

L'ANALYSE FACTORIELLE MULTIPLE DUALE APPLIQUEE A L'ANALYSE DE LA COLLECTE MULTIMODE

Marine GUILLERM(), Tiaray RAZAFINDRANOVA(**)*

()Insee, Division méthodes appliquées de l'économétrie et de l'évaluation*

*(**)Insee, Division recueil et traitement de l'information*

Résumé

Les organismes producteurs d'information statistique envisagent de développer les enquêtes par internet et la collecte multimode notamment pour des raisons budgétaires. Mais de nombreuses questions se posent et en particulier celle des effets de mode : une personne enquêtée ne répond pas forcément de la même manière selon le mode de collecte. Pour étudier ce type de problème d'ordre méthodologique, une enquête expérimentale auto-administrée (internet ou papier) "Vols, violences et sécurité" a été réalisée par l'Insee dans un objectif de comparaison avec l'enquête de victimation en face à face "Cadre de vie et sécurité". L'analyse porte ici sur les réponses aux questions d'opinion en matière de sécurité particulièrement susceptibles d'être affectées par des effets de mode.

Alors que l'étude des effets de modes repose habituellement sur de l'analyse univariée, une autre optique est proposée ici en s'intéressant aux liaisons entre variables. L'analyse factorielle multiple duale (AFMD) proposée par Lê et Pagès (2010) est particulièrement adaptée à ce type d'analyse, où l'on dispose de plusieurs tableaux de données (un tableau par groupe : ici, le mode de collecte) dont les variables sont communes, et qui peuvent par conséquent être empilés pour former un multi-tableau de données. L'AFMD construit un unique jeu de composantes principales et assure que la contribution des différents groupes d'individus dans sa construction est équilibrée. L'apport de l'AFMD est de disposer d'un repère unique adapté à la représentation des corrélations entre variables aux niveaux global (pour l'ensemble des individus) d'une part et partiels (au sein de chaque groupe d'individus) d'autre part. La superposition des nuages global et partiels des variables a un sens et permet de visualiser les différences de corrélations entre variables d'un groupe d'individus à un autre. Le principal résultat de l'analyse est que les opinions exprimées apparaissent plus tranchées dans le cadre auto-administré. Les comportements de non-réponse apparaissent également plus répétés quand l'individu répond sur questionnaire papier par rapport à une interrogation en face à face.

Les différences constatées peuvent néanmoins relever, pour partie, de phénomènes liés à

la sélection. Un contrôle de la sélection a ainsi été effectué via un *matching* sur caractéristiques sociodémographiques préalable à l'AFMD pour jauger la robustesse de nos résultats. Ce contrôle de la sélection sur les caractéristiques observables modifie peu les résultats.

Abstract

Before generalizing the use of the internet to collect data on households, probably in a mixed modes way, some issues need a bit of investigation. In particular mode effects have to be clearly understood and that's why Insee has decided to launch a series of experimental web surveys. An experimental safety survey on web and paper is part of this program of experiments and is compared to the traditional face-to-face safety survey : key variables of the analysis of mode effects are opinion questions on safety.

The dual multiple factor analysis (DMFA) is a relevant technique to make such investigations as the same set of variables is measured on different subsets of persons. The DMFA allows graphical representations of the association structures on the whole set of individuals and in the different subsets (modes of data collection) within the same factor space.

The main result of the analysis is that opinions are more clear-cut in the auto-administered configuration. Non-response to an item is a more repeated behavior in the paper survey than in the face-to-face survey. To control for selection effects, matching on sociodemographic characteristics has been made before another DMFA : conclusions are still the same and it sounds quite relevant to interpret results as pure mode effects.

Mots-clés : effets de mode, analyse factorielle multiple duale, multimode, collecte

Ce document a bénéficié des relectures et remarques de Brigitte Gelein, Pauline Givord et Gaël de Peretti. Nous les en remercions. Nous restons seuls responsables des erreurs qui pourraient y demeurer.

Table des matières

| | | |
|----------|--|-----------|
| 1 | L'analyse factorielle multiple duale | 6 |
| 1.1 | Pour quelles données et quelle problématique ? | 6 |
| 1.2 | Principe général de l'AFMD | 8 |
| 1.2.1 | Etape préliminaire : Centrer et réduire les données à l'intérieur de chaque groupe | 8 |
| 1.2.2 | AFMD = ACP sur le tableau des données centrées et réduites par groupe | 8 |
| 1.2.3 | Les tableaux \tilde{Z} en tant qu'éléments supplémentaires | 9 |
| 1.3 | Les caractéristiques de l'AFMD | 10 |
| 1.3.1 | Une analyse qui équilibre l'influence des groupes | 10 |
| 1.3.2 | Une analyse qui offre plusieurs points de vue | 11 |
| 1.4 | Lecture et interprétation des résultats | 12 |
| 1.4.1 | Le point de vue global | 12 |
| 1.4.2 | Les nuages partiels des variables | 13 |
| 1.4.3 | Les groupes d'individus | 14 |
| 1.4.4 | L'étude de l'inter-groupe | 15 |
| 1.4.5 | Variables supplémentaires | 16 |
| 1.5 | L'extension aux variables catégorielles | 16 |
| 1.6 | En pratique | 16 |
| 2 | Application de l'AFMD à la collecte multimode | 19 |
| 2.1 | Présentation des données | 19 |
| 2.1.1 | L'enquête Cadre de vie et sécurité | 19 |
| 2.1.2 | L'enquête expérimentale Vols, violence et sécurité | 19 |
| 2.1.3 | Quelques précisions sur le taux de réponse aux enquêtes | 20 |
| 2.1.4 | Les variables d'intérêt | 20 |
| 2.2 | L'AFMD sur les données des enquêtes CVS et VVS | 21 |
| 2.2.1 | Le nuage global des modalités | 21 |
| 2.2.2 | Les nuages partiels : corrélations intra-groupe | 25 |
| 2.2.3 | Les nuages partiels : étude de l'inter-groupe | 26 |
| 2.2.4 | D'une analyse de données groupées par mode... à une analyse de données groupées par enquête | 32 |
| 2.2.5 | Projection de variables supplémentaires | 34 |
| 2.2.6 | Robustesse des résultats après appariement (ou <i>matching</i>) | 35 |
| | Annexes | 44 |
| 1 | Annexe : le nuage des groupes d'individus | 44 |
| 1.1 | Le produit scalaire de Hilbert-Schmidt | 44 |
| 1.2 | La base des matrices U_s | 44 |
| 2 | Questionnaire (extrait) | 46 |
| 3 | Principaux résultats des AFMD | 47 |
| 3.1 | Les valeurs propres | 47 |
| 3.2 | Les nuages partiels dans la première AFMD | 47 |
| 3.3 | Les coordonnées des groupes d'individus | 49 |

| | | |
|----------|---|-----------|
| 3.4 | Les coordonnées des variables supplémentaires | 50 |
| 4 | Quelques matrices de corrélations | 52 |

Introduction

La collecte de données d'enquêtes auprès des ménages réalisée à l'Insee est très largement informatisée depuis le début des années 1990. En face à face (*Capi* pour *Computer assisted personal interviewing*) ou par téléphone (*Cati* pour *Computer assisted telephone interviewing*), les enquêteurs saisissent directement les réponses du ménage sur leur ordinateur et la transmission des données peut ainsi s'effectuer sans délai. Cette manière de collecter l'information a également l'avantage de minimiser les possibilités d'erreur liées à la passation de questionnaire grâce à l'automatisation des filtres et des contrôles.

