)
 - $H_1 : \beta_1 \neq 0$
2. Calcul de la statistique F de Fisher :

$$F = \frac{CME}{CMR} = \frac{SCE/1}{SCR/(n-2)} = \frac{\frac{R^2}{1}}{\frac{1-R^2}{n-2}}$$

où :

Source de variation	Somme des carrés	Degrés de liberté	Carrés moyens
Expliquée	$SCE = \sum (\hat{y}_i - \bar{y})^2$	1	$CME = \frac{SCE}{1}$
Résiduelle	$SCR = \sum (y_i - \hat{y}_i)^2$	$n - 2$	$CMR = \frac{SCR}{n-2}$
Totale	$SCT = \sum (y_i - \bar{y})^2$	$n - 1$	-

TABLE 4.1 – Tableau d'analyse de la variance pour la régression linéaire simple

Selon le type de variation on remarque que :

- Pour la régression : 1 (car une seule variable explicative),
- Pour l'erreur (résiduelle) : $n - 2$ (dans le cas d'une régression linéaire simple).
- R^2 est le coefficient de détermination, mesurant la proportion de la variance totale de Y expliquée par le modèle.

3. Choix du risque α :

Généralement, le choix du risque α est fixé à 5%.

4. Lecture de la valeur critique dans la table de Fisher :

Elle est obtenue à partir de la table de Fisher avec $(1, n - 2)$ degrés de liberté.

- Par exemple, pour un échantillon de taille $n = 6$, on a $ddl = (1, 4)$, et la valeur critique est $F_{\text{critique}} = 7,71$.

5. Comparaison :

- Si $F_{\text{calculé}} > F_{\text{critique}}$, alors on rejette l'hypothèse nulle H_0 : le modèle est significatif à 95%, il existe une relation entre X et Y .
- Si la p -value (généralement fournie par les logiciels) est inférieure ou égale à α , on rejette également H_0 .

v2 \ v1	1	2	3	4	5	6
1	161.45	199.50	215.70	224.58	230.16	234.00
2	18.51	19.00	19.16	19.25	19.30	19.34
3	10.13	9.55	9.28	9.12	9.01	8.94
4	7.71	6.94	6.59	6.39	6.26	6.16
5	6.61	5.79	5.41	5.19	5.05	4.95
6	5.99	5.14	4.76	4.53	4.39	4.28
7	5.59	4.74	4.35	4.12	3.98	3.87
8	5.32	4.46	4.07	3.84	3.69	3.58

TABLE 2.1 – La table ANOVA à $\alpha = 5\%$

2.5.3 Test de Student

Le test de Student (ou test t) est utilisé pour évaluer la significativité de chaque coefficient individuellement. Plus particulièrement, pour la pente β_1 , le test t permet de déterminer si la variable explicative X a une influence linéaire significative sur la variable dépendante Y .

Généralement, pour un coefficient β_i (où i est 0 ou 1), le test de Student s'articule autour des hypothèses suivantes :

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases} \quad \text{pour } i = 0 \text{ ou } 1.$$

On rejete H_0 si

$$\left| \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \right| \geq t_{n-2, 1-\alpha/2}$$

avec $t_{n-2, 1-\alpha/2}$ est le quantile de la loi de Student à $n - 2$ degrés de liberté

– **Test sur β_1 :**

Sous l'hypothèse $H_0 : \beta_1 = 0$, on a :

$$T_n = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim \mathcal{T}(n - 2).$$

Pour une hypothèse alternative

$H_1 : \beta_1 \neq 0$ (test bilatéral), on rejette H_0 avec un risque $0 \leq \alpha \leq 1$ si :

$$|t| \geq t_{n-2, 1-\alpha/2} \quad \text{où } t \text{ est la réalisation de } T_n.$$

Dans ce cas, nous disons que la relation linéaire entre X et Y est significative au seuil α .

Si

$$|t| < t_{n-2, 1-\alpha/2}$$

dans ce cas Y ne dépend pas linéairement de X . Le modèle devient alors :

$$Y_i = \beta_0 + \epsilon_i$$

Le modèle proposé $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ est inadéquat. Nous testons alors un nouveau modèle.

– **Test sur β_0 :**

Si $\beta_0 = 0$, alors la relation entre les variables dépend uniquement de β_1 , et la droite de régression passe par l'origine.

