

A. Labenne^a, M. Chavent^{b,c}, V. Kuentz-Simonet^a, T. Rambonilaza^a and J. Saracco^{b,c}

^aIRSTEA, UR ADBX
33612 Cestas Cedex, France
amaury.labenne@irstea.fr

^bUniv. Bordeaux, IMB, UMR 5251
F-33400 Talence

^cINRIA, CQFD
F-33400 Talence

Mots clefs : analyse factorielle multiple, méthode PCAMIX, analyse de la qualité de vie

1 Introduction

L'Analyse Factorielle Multiple (AFM) est une méthode de réduction de dimension qui permet de prendre en compte le fait que les individus sont décrits par des variables naturellement structurées en groupes ou thématiques. Initialement l'AFM a été mise en place pour l'analyse de variables quantitatives (Escofier et Pagès [1]). Elle a ensuite été élargie à l'analyse de groupes de variables qualitatives (Escofier et Pagès [2]) puis à l'étude d'un tableau de données que l'on qualifiera de "semi-mixte", où chaque groupe peut être soit de type quantitatif, soit de type qualitatif. Cette dernière extension de la méthode, proposée par Pagès [3], permet la réduction de dimension dans un contexte où les bases de données deviennent de plus en plus composites. A ce titre, nous sommes confrontés à une variété de données complexes dans les travaux relatifs à la construction d'indicateurs du développement durable, il faut entendre par là l'état de l'environnement, de l'économie, de la santé, des conditions sociales des individus comme des communautés. Pour cela, nous optons pour une approche en termes de qualité de vie : en effet, l'analyse et la mesure du bien être et de ses différentes composantes constituent un indicateur pertinent pour l'évaluation des états des sociétés. Face à la multitude de variables issues de thématiques différentes (environnement, social, économie, démographie, etc.) disponibles pour décrire la qualité de vie, les méthodes multi-tableaux telles l'AFM sont une réponse pertinente pour l'analyse de ces données structurées en groupes. Dans cette problématique, les variables au sein d'une même thématique ne sont pas homogènes, mais mixtes dans le sens où elles peuvent être quantitatives ou qualitatives. L'écriture actuelle de l'AFM et son implémentation dans le package R FactoMineR (Husson et al. [4]) ne permettant pas d'intégrer des thématiques mixtes dans l'analyse, nous proposons une extension de l'AFM qui permet l'analyse de groupes mixtes via l'utilisation de la méthode PCAMIX, voir par exemple Chavent et al. [5].

2 Rappels sur la méthode PCAMIX

La méthode PCAMIX présentée par exemple dans Chavent et al. [5] est une méthode d'analyse factorielle pour des données mixtes, c'est à dire un mélange de variables quantitatives et qua-

litatives. Elle est similaire à la méthode d'analyse factorielle de données mixtes d'Escofier [6] et Pagès [7] dans la manière dont sont recodées les variables qualitatives et quantitatives. Cependant, Chavent et al. [5] propose une formulation de PCAMIX à l'aide d'une décomposition en valeurs singulières (SVD) des données préalablement transformées. Cela permet d'obtenir directement les composantes principales, les "loadings" des variables quantitatives (corrélations avec les composantes principales), les rapports de corrélation entre les variables qualitatives et les composantes, ainsi que les coordonnées principales des modalités des variables qualitatives.

3 La méthode MFAMIX

Cette méthode se distingue d'une ACP ou d'une ACM globale appliquée à l'ensemble des données dans la mesure où elle permet de prendre en compte la structure en groupes de l'ensemble des variables. Pour cela, l'AFM applique une pondération particulière aux variables selon leur appartenance aux différentes thématiques. Ainsi l'influence des groupes est équilibrée dans la construction des composantes principales globales.

Le principe général de l'AFM mixte (appelée MFAMIX dans la suite) repose essentiellement sur deux étapes. Tout d'abord, on analyse chaque thématique prise séparément avec la méthode PCAMIX. On obtient ainsi la plus grande valeur propre correspondant à chaque sous-tableau. Puis, on applique PCAMIX sur l'ensemble de toutes les variables prises en commun où chaque variable est pondérée par l'inverse de la première valeur propre de la thématique dont elle est issue. Ainsi l'influence de chaque groupe est équilibrée dans la construction des composantes principales globales. Nous proposons une écriture sous forme de SVD pour MFAMIX. Les codes R sont disponibles auprès des auteurs et feront l'objet d'un package R.

L'application de la méthode via les codes R associés sera illustrée sur des données socio-économiques relatives à la qualité de vie d'un ensemble de communes du bassin versant Adour-Garonne.

Références

- [1] Escofier B et Pagès J (1983), Méthode pour l'analyse de plusieurs groupes de variables. Application à la caractérisation des vins rouges du Val de Loire, *Revue de statistique appliquée*, 31(2) : 43-59.
- [2] Escofier B et Pagès J (1998), *Analyses factorielles simples et multiples*, Dunod, 3^e ed.
- [3] Pagès J (2002), Analyse factorielle multiple appliquée aux variables qualitatives et aux données mixtes, *Revue de statistique appliquée*, 50(4) : 5-37.
- [4] Husson F, Josse J, Lê S et Mazet J (2012). FactoMineR : Multivariate Exploratory Data Analysis and Data Mining with R, *R package version 1.20*. <http://CRAN.R-project.org/package=FactoMineR>.
- [5] Chavent M, Kuentz-Simonet V, Saracco J, (2012), Orthogonal rotation in PCAMIX, *Advances in Data Analysis and Classification*, 6 : 131-146.
- [6] Escofier B (1979), Traitement simultané de variables qualitatives et quantitatives en analyse factorielle, *Cahiers de l'Analyse des données*, 4(2) : 137-146.
- [7] Pagès J (2004), Analyse factorielle de données mixtes, *Revue de statistique appliquée*, 52(4) : 93-111.