Régression Linéaire Multiple

Bruno Pelletier

Outline

Modèle

Rappel : Modèle Linéaire Standard

On considère deux variables X et Y et l'on dispose d'observations (x_i, Y_i), $i = 1, \dots, n$. Ecriture analytique :

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \forall i = 1, \dots, n.$$

Hypothèses :

- Centrage : $\mathbb{E}[\varepsilon_i] = 0$,
- Homoscédasticité : $\mathrm{Var}(\epsilon_i) = \sigma^2, \quad \forall i$,
- Décorrélation : $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0, \quad \forall i \neq j$.

Ecriture matricielle :

$$\mathbf{Y} = \mathbb{X}\beta + \epsilon,$$

avec

$$\mathbb{E}[\epsilon] = 0 \quad \text{et} \quad \mathbb{V}(\epsilon) = \sigma^2 \mathbf{I}_n.$$

Contexte multivarié

- Une variable réponse Y,
- Variables explicatives : $X_1, \ldots, X_{p-1}$.
- On inclut généralement une variable constante : $X_0 = 1$.
- Observations : (x_i, Y_i), pour $i = 1, \ldots, n$ avec

$$x_i = (x_{i,0}, \ldots, x_{i,p-1}) \qquad \text{(vecteur ligne).}$$

- Modèle (écriture analytique) :

$$Y_i = \beta_0 x_{i,0} + \beta_1 x_{i,1} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i,$$

sous les hypothèses standard :

(i) $\mathbb{E}[\epsilon_i] = 0 \ \forall i$,

(ii) $\mathbb{V}(\epsilon_i) = \sigma^2 \ \forall i$,

(iii) $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0 \ \forall i \neq j$.

Ecriture matricielle du modèle

On pose :

$$\begin{array}{rcll}
\mathbf{Y} & = & (Y_1, \ldots, Y_n)', & \text{(vecteur aléatoire de } \mathbb{R}^n\text{)} \\
\epsilon & = & (\epsilon_1, \ldots, \epsilon_n)', & \text{(vecteur aléatoire de } \mathbb{R}^n\text{)} \\
\beta & = & (\beta_0, \ldots, \beta_{p-1})', & \text{(vecteur de paramètres de } \mathbb{R}^p\text{)} \\
\mathbb{X} & = & \begin{bmatrix} x_{1,0} & \ldots & x_{1,p-1} \\ \vdots & \vdots & \vdots \\ x_{n,0} & \ldots & x_{n,p-1} \end{bmatrix} & \text{(matrice de taille } n \times p\text{)}
\end{array}$$

Modèle (écriture matricielle) :

$$\mathbf{Y} = \mathbb{X}\beta + \epsilon,$$

sous les hypothèses standard :

$$\mathbb{E}[\epsilon] = 0 \quad \text{et} \quad \mathbb{V}(\epsilon) = \sigma^2 \mathbf{I}_n.$$

Estimation

Estimation par Moindres Carrés Ordinaires

Le critère des moindres carrés s'écrit :

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} \left(Y_i - (\beta_0 x_{i,0} + \beta_1 x_{i,1} + \cdots + \beta_{p-1} x_{i,p-1})\right)^2 .$$

En notations matricielles :

$$\mathcal{L}(\beta) = \|\mathbf{Y} - \mathbb{X}\beta\|^2.$$

On cherche à construire un estimateur $\hat{\beta}$ en minimisant $\mathcal{L}(\beta)$ par rapport à β sur $\mathbb{R}^p$.

- $\mathcal{L}$ admet-elle un minimum sur $\mathbb{R}^p$? Si oui, est-il unique ?

Différentiation de $\mathcal{L}(\beta)$

On a

$$\begin{aligned}
\mathcal{L}(\beta) &= \|\mathbf{Y} - \mathbb{X}\beta\|^2 \\
&= \|\mathbf{Y}\|^2 - 2\langle \mathbf{Y}, \mathbb{X}\beta \rangle + \|\mathbb{X}\beta\|^2 \\
&= \|\mathbf{Y}\|^2 - 2\mathbf{Y}'\mathbb{X}\beta + (\mathbb{X}\beta)'(\mathbb{X}\beta) \\
&= \|\mathbf{Y}\|^2 - 2\mathbf{Y}'\mathbb{X}\beta + \beta'\mathbb{X}'\mathbb{X}\beta.
\end{aligned}$$

D'où l'expression du gradient de $\mathcal{L}$:

$$\nabla\mathcal{L}(\beta) = 2\mathbb{X}'\mathbb{X}\beta - 2\mathbb{X}'\mathbf{Y}.$$

Remarque : différentiation d'une forme linéaire et d'une forme quadratique.