Pour pousser encore plus loin l'intégration de la collecte dans les processus informatiques de production, le recours à internet pour enquêter auprès des ménages est envisagé. Ce moyen de communication est en effet aujourd'hui très largement ancré dans le quotidien des personnes même si tout le monde n'y a pas encore accès, et son utilisation laisserait entrevoir des avantages en termes de coût financier et temporel (de Peretti et Razafindranovona, 2014).

Mais l'un des problèmes méthodologiques majeurs que posent les enquêtes par internet est celui des effets de mode : les réponses collectées via ce mode auto-administré sont susceptibles de différer de celles résultant d'une passation de questionnaire avec intermédiation d'un enquêteur. Ces différences peuvent s'expliquer en partie par des phénomènes de type *satisficing* (répondant moins impliqué en l'absence d'enquêteur) ou par la désirabilité sociale (réponses plus conformes à la norme sociale en présence d'enquêteur).

La littérature autour des effets de mode est abondante, néanmoins, comme le dit Couper, "il n'y a pas encore de consensus sur l'étendue des différences d'erreurs de mesure entre les différents modes et sur ce (si besoin est) qu'il est nécessaire de faire à ce propos" (Couper, 2011). Par ailleurs, l'une des principales difficultés de l'analyse des effets de mode est qu'il peut s'avérer assez compliqué de distinguer ce qui relève de l'effet de mode strictement lié à la mesure, des effets de sélection, d'autant plus lorsque ces effets relèvent de caractéristiques inobservables (Razafindranovona et al., 2014).

La mesure des effets de mode telle qu'elle est pratiquée habituellement repose sur de l'analyse univariée. Elle compare des estimations de paramètres d'intérêt à partir de données d'enquête provenant de modes de collecte différents. Ainsi, la comparaison va par exemple porter sur les fréquences prises par les différentes modalités de réponse à une question ou sur la distribution d'une variable numérique. Mais comme les données d'enquêtes ne sont pas utilisées à ces seules fins, la question de l'effet de mode dans les associations entre variables devrait également se poser.

Les liaisons entre variables sont en effet susceptibles de changer d'un mode de collecte à un autre. L'analyse géométrique des données peut alors être mobilisée pour fournir une vision synthétique des liaisons entre variables. L'analyse factorielle multiple duale (AFMD) (Lê et Pagès, 2010) permet d'aller plus loin que les analyses factorielles classiques, du type analyse en composantes principales (ACP) ou analyse des correspondances multiples (ACM) en permettant d'analyser simultanément les liaisons entre variables pour chaque mode de collecte et de les comparer entre eux.

Notre analyse porte sur les principales questions d'opinion communes aux enquêtes Cadre de vie et sécurité (CVS) et Vols, violences et sécurité (VVS). Ce type de questions est en effet particulièrement susceptible d'être affecté par des effets de mode. La question est de savoir dans quelle mesure existent des différences selon le mode de collecte

dans les liaisons entre modalités renseignées. La comparaison brute (sans correction sur la sélection des répondants) des réponses selon les différents modes de collecte en utilisant l'AFMD permet ainsi de constater quel genre d'effets de mode globaux existent dans les associations entre variables. Une discussion sur la robustesse des résultats via un contrôle de la sélection sur caractéristiques observables (par *matching*) donnera des éléments de réponse quant à l'existence d'effets de mode de ce type intrinsèquement liés à la mesure.

Ce document présente dans la section 1 la théorie de l'AFMD, puis la section 2 son application à l'analyse de la collecte multimode.

1 L'analyse factorielle multiple duale

1.1 Pour quelles données et quelle problématique ?

Les analyses factorielles classiques, analyse en composantes principales et analyse des correspondances multiples, permettent l'analyse des liaisons entre variables d'un tableau de données individus \times variables, en offrent une représentation graphique et font ressortir les plus saillantes. Parfois ces tableaux comprennent une dimension supplémentaire lorsque les individus observés sont organisés en groupes. C'est le cas par exemple lorsque la même enquête est réalisée à des dates différentes, ou que les individus appartiennent à des pays différents, ou encore, comme c'est le cas ici, lorsque les individus sont enquêtés via des modes de collecte différents. Cette structure en groupe des individus soulève de nouvelles questions. On peut en effet vouloir analyser les liaisons entre variables à l'intérieur de chaque groupe d'individus et étudier comment celles-ci évoluent d'un groupe d'individus à un autre.

Les analyses factorielles classiques ne sont pas adaptées à une telle problématique. Mener autant d'analyses factorielles sur chacun des groupes d'individus ne répond que partiellement à cet objectif. Elles permettent certes d'analyser les corrélations entre variables à l'intérieur de chaque groupe d'individus mais étant indépendantes les unes des autres, elles sont par conséquent difficilement comparables.

L'analyse factorielle multiple duale (AFMD) (Lê et Pagès, 2010) est adaptée à cette problématique qui est celle de l'analyse d'un multi-tableau se présentant sous la forme d'un empilement de sous-tableaux de données, les variables de ces sous-tableaux étant communes (cf. schéma 1). Comme toute analyse factorielle, l'AFMD offre un repère permettant de représenter au mieux les liaisons entre variables au niveau global, c'est-à-dire mesurées sur l'ensemble des individus : le nuage global des variables. Ce repère est également adapté pour décliner l'analyse à chaque niveau partiel, c'est-à-dire au sein de chaque groupe d'individus, offrant autant de nuages des variables partiels que de groupes d'individus. Enfin, il autorise la superposition de ces nuages global et partiels, particulièrement adaptée à la comparaison des liaisons entre variables d'un groupe d'individus à un autre. En outre, l'AFMD équilibre l'influence des groupes d'individus dans l'analyse.

L'AFMD a été élaborée par Lê et Pagès (2010) sur variables quantitatives. Son extension aux variables catégorielles demande quelques ajustements qui sont présentés au paragraphe 1.5.

