

# Extension de l'ACPVI multibloc à la discrimination qualitative. Application en épidémiologie vétérinaire

Stéphanie Bougeard<sup>1</sup>, El Mostafa Qannari<sup>2</sup> ; Christelle Fablet<sup>1</sup>

<sup>1</sup> AFSSA, Département d'épidémiologie animale - Zoopôle, BP53, 22440 Ploufragan  
[s.bougeard@afssa.fr, c.fablet@afssa.fr]

<sup>2</sup> ENITIAA, Unité de Sensométrie et Chimiométrie - Rue de la Géraudière BP 82225, 44322 Nantes  
[elmostafa.qannari@enitiaa-nantes.fr]

**Résumé** Une nouvelle présentation de l'Analyse en Composantes Principales sur Variables Instrumentales Multibloc, dont l'objectif est de prédire un tableau  $Y$  à partir de plusieurs tableaux  $(X_1, \dots, X_K)$ , est proposée. Elle est basée sur la détermination, pas à pas, de composantes dans l'espace des variables  $Y$ . Chaque composante est projetée sur les espaces engendrés respectivement par les variables des tableaux  $X_k$  ( $k = 1, \dots, K$ ). L'ACPVI multibloc consiste à maximiser, en moyenne, la variance restituée par ces projections. Cette méthode multibloc est ensuite appliquée au cadre de la description et la prédiction d'une variable qualitative  $y$  par un ensemble de variables qualitatives  $(x_1, \dots, x_K)$ , chaque variable étant codée en un tableau contenant les indicatrices de ses modalités. La discrimination est opérée sur la base de composantes globales mutuellement orthogonales résumant l'ensemble des variables explicatives. La démarche d'analyse est comparée à d'autres méthodes de discrimination qualitative et illustrée sur la base d'une étude de cas en épidémiologie vétérinaire.

**Mots clés :** Analyse canonique généralisée, analyse en composantes principales sur variables instrumentales multibloc, analyse des correspondances multiples avec un tableau de référence, discrimination qualitative

**Résumé** A new presentation of Multiblock Redundancy Analysis is discussed. This method of analysis aims at predicting a set of variables  $Y$  from several blocks of variables  $(X_1, \dots, X_K)$ . It consists in determining, step by step, components in the  $Y$  space which are projected upon the spaces spanned by the variables in  $X_k$  ( $k = 1, \dots, K$ ). At each step, the component is sought in such a way so as to maximize the averaged variance explained by the projections. Thereafter, the method of analysis is applied to the case of categorical variables, each variable being coded by the indicators of its categories. The discrimination and classification is achieved using orthogonal components from the predictive variables. We also outline how the method is related to other qualitative discriminant techniques. The interest of the method is illustrated on the basis of a case study in the field of veterinary epidemiology.

**Keywords :** Generalized canonical analysis, multiblock redundancy analysis, multiple correspondence analysis with a reference table, discrimination with categorical variables