IV.- Espérance mathématique de l'estimateur $\hat{A}$:

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

alors l'espérance mathématique sera : $E(\hat{A}) = E(A) + E\left[(X'X)^{-1} X'\varepsilon\right]$,

soit $E(\hat{A}) = A + (X'X)^{-1} X'E[\varepsilon]$, car A est une quantité certaine (non-aléatoire).

Par hypothèse fondamentale, on a $E[\varepsilon] = 0$,

d'où $E(\hat{A}) = A$. L'estimateur $\hat{A}$ est non-biaisé.

V.- Matrice des variances-covariances de l'estimateur $\hat{A}$:

La matrice $\Omega_{\hat{A}}$ des variances-covariances de $\hat{A}$ est définie par :

$$\Omega_{\hat{A}} = \begin{pmatrix} Var(\hat{a}_1) & Cov(\hat{a}_1, \hat{a}_2) & \cdots & Cov(\hat{a}_1, \hat{a}_k) \\ Cov(\hat{a}_2, \hat{a}_1) & Var(\hat{a}_1) & \cdots & Cov(\hat{a}_2, \hat{a}_k) \\ \vdots & & \ddots & \vdots \\ Cov(\hat{a}_k, \hat{a}_1) & Cov(\hat{a}_k, \hat{a}_2) & \cdots & Var(\hat{a}_1) \end{pmatrix}$$

En remarquant que : $Var(\hat{a}_i) = E\left[(\hat{a}_i - E(\hat{a}_i))^2\right]$

$= E\left[(\hat{a}_i - a_i)^2\right]$ car $E(\hat{a}_i) = a_i$

$= E[(\hat{a}_i - a_i)(\hat{a}_i - a_i)]$

et que $Cov(\hat{a}_i, \hat{a}_j) = E\left[(\hat{a}_i - E(\hat{a}_i))(\hat{a}_j - E(\hat{a}_j))\right]$

$= E\left[(\hat{a}_i - a_i)(\hat{a}_j - a_j)\right]$

On peut mettre $\Omega_{\hat{A}}$ sous forme :

$$\Omega_{\hat{A}} = E\left[\left(\hat{A} - E(\hat{A})\right)\left(\hat{A} - E(\hat{A})\right)'\right]$$

$$\Omega_{\hat{A}} = E\left[\left(\hat{A} - A\right)\left(\hat{A} - A\right)'\right]$$

Or $\hat{A} - A = (X'X)^{-1} X'\varepsilon$ et $\left[\hat{A} - A\right]' = \left[(X'X)^{-1} X'\varepsilon\right]'$

$\left[\hat{A}-A\right]' = \varepsilon'(X')'\left[(X'X)^{-1}\right]' = \varepsilon' X(X'X)^{-1}$ puisque $\left[(X'X)^{-1}\right]' = \left[(X'X)'\right]^{-1} = [X'X]^{-1}$

Alors, on aura :

$$\Omega_{\hat{A}} = E\left[\left((X'X)^{-1}X'\varepsilon\right)\left(\varepsilon' X(X'X)^{-1}\right)\right]$$

soit

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

Remplaçons $E(\varepsilon\varepsilon')$ par sa valeur $\sigma^2 I$, on obtient:

$$\Omega_{\hat{A}} = (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} = \sigma^2(X'X)^{-1}(X'X)(X'X)^{-1}$$

Finalement : $\boxed{\Omega_{\hat{A}} = \sigma^2(X'X)^{-1}}$

La matrice $(X'X)^{-1}$ est une matrice symétrique de type (k,k).

Pour que l'estimateur $\hat{A}$ soit convergent, il est nécessaire que la matrice $(X'X)^{-1}$ tend vers zéro lorsque le nombre d'observations n augmente indéfiniment. Habituellement, ceci se pose comme une hypothèse technique de la régression multiple.

VI.- Théorème de Gauss-Markov : L'estimateur Â est « BLUE » :

L'estimateur $\hat{A} = \left[(X'X)^{-1}X'\right]Y$ est un estimateur **BLUE** « **B**est **L**inear **U**nbiased **E**stimator » de A, c'est-à-dire que la variance de Â est la plus petite entre tous les estimateurs sans biais de A.

Démonstration :

Considérons un estimateur quelconque α de A.

a) L'estimateur α doit être linéaire en Y :
   $\alpha = MY$

   A est de type $(k,1)$. Y est de type $(n,1)$. La matrice M est ainsi de type (k,n).
   Remplaçons Y par sa valeur, on obtient : $\alpha = M(XA+\varepsilon)$

   $$\alpha = MXA + M\varepsilon$$

b) Il faut montrer que $M=(X'X)^{-1}X'$:

   Pour cela, on peut considérer l'expression : $M=(X'X)^{-1}X'+N$ et montrer que la matrice N, de type (k,n) est égale à zéro : $N=0$.

c) L'estimateur α doit être non-biaisé, c'est-à-dire qu'il doit vérifier $E(\alpha)=A$:

$E(MXA+M\varepsilon)=A \Rightarrow MXA+M\,E(\varepsilon)=A$

$\Rightarrow MXA=A$ puisque $E(\varepsilon)=0$.

