



HAL
open science

Les analyses canoniques simple et généralisée linéaires : application à des données psychosociales.

Victor Nzobounsana, Sandrine Gaymard

► **To cite this version:**

Victor Nzobounsana, Sandrine Gaymard. Les analyses canoniques simple et généralisée linéaires : application à des données psychosociales.. *Mathématiques et Sciences Humaines*, 2010, 189 (1), pp.69-101. 10.4000/msh.11653 . hal-03355921

HAL Id: hal-03355921

<https://univ-angers.hal.science/hal-03355921v1>

Submitted on 27 Sep 2021

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.

LES ANALYSES CANONIQUES SIMPLE ET GÉNÉRALISÉE LINÉAIRES : APPLICATIONS À DES DONNÉES PSYCHOSOCIALES

Victor NZOBOUNSA¹, Sandrine GAYMARD²

RÉSUMÉ – *L'analyse des corrélations canoniques est une méthode statistique, proposée en 1936 par Hotelling, surtout connue pour ses qualités théoriques, puisqu'elle englobe de nombreuses autres méthodes statistiques. Elle permet de décrire les relations linéaires qui existent entre deux ensembles de variables mesurées sur les mêmes individus. Plusieurs généralisations de cette méthode, regroupées sous le nom d'analyse canonique généralisée, ont été proposées dans la littérature scientifique pour étudier conjointement les relations linéaires qu'il y a entre plusieurs groupes de variables mesurées sur les mêmes individus. Dans cet article, après avoir décrit l'analyse canonique linéaire simple et généralisée, nous comparons trois méthodes : la méthode Sumcor, la méthode Ssqcor et la méthode de Carroll. Afin d'illustrer l'application de ces méthodes, nous partons d'un questionnaire psychosocial élaboré dans le champ théorique des représentations sociales et portant sur l'étude des conditions de transgressions routières auprès de jeunes conducteurs. Les résultats observés dans cette étude montrent que la méthode d'analyse des corrélations canoniques et ses généralisations sont des outils statistiques adaptés pour décrire et analyser les données recueillies dans le champ des sciences humaines. De plus, dans l'étude des représentations sociales, cette approche permet d'illustrer au travers des « patterns » de réponses, la structuration interne des transgressions légitimes.*

MOTS-CLÉS – Analyse canonique généralisée, Analyse des corrélations canoniques, Conduite automobile, Questionnaire des Scripts Conditionnels (CSQ), Représentations sociales, Transgressions légitimes, Variables canoniques

SUMMARY – Simple and generalized linear canonical analysis: applications to psychosocial data

Canonical correlation analysis is a statistical method proposed by Hotelling in 1936, best known for its theoretical qualities, since it encompasses many other statistical methods. It allows the linear relationships between two sets of variables measured on the same individuals to be described. Several generalizations of this method, grouped under the name Generalized Canonical Analysis, have been proposed in the literature to study jointly the linear relationships existing between several groups of variables measured on the same individuals. In this article, after describing linear canonical analysis and its generalizations, we compare three methods of generalized canonical analysis: the Sumcor method, the Ssqcor method and the Carroll method. To illustrate the application of these methods we start with a psychosocial questionnaire developed in the theory of social representations and concerning the study of disobedience of traffic rules among young drivers. The results observed

¹Laboratoire de psychologie UPRES EA 2646, Université d'Angers, UFR Lettres Langues et Sciences Humaines, 11 boulevard Lavoisier 49045 Angers Cedex 01 et Rectorat de l'Académie de Reims, 1 rue Navier 51082 Reims cedex, nzobounsana@hotmail.com

²Laboratoire de psychologie UPRES EA 2646, Université d'Angers, UFR Lettres Langues et Sciences Humaines, 11 boulevard Lavoisier 49045 Angers Cedex 01, sandrine.gaynard@univ-angers.fr

in this study show that the method of canonical correlation analysis and its generalizations are appropriate statistical tools for describing and analyzing the data collected in the field of social sciences. Furthermore in the study of social representations this approach allows us to illustrate through the “patterns” of answers, the internal structuring of the legitimate transgressions.

KEYWORDS – Canonical correlation analysis, Canonical variables, Conditional Scripts Questionnaire (CSQ), Driving, Generalized canonical analysis, Legitimate transgressions, Social representations

1. INTRODUCTION

L’analyse des corrélations canoniques ou encore analyse canonique simple est une méthode statistique, proposée en 1936 par Hotelling, surtout connue pour ses qualités théoriques, puisqu’elle englobe de nombreuses autres méthodes statistiques (par exemple, l’analyse factorielle des correspondances). Elle permet de décrire les relations linéaires qui existent entre deux ensembles de variables mesurées sur les mêmes individus.

Plusieurs méthodes ont été proposées pour généraliser l’analyse des corrélations canoniques. Ces différentes méthodes sont regroupées sous le nom d’analyse canonique généralisée lorsqu’elles donnent, dans le cas de deux groupes de variables, les mêmes résultats que l’analyse canonique simple [Hotelling, 1936].

En général, on cherche des combinaisons linéaires des variables de chaque groupe appelées variables canoniques qui optimisent des fonctions construites à partir de leur matrice de corrélation ou de variance covariance. Parmi ces méthodes, on peut citer:

1. la méthode Genvar [Stell, 1951] qui minimise le déterminant de la matrice de corrélation de l’ensemble des variables canoniques ;
2. la méthode Minvar [Kenttenring, 1971, 1983] qui minimise la plus petite valeur propre de la matrice de corrélation de l’ensemble des variables canoniques ;
3. la méthode Sumcor [Horst, 1961(a), 1961(b)] qui consiste à maximiser la somme des inters corrélations entre les variables canoniques ;
4. la méthode Maxvar [Horst, 1961(a), 1961(b)] qui cherche à maximiser la plus grande valeur propre de la matrice de corrélation de l’ensemble des variables canoniques ;
5. la méthode de Carroll [Carroll, 1968] qui cherche à maximiser la somme des carrés des corrélations multiples entre le facteur commun, appelé compromis canonique (combinaison linéaire de toutes les variables à étudier), et les variables de chaque groupe.
6. la méthode Ssqcor [Kenttenring, 1971, 1983] qui consiste à maximiser la somme des carrés des corrélations des couples des variables canoniques ;
7. la méthode Ecart [Nzobounsana, 2001 ; Nzobounsana et Dhorne, 2003] qui consiste à minimiser l’écart entre la plus grande et la plus petite valeurs propres de la matrice des corrélations des couples des variables canoniques ;

Le but de cet article est, dans un premier temps, de décrire l'analyse des corrélations canoniques, l'analyse canonique généralisée et de faire une comparaison de ces méthodes (la méthode Sumcor, la méthode Ssqcor et, la méthode de Carroll).

Dans un second temps nous présentons un questionnaire sur la représentation de la conduite automobile dans le champ de la psychologie sociale. Le questionnaire des scripts conditionnels [Gaymard, 2007], a été élaboré afin d'identifier les conditions spécifiques qui permettent de justifier le non respect des règles de conduite. Il s'inscrit dans le courant structuraliste des représentations sociales et s'appuie en particulier sur les travaux démarrés par Flament [1989, 1994(a), 1994(b)] portant sur la conditionnalité des normes. Le questionnaire des scripts conditionnels comporte à l'origine huit scénarii routiers (feu rouge, feu orange, limite de vitesse, stop, sens interdit, ligne blanche continue et respect général) et 127 situations de conduite qui vont plus ou moins justifier les transgressions selon la théorie de la conditionnalité [Flament, 1994(a), 1994(b)]. Une partie de ce questionnaire nous servira de base d'application des méthodes d'analyses canoniques simple et généralisée. Il s'agira de montrer que l'approche proposée par Hotelling et les généralisations sont des outils statistiques adaptés pour décrire et analyser des données recueillies dans le champ des sciences humaines et sociales. De plus, cette approche méthodologique nous permettra de développer une réflexion en lien avec les dernières recherches dans le domaine des représentations sociales et des aspects normatifs.

2. ANALYSE CANONIQUE LINÉAIRE SIMPLE

2.1. INTÉRÊT

L'analyse canonique linéaire simple est une méthode factorielle dont l'origine remonte à [Hotelling, 1936]. Elle a pour intérêt, de mesurer et de caractériser la dépendance linéaire qu'il y a ou non entre deux groupes de variables mesurées sur les mêmes individus. La notion de dépendance est mesurée ici par le coefficient de corrélation maximal entre les combinaisons linéaires des variables du premier groupe et les combinaisons linéaires des variables du deuxième groupe. Ce coefficient est appelé *coefficient de corrélation canonique*.

2.2. NOTATIONS

On considère deux groupes de variables X^1 et X^2 mesurées sur $n > 1$ mêmes individus et possédant respectivement m_1 et m_2 variables ($m_1 > 1$, $m_2 > 1$) tels que les matrices de variances covariances de chaque groupe de variables existent. On note par $Im(X^i)$ le sous-espace vectoriel engendré par les variables de X^i et par $X = [X^1|X^2]$ une matrice de dimension $n * (m_1 + m_2)$ obtenue en juxtaposant cote à cote X^1 et X^2 .

Dans la suite, on postule simplement que la moyenne et les covariances des variables de X sont connues et on les note μ_x et Σ_x . Pour simplifier l'exposé, on suppose que les variables des deux groupes ont une moyenne nulle sans que ceci ne soit restrictif et que Σ_{x^i} (matrice des covariances entre les variables de X^i , $i \in \{1, 2\}$) est inversible.

La matrice de variance covariance de X peut quant à elle être décomposée par rapport à X^1 et X^2 de la façon suivante :

$$\Sigma_x = \begin{pmatrix} \Sigma_{x^1} & \Sigma_{x^1x^2} \\ \Sigma_{x^2x^1} & \Sigma_{x^2} \end{pmatrix}$$

où $\Sigma_{x^2x^1}$ (respectivement $\Sigma_{x^1x^2}$) désigne la matrice des covariances entre les variables de X^2 et celles de X^1 (respectivement entre les variables de X^1 et celles de X^2).

DÉFINITION 1. *On appelle dès lors l'analyse canonique linéaire simple de X^1 et X^2 , la recherche d'une combinaison linéaire normée des variables de X^1 notée $X^1P_1 \in \text{Im}(X^1)$ (variable canonique associée à X^1) avec $P_1 \in R^{m_1}$ et une combinaison linéaire normée des variables de X^2 (variable canonique de X^2) notée $X^2P_2 \in \text{Im}(X^2)$ avec $P_2 \in R^{m_2}$ telle que la corrélation linéaire entre ces deux variables soit maximale. Mathématiquement cela est équivalent à trouver les deux combinaisons linéaires telles que :*

$$\{P_1, P_2\} = \arg. \max_{P_1, P_2} \{Cor(X^1P_1, X^2P_2)\} \quad (1)$$

sous les contraintes

$$\text{Var}(X^1P_1) = 1 \quad \text{et} \quad \text{Var}(X^2P_2) = 1 \quad (2)$$

Une fois déterminé la première corrélation canonique (le maximum du carré de la corrélation linéaire entre X^1P_1 et X^2P_2 notée ρ_1 ainsi que le premier couple, $(Z_1^{(1)}, Z_2^{(1)})$, des variables canoniques, les autres corrélations canoniques et variables canoniques sont déterminées par récurrence en maximisant :

$$\rho_i = Cor(Z_1^{(i)}, Z_2^{(i)}) \quad (3)$$

sous les contraintes $\text{Var}(Z_1^{(i)}) = \text{Var}(Z_2^{(i)}) = 1$ et des contraintes additionnelles:

$$Cor(Z_1^{(s)}, Z_1^{(k)}) = Cor(Z_2^{(s)}, Z_2^{(k)}) = 0 \quad \text{pour tout } k \in \{1, 2, \dots, (s-1)\}$$

PROPOSITION 1. *Sous la contrainte (2), les vecteurs canoniques notés P_1 et P_2 solutions de (1) vérifient les deux équations suivantes :*

$$\Sigma_{x^1}^{-1} \Sigma_{x^1x^2} \Sigma_{x^2}^{-1} \Sigma_{x^2x^1} P_1 = \rho^2 P_1 \quad \text{et} \quad \Sigma_{x^2}^{-1} \Sigma_{x^2x^1} \Sigma_{x^1}^{-1} \Sigma_{x^1x^2} P_2 = \rho^2 P_2 \quad (4)$$

c'est-à-dire que P_1 (respectivement P_2) est un vecteur propre normé de $\Sigma_{x^1}^{-1} \Sigma_{x^1x^2} \Sigma_{x^2}^{-1} \Sigma_{x^2x^1}$ (respectivement de $\Sigma_{x^2}^{-1} \Sigma_{x^2x^1} \Sigma_{x^1}^{-1} \Sigma_{x^1x^2}$) associé à la plus grande valeur propre commune aux deux matrices et notée ρ^2 .

