

## **M2 Économétrie appliquée**

### **Étude de cas – Défaut de remboursement**

L'objectif de l'analyse est de construire un modèle permettant de prédire le risque de défaut de remboursement des clients auprès d'un organisme de crédit (et de son partenaire commercial en lien avec le projet d'achat du client) au moment de l'étude de leur demande de prêt. Si ce risque est trop élevé, l'organisme de crédit refusera d'accorder un prêt au client. Il s'agit donc de proposer une méthode pour calculer un « score » correspondant à la probabilité individuelle de défaut.

#### **1. Description de la base de données**

La base de données "*credit*" contient des informations sur 64790 clients d'un organisme de crédit dont certains sont classés Mauvais (n'ont pas pu rembourser leur emprunt), d'autres Intermédiaires (retards de paiement) et d'autres encore Bons (pas d'incidents de paiement). Le dictionnaire des variables est en fichier joint.

- 1) Décrire la base de données et faire des statistiques descriptives sur toutes les variables.
- 2) Créer une variable dichotomique « défaut de paiement ».
- 3) Tracer le nuage de points de la relation linéaire entre la probabilité de défaut de remboursement et le montant du prêt. Que concluez-vous ?
- 4) Idem avec d'autres variables disponibles.
- 5) Écrire un modèle binaire pour décrire la relation entre la probabilité de défaut et les caractéristiques du client.
- 6) Quelles variables explicatives incluriez-vous dans le modèle ?

#### **2. Construction de l'échantillon d'apprentissage**

Séparer l'échantillon complet en deux échantillons :

- un échantillon d'apprentissage, sur lequel se feront l'estimation et le choix du modèle
- et un échantillon test qui permettra la validation du modèle.

- 7) Réaliser un tirage aléatoire de 40000 individus sur l'échantillon complet. Créer l'échantillon « apprentissage » (40000 observations) et l'échantillon « test » (les observations restantes).
- 8) Estimer votre modèle à l'aide d'une régression logistique sur l'échantillon « apprentissage ». Interpréter les résultats.
- 9) Créez la variable « score » qui contient la probabilité de défaut estimée par le modèle pour chaque client de l'échantillon « apprentissage ».
- 10) Si l'on considère que le client a fait défaut si la probabilité estimée de défaut est supérieure à 0,5, quelle est la part d'observations bien classées (i.e. dont la vraie valeur de défaut correspond à l'estimation) ? Quelle est la part d'observations mal classées ?
- 11) Tracez la courbe ROC. Que concluez-vous quant à la qualité de votre modèle ?

#### **3. Validation du modèle sur l'échantillon test**

- 12) Estimer le modèle sur l'échantillon « test ».
- 13) Faites l'analyse du modèle (questions 9 à 11) sur l'échantillon « test ».

#### **4. Rédaction du rapport**

- 14) Rédiger un rapport à destination de l'organisme de crédit (intelligible donc pour des professionnels) reprenant les principaux éléments de votre analyse (description de la base de données, présentation du modèle, choix des variables et raisons de votre choix, résultats de l'estimation et performance du modèle).