

# Analyse en Composantes Principales Mixte

Rafik Abdesselam

CREM UMR CNRS 6211

Université de Caen, Esplanade de la Paix, F-14032 Caen, France

rafik.abdesselam@unicaen.fr,

<http://www.unicaen.fr/crem/membres.php>

**Résumé.** Le traitement simultané de données mixtes (quantitatives et qualitatives) ne peut pas se réaliser directement par les méthodes classiques de la statistique exploratoire multidimensionnelle. Dans ce travail, l'analyse factorielle sur données mixtes proposée est une analyse en composantes principales normée après transformation des indicatrices des variables qualitatives en variables quantitatives au travers de projections de nuages de points dans l'espace des individus correspondant à des analyses de la variance multivariée. La méthode est évaluée sur la base d'une application sur données réelles mixtes.

## 1 Introduction

Dans le cadre d'un traitement de données multidimensionnelles, il est très fréquent que le thème homogène de variables à analyser soit constitué de données mixtes (variables quantitatives et qualitatives). La méthodologie usuelle de traitement consiste, soit à mettre les variables qualitatives [resp. quantitatives] en éléments illustratifs dans une Analyse en Composantes Principales (ACP) [resp. Analyse des Correspondances Multiples (ACM)], soit encore de discrétiser les variables quantitatives du thème en variables qualitatives en vue d'une ACM, ce qui introduit très souvent un biais dû au choix du nombre de classes et de leurs amplitudes égales ou différentes, et qui occasionne une perte d'information. De nombreux chercheurs se sont intéressés à cette problématique et ont proposé des méthodes qui traitent simultanément les deux types de variables en éléments actifs : l'ACP avec indicatrices introduite par Tenenhaus (1977); Escofier et Pagès (1979); Saporta (1990); Young (1981); Pagès (2002) et plus récemment l'Analyse Factorielle de Données Mixtes (AFDM) proposée par Pagès (2004).

L'Analyse en Composantes Principales Mixte (ACPM) proposée est une ACP normée des variables quantitatives et des indicatrices des variables qualitatives transformées en variables combinaisons linéaires des variables quantitatives, à partir de projections orthogonales de nuages de points dans l'espace des individus muni d'un produit scalaire relationnel. Chaque variable qualitative est quantifiée séparément, puis traitée de la même façon que les variables quantitatives.