

Chapitre 7

Introduction à la régression linéaire multiple

1. MODÈLE LINÉAIRE MULTIPLE.

Modèle :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

$\beta_1, \beta_2, \dots, \beta_p$: coefficients de régression théoriques

ε : variable résiduelle.

Définitions :

- coefficient de corrélation multiple : coefficient de corrélation linéaire R entre la variable expliquée Y et son estimation par le modèle.
- coefficient de détermination : carré R^2 du coefficient de corrélation multiple..
- variance des résidus : S^2 .

$$S^2 = (1 - R^2) s_y^2$$

- estimation sans biais de la variance résiduelle σ_ε^2 :

$$S^{*2} = \frac{n}{(n - p - 1)} S^2$$

2. APPLICATIONS AUX MODÈLES ÉCONOMÉTRIQUES

2.1 les variables explicatives de la forme $X_j = t_j$.

Modèle :

$$Y_t = \beta_0 + \sum_{j=1}^p \beta_j t_j + \varepsilon$$

2.2 variables explicatives de la forme $X_j = Y_{t-j}$.

Modèle :

$$Y_t = \beta_0 + \sum_{j=1}^p \beta_j Y_{t-j} + \varepsilon_t$$

2.3 analyse des résidus.

Propriété à vérifier : répartition gaussienne et non corrélation des résidus.

graphique : on reporte simplement le temps en abscisse et les résidus en ordonnées, de la même façon que l'on représente toute série chronologique.

tests : test sur le coefficient d'autocorrélation d'ordre 1, entre les séries e_t et e_{t-1} .
test de Durbin et Watson, test « portmanteau ».

3. SÉLECTION DE VARIABLES EXPLICATIVES.

Définition :

Le coefficient de corrélation partielle de Y et X_2 sachant X_1 mesure l'information apportée par X_2 à Y après la régression de Y par X_1 .

3.1 le meilleur ensemble possible de variables explicatives

Certains logiciels donnent directement le meilleur ensemble de variables explicatives possible. On choisit en général comme critère l'estimation sans biais de la variance résiduelle.

3.2 algorithme ascendant.

- On choisit comme première variable explicative celle qui minimise la somme des carrés des résidus, ou leur variance.
- On définit ensuite comme deuxième variable explicative celle qui apporte l'information complémentaire la plus importante.
- On continue cette démarche jusqu'à ce que l'information complémentaire apportée soit non significative, par un test du F sur le coefficient de corrélation partielle.

3.3 algorithme descendant.

La procédure est l'inverse de la précédente.

- On effectue la régression par la totalité des variables explicatives disponibles.
- On considère les variables explicatives dont le coefficient de corrélation partielle avec la variable expliquée conditionnellement aux autres n'est pas significatif. Parmi elles, on élimine celle dont le coefficient de corrélation partielle est le plus petit en valeur absolue.

- On effectue la régression avec les variables explicatives sauf celle qui a été éliminée, et on recommence la procédure d'exclusion.
- On continue cette démarche jusqu'à ce que le coefficient de corrélation partielle de toutes les variables explicatives restantes avec la variable expliquée soit significatif.

3.4 algorithme stepwise.

La procédure consiste à introduire et à exclure des variables explicatives.

- L'introduction d'une variable explicative est effectuée suivant l'algorithme ascendant.
- Après chaque introduction, on effectue l'algorithme descendant pour exclure une variable dont le coefficient de corrélation partielle serait devenu non significatif.