n/a	90 %	80 %	70 %	60 %	50 %	40 %	30 %	20 %	10 %	5 %	2 %	1 %
1	0.1584	0.3249	0.5095	0.7265	1.0000	1.3764	1.9626	3.0777	6.3138	12.7062	31.8205	63.6567
2	0.1421	0.2887	0.4447	0.6172	0.8165	1.0607	1.3862	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.1366	0.2767	0.4242	0.5844	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.1338	0.2707	0.4142	0.5686	0.7407	0.9410	1.1896	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.1322	0.2672	0.4082	0.5594	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.1311	0.2648	0.4043	0.5534	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.1303	0.2632	0.4015	0.5491	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.1297	0.2619	0.3995	0.5459	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.1293	0.2610	0.3979	0.5435	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.1289	0.2602	0.3966	0.5415	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.1286	0.2596	0.3956	0.5399	0.6974	0.8755	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.1283	0.2590	0.3947	0.5386	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.1281	0.2586	0.3940	0.5375	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.1280	0.2582	0.3933	0.5366	0.6924	0.8681	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.1278	0.2579	0.3928	0.5357	0.6912	0.8662	1.0735	1.3406	1.7531	2.1314	2.6025	2.9467
16	0.1277	0.2576	0.3923	0.5350	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.1276	0.2573	0.3919	0.5344	0.6892	0.8633	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.1274	0.2571	0.3915	0.5338	0.6884	0.8620	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.1274	0.2569	0.3912	0.5333	0.6876	0.8610	1.0655	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.1273	0.2567	0.3909	0.5329	0.6870	0.8600	1.0640	1.3253	1.7247	2.0860	2.528	2.8453
21	0.1272	0.2566	0.3906	0.5325	0.6864	0.8591	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.1271	0.2564	0.3904	0.5321	0.6858	0.8583	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.1271	0.2563	0.3902	0.5317	0.6853	0.8575	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.1270	0.2562	0.3900	0.5314	0.6848	0.8569	1.0593	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.1269	0.2561	0.3898	0.5312	0.6844	0.8562	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.1269	0.2560	0.3896	0.5309	0.6840	0.8557	1.0575	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.1268	0.2559	0.3894	0.5306	0.6837	0.8551	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.1268	0.2558	0.3893	0.5304	0.6834	0.8546	1.0560	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.1268	0.2557	0.3892	0.5302	0.6830	0.8542	1.0553	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.1267	0.2556	0.3890	0.5300	0.6828	0.8538	1.0547	1.3104	1.6973	2.0423	2.4573	2.7500
35	0.1266	0.2553	0.3885	0.5292	0.6816	0.8520	1.0520	1.3062	1.6896	2.0301	2.4377	2.7238
40	0.1265	0.2550	0.3881	0.5286	0.6807	0.8507	1.0500	1.3031	1.6839	2.0211	2.4233	2.7045
45	0.1264	0.2549	0.3878	0.5281	0.6800	0.8497	1.0485	1.3006	1.6794	2.0141	2.4121	2.6896
50	0.1263	0.2547	0.3875	0.5278	0.6794	0.8489	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778
55	0.1262	0.2546	0.3873	0.5275	0.6790	0.8482	1.0463	1.2971	1.6730	2.0040	2.3961	2.6682
60	0.1262	0.2545	0.3872	0.5272	0.6786	0.8477	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603
70	0.1261	0.2543	0.3869	0.5268	0.6780	0.8468	1.0442	1.2938	1.6669	1.9944	2.3808	2.6479
80	0.1261	0.2542	0.3867	0.5265	0.6776	0.8461	1.0432	1.2922	1.6641	1.9901	2.3739	2.6387
90	0.1260	0.2541	0.3866	0.5263	0.6772	0.8456	1.0424	1.2910	1.6620	1.9867	2.3685	2.6316
100	0.1260	0.2540	0.3864	0.5261	0.6770	0.8452	1.0418	1.2901	1.6602	1.9840	2.3642	2.6259
150	0.1259	0.2538	0.3861	0.5255	0.6761	0.8440	1.0400	1.2872	1.6551	1.9759	2.3515	2.6090
200	0.1258	0.2537	0.3859	0.5252	0.6757	0.8434	1.0391	1.2858	1.6525	1.9719	2.3451	2.6006

FIGURE 2.4 – Table de la loi de Student

2.6 NumPy : bibliothèque de calcul scientifique

NumPy (Numerical Python) est une bibliothèque fondamentale pour le calcul scientifique en langage Python. Elle fournit des structures de données performantes, notamment les tableaux multidimensionnels (`ndarray`), ainsi qu'un large ensemble de fonctions vectorisées.

Dans le contexte de la régression linéaire simple, NumPy permet de représenter efficacement les séries de données statistiques (x_i, y_i) sous forme de vecteurs. Les opérations nécessaires à l'estimation des paramètres (moyennes, écarts, produits scalaires, sommes) sont directement implémentées dans cette bibliothèque.

Les principales fonctionnalités de NumPy pertinentes pour l'analyse de régression sont :

- La création de tableaux à partir de données statistiques.
- Les fonctions `mean()`, `sum()`, `std()` pour les calculs descriptifs de base.
- Les opérations arithmétiques vectorisées (addition, multiplication, soustraction) sans recours à des boucles explicites.
- La gestion des matrices et des produits matriciels via `dot()` ou l'opérateur `@`.

Ces fonctionnalités font de NumPy un outil de choix pour implémenter les méthodes des moindres carrés, qu'il s'agisse de la forme algébrique classique ou de la formulation matricielle $\hat{\beta} = (X^T X)^{-1} X^T y$.