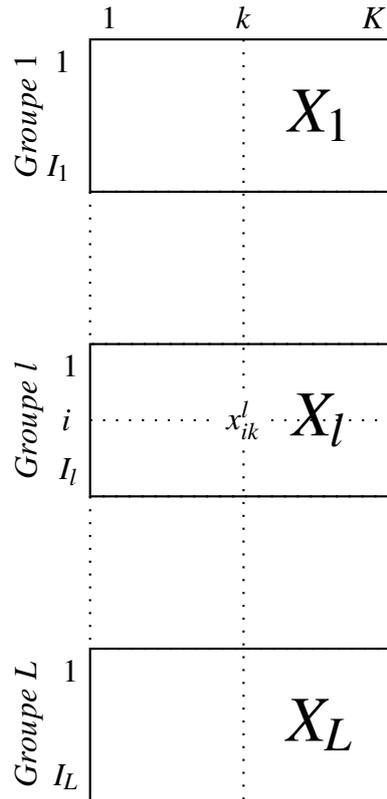


Schéma 1 – Tableau de données

L'AFMD se trouve dans la lignée de l'analyse factorielle multiple (AFM). De même que l'AFMD, l'AFM permet d'analyser les multi-tableaux de données, mais ceux-ci se présentent sous la forme d'une juxtaposition de sous-tableaux de données : les individus sont communs d'un sous-tableau à un autre et les variables sont organisées en groupe (par exemple par thème ou dans le cas de données de panel par date d'observation). L'objectif est d'analyser les différences entre individus selon plusieurs points de vue : globalement (sur l'ensemble des variables), et partiellement (selon un groupe de variables). Il est ainsi possible d'analyser si un individu se distingue des autres lorsqu'il est décrit par l'ensemble des variables, ou s'ils s'en éloignent particulièrement pour un groupe de variables (voir par exemple Gelein et Sautory (2002) et Escofier et Pagès (1998) pour une présentation plus détaillée de cette méthode).

Les notations : N individus (lignes) sont observés et décrits à travers K variables quantitatives (colonnes) notées k . Ces individus se répartissent en L groupes. Le tableau de données correspondant aux N^l individus du groupe l ($l = 1, \dots, L$) est noté X_l . On note k^l la restriction de la variable k au groupe d'individus l . La valeur de la variable k pour l'individu i appartenant au groupe l est notée x_{ik}^l . Ces L tableaux de données peuvent être empilés, les variables observées étant communes d'un groupe d'individus à un autre. Cet empilement forme ainsi un unique tableau de données noté X (cf schéma 1).

1.2 Principe général de l'AFMD

Techniquement, l'AFMD consiste à mener une ACP particulière. Elle se déroule en plusieurs étapes qui sont développées ici.

Pour ne pas alourdir les notations, nous supposons ici que tous les individus ont le même poids.

1.2.1 Etape préliminaire : Centrer et réduire les données à l'intérieur de chaque groupe

La première étape de l'AFMD consiste à centrer et réduire les données à l'intérieur de chaque groupe d'individus. Avec les notations décrites au paragraphe précédent, on travaille ainsi sur les variables z_k telles que :

$$z_{ik}^l = \frac{x_{ik}^l - \bar{x}_k^l}{\sigma_k^l} \quad \text{avec} \quad \bar{x}_k^l = \frac{1}{N^l} \sum_{i \in I^l} x_{ik}^l \quad \text{et} \quad \sigma_k^{l^2} = \frac{1}{N^l} \sum_{i \in I^l} (x_{ik}^l - \bar{x}_k^l)^2$$

On note Z_l le sous-tableau de données du groupe l ainsi transformé. L'empilement de ces sous-tableaux forme le tableau de données Z . On note z_k ses variables et z_k^l leur restriction au groupe d'individus l .

Les variables z du tableau de données Z sont ainsi centrées et réduites au niveau global. Cette transformation ne modifie pas les corrélations entre variables à l'intérieur des groupes. Enfin elle permet d'équilibrer l'influence des groupes d'individus dans la construction des axes de l'analyse. Cet aspect est présenté plus en détail à la section 1.3.1.

1.2.2 AFMD = ACP sur le tableau des données centrées et réduites par groupe

L'AFMD consiste à mener une ACP sur le tableau de données Z , centré et réduit par groupe. L'AFMD est ainsi une analyse factorielle sur un tableau de données particulier. On retrouve donc naturellement les principes généraux d'une analyse factorielle. Nous en rappelons ici les grandes lignes. Pour plus de détails sur les analyses factorielles, le lecteur peut se référer par exemple à Escofier et Pagès (1998). Les sections suivantes insistent plus particulièrement sur ce qui est spécifique à l'AFMD.

Comme toute analyse factorielle, l'AFMD fournit un jeu de K composantes principales notées F_1, \dots, F_K . Il s'agit de nouvelles variables construites de sorte à restituer au mieux l'information contenue dans le tableau de données en un nombre réduit de dimensions. Elles correspondent à des combinaisons linéaires des variables initiales z et sont deux à deux orthogonales. On note $u_s(z)$ le coefficient associé à la variable z dans la composante F_s , de tel sorte que $F_s = \sum_{k=1}^K u_s(k)z_k$. Ces composantes principales définissent un nouveau repère dans lequel sont représentées les données :

– Le nuage des individus

Les individus sont initialement décrits à travers les K variables initiales et sont donc représentés dans un espace de dimension K , difficile à appréhender. On voudrait pouvoir repérer plus aisément les individus proches (qui renseignent des valeurs proches pour les variables) ou au contraire éloignés. Les composantes principales répondent à cette problématique. Elles définissent un nouveau repère pour représenter les individus et sont construites de sorte à permettre de visualiser au mieux le nuage des individus en un nombre restreint de dimensions. Les premiers plans factoriels offrent ainsi une information résumée et plus lisible. La perte d'information est minimisée.

Les individus sont donc décrits à travers ces nouvelles variables. On note $F_s(i)$ la coordonnée de l'individu i sur l'axe s engendré par la $s^{\text{ème}}$ composante principale. On a ainsi $F_s(i) = \sum_{k=1}^K u_s(k)z_{ik}$. Le vecteur $(F_s(1), \dots, F_s(N))'$ de dimension N regroupant les coordonnées des N individus est appelé $s^{\text{ème}}$ facteur principal. $\Delta_s = \text{vect}(u_s(1), \dots, u_s(K))$ est le $s^{\text{ème}}$ axe factoriel.

– Le nuage des variables

Le nuage des variables offre une représentation graphique des corrélations entre les variables. On s'intéresse donc aux colonnes du tableau de données. Sur ce nuage, chaque variable est représentée par un vecteur situé sur la sphère unité, les variables étant réduites. Ce nuage s'interprète en analysant l'angle entre les vecteurs de deux variables, son cosinus correspondant à leur corrélation. L'angle entre deux variables est ainsi d'autant plus petit qu'elles sont corrélées. Une problématique similaire à celle rencontrée pour le nuage des individus se pose : visualiser au mieux les corrélations entre variables. Il s'agit donc de construire un repère sur lequel projeter les données tout en déformant le moins possible les angles entre variables. Les axes de ce repère optimal sont engendrés par les facteurs F_1, \dots, F_K . La coordonnée d'une variable z_k sur l'axe Δ_s correspond à sa corrélation avec le facteur principal F_s .

Nuages des individus et nuages de variables sont construits à partir des mêmes composantes principales et s'interprètent donc conjointement.

