



Rotation orthogonale en ACP de données mixtes. Le package PCAmixdata et une application en sociologie culturelle.

Marie Chavent, Vanessa Kuentz-Simonet, Zoltan Lakatos, Jerome Saracco

► To cite this version:

Marie Chavent, Vanessa Kuentz-Simonet, Zoltan Lakatos, Jerome Saracco. Rotation orthogonale en ACP de données mixtes. Le package PCAmixdata et une application en sociologie culturelle.. 1ères Rencontres R, Jul 2012, Bordeaux, France. hal-00717515

HAL Id: hal-00717515

<https://hal.archives-ouvertes.fr/hal-00717515>

Submitted on 13 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Rotation orthogonale en ACP de données mixtes.
Le package PCAmixdata et une application en sociologie culturelle.**

M. Chavent^a, V. Kuentz-Simonet^b, Z. Lakatos^c and J. Saracco^a

^a INRIA Bordeaux Sud Ouest,
Institut de Mathématiques de Bordeaux,
Université de Bordeaux,
351 cours de la libération, 33405 Talence Cedex, France
{marie.chavent, jerome.saracco}@math.u-bordeaux1.fr

^bIrstea, UR ADBX,
50 avenue de Verdun, 33612 Cestas Cedex, France
vanessa.kuentz-simonet@irstea.fr

^c BME-GTK
Université Polytechnique et Economique de Budapest,
Faculté des Sciences Economiques et Sociales,
Stoczek utca 2, 1111 Budapest, Hongrie
sultan.lakatos@gmail.com

Mots clefs : Données mixtes, mélange de variables quantitatives et qualitatives, analyse en composantes principales, analyse des correspondances multiples, rotation, sociologie culturelle.

La méthode PCAMIX est une méthode d'analyse de données qui permet en particulier de déterminer des composantes principales pour des données mixtes, c'est à dire pour un mélange de variables quantitatives et de variables qualitatives. L'approche PCAMIX inclut ainsi comme cas particulier l'analyse en composantes principales (ACP) et l'analyse des correspondances multiples (ACM). Kiers (1991) a présenté une manière de faire de la rotation orthogonale dans le cadre de la méthode PCAMIX.

Dans cette communication, nous donnons une nouvelle présentation de la méthode PCAMIX dans laquelle les composantes principales et les "squared loadings" sont obtenus via une décomposition aux valeurs singulières. Nous proposons alors une procédure efficace pour faire de la rotation de type Varimax en PCAMIX. Nous donnons en particulier une solution directe de l'angle optimal de rotation. Un package appelé "PCAmixdata" comprenant la méthode PCAMIX ainsi que la rotation en PCAMIX a été implémenté dans R. Ce package disponible auprès des auteurs et sera bientôt sur le site du CRAN. Notons que ce package permet aussi de gérer des données manquantes.

Nous illustrerons avec des données simulées le bon comportement numérique de l'algorithme proposé dans ce package et le gain en terme de temps de calculs par rapport à l'approche de Kiers (1991).

Nous présenterons aussi une application sur des données réelles permettant d'illustrer l'intérêt de faire de la rotation en ACM (cas particulier de PCAMIX lorsque toutes les variables sont qualitatives). Cette application est issue d'une collaboration avec un doctorant hongrois en sociologie, Zoltan Lakatos, dont les travaux portent sur l'évolution des valeurs culturelles dans les sociétés. Plus précisément, il s'intéresse à une critique empirique de la thèse sociologique du post- matérialisme du politologue américain Ronald Inglehart. Les données sur lesquelles il travaille sont issues d'une enquête globale (World Values Survey) sur les valeurs culturelles, initiée et dirigée par Ronald Inglehart (enquêtes individuelles menées au niveau national dans une centaine de pays, par vagues successives depuis 1981). Une ACM au niveau des répondants est appliquée pour mettre en évidence les limitations des thèses d'Inglehart sur le contenu et l'évolution des valeurs culturelles. Grâce à la rotation, il devient possible d'identifier deux dimensions distinctes, à savoir "religieux vs. laïque" et "autoritaire vs. libertaire" que les méthodes d'Inglehart (ACP appliquées aux moyennes nationales) traitent comme appartenant à une seule et même dimension sous l'étiquette "valeurs traditionnelles vs. modernes". Les résultats obtenus par l'ACM avec rotation constituent un pas important vers le dépassement des typologies

réductionnistes en sociologie des valeurs présentant un schéma unidirectionnel et linéaire de l'évolution socioculturelle.

Références

[1] Chavent, M., Kuentz-Simonet, V. Saracco, J. (2012). Orthogonal rotation in PCAMIX. To appear in *Advances in Data Analysis and Classification*.

[2] Kiers, H.A.L. (1991) Simple structure in component analysis techniques for mixtures of qualitative and quantitative variables. *Psychometrika*, 56:197-212.

[3] LAKATOS, Zoltan (2012) The Cultural Values-Economic Growth Nexus: A Critical Reassessment. Doctoral thesis: Faculty of Social Sciences, Eotvos Lorand University of Sciences (ELTE TaTK), Budapest.