

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

Stéphanie Bougeard¹, El Mostafa Qannari² ; Christelle Fablet¹

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

² 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

1 Introduction

Dans le domaine vétérinaire, l'épidémiologie analytique est un outil permettant l'étude des causes apparentes et des événements directement ou indirectement associés au phénomène de santé étudié (22). Ce type d'étude vise à mesurer l'intensité de la liaison entre les facteurs étudiés et la maladie, et ainsi déterminer les facteurs de risque associés au développement de cette maladie. Les données sont obtenues le plus souvent au travers d'enquêtes conduites en élevage, à l'abattoir et dans des laboratoires de diagnostic. La maladie étudiée est soit une variable dichotomique (non malade, malade), soit une variable nominale (non malade, un peu malade, malade). L'objectif est d'expliquer une variable qualitative par un grand ensemble de variables également qualitatives. Cette problématique statistique est classique et de nombreux ouvrages et thèses ont été écrits sur le sujet (6; 9). Du fait du grand nombre de variables explicatives et de leurs liaisons structurelles, nous choisissons de nous situer dans le cadre des méthodes factorielles propices à l'étude de ce type de données. Les méthodes entrant dans le cadre des cas particuliers des modèles linéaires généralisés (régression logistique, modèles log-linéaires, etc.) ou de la segmentation ne sont pas évoquées ici. Notre objectif de traitement est à la fois descriptif et prédictif. Pour cela, nous cherchons des espaces de représentation permettant de mettre en évidence la structure de dépendance entre l'ensemble des variables explicatives et la variable à expliquer. L'objectif prédictif, qui revient à expliquer la variable y par les variables (x_1, \dots, x_K) , est un problème de discrimination qualitative. Pour cela, des composantes (*i.e.* variables latentes), résumés des variables explicatives contraintes à être orthogonales mutuellement, sont déterminées et utilisées pour expliquer la variable y . L'interprétation des résultats conduit en particulier à cerner les facteurs de risque et les facteurs à effet protecteur de la maladie.

Dans une publication précédente, l'ACPVI multibloc est proposée pour décrire les liens entre K tableaux (X_1, \dots, X_K) orientés vers l'explication d'un tableau Y , ces variables quantitatives étant mesurées sur les mêmes individus (2). Cette méthode constitue une extension de l'Analyse en Composantes Principales sur Variables Instrumentales (19; 24; 20) au cas de données structurées en plusieurs blocs. L'ACPVI multibloc est initialement présentée au travers de plusieurs critères qui reflètent différentes facettes de cette méthode. Nous présentons ici une nouvelle façon d'introduire l'ACPVI multibloc qui souligne notamment ses liens avec l'analyse canonique généralisée (11; 3). Elle peut être adaptée au cadre qualitatif en considérant que chaque bloc est constitué par l'ensemble des modalités d'une variable qualitative codée de façon disjonctive. Par la suite, la méthode proposée sera désignée par ACPVI multibloc qualitative.

2 Méthode

2.1 ACPVI multibloc

Soit un tableau de variables quantitatives X partitionné en K blocs $X = [X_1 | \dots | X_K]$. Chaque tableau X_k est un tableau $(N \times P_k)$ dont les lignes correspondent aux mêmes N individus. Toutes les variables sont supposées centrées. On désigne par $P_{X_k} = X_k(X_k'X_k)^{-1}X_k'$ le projecteur associé au sous-espace engendré par les variables de X_k . L'analyse canonique généralisée (ACG) peut être définie comme une démarche pas à pas, qui consiste à déterminer à chaque étape une variable canonique, *i.e.* une composante normée t associée au tableau concaténé X , définie par $t = Xw$ et liée de manière optimale aux différents tableaux (X_1, \dots, X_K) . Pour cela, t est projetée sur chaque sous-espace engendré par les variables de X_k . L'ACG recherche à maximiser en moyenne les variances restituées par ces projections, ce qui revient du fait des propriétés

de symétrie et d'idempotence du projecteur, pour chaque solution d'ordre $h = (1, \dots, H)$, à maximiser le critère (1).

$$\text{Maximiser } \sum_{k=1}^K \text{var}(P_{X_k} t^{(h)}) = \frac{1}{N} t^{(h)'} \sum_k P_{X_k} t^{(h)} \quad (1)$$

La solution de ce problème revient à considérer $(t^{(1)}, \dots, t^{(H)})$ comme des vecteurs propres normés successifs de la matrice $\sum_k P_{X_k}$. Par la suite, K composantes partielles associées aux tableaux X_k peuvent être exhibées par $t_k^{(h)} = P_{X_k} t^{(h)} / \|P_{X_k} t^{(h)}\|$.

Soit à présent un tableau de variables quantitatives X partitionné en K blocs $X = [X_1 | \dots | X_K]$ mesurés sur N individus, chaque bloc contenant P_k variables ($k = 1, \dots, K$) et un tableau Y contenant Q variables quantitatives à expliquer, mesurées sur les mêmes N individus. Les objectifs de la méthode sont de décrire les liens entre l'ensemble des variables et d'expliquer le tableau Y à partir des K tableaux X_k ($k = 1, \dots, K$). Pour cela, nous cherchons désormais une composante u , contrainte à être la combinaison linéaire des variables Y , définie par $u = Yv$. Cette composante est projetée sur chaque sous-espace engendré par les variables de X_k . Nous souhaitons qu'en moyenne, les variances restituées par ces projections soient maximales. Ainsi, la solution d'ordre un revient à maximiser le critère (2).

$$\text{Maximiser } \sum_{k=1}^K \text{var}(P_{X_k} u^{(1)}) = \frac{1}{N} u^{(1)'} \sum_k P_{X_k} u^{(1)} = \frac{1}{N} v^{(1)'} Y' P_{X_k} Y v^{(1)} \quad (2)$$

En adoptant la contrainte de norme $\|v^{(1)}\| = 1$, il s'ensuit que $v^{(1)}$ est un vecteur propre de la matrice $1/N(Y' \sum_k P_{X_k} Y)$ associé à la plus grande valeur propre, ce qui conduit à la solution de l'ACPVI multibloc (2). Des composantes partielles associées à chaque tableau X_k sont obtenues à partir des projections de $u^{(1)}$ sur chaque espace engendré par les variables de X_k : $t_k^{(1)} = P_{X_k} u^{(1)} / \|P_{X_k} u^{(1)}\|$. Nous pouvons montrer que ces composantes sont liées de manière optimale au tableau Y , en ce sens qu'elles maximisent le critère (3).

$$\sum_{k=1}^K \text{cov}^2(u^{(1)}, t_k^{(1)}) \quad \text{avec } t_k^{(1)} = X_k w_k^{(1)}, \quad u^{(1)} = Y v^{(1)}, \quad \|t_k^{(1)}\| = \|v^{(1)}\| = 1 \quad (3)$$

Une composante globale associée au tableau concaténé X est donnée par $t^{(1)} = \sum_k a_k^{(1)} t_k^{(1)}$, sous la contrainte $\sum_k a_k^{(1)2} = 1$, avec les pondérations $a_k^{(1)}$ données par l'équation (4).