1.2.3 Les tableaux \tilde{Z} en tant qu'éléments supplémentaires

Jusqu'à présent il n'a pas été fait référence à la structure en groupes des individus. Les corrélations entre variables analysées sont calculées sur l'ensemble des individus. L'AFMD se donne aussi pour objectifs : d'une part de représenter les corrélations entre variables à l'intérieur de chaque groupe d'individus et d'autre part d'analyser comment elles évoluent d'un groupe d'individus à un autre. Mener des analyses factorielles séparées sur chacun des sous-tableaux de données permettrait de répondre au premier de ces objectifs. Mais ces analyses sont indépendantes et par conséquent difficilement comparables entre elles. Il n'est pas possible notamment de superposer les nuages des variables issus de ces différentes analyses, les composantes principales n'ayant *a priori* pas de raison d'être les mêmes d'une analyse à une autre.

Dans l'AFMD, les composantes principales (construites au niveau global) F_1, \dots, F_K sont utilisées pour représenter les corrélations entre variables à l'intérieur des groupes d'individus : les nuages partiels des variables. On note F_s^l la restriction de la composante F_s au groupe d'individus l . On a donc $F_s^l = \sum_{k=1}^K u_s(k)z_k^l$.

$\tilde{Z}_l =$

| |
|-------|
| 0 |
| z_1 |
| 0 |

Schéma 2 – Tableau de données \tilde{Z}

En pratique, les coordonnées des variables partielles peuvent être obtenues, à un coefficient multiplicatif près, en intégrant dans l'ACP sur le tableau Z , les éléments \tilde{Z} en tant qu'éléments supplémentaires. Le tableau \tilde{Z} correspond au tableau Z_l complété par des zéros (cf schéma 2).

Dans le cadre de l'AFMD, la superposition des nuages global et partiels a un sens : les axes des nuages global et partiels sont issus des mêmes composantes principales. Celles-ci ont la même interprétation d'un groupe d'individus à un autre : la contribution de la variable z_k^l à la construction de la variable F_s^l vaut $u_s(k)^2/\lambda_s$ et est ainsi la même au niveau global et pour chaque groupe d'individus.

Chaque variable est ainsi représentée par $L + 1$ points : L points partiels avec à leur barycentre (pondéré) un point global. La trajectoire de ces points homologues (associés à une même variable mais à des groupes d'individus différents) s'interprète comme des différences de corrélations entre variables au sein des groupes d'individus (cf section 1.4.4).

1.3 Les caractéristiques de l'AFMD

1.3.1 Une analyse qui équilibre l'influence des groupes

Un des objectifs de l'AFMD est d'équilibrer l'influence des différents groupes d'individus dans la construction des axes. Ceci est rendu possible par l'opération préalable sur les données qui consiste à les centrer et les réduire par groupe.

Pour le comprendre, rappelons le calcul sous-jacent à la construction des composantes principales. Les composantes principales sont construites itérativement. À chaque étape, on cherche la variable orthogonale à celles construites précédemment et résumant au

mieux l'information. Pour cela, le critère retenu est l'inertie expliquée par un axe qui correspond à la moyenne des carrés des coordonnées des individus projetés sur cet axe, ou matriciellement à $u' \left[\frac{1}{N} Z' Z \right] u$ avec $u \in \mathbb{R}^K$. On montre que les axes factoriels sont engendrés par les vecteurs propres normés de la matrice d'inertie $\frac{1}{N} Z' Z$, notés u_1, \dots, u_K ordonnés de manière décroissante selon les valeurs propres associées $\lambda_1 > \lambda_2 \dots > \lambda_K$. La valeur propre λ_s correspond à l'inertie projetée des individus sur l'axe Δ_s engendré par le vecteur u_s .

En écrivant le programme de maximisation sur lequel repose la recherche d'un vecteur unitaire u_1 générateur du premier axe factoriel Δ_1 ainsi,

$$u_1 = \operatorname{argmax}_{\substack{u \in \mathbb{R}^K \\ \|u\|=1}} u' \left[\frac{1}{N} Z' Z \right] u = \operatorname{argmax}_{\substack{u \in \mathbb{R}^K \\ \|u\|=1}} \sum_{l=1}^L \frac{N^l}{N} \times \underbrace{u' \left[\frac{1}{N^l} Z'_l Z_l \right] u}_{\substack{\text{Inertie de la projection du} \\ \text{nuage d'individus} \\ \text{du groupe } l \text{ sur } \Delta = \text{Vect } u}} \quad (1)$$

on fait apparaître les axes factoriels de l'AFMD comme le résultat d'un compromis entre les différents groupes d'individus¹. Chaque groupe d'individus a un axe factoriel optimal qui lui est propre (celui qui serait obtenu si on menait une ACP sur les seuls individus de ce groupe). Le premier axe factoriel construit par l'AFMD est un compromis entre ces différents axes. Chaque groupe d'individus contribue de manière équilibrée à la construction, les inerties projetées des différents groupes d'individus ayant toutes le même intervalle de variation (entre zéro et K), les données ayant été préalablement centrées et réduites au sein de chaque groupe.

Équilibrer la contribution des groupes dans la construction des axes est un enjeu important dans une telle analyse. On rappelle que l'AFMD se donne comme objectif l'analyse des corrélations entre variables au niveau global, mais aussi au sein de chaque groupe d'individus. Si les groupes n'intervenaient pas de manière équilibrée pour construire les axes, par exemple si un groupe déterminait à lui seul la construction d'un des premiers axes de l'analyse, le repère construit ne permettrait pas d'offrir une bonne représentation des corrélations entre variables au sein des autres groupes, au sens où il en offrirait une représentation trop déformée. Cet axe ne mettrait en évidence que des corrélations saillantes entre variables spécifiques à un groupe d'individus.

Cette opération qui consiste à centrer et réduire les données par groupe a d'autres propriétés qui répondent aux objectifs de l'AFMD. Elles sont développées plus loin.

1.3.2 Une analyse qui offre plusieurs points de vue

L'AFMD permet de représenter les corrélations entre variables aux niveaux global et partiels à partir des mêmes composantes principales. Ceci justifie la superposition des nuages global et partiels des variables. Chaque variable z est représentée par $L + 1$ points dans le nuage des variables. Le point global se trouve au barycentre (pondéré) des points partiels. Ce nuage offre plusieurs points de vue :

1. Les autres facteurs sont obtenus par le même type de programme de maximisation, mais avec la contrainte supplémentaire d'être orthogonaux aux axes créés précédemment.

