Cours et exemples sur la régression linéaire multiple

Introduction Ce cours consiste à expliquer comment établir la droite de régression (un ajustement affine) par la méthode de régression multiple. La régression linéaire multiple est identique à la régression linéaire simple sauf que dans cette régression, on va expliquer la variable à expliquer par plusieurs variables explicatives. Pour bien comprendre la méthode, on va procéder directement par un exemple explicatif.

Exemple 1

Dans cet exemple, on va prédire la déviation de la conduite sur un parcours (en cm) sur la base du taux d'alcool (en g par litre de sang) et la vitesse (en Km/H).

Premier objectif de cet exemple est de prédire la variable expliquée.

Deuxième objectif de l'exemple est d'évaluer l'importance relative de chaque variable explicative.

Les données sont :

Déviation	Taux d'alcool	Vitesse
61	0,0099	99
58	0,01302	93
67	0,03094	91
41	0,0242	110
44	0,00819	91
52	0,00636	106

Question : Estimer les paramètres de ce modèle.

Réponse : On commence par mettre la formule générale qui est la suivante :

$$y = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \hat{\alpha}_2 x_2 + \epsilon$$

Donc on doit trouver les valeurs du vecteur $\hat{\alpha} = \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix}$.

Premièrement, on pose $\epsilon = 0$, on considère aussi que x_1 représente les valeurs du taux d'alcool et x_2 représente les valeurs de la vitesse.

$$\begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = (x'x)^{-1} \times (x'y) \text{ et le } n = \text{au nombre d'observations qui est égale à } 6 \text{ (n = 6)}.$$

La matrice x est calculée comme suit :

$$x = \begin{pmatrix} 1 & 0,0099 & 99 \\ 1 & 0,01302 & 93 \\ 1 & 0,03094 & 91 \\ 1 & 0,0242 & 110 \\ 1 & 0,00819 & 91 \\ 1 & 0,00636 & 106 \end{pmatrix}.$$

La matrice transposée de x notée x' est donc calculée comme suit :

On inverse les colonnes et les lignes.

$$x' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0,0099 & 0,01302 & 0,03094 & 0,0242 & 0,00819 & 0,00636 \\ 99 & 93 & 91 & 110 & 91 & 106 \end{pmatrix}.$$

$$x'x = \begin{pmatrix} 1 & 0,0099 & 99 \\ 1 & 0,01302 & 93 \\ 1 & 0,03094 & 91 \\ 1 & 0,0242 & 110 \\ 1 & 0,00819 & 91 \\ 1 & 0,00636 & 106 \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0,0099 & 0,01302 & 0,03094 & 0,0242 & 0,00819 & 0,00636 \\ 99 & 93 & 91 & 110 & 91 & 106 \end{pmatrix}.$$

$$x'x = \begin{pmatrix} 6 & 0,09261 & 590 \\ 0,09261 & 0,001917987 & 9,08795 \\ 590 & 9,08795 & 58348 \end{pmatrix}.$$

On remarque que la matrice $x'x$ est une matrice symétrique.

La matrice $(x'x)^{-1} = \frac{1}{\det(x'x)} \times Co(x'x)'$, sachant que $\det(x'x)$ est le déterminant de la matrice $x'x$.

Pour calculer le déterminant de la matrice $x'x$ il faut mettre des signes (+ et -) sur les lignes de cette matrice de la manière suivante :

$$\begin{pmatrix} + \\ - \\ + \end{pmatrix} \begin{pmatrix} 6 & 0,09261 & 590 \\ 0,09261 & 0,001917987 & 9,08795 \\ 590 & 9,08795 & 58348 \end{pmatrix}.$$

Puis on prend les éléments de la première colonne un par un et à chaque fois on barre la ligne et la colonne correspondante (dont l'élément barré forme l'intersection) de la manière suivante :

$$\det(x'x) = +6 \times \begin{vmatrix} 0,001917987 & 9,08795 \\ 9,08795 & 58348 \end{vmatrix} - 0,09261 \times \begin{vmatrix} 0,09261 & 9,08795 \\ 590 & 58348 \end{vmatrix} + 590 \times \begin{vmatrix} 0,09261 & 0,001917987 \\ 590 & 9,08795 \end{vmatrix}$$

$$\Rightarrow \det(x'x) = 175,573 - 3,8634 - 170,7430 = 0,969128.$$

$$Co(x'x) = \begin{vmatrix} + \begin{vmatrix} 0,001917987 & 9,08795 \\ 9,08795 & 58348 \end{vmatrix} & - \begin{vmatrix} 0,09261 & 9,08795 \\ 590 & 58348 \end{vmatrix} & + \begin{vmatrix} 0,09261 & 0,001917987 \\ 590 & 9,08795 \end{vmatrix} \\ - \begin{vmatrix} 0,09261 & 590 \\ 9,08795 & 58348 \end{vmatrix} & + \begin{vmatrix} 6 & 590 \\ 590 & 58348 \end{vmatrix} & - \begin{vmatrix} 6 & 0,09261 \\ 590 & 9,08795 \end{vmatrix} \\ + \begin{vmatrix} 0,09261 & 590 \\ 0,001917987 & 9,08795 \end{vmatrix} & - \begin{vmatrix} 6 & 590 \\ 0,09261 & 9,08795 \end{vmatrix} & + \begin{vmatrix} 6 & 0,09261 \\ 0,09261 & 0,001917987 \end{vmatrix} \end{vmatrix}$$

$$= \begin{pmatrix} 29,2622 & -41,2547 & -0,2893 \\ -41,2547 & 1998 & 0,1122 \\ -0,2893 & 0,1122 & 0,00292 \end{pmatrix}$$

Remarque : La matrice $Co(x'x)$ est équivalente à la matrice $Co(x'x)'$.

