

STNUM

TP



École des Ponts
ParisTech

Consignes

- Le TP est à faire en binôme.
- Pour réaliser ce TP, il est vivement conseillé de s'appuyer sur les codes fournis dans le cadre du cours (sur Educnet).
- Deux fichiers distincts doivent être transmis par mail aux responsables de PC :
 1. Le code en Python ou R (sous forme de script ou de notebook).
 2. Les réponses rédigées dans un fichier pdf (issu d'Office ou de L^AT_EX) : celui-ci ne comprend aucun code et intègre uniquement les sorties logiciel commentées (graphiques, extraits **informatifs** de listing).
- La rédaction a un poids majeur dans la notation. Ne pas oublier de commenter l'objectif suivi, la méthode appliquée (et ses conditions d'applicabilité) et les résultats.
- Les exercices et questions n'ont pas le même poids.

La résistance à la compression du béton est une fonction de l'âge et de sa composition (ciment, laitier de haut fourneau, cendres volantes, eau, superplastifiant, granulats grossiers, granulats fins). Le jeu de données `concrete_compressive_strength` contient 1030 observations relatives à des bétons, issues de (Yeh, 1998), avec les 9 variables suivantes :

- `cement` : ciment (en *kg* dans un mélange de 1 m^3),
- `blast_furnace_slag` : laitier de haut fourneau (en *kg* dans un mélange de 1 m^3),
- `fly_ash` : cendres volantes (en *kg* dans un mélange de 1 m^3),
- `water` : eau (en *kg* dans un mélange de 1 m^3),
- `superplasticizer` : superplastifiant (en *kg* dans un mélange de 1 m^3),
- `coarse_aggregate` : granulats grossiers (en *kg* dans un mélange de 1 m^3),
- `fine_aggregate` : granulats fins (en *kg* dans un mélange de 1 m^3),
- `age` : âge (en jours),
- `compressive_strength` : résistance à la compression du béton (en *MPa*).

Exercice 1 (Analyse exploratoire)

1. Représenter les distributions des 9 variables et calculer leurs principaux résumés numériques. Commenter.
2. Représenter les nuages de points de la résistance à la compression du béton en fonction des 8 covariables et calculer la matrice de corrélations. Commenter.
3. Effectuer une ACP sur les 8 covariables (toutes sauf `compressive_strength`). Analyser les composantes principales et la projection des points.
4. Effectuer un clustering, de votre choix, sur les 8 mêmes covariables. Décrire les clusters obtenus.

Exercice 2 (Statistique décisionnelle)

Comparer les résistances moyennes à la compression du béton en fonction de la présence ou non de cendres volantes.

Exercice 3 (Régression)

On souhaite modéliser résistance à la compression du béton en fonction des covariables.

- 1. Effectuer une régression linéaire multiple. Commenter les résultats obtenus (significativité des paramètres, qualité de la régression).*
- 2. Déterminer un modèle, désigné par modèle (*), ayant tous ses paramètres significatifs.*
- 3. Effectuer une sélection automatique de modèles à l'aide d'une procédures automatique (backward, forward ou stepwise) . Commenter.*
- 4. Séparer aléatoirement l'échantillon en 2 parties : 80% pour l'échantillon d'apprentissage, 20% pour l'échantillon test. Evaluer la qualité prédictive du modèle (*). Commenter.*
- 5. Répéter la question précédente pour 2 autres séparations apprentissages-test (en conservant la proportion 80%-20%). Commenter.*

Bibliographie

Yeh, I.-C. 1998, «Modeling of strength of high performance concrete using artificial neural networks», *Cement and Concrete Research*, vol. 28, n° 12, p. 1797–1808.