- Le point de vue global ne s’intéresse qu’aux points globaux. Il analyse les corrélations entre variables (transformées) calculées sur l’ensemble des individus, sans tenir compte du fait que les individus sont structurés en groupes.
- Les points de vue partiels s’intéressent aux corrélations entre variables observées selon chaque groupe d’individus pris isolément. On analyse le nuage des points associés à un groupe d’individus en particulier.
Les restrictions des composantes principales à un groupe d’individus donné ne sont en général pas optimales pour visualiser les corrélations entre variables à l’intérieur du groupe d’individus considéré : elles diffèrent des composantes qu’on obtiendrait en menant une ACP sur le sous-tableau de données correspondant. Cependant, le fait d’avoir centré et réduit les variables par groupe garantit une certaine contribution de chaque groupe l à la construction des composantes de l’AFMD. Le repère constitue un compromis entre les différents groupes d’individus. Pour certains groupes présentant une structure de corrélation très différente des autres groupes, il peut être mal adapté. Il est possible de le vérifier au cas par cas (cf analyse des groupes à la section 1.4.3).
- Le point de vue “inter-groupe” étudie la trajectoire des points homologues (les points associés à une même variable). Leurs positions relatives renseignent en effet sur les différences de corrélations entre variables d’un groupe d’individus à un autre (voir paragraphe 1.4.4).

Ce point de vue est l’apport essentiel de l’AFMD. Sans l’objectif de comparer les groupes d’individus entre eux, les analyses globale et partielles auraient en effet pu être menées par des ACP indépendantes sur le tableau complet et sur chacun des sous-tableaux. L’AFMD fournit un repère commun à ces différentes analyses et adapté à leur comparaison.

1.4 Lecture et interprétation des résultats

L’AFMD est une analyse factorielle. On retrouve naturellement les mêmes aides à l’interprétation des résultats que dans une analyse factorielle classique. On ne présente ici en détail que ce qui est spécifique à l’AFMD. Le lecteur retrouvera une présentation détaillée des outils classiques d’une analyse factorielle dans Escofier et Pagès (1998) par exemple.

1.4.1 Le point de vue global

Le point de vue global s’intéresse à l’ensemble des individus sans tenir compte du fait qu’ils soient structurés en groupes.

Les valeurs propres

Les valeurs propres correspondent aux inerties du nuage des individus projeté sur les axes factoriels. Comme classiquement dans une analyse factorielle, l’examen de l’histogramme des valeurs propres peut permettre de sélectionner le nombre d’axes à interpréter (critère de l’inertie moyenne ou critère du coude). Les variables étant centrées et réduites, l’inertie totale du nuage des individus correspond au nombre de variables actives incluses dans l’analyse.

Dans le cadre de l'AFMD, l'examen de l'histogramme des valeurs propres peut être complété. Du programme de maximisation sous-jacent à la recherche des axes dans l'AFMD (1), on en déduit que la valeur maximale de la première valeur propre est $\sum_{l=1}^L \frac{N^l}{N} \lambda_1^l$ où λ_1^l correspond à l'inertie maximale du groupe d'individus l projeté sur un axe, soit la première valeur propre qui serait obtenue dans l'ACP normée restreinte à ce groupe. Une première valeur propre proche de $\sum_{l=1}^L \frac{N^l}{N} \lambda_1^l$ correspond à une situation dans laquelle le premier axe issu de l'AFMD restitue pour chaque groupe une inertie proche de l'inertie maximale. Ce premier axe de l'AFMD correspond donc à une direction d'inertie commune aux différents groupes d'individus. Quand la première valeur propre est inférieure à cette valeur, le premier axe factoriel est un compromis entre les groupes d'individus, qui peuvent par conséquent être plus ou moins bien représentés. Des indicateurs présentés au paragraphe 1.4.3 permettent de compléter ce diagnostic.

Le nuage global des variables

Comme dans une analyse factorielle classique, les corrélations entre variables au niveau global (nuage global des variables) sont représentées à partir des composantes principales $F_s = \sum_{k=1}^K u_s(k)k$. La coordonnée d'une variable sur le $s^{\text{ème}}$ axe principal correspond à sa corrélation avec la composante principale. L'axe s'interprète de manière classique en repérant les variables qui contribuent le plus à la construction de l'axe. Sur ce nuage, les variables appartiennent à la sphère unité et sont d'autant plus proches qu'elles sont corrélées (le cosinus de leur angle correspond à leur corrélation).

Le nuage des individus

Le nuage des individus représente les individus avec leurs nouvelles coordonnées. Sur ce nuage, deux individus sont d'autant plus proches qu'ils renseignent des valeurs proches. Les variables étant centrées réduites par groupe, il est difficile d'interpréter la distance entre deux individus de deux groupes différents.

1.4.2 Les nuages partiels des variables

Les nuages partiels des variables représentent les corrélations entre variables à l'intérieur de chaque groupe d'individus. Dans le cadre de l'AFMD, on utilise les composantes principales restreintes au groupe d'individus considéré, soit les variables $F_s^l = \sum_{k=1}^K u_s(k)k^l$. De même que pour le nuage global des variables, la coordonnée partielle d'une variable sur un axe correspond à la corrélation de cette variable avec la composante principale restreinte au groupe.

Les composantes principales restreintes ont la même interprétation d'un groupe à un autre. La contribution d'une variable z^l à la construction de la composante principale restreinte F_s^l vaut $u_s(k)/\sum_k u_s(k)^2$ et est ainsi la même d'un groupe d'individus à un autre, et la même qu'au niveau global.

Les composantes principales F_s forment une base orthogonale pour le produit scalaire considéré (elles sont deux à deux non corrélées). Cela n'est pas nécessairement le cas des restrictions F_s^l . En toute rigueur, les axes du nuage des variables engendrés par les composantes principales restreintes F_s^l ne sont donc pas orthogonaux. Ils sont pourtant

représentés comme tel, ce qui donne une image plus ou moins déformée du nuage partiel des variables. Les graphiques proposés par la fonction `DMFA` du package `FactoMineR` sous R représentent une ellipse qui permet d’apprécier visuellement la distorsion.

1.4.3 Les groupes d’individus

Une troisième représentation des données spécifique à l’AFMD est proposée : le nuage des groupes d’individus. Il s’agit de repérer les groupes d’individus présentant des corrélations entre variables similaires. Ce nuage repose sur la métrique associée au produit scalaire de Hilbert-Schmidt (pour plus de détails voir annexe 1). Chaque groupe d’individus y est représenté à travers sa matrice d’inertie $C_l = \frac{1}{N_l} Z_l' Z_l$. Sur ce nuage, deux groupes d’individus sont ainsi d’autant plus proches que les corrélations entre variables sont semblables d’un groupe à un autre, et en particulier pour les variables les plus corrélées entre elles au niveau global (voir annexe page 44).

On représente les groupes d’individus dans le repère formé par les matrices $U_s = u_s u_s'$ avec $u_s = (u_s(1), \dots, u_s(K))'$ associé aux composantes principales construits dans le cadre de l’AFMD. On montre que ces matrices forment une base orthonormée pour le produit scalaire de Hilbert-Schmidt, optimale pour la représentation des groupes d’individus.

