Chapitre 2 : **REGRESSION LINÉAIRE MULTIPLE**

Plan du Chapitre :

I.- Position du problème

II.- Hypothèses d'application de la Méthode de moindres carrées

III.- Estimation des composants du vecteur A

IV.- Espérance mathématique et matrices des variances covariances des estimateurs

…à suivre…

I.- Position du problème :

Le modèle de régression linéaire multiple est de la forme :

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_k X_k + \varepsilon ,$$

où Y est la variable à expliquer,.

$X_1, X_2, \ldots, X_k$ ce sont les variables explicatives.

$a_1, a_2, \ldots, a_k$ ce sont les paramètres du modèle.

ε est l'erreur aléatoire, inconnue et centrée.

Les variables Y, $X_1, X_2, \ldots, X_k$ et ε sont des vecteurs de n composantes.

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}; \qquad X_1 = \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{i1} \\ \vdots \\ x_{n1} \end{pmatrix}; \cdots\cdots\cdots; \quad X_k = \begin{pmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{ik} \\ \vdots \\ x_{nk} \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Le vecteur aléatoire Y est connu. Les vecteurs $X_1, X_2, \ldots, X_k$ sont connus et non aléatoires.

Résoudre le problème consiste en estimer les paramètres $a_1, a_2, \ldots, a_k$ qui sont inconnus. Notons par $\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_k$ leurs estimateurs. Le modèle s'écrit, alors :

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = a_1 \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{i1} \\ \vdots \\ x_{n1} \end{pmatrix} + \cdots\cdots + a_k \begin{pmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{ik} \\ \vdots \\ x_{nk} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

ou encore sous forme matricielle,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Soit finalement, en posant,

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \quad \text{et} \quad A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}$$

sous forme d'une équation matricielle :

$$\boxed{Y = XA + \varepsilon}$$

Matrices de types $(n,1)=(n,k)(k,1)+(n,1)$

Remarquons qu'un modèle écrit sous forme non homogène, c'est-à-dire avec un terme constant :

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_k X_k + \varepsilon$$

Revient au cas général précédent en supposant que la constante a_0 est multipliée par un vecteur unitaire X_0 :

$$Y = a_0 X_0 + a_1 X_1 + a_2 X_2 + \cdots + a_k X_k + \varepsilon$$

La matrice X admet, alors, une colonne complètement des « 1 » :

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \quad \text{et} \quad A = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{pmatrix}$$

Par la suite, nous allons considérer le cas homogène et nous allons estimer les termes du vecteur A en appliquant la méthode des moindres carrées.

II.- Hypothèses d'application de la méthode des moindres carrées :

On suppose que :

1°) Le nombre des composantes (n) est plus grand que le nombre des variables explicatives (k) : $n>k$.

Les variables explicatives sont connues et non-aléatoires. En d'autres termes, commençant de nouveau par l'échantillonnage, l'unique source de variation pour Y provient de l'erreur aléatoire ε.

2°) La matrice X de type (n,k) est de rang k, c'est-à-dire que aucune variable explicative n'est linéairement dépendante des autres.

Dans le cas où le rang de X est inférieur à k, la procédure d'estimation n'est plus valable. En effet, le rang de la matrice $X'X$, ou X' désigne la matrice transposée de X, est alors inférieur à k, et la matrice $(X'X)^{-1}$, matrice inverse de $X'X$ n'existe pas.

3°) L'espérance mathématique du vecteur résiduel ε est nulle :

$E(\varepsilon)=0$ (hypothèse fondamentale)

Le zéro représente le vecteur 0 de n composantes. En effet, pour tout i, on a $E(\varepsilon_i)=0$, c'est-à-dire :

$$\begin{cases} E(\varepsilon_1)=0 \\ E(\varepsilon_2)=0 \\ \quad\vdots \\ E(\varepsilon_n)=0 \end{cases} \Leftrightarrow E\left[\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}\right] = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Leftrightarrow E(\varepsilon)=0$$

4°) L'espérance mathématique de la matrice formée par le produit du vecteur ε et de sa transposée ε' est égale à $\sigma^2 I$, où pour tout i, $\sigma^2 = E(\varepsilon_i^2)$ et où I est la matrice identité de type (n,n).

En effet,

$$\varepsilon\varepsilon' = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n) = \begin{pmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \cdots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \cdots & \varepsilon_2\varepsilon_n \\ \vdots & & \ddots & \vdots \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \cdots & \varepsilon_n^2 \end{pmatrix}$$

de types (n,1) (1,n) (n,n)

Alors, l'espérance de cette matrice sera :

$$E(\varepsilon\varepsilon') = \begin{pmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \cdots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & \cdots & E(\varepsilon_2\varepsilon_n) \\ \vdots & & \ddots & \vdots \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \cdots & E(\varepsilon_n^2) \end{pmatrix} = \Omega_\varepsilon$$

Cette matrice est la matrice des variances-covariances des erreurs que l'on note Ω_ε.