La matrice $(x'x)^{-1} = \frac{1}{0,969128} \times Co(x'x)'$

$$\Rightarrow (x'x)^{-1} = \begin{pmatrix} 30,2534273 & -43,0467171 & -0,299210182 \\ -43,0467171 & 2051,3285582 & 0,115774177 \\ -0,2992102 & 0,1157742 & 0,003024643 \end{pmatrix}$$

$$\text{La matrice } (x'y) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0,0099 & 0,01302 & 0,03094 & 0,0242 & 0,00819 & 0,00636 \\ 99 & 93 & 91 & 110 & 91 & 106 \end{pmatrix} \times \begin{pmatrix} 61 \\ 58 \\ 67 \\ 41 \\ 44 \\ 52 \end{pmatrix}$$

$$= \begin{pmatrix} 323 \\ 5,11532 \\ 31556 \end{pmatrix}$$

Donc

$$\begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \frac{1}{0,969128} \times \begin{pmatrix} 29,2622 & -41,2547 & -0,2893 \\ -41,2547 & 1998 & 0,1122 \\ -0,2893 & 0,1122 & 0,00292 \end{pmatrix} \times \begin{pmatrix} 323 \\ 5,11532 \\ 31556 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} 109,782785 \\ 242,482328 \\ -0,607039 \end{pmatrix}$$

$\hat{y} = 109,782785 + 242,482328x_1 - 0,607039x_2$ qui représente la droite de la régression multiple.

A partir de cette formule, on peut calculer les y estimées notées $\hat{y}_i$ comme suit :

$$\hat{y}_1 = 109,782785 + 242,482328 \times 0,0099 - 0,607039 \times 99 = 52,08650$$

$$\hat{y}_2 = 109,782785 + 242,482328 \times 0,01302 - 0,607039 \times 93 = 56,48528$$

$$\hat{y}_3 = 109,782785 + 242,482328 \times 0,03094 - 0,607039 \times 91 = 62,04464$$

$$\hat{y}_4 = 109,782785 + 242,482328 \times 0,0242 - 0,607039 \times 110 = 48,87657$$

$$\hat{y}_5 = 109,782785 + 242,482328 \times 0,00819 - 0,607039 \times 91 = 56,52817$$

$$\hat{y}_6 = 109,782785 + 242,482328 \times 0,00636 - 0,607039 \times 106 = 46,97884$$

Exemple 2

On suppose qu'on a le tableau suivant :

Année	y	x_1	x_2
2000	4	1	6
2001	4	3	6
2002	8	4	7
2003	12	6	7
2004	12	6	9

Question : Estimer les paramètres de ce modèle.

Réponse : On commence par mettre la formule générale qui est la suivante :

$$y = \hat{\alpha}_0 + \hat{\alpha}_1x_1 + \hat{\alpha}_2x_2 + \epsilon$$

Les paramètres du modèle $\begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = (x'x)^{-1} \times (x'y)$ et le n = au nombre d'observations qui est égale à 5 (n = 5).

$$x = \begin{pmatrix} 1 & 1 & 6 \\ 1 & 3 & 6 \\ 1 & 4 & 7 \\ 1 & 6 & 7 \\ 1 & 6 & 9 \end{pmatrix}.$$

La matrice transposée de x notée x' est donc calculée comme suit :
 On inverse les colonnes et les lignes.

$$x' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 6 \\ 6 & 6 & 7 & 7 & 9 \end{pmatrix}.$$

$$(x'x) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 6 \\ 6 & 6 & 7 & 7 & 9 \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 6 \\ 1 & 3 & 6 \\ 1 & 4 & 7 \\ 1 & 6 & 7 \\ 1 & 6 & 9 \end{pmatrix}$$

$$\Rightarrow (x'x) = \begin{pmatrix} 5 & 20 & 35 \\ 20 & 98 & 148 \\ 35 & 148 & 251 \end{pmatrix}$$

$$(x'x)^{-1} = \frac{1}{220} \times \begin{pmatrix} 2694 & 160 & -470 \\ 160 & 30 & -40 \\ -470 & -40 & 90 \end{pmatrix}.$$

$$x'y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 6 \\ 6 & 6 & 7 & 7 & 9 \end{pmatrix} \times \begin{pmatrix} 4 \\ 4 \\ 8 \\ 12 \\ 12 \end{pmatrix}$$

$$= \begin{pmatrix} 40 \\ 192 \\ 296 \end{pmatrix}.$$

$$\begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \frac{1}{220} \times \begin{pmatrix} 2694 & 160 & -470 \\ 160 & 30 & -40 \\ -470 & -40 & 90 \end{pmatrix} \times \begin{pmatrix} 40 \\ 192 \\ 296 \end{pmatrix}$$

$$= \frac{1}{220} \times \begin{pmatrix} -640 \\ 310 \\ 160 \end{pmatrix}.$$

$$\Rightarrow \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} -2,91 \\ 1,455 \\ 0,727 \end{pmatrix}.$$

$$\hat{y} = -2,91 + 1,455x_1 + 0,727x_2.$$

A partir de cette formule, on peut calculer les y estimées notées $\hat{y}_i$ comme suit :

$$\hat{y}_1 = -2,91 + 1,455 \times 1 + 0,727 \times 6 = 2,90$$

$$\hat{y}_2 = -2,91 + 1,455 \times 3 + 0,727 \times 6 = 5,81$$

$$\hat{y}_3 = -2,91 + 1,455 \times 4 + 0,727 \times 7 = 8,00$$

$$\hat{y}_4 = -2,91 + 1,455 \times 6 + 0,727 \times 7 = 10,90$$

$$\hat{y}_5 = -2,91 + 1,455 \times 6 + 0,727 \times 9 = 12,36$$