Modèle Linéaire 2016-2017

Devoir Maison – Régression linéaire multiple

On s'intéresse à la base de données `Prestige` du paquet `car`. On souhaite expliquer le prestige des professions, donné par la variable `prestige` en fonction des caractéristiques de ces professions. On ne regardera que les professions salariées.

1. Créer un nouveau `data.frame` ne conservant que les professions salariées à l'aide de la fonction `subset`.
2. Décrire les variables. Pourquoi enlève-t-on la variable `census` de l'étude ?
3. On s'intéresse dans un premier temps aux corrélations entre les variables.
   (a) Faire les représentations bivariées des variables.
   (b) Regarder la matrice de corrélation des variables. Que signifie la corrélation négative entre les variables `income` et `women` ?
   (c) Donner également la matrice des corrélations partielles, qui peut être obtenue à l'aide de la fonction `pcor` du paquet `ppcor`. Commenter la corrélation et la corrélation partielle obtenues entre les variables `prestige` et `women`.
4. On souhaite mettre en place un modèle linéaire multiple expliquant la variable `prestige`.
   (a) Mettre en œuvre une sélection de variables ascendante à l'aide de la fonction `add1`.
   (b) Mettre en œuvre une sélection de variables pas à pas à l'aide de la fonction `step`. Commenter. Faire un test de Fisher de comparaison des modèles si nécessaire.
   (c) Commenter les tests de significativité des coefficients et de pertinence du modèle conservé.
5. On souhaite maintenant mettre en place un deuxième modèle linéaire multiple sans constante.
   (a) Mettre en œuvre une sélection de variables ascendante à l'aide de la fonction `add1`.
   (b) Mettre en œuvre une sélection de variables pas à pas à l'aide de la fonction `step`. Commenter. Faire un test de Fisher de comparaison des modèles si nécessaire.
   (c) Commenter les tests de significativité des coefficients et de pertinence du modèle conservé.
6. Comparer les deux modèles obtenus, avec et sans constante. Quel modèle conservez vous ?
7. Vérifier la corrélation des variables explicatives. Représenter graphiquement le nuage de points (on pourra utiliser la fonction `point3D`). Si vous disposez du paquet `rgl` vous pouvez également représenter la surface de réponse du modèle avec l'instruction `scatter3d(z ~ x+y)`.
8. Vérifier graphiquement les hypothèses du modèle. Faire un test de normalité des résidus.
9. Les métiers en sortie de Master statistique ont un salaire moyen de 40000 euros, soit environ 7800 dollars en 1971 en tenant compte de l'inflation, et une proportion d'environ 37% de femmes (donnée fictive). Donner l'intervalle de confiance du niveau de prestige donné par le modèle. Comparer avec les professions ministres et economistes (`ministers` et `economists`). Critiquez cette comparaison : a-t-elle réellement un sens ?