Mais, par l'hypothèse d'homoscédasticité, on a : $E(\varepsilon_1^2) = E(\varepsilon_2^2) = \cdots = E(\varepsilon_n^2) = \sigma^2$

et par l'hypothèse de non-corrélation des erreurs : $E(\varepsilon_i\varepsilon_j) = 0, \quad si\ i \neq j$

D'où $$E(\varepsilon'\varepsilon) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} = \sigma^2 I$$

Finalement, $$E(\varepsilon\varepsilon') = \Omega_\varepsilon = \sigma^2 I .$$

III.- Estimation des composants du vecteur A :

Soit le modèle $Y=XA+\varepsilon$, où ε est une variable aléatoire inconnue.

Soit $\hat{A}$ le vecteur estimateur de A.

Le modèle estimé est alors : $\hat{Y} = X\,\hat{A}$.

Désignons par e le résidu qui existe entre la valeur réelle de Y et la valeur estimé $\hat{Y}$:

$$e = Y - \hat{Y} = Y - X\,\hat{A}$$

Le vecteur e est un vecteur colonne de n composantes $e_1, e_2, \ldots, e_n$.

Calculons selon la méthode des moindres carrées, la somme des carrées des résidus :

$$\sum_{i=1}^{n} e_i^2 = e'.e$$

où e' est le vecteur ligne transposée de e.

On a : $\sum_{i=1}^{n} e_i^2 = \left(Y - X\,\hat{A}\right)'\left(Y - X\,\hat{A}\right)$

or $\left(Y - X\,\hat{A}\right)' = Y' - \left(X\,\hat{A}\right)' = Y' - \hat{A}'X'$

alors $\sum_{i=1}^{n} e_i^2 = (Y' - \hat{A}'X')(Y - X\hat{A}) = Y'Y - Y'X\hat{A} - \hat{A}'X'Y + \hat{A}'X'X\hat{A}$

les termes de cette somme sont des matrices de type (1,1), c'est-à-dire des scalaires. Or, la transposée d'un scalaire est le même scalaire. Ainsi, $Y'X\hat{A} = (Y'X\hat{A})'$

de type (1,n)(n,k)(k,1)=(1,1)

mais aussi $(Y'X\hat{A})' = \hat{A}'X'(Y')' = \hat{A}'X'Y$

Finalement $\sum_{i=1}^{n} e_i^2 = Y'Y - 2\hat{A}'X'Y + \hat{A}'X'X\hat{A}$

Le 2ème terme du 2ème membre est une forme linéaire en $\hat{A}'$, le 3ème est une forme quadratique en $\hat{A}$ (la matrice *X'X* est une matrice de type (*k,k*) symétrique).

Pour déterminer le minimum de $\sum_{i=1}^{n} e_i^2$, nous la dérivons par rapport au paramètre à estimer $\hat{A}$

On a : $\frac{\partial(\hat{A}'X'Y)}{\partial\hat{A}} = X'Y$ et $\frac{\partial(\hat{A}'X'X\hat{A})}{\partial\hat{A}} = 2X'X\hat{A}$

Donc, la dérivée de la somme des carrées des résidus est :

$$\frac{\partial\left(\sum_{i=1}^{n} e_i^2\right)}{\partial\hat{A}} = 0 - 2(X'Y) + 2X'X\hat{A}$$

Alors ; $\frac{\partial\left(\sum_{i=1}^{n} e_i^2\right)}{\partial\hat{A}} = 0 \Leftrightarrow (X'X)\hat{A} = X'Y$

Si la matrice *X* est de rang k, la matrice $(X'X)^{-1}$ existe, alors ; $\boxed{\hat{A} = (X'X)^{-1}X'Y}$

Le vecteur $\hat{A}$ est aléatoire. En effet, si on remplace Y par sa valeur $Y=XA+\varepsilon$, on aura :

$$\hat{A} = (X'X)^{-1}X'(XA+\varepsilon)$$

$$\hat{A} = (X'X)^{-1}(X'X)A + (X'X)^{-1}X'\varepsilon$$

$$\boxed{\hat{A} = A + (X'X)^{-1}X'\varepsilon}$$

Le vecteur $\hat{A}$ dépend linéairement du vecteur aléatoire ε.