C'est-à-dire la condition $MX=I$.

Ou encore $[(X'X)^{-1}X'+N]X=I \Rightarrow (X'X)^{-1}(X'X)+NX=I$

$\Rightarrow I+NX=I \Rightarrow NX=0$

d) L'estimateur α doit être de variance minimale :

La matrice Ω_α des variances-covariances de α est : $\Omega_\alpha = E\left[(\alpha - E(\alpha))(\alpha - E(\alpha))'\right]$

Mais $\alpha = MXA + M\varepsilon$

$$= I\,A + M\varepsilon$$
$$= A + M\varepsilon$$
$$= E(\alpha) + M\varepsilon$$

$$\Rightarrow \alpha - E(\alpha) = M\varepsilon$$

$$\Omega_\alpha = E\left[(M\varepsilon)(M\varepsilon)'\right] = E[M\varepsilon\varepsilon' M'] = ME(\varepsilon\varepsilon')M' = M\sigma^2 IM'$$

$$\Rightarrow \Omega_\alpha = \sigma^2 MM'$$

Cette expression peut aussi s'écrire :

$$\Omega_\alpha = \sigma^2\left[(X'X)^{-1}X' + N\right]\left[(X'X)^{-1}X' + N\right]'$$

$$= \sigma^2\left[(X'X)^{-1}X' + N\right]\left[X'(X'X)^{-1} + N'\right]$$

$$= \sigma^2\left[(X'X)^{-1}(X'X)(X'X)^{-1} + (X'X)^{-1}X'N' + NX(X'X)^{-1} + NN'\right]$$

$$= \sigma^2\left[(X'X)^{-1} + (X'X)^{-1}(NX)' + NX(X'X)^{-1} + NN'\right]$$

$$= \sigma^2\left[(X'X)^{-1} + NN'\right] \qquad \text{puisque } NX=0$$

Les variances se trouvent sur la diagonale principale de Ω_α. Il faut minimiser ces termes. Mais, σ^2 et $(X'X)^{-1}$ étant des constantes, il faut donc minimiser les éléments de la diagonale de NN'. Ces éléments qui sont des variances, sont obligatoirement positifs ou nuls.

$$NN'=\begin{pmatrix}\omega_{11} & \omega_{12} & \cdots & \omega_{1n}\\ \vdots & & \ddots & \vdots\\ \omega_{k1} & \omega_{k2} & \cdots & \omega_{kn}\end{pmatrix}\begin{pmatrix}\omega_{11} & \cdots & \omega_{k1}\\ \omega_{12} & \cdots & \omega_{k2}\\ \vdots & \ddots & \vdots\\ \omega_{1n} & \cdots & \omega_{kn}\end{pmatrix}=\begin{pmatrix}\sum_{j=1}^{n}\omega_{1j}^2 & & & & \\ & \ddots & & & \\ & & \sum_{j=1}^{n}\omega_{ij} & & \\ & & & \ddots & \\ & & & & \sum_{j=1}^{n}\omega_{kj}\end{pmatrix}$$

Ils sont minimums lorsqu'ils sont nuls, c'est-à-dire $\sum_{j=1}^{n}\omega_{ij}^2=0;\ \forall i$

Cette dernière relation est vérifiée si et seulement si $\omega_{ij}=0$, pour tout i et pour tout j, en d'autres termes si $N=0$.

Conclusion : L'estimateur Â est ainsi « BLUE » .

VII.-Estimateur de σ^2 :

σ^2 est lui-même inconnu, alors on l'estime. On peut montrer que

$$\hat{\sigma}^2=\frac{e'\cdot e}{n-k}$$

où ***e'.e*** est la somme des carrées des résidus, c'est-à-dire : $e'\cdot e=\sum_{i=1}^{n}e_i^2$ et (n-k) est le degré de liberté.

L'estimateur $\hat{\sigma}$ est sans biais, $E\left[\hat{\sigma}^2\right]=\sigma^2$.

On peut montrer aussi que :

$$\hat{\sigma}^2=\frac{Y'\Gamma Y}{n-k} \quad \text{avec} \quad \Gamma=I-X\left(X'X\right)^{-1}X'$$

Γ est une matrice carrée d'ordre n (c'est la différence de deux matrices), symétrique ($\Gamma'=\Gamma$) et idempotente ($\Gamma^2=\Gamma$). Bien sur, $Y'\Gamma Y=e'.e=\Sigma e_i^2$.

