

Acquisition et analyse de données (BING-F4002):

travaux pratiques

Marius Gilbert & Marc Dufrêne

Année académique 2015-2016

Séance VII. Régression linéaire

- Chargez le jeu de données ‘faithful’ et faites un graphique de la durée des éruption du geyser en fonction du temps d’attente.
 - La fonction ‘lm()’ qui génère un objet de type « régression linéaire » s’écrit de la même manière que la fonction ‘aov()’ utilisée précédemment. Utilisez cette fonction ainsi que la fonction `abline()` pour construire une régression linéaire.
 - En fonction des éléments vus au cours, détaillez les éléments fournis par la fonction ‘summary()’ appliqué à l’objet régression.
-
- Chargez le jeu de données « longley ». A l’aide de la fonction ‘pairs()’, examinez les relations bi-variées entre ces variables. Quelle variables selon vous devrait permettre de prédire le nombre de personnes ayant un emploi ? (variable « Employed »)
 - La fonction ‘lm()’ qui génère un objet de type « régression linéaire » s’écrit de la même manière que la fonction ‘aov()’ utilisée précédemment. A l’aide de cette fonction, faites une régression du nombre de personnes employées en fonction de l’année d’une part (« Year »), et du PNB d’autre part (variable « GNP »). Pour visualiser vos deux régression, vous pouvez utiliser la fonction ‘plot(Employed ~ Year, data = longley)’ suivie de ‘abline(myReg)’, ou myReg est votre objet régression. Comment interprétez-vous ces deux régressions ? L’une vous semble-t-elle préférable à l’autre ? Pourquoi ?
 - Ajoutez la variable « Armed.Forces » à votre modèle de régression. Cela améliore-t-il le modèle ? A-t-on des raisons de conserver cette variable ?
 - Ajoutez la variable « Population » à votre modèle de régression. Cela améliore-t-il le modèle ? A-t-on des raisons de conserver cette variable ?
 - Calculez le taux d’emploi (personnes employées / population active), et étudiez si l’année, la population totale, les personnes employées par l’armée, et le PNB permettent de prédire le taux d’emplois. Peut-on simplifier ce modèle ?
 - A l’aide la fonction `plot(myReg, which = 1)` et `plot(myReg, which = 2)` appliquée à votre objet régression, et du test de Shapiro, vérifiez les conditions d’application de votre modèle. Celles-ci vous semblent-elles rencontrées ?