Point critique de $\mathcal{L}(\beta)$

$$\nabla \mathcal{L}(\beta) = 0 \iff \underbrace{\mathbb{X}'\mathbb{X}\beta = \mathbb{X}'\mathbf{Y}}_{\text{Equations normales}} .$$

$\mathbb{X}'\mathbb{X}$ est une matrice symétrique de taille $p \times p$, et si elle est inversible, on obtient l'estimateur (unique) des MCO de β :

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

Remarques :

- $(\mathbb{X}'\mathbb{X})$ est inversible si $\mathbb{X}$ est de rang plein (i.e., p; on suppose $n \geq p$).
- On vérifie qu'il s'agit bien d'un minimum.

Valeurs ajustées

Pour la $i^{\text{ème}}$ observation, la valeur prédite par le modèle est :

$$\hat{Y}_i = \hat{\beta}_0 x_{i,0} + \hat{\beta}_1 x_{i,1} + \cdots + \hat{\beta}_{p-1} x_{i,p-1}.$$

On introduit le vecteur des valeurs ajustées :

$$\hat{\mathbf{Y}} = \left(\hat{Y}_1, \dots, \hat{Y}_n\right)' = \mathbb{X}\hat{\beta}.$$

D'où :

$$\begin{aligned} \hat{\mathbf{Y}} &= \mathbb{X}\left(\mathbb{X}'\mathbb{X}\right)^{-1}\mathbb{X}'\mathbf{Y} \\ &= \Pi_{\mathrm{Im}(\mathbb{X})}\mathbf{Y}, \end{aligned}$$

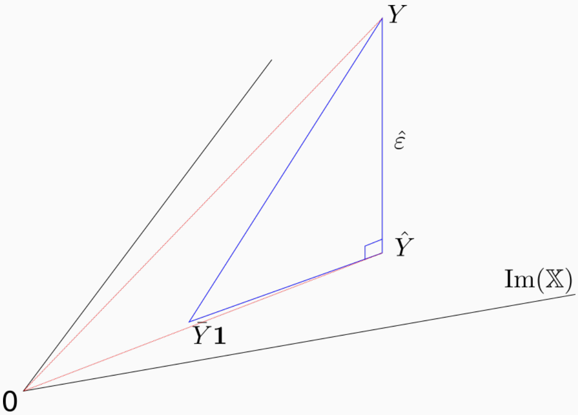
où

$$\Pi_{\mathrm{Im}(\mathbb{X})} = \mathbb{X}\left(\mathbb{X}'\mathbb{X}\right)^{-1}\mathbb{X}'$$

est la matrice de projection sur $\mathrm{Im}(\mathbb{X})$ dans $\mathbb{R}^n$.

Résidus

Vecteur des résidus : $\hat{\epsilon} = (\epsilon_1, \dots, \epsilon_n)' = \mathbf{Y} - \hat{\mathbf{Y}}$.



$$
\begin{aligned}
\|\mathbf{Y}\|^2 &= \|\hat{\mathbf{Y}}\|^2 + \|\hat{\epsilon}\|^2 && \text{(toujours vraie)} \\
\|\mathbf{Y} - \bar{Y}\mathbf{1}\|^2 &= \|\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}\|^2 + \|\hat{\epsilon}\|^2 && \text{(lorsque modèle avec constante)}
\end{aligned}
$$

Coefficient de détermination R^2

$$
\begin{aligned}
R^2 &= \frac{\mathrm{SCE}}{\mathrm{SCT}} = 1 - \frac{\mathrm{SCR}}{\mathrm{SCT}} \\
&= \begin{cases} \frac{\|\hat{\mathbf{Y}}\|^2}{\|\mathbf{Y}\|^2} & \text{si modèle sans constante} \\ \frac{\|\hat{\mathbf{Y}}-\bar{Y}\mathbf{1}\|^2}{\|\mathbf{Y}-\bar{Y}\mathbf{1}\|^2} & \text{si modèle avec constante} \end{cases}
\end{aligned}
$$

Le R^2 ne tient pas compte du nombre de paramètres du modèle. Coefficient de détermination ajusté R_a^2 :

$$
R_a^2 = \begin{cases} 1 - \frac{n}{n-p}\frac{\mathrm{SCR}}{\mathrm{SCT}} = 1 - \frac{n}{n-p}(1-R^2) & \text{modèle sans constante} \\ 1 - \frac{n-1}{n-p}\frac{\mathrm{SCR}}{\mathrm{SCT}} = 1 - \frac{n-1}{n-p}(1-R^2) & \text{modèle avec constante} \end{cases}
$$

Remarque : Logiciel R : ”R-squared” et ”Multiple R-squared”.

