

Fiche d'interprétation d'une régression linéaire multiple déjà faite par un autre

Que regarder :

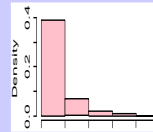
- Ajustement :
 - Souhaiter un R^2 élevé (0-1)
 - Discuter la parcimonie
- Ecrire le modèle
 - /! pour écrire, garder tous les coefficients, même les non-significatifs
 - Interpréter le modèle
 - /! ne pas conclure que les variables avec coefficients non-significatifs sont indépendantes de Y
 - Recherche multi-colinéarité
- Analyse des résidus :
 - Distribution normale
 - En fonction de \hat{Y} :
 - Moyenne indépendante de \hat{Y}
 - Variance indépendante de \hat{Y} (homoscedasticité)
 - (+/- idem en fonction de chaque X)
- Recherche d'individus influents
 - Distance Cook des individus

Problèmes fréquents :

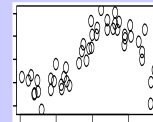
R^2 faible => régression peu utile

Trop de X_i => risque de surajustement

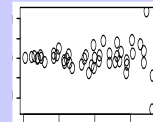
Variables X_i très corrélées (multicolinéarité) => pouvoir prédictif OK, mais mauvaise compréhension des relations entre X_i et Y



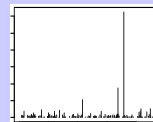
Distribution non-normale => transformer Y



Moyenne des résidus dépend de \hat{Y} => relation non-linéaire



Variance des résidus dépend de \hat{Y} (hétéroscédasticité) => relation non-linéaire



Quelques individus trop influents => discuter leur éviction