$$a_k^{(1)} = \frac{\|P_{X_k} u^{(1)}\|}{\sqrt{\sum_l \|P_{X_l} u^{(1)}\|^2}} = \frac{\text{cov}(u^{(1)}, t_k^{(1)})}{\sqrt{\sum_l \text{cov}^2(u^{(1)}, t_l^{(1)})}} \quad (4)$$

Ainsi, les coefficients $a_k^{(1)}$ reflètent le lien entre le tableau Y et les tableaux X_k . Nous montrons que la composante $t^{(1)}$ est liée de façon optimale au tableau Y car elle est également solution du problème consistant à maximiser le critère (5).

$$\text{cov}^2(u^{(1)}, t^{(1)}) \quad \text{avec } t^{(1)} = \sum_{k=1}^K a_k^{(1)} t_k^{(1)}, \quad t_k^{(1)} = X_k w_k^{(1)}, \quad u^{(1)} = Y v^{(1)}, \quad (5)$$

$$\sum_{k=1}^K a_k^{(1)2} = 1, \quad \|t_k^{(1)}\| = \|v^{(1)}\| = 1$$

Il faut souligner que la composante globale t , résumé du tableau concaténé $X = [X_1 | \dots | X_K]$ et synthèse des K composantes partielles t_k qui constituent elles-mêmes des résumés des tableaux X_k ($k = 1, \dots, K$), joue un rôle important, aussi bien pour la description des données que pour la prédiction. Il est préconisé de déterminer la solution d'ordre suivant en effectuant la même démarche en remplaçant les tableaux X_k par leurs résidus respectifs de la régression sur la première composante globale $t^{(1)}$. Cette procédure est répétée plusieurs fois pour obtenir des composantes globales ($t^{(1)}, \dots, t^{(H)}$) ainsi que les composantes partielles associées. Les composantes globales ainsi obtenues peuvent servir à des fins de prédiction, en régressant les variables Y sur celles-ci. Cette procédure de déflation conduit à l'obtention de composantes globales, résumés de l'ensemble des variables explicatives, mutuellement orthogonales, ce qui améliore la qualité de prédiction du modèle (25).

2.2 ACPVI multibloc qualitative

2.2.1 Données et notations

Soit la variable à expliquer y , une variable qualitative, mesurée sur N individus, comportant Q modalités. Par la suite, y est codée en un tableau Y , de taille $(N \times Q)$ contenant les indicatrices de ses modalités. Soit le tableau des variables qualitatives explicatives constitué par K variables (x_1, \dots, x_K) mesurées sur les mêmes N individus. Chacune de ces variables x_k , comportant P_k modalités, est codée en un tableau X_k , de taille $(N \times P_k)$ contenant les indicatrices. Chaque tableau X_k constitue à présent un bloc de variables, et de ce fait le tableau concaténé $X = [X_1 | \dots | X_K]$ est partitionné en K blocs. Les tableaux $D_k = X_k'X_k$ (resp. $D_Y = Y'Y$) sont les matrices diagonales dont les éléments sont les effectifs des modalités associées à la variable x_k (resp. y). Nous désignons par $P_{X_k} = X_k D_k^{-1} X_k'$ (resp. $P_Y = Y D_Y^{-1} Y'$) le projecteur associé au sous-espace engendré par les indicatrices des modalités de x_k (resp. y).

2.2.2 Solution d'ordre un

Les objectifs de l'ACPVI multibloc qualitative sont de décrire les liens entre l'ensemble des variables et d'expliquer la variable y à partir des variables (x_1, \dots, x_K) . Pour cela, la méthode cherche une composante u , contrainte à être une combinaison linéaire des indicatrices des modalités de y et liée de manière optimale aux différentes variables (x_1, \dots, x_K) . Nous supposons que le tableau Y est préalablement standardisé : $\tilde{Y} = Y D_Y^{-1/2}$. Ce choix est désormais courant dans le cadre du traitement des variables qualitatives et permet de limiter l'impact de la taille des groupes qui peut être différente selon la modalité. La composante u , donnée par $u = Y D_Y^{-1/2} v$, est projetée sur chaque sous-espace engendré par les indicatrices des modalités de x_k . Nous souhaitons qu'en moyenne, les normes de ces projections soient les plus grandes possibles. De manière formelle, nous souhaitons déterminer la composante u de manière à maximiser le critère (6).

$$\text{Max. } \sum_{k=1}^K \|P_{X_k} u^{(1)}\|^2 = u^{(1)'} \sum_k P_{X_k} u^{(1)} = v^{(1)'} D_Y^{-1/2} Y' \sum_k P_{X_k} Y D_Y^{-1/2} v^{(1)} \quad (6)$$

En adoptant la contrainte de norme $\|v^{(1)}\| = 1$, il s'ensuit que $v^{(1)}$ est un vecteur propre normé, associé à la plus grande valeur propre $\lambda^{(1)}$, de la matrice $D_Y^{-1/2} Y' \sum_k P_{X_k} Y D_Y^{-1/2}$. Il découle

l'équation (7).

$$D_Y^{-1/2} Y' \sum_k P_{X_k} Y D_Y^{-1/2} v^{(1)} = \lambda^{(1)} v^{(1)} \quad (7)$$

De là, nous pouvons déduire :

$$\begin{aligned} Y D_Y^{-1} Y' \sum_k P_{X_k} Y D_Y^{-1/2} v^{(1)} &= \lambda^{(1)} Y D_Y^{-1/2} v^{(1)} \\ \Leftrightarrow P_Y \sum_k P_{X_k} u^{(1)} &= \lambda^{(1)} u^{(1)} \end{aligned} \quad (8)$$

Ainsi, la composante $u^{(1)}$ est un vecteur propre de la matrice $P_Y \sum_k P_{X_k}$ associé à la valeur propre $\lambda^{(1)}$. En réalité, il est facile de vérifier que la première solution ainsi déterminée est proportionnelle au vecteur $\mathbf{1}$ dont toutes les coordonnées sont égales à un. Cette solution triviale est par conséquent écartée et nous retenons le vecteur propre de la matrice $P_Y \sum_k P_{X_k}$ associé à la deuxième plus grande valeur propre. En remarquant que $P_Y u^{(1)} = u^{(1)}$, nous en déduisons que les vecteurs propres de la matrice $P_Y \sum_k P_{X_k}$ sont également vecteurs propres de la matrice symétrique $P_Y \sum_k P_{X_k} P_Y$. Il en découle que la composante $u^{(1)}$ retenue est centrée car orthogonale au vecteur $\mathbf{1}$, également vecteur propre de cette matrice.

Du problème de maximisation (6), nous avons la relation (9).

$$\lambda^{(1)} = \sum_k \|P_{X_k} u^{(1)}\|^2 = N \sum_k \text{var}(P_{X_k} u^{(1)}) \quad (9)$$

Dans cette expression, $\text{var}(P_{X_k} u^{(1)})$ représente la variance inter-groupe de $u^{(1)}$, les groupes étant associés à chacune des variables x_k . Rappelons ici que $u^{(1)}$ représente elle-même un codage de y , en ce sens que les individus ayant la même modalité de y sont associés à une même valeur. Ainsi, il apparaît que la méthode d'analyse cherche à discriminer les groupes formés par le croisement des modalités de y d'un côté et des modalités de chaque variable x_k de l'autre.