La coordonnée du groupe d’individus l sur l’axe s engendré par la matrice U_s correspond à l’inertie projetée du nuage des individus de ce groupe sur cet axe (on se place alors dans l’espace des individus). Elle est ainsi comprise entre 0 et λ_1^l l’inertie maximale de la projection du nuage des individus du groupe l sur un axe (ou encore l’inertie associée au premier axe factoriel issu de l’ACP sur ce groupe d’individus). Elle mesure l’importance de l’axe pour le groupe d’individus, au sens où il correspond à une direction importante d’inertie. Cette coordonnée peut être normalisée, en la divisant par λ_1^l .

L’interprétation de la coordonnée d’un groupe d’individus sur un axe en terme d’importance de l’axe pour le groupe d’individus considéré apparaît également quand on l’écrit sous cette forme :

$$\frac{1}{\lambda_s} \sum_{h,k} r(F_s, k) r(F_s, h) r(k^l, h^l)$$

Elle est ainsi d’autant plus grande que pour chaque couple de variables h et k :

1. La corrélation entre ces deux variables restreintes au groupe l est grande.
2. Leur corrélation calculée sur l’ensemble des individus est mise en avant par l’axe s .

La coordonnée du groupe l sur le $s^{\text{ème}}$ axe est ainsi d’autant plus grande que les corrélations entre variables mises en évidence par l’axe sont des corrélations importantes au sein du groupe d’individus l .

La distance entre les coordonnées de deux groupes d’individus différents sur un axe,

$$\langle C_l, U_s \rangle - \langle C_{l'}, U_s \rangle = \frac{1}{\lambda_s} \sum_{h,k} r(F_s, k) r(F_s, h) \left[r(k^l, h^l) - r(k^{l'}, h^{l'}) \right]$$

est d’autant plus grande que leur structure de corrélation diffèrent, en particulier pour les variables qui contribuent le plus à la construction de l’axe.

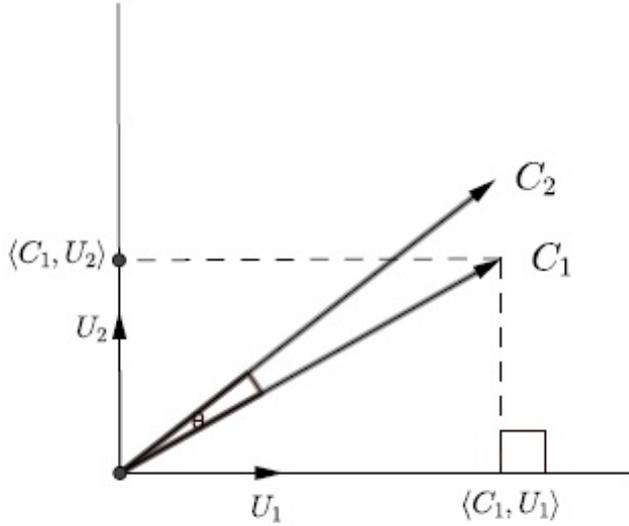


Schéma 3 – Nuage des groupes d'individus

1.4.4 L'étude de l'inter-groupe

L'analyse des différences de coordonnées des groupes d'individus permet de comparer globalement les groupes d'individus entre eux. Pour en faire une analyse plus fine, les nuages global et partiels peuvent être superposés.

Pour rappel, cette superposition a un sens dans le cadre de l'AFMD : les nuages partiels des variables sont construits à partir des mêmes composantes principales et ont la même interprétation d'un groupe à un autre. D'autre part, les coordonnées partielles se déduisent des coordonnées globales.

Chaque variable est ainsi représentée par $L + 1$ points : un point global et L points partiels. La coordonnée du point global sur un axe se trouve au barycentre (pondéré) des coordonnées des points partiels. Il faut noter que le nombre de points représentés sur ce nuage peut être rapidement élevé même avec un nombre raisonnable de groupes d'individus et de variables analysées.

La trajectoire des points homologues, c'est-à-dire les points partiels associés à la même variable mais à des groupes d'individus différents, offre une analyse des différences de corrélations entre variables d'un groupe d'individus à un autre. On note $G_s(k)$ (respectivement $G_s^l(k)$) la coordonnée de la variable z_k (respectivement z_k^l) sur l'axe de rang s . En exprimant la coordonnée $G_s^l(k)$ sous la forme :

$$\begin{aligned} G_s^l(k) &= \frac{1}{\sqrt{\lambda_s}} \frac{1}{\sqrt{\text{Var}(F_s^l)}} \sum_{h=1}^K r(z_k^l, z_h^l) G_s(z_h) \\ &= \frac{1}{\sqrt{\lambda_s}} \frac{1}{\sqrt{\text{Var}(F_s^l)}} \sum_{h=1}^K r(z_k^l, z_h^l) r(z_h, F_s) \end{aligned}$$

on remarque que la coordonnée $G_s^l(k)$ se déduit des coordonnées de toutes les variables globales sur l'axe s par combinaison linéaire. Une variable intervient d'autant plus, qu'à

l'intérieur du groupe d'individus l elle est corrélée avec la variable z_k^l . La proximité entre variables homologues traduit ainsi une ressemblance des deux groupes quant aux corrélations associées à la variable z_k .

1.4.5 Variables supplémentaires

Comme dans une analyse factorielle classique, il est possible d'inclure dans l'analyse des variables illustratives (encore appelées variables supplémentaires). De même qu'une variable active, une variable illustrative est représentée par $L + 1$ points. La trajectoire de ces points homologues permet de repérer des différences de corrélations d'un groupe d'individus à un autre.

1.5 L'extension aux variables catégorielles

L'AFMD a été initialement présentée pour l'analyse de variables continues. Son extension aux variables catégorielles a été proposée par Abascal et al. (2013). Ils proposent de mener l'AFMD sur les données du tableau disjonctif complet centré et réduit par groupe. Cette méthode s'éloigne cependant beaucoup de l'ACM où les modalités les plus rares ont un poids plus important. Autrement dit, la distance entre deux individus consiste à compter le nombre de modalités non concordantes et à attribuer plus de poids quand l'un renseigne une modalité rare et pas l'autre.

Une manière de ne pas s'éloigner trop d'une ACM consiste à appliquer cette pondération aux variables du tableau disjonctif complètes et réduites par groupe. Le poids d'une modalité correspond à la part des individus qui ne possèdent pas cette modalité.

L'AFMD sur variables catégorielles est équivalente à mener une ACP sur ce tableau de données. Les logiciels statistiques n'offrent pas toujours la possibilité de pondérer les variables. On montre qu'il est équivalent de mener une ACP (non normée) sur les indicatrices du tableau disjonctif centrées et réduites par groupe et multipliées par la racine carrée de leur poids, soit :

$$z_{ik}^l = \sqrt{\frac{N - N_k}{N}} \frac{y_{ik}^l - \frac{N_k^l}{N^l}}{\sqrt{\frac{N_k^l}{N^l} \left(1 - \frac{N_k^l}{N^l}\right)}} \quad i \in l$$

y_{ik}^l est l'indicatrice qui vaut 1 si l'individu i appartenant au groupe d'individus l possède la modalité k et vaut 0 sinon. N est le nombre total d'individus et N_k le nombre de ceux qui possèdent la modalité k . N^l le nombre total d'individus du groupe l et N_k^l le nombre d'entre eux possédant la modalité k .

