

# Analyse des Correspondances Multiples.

Travaux pratiques (statistiques exploratoires).

Les jeux de données étudiés sont disponibles sur <http://math.univ-lille1.fr/~marbaclo/>

## 1 Crédits à consommation sous R

Le jeu de données `credit.csv`<sup>1</sup> contient 66 clients ayant souscrit un crédit à la consommation dans un organisme de crédit. Les 11 variables qualitatives et les modalités associées à cet exemple sont les suivantes :

- Marché : rénovation d'un bien, voiture, scooter, moto, mobilier-ameublement, side-car. Cette variable indique le bien pour lequel les clients ont réalisé un emprunt.
- Apport : oui, non. Cette variable indique si les clients possèdent un apport personnel avant de réaliser l'emprunt. Un apport personnel représente une garantie pour l'organisme de crédit.
- Impayé : 0, 1 ou 2, 3 et plus. Cette variable indique le nombre d'échéances impayées par le client.
- Taux d'endettement : 1 (faible), 2, 3, 4 (fort). Cette variable indique le niveau d'endettement du client. Le taux d'endettement est calculé comme le rapport entre les charges (ensemble des dépenses) et le revenu. Ce taux a été discrétisé en 4 classes.
- Assurance : sans assurance, AID (assurance invalidité et décès), AID + chômage, Senior (pour les plus de 60 ans). Cette variable indique le type d'assurance à laquelle le client a souscrit.
- Famille : union libre, marié, veuf, célibataire, divorcé.
- Enfants à charge : 0, 1, 2, 3, 4 et plus.
- Logement : propriétaire, accédant à la propriété (personne qui n'a pas encore fini de rembourser son emprunt immobilier), locataire, logé par la famille, logé par l'employeur.
- Profession : ouvrier non qualifié, ouvrier qualifié, retraité, cadre moyen, cadre supérieur.
- Intitulé : M, Mme, Mlle.
- Age : 20 (18 à 29 ans), 30 (30 à 39), 40 (40 à 49), 50 (50 à 59), 60 et plus.

Le but de cette étude est de caractériser la clientèle de l'organisme de crédit. Nous voulons dans un premier temps mettre en évidence différents profils de comportement bancaires, c'est-à-dire effectuer une typologie des individus. Nous voulons ensuite étudier la liaison entre la signalétique (CSP, âge, etc.) et les principaux facteurs de variabilité des profils de comportement bancaires (i.e. caractériser les clients aux comportements particuliers).

On vous propose d'effectuer les étapes suivantes :

1. Importer le jeu de données
2. Effectuer les statistiques descriptives des variables. Vérifier la pertinence des modalités proposées.
3. Effectuer l'ACM en choisissant de manière pertinente les variables actives et les variables supplémentaires. (rappel : on souhaite déterminer les profils des comportements bancaires).
4. Choisir le nombre d'axe.
5. Interpréter les résultats. Les représentations graphiques des ACM sont vites surchargées. Pour obtenir des représentations séparées, il faut utiliser l'argument `invisible` de la fonction `plot`.

## 2 Races des chiens sous R

Les données présentées dans `chiens.txt`<sup>2</sup> décrivent les caractéristiques de 27 races de chiens au moyen de variables qualitatives. Faites l'ACM de ce tableau de données en considérant la variable `fonction` comme supplémentaire.

1. P.A. Cornillon et al. Statistiques avec R, 2nd édition. PUR éditions. 2012

2. G. Saporta. Probabilité, analyse des données et statistiques. Éditions Technip. 1990

Le codage des différentes modalités a la signification suivante :

Variable	Modalité		
	1	2	3
Taille	petite taille	taille moyenne	grande taille
Poids	petit poids	poids moyen	poids élevé
Vélocité	lent	assez rapide	très rapide
Intelligence	peu intelligent	intelligence moyenne	très intelligent
Affection	peu affectueux	affectueux	
Agressivité	peu agressif	agressif	
Fonction	chien de compagnie	chien de chasse	utilité

TABLE 1 – Codage des différentes modalités pour les races de chiens.

### 3 Races des chiens sous SAS

On reprend le jeu de données précédent.

1. Importer ces données dans SAS.
2. Programmer en SAS une étape **DATA** permettant de renommer les modalités des variables. Par exemple la modalité 1 de la variable Taille deviendra TAILLE1. Ce sera utile pour s'y retrouver dans les tableaux et les figures de l'ACM.
3. Effectuer l'ACM avec l'option **tableau de burt**. Ne pas oublier aussi de projeter la variable supplémentaire.
4. Effectuer l'ACM avec l'option **tableau disjonctif complet**. Ne pas oublier aussi de projeter la variable supplémentaire.
5. Comparer les résultats produits par les deux analyses.