Des composantes partielles $(t_1^{(1)}, \dots, t_K^{(1)})$ associées à chaque tableau X_k peuvent être exhibées par projection de la composante $u^{(1)}$ sur les sous-espaces engendrés par les indicatrices des modalités de x_k selon l'équation (10).

$$t_k^{(1)} = \frac{P_{X_k} u^{(1)}}{\|P_{X_k} u^{(1)}\|} \quad \text{pour } k = (1, \dots, K) \quad (10)$$

En reprenant les équations (8) et (10), il s'ensuit l'égalité (11) qui fournit un lien entre les composantes partielles $(t_1^{(1)}, \dots, t_K^{(1)})$ et la composante $u^{(1)}$.

$$\sum_k \|P_{X_k} u^{(1)}\| P_Y t_k^{(1)} = \lambda^{(1)} u^{(1)} \quad (11)$$

Dans un objectif de synthèse, nous allons rechercher à présent une composante globale $t^{(1)}$, synthèse des composantes partielles $(t_1^{(1)}, \dots, t_K^{(1)})$. Cette composante est définie par $t^{(1)} = \sum_k a_k^{(1)} t_k^{(1)}$. De l'équation (4), nous savons que les coefficients $a_k^{(1)}$ sont donnés par :

$$a_k^{(1)} = \frac{\|P_{X_k} u^{(1)}\|}{\sqrt{\sum_l \|P_{X_l} u^{(1)}\|^2}}$$

Ainsi, les coefficients $a_k^{(1)}$ sont d'autant plus grands que la composante partielle $t_k^{(1)}$ est liée à $u^{(1)}$, en ce sens que la variance de $u^{(1)}$ expliquée par $t_k^{(1)}$ est grande.

Nous pouvons ainsi, à partir des équations (4), (9) et (10), déterminer une relation directe entre $t^{(1)}$ et $u^{(1)}$, donnée par l'équation (12).

$$t^{(1)} = \sum_{k=1}^K \frac{\|P_{X_k} u^{(1)}\|}{\sqrt{\sum_l \|P_{X_l} u^{(1)}\|^2}} t_k^{(1)} = \frac{\sum_k P_{X_k} u^{(1)}}{\sqrt{\sum_l \|P_{X_l} u^{(1)}\|^2}} = \frac{1}{\sqrt{\lambda^{(1)}}} \sum_k P_{X_k} u^{(1)} \quad (12)$$

Des égalités (8) et (12), il s'ensuit l'égalité (13) qui donne une autre vision du lien entre les composantes $t^{(1)}$ et $u^{(1)}$.

$$P_Y t^{(1)} = \frac{1}{\sqrt{\lambda^{(1)}}} P_Y \sum_{k=1}^K P_{X_k} u^{(1)} = \sqrt{\lambda^{(1)}} u^{(1)} \quad (13)$$

En utilisant les équations (12) et (13), une autre propriété de la composante $t^{(1)}$ peut être donnée par :

$$\sum_{k=1}^K P_{X_k} P_Y t^{(1)} = \sqrt{\lambda^{(1)}} \sum_k P_{X_k} u^{(1)} = \sum_k P_{X_k} P_Y t^{(1)} = \lambda^{(1)} t^{(1)} \quad (14)$$

Ainsi, $t^{(1)}$ est un vecteur propre de la matrice $\sum_k P_{X_k} P_Y$ associé à la valeur propre $\lambda^{(1)}$.

2.2.3 Solutions d'ordres supérieurs et modèle de prédiction

Notre objectif étant aussi bien descriptif que prédictif, nous choisissons, comme dans le cadre quantitatif détaillé dans le paragraphe 2.1, de déterminer les composantes t , résumés de l'ensemble des variables explicatives de manière à ce qu'elle soient mutuellement orthogonales, afin d'améliorer la qualité de prédiction du modèle (25). La démarche décrite pour la solution d'ordre un est effectuée à nouveau en remplaçant les tableaux (X_1, \dots, X_K) par leurs résidus respectifs de la régression sur la première composante globale $t^{(1)}$. Cette procédure est répétée plusieurs fois pour obtenir les composantes suivantes $(t^{(1)}, \dots, t^{(H)})$.

La qualité de prédiction de la méthode, ainsi que le choix du nombre de composantes à retenir dans le modèle, sont évalués par la qualité de la règle de classement issue du modèle, mesurée par la probabilité d'affectation correcte de nouveaux individus, au vu des valeurs prises par leurs variables explicatives. De nombreuses méthodes d'évaluation de la qualité sont possibles (6). Les techniques de ré-échantillonnage offrent une solution qui permet d'évaluer le taux apparent d'individus bien classés en se basant sur l'échantillon de calibration, mais aussi le taux théorique d'individus bien classés calculé sur l'échantillon de validation. Ces taux représentent les pourcentages d'individus bien classés par le modèle, donnés soit pour chaque classe soit pour l'ensemble des classes. Un bon modèle est celui qui, pour un nombre relativement restreint de composantes, donne des taux apparents et théoriques élevés. La règle d'affectation utilisée est basée sur le calcul des distances aux centres de gravité associés aux Q modalités de y , sur la base des coordonnées factorielles des composantes t . Ces distances sont définies, pour chaque individu, chaque modalité de la variable à expliquer $q = (1, \dots, Q)$ et pour des modèles comportant $h = (1, \dots, H)$ composantes. Les individus, pour chaque modèle comportant h composantes, sont ensuite affectés à la classe associée à la plus petite distance.

2.3 Méthodes alternatives

L'ACPVI multibloc qualitative peut tout d'abord être comparée à l'analyse des correspondances et ses extensions. L'analyse des correspondances multiples, ACM (1; 10) permet de décrire et synthétiser de façon optimale les variables explicatives qualitatives (x_1, \dots, x_K) , les indicatrices de la variable à expliquer y pouvant être projetées en tant que variables supplémentaires sur le même espace. Afin de discriminer les classes induites par la variable y à partir des variables (x_1, \dots, x_K) , une analyse discriminante est réalisée, en remplaçant le tableau X par les composantes d'ACM issues de ce tableau et sélectionnées au préalable. Cette méthode nommée *Disqual* (21) est simple et intéressante. Cependant elle présente les limites classiques de la régression sur composantes principales (18), à savoir que les premières composantes qui reflètent bien les liaisons entre les variables explicatives ne sont pas nécessairement les plus liées à la variable à expliquer. L'analyse discriminante barycentrique (17; 21; 16; 5) lève la première limite en décrivant l'ensemble des variables dans un même espace grâce à l'analyse des correspondances du tableau de contingence $X'Y$, croisant l'ensemble des modalités codées sous forme disjonctive des variables explicatives X et à expliquer Y . L'étape prédictive associée peut être donnée par la régression *PLS* après ACM (4) et aboutit à une solution comparable à celle de l'ACPVI multibloc qualitative.