Propriétés des MCO

Espérance de $\hat{\beta}$

$$\mathbb{E}\left[\hat{\beta}\right] = \mathbb{E}[\underbrace{(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'}_{\text{déterministe}}\mathbf{Y}] = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{E}[\mathbf{Y}].$$

$$\mathbb{E}[\mathbf{Y}] = \mathbb{E}[\mathbb{X}\beta + \epsilon] = \mathbb{X}\beta + \mathbb{E}[\epsilon] = \mathbb{X}\beta.$$

D'où :

$$\mathbb{E}[\hat{\beta}] = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{X}\beta = \beta.$$

Conclusion : $\hat{\beta}$ est un estimateur sans biais de β.

Matrice de covariance de $\hat{\beta}$

Rappel : Covariance
Pour $\mathbf{A}$ une matrice déterministe et $\mathbf{Y}$ un vecteur aléaoire :

$$\begin{aligned} \mathbb{V}(\mathbf{Y}) &= \mathbb{E}[\mathbf{Y}\mathbf{Y}'] - \mathbb{E}[\mathbf{Y}]\mathbb{E}[\mathbf{Y}]' \\ \mathbb{V}(\mathbf{A}\mathbf{Y}) &= \mathbf{A}\,\mathbb{V}(\mathbf{Y})\mathbf{A}'. \end{aligned}$$

$$\mathbb{V}\left(\hat{\beta}\right) = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\,\mathbb{V}(\mathbf{Y})\left[(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\right]'.$$

Or : $\quad \mathbb{V}(\mathbf{Y}) = \mathbb{V}(\mathbb{X}\beta + \epsilon) = \mathbb{V}(\epsilon) = \sigma^2\mathbf{I}_n,$

et : $\quad \sigma^2(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\left[(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\right]' = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{X}\left[(\mathbb{X}'\mathbb{X})^{-1}\right]' = (\mathbb{X}'\mathbb{X})^{-1}.$

D'où : $\quad \mathbb{V}\left(\hat{\beta}\right) = \sigma^2(\mathbb{X}'\mathbb{X})^{-1}$

Prévision

Prévision

- Modèle : $Y_i = \beta_0 x_{i,0} + \beta_1 x_{i,1} + \cdots + \beta_{i,p-1} x_{i,p-1} + \epsilon_i$.
- Nouvelle donnée : $x_{n+1} = (x_{n+1,0}, \ldots, x_{n+1,p-1})$, observée.
- Modèle étendu :
  $Y_{n+1} = \beta_0 x_{n+1,0} + \cdots + \beta_{p-1} x_{n+1,p-1} + \epsilon_{n+1}$, (non observée).
- Objectif : prédire la valeur de Y_{n+1} à partir de x_{n+1}.
- Hypothèses : $\mathbb{E}\left[\epsilon_{n+1}\right] = 0$, $\mathbb{V}\left(\epsilon_{n+1}\right) = \sigma^2$, $\mathrm{Cov}\left(\epsilon_i, \epsilon_{n+1}\right) = 0$ $\forall\ i = 1, \ldots, n$.
- Valeur prédite :

$$\hat{Y}_{n+1}^{(p)} = x_{n+1}\beta$$

- Erreur de prévision :

$$\hat{\epsilon}_{n+1}^{(p)} = Y_{n+1} - \hat{Y}_{n+1}^{(p)}.$$

Propriétés de l'erreur de prévision

- $\mathbb{E}\left[\widehat{\epsilon}_{n+1}^{(p)}\right] = \mathbb{E}\left[Y_{n+1}\right] - \mathbb{E}\left[\widehat{Y}_{n+1}^{(p)}\right] = x_{n+1}\beta - x_{n+1}\underbrace{\mathbb{E}\left[\widehat{\beta}\right]}_{=\beta} = 0.$

- $\mathbb{V}\left(\widehat{\epsilon}_{n+1}^{(p)}\right) = \mathbb{V}\left(Y_{n+1}\right) + \mathbb{V}\left(\widehat{Y}_{n+1}^{(p)}\right) - 2\underbrace{\mathrm{Cov}\left(Y_{n+1}, \widehat{Y}_{n+1}^{(p)}\right)}_{=0}$

- En effet : $\widehat{Y}_{n+1}^{(p)} \in \sigma\{\epsilon_1, \dots, \epsilon_n\}$ tandis que $Y_{n+1} \in \sigma\{\epsilon_{n+1}\}$.

- $$\begin{aligned} \mathbb{V}\left(\widehat{\epsilon}_{n+1}^{(p)}\right) &= \sigma^2 + \mathbb{V}\left(x_{n+1}\widehat{\beta}\right) = \sigma^2 + x_{n+1}\,\mathbb{V}\left(\widehat{\beta}\right)x_{n+1}' \\ &= \sigma^2 + x_{n+1}\sigma^2(\mathbb{X}'\mathbb{X})^{-1}x_{n+1}' \\ &= \sigma^2\left[1 + x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x_{n+1}'\right]. \end{aligned}$$

Identifiabilité

Identifiabilité

- Modèle : $\mathbf{Y} = \mathbb{X}\beta + \epsilon$.
- MCO : $\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbf{Y}$.
- Est-il toujours possible d'estimer β ?
- Exemple : Soit la matrice de design suivante

$$\mathbb{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$\mathbb{X} \sim \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

$\mathbb{X}$ est de rang 2, donc $\mathbb{X}'\mathbb{X}$ n'est pas inversible.

