Regression Logistique

PINARD Paul

24 septembre 2025

Résumé

Ce document a pour but d'expliquer la régression logistique et ses caractéristiques d'analyses. Des exemples seront fournis tout au long du document. La section 'Exemple Concret' montre l'ensemble des calculs pour obtenir les différents résultats des sorties des logiciels de Statistique (R,Python notamment).

1 Exemple Concret

Pour cet exemple, nous prendrons une base de données de 8 individus et 3 variables :

- Le sexe de l'individu, qui sera notre variable dépendante Y.
- La tendance de l'individu à boire de l'alcool (codé en 3 modalités).
- La pratique sportive de l'indivudu (codé en Oui/Non).

ID	SEXE	PRAT SPORT	ALCOOL
1	Homme	Oui	Jamais
2	Femme	Oui	Jamais
3	Femme	Non	Jamais
4	Homme	Non	Hebdomadaire
5	Homme	Non	Journalier
6	Homme	Oui	Jamais
7	Femme	Non	Hebdomadaire
8	Femme	Oui	Hebdomadaire

Table 1 – Jeu de données jouet pour notre exemple.

1.1 1$^{\text{ère}}$ étape : Initialisation

Notre base de données ($\mathbb{X}$) sera présentée sous une matrice des constantes. Des variables 'dummys' seront créées par la même occasion. Les modalités qui seront mises en place en tant que référente sont :

- La modalité "Oui" pour PRAT SPORT
- La modalité "Journalier" pour ALCOOL
- La modalité "Femme" pour SEXE

On obtient donc une matrice avec 5 colonnes (la première correspond à l'intercept, la seconde à la colonne SEXE = NON, la troisième pour PRAT SPORT = NON, et la quatrième et cinquième colonne pour les modalités ALCOOL).

On cherche à résoudre l'équation : $\beta^{i+1} = \beta^i + (\mathbb{X}^T W \mathbb{X})^{-1}\mathbb{X}^T(y-\pi)$ où :

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

- $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$
- La probabilité $\pi_i = \frac{1}{1+e^{x_i^T * \beta}}$. On transpose x_i car impossible de multiplier une matrice $(j * 1)$ avec une autre matrice $(j * 1)$.
- La matrice W est une matrice diagonale avec $\pi_i(1 - \pi_i)$

Comme nous avons pu en parler plus haut, la création des estimateurs se fait par itération. C'est un choix que l'utilisateur fait pour définir la première initialisation.Elle se fait souvent avec une matrice β nulle. On aura donc par la suite $\beta^{(0)T} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$.

Donc :

- $x_i^T * \beta = 0 \Rightarrow \pi_i = \frac{1}{1+e^0} = 0.5$
- $\pi^T = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix}$
- $W = diag(0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25)$ car $\pi_i(1 - \pi_i) = 0.5 * (1 - 0.5) = 0.25$

1.2 2nd étape : Calcul mise à jour

- Résidus $y - \pi$:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -0.5 \\ -0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ -0.5 \\ -0.5 \end{bmatrix}$$

- $X^T W X$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0.5 & 0.5 & 0.25 \\ 1 & 0.5 & 1 & 0.25 & 0.5 \\ 1 & 0.5 & 0.25 & 1 & 0 \\ 0.75 & 0.25 & 0.5 & 0 & 0.75 \end{bmatrix}$$

- $X^T(y - \pi)$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} * \begin{bmatrix} 0.5 \\ -0.5 \\ -0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ -0.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \\ -0.5 \end{bmatrix}$$

- Puis la transposée de $X^T W T$:

$$(\begin{bmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0.5 & 0.5 & 0.25 \\ 1 & 0.5 & 1 & 0.25 & 0.5 \\ 1 & 0.5 & 0.25 & 1 & 0 \\ 0.75 & 0.25 & 0.5 & 0 & 0.75 \end{bmatrix})^T = \begin{bmatrix} 2 & 1 & 1 & 1 & 0.75 \\ 1 & 1 & 0.5 & 0.5 & 0.25 \\ 1 & 0.5 & 1 & 0.25 & 0.5 \\ 1 & 0.5 & 0.25 & 1 & 0 \\ 1 & 0.25 & 0.5 & 0 & 0.75 \end{bmatrix}$$

1.3 3$^{\text{ème}}$ étape : Calcul des estimateurs β^1

$$\beta^1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 & 1 & 1 & 1 & 0.75 \\ 1 & 1 & 0.5 & 0.5 & 0.25 \\ 1 & 0.5 & 1 & 0.25 & 0.5 \\ 1 & 0.5 & 0.25 & 1 & 0 \\ 1 & 0.25 & 0.5 & 0 & 0.75 \end{bmatrix} * \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 1.625 \\ 1.875 \\ 0.75 \\ 1 \\ 0.125 \end{bmatrix}$$

On repart à l'étape 1 avec cette fois-ci β^1 à la place de β^0. Les itérations continuent jusqu'à ce que la différence entre β^{i+1} ne soit plus significative avec β^i. Le calcul des p-values de chaque variable/modalité se réalise grâce aux estimateurs qui maximisent $l(\beta)$.

On calcule la statistique de Wald :

- $SE(\hat{\beta}_j) = \sqrt{(X^TWX)^{-1}}$
- $z_j = \frac{\hat{b_j}}{SE(\hat{b_j})}$
- p-value= $2 * \mathbb{P}(Z > |z_j|)$