L'ACPVI multibloc qualitative a des similarités marquantes avec l'Analyse des Correspondances Multiples avec un Tableau de Référence, ACMTR (12) et avec l'Analyse Non-Symétrique des Correspondances Multiples, ANSCM (14; 13; 15). En effet, la démarche préconisée pour l'ACMTR se place dans le cadre d'une analyse canonique généralisée sous contrainte, *i.e.* l'Analyse Canonique Généralisée avec un Tableau de Référence (12), alors que la démarche de l'ACPVI multibloc diffère dans le sens où elle privilégie la restitution de la variabilité du tableau Y . Cependant, dans le cas particulier des variables qualitatives et du fait de la standardisation des indicatrices des modalités de y adoptée, les solutions d'ordre un de l'ACPVI multibloc qualitative et de l'ACMTR sont identiques. Aux étapes d'ordre suivant, l'ACMTR adopte une stratégie de déflation à l'intérieur de chaque bloc de modalités, *i.e.* la même démarche est réitérée en remplaçant les tableaux initiaux par leur projection sur l'orthogonal de l'espace engendré par les composantes partielles associées respectivement à ces tableaux. Dans le cadre de l'ACPVI multibloc qualitative, une stratégie de déflation par rapport aux composantes globales est adoptée, ce qui conduit à privilégier l'aspect prédictif (25). De plus, les composantes globales étant orthogonales, elles permettent d'établir des représentations synthétiques des relations entre les variables associées aux différents tableaux.

Un lien entre l'ACPVI multibloc qualitative et l'analyse non-symétrique des correspondances étendue au cas de tableaux de contingence multiples (ANSCM) peut aussi être mis en évidence. Les solutions d'ordre un, *i.e.* premières composantes associées à la variable y et au tableau concaténé regroupant toutes les modalités des variables (x_1, \dots, x_K) de l'ANSCM, peuvent être retrouvées en adoptant la même démarche que celle de l'ACPVI multibloc qualitative, à condition de pas standardiser le tableau des modalités de y . Pour les étapes suivantes, l'ANSCM ne préconise pas de déflations mais impose des contraintes d'orthogonalité aux composantes u associées à la variable y . Ces contraintes d'orthogonalité ne se traduisent pas par une orthogonalité des composantes globales t .

3 Application

3.1 Données d'épidémiologie vétérinaire

Les données d'épidémiologie vétérinaire sont issues d'une enquête analytique portant sur les pathologies respiratoires du porc à l'engrais. Le jeu de données qui en est issu est plus particulièrement orienté vers l'étude des facteurs associés à la pneumonie. Le tableau comporte 105 élevages sur lesquels sont mesurées, après sélection, 19 variables qualitatives, soit 18 variables explicatives et une variable à expliquer. Dans chaque élevage, un lot d'animaux est étudié. La variable à expliquer (*PNEU*), qui comporte 3 modalités, mesure la sévérité de la pneumonie du lot. Cette sévérité est mesurée par une catégorisation en trois classes établie à partir de l'observation de 30 poumons de porcs d'un même lot observés à l'abattoir. Le premier objectif de l'étude est de décrire les liens entre les variables explicatives et la variable à expliquer. Le second objectif est à la fois de déterminer, parmi les variables de X , celles qui sont facteurs de risque ou, au contraire, facteurs protecteurs de la pneumonie. Les variables et leurs modalités sont décrites dans la Table 1.

TAB. 1 – Description des variables et de leurs modalités

Variables	Description
<i>PNEU</i>	Note médiane de pneumonie du lot observé à l'abattoir (1=note \leq 0,25, 2=0,25 <note \leq 3,75, 3=note $>$ 3,75)
<i>Pedi</i>	Présence de pédiluves à l'entrée de chaque secteur de l'élevage (1=non, 2=oui)
<i>Surf</i>	Surface par porc en fin d'engraissement, en $m^2/porc$ (1=surf. \leq 0.95, 2=surf. $>$ 0.95)
<i>NbPorc</i>	Nombre de porcs en fin d'engraissement dans la salle observée (1=nb. \leq 90, 2=nb. $>$ 90)
<i>Comble</i>	Existence de combles dans la salle d'engraissement (1=non, 2=oui)
<i>Chauff</i>	Préchauffage des salles d'engraissement avant l'entrée des porcs en hiver (1=non, 2=oui)
<i>AirEng</i>	Type d'entrée d'air dans la salle d'engraissement (1=plafond diffuseur, 2=autre)
<i>NbVent</i>	Nombre de ventilateurs dans la salle d'engraissement (1=inférieur ou égal à 1, 2=plus d'un)
<i>OrAirMat</i>	Origine de l'air incorporé dans les salles de maternité (1=extérieur, 2=combles, 3=couloir)
<i>AirMat</i>	Type d'entrée d'air dans les salles de maternité (1=plafond diffuseur, 2=autre)
<i>VentMat</i>	Plage de ventilation en maternité l'hiver (1=plage \leq 4°C, 2=plage $>$ 4°C)
<i>Dent</i>	Mode de soin des dents des porcelets (1=pas, 2=meulage, 3=épointage)
<i>SevPrec</i>	Réalisation d'un sevrage précoce (1=non, 2=oui)
<i>IntBand</i>	Intervalles entre les bandes (2=int. \leq 2 sem., 3=int. de 3 sem., 4=int. \geq 4 sem.)
<i>Tremp</i>	Trempage des salles de post-sevrage avant le lavage (1=non, 2=oui)
<i>DetPs</i>	Application d'un détergent après lavage dans les salles de post-sevrage (1=non, 2=oui)
<i>DetEng</i>	Application d'un détergent après lavage dans les salles d'engraissement (1=non, 2=oui)
<i>Detass</i>	Détassage des porcs au cours de la période d'engraissement (1=non, 2=oui)
<i>Appro</i>	Nombre d'élevages approvisionnant pour les truies reproductrices (1=nb.= 1, 2=nb. $>$ 1)

3.2 Description des individus et des liens entre variables

La Table 2 donne les valeurs propres $\lambda^{(h)}$ de l'ACPVI multibloc qualitative. Rappelons que ces valeurs propres reflètent la qualité de discrimination de la variable y par les composantes partielles ou la composante globale. La Table 2 donne également les valeurs propres associées à l'ACM. Les deux séries de valeurs propres ne sont pas nécessairement comparables car elles concernent *a priori* des problématiques différentes. Les deux premières composantes

de l'ACPVI multibloc qualitative sont suffisantes pour séparer les modalités de la variable y (86.0% de la variance inter-groupe de y est restituée).

TAB. 2 – Valeurs propres $\lambda^{(h)}$ et pourcentage cumulé des valeurs propres pour les cinq premières dimensions h

Dimension	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$...	Total
$\lambda^{(h)}$	0,550	0,373	0,063	0,052	0,016	...	1,072
$\%_{cum.}\lambda^{(h)}$	51,3%	86,0%	92,0%	96,8%	98,3%	...	100%

(a) ACPVI multibloc qualitative

Dimension	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$...	Total
$\lambda^{(h)}$	0,134	0,113	0,099	0,083	0,079	...	1,111
$\%_{cum.}\lambda^{(h)}$	12,1%	22,3%	31,2%	38,7%	45,8%	...	100%

(b) ACM

La Figure 1 donne l'importance de chaque variable explicative dans la construction des composantes globales $t^{(h)}$. Cette importance est reflétée par les coefficients $a_k^{(h)^2}$. Ces coefficients peuvent ainsi être utilisés pour sélectionner les variables explicatives permettant de séparer au mieux les modalités de la variable à expliquer. La première composante $t^{(1)}$ de l'ACPVI multibloc qualitative est principalement le reflet du nombre de porcs en fin d'engraissement (*NbPorc*) et du mode de soin des dents des porcelets (*Dent*). La seconde composante $t^{(2)}$ est principalement déterminée par l'origine de l'air incorporé dans les salles de maternité (*OrAirMat*).