Equations normales

- Critère des MCO : $\mathcal{L}(\beta) = \|\mathbf{Y} - \mathbb{X}\beta\|^2$.
- $\nabla\mathcal{L}(\beta) = 0 \iff \mathbb{X}'\mathbb{X}\beta = \mathbb{X}'\mathbf{Y}$.
- Lorsque $\mathbb{X}'\mathbb{X}$ n'est pas inversible les équations normales admettent une infinité de solutions (un s.e.v.).
- Le modèle est dit identifiable si $\mathbb{X}\alpha_1 = \mathbb{X}\alpha_2 \implies \alpha_1 = \alpha_2$ $\forall \alpha_1, \alpha_2 \in \mathbb{R}^p$.
- Néanmoins, peut-on estimer certaines composantes de β? Ou bien certaines combinaisons linéaires de β?

Il est possbile d'estimer $\langle a, \beta\rangle$ si cette combinaison est identifiable, i.e., si $\mathbb{X}\alpha_1 = \mathbb{X}\alpha_2 \implies \langle a, \alpha_1\rangle = \langle a, \alpha_2\rangle$, $\forall \alpha_1, \alpha_2 \in \mathbb{R}^p$.

Identifiabilité d'une combinaison linéaire $\langle a, \beta \rangle$

$$[\mathbb{X}\alpha_1 = \mathbb{X}\alpha_2 \implies \langle a, \alpha_1 \rangle = \langle a, \alpha_2 \rangle, \quad \forall \alpha_1, \alpha_2 \in \mathbb{R}^p]$$

$$\iff [\mathbb{X}(\alpha_1 - \alpha_2) = 0 \implies \langle a, \alpha_1 - \alpha_2 \rangle = 0, \quad \forall \alpha_1, \alpha_2 \in \mathbb{R}^p]$$

$$\iff [\mathbb{X}u = 0 \implies \langle a, u \rangle = 0, \quad \forall u \in \mathbb{R}^p]$$

$$\iff [\langle a, u \rangle = 0, \quad \forall u \in \mathrm{Ker}(\mathbb{X})]$$

$$\iff a \in \mathrm{Ker}(\mathbb{X})^\perp$$

$$\iff a \in \mathrm{Row}(\mathbb{X}).$$

La combinaison linéaire $\langle a, \beta \rangle$ est identifiable si a appartient à l'espace engendré par les lignes de $\mathbb{X}$.

Exemple

$$\mathbb{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Donc

$$\mathrm{Row}(\mathbb{X}) = \mathrm{Vect}\left\{[1\ 1\ 0], [0\ 1\ {-1}]\right\}.$$

- Par exemple, il est possible d'estimer $\beta_0 + \beta_1$ car

$$\beta_0 + \beta_1 = \langle \beta, [1\ 1\ 0]' \rangle.$$

- Est-il possible d'estimer β_0 ?

De manière équivalente : $[1\ 0\ 0]$ appartient-il à $\mathrm{Row}(\mathbb{X})$? ...Non.

Exercice

Exercice 1 (Deux variables explicatives)

On souhaite expliquer une variable réponse y à partir de deux variables explicatives x et z. Soit $\mathbb{X} = [\mathbf{1}\; x\; z]$ la matrice $n \times 3$ du plan d'expériences qui satisfait les relations suivantes :

$$\mathbb{X}'\mathbb{X} = \begin{pmatrix} 25 & 0 & 0 \\ ? & 9.3 & 5.4 \\ ? & ? & 12.7 \end{pmatrix} \quad (\mathbb{X}'\mathbb{X})^{-1} = \begin{pmatrix} 0.04 & 0 & 0 \\ 0 & 0.1428 & -0.0607 \\ 0 & -0.0607 & 0.1046 \end{pmatrix}.$$

1. Donner les valeurs manquantes.
2. Que vaut n ?
3. Calculer le coefficient de corrélation empirique entre x et z.
4. L'ajustement linéaire donne

$$y = -1.61\mathbf{1} + 0.61x + 0.46z + \hat{\epsilon}, \quad SCR = \|\hat{\epsilon}\|^2 = 0.3.$$

- Déterminer la valeur de la moyenne empirique $\bar{y}$.
- Calculer la somme des carrés expliqués (SCE), la somme des carrés totale (SCT), le coefficient de détermination et le coefficient de détermination ajusté.