PROPOSITION 2. [Hotelling, 1936]. *L'analyse canonique entre X^1 et X^2 vérifie les propriétés suivantes:*

- Si toutes les corrélations canoniques sont nulles, les deux groupes de variables sont complètement indépendants. Dans cette situation, il est inutile de vouloir prédire par exemple X^1 par X^2 et réciproquement.
- Si deux corrélations canoniques sont égales à 1, tout un ensemble de variables d'un groupe peut être prédit parfaitement à l'aide des variables de l'autre groupe.
- Les vecteurs canoniques du sous-espace $Im(X^1)$ notés $P_1^{(1)}, P_1^{(2)}, \dots, P_1^{(s)}$ et ceux de $Im(X^2)$ notés $P_2^{(1)}, P_2^{(2)}, \dots, P_2^{(s)}$ sont liés par la relation suivante

$$P_1^{(k)} = \frac{1}{\rho_k} \Sigma_{x^1}^{-1} \Sigma_{x^1 x^2} P_2^{(k)}$$

PROPOSITION 3. Les éléments propres de l'analyse canonique entre X^1 et X^2 vérifient les propriétés suivantes :

- Les deux matrices $\Sigma_{x^2}^{-1} \Sigma_{x^2 x^1} \Sigma_{x^1}^{-1} \Sigma_{x^1 x^2}$ et $\Sigma_{x^1}^{-1} \Sigma_{x^1 x^2} \Sigma_{x^2}^{-1} \Sigma_{x^2 x^1}$ ont en commun μ valeurs propres strictement positives dont ν ($\nu < \mu$) valeurs propres sont égales à 1. Les $p_1 - \mu$ (respect. les $p_2 - \mu$ autres valeurs propres de M_{11} (respect. de M_{22}) sont nulles.
- Dans l'espace $Im(X^1)$, les p_1 vecteurs propres $P_1^{(1)}, P_1^{(2)}, \dots, P_1^{(p_1)}$ de M_{11} , normés, appelés vecteurs canoniques de $Im(X^1)$, constituent une base orthonormée de $Im(X^1)$ appelée base de $Im(X^1)$ canonique par rapport à $Im(X^2)$ tandis que, les p_2 vecteurs propres $P_2^{(1)}, P_2^{(2)}, \dots, P_2^{(p_2)}$ de M_{22} normés et appelés vecteurs canoniques de $Im(X^2)$, constituent une base orthonormée de $Im(X^2)$ appelée base de $Im(X^2)$ canonique par rapport à $Im(X^1)$.
- Pour tout i , le couple $(P_1^{(i)}, P_2^{(i)})$ est appelé i -ème couple canonique de l'analyse canonique entre X^1 et X^2 et vérifient les deux équations suivantes :

$$P_1^{(i)} = \frac{1}{\rho_i} \Sigma_{x^1}^{-1} \Sigma_{x^1 x^2} P_2^{(i)} \quad \text{et} \quad P_2^{(i)} = \frac{1}{\rho_i} \Sigma_{x^2}^{-1} \Sigma_{x^2 x^1} P_1^{(i)} \quad (5)$$

pour ρ_i différent de zéro.

La proposition suivante fait un lien entre le critère de corrélation linéaire et celui lié aux corrélations multiples.

PROPOSITION 4. Soient X^1 et X^2 deux groupes de variables de dimension m_1 et m_2 dont les variables sont mesurées sur les mêmes individus. Les solutions (le vecteur normé $Z_1 = X^1 P_1 \in Im(X^1)$ et le vecteur normé $Z_2 = X^2 P_2 \in Im(X^2)$) de l'analyse canonique au sens de Hotelling de X^1 et X^2 peuvent être obtenues en cherchant un vecteur normé noté $Z = X P$ appelé compromis canonique de X^1 et X^2 (avec $P \in R^{m_1+m_2}$ et Z , une combinaison linéaire des variables de $X = [X^1|X^2]$) qui maximise le critère :

$$Cor^2(Z, X^1) + Cor^2(Z, X^2) \quad (6)$$

où $\forall i \in \{1, 2\}$, $Cor(Z, X^i)$ désigne la corrélation multiple entre le vecteur Z et les variables de X^i .

DÉMONSTRATION 1. La déduction des solutions (le vecteur normé $Z_1 = X^1 P_1 \in Im(X^1)$ et le vecteur normé $Z_2 = X^2 P_2 \in Im(X^2)$) de l'analyse canonique au sens de Hotelling de X^1 et X^2 à partir du critère (6) se démontre comme suit :

Pour tout i , la valeur maximale du carré de la corrélation multiple entre le vecteur Z et les variables de X^i est donnée par :

$$Cor^2(Z, X^i) = \frac{{}^t Z X^i ({}^t X^i X^i)^{-1} {}^t X^i Z}{Var(Z)}, \quad i \in \{1, 2\} \quad (7)$$

sous la contrainte que l'inverse de $({}^t X^i X^i)$ existe. En reportant l'égalité (7) dans (6) on obtient :

$$Cor^2(Z, X^1) + Cor^2(Z, X^2) = \frac{{}^t Z X^1 ({}^t X^1 X^1)^{-1} {}^t X^1 Z + {}^t Z X^2 ({}^t X^2 X^2)^{-1} {}^t X^2 Z}{Var(Z)} = \frac{{}^t Z Q Z}{Var(Z)} \quad (8)$$

où Q est une matrice définie par :

$$Q = X^1 ({}^t X^1 X^1)^{-1} {}^t X^1 + X^2 ({}^t X^2 X^2)^{-1} {}^t X^2 = X M^t X \quad (9)$$

et M est une matrice $(m_1 + m_2) \times (m_1 + m_2)$ définie par :

$$M = \begin{pmatrix} \Sigma_{x^1}^{-1} & 0 \\ 0 & \Sigma_{x^2}^{-1} \end{pmatrix}$$

En d'autres termes, le vecteur $Z = X P$ normé qui maximise (6), est un vecteur propre de Q associé à sa plus grande valeur propre notée λ . De (8), il s'ensuit que le vecteur $P \in R^{m_1+m_2}$ est un vecteur propre de $M \Sigma_x = M ({}^t X X)$ associé à la même valeur propre puisque :

$$X M^t X Z = \lambda Z \iff X M^t X X P = \lambda X P \iff M^t X X P = \lambda P \iff \Sigma_x P = \lambda M^{-1} P \quad (10)$$

De l'égalité (10), il découle, après transformation, que les sous-vecteurs P_1 et P_2 de $P = ({}^t P_1, {}^t P_2)$ avec $P_1 \in R^{m_1}$ et $P_2 \in R^{m_2}$ vérifient les deux égalités ci-dessous :

$$\Sigma_{x^1 x^2} P_2 = (\lambda - 1) \Sigma_{x^1} P_1 \quad \text{et} \quad \Sigma_{x^2 x^1} P_1 = (\lambda - 1) \Sigma_{x^2} P_2 \iff \quad (11)$$

$$\Sigma_{x^1}^{-1} \Sigma_{x^1 x^2} P_2 = (\lambda - 1) P_1 \quad \text{et} \quad \Sigma_{x^2}^{-1} \Sigma_{x^2 x^1} P_1 = (\lambda - 1) P_2 \iff \quad (12)$$

$$\Sigma_{x^2}^{-1} \Sigma_{x^2 x^1} \Sigma_{x^1}^{-1} \Sigma_{x^1 x^2} P_2 = (\lambda - 1)^2 P_2 \quad \text{et} \quad \Sigma_{x^1}^{-1} \Sigma_{x^1 x^2} \Sigma_{x^2}^{-1} \Sigma_{x^2 x^1} P_1 = (\lambda - 1)^2 P_1 \quad (13)$$

(4) et (13) étant identiques ($\rho^2 = (1 - \lambda)^2$), cela entraîne que l'analyse canonique de Hotelling qui correspond à chercher deux vecteurs $Z_1 = X^1 P_1$ et $Z_2 = X^2 P_2$ normés tels que $Cor^2(Z_1, Z_2)$ est maximale est équivalente à la recherche d'un vecteur $Z = X P$ normé (avec $P = {}^t({}^t P_1, {}^t P_2)$, $P_1 \in R^{m_1}$ et $P_2 \in R^{m_2}$) tel que le critère (6) soit maximal.

PROPOSITION 5. *L'analyse canonique généralise la régression multiple et admet comme cas particulier (cf. [Saporta, 1990] p. 211-214 et chapitre 18) l'analyse factorielle des correspondances (X^1 et X^2 sont des variables qualitatives ayant respectivement $m_1 > 1$ et $m_2 > 1$ modalités; $\Sigma_{x^2 x^1}$ est un tableau de fréquences, Σ_{x^1} et Σ_{x^2} sont des matrices diagonales dont les éléments sont les fréquences marginales du tableau $\Sigma_{x^2 x^1}$) et l'analyse discriminante (cas où X^1 est une variable qualitative à $m_1 > 1$ modalités et X^2 est formé à partir de $m_2 > 1$ variables quantitatives).*

3. ANALYSE CANONIQUE GÉNÉRALISÉE

Dans cette section, nous présentons 3 généralisations de l'analyse des corrélations canoniques de deux groupes de variables. Il s'agit de la généralisation proposée: par Horst (la méthode *Sumcor*); par Kettenring (la méthode *Ssqcor*) et enfin, par Carroll.

Dans la suite, nous supposons que l'on dispose de p ($p > 2$) groupes de variables notés X^1, X^2, \dots, X^p dont les variables sont mesurées sur n mêmes individus et, nous notons par m_1, m_2, \dots, m_p le nombre total des variables dans les groupes.

Nous supposons aussi que la matrice de variance covariance (respectivement, de corrélation lorsque les variables sont centrées et réduites) de $X = [X^1 | X^2 | \dots | X^p]$ (juxtaposition des p groupes de variables) est décomposée par rapport aux p groupes de la façon suivante:

$$\Sigma_x = \begin{pmatrix} \Sigma_{x^1} & \Sigma_{x^1 x^2} & \cdots & \Sigma_{x^1 x^p} \\ \Sigma_{x^2 x^1} & \Sigma_{x^2} & \cdots & \Sigma_{x^2 x^p} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{x^p x^1} & \Sigma_{x^p x^2} & \cdots & \Sigma_{x^p} \end{pmatrix}$$

Soient Z_1, Z_2, \dots, Z_p p combinaisons linéaires définies à partir de X^1, X^2, \dots, X^p où $Z_i = X^i P_i$ et $P_i \in R^{m_i}$ pour tout $i \in \{1, 2, \dots, p\}$. Z une combinaison linéaire des variables de X définie par $Z = X P$, $P \in R^m$ et $m = m_1 + m_2 + \dots + m_p$ (le nombre total des variables de X).

3.1. BUT ET INTÉRÊT DE L'ANALYSE CANONIQUE GÉNÉRALISÉE

Les méthodes d'analyse canonique généralisée ont pour but, de décrire et de résumer les liaisons linéaires existantes entre plusieurs groupes de variables mesurées sur les mêmes individus.

3.2. PRÉSENTATION DE TROIS FONCTIONS OPTIMISÉES EN ACG.

Méthode 1 : Maximisation de la somme des corrélations des couples de variables canoniques: c'est la méthode *Sumcor*. Cette méthode a été proposée par Horst en 1961 et consiste à trouver les vecteurs P_1, P_2, \dots, P_p qui maximisent sous des contraintes (cf. sous-section suivante) la fonction suivante :

$$f_1(P_1, P_2, \dots, P_p) = \sum_{i=1}^p \sum_{j=1}^p Cor(Z_i, Z_j) \quad [\text{Horst, 1961(a)}] \quad (14)$$

Méthode 2 : Maximisation de la somme des carrés des corrélations des couples de variables canoniques: c'est la méthode *Ssqcor* proposée par Kettenring en 1971 et qui consiste à trouver les vecteurs P_1, P_2, \dots, P_p qui maximisent sous des contraintes (cf. sous-section suivante) la fonction suivante :

$$f_2(P_1, P_2, \dots, P_p) = \sum_{i=1}^p \sum_{j=1}^p Cor^2(Z_i, Z_j) \quad [\text{Kettering, 1983}] \quad (15)$$

Méthode 3 : Maximisation de la somme de corrélations multiples entre un vecteur appelé compromis (combinaison linéaire des variables de X) et les variables de chaque groupe: c'est la généralisation de l'analyse canonique proposée par Carroll en 1968 qui consiste à trouver un vecteur P qui maximise sous des contraintes (cf. sous-section suivante) la fonction suivante :

$$f_3(P) = \sum_{i=1}^p Cor^2(Z, X^i) \quad [\text{Carroll, 1968}] \quad (16)$$

3.3. TYPES DE CONTRAINTES IMPOSÉES AUX VARIABLES CANONIQUES

Dans la plupart des problèmes d'optimisation définis dans la section précédente, le calcul des variables et vecteurs canoniques (Z_1, Z_2, \dots, Z_p et P_1, P_2, \dots, P_p) se fait pas à pas. Dans certaines généralisations, il nécessite la mise en oeuvre d'un algorithme itératif.

À chaque pas, les variables canoniques obtenues sont conservées en vue de construire une base orthonormée pour chacun des sous-espaces vectoriels, $Im(X^i)$ pour $i \in \{1, 2, \dots, p\}$, engendrés par les variables de chaque X^i .

Soient $Z_1^{(s)}, Z_2^{(s)}, \dots, Z_p^{(s)}$, l'ensemble des variables canoniques obtenues au pas s ; au pas $(s + 1)$, celles-ci sont obtenues selon l'une des familles de contraintes suivantes :

- *Des contraintes générales.* Pour un groupe de variables X^i , si l'on veut avoir une base orthonormée de $Im(X^i)$, on optimise les fonctions définies dans la sous-section précédente sous les contraintes :

$$Cor(Z_i^{(s+1)}, Z_i^{(k)}) = 0, \quad i \in \{1, \dots, p\} \quad \text{et} \quad k \in \{1, \dots, s\} \quad (17)$$

Avec ces contraintes, on arrête la recherche des variables canoniques dès qu'on a obtenu une base orthonormée de chacun des $Im(X^i)$.

- *Des contraintes flexibles.* Elles sont définies :

$$Cor(Z_i^{(s+1)}, Z_i^{(k)}) = 0, \quad i \in F^l \quad \text{et} \quad k \in \{1, 2, \dots, s\} \quad (18)$$

où F^l est un sous ensemble de $\{1, 2, \dots, p\}$.