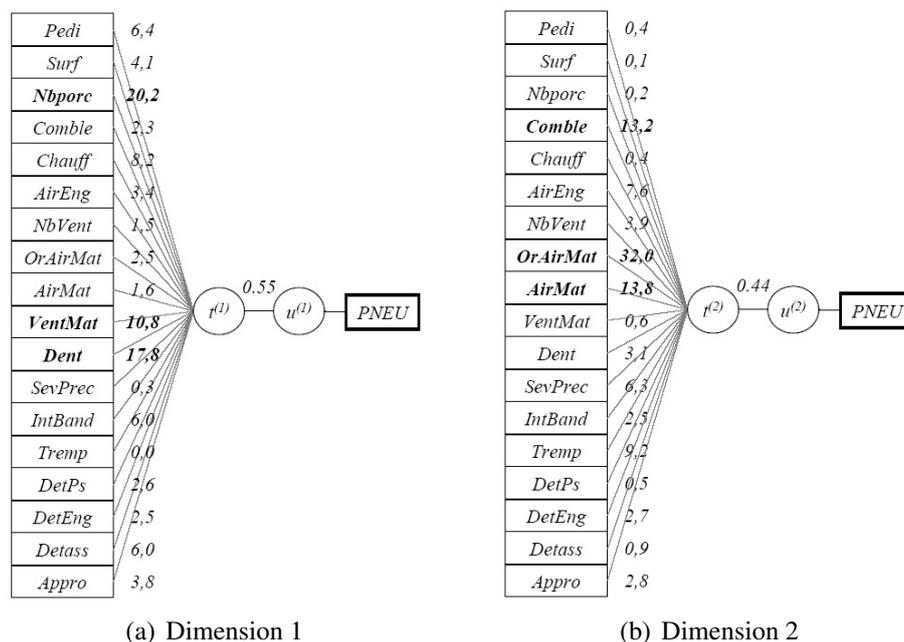


FIG. 1 – Contribution relative (coefficients $a_k^{(h)^2} \times 100$) des variables explicatives dans la construction des composantes $t^{(1)}$ et $t^{(2)}$ de l'ACPVI multibloc qualitative. Le lien entre les composantes t et u est donné par leur coefficient de corrélation

Les modalités des variables (x_1, \dots, x_K) et y peuvent être représentées sur le plan des composantes $t^{(h)}$, qui sont par construction mutuellement orthogonales. La Figure 2 illustre cette

représentation pour les deux premières dimensions $t^{(1)}$ et $t^{(2)}$. La modalité ($PNEU = 1$), correspondant aux élevages faiblement atteints par la pneumonie, est surtout associée au fait de ne pas ép pointer les dents des porcelets ($Dent = 1$) et à un nombre réduit de porcs en fin d'engraissement dans la salle ($NbPorc = 1$). La modalité ($PNEU = 2$), qui représente les élevages moyennement affectés par la pneumonie, est principalement liée à de l'air incorporé dans les salles de maternité provenant des combles ($OrAirMat = 2$) et à une plage de ventilation en maternité l'hiver supérieure à $4^{\circ}C$ ($Ventmat = 2$). La modalité ($PNEU = 3$), qui représente les élevages les plus touchés par la pneumonie, est associée à de l'air incorporé dans les salles de maternité ne provenant pas des combles ($OrAirMat = 1 - 3$), à une entrée d'air dans les salles de maternité ne provenant pas d'un plafond diffuseur ($AirMat = 2$) et au fait de ne pas tremper les salles de post-sevrage avant le lavage ($Tremp = 1$). Il est intéressant d'associer à la représentation factorielle des variables, celle des individus (élevages) pour le même plan (Figure 3). Ces individus sont labellisés selon leur appartenance aux modalités de la variable à expliquer, *i.e.* la sévérité de la pneumonie du lot de porcs. La coloration du plan factoriel illustre les distances entre chaque point du plan et les centres de gravités des classes, ce qui permet de mieux appréhender les séparations entre classes. Afin d'appuyer l'interprétation, la significativité du pouvoir discriminant des composantes $t^{(h)}$, pour chaque dimension h , est donnée par un modèle d'analyse de variance à un facteur, où l'objectif est de tester l'influence de la variable qualitative y sur chaque composante $t^{(h)}$ (8). Un test de comparaison multiple de moyennes de Newman & Keuls détermine les modalités de y significativement différentes des autres (7). Il s'ensuit que la composante $t^{(1)}$ différencie de façon significative les élevages les moins atteints ($PNEU = 1$) des autres et la composante $t^{(2)}$ les élevages fortement atteints ($PNEU = 3$) des autres. Les composantes suivantes ne différencient pas significativement les trois groupes d'élevages.

3.3 Discrimination qualitative

Le nombre optimal de composantes à retenir dans le modèle de prédiction de y par les variables (x_1, \dots, x_K) est évalué par validation croisée. Cette procédure, répétée ($m = 500$) fois, est basée sur un échantillon de calibration comprenant $2/3$ des individus et un échantillon de validation comprenant le $1/3$ restant. La Figure 4 illustre l'évolution des taux apparent et théorique d'individus bien classés pour l'ensemble des classes, selon le nombre de composantes $(t^{(1)}, \dots, t^{(H)})$ introduites dans le modèle, pour l'ACPVI multibloc qualitative et l'ACM associée à la procédure *Disqual*. Le nombre optimal de composantes à retenir dans le modèle est celui qui fournit un bon compromis entre qualité de modélisation et de prédiction. Pour les deux méthodes, un modèle comportant deux composantes semble approprié. Pour l'ACPVI multibloc qualitative, ce modèle permet d'obtenir 65.2% d'individus bien classés au total lors de la modélisation et 52.8% lors de la prédiction. L'ACM associée à la procédure *Disqual* donne de moins bons résultats en terme de qualité de modélisation (52.3% d'individus bien classés au total) et de prédiction (48.9% d'individus bien classés).