VIII.-Formule de l'analyse de la variance, coefficient de détermination et coefficient de corrélation linéaire multiple empirique :

1) Formule de l'analyse de la variance :

D'après les hypothèses de la régression et comme on a vu dans le chapitre précédent :

$$\begin{pmatrix} Variance \\ totale \end{pmatrix} = \begin{pmatrix} Variance \\ \exp liquée \end{pmatrix} + \begin{pmatrix} Variance \\ résiduelle \end{pmatrix}$$

Cette même relation, on peut l'exprimer par :

$$(Y'Y) = (\hat{A}'X'Y) + (Y'\Gamma Y)$$

2) Coefficient de détermination R^2 :

Par définition, on appelle coefficient de détermination, noté R^2 le rapport :

$$R^2 = \frac{Variance \exp liquée}{Variance\, totale} = \frac{\hat{A}'X'Y}{Y'Y}$$

R^2 représente la proportion de la variance de Y expliquée par l'influence linéaire des variables exogènes *X*.

On rappelle aussi que : $0 \leq R^2 \leq 1$.

Il est autant proche de 1 que le « nuage de points » est plus alongé et étiré. Néanmoins, ce coefficient dépend essentiellement de l'échantillon observé, d'où l'adjectif « empirique ». En effet, R^2 est une variable aléatoire puisqu'il dépend d'un échantillon aléatoire.

Ainsi, l'information fournie par R^2 n'est pas primordiale pour l'étude du modèle considéré.

3) Coefficient de corrélation linéaire empirique multiple R :

La racine carrée du coefficient de détermination R^2 est le coefficient de corrélation linéaire multiple R.

Comme on a déjà vu $-1 \leq R \leq 1$.

Le coefficient R est lui-même empirique car il dépend de l'échantillon aléatoire choisi.

IX.- Utilisation des tests statistiques :

La théorie statistique classique des tests consiste en établir une hypothèse au paravent. Le vecteur aléatoire ε suit une loi normale de moyenne nulle et de variance $\sigma^2 I$.

$$\varepsilon \to N(0;\sigma^2 I)$$

Mais, nous avons vu que $\hat{A}$ dépend linéairement de ε. En effet,

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

avec E(Â)=A et $\Omega_{\hat{A}} = \sigma^2 (X'X)^{-1}$. Donc $\hat{A} \to N\left(A;\sigma^2 (X'X)^{-1}\right)$

Il est possible, alors, de construire par les méthodes usuelles de l'analyse de la variance un test de signification des coefficients estimés précédemment.

Les tests « classiques » sont les tests de t de Student et celui de F de Fisher-Snedecor.

1) Test de t de Student :

Il consiste en tester si $a_i = 0$ (un seul parmi $i=1,2,\ldots,k$)

On calcul l'indicateur (ou la statistique) :

$$t = \frac{\hat{a}_i - a_i}{\sigma_{\hat{a}_i}}$$

où $\sigma_{\hat{a}_i}$ est l'écart-type estimé, obtenu à partir de la matrice des variances-covariances $\Omega_{\hat{A}} = \sigma^2 (X'X)^{-1}$ ainsi $\sigma_{\hat{a}_i} = \sigma\sqrt{v_{ii}}$ où vii est l'i$^{\text{ème}}$ terme de la diagonale de la matrice $(X'X)^{-1}$ et σ est estimé par $\hat{\sigma}^2 = \frac{e' \cdot e}{n-k}$

on compare t avec t_α lue dans la table de la loi de Student, pour un seuil (ou niveau de confiance) α, et $n-k-1$ degré de liberté.

Si $|t|>t_\alpha$ alors ce a_i est différent de zéro.

Si $|t|<t_\alpha$ alors ce a_i est égale à zéro.

2) Test de F de Fisher-Snedecor:

Il consiste en tester si $a_1=a_2=\ldots=a_i=\ldots=a_k=0$ (tous).

On calcul l'indicateur (ou la statistique) : $F=\dfrac{R^2/k}{(1-R^2)/(n-k-1)}$

Où R^2 est le coefficient de détermination.

On compare la valeur de F avec une valeur F_α lue dans la table de Fisher à un seuil α et en fonction des degrés de liberté k et $(n-k-1)$.

Si $F>F_\alpha$ alors au moins un des coefficients a_i est différent de zéro.

On peut aussi utiliser le tableau de l'analyse de la variance (pour ce cas de régression linéaire multiple) :

Source de variation	**Somme des carrées**	**Degré de liberté**	**Moyenne des carrées**
Régression	$\sum \hat{Y}_i^2$	k	$MCR=\left(\sum \hat{Y}_i^2 / k\right)$
Résidus	$\sum e_i^2$	$n-k-1$	$MCE=\left(\sum e_i^2 / (n-k-1)\right)$
Total	$\sum Y_i^2$	$n-1$	

On calcul $F=\dfrac{MCR}{MCE}=\dfrac{\sum \hat{Y}_i^2/k}{\sum e_i^2/(n-k-1)}$

Que l'on compare avec F_α lue dans la table de Fisher.

Si $F>F_\alpha$, alors cela veut dire qu'il existe au moins un des a_i qui est différent de zéro.