Pour cette famille de contraintes, le nombre de pas à effectuer pour extraire l'ensemble des variables canoniques est égal à :

$$p_i \quad \text{ou} \quad i = \min\{k, k \in F^l\}$$

- *Des contraintes rigides.* Lorsqu'on impose l'orthogonalité entre toutes les variables canoniques de $Im(X^i)$ et celles de $Im(X^j)$ pour tout $i \neq j$ et $i, j \in \{1, 2, \dots, p\}$, on utilise les contraintes suivantes :

$$Cor((Z_1^{(s+1)}, \dots, Z_p^{(s+1)}), (Z_1^{(k)}, Z_2^{(k)}, \dots, Z_p^{(k)})) = 0, \quad k \in \{1, 2, \dots, s\} \quad (19)$$

- *Les contraintes particulières.* Si l'on veut avoir une base orthonormée des vecteurs canoniques associés aux variables canoniques cherchées, on utilise les contraintes d'orthogonalités suivantes :

$${}^t P_i^{(k)} P_i^{(s)} = 0, \quad k \in \{1, 2, \dots, s\} \quad \text{et} \quad i \in \{1, 2, \dots, p\} \quad (20)$$

3.4. SOLUTIONS DES DIFFÉRENTES MÉTHODES D'ACG

Sans perdre de généralité (puisque toutes les fonctions optimisées en analyse canonique généralisée et qui sont décrites dans la sous-section 3.2., sont invariantes par rapport aux changements d'échelles et aux transformations régulières des variables dans les groupes), nous supposons dans les développements théoriques qui suivent que les variables dans les groupes sont orthogonales deux à deux et ont une norme égale à 1. De cette restriction, il en résulte que la matrice de corrélation de $X = (X^1 | X^2 | \dots | X^p)$ est de la forme :

$$\Sigma = \begin{pmatrix} Id_{m_1} & \Sigma_{x^1 x^2} & \cdots & \Sigma_{x^1 x^p} \\ \Sigma_{x^2 x^1} & Id_{m_2} & \cdots & \Sigma_{x^2 x^p} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{x^p x^1} & \Sigma_{x^p x^2} & \cdots & Id_{m_p} \end{pmatrix}$$

et celle des variables canoniques est donnée par

$$\Phi = \begin{pmatrix} {}^t P_1 & 0 & \cdots & 0 \\ 0 & {}^t P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {}^t P_p \end{pmatrix} \begin{pmatrix} Id_{m_1} & \Sigma_{x^1 x^2} & \cdots & \Sigma_{x^1 x^p} \\ \Sigma_{x^2 x^1} & Id_{m_2} & \cdots & \Sigma_{x^2 x^p} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{x^p x^1} & \Sigma_{x^p x^2} & \cdots & Id_{m_p} \end{pmatrix} \begin{pmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_p \end{pmatrix}$$

3.4.1. Méthode Sumcor

Le but de la méthode Sumcor, est d'étudier les liaisons entre plusieurs groupes de variables en maximisant [14]. Elle a pour intérêt, d'expliciter d'une part les liaisons deux à deux entre les p groupes de variables et d'autre part, les liaisons entre chaque groupe de variable et les $(p - 1)$ autres restants. Les solutions (variables et vecteurs canoniques) de cette méthode sont obtenues pas à pas.

- *Calcul des variables canoniques d'ordre un.* Les solutions d'ordre 1 de cette méthode sont obtenues en recherchant, pour tout $i \in \{1, 2, \dots, p\}$, les vecteurs Z_1, Z_2, \dots, Z_p qui maximisent la fonction :

$$f_1(P_1, P_2, \dots, P_p) = \sum_{i=1}^p \sum_{i \neq j}^p Cor(Z_i, Z_j) \quad (21)$$

sous les contraintes :

$$Var(Z_1) = Var(Z_2) = \dots = Var(Z_p) = 1. \quad (22)$$

Comme on a supposé que les variables dans les groupes sont deux à deux orthogonales, la maximisation de (21) sous les contraintes (22) est équivalente à trouver p vecteurs canoniques, P_1, P_2, \dots, P_p tels que :

$$(P_1, P_2, \dots, P_p) = arg. \max_{P_1, P_2, \dots, P_p} \left\{ \sum_{i=1}^p \sum_{i \neq j}^p {}^t P_i \Sigma_{x^i x^j} P_j \right\} \quad (23)$$

sous les contraintes :

$${}^t P_1 P_1 = {}^t P_2 P_2 = \dots = {}^t P_p P_p = 1. \quad (24)$$

C'est-à-dire que chaque vecteur canonique doit être de norme égale à 1.

PROPOSITION 6. *Les solutions de l'analyse canonique généralisée selon la méthode Sumcor vérifient les égalités suivantes (pour tout $i \neq j$ et $i, j \in \{1, 2, \dots, p\}$)*

$$\lambda_i P_i = \sum_{j=1, i \neq j}^p \Sigma_{x^i x^j} P_j \quad (25)$$

où les λ_i sont des constantes de normalisation des vecteurs canoniques P_1, P_2, \dots, P_p .

Ce résultat s'obtient en utilisant la méthode des multiplicateurs de Lagrange. Il s'agit de trouver P_1, P_2, \dots, P_p solutions du système d'équations :

$$\begin{pmatrix} I_{m_1} & \Sigma_{x^1x^2} & \cdots & \Sigma_{x^1x^p} \\ \Sigma_{x^2x^1} & I_{m_2} & \cdots & \Sigma_{x^2x^p} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{x^px^1} & \Sigma_{x^px^2} & \cdots & I_{m_p} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_p \end{pmatrix} = \begin{pmatrix} \lambda_1 + 1 & 0 & \cdots & 0 \\ 0 & \lambda_2 + 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p + 1 \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_p \end{pmatrix}$$

Remarque 1. Lorsque les constantes de normalisations, $\lambda_1, \lambda_2, \dots, \lambda_p$, sont égales, la première solution de la méthode Sumcor est obtenue en faisant la décomposition spectrale de la matrice

$$\begin{pmatrix} I_{m_1} & \Sigma_{x^1x^2} & \cdots & \Sigma_{x^1x^p} \\ \Sigma_{x^2x^1} & I_{m_2} & \cdots & \Sigma_{x^2x^p} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{x^px^1} & \Sigma_{x^px^2} & \cdots & I_{m_p} \end{pmatrix}$$

Les vecteurs canoniques, P_1, \dots, P_p , sont alors des sous vecteurs de $P = {}^t(P_1, \dots, P_p)$ (premier vecteur propre de Σ) associé à la plus grande valeur propre.

Dans le cas général, les solutions du problème de l'analyse canonique généralisée selon la méthode Sumcor s'obtiennent par un processus itératif (itérations successives) de la façon suivantes :

À partir d'une solution initiale $P_1^{(0)}, P_2^{(0)}, \dots, P_p^{(0)}$ quelconque par exemple appartenant à $Im(X^1), Im(X^2), \dots$ et $Im(X^p)$.

On définit la solution d'ordre 1 du problème de la manière suivante: c'est l'ensemble des vecteurs $P_1^{(1)}, P_2^{(1)}, \dots, P_p^{(1)}$ normés solutions du système d'équations ci-dessous :

$$S_1 = \begin{cases} \sum_{j=1, j \neq 1}^p \Sigma_{x^1x^j} P_j = \lambda_1 P_1, & \lambda_1 = \frac{P_1}{{}^t P_1 P_1} \\ \sum_{j=1, j \neq 2}^p \Sigma_{x^2x^j} P_j = \lambda_2 P_2, & \lambda_2 = \frac{P_2}{{}^t P_2 P_2} \\ \dots \\ \sum_{j=1, j \neq p}^p \Sigma_{x^px^j} P_j = \lambda_p P_p, & \lambda_p = \frac{P_p}{{}^t P_p P_p} \end{cases}$$

Soient $P_1^{(1)}, P_2^{(1)}, \dots, P_p^{(1)}$ l'ensemble des p vecteurs canoniques normés dans $R^{m_1}, R^{m_2}, \dots, R^{m_p}$ solutions du système d'équations S_1 , on a les propriétés suivantes :

PROPOSITION 7. *Les variables canoniques d'ordre un de la méthode Sumcor sont normées (pour tout i , $Var(Z_i^{(1)}) = 1$). Elles vérifient les égalités :*

$$\begin{pmatrix} a_{11}I_{m_1} & \Sigma_{x^1x^2} & \cdots & \Sigma_{x^1x^p} \\ \Sigma_{x^2x^1} & a_{22}I_{m_2} & \cdots & \Sigma_{x^2x^p} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{x^px^1} & r_{p2}\Sigma_{x^px^2} & \cdots & a_{pp}I_{m_p} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_p \end{pmatrix} = \lambda \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_p \end{pmatrix}$$

où

$$\begin{aligned}
a_{ii} &= \sum_{k=1, k \neq i}^p \sum_{j=1, j \neq k}^p \text{Cor}(Z_k, Z_j) \quad , \quad \lambda = \lambda_1 + \lambda_2 + \dots + \lambda_p = a_{ii} + \lambda_i \quad \text{pour tout } i \\
\lambda_i Z_i^{(1)} &= \Pi_{x^i} \left(\sum_{i \neq j, j=1}^p Z_j^{(1)} \right) \quad i \in \{1, 2, \dots, p\}
\end{aligned} \tag{26}$$

De plus, les multiplicateurs de Lagrange, $\lambda_1, \lambda_2, \dots, \lambda_p$, vérifient les égalités :

$$\lambda_i = \text{Cov}(Z_i^{(1)}, \sum_{j=1, i \neq j}^p Z_j^{(1)}) \quad i \in \{1, 2, \dots, p\}$$

où :

– Π_{x^i} désigne la matrice de projection orthogonale associée au sous-espace vectoriel, $\text{Im}(X^i)$, engendré par les variables de X^i .

– $Z_i^{(1)}$, une variable canonique d'ordre un de X^i .

• *Calcul des variables canoniques d'ordre 2.* La procédure de calcul des variables canoniques d'ordre deux est similaire à celle qui a permis d'avoir les solutions d'ordre un. Il s'agit encore de maximiser la fonction $f_1(P_1, P_2, \dots, P_p)$ par rapport à chaque vecteur P_i sous la contrainte que les variables canoniques sont normées et des contraintes additionnelles $\text{Cor}(Z_i, Z_i^{(1)}) = 0$ (la solution d'ordre deux est orthogonale à la solution d'ordre un). Plus précisément, il s'agit de trouver P_1, P_2, \dots, P_p solutions du problème

$$\begin{cases} (P_1, P_2, \dots, P_p) = \arg.\max[\sum_{i=1}^p \sum_{j=1}^p \text{Cor}(Z_i, Z_j) \\ \text{Var}(Z_1) = \text{Var}(Z_2) = \dots = \text{Var}(Z_p) = 1 \\ \text{Cor}(Z_1, Z_1^{(1)}) = \text{Cor}(Z_2, Z_2^{(1)}) = \dots = \text{Cor}(Z_p, Z_p^{(1)}) = 0 \end{cases}$$

PROPOSITION 8. Les solutions d'ordre deux de l'analyse canonique généralisée selon la méthode Sumcor vérifient les égalités suivantes :

$$\lambda_i P_i = \sum_{j=1}^p [I_{m_i} - \Pi_{P_i^{(1)}}] \Sigma_{x^i x^j} [I_{m_j} - \Pi_{P_j^{(1)}}] P_j \quad \text{pour tout } i \neq j \tag{27}$$

où $\Pi_{P_i^{(1)}}$ (respectivement λ_i) désigne la matrice de projection orthogonale associée au sous-espace $\text{Im}(P_i^{(1)})$ (respectivement des constantes de Lagrange).

DÉMONSTRATION 2. Soient $\lambda_1, \lambda_2, \dots, \lambda_p$ et $\mu_1, \mu_2, \dots, \mu_p$, $2p$ multiplicateurs de Lagrange. Pour trouver les solutions d'ordre deux de la méthode Sumcor, on doit résoudre (puisque pour tout i , $\Sigma_{x^i} = Id_{m_i}$) le système des équations suivantes:

$$\frac{\partial}{\partial P_i} \left\{ \sum_{i=1}^p \left\{ \sum_{i \neq j}^p {}^t P_i \Sigma_{x^i x^j} P_j - \lambda_i ({}^t P_i P_i - 1) - \mu_i {}^t P_i P_i^{(1)} \right\} \right\} = 0 \tag{28}$$

lesquelles après simplification et quelques transformations algébriques, conduisent à

$$\lambda_i P_i = \sum_{j=1}^p [I_{m_i} - \Pi_{p_i^{(1)}}] \Sigma_{x^i x^j} [I_{m_j} - \Pi_{p_j^{(1)}}] P_j \quad \text{pour tout } i \neq j$$

où $\Pi_{p_i^{(1)}}$ désigne la matrice de projection orthogonale associée au sous-espace $\text{Im}(P_i^{(1)})$.

Comme pour les solutions d'ordre 1, les vecteurs canoniques $P_1^{(2)}, P_2^{(2)}, \dots, P_p^{(2)}$, solutions du problème (14) sont obtenues par un processus itératif similaire à S_1 . Il suffit de remplacer les matrice $\Sigma_{x^i x^j}$ du système S_1 par les nouvelles matrices $[I_{m_i} - \Pi_{p_i^{(1)}}] \Sigma_{x^i x^j}$.

• Calcul des variables canoniques d'ordre s ($s > 2$). Soient $P_1^{(s-1)}, P_2^{(s-1)}, \dots, P_p^{(s-1)}$, p vecteurs normés de $R^{m_1}, R^{m_2}, \dots, R^{m_p}$ qui maximisent la fonction $f_1(P_1, P_2, \dots, P_p)$ à l'ordre $(s-1)$.

À l'ordre s ($s \geq 3$), on cherche p nouveaux vecteurs canoniques notés $P_1^{(s)}, P_2^{(s)}, \dots, P_p^{(s)}$ normés qui maximisent la fonction f_1 sous la contrainte additionnelle :

$${}^t P_i^{(s)} P_i^{(k)} = 0 \quad \text{pour tout } k \in \{1, 2, \dots, (s-1)\} \quad (29)$$

En prenant le Lagrangien associé à ce problème de maximisation, on est amené à maximiser la fonction

$$L_{(\lambda, \beta)}(P) = \sum_{i=1}^p \left\{ \sum_{j=1}^p \text{Cor}(Z_i^{(s)}, Z_j^{(s)}) - \lambda_i ({}^t P_i^{(s)} P_i^{(s)} - 1) - \sum_{k=1}^{(s-1)} \beta_i^{kt} P_i^{(s)} P_i^{(k)} \right\} \quad (30)$$

où $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_p\}$ et $\beta = \{\beta_1^k, \beta_2^k, \dots, \beta_p^k\}$ sont deux vecteurs formés de p réels appelés multiplicateurs de Lagrange. Enfin le résultat suivant permet le calcul numérique des solutions du problème.