Pour un modèle optimal comportant deux composantes, il est possible de préciser pour chaque classe le pourcentage d'individus bien classés, moyenné sur l'ensemble des simulations. Les résultats, pour l'ACPVI multibloc qualitative et l'ACM – *Disqual*, sont reportés dans la Table 3, à la fois pour la qualité de modélisation et de prédiction. Pour l'ACPVI multibloc qualitative, les élevages ayant la modalité ($PNEU = 2$) sont à la fois les mieux expliqués et les mieux prédits par le modèle, soit 69.1% pour la modélisation et 79.5% pour la prédiction. Ceci pourrait s'expliquer par le fait que cette modalité joue un rôle central car elle est intermédiaire

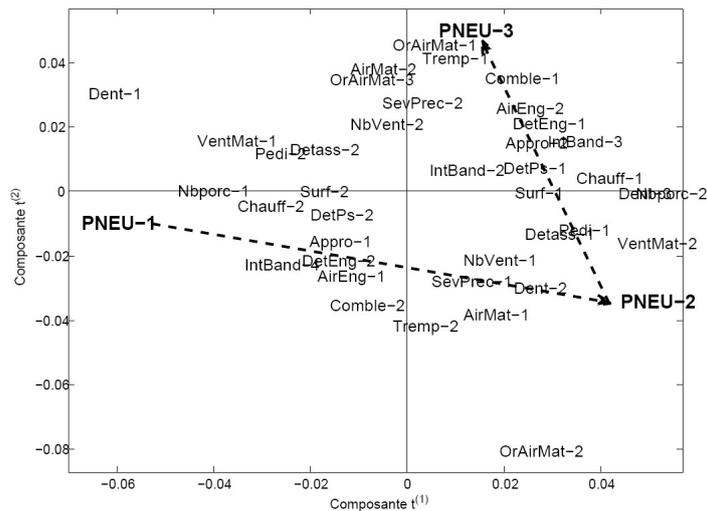


FIG. 2 – Représentation factorielle des modalités des variables (x_1, \dots, x_K) et y sur le plan des composantes $t^{(1)}$ et $t^{(2)}$.

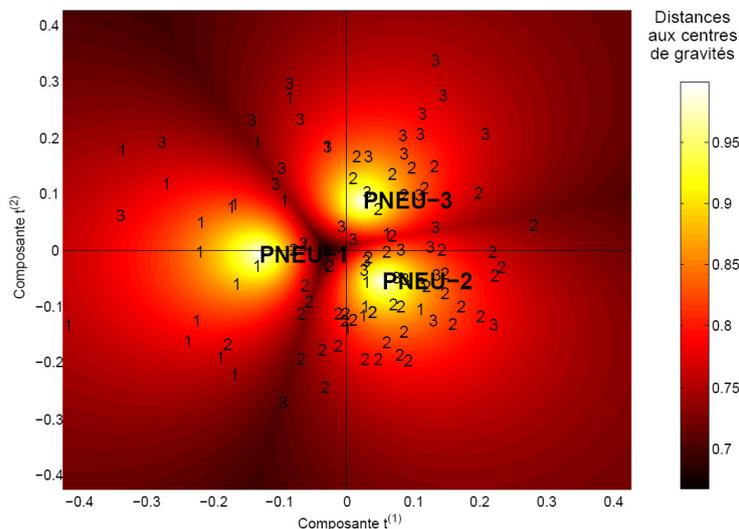


FIG. 3 – Représentation factorielle des individus sur le plan des composantes $t^{(1)}$ et $t^{(2)}$. Les individus sont labellisés selon leur appartenance aux modalités de la variable à expliquer. La coloration du plan reflète les distances de chaque point du plan au centre de gravité le plus proche

entre les deux autres modalités. Ce constat est corroboré par le fait que pour les individus mal classés des modalités ($PNEU = 1$) et ($PNEU = 3$), la confusion est plus grande avec les individus de cette modalité intermédiaire que les uns avec les autres. Dans le cas de l'ACM – *Disqual*, la qualité de modélisation et de prédiction des trois groupes est plus équilibrée, mais les confusions entre classes éloignées sont relativement plus importantes que pour l'ACPVI multibloc qualitative.

Les coefficients de régression des modalités des variables explicatives associés aux différentes classes sont calculés, pour le nombre optimal de dimensions du modèle. Les écart-types et intervalles de confiance à 95% associés à chacun de ces coefficients sont aussi calculés à partir des ($m = 500$) simulations de la procédure de validation croisée. L'ensemble de ces résultats est

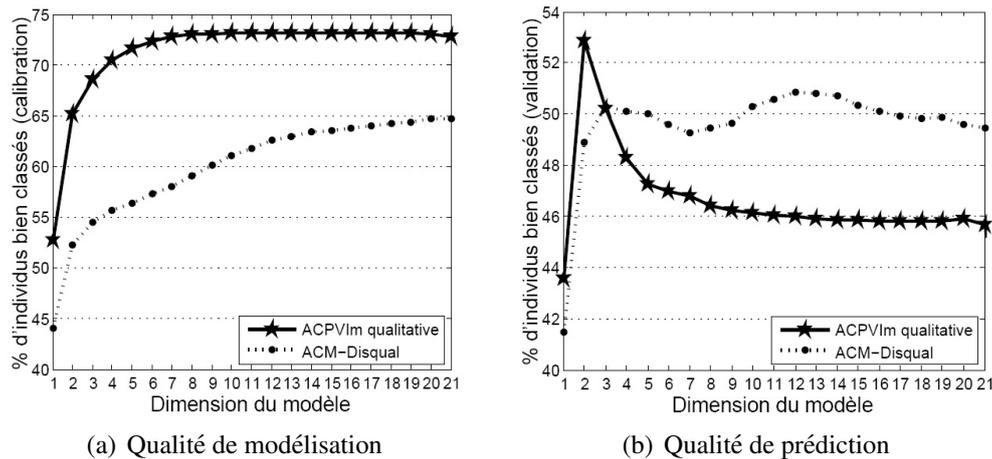


FIG. 4 – Evolution de la qualité de modélisation et de prédiction du modèle en fonction du nombre de composantes t introduites dans le modèle, sur la base du pourcentage total d’individus bien classés. Comparaison des résultats de l’ACPVI multibloc qualitative et de l’ACM – Disqual

reporté dans la Table 4. Une synthèse est donnée, sous forme de recommandations aux éleveurs pour limiter les risques de pneumonie des porcs à l’engrais :

- Règles d’hygiène : utiliser un pédiluve à l’entrée de chaque secteur de l’élevage, tremper les salles de post-sevrage avant le lavage,
- Conduite générale d’élevage : ne pas épointer les dents des porcelets, limiter la densité des porcs en fin d’engraissement en pratiquant par exemple le détassage, respecter un intervalle de temps suffisant entre les bandes (plutôt 4 semaines que 3) et limiter le nombre d’élevages approvisionnant en truies reproductrices,
- En maternité : utiliser l’air provenant des combles, le faire arriver par un plafond diffuseur et régler la plage de ventilation à moins de 4°C l’hiver,
- En engraissement : introduire l’air par des combles et un plafond diffuseur, préchauffer les salles avant l’entrée des porcs en hiver.