Les sorties s'interprètent ensuite de la même façon que dans le cas continu développé dans les paragraphes précédents.

1.6 En pratique

Le package `Factominer` sous R propose la fonction `DMFA` permettant de mettre en œuvre une AFMD. Le site <http://factominer.free.fr/> offre une présentation de ce

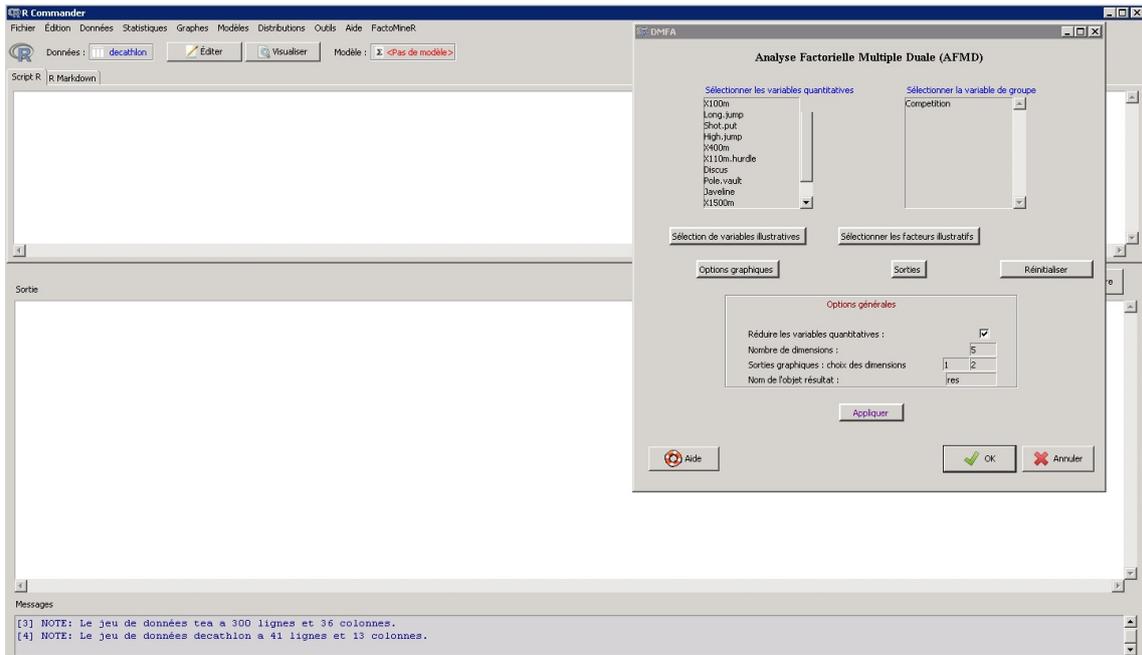


Schéma 4 – Interface de l’AFMD sous FactoMineR

package.

Exemple de programme :

```
library(FactoMineR)
source("U:/R/FactoMineR/install-facto-fr.r")
```

Cette commande permet d’accéder à une interface (voir capture d’écran ci-dessous). La fonction AFMD se trouve dans le menu “FactoMineR -> Analyse Factorielle Multiple Duale (AFMD)”.

Il est aussi possible d’appeler directement la fonction DMFA dans la console de R. Exemple de programme :

```
res.DMFA=DMFA(don, num.fact = ncol(data), scale.unit = TRUE, ncp = 5,
  quanti.sup = NULL, quali.sup = NULL, graph = TRUE, axes=c(1,2))
```

don est la table dont les colonnes sont :

- Le groupe d’appartenance des individus, dont le numéro de la colonne doit être spécifié dans le paramètre `num.fact=` (ici la dernière colonne de la table).
- Les variables supplémentaires qualitatives et quantitatives, dont les numéros de colonne doivent être spécifiés respectivement dans les paramètres `quali.sup` et `quanti.sup`.
- Les variables actives correspondent aux colonnes restantes.

Les autres paramètres définissent les sorties demandées : `ncp` le nombre d’axes décrits dans les sorties, `graph=TRUE` pour obtenir les nuages global, partiels et des groupes d’individus, `axes` spécifie les deux axes retenus pour la représentation graphique.

Descriptif (non exhaustif) des sorties :

| Les valeurs propres | |
|--|---|
| res.DMFA\$eig | la première colonne du tableau contient les valeurs propres de la matrice d'inertie $\frac{1}{n}Z'Z$ présentées par ordre décroissant (qui correspondent aussi aux inerties projetées). La deuxième colonne correspond à l'inertie projetée sur l'axe, en pourcentage de l'inertie totale (égale à K , le nombre de variables). La troisième colonne correspond à ces pourcentages cumulés. |
| Le nuage global des variables | |
| res.DMFA\$var\$coord res.DMFA\$var\$cor | les coordonnées des variables sur les différents axes. Corrélations des variables (globales) avec les composantes principales. Dans une ACP normée, ces corrélations sont égales aux coordonnées des variables sur les axes. |
| res.DMFA\$var\$cos2 | correspond au carré de l'angle entre le vecteur et l'axe. |
| res.DMFA\$var\$contrib | correspond à coordonnée de la variable au carré. est la contribution de la variable à la construction de l'axe en pourcentage. Les variables étant centrées et réduites, elle correspond à $100 \times G_s(k)^2 / \sum_k G_s(k)^2 = 100 \times G_s(k)^2 / \lambda_s$. |
| Les nuages partiels des variables | |
| res.DMFA\$var.partiel | les coordonnées partielles des variables pour chaque groupe d'individus. |
| Le nuage des individus | |
| res.DMFA\$ind\$coord | Les coordonnées des individus sur les axes $F_s(i)$. |
| res.DMFA\$ind\$cos2 | Le carré du cosinus de l'angle entre le vecteur de l'individu i et l'axe. |
| res.DMFA\$ind\$contrib | La contribution de l'individu i à la construction de l'axe, soit $100 \times F_s(i)^2 / \sum_i F_s(i)^2$ |
| res.DMFA\$ind\$dist | Norme du vecteur $\sqrt{\sum_s F_s(i)^2}$ |
| Le nuage des groupes d'individus | |
| res.DMFA\$group\$coord | Coordonnées des groupes d'individus |
| res.DMFA\$group\$coord.n | Coordonnées normalisées des groupes d'individus |
| res.DMFA\$group\$cos2 | Cosinus carré de l'angle entre le groupe et l'axe |

2 Application de l'AFMD à la collecte multimode

2.1 Présentation des données

L'analyse effectuée ici s'appuie sur les données de deux enquêtes dites de victimation dont la collecte a été réalisée en 2013 : l'enquête Cadre de vie et sécurité (CVS) qui interroge des personnes en face à face et l'enquête expérimentale auto-administrée Vols, violences et sécurité (VVS). Par construction, comme l'enquête expérimentale VVS a été conçue à des fins de comparaison avec CVS, il existe ainsi un certain nombre de variables communes (même libellé de question, mêmes modalités de réponse) aux deux enquêtes.