PROPOSITION 9. À l'ordre s , sous la contrainte additionnelle 29, la fonction scalaire $f_1(P_1, P_2, \dots, P_p)$ est maximisée si et seulement si les vecteurs canoniques cherchés vérifient le système suivant :

$$S_1^{(s)} = \begin{cases} \sum_{j=1, j \neq 1}^p M_1^{(s)} \Sigma_{x^1 x^j} M_j^{(s)} P_j^{(s)} = \lambda_1 P_1^{(s)} \\ \sum_{j=1, j \neq 2}^p M_2^{(s)} \Sigma_{x^2 x^j} M_j^{(s)} P_j^{(s)} = \lambda_2 P_2^{(s)} \\ \dots \dots \dots \\ \sum_{j=1, j \neq p}^p M_p^{(s)} \Sigma_{x^p x^j} M_j^{(s)} P_j^{(s)} = \lambda_p P_p^{(s)} \end{cases}$$

où la matrice $M_i^{(s)}$ pour tout i est définie par :

$$M_i^{(s)} = I_{m_i} - \Pi_A \quad \text{et } A = \text{Im}\{P_1^{(s-1)}, P_2^{(s-1)}, \dots, P_p^{(s-1)}\}$$

et Π_A désigne la matrice de projection orthogonale associée à A .

DÉMONSTRATION 3. Pour avoir les résultats de la proposition précédente, on dérive l'égalité (30) par rapport aux p vecteurs $P_1^{(s)}, P_2^{(s)}, \dots, P_p^{(s)}$. En dérivant (30) par rapport à $P_1^{(s)}, P_2^{(s)}, \dots$, et $P_p^{(s)}$ et en égalant à zéro le résultat, on trouve :

$$\sum_{j=1, j \neq i}^p \Sigma_{x^i x^j} P_j^{(s)} - \lambda_i P_i^{(s)} - \sum_{k=1}^{(s-1)} \beta_i^k P_i^{(k)} = 0 \quad \forall i \quad (31)$$

Ce qui implique, en multipliant cette dernière équation par ${}^t P_i^{(k)}$:

$$\sum_{j=1, j \neq i}^p {}^t P_i^{(k)} \Sigma_{x^i x^j} P_j^{(s)} = \beta_i^k \quad \forall i \text{ et } k \quad (32)$$

En remplaçant β_i^k dans l'équation (31), on montre que :

$$\sum_{j=1, j \neq i}^p [I_{m_i} - \sum_{k=1}^{(s-1)} P_i^{kt} P_i^{(k)}] \Sigma_{x^i x^j} P_j^{(s)} = \lambda_i P_i^{(s)} \quad (33)$$

c'est-à-dire que les vecteurs $P_1^{(s)}, P_2^{(s)}, \dots, P_p^{(s)}$ vérifient bien le système d'équation $S^{(s)}$ (puisque $M_j^{(s)} P_j^{(s)} = P_j^{(s)}$).

3.4.2. Méthode Ssqcor

Cette généralisation de l'analyse canonique est due à [Kettering, 1971]. Elle consiste à trouver p combinaisons linéaires normées, Z_1, Z_2, \dots, Z_p , appelées variables canoniques qui maximisent la fonction

$$f_2(P_1, P_2, \dots, P_p) = \sum_{i=1}^p \sum_{j \neq i}^p Cor^2(X^i P_i, X^j P_j). \quad (34)$$

Les solutions de ce problème d'optimisation sont obtenues de la manière suivante.

- *Solutions d'ordre un du problème.* Au premier pas, il s'agit de trouver p vecteurs canoniques, P_1, P_2, \dots, P_p , normés (puisque l'on a supposé que dans les groupes, les variables sont deux à deux orthogonales et normées) associés aux p variables canoniques, Z_1, Z_2, \dots, Z_p , qui maximisent la fonction $f_2(P_1, P_2, \dots, P_p)$. Pour ce faire, on utilise la méthode des multiplicateurs de Lagrange. Soient $\lambda_1, \lambda_2, \dots, \lambda_p$ p multiplicateurs de Lagrange, cela conduit, pour tout i , à résoudre le système d'équations:

$$\frac{\partial L(P_1, \dots, P_p)}{\partial P_i} = 0 \quad \text{où} \quad L(P_1, \dots, P_p) = \sum_{i=1}^p \sum_{j \neq i}^p Cor^2(Z_i, Z_j) - \sum_{i=1}^p \lambda_i ({}^t P_i P_i - 1) \quad (35)$$

Un raisonnement analogue à celui utilisé avec la méthode Sumcor permet de voir que sous la contrainte ${}^t P_i P_i = 1$, les vecteurs canoniques qui maximisent la fonction $f_2(P_1, P_2, \dots, P_p)$ vérifient l'équation matricielle :

$$\begin{pmatrix} I_{m_1} & r_{12}\Sigma_{x^1x^2} & \cdots & r_{1p}\Sigma_{x^1x^p} \\ r_{21}\Sigma_{x^2x^1} & I_{m_2} & \cdots & r_{2p}\Sigma_{x^2x^p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1}\Sigma_{x^px^1} & r_{p2}\Sigma_{x^px^2} & \cdots & I_{m_p} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_p \end{pmatrix} = \begin{pmatrix} \lambda_1^* & 0 & \cdots & 0 \\ 0 & \lambda_2^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p^* \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_p \end{pmatrix}$$

où pour tout i et j , $r_{ij} = \text{Cor}(Z_i, Z_j)$ et $\lambda_i^* = \lambda_i + 1$.

De manière explicite, l'égalité matricielle précédente est équivalente aux p systèmes d'équations suivantes :

$$S_2 = \begin{cases} \sum_{j=1, j \neq 1}^p r_{1j}\Sigma_{x^1x^j} P_j = \lambda_1^* P_1 \\ \sum_{j=1, j \neq 2}^p r_{2j}\Sigma_{x^2x^j} P_j = \lambda_2^* P_2 \\ \dots \dots \dots \\ \sum_{j=1, j \neq p}^p r_{jp}\Sigma_{x^jx^p} P_j = \lambda_p^* P_p \end{cases}$$

On montre aussi que si P_1, P_2, \dots, P_p sont solutions de S_2 alors les variables canoniques associées, Z_1, Z_2, \dots, Z_p , vérifient les égalités suivantes.

PROPOSITION 10. Soient P_1, P_2, \dots, P_p les solutions du système S_2 , alors les p variables canoniques Z_1, Z_2, \dots, Z_p associées et les constantes de normalisation notées, $\lambda_1, \lambda_2, \dots, \lambda_p$, vérifient les égalités suivantes:

$$\begin{pmatrix} a_{11}I_{m_1} & r_{12}\Sigma_{x^1x^2} & \cdots & r_{1p}\Sigma_{1p} \\ r_{21}\Sigma_{x^2x^1} & a_{22}I_{m_2} & \cdots & r_{2p}\Sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1}\Sigma_{x^px^1} & r_{p2}\Sigma_{x^px^2} & \cdots & a_{pp}I_{m_p} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_p \end{pmatrix} = \lambda \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_p \end{pmatrix}$$

où

$$\begin{aligned} a_{ii} &= \sum_{k=1, k \neq i}^p \sum_{j=1, j \neq k}^p \text{Cor}^2(Z_k, Z_j), \quad \lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_p = a_{ii} + \lambda_i \quad \text{pour tout } i \\ \text{Var}(Z_i) &= 1, \quad \lambda_i Z_i = \left\{ \sum_{j=1, j \neq i}^p r_{j,i} \Pi_{x^i} Z_j \right\}, \quad \lambda_i = \sum_{j=1, j \neq i}^p \text{Cor}^2(Z_j, Z_i) \end{aligned} \quad (36)$$

c'est-à-dire $\forall i$, la variable canonique Z_i associée à X^i est une combinaison linéaire normée des $r_{j,i} \Pi_{x^i} Z_j$. Π_{x^i} est la matrice de projection orthogonale associée à $\text{Im}(X^i)$.

• *Calcul des variables canoniques d'ordre s .* Soient $P_i^{(1)}, P_i^{(2)}, \dots, P_i^{(s-1)}$ les $(s-1)$ premières variables canoniques de X^i qui maximisent la fonction $f_2(P_1^{(s)}, P_2^{(s)}, \dots, P_p^{(s)})$ à l'ordre 1, 2, \dots , $(s-1)$.

À l'ordre s , on cherche p combinaisons linéaires $Z_1^{(s)}, Z_2^{(s)}, \dots, Z_p^{(s)}$, où $\forall i$ le vecteur $Z_i^{(s)}$ est défini par $Z_i^{(s)} = Y^i P_i^{(s)}$, qui maximisent la fonction $f_2(P_1^{(s)}, \dots, P_p^{(s)})$ sous la contrainte ${}^t P_i^{(s)} P_i^{(s)} = 1$ et une contrainte additionnelle ${}^t P_i^{(s)} P_i^{(k)} = 0$ pour tout $k \in \{1, 2, \dots, (s-1)\}$.

Nzobounsana [2001] montre qu'à l'ordre s , les vecteurs canoniques (les poids des variables canoniques) solutions de la méthode Ssqcor s'obtiennent par un processus itératif. Concrètement, on remplace dans S_2 , les matrices $\Sigma_{x^i x^j}$ par les matrices $A_i \Sigma_{x^i x^j} \forall i, j \in \{1, 2, \dots, p\}$ avec $A_i = I_{p_i} - \Pi_i^{(s-1)}$. $\Pi_i^{(s-1)}$ désigne la matrice de projection orthogonale sur le sous-espace vectoriel de R^{m_i} engendré par les $(s-1)$ premiers vecteurs canoniques associés au $(s-1)$ premières variables canoniques de X^i .

En conclusion, nous pouvons affirmer que l'analyse canonique généralisée au sens de la méthode Ssqcor, n'est autre que l'analyse canonique généralisée au sens de la méthode Sumcor où les matrices de corrélations $\Sigma_{x^i x^j}$ sont multipliées par les réels $r_{ij} = \text{Cor}(Z_i, Z_j)$. Ces réels varient en fonction des données et des vecteurs canoniques [cf. S_1 et S_2].

3.4.3. Méthode de Carroll

Intérêt. Comme pour les deux premières généralisations, l'analyse canonique généralisée proposée par Carroll a pour intérêt de mesurer et de caractériser les liaisons linéaires qui existent entre p groupes de variables. Elle s'appuie sur la notion de proximité entre un sous-espace vectoriel dit global (sous-espace vectoriel engendré par les variables de $X = [X^1, X^2, \dots, X^p]$) et des sous-espaces vectoriels dits partiels définis par $\text{Im}(X^1), \dots, \text{Im}(X^p)$ où $\text{Im}(X^i)$ est le sous-espace vectoriel engendré par les variables de X^i .

DÉFINITION 2. Soient X^1, X^2, \dots, X^p , p groupes de variables mesurés sur les mêmes individus. Soit m_i pour tout i , le nombre de variables du groupe X^i et $m = \sum_{i=1}^p m_i$ le nombre total des variables de $X = [X^1, X^2, \dots, X^p]$. On appelle analyse canonique généralisée au sens de Carroll, la recherche de $q = \inf\{m_1, m_2, \dots, m_p\}$ combinaisons linéaires notées $Z^{(k)} = XP^{(k)}$ avec $k \in \{1, 2, \dots, q\}$ et $P^k \in R^p$, $p = m_1 + \dots + m_p$ et appelées Compromis canoniques de $X = [X^1, \dots, X^p]$ solutions des problèmes suivants :

$$P^{(k)} = \arg. \max_P \left\{ \sum_{i=1}^p \text{Cor}^2(Z^{(k)}, X^i) \right\} \quad (37)$$

sous les contraintes :

$$\text{Var}(Z^{(k)}) = 1 \quad \text{et} \quad \text{Cor}(Z^{(s)}, Z^{(k)}) = 0$$

$\forall k \in \{1, 2, \dots, q\}$ et pour $s \neq 1$ et $k \in \{1, 2, \dots, q\}$. $\text{Cor}(Z^{(k)}, X^i)$ désigne la corrélation multiple entre le compromis canonique d'ordre k , $Z^{(k)}$, et les variables du groupe i .

- *Solutions de l'A.C.G. au sens de Carroll.* Le problème (37) qui consiste à rendre maximale la somme des carrés des corrélations multiples entre un vecteur compromis et des variables de chaque groupe sous la contrainte $\text{Var}(Z) = 1$ est résolu par les mathématiques classiques au moyen des multiplicateurs de Lagrange. On montre qu'à l'ordre 1, $Z^{(1)} = Z = XP$ (respectivement, $P^{(1)} = P$) est solution de l'équation

$$\left\{ \sum_{i=1}^p \Pi_{x^i} - \lambda I_{n \times n} \right\} Z = 0 \iff \left\{ \sum_{i=1}^p \Pi_{x^i} \right\} Z = \lambda Z = 0 \iff XM^t XZ = \lambda Z = 0 \quad (38)$$

où $M = \text{diag.bloc}(\Sigma_{x^1}^{-1}, \dots, \Sigma_{x^p}^{-1})$ (respectivement, de l'équation $M\Sigma_x P = \lambda P = 0$) et que les variables canoniques notées Z_i , pour tout i , vérifient les égalités : $Z_i = kk\Pi_{x^i} Z$ où kk est un paramètre de normalisation et $\Pi_{x^i} = X^i \Sigma_{x^i}^{-1t} X^i$.