4 Conclusion et perspectives

Les méthodes étudiées dans cet article se placent dans le cadre de l’explication d’une variable qualitative y par un ensemble de variables qualitatives (x_1, \dots, x_K) . Nous appliquons l’ACPVI multibloc au cadre qualitatif, pour lequel chaque bloc est constitué par l’ensemble des modalités d’une variable qualitative codée de façon disjonctive. Ainsi, l’application de méthodes multiblocs, initialement développées pour le traitement de données quantitatives organisées en $(K + 1)$ tableaux, ouvre sur de nouvelles voies de recherche pour les problématiques de discrimination qualitative. L’avantage direct de ce type d’application est que l’ensemble des modalités d’une variable constitue un bloc, ce qui permet une prise en compte de manière appropriée de la structure des données et procure des aides à l’interprétation de nature à aider l’utilisateur, *i.e.* composantes globales, composantes par bloc, coefficients reflétant l’importance de chaque bloc dans la discrimination. Nous montrons que la solution d’ordre un de cette méthode est celle de l’Analyse des Correspondances Multiples avec un Tableau de Référence (12). Les solutions d’ordre suivant permettent d’optimiser l’aspect prédictif de la méthode, et reprennent

TAB. 3 – Matrices de confusion pour l'échantillon de calibration (qualité de modélisation) et de validation (qualité de prédiction) données par l'ACPVI multibloc qualitative et l'ACM – *Disqual*, pour un modèle comportant deux composantes

	$PNEU_{pr.} = 1$	$PNEU_{pr.} = 2$	$PNEU_{pr.} = 3$	Total
$PNEU_{obs.} = 1$	65,7	20,8	13,5	100%
$PNEU_{obs.} = 2$	10,0	69,1	20,8	100%
$PNEU_{obs.} = 3$	11,1	30,0	58,9	100%

(a) Qualité de modélisation de l'ACPVI multibloc qualitative

	$PNEU_{pr.} = 1$	$PNEU_{pr.} = 2$	$PNEU_{pr.} = 3$	Total
$PNEU_{obs.} = 1$	27,1	53,6	19,3	100%
$PNEU_{obs.} = 2$	2,9	79,5	17,5	100%
$PNEU_{obs.} = 3$	8,9	55,8	35,3	100%

(b) Qualité de prédiction de l'ACPVI multibloc qualitative

	$PNEU_{pr.} = 1$	$PNEU_{pr.} = 2$	$PNEU_{pr.} = 3$	Total
$PNEU_{obs.} = 1$	50,2	29,6	20,2	100%
$PNEU_{obs.} = 2$	19,8	53,0	27,2	100%
$PNEU_{obs.} = 3$	15,7	31,3	52,9	100%

(c) Qualité de modélisation de l'ACM – *Disqual*

	$PNEU_{pr.} = 1$	$PNEU_{pr.} = 2$	$PNEU_{pr.} = 3$	Total
$PNEU_{obs.} = 1$	45,5	31,6	23,0	100%
$PNEU_{obs.} = 2$	21,0	49,1	29,9	100%
$PNEU_{obs.} = 3$	16,3	31,9	51,8	100%

(d) Qualité de prédiction de l'ACM – *Disqual*

les idées de déflations sur une composante résumant l'ensemble des variables explicatives (4).

L'ACPVI multibloc qualitative peut être étendue au cas de plusieurs tableaux Y à expliquer, ce qui permettrait d'expliquer simultanément plusieurs variables qualitatives, sans les synthétiser en une seule variable. Dans le cadre des données d'épidémiologie animale, il serait possible d'expliquer une maladie caractérisée par plusieurs variables qualitatives, ou l'évolution d'une maladie, caractérisée par une seule variable et évoluant au cours du temps. De plus, il faut considérer que des variables externes peuvent avoir une influence sur l'indépendance des individus étudiés. En épidémiologie vétérinaire par exemple, des tirages au sort stratifiés sur certains critères d'importance (la taille de l'élevage par exemple) peuvent être réalisés et il est important d'ajuster le lien entre les variables explicatives et la variable à expliquer sur ces variables. La décomposition de matrices de données selon des informations externes, abordée pour d'autres méthodes d'analyse, pourrait être appliquée (14; 23).

Références

- [1] Benzecri, J.P. : L'analyse des données. Tome 1 : La taxonomie. Tome 2 : L'analyse des correspondances, Dunod, Paris (1973)
- [2] Bougeard, S., Hanafi, M., Qannari, E.M. : ACPVI multibloc. Application à des données d'épidémiologie animale. Journal de la Société Française de Statistique, **148** (2007) 77–94

- [3] Carroll, J. D. : A generalization of canonical correlation analysis to three or more sets of variables. 76th annual convention of the American psychological association (1968) 227–228
- [4] Cazes, P. : Adaptation de la régression *PLS* au cas de la régression après analyse des correspondances multiples. *Revue de statistique appliquée*, **XLV** (1997) 89–99
- [5] Cazes, P., Baumerder, A., Bonnefous, S., Pages, J.P. : Codage et analyse des tableaux logiques - Introduction à la pratique des variables qualitatives. *Cahiers du bureau universitaire de recherche opérationnelle* (1977)
- [6] Celeux, G., Nakache, J.P. : *Analyse discriminante sur variables qualitatives*, Polytechnica, Paris (1994)
- [7] Dagnélie, P. : *Théorie et méthodes statistiques. Applications agronomiques. Les méthodes de l'inférence statistique*, Les presses agronomiques de Gembloux, Gembloux (1975)
- [8] Desbois, D. : Une introduction à l'analyse discriminante avec SPSS pour Windows. *La revue MODULAD*, **30** (2003) 19–49
- [9] Dreesbeke, J.J., Lejeune, M., Saporta, G. : *Modèles statistiques pour données qualitatives*, Technip, Paris (2005)
- [10] Greenacre, M.J. : *Theory and applications of correspondence analysis*, Academic press, London (1984)
- [11] Horst, P. : Relations among m sets of measures. *Psychometrika*, **26** (1961) 129–149
- [12] Kissita, G. : *Les analyses canoniques généralisées avec tableau de référence généralisé : éléments théoriques et appliqués*, Université Paris Dauphine IX, Paris (2003)
- [13] Lauro, C., Balbi, S. : The analysis of structured qualitative data. *Applied Stochastic Models and Data Analysis*, **15** (1999) 1–27
- [14] Lauro, C., D'Ambra, L. : L'analyse non symétrique des correspondances. *Data Analysis and Informatics*, **III** (1984) 433–446
- [15] Lauro, C., Siciliano, R. : *Analyse non symétrique des correspondances pour des tables de contingences. L'analyse des correspondances et les techniques annexes. Approches nouvelles pour l'analyse des données*, Springer Ed., (2000) 183–210
- [16] Leclerc, A. : L'analyse des correspondances sur juxtaposition de tableaux de contingence. *Revue de statistique appliquée*, **23** (1975) 5–16
- [17] Masson, M. : *Processus linéaires et analyse de données non-linéaires*, Université Paris VI, Paris (1974)
- [18] Massy, W. F. : Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, **60** (1965) 234–256
- [19] Rao, C. R. : The use and interpretation of principal component analysis in applied research. *Sankhya, A.* **26** (1964) 329–358

- [20] Sabatier, R. : Analyse factorielle de données structurées et métriques. *Statistique et analyse des données*, **12** (1987) 75–96
- [21] Saporta, G. : Liaisons entre plusieurs ensembles de variables et codage de données qualitatives, Université Paris VI, Paris (1975)
- [22] Toma, B., Dufour, B., Sanaa, M., Benet, J.J., Ellis, P., Moutou, F., Louza, A. : *Epidémiologie appliquée à la lutte collective contre les maladies animales transmissibles majeures*, AEEMA, Maisons Alfort (1996)
- [23] Takane, Y., Shibayama, T. : Principal component analysis with external information on both subjects and variables. *Psychometrika*, **56** (1991) 97–120
- [24] Van Den Wollenberg, A. : Redundancy analysis : an alternative for canonical correlation analysis. *Psychometrika*, **42** (1977) 207–219
- [25] Westerhuis, J.A., Smilde, A.K. : Deflation in multiblock *PLS*. *Journal of chemometrics*, **15** (2001) 485–493

