

Classification de variables qualitatives pour la compréhension de la prise en compte de l'environnement par les agriculteurs

Jacqueline CANDAU, Marie CHAVENT, Philippe DEUFFIC, Vanessa KUENTZ, Sandrine LYSER, Jérôme SARACCO¹

Depuis une trentaine d'années, l'agriculture française et européenne a remis en cause son modèle établi dans les années 60. Certes aujourd'hui, la finalité de l'agriculture est toujours de fournir des biens alimentaires aux hommes mais d'autres objectifs viennent s'ajouter, comme la protection et le renouvellement des ressources naturelles, l'emploi ou le maintien d'un tissu économique et social dans les territoires ruraux.

C'est autour de ce concept de « multifonctionnalité » de l'agriculture, apparu au sommet de Rio en 1992, qu'une enquête auprès de 1000 agriculteurs français a été réalisée en 2005 par une équipe de sociologues du Cemagref de Bordeaux, afin de « cerner la façon dont les agriculteurs conçoivent la protection de l'environnement en relation avec leur activité » (Candau et al., 2005). Différentes questions ont été posées en relation à quatre grandes thématiques : conception du métier, de l'environnement, des mesures agro-environnementales et de la nature ; et ont permis d'aboutir à une centaine de variables qualitatives. Les méthodes d'analyse des données sont particulièrement adaptées pour répondre à ce type de questionnement. Une stratégie classique pour obtenir des profils-types d'individus consiste à réaliser une analyse factorielle des données (ACM dans notre cas) puis à appliquer une méthode de classification sur les composantes principales obtenues.

Dans cette contribution, nous proposons de remplacer la première étape par un outil novateur qu'est la classification de variables. L'objectif de la classification de variables est de construire des classes de variables liées et de supprimer l'information redondante. Des algorithmes spécifiques ont été développés pour la classification de variables. Nous utilisons l'approche proposée par Chavent et al. (2011) et implémentée dans le package *ClustOfVar* du logiciel R.

Dans cette méthode, le critère d'homogénéité d'une classe est la somme des carrés des corrélations (pour les variables quantitatives) et des rapports de corrélations (pour les variables qualitatives) à une variable synthétique (quantitative) résumant au mieux les variables de la classe. La variable synthétique qui maximise ce critère est la première composante principale calculée par la méthode PCAMIX. Cette méthode d'analyse factorielle de données mixtes (Kiers, 1991), aussi appelée AFDM (Pagès, 2004), inclut l'ACP et l'ACM comme cas particuliers. Un algorithme de classification ascendante hiérarchique est utilisé pour maximiser le critère de partitionnement. Une approche basée sur le bootstrap est également développée pour aider au choix du nombre de variables synthétiques.

Dans notre étude, cette approche permet de mettre en évidence une structure dans les réponses des agriculteurs, en résumant l'information par 9 variables synthétiques (relatives au cadre de vie, au patrimoine, à la perception du métier, aux problèmes de l'environnement, etc.). Dans un second temps, cette analyse est complétée par une classification ascendante hiérarchique (CAH) sur les variables synthétiques afin d'établir des profils-types d'agriculteurs. Cette typologie nous conduit à mieux appréhender la perception de l'environnement par cette catégorie socioprofessionnelle et à mesurer ainsi l'importance accordée aux problématiques environnementales parmi les autres préoccupations actuelles.

¹ jacqueline.candau ; philippe.deuffic ; vanessa.kuentz ; sandrine.lyser}@cemagref.fr
marie.chavent ; [jerome.saracco](mailto:jerome.saracco@math.u-bordeaux1.fr)@math.u-bordeaux1.
Cemagref, UR ADBX, F-33612 Cestas Cedex, France Université de Bordeaux, France

Notons que l'utilisation de la classification de variables, comparativement à l'approche plus classique d'ACM, permet de mieux caractériser et interpréter la formation des groupes d'agriculteurs (intéressés par le changement, attentifs à la protection de l'environnement, adeptes de la déprise agricole, etc.)

Références

- Candau, J., Deuffic, P., Ginelli, L., Lewis, N., Lyser, S. (2005). La prise en compte de l'environnement par les agriculteurs. *Résultats d'enquête. CNASEA*, 83 p.
- Chavent, M., Kuentz, V., Liquet, B., Saracco, J. (2011). Classification de variables : le package ClustOfVar. *43èmes Journées de Statistique (SFdS)*, Tunisie.
- Kiers, H.A.L., (1991). Simple structure in Component Analysis Techniques for mixtures of qualitative and quantitative variables, *Psychometrika*, 56, 197-212.
- Pagès, J. (2004). Analyse factorielle de données mixtes, *Revue de statistique appliquée*, 52(4), 93-111.