Remarque 2. D'après ce qui précède, il s'ensuit que les solutions d'ordre s ($s > 1$) qui maximisent le critère de l'analyse canonique généralisée au sens de Carroll sous les contraintes que $Z^{(s)}$ est orthogonale à $Z^{(k)}$ et $Var(Z^{(s)}) = Var(Z^{(k)}) = 1$, sont les s premiers vecteurs propres de la matrice $XM^t X$ associés aux s premières valeurs propres. En d'autres termes, l'ACG proposée par Carroll est une analyse en composantes principales munie d'une métrique M bloc-diagonale définie par $M = \text{diag.bloc}(\Sigma_{x^1}^{-1}, \dots, \Sigma_{x^p}^{-1})$.

Les variables canoniques de cette ACG notées $Z_i^{(1)}, Z_i^{(2)}, \dots, Z_i^{(k)}$ où $\forall i, Z_i^{(k)} = X^i P_i^{(k)}$, sont les projections des vecteurs $Z^{(1)}, Z^{(2)}, \dots, Z^{(k)}$ (Compromis canoniques des groupes) sur le sous-espace vectoriel engendré par les variables de X^i .

3.5. COMPARAISON DES TROIS MÉTHODES

Dans cette section, nous comparons à partir d'un système d'équations dites généralisées les solutions des méthodes Sumcor, Ssqcor et de Carroll.

DÉFINITION 3. Soient A et B deux matrices de même dimension $p \times p$, on appelle produit matriciel des matrices A et B au sens d'Hadamard, une matrice notée $C = (c_{ij})_{ij}$ de dimension $p \times p$ définie par $c_{ij} = (A \times B)_{ij} = a_{ij}b_{ij}$ où a_{ij} (respectivement b_{ij}) correspond à la i -ème composante de la colonne j de A (respectivement de B).

DÉFINITION 4. Soient X^1, X^2, \dots, X^p p groupes de variables mesurées sur les mêmes individus. On appelle système d'équations généralisées associé aux différents algorithmes présentés dans la section 3, le système d'équations définis par :

$$\Delta_\lambda P = (A \times \Sigma_x) P \quad (39)$$

où $\Delta_\lambda = \text{diag}(\lambda_1 Id_{m_1}, \lambda_2 Id_{m_2}, \dots, \lambda_p Id_{m_1})$ est une matrice diagonale et $\forall i$, les $\lambda_p \in R^+$, $\Sigma_x = (\Sigma_{x^i x^j})_{ij}$ est la matrice de corrélation de $X = [X^1 | X^2 | \dots | X^p]$, $\Sigma_{x^i x^j}$ la matrice de corrélation du couple (X^i, X^j) , $P = {}^t\{P_1, P_2, \dots, P_p\}$ un vecteur de R^m , P_i un vecteur poids associé aux variables de X^i de dimension $p \times 1$ et enfin $A = (A_{ij})_{ij}$ une matrice de dimension $p \times p$ ayant pour élément générique la sous-matrice $A_{ij} = a_{ij} K_{ij}$, a_{ij} sont des réels et K_{ij} est une matrice ayant tous ces éléments égaux à 1 et de dimension $m_i \times m_j$.

La proposition suivante permet de faire un lien entre le système d'équations généralisées [39] et les méthodes Sumcor, Ssqcor et de Carroll.

PROPOSITION 11.

1. Lorsque $\lambda_i = \lambda_j$ pour tout $i \neq j$ et $A_{i,j} = K_{i,j}$ (c'est-à-dire $a_{ij} = 1$ pour tout i et j) le système [39] correspond au système d'équations qui permet de déterminer les vecteurs canoniques de la méthode de Carroll. En d'autres termes, c'est le système qui permet de calculer les valeurs et vecteurs propres de la matrice Σ .
2. Lorsque $A_{i,j} = K_{i,j}$ et $\lambda_1, \lambda_2, \dots, \lambda_p$ sont quelconques, le système [39] correspond au système d'équations qui permet de déterminer les solutions de la méthode Sumcor. En d'autres termes, c'est le système qui permet de calculer les vecteurs propres d'une matrice sous la contrainte que les sous-vecteurs constituant chaque vecteur propre sont normés à chaque itération.
3. Lorsque la matrice A est définie par blocs et que les sous-matrices A_{ij} vérifient les égalités $A_{ij} = a_{ij}K_{i,j}$ avec $a_{ij} = \text{Cor}(Z_i, Z_j)$, $\lambda_1, \lambda_2, \dots, \lambda_p$ quelconques, et K_{ij} est une matrice de dimension $m_i \times m_j$ ayant tous ces termes égaux à 1, le système [39] correspond au système qui permet de déterminer les solutions de la méthode Ssqcor.

4. PRÉSENTATION DU QUESTIONNAIRE PSYCHOSOCIAL

Afin d'illustrer l'application des analyses canoniques, nous partons du questionnaire des scripts conditionnels élaboré par Gaymard (cf. [Gaymard, 2007]) et des données (cf. Tableau 1) recueillies auprès de 43 jeunes conducteurs (moyenne d'âge = 20, 4 ans).

Ces travaux s'inscrivent dans le champ théorique des représentations sociales et de la conditionnalité des normes [Flament, 1999, 2001] ; [Gaymard, 1999, 2002]. Selon la théorie du noyau central des représentations sociales, la représentation est caractérisée par des éléments centraux « absolus » et par des éléments périphériques conditionnels qui sont liés aux pratiques conditionnelles. C'est l'intérêt porté à ces derniers qui est à l'origine de la théorie de la conditionnalité développée par Flament [1994(a), 1994(b)]. Celle-ci articule la notion de prescription et la notion de condition (« Dans telle condition il faut faire ceci ou cela... »). Ainsi Flament [1994b] constate dans le domaine routier que le refus de transgression est rare alors que les transgressions conditionnelles sont importantes (« Je roule à plus de 50km/heure si c'est la nuit »). Pour faire émerger cette conditionnalité liée aux pratiques dans le domaine de la conduite automobile, Gaymard [2007] a construit un questionnaire permettant d'identifier les conditions spécifiques dans lesquelles les conducteurs trouvent qu'il est acceptable de transgresser les règles [Gaymard, 2009] [Verkuyten, Rood-Pijpers, Elffers, Hessing, 1994]. L'intégralité du questionnaire est constitué de 127 situations de conduite réparties sur 8 scénarii présentant une bonne consistance interne ($\alpha_{global} = 0,91$). Les sujets sont interrogés sur leurs possibles transgressions concernant des situations spécifiques. Par exemple dans le cas du scénario du feu rouge : « il vous arrive de ne pas vous arrêter au feu rouge si le véhicule suivant vous colle... ».

Chaque scénario inclut plusieurs situations évaluées sur une échelle ordinale comportant 6 échelons dans l'ordre suivant: respect absolu (échelon 1 : absolument jamais), respect (échelon 2 : jamais), respect conditionnel (échelon 3 : plutôt jamais), transgression conditionnelle (échelon 4 : plutôt tout le temps), transgression (échelon 5 : tout le temps), transgression absolue (échelon 6 : absolument tout le temps). Les différentes situations proposées renvoient à des circonstances multiples comme les autres conducteurs (par exemple « le véhicule suivant vous colle »), l'infrastructure (par exemple « vous êtes sur une ligne droite »), le moment de la journée (par exemple « c'est tard le soir ou la nuit »), etc. Ce questionnaire permet ainsi de comparer la conditionnalité entre les différents scénarii et de mettre en exergue les situations les plus conditionnelles c'est-à-dire celles qui justifient le plus la transgression du scénario et à l'opposé les situations les moins conditionnelles, c'est-à-dire les plus vertueuses. L'analyse des tableaux de comptages [Gaymard, 2007] a révélé des distributions significativement différentes en fonction des scénarii. Tous les scénarii sont conditionnels à des degrés divers excepté le scénario du port de la ceinture de sécurité (à l'avant) pour lequel les conducteurs n'ont trouvé aucune circonstance justifiant le non respect de la règle (respect absolu). L'analyse par quartiles a permis de dégager les scénarii les plus conditionnels (limite de vitesse et feu orange). La méthode des quartiles reste cependant limitée puisqu'elle ne permet pas de mettre en évidence le lien entre les scénarii et d'autre part le lien entre des situations et les scénarii en tenant compte des inter-relations. Or cette approche peut s'avérer intéressante dans une perspective de mettre à jour méthodologiquement les patterns de réponses qui renverraient à la structuration des aspects normatifs de la représentation. En effet certaines recherches (Gaymard, Allain, Osiurak et Le Gall, 2006) montrent qu'il existe des modèles de transgressions légitimes. Avec une autre approche statistique (Anova), la comparaison de conducteurs jeunes et âgés au questionnaire des scripts conditionnels révèle une certaine constance dans la conditionnalité observée. En effet, les sujets jeunes et âgés évalués ne se différencient pas dans la représentation des scénarii les plus vertueux ou les plus conditionnels. Pour les deux groupes le port de la ceinture est le scénario le plus respecté et la limite de vitesse et le feu orange, les scénarii les moins respectés. Il semble donc exister des "modèles" de transgressions légitimes qui relèvent d'un consensus sur les règles qui doivent être respectées et sur celles que l'on juge acceptable de transgresser [Verkuyten, Rood-Pijpers, Elffers et Hessing, 1994]. Plus généralement les recherches sur les aspects normatifs des représentations sociales ont montré l'importance des modèles de référence avec l'analyse de régression multiple [Flament, 1999 ; Gaymard, 2003, 2009]. Plus récemment cette réflexion s'est articulée avec la mise en évidence d'un effet Guttman [Gaymard et Andrés, 2009], Flament, Guimelli et Abric [2006] faisant référence aux différentes facettes d'une représentation sociale.

Afin d'appliquer au mieux le questionnaire des scripts conditionnels à l'analyse des corrélations simple et généralisée, nous avons isolé 3 scénarii et 10 situations qui leurs étaient communes.

Nous faisons l'hypothèse que l'analyse des corrélations simple et généralisée vont permettre de dégager le lien entre les scénarii et les situations plus ou moins conditionnelles.

Questionnaire psychosocial

*Il vous arrive de brûler un feu rouge si : répondez à ces propositions en cochant une case de 1 à 6.
Il vous arrive de passer au feu orange si : répondez à ces propositions en cochant une case de 1 à 6.*

Il vous arrive de dépasser la limite de vitesse si : répondez à ces propositions en cochant une case de 1 à 6.

➤ Vous êtes avec des amis (situation 1)

1	2	3	4	5	6
Absolument jamais	Jamais	Plutôt jamais	Plutôt tout le temps	Tout le temps	Absolument tout le temps

➤ Vous êtes distrait (situation 2)

1	2	3	4	5	6
Absolument jamais	Jamais	Plutôt jamais	Plutôt tout le temps	Tout le temps	Absolument tout le temps

➤ Vous avez bu (situation 3)

1	2	3	4	5	6
Absolument jamais	Jamais	Plutôt jamais	Plutôt tout le temps	Tout le temps	Absolument tout le temps

➤ C'est tard le soir (situation 4)

1	2	3	4	5	6
Absolument jamais	Jamais	Plutôt jamais	Plutôt tout le temps	Tout le temps	Absolument tout le temps

➤ Vous êtes seul(e) dans la voiture (situation 5)

1	2	3	4	5	6
Absolument jamais	Jamais	Plutôt jamais	Plutôt tout le temps	Tout le temps	Absolument tout le temps

➤ Vous êtes pressé(e) (situation 6)

1	2	3	4	5	6
Absolument jamais	Jamais	Plutôt jamais	Plutôt tout le temps	Tout le temps	Absolument tout le temps

▶ Vous avez des enfants en bas âge dans la voiture (situation 7)

1	2	3	4	5	6
Absolument jamais	Jamais	Plutôt jamais	Plutôt tout le temps	Tout le temps	Absolument tout le temps

➤ Le véhicule suivant vous colle (situation 8)

1	2	3	4	5	6
Absolument jamais	Jamais	Plutôt jamais	Plutôt tout le temps	Tout le temps	Absolument tout le temps

➤ Vous êtes sur une ligne droite (situation 9)

1	2	3	4	5	6
Absolument jamais	Jamais	Plutôt jamais	Plutôt tout le temps	Tout le temps	Absolument tout le temps

➤ Vous vous rendez au travail (situation 10)

1	2	3	4	5	6
Absolument jamais	Jamais	Plutôt jamais	Plutôt tout le temps	Tout le temps	Absolument tout le temp

Les situations communes proposées dans le cas des 3 scénarii

Tableau 1 : scénario feu rouge (FR)						
	Absolument jamais	Jamais	Plutôt jamais	Plutôt Tout le temps	Tout le temps	Absolument tout le temps
Situation 1	21	4	7	8	2	1
Situation 2	37	3	0	3	0	0
Situation 3	40	3	0	0	0	0
Situation 4	27	9	4	1	2	0
Situation 5	35	4	2	1	0	1
Situation 6	20	3	2	14	2	2
Situation 7	41	2	0	0	0	0
Situation 8	25	7	5	6	0	0
Situation 9	31	7	2	3	0	0
Situation 10	30	4	2	5	2	0
Tableau 2 : scénario feu orange (FO)						
Situation 1	21	7	8	5	0	2
Situation 2	4	5	8	14	6	6
Situation 3	38	3	1	1	0	0
Situation 4	12	4	5	15	5	2
Situation 5	12	4	6	14	7	0
Situation 6	5	2	3	23	6	4
Situation 7	34	6	2	1	0	0
Situation 8	9	8	2	14	5	5
Situation 9	15	6	8	10	3	1
Situation 10	18	5	6	12	2	0
Tableau 3 : scénario limite de vitesse (LV)						
Situation 1	13	10	9	6	4	1
Situation 2	10	3	3	13	11	3
Situation 3	36	3	1	3	0	0
Situation 4	8	7	8	13	4	3
Situation 5	3	3	5	20	5	7
Situation 6	1	0	4	19	12	7
Situation 7	33	6	2	2	0	0
Situation 8	11	8	6	10	6	2
Situation 9	11	1	10	13	5	3
Situation 10	7	2	6	15	9	4

Table 1. Tableaux des données. *Ce sont des tableaux de comptages croisant 10 situations et les réponses aux 10 situations obtenues auprès de 43 jeunes conducteurs. Pour l'application de l'analyse canonique simple et généralisée à ces données, on considère que les individus (les lignes des tableaux) sont les 10 situations tandis que, les variables au nombre de 6 (Absolument jamais, Jamais, Plutôt jamais, Plutôt Tout le temps, Tout le temps, Absolument tout le temps), par tableau, correspondent aux réponses aux 10 situations. La valeur d'une variable sur un individu correspond à des effectifs observés (le nombre de fois qu'une même réponse a été donnée dans une situation).*

5. RÉSULTATS

5.1. PREMIÈRE ÉTAPE : ANALYSES SÉPARÉES

Il s'agit d'utiliser l'analyse des corrélations canoniques pour mettre en relation, pour chacun des 3 scénarii, les 10 situations et les 6 réponses obtenues à ces 10 situations. Ces analyses de corrélations canoniques sont équivalentes à des analyses factorielles de correspondances réalisées à partir des Tableaux 1, 2 et 3 (cf. Tableau 1), et les résultats de la Proposition 5.