TAB. 4 – Coefficients de régression associés aux modalités des variables explicatives en vue d’expliquer les modalités de la variable y , ainsi que leurs intervalles de confiance à 95% associés, pour un modèle à deux composantes. Les coefficients de régression dont l’intervalle de confiance ne contient pas zéro sont représentés en gras

Variable	$PNEU = 1$	$PNEU = 2$	$PNEU = 3$
<i>Pedi</i> – 1	-0,11 [-0,21 ; 0,00]	0,08 [0,01 ; 0,16]	-0,01 [-0,09 ; 0,07]
<i>Pedi</i> – 2	0,11 [0,00 ; 0,21]	-0,08 [-0,16 ; -0,01]	0,01 [-0,07 ; 0,09]
<i>Surf</i> – 1	-0,08 [-0,15 ; 0,00]	0,04 [-0,02 ; 0,11]	0,02 [-0,05 ; 0,08]
<i>Surf</i> – 2	0,08 [0,00 ; 0,15]	-0,04 [-0,11 ; 0,02]	-0,02 [-0,08 ; 0,05]
<i>Nbporc</i> – 1	0,16 [0,09 ; 0,24]	-0,09 [-0,16 ; -0,03]	-0,03 [-0,11 ; 0,04]
<i>Nbporc</i> – 2	-0,16 [-0,24 ; -0,09]	0,09 [0,03 ; 0,16]	0,03 [-0,04 ; 0,11]
<i>Comble</i> – 1	-0,06 [-0,12 ; 0,00]	-0,04 [-0,09 ; 0,02]	0,10 [0,04 ; 0,16]
<i>Comble</i> – 2	0,06 [0,00 ; 0,12]	0,04 [-0,02 ; 0,09]	-0,10 [-0,16 ; -0,04]
<i>Chauff</i> – 1	-0,12 [-0,17 ; -0,07]	0,06 [-0,02 ; 0,14]	0,04 [-0,05 ; 0,12]
<i>Chauff</i> – 2	0,12 [0,07 ; 0,17]	-0,06 [-0,14 ; 0,02]	-0,04 [-0,12 ; 0,05]
<i>AirEng</i> – 1	0,07 [0,02 ; 0,12]	0,01 [-0,03 ; 0,06]	-0,08 [-0,13 ; -0,03]
<i>AirEng</i> – 2	-0,07 [-0,12 ; -0,02]	-0,01 [-0,06 ; 0,03]	0,08 [0,03 ; 0,13]
<i>NbVent</i> – 1	-0,04 [-0,10 ; 0,03]	0,06 [0,00 ; 0,12]	-0,04 [-0,10 ; 0,01]
<i>NbVent</i> – 2	0,04 [-0,03 ; 0,10]	-0,06 [-0,12 ; 0,00]	0,04 [-0,01 ; 0,10]
<i>OrAirMat</i> – 1	0,00 [-0,11 ; 0,11]	-0,09 [-0,20 ; 0,01]	0,11 [-0,01 ; 0,24]
<i>OrAirMat</i> – 2	-0,05 [-0,15 ; 0,05]	0,19 [0,11 ; 0,27]	-0,19 [-0,29 ; -0,09]
<i>OrAirMat</i> – 3	0,05 [-0,06 ; 0,16]	-0,10 [-0,19 ; 0,00]	0,08 [-0,05 ; 0,20]
<i>AirMat</i> – 1	-0,03 [-0,09 ; 0,03]	0,10 [0,04 ; 0,15]	-0,09 [-0,15 ; -0,03]
<i>AirMat</i> – 2	0,03 [-0,03 ; 0,094]	-0,10 [-0,15 ; -0,04]	0,09 [0,03 ; 0,15]
<i>VentMat</i> – 1	0,15 [0,03 ; 0,26]	-0,12 [-0,20 ; -0,04]	0,01 [-0,08 ; 0,10]
<i>VentMat</i> – 2	-0,15 [-0,26 ; -0,03]	0,12 [0,04 ; 0,20]	-0,01 [-0,10 ; 0,08]
<i>Dent</i> – 1	0,22 [0,10 ; 0,35]	-0,19 [-0,28 ; -0,10]	0,03 [-0,08 ; 0,14]
<i>Dent</i> – 2	-0,07 [-0,1 ; 0,03]	0,10 [0,02 ; 0,18]	-0,06 [-0,14 ; 0,02]
<i>Dent</i> – 3	-0,15 [-0,24 ; -0,06]	0,09 [0,01 ; 0,17]	0,03 [-0,05 ; 0,11]
<i>SevPrec</i> – 1	-0,01 [-0,09 ; 0,06]	0,06 [-0,01 ; 0,13]	-0,07 [-0,13 ; 0,00]
<i>SevPrec</i> – 2	0,01 [-0,06 ; 0,09]	-0,06 [-0,13 ; 0,01]	0,07 [0,00 ; 0,13]
<i>IntBand</i> – 2	-0,02 [-0,13 ; 0,10]	0,00 [-0,10 ; 0,09]	0,02 [-0,08 ; 0,12]
<i>IntBand</i> – 3	-0,10 [-0,19 ; -0,02]	0,03 [-0,06 ; 0,11]	0,06 [-0,03 ; 0,15]
<i>IntBand</i> – 4	0,12 [0,00 ; 0,24]	-0,02 [-0,12 ; 0,07]	-0,08 [-0,18 ; 0,02]
<i>Tremp</i> – 1	-0,02 [-0,11 ; 0,04]	-0,07 [-0,16 ; 0,01]	0,11 [0,01 ; 0,21]
<i>Tremp</i> – 2	0,02 [-0,07 ; 0,11]	0,07 [-0,01 ; 0,16]	-0,11 [-0,21 ; -0,01]
<i>DetPs</i> – 1	-0,07 [-0,15 ; 0,01]	0,03 [-0,05 ; 0,10]	0,03 [-0,05 ; 0,11]
<i>DetPs</i> – 2	0,07 [-0,01 ; 0,15]	-0,03 [-0,10 ; 0,05]	-0,03 [-0,11 ; 0,05]
<i>DetEng</i> – 1	-0,08 [-0,18 ; 0,02]	0,00 [-0,08 ; 0,08]	0,07 [-0,01 ; 0,15]
<i>DetEng</i> – 2	0,08 [-0,02 ; 0,18]	0,00 [-0,08 ; 0,08]	-0,07 [-0,15 ; 0,01]
<i>Detass</i> – 1	-0,08 [-0,14 ; -0,02]	0,07 [0,02 ; 0,13]	-0,02 [-0,08 ; 0,05]
<i>Detass</i> – 2	0,08 [0,02 ; 0,14]	-0,07 [-0,13 ; -0,02]	0,02 [-0,05 ; 0,08]
<i>Appro</i> – 1	0,07 [0,01 ; 0,13]	-0,01 [-0,07 ; 0,05]	-0,05 [-0,12 ; 0,01]
<i>Appro</i> – 2	-0,07 [-0,13 ; -0,01]	0,01 [-0,05 ; 0,07]	0,05 [-0,01 ; 0,12]