
Le Modèle de Régression Linéaire Multiple

Mars 2020

1 Exercice théorique

1.1 Exercice 1

Montrez que l'estimateur de la variance est sans biais :

$$\mathbb{E}(\hat{\sigma}_\varepsilon^2 | X) = \sigma^2$$

1.2 Exercice 2

Supposez que le modèle suivant est estimé par MCO :

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t$$

- (a) Calculez le R^2 de la régression de ε sur x_1 et x_2 .
- (b) Supposez que vous régressez y sur une constante et \hat{y} . Qu'obtenez-vous pour les estimateurs de la constante et du coefficient ? Quel est le lien entre le R^2 de cette régression et celui du modèle original ?
- (c) Supposez que vous régressez y sur une constante et ε . Qu'obtenez-vous pour les estimateurs de la constante et du coefficient ? Quel est le lien entre le R^2 de cette régression et celui du modèle original ?
- (d) Ajoutez une variable explicative additionnelle x_3 au modèle et ré-estimez les paramètres par MCO. Montrez que l'estimateur de x_3 et sa déviation standard sont les mêmes que dans la régression de ε sur x_1 , x_2 et x_3 .

2 Application

Dans cet exercice, nous nous intéressons à l'étude du lien entre volume des ventes d'une firme et dépenses de publicité. Plus spécifiquement, nous travaillons avec des données relatives au volume de ventes d'une part, ainsi qu'aux dépenses publicitaires effectuées par la firme à destination des réseaux sociaux (YouTube et Facebook) et des outils de la presse traditionnelle. L'objectif est de déterminer quels outils publicitaires ont le plus d'impact sur les ventes sur la période d'étude.

*Q1 : Importez le fichier **marketing.xls**. Réalisez les analyses graphiques usuelles (courbes et*

nuages de point entre “sales ”et les 3 autres variables) ainsi que les statistiques descriptives.

Q2 : Procédez à la régression linéaire du volume des ventes sales sur les autres variables de la base de données, i.e. youtube, facebook et newspaper qui capturent les dépenses publicitaires dans les médias sociaux Youtube et Facebook vs. les dépenses dans le secteur de la presse traditionnelle resp. :

$$sales_t = \beta_0 + \beta_1 \cdot youtube_t + \beta_2 \cdot facebook_t + \beta_3 \cdot newspaper_t + \varepsilon_t$$

Analysez la régression à l'aide des éléments de synthèse proposés dans le summary : les coefficients estimés sont-ils tous significatifs? que vous indiquent les métrique et test(s) permettant de mesurer la qualité globale du modèle? Justifiez.

Calculez ensuite pour chacun des coefficients du modèle un intervalle de confiance au niveau $1 - \alpha = 90\%$? (commande `confint`) Ces résultats confirment-ils vos analyses à partir des t-stats réalisés précédemment? Justifiez.

Q3 : Pour tester la stabilité des coefficients, on se propose de réaliser un test de Chow sur 2 sous-périodes obtenues en coupant l'échantillon en 2 au point milieu (i.e. on analyse d'abord la régression sur une base allant de la 1^{ère} à la 100^{ème} observation, puis une autre sur des données allant de la 101^{ème} à la 200^{ème} observation).

*Comparez la réalisation du test à la main (analyse des régressions contrainte vs. non contrainte, calcul de la statistique F et de la P-value associée) vs. les résultats issus de l'utilisation du package **gap**. Les résultats et vos conclusions sont-elles bien identiques dans les deux cas?*

Q4 : Retrouvez “à la main” les coefficients du modèle.