Les résultats obtenus avec les trois analyses des corrélations canoniques (cf. Figure 1) montrent qu'il y a un lien global, quel que soit le scénario, entre l'ensemble des situations et les réponses données par les 43 jeunes conducteurs à chacune des 10 situations :

feu rouge : $\text{Cor}(Z_1, Z_2) = 0.41$, $\chi^2_{(ddl=45, n_1=10, n_2=6)} = 111.91$, $P_{value} = 0$;

feu orange : $\text{Cor}(Z_1, Z_2) = 0.56$, $\chi^2_{(ddl=45, n_1=10, n_2=6)} = 171.84$, $P_{value} = 0$;

limite de vitesse : $\text{Cor}(Z_1, Z_2) = 0.61$, $\chi^2_{(ddl=45, n_1=10, n_2=6)} = 206.79$ et $P_{value} = 0$).

Scénario feu rouge : Pour ce scénario du feu rouge, le premier axe factoriel explique 65.62 % d'inertie tandis que le deuxième explique 21.36 % d'inertie. La carte factorielle obtenue permet de dégager trois classes. Une 1^e classe formée par des situations justifiant le respect absolu des règles de conduite : « vous avez des enfants en bas âge dans le véhicule », « vous avez bu », « vous êtes distrait », « vous êtes seul dans la voiture ». Une 2^e classe constituée par des situations qui renvoient au respect et au respect conditionnel des règles : « vous vous rendez au travail », « vous êtes avec des amis », « c'est tard le soir », « le véhicule suivant vous colle ». Et enfin, une 3^e classe regroupant les situations justifiant la transgression des règles et qui concerne la situation : « vous êtes pressé ». Dans l'étude de [Gaymard, 2007] le scénario « feu rouge » est peu conditionnel si ce n'est pour 2 situations : « vous êtes pressé » et « vous êtes avec des amis » dont les quartiles (q_{75} %) sont respectivement de 3.52 et 3.03.

Scénario feu orange : Le 1^{er} axe factoriel explique 77.86 % d'inertie tandis que le 2^e explique 8.78 % d'inertie. La carte factorielle obtenue avec les données du scénario feu orange montre un regroupement des situations en trois classes. Une première classe formée par des situations justifiant le respect absolu des règles de conduite, il s'agit des situations également respectées dans le cadre du scénario feu rouge : « vous avez des enfants en bas âge dans le véhicule », « vous avez bu ». Une 2^e classe formée à partir des situations justifiant le respect des règles de conduite : le fait de se rendre au travail ou d'être sur une ligne droite. Une 3^e classe regroupant des situations justifiant la transgression des règles de conduite. Il s'agit des situations : « vous êtes pressé », « vous êtes distrait », « c'est tard le soir », « le véhicule suivant vous colle » et « vous êtes seul dans la voiture ».

Scénario limite de vitesse : pour ce scénario, le 1^{er} axe factoriel explique 76.44 % d'inertie tandis que le 2^e explique 15.02 % d'inertie. La carte factorielle obtenue avec ce scénario montre un regroupement des situations en trois classes. Une 1^e classe formée par des situations justifiant le respect absolu des règles de conduite : « vous avez des enfants en bas âge dans le véhicule », « vous avez bu ». Une 2^e classe formée à partir des situations qui justifient le respect et le respect conditionnel des règles de conduite : « vous êtes avec des amis », « c'est tard le soir », « le véhicule suivant vous colle » et enfin, une 3^e classe regroupant les situations justifiant la transgression des règles de conduite : « vous êtes distrait », « vous vous rendez au travail », « vous êtes sur une ligne droite », « vous êtes seul dans la voiture », « vous êtes pressé ». Dans l'étude de [Gaymard, 2007], ce scénario est le plus conditionnel et les quartiles sont élevés.

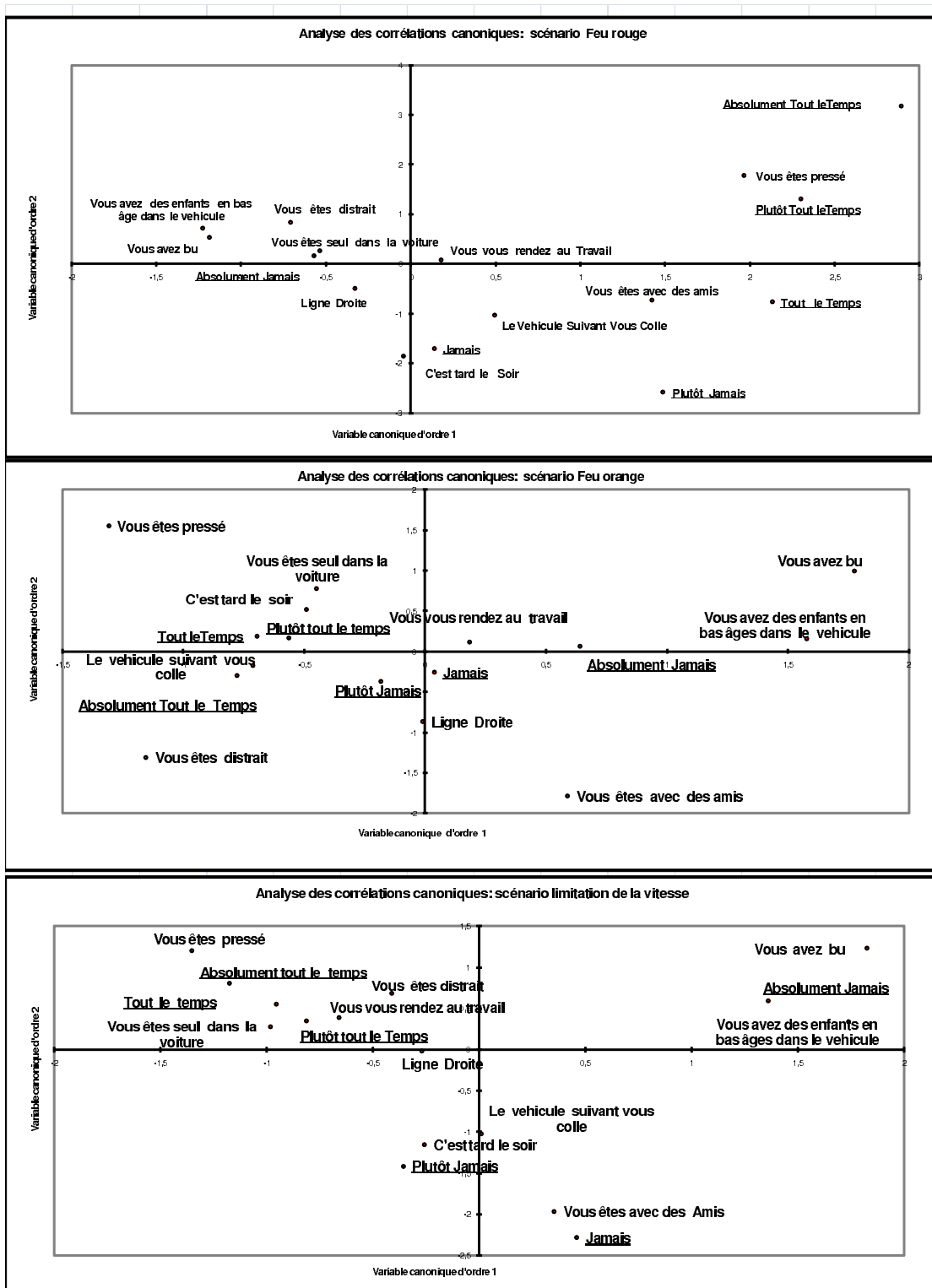


FIGURE 1. Analyse canonique simple : analyse de chaque scénario

5.2. DEUXIÈME ÉTAPE : ANALYSE GLOBALE

Il s'agit de comparer les 3 scénarii en utilisant les méthodes d'analyse canonique généralisée (la méthode Sumcor, la méthode Ssqcor et la méthode de Carroll). Les

résultats de chaque analyse vont nous permettre d'une part de mettre en évidence, de façon globale, les liaisons entre les scénarii et d'autre part, de mettre en évidence, pour chaque scénario, les proximités qu'il y a, en tenant compte des inter-relations des scénarii, entre les situations et les degrés d'échelle (mise en évidence des situations plus ou moins transgressées).

Méthodes Sumcor et Ssqcor. Pour ces deux méthodes, nous allons nous intéresser d'une part aux corrélations (> 0.5) entre les variables initiales et les variables canoniques d'ordres 1 et 2 (respectivement, aux corrélations entre les groupes de variables) et d'autre part, aux proximités qu'il y a entre les situations et les réponses associées. L'examen des résultats obtenus (cf. Tableaux 2 et 3) montre l'existence d'une forte liaison linéaire entre les 3 scénarii (toutes les corrélations, en valeurs absolues, entre les variables canoniques sont supérieures à 0.7).

Méthode Sumcor						
	FR	FO	LV	FR	FO	LV
FR	1.0000	0.9717	0.9063	1.0000	0.8182	0.7374
FO	0.9717	1.0000	0.9522	0.8182	1.0000	0.8865
LV	0.9063	0.9522	1.0000	0.7374	0.8865	1.0000

Table 2. Corrélations entre les variables canoniques d'ordre 1 (colonnes 1-3) et d'ordre 2 (colonnes 4-6). FR : feu rouge ; FO : feu orange ; LV : limite de vitesse

Méthode Ssqcor						
	FR	FO	LV	FR	FO	LV
FR	1.0000	-0.8921	0.8137	1.0000	0.7017	0.6953
FO	-0.8921	1.0000	-0.9589	0.7017	1.0000	0.9522
LV	0.8137	-0.9589	1.0000	0.6953	0.9522	1.0000

Table 3. Corrélations entre les variables canoniques d'ordre 1 (colonnes 1-3) et d'ordre 2 (colonnes 4-6). FR : feu rouge ; FO : feu orange ; LV : limite de vitesse

Concernant le scénario Feu Rouge, la méthode Sumcor et la méthode Ssqcor montrent, d'après les Figures 2 et 3, une opposition entre les situations qui justifient la transgression et les situations davantage respectées. La méthode Sumcor montre une opposition entre la transgression (modalité tout le temps), la transgression conditionnelle (modalité plutôt tout le temps) et le respect conditionnel (modalité plutôt jamais) concernant les situations : « vous êtes pressé », « vous êtes avec des amis », « vous vous rendez au travail », « le véhicule suivant vous colle » et le respect absolu avec les situations « vous avez bu » ; « vous avez des enfants en bas âge dans le véhicule » et « vous êtes seul ». La méthode Ssqcor oppose le respect absolu avec les modalités « vous avez bu » ; « vous avez des enfants en bas âge dans le véhicule », « vous êtes distrait » à la transgression, la transgression absolue et au respect conditionnel associés aux situations « vous êtes pressé », « le véhicule suivant vous colle », « vous êtes avec des amis », « vous vous rendez au travail ».

