

Ce travail est à réaliser en équipe (**minimum deux, maximum quatre** personnes). On cherche à estimer le total du solde de la carte de crédit VISA Première (credm , en francs) à l'aide d'autres variables explicatives présentes dans la base de données visacredm ; 25 données sont intentionnellement manquantes pour permettre d'évaluer vos modèles finaux et faire un classement des équipes.

1. Faites une brève analyse exploratoire des données. Quelles variables seraient d'après-vous le plus utile pour modéliser credm ? Utilisez des graphiques pour déterminer si vous devriez inclure des transformations (logarithmiques, indicateurs de points de ruptures, interactions, polynômes) des variables explicatives. Résumez votre analyse en une page maximum (texte et graphiques)
2. **Sélection de variables** : Essayez les méthodes de sélection suivantes (choose) pour effectuer votre sélection pour ces données avec les trois options suivantes :
 - (a) sélection avec critères d'information (pénalisation)
 - (b) validation croisée à cinq plis
 - (c) séparation de la base de données en échantillons d'entraînement (2/3) et de validation (1/3)

Pour chacune des méthodes de sélection, utilisez les procédures suivantes pour la recherche de modèles :

- Sélection séquentielle
- Régression avec pénalité q_1 (LASSO) — choisissez la valeur du paramètres de pénalité λ par validation croisée avec votre échantillon d'apprentissage.
- Moyenne de modèles

Rapportez l'erreur moyenne quadratique pour les différents modèles avec les méthodes (b) et (c) dans un tableau pour chaque modèle estimé.

3. Écrivez un rapport résumant vos trouvailles et détaillant votre démarche. Fournissez votre code SAS et une base de données au format `d4_matricule.sas7bdat` avec deux colonnes et les 25 lignes correspondant aux données manquantes pour credm : la première colonne contiendra les matricules (`matric`), la deuxième colonnes les prédictions (`predict`) selon votre modèle préféré parmi tous ceux essayés. Justifiez votre choix adéquatement.

Vous serez évalués sur votre méthodologie, et non pas la performance relative de votre modèle par rapport à celles des autres étudiant(e)s. Vous devez expliquer clairement votre démarche (méthodologie) dans votre rapport et décrire le modèle que vous avez retenu (les variables utilisées).