Vérifier que même sans les pts aberrants, il n’y a pas de relation entre les variables

Modèle incomplet ou manque de variable explicative si R2 proche de zéro

Regardez pour valider une régression 🡪 R2 (pas tjrs suffisant) ET  le Cp de Mallows(mieux que le R2 pour comparer les modèles).

Il faudrait connaitre l’erreur de prédiction des modèles, mais c’est un OPM ! on l’estime avec des observations, c’est soit le Cp soit l’erreur de validation croisée (err\_vc).

Pour expliquer, regardez également ces 4cvariables :

* Effet levier
* L’erreur
* (E,h,r et c)

Un bon modèle c’est quoi ?

* Modèle qui va avoir un bon pouvoir prédictif

Pouvoir prédictif d’un modèle ?

* Capable de prédire Y pour u nouveau X

Comment arriver à mesurer le pouvoir prédictif d’un modèle ?

* Séparer les observations en deux ensembles.

Un ensemble dit d’estimation (celle qui construit X) et un ensemble dit de testn(

Comment évaluer l’effet d’une variable sur une autre ?

* Regardez le coefficient directeur sur cette variable dans la régression (si proche de zero alors on peut penser qu’elle est inutile
* Il faut regarder individuellement chaque variable

Est-ce que toutes les variables sont pertinentes pour expliquer une autre ?

* Question qu’il faut se poser
* **Il faut tester toutes les combinaisons possibles de variables**

**S’il y a trop de variable, utiliser la stratégie : forward backward et le dernier**

Le tp est ultra important, bien refaire ces techniques