Comme pour le feu rouge, les deux méthodes montrent pour le scénario du feu orange une opposition entre la transgression et le respect de la règle. La méthode

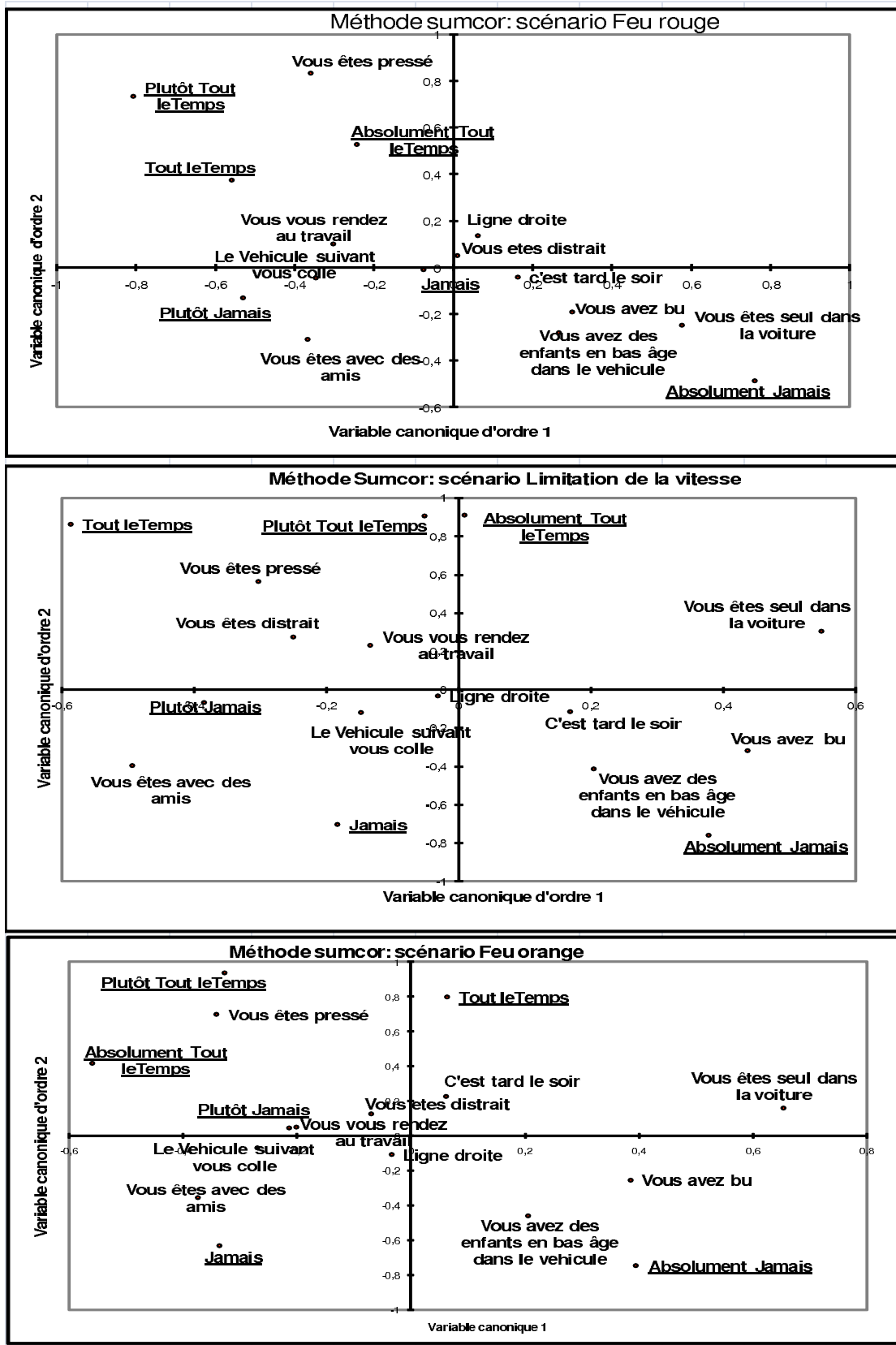


FIGURE 2. Analyse canonique généralisée : méthode Sumcor

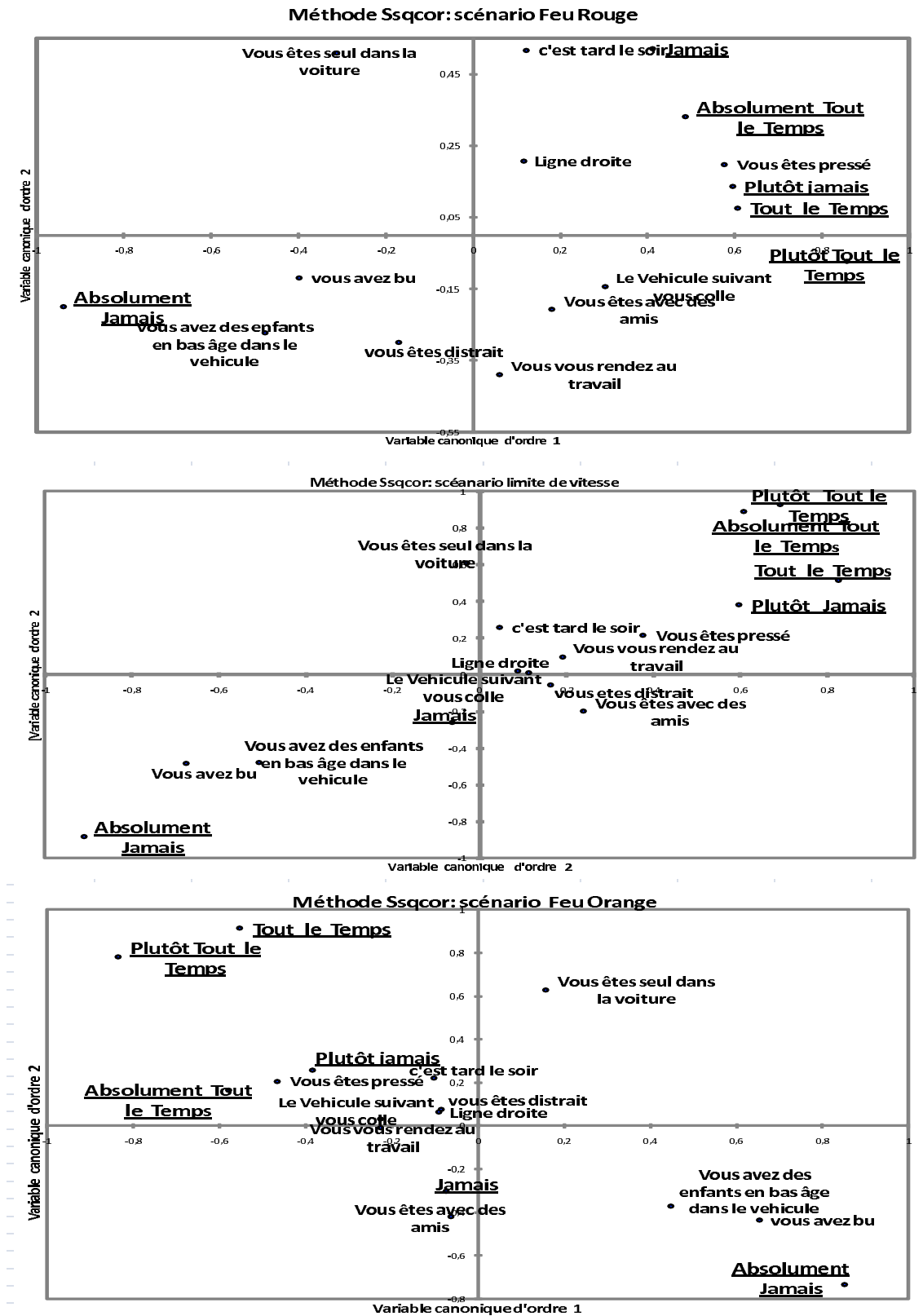


FIGURE 3. Analyse canonique généralisée : méthode Ssqcor

Corrélations entre les variables initiales et les variables canoniques d'ordres 1 et 2						
	Méthode Sumcor					
	FR		FO		LV	
	Vc1	Vc2	Vc1	Vc2	Vc1	Vc2
Absolument jamais	0.76	-0.49	0.39	-0.75	0.38	-0.76
Jamais	-0.08	-0.01	-0.34	-0.63	-0.18	-0.70
Plutôt jamais	-0.53	-0.13	-0.20	0.05	-0.39	-0.06
Plutôt tout le temps	-0.81	0.74	-0.33	0.94	-0.05	0.91
Tout le temps	-0.56	0.37	0.06	0.80	-0.59	0.86
Absolument tout le temps	-0.24	0.53	-0.56	0.41	0.01	0.91
	Méthode Ssqcor					
	FR		FO		LV	
	Vc1	Vc2	Vc1	Vc2	Vc1	Vc2
Absolument jamais	-0.94	-0.20	0.85	-0.73	-0.91	-0.88
Jamais	0.41	0.52	-0.07	-0.30	-0.06	-0.26
Plutôt jamais	0.59	0.14	-0.38	0.26	0.60	0.38
Plutôt tout le temps	0.86	-0.07	-0.84	0.78	0.69	0.93
Tout le temps	0.61	0.08	-0.55	0.91	0.83	0.52
Absolument tout le temps	0.49	0.33	-0.58	0.16	0.61	0.89

Table 4. Corrélations entre les variables initiales et les variables canoniques d'ordres 1 et 2. FR : feu rouge ; FO : feu orange ; LV : limite de vitesse

Sumcor oppose les modalités de transgression absolue et transgression conditionnelle associées à la situation « vous êtes pressé » au respect absolu qui concerne les situations « vous avez bu », « vous avez des enfants en bas âge dans le véhicule » tandis que, la méthode Ssqcor oppose la transgression associée aux modalités « vous êtes pressé », « le véhicule suivant vous colle », « vous vous rendez au travail » au respect absolu associé aux modalités « vous avez bu », « vous avez des enfants en bas âge dans le véhicule ».

Nous retrouvons avec le scénario Limite de vitesse et pour les deux méthodes, l'opposition entre le respect et la transgression. La méthode Sumcor oppose la transgression associée aux modalités « vous êtes pressé », « vous êtes distrait », « vous vous rendez au travail » et le respect absolu associé aux modalités « vous avez bu », « vous avez des enfants en bas âge dans le véhicule ». À l'inverse, la méthode Ssqcor montre une opposition entre le respect associé aux mêmes situations « vous avez bu », « vous avez des enfants en bas âge dans le véhicule » et la transgression associée aux situations « vous êtes pressé », « vous êtes seul dans la voiture », « c'est tard le soir ».

b) Analyse canonique généralisée de Carroll. La généralisation de l'analyse canonique proposée par Carroll a permis d'obtenir une base orthogonale de dimension 10 engendrée par l'ensemble des 18 variables (6 variables pour le scénario Limite de vitesse, 6 variables pour le scénario Feu orange et, 6 variables pour le scénario Feu rouge). Avec 10 valeurs propres non nulles (3, 2.893, 2.773, 2.529, 2.051, 1.842, 1.156, 1.058, 0.587, 0.111), ce sous-espace se révèle être de dimension 10. Les 10 compromis canoniques associés à ces 10 valeurs propres définissent des sous-espaces factoriels, en particulier des plans factoriels permettant d'obtenir des représenta-

tions synthétiques de l'ensemble des 18 variables (6 variables par scénario) ou de l'ensemble des 10 situations (individus). Pour cette méthode, la première valeur propre est égale à 3 (le nombre de scénarii à analyser). Cela nous conduit à la première conclusion suivante: le premier compromis entre l'ensemble des variables passe par les axes d'allongement maximal des 3 groupes c'est-à-dire qu'il existe une forte relation linéaire entre les sous-espaces associés aux tableaux des scénarii.

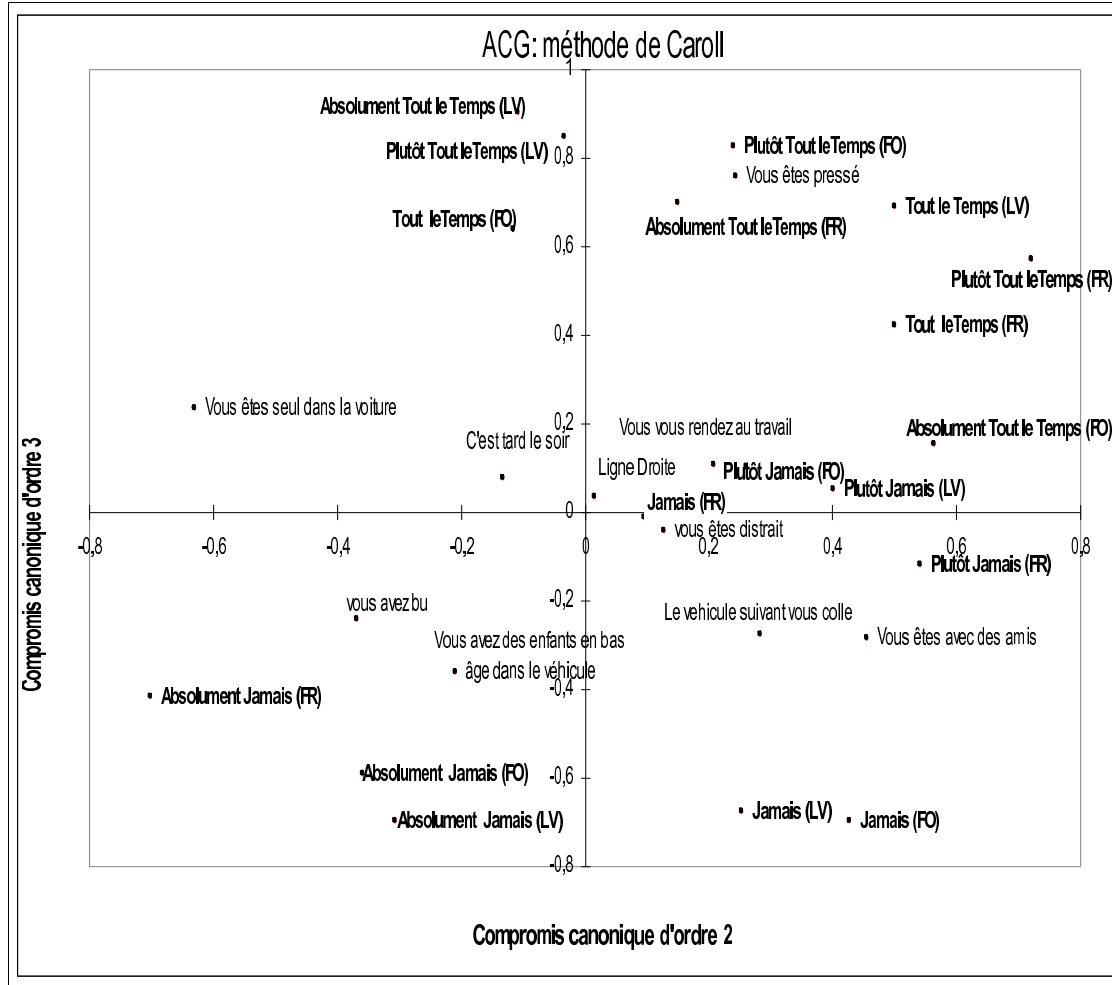


FIGURE 4. Représentation des situations et des réponses dans le 2^e plan factoriel

Corrélations entre variables canoniques d'ordre 1				Corrélations entre les variables canoniques d'ordre 2.		
Méthode de Carroll						
	FR	FO	LV	FR	FO	LV
FR	1.0000	0.9337	0.8183	1.0000	0.9572	0.9038
FO	0.9337	1.0000	0.7451	0.9572	1.0000	0.9774
LV	0.8183	0.7451	1.0000	0.9038	0.9774	1.0000

Table 5. Méthode de Carroll. Mesures de liaison entre les 3 scénarii

Le Tableau 5 (respectivement, le Tableau 6) donne les corrélations (deux-à-deux) entre les variables canoniques (respectivement, les corrélations entre les variables initiales et les compromis canoniques) tandis que, le Tableau 7 donne les coordonnées des 10 situations sur les 3 premiers compromis canoniques. Enfin, dans la Figure 4 nous montrons les liens qui existent entre les 10 situations et les réponses associées.

On notera, à partir du Tableau 5, que toutes les corrélations (deux-à-deux) entre les variables canoniques d'ordre 1 (respectivement, d'ordre 2) sont supérieures à 0.7 et donc, il existe bien une relation linéaire entre les différents scénarii.

Les résultats présentés dans les Tableaux 6, 7 et la Figure 4 montrent que la situation « vous êtes seul dans la voiture » est associée à la transgression pour la limite de vitesse et le feu orange. D'autre part, les situations « vous avez des enfants en bas âge » et « vous avez bu » sont associées au respect absolu dans les trois scénarii. Les situations « vous êtes avec des amis » et « le véhicule suivant vous colle » sont associées au respect de limite de vitesse et du feu orange et au respect conditionnel pour le feu rouge. La modalité « vous êtes pressé » est associée à la transgression absolue pour le feu rouge et orange, à la transgression pour le feu orange et la limite de vitesse et à la transgression conditionnelle pour les feux orange et rouge.

Corrélations entre les variables et les 3 premiers compromis canoniques									
	Groupe1 (FR)			Groupe2 (FO)			Groupe3 (LV)		
	Cp1	Cp2	Cp3	Cp1	Cp2	Cp3	Cp1	Cp2	Cp3
Absolument jamais	0.52	-0.70	-0.42	0.61	-0.36	-0.59	0.64	-0.31	-0.70
Jamais	0.10	0.10	-0.01	0.55	0.43	-0.70	0.66	0.25	-0.68
Plutôt jamais	0.09	0.54	-0.12	-0.14	0.26	0.10	-0.06	0.40	0.05
Plutôt tout le temps	-0.77	0.72	0.57	-0.79	0.24	0.83	-0.70	-0.03	0.85
Tout le temps	-0.47	0.50	0.42	-0.48	-0.11	0.64	-0.84	0.50	0.69
Absolument tout le temps	-0.64	0.15	0.70	-0.36	0.56	0.15	-0.73	-0.11	0.90

Table 6. Corrélations entre les variables initiales (les lignes du tableau) et les 3 premiers compromis canoniques (les colonnes du tableaux). Cp: compromis canonique

Situation	Cp (1)	Cp (2)	Cp (3)
Vous êtes avec des amis	-0.316	0.456	-0.287
Vous êtes distrait	-0.316	0.128	-0.041
Vous avez bu	-0.316	-0.367	-0.243
C'est tard le soir	-0.316	-0.132	0.075
Vous êtes seul dans la voiture	-0.316	-0.629	0.235
Vous êtes pressé	-0.316	0.244	0.759
Vous avez des enfants en bas âge dans le véhicule	-0.316	-0.208	-0.363
Le véhicule suivant vous colle	-0.316	0.284	-0.275
Ligne droite	-0.316	0.016	0.033
Vous vous rendez au travail	-0.316	0.209	0.105

Table 7. Coordonnées des 10 situations sur les 3 premiers compromis canoniques

6. CONCLUSION

Dans cet article nous avons, d'une part décrit les méthodes d'analyse des corrélations simple et généralisée et, d'autre part réalisé une comparaison théorique et numérique de trois méthodes d'analyse des corrélations canoniques généralisées (la méthode Sumcor, la méthode Ssqcor et la méthode de Carroll). Les résultats numériques obtenus montrent que la méthode d'analyse canonique et ses généralisations sont des outils adaptés pour décrire et analyser les données recueillies dans le champ des sciences humaines. Par ailleurs ces données ayant été préalablement soumises à une analyse par quartile [Gaymard, 2007], nous pouvons croiser les résultats et dégager l'apport et la spécificité de la méthode généralisée qui permet de traiter plusieurs ensembles de variables. Concernant les analyses séparées, elles montrent pour le feu rouge, le feu orange et la limite de vitesse, une opposition entre le respect et la transgression qui au travers de l'analyse par quartiles renvoyait aux scenarii les moins conditionnels (quartiles les moins élevés) et aux scenarii les plus conditionnels (quartiles les plus élevés). Néanmoins l'analyse canonique permet un regroupement en classes par les liaisons linéaires et une lecture plus globale au travers de la carte factorielle. Pour le scénario du feu rouge, l'analyse par quartile avait montré qu'il était peu conditionnel, les situations les moins transgressables étant représentées par les modalités « vous avez des enfants en bas âge » et « vous avez bu », les plus conditionnelles par les modalités « vous êtes pressé » et « vous êtes avec des amis ». L'analyse canonique simple isole trois classes, deux justifiant le respect des règles et la dernière étant liée à la transgression. Une classe est constituée de quatre situations proches du respect absolu, une autre de quatre situations proches du respect conditionnel et la dernière regroupe les deux modalités qui justifient davantage la transgression. Comme cela avait été observé avec l'analyse par quartiles, ce découpage montre que le feu rouge est peu conditionnel.

L'analyse canonique simple du feu orange permet de retrouver les situations les plus respectées mises en évidence dans l'analyse par quartiles. À la différence du feu rouge les deux modalités « vous avez des enfants en bas âge » et « vous avez bu » forment une classe unique. Une autre classe est constituée de deux situations renvoyant au respect conditionnel. Enfin il y a cinq situations qui justifient les transgressions rappelant que ce scénario est plus conditionnel que le feu rouge. L'analyse canonique simple de la limite de vitesse, scénario le plus conditionnel permet de dégager également 3 classes : deux justifiant le respect des règles et une la transgression. Comme pour le feu orange, les deux modalités : « vous avez des enfants en bas âge » et « vous avez bu » constituent la classe du respect absolu de la règle. Une seconde classe regroupe les trois situations qui illustrent le respect conditionnel, enfin la troisième classe regroupe les cinq situations justifiant le plus les transgressions de la règle et dont les quartiles $q_{0,75}$ étaient les plus élevés.

Concernant l'analyse globale pour le scénario feu rouge, les méthodes Sumcor et Ssqcor montrent une opposition entre les différents niveaux de transgression (absolue, conditionnelle) et le respect absolu. Pour le scénario du feu orange et de la limite de vitesse, nous retrouvons cette opposition, la zone du respect absolue n'étant plus que représentée par les deux modalités « vous avez bu » et « vous avez des enfants en bas âge dans la voiture ». L'analyse globale montre ainsi pour les 3

scénarii une structuration identique, le feu rouge étant moins conditionnel. L'analyse canonique généralisée de Carroll permet de mettre en relation pour les 3 scénarii, les situations qui s'inscrivent conjointement dans le respect ou la transgression. Nous observons ainsi que les 2 situations « vous avez bu » et « vous avez des enfants en bas âge dans la voiture » sont associées au respect absolu pour les 3 scénarii. La situation « vous êtes seul dans la voiture » justifie la transgression du feu orange et de la limite de vitesse mais l'analyse généralisée montre que comparativement à la modalité « vous êtes pressé », les modalités « le véhicule suivant vous colle » et « vous êtes avec des amis » justifient peu les transgressions des 3 scénarii. Ainsi à la différence de l'analyse par quartiles, nous pouvons avec cette méthode mettre en évidence le lien entre les scénarii et le lien entre les situations et les scénarii en tenant compte des inter-relations. Cette approche permet d'illustrer par le biais des *patterns* de réponses, la structuration interne des transgressions légitimes au travers de la relation linéaire entre les différents scénarii. Nous voyons avec cette méthode qu'il n'existe pas de modalité *contradictoire* c'est-à-dire s'inscrivant dans le respect pour un scénario et la transgression pour un autre. Il apparaît ainsi que les dix situations ont la même *valeur* dans les 3 scénarii. Par ailleurs la méthode généralisée apporte une autre lecture de la transgression. Par exemple la modalité « vous êtes avec des amis » faisait partie des plus conditionnelles dans l'analyse par quartile du feu rouge. De même la modalité « vous êtes distrait » faisait partie des plus conditionnelles dans l'analyse du feu orange [Gaymard, 2007]. L'analyse conjointe des scénarii et des situations place ces deux dernières dans le respect et le respect conditionnel (méthode de Carroll). Cette méthode permet donc d'apporter une autre interprétation de la conditionnalité en tenant compte à la fois des scénarii et des situations. Il faut cependant garder à l'esprit que les scénarii proposés à l'origine n'étaient pas identiques [Gaymard, 2007] car pour appliquer cette méthode il nous a fallu isoler 3 scénarii et 10 situations communes. Néanmoins dans une perspective future il nous semble pertinent de réfléchir en amont à la construction du questionnaire comme celui des scripts conditionnels pour une analyse canonique généralisée. L'analyse d'une situation sociale doit inclure plusieurs facteurs et leurs relations. La mise en évidence de modèles de transgressions légitimes au niveau des scénarii [Gaymard, Allain, Osiurak, Le Gall, 2006] peut être affinée en intégrant les situations. Enfin il nous semble que cette méthode s'inscrit dans la démarche méthodologique de mettre à jour les aspects normatifs des représentations sociales comme les dernières recherches sur l'effet Guttman [Flament, Guimelli et Abric, 2006 ; Gaymard et Andrés, 2009].

BIBLIOGRAPHIE

- CARROLL J.D., "Generalization of canonical correlation analysis to three or more sets of columns", *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, p. 227-228.
- FLAMENT C., « Structure et dynamique des représentations sociales », *Les représentations sociales*, D. Jodelet (éd.), Paris, Presses Universitaires de France, 1989, p. 204-219.
- FLAMENT C., « Structure, dynamique et transformation des représentations sociales », *Pratiques sociales et représentations*, J.-C. Abric (éd.), Paris, Presses Universitaires de France, 1994(a), p. 37-57.
- FLAMENT C., « Aspects périphériques des représentations sociales », *Structures et transformations sociales*, C. Guimelli (éd), Lausanne, Delachaux et Niestlé, 1994(b), p. 85-118.
- FLAMENT C., « La représentation sociale comme système normatif », *Psychologie et société* 1, 1999, p.29-53.
- FLAMENT C., « Représentation sociale et normativité : quelques pistes », *Penser la vie, le social, la nature, Mélanges en l'honneur de Serge Moscovici*, F. Buschini, N. Kalampakis (éds.), Paris, Publications de la Maison des sciences de l'Homme, 2001, p. 257-261.
- FLAMENT C., GUIMELLI C., ABRIC J.-C., « Effets de masquage dans l'expression d'une représentation sociale », *Les Cahiers Internationaux de Psychologie Sociale* 69, 2006, p. 15-31.
- GAYMARD S., *Les études supérieures comme enjeu dans un contexte de négociation implicite entre les filles d'origine maghrébine et leurs parents : aspects conditionnels et normatifs des représentations sociales dans une situation de biculturalisme*, Thèse de Doctorat publiée, Aix-en-Provence, Université de Provence, 1999.
- GAYMARD S., « Représentation sociale et algèbre de Boole : une étude des modèles normatifs dans une situation de biculturalisme », *Revue Internationale de Psychologie Sociale* 15(1), 2002, p.163-184.
- GAYMARD S., « Représentation sociale et modèles normatifs dans un contexte biculturel », *Les Cahiers Internationaux de Psychologie Sociale* 58, 2003, p. 50-58.
- GAYMARD S., « La représentation de la conduite chez les jeunes conducteurs. Une étude de la conditionnalité routière », *Recherche Transports Sécurité* 97, 2007, p. 339-359.
- GAYMARD S., "Norms in social representations: two studies with French young drivers", *The European Journal of Psychology Applied to Legal Context* 1(2), 2009, p. 165-181.
- GAYMARD S., ALLAIN P., OSIURAK F., LE GALL D., « Étude de la conditionnalité routière chez des sujets jeunes et âgés », *Actes du 6^e congrès international de psychologie sociale de la langue française*, 2006, page 258.
- GAYMARD S., ANDRES S., « Représentations, modèles normatifs et compétences sociales : une étude exploratoire dans un lycée sensible en Espagne », *Revue Internationale de Psychologie Sociale* 2, 2009, p. 43-69.
- HORST P., "Relations among m sets of measures", *Psychometrika* 26(2), 1961(a), p. 129-149.
- HORST P., "Generalized canonical correlations and their applications to experimental data", *Journal of clinical psychology, Monograph supplement* 14, 1961(b), p. 331-347.
- HOTELLING H., "Relations between two sets variables", *Biometrika* 28, 1936, p. 321-377.
- KETTERING R.J., "Canonical analysis of several sets of variables", *Biometrika* 58(3), 1971, p. 433-451.

KETTERING R.J., "Canonical analysis", *Encyclopedia of statistical Sciences*, S. Kotz, N.L. Johnson (eds.), New York, Wiley, 1983, p. 354-365.

NZOBOUNSA V., *Analyse canonique généralisée : méthode, applications et sélection des variables dans les groupes*, Thèse de doctorat publiée, Rennes, Université de Rennes 2, 2001.

NZOBOUNSA V., DHORNET T., « Écart : une nouvelle méthode d'analyse canonique généralisée (ACG) », *Revue de Statistique Appliquée* 51(4), 2003, p. 57-82.

SAPORTA G., *Probabilité, analyse des données et statistique*, première édition, Paris, Éditions Technip, 1990, 493 pages.

STEEL R.G., "Minimum generalized variance for a set of linear functions", *The Annals of Mathematical Statistics* 22, 1951, p. 456-60.

VERKUYTEN M., ROOD-PIJPERS E., ELFFERS H., HESSING D.J., "Rules for breaking formal rules: Social representations and everyday rule-governed behavior", *The Journal of Psychology* 128(5), 1994, p. 485-497.