2.1.1 L'enquête Cadre de vie et sécurité

Les enquêtes de victimation sont apparues aux États-Unis dans les années 1960 et offrent une approche alternative de la mesure de la délinquance à celle fondée sur les seuls comptages administratifs (Zauberman et Robert, 2011). En effet, les données administratives présentent certaines lacunes bien connues comme leur sensibilité aux pratiques d'enregistrement ainsi qu'aux modifications d'ordre législatif. Les évolutions mesurées à partir d'enquêtes de victimation sont *a priori* moins sensibles aux modifications du cadre législatif ou administratif que celles résultant de l'exploitation des remontées provenant des services de police ou de gendarmerie.

En France, les enquêtes de victimation sont utilisées depuis environ 40 ans et depuis 1996, l'Insee est un acteur majeur dans la mesure des victimations à partir de ce type d'enquêtes. Cette mesure était auparavant réalisée via le module de victimation des enquêtes EPCV (Enquête permanente sur les conditions de vie des ménages) remplacé aujourd'hui par une enquête à part entière, appelée Cadre de vie et sécurité. L'enquête annuelle CVS est une enquête obligatoire en population générale, qui vise à connaître les faits de délinquance dont les ménages et leurs membres ont pu être victimes dans les deux années précédant l'enquête. Elle permet donc de mesurer la prévalence et l'incidence de certaines atteintes aux personnes et aux biens. Elle porte également sur l'opinion des personnes concernant leur cadre de vie et la sécurité.

La passation du questionnaire est réalisée pour sa plus grande partie en face à face, par un enquêteur équipé d'un ordinateur (Capi). Pour les questions les plus sensibles (violences sexuelles et violences au sein du ménage), il est demandé au répondant de saisir lui-même ses réponses sur l'ordinateur au moyen d'un protocole sous casque (Casi pour *Computer Assisted Self Interviewing*). Après verrouillage du questionnaire par l'enquêteur, les réponses à ces questions sensibles ne sont plus accessibles à l'enquêteur ce qui garantit une complète confidentialité sur cette partie du questionnaire.

2.1.2 L'enquête expérimentale Vols, violence et sécurité

L'enquête expérimentale multimode sur le thème de la victimation s'intitule Vols, violences et sécurité. Cette enquête méthodologique auto-administrée a pour objectif principal de comparer les résultats obtenus avec ceux de l'enquête Cadre de vie et sécurité.

Pour l'enquête VVS, deux modes de collecte coexistent : l'enquêté peut répondre sur internet ou en renvoyant le questionnaire papier par courrier. Le protocole de cette enquête est le protocole "standard" utilisé pour les expérimentations multimodes de l'Insee :

- envoi d'une lettre-avis avec les données de connexion ;
- 1ère relance (3 semaines après) avec envoi d'une lettre contenant les données de connexion, un questionnaire papier ainsi qu'une enveloppe pré-affranchie pour la réponse ;
- 2ème relance avec les données de connexion.

La possibilité de répondre sur papier est offerte afin de ne pas exclure d'emblée de l'enquête une partie de la population, d'autant plus quand la réponse est obligatoire, comme dans l'enquête principale en face à face. En revanche, cette possibilité n'est offerte que dans un deuxième temps afin de favoriser la réponse par internet sans que cela n'ait d'incidence négative sur le taux de réponse global.

L'enquête expérimentale VVS reprend une partie des thématiques de CVS en s'appuyant pour l'essentiel sur les mêmes questions. Que ce soit dans le questionnaire web ou sa version papier, il n'y a pas de questionnement sur les violences au sein du ménage. On considère, en effet, que ces protocoles n'offrent pas des garanties suffisantes de confidentialité et de sécurité au répondant.

2.1.3 Quelques précisions sur le taux de réponse aux enquêtes

Pour des raisons qui tiennent principalement aux différences dans le protocole de contact et à la motivation apportée par l'enquêteur, les taux de réponse des enquêtes en face à face sont généralement très supérieurs à ceux des enquêtes auto-administrées. La comparaison entre les deux enquêtes de victimation ici analysées est une très bonne illustration de cette tendance : alors que le taux de réponse est d'environ 70 % pour CVS (environ 14 600 questionnaires exploitables), il est plus de deux fois moindre pour VVS puisqu'il est de 32 % (environ 12 900 questionnaires exploitables) avec une répartition plutôt équilibrée pour cette dernière entre réponse sur internet (47 %) et par papier (53 %).

2.1.4 Les variables d'intérêt

L'analyse va porter ici principalement sur les réponses données à des questions d'opinion sur le cadre de vie, la sécurité ou les institutions (voir annexe 2 pour le libellé exact des questions ainsi que des modalités de réponse) qui interviennent dans CVS comme dans VVS :

- opinion générale sur l'action de la police ou de la gendarmerie ;
- opinion générale sur l'action de la justice ;
- sentiment d'insécurité au domicile ;
- sentiment d'insécurité dans le quartier (ou village) ;
- avis sur la présence de la police ou de la gendarmerie dans le quartier (ou village) ;
- avis sur l'efficacité de la police ou de la gendarmerie dans le quartier (ou village).

Pour mener à bien notre analyse nous disposons également d'autres variables communes aux deux enquêtes que sont des caractéristiques socio-démographiques, une question d'opinion sur les principaux problèmes dans la société française et les variables de

victimation (par exemple : avoir été victime d'un vol sans violence au cours des deux dernières années) qui sont les variables clés de ces enquêtes.

2.2 L'AFMD sur les données des enquêtes CVS et VVS

La mise en oeuvre pratique a été réalisée sous R avec la fonction DMFA du *package* FactoMineR. La table de données sur laquelle l'analyse a été effectuée est constituée d'environ 27 500 observations : 14 600 proviennent de CVS (interrogation en face à face) et 12 900 de VVS (6 100 réponses sur internet et 6 800 réponses par papier).

2.2.1 Le nuage global des modalités

Nous nous intéressons ici aux questions d'opinion posées dans les enquêtes CVS et VVS listées au paragraphe 2.1.4. La représentation graphique du nuage global des modalités sur le premier plan factoriel est fournie ci-après (graphique 5). La coordonnée d'une modalité correspond à sa corrélation avec le facteur. Dans cette première analyse, les données sont groupées selon les trois modes de collecte différents : face à face (CVS), internet et papier (VVS).

Rappelons que comme dans une analyse de correspondances multiples classique, deux modalités sont d'autant plus proches qu'elles sont prises ou non prises simultanément par un grand nombre de répondants. L'interprétation des axes s'effectue en repérant les variables les plus corrélées, à savoir celles qui sont les plus excentrées sur les axes. Il s'avère ainsi que l'on peut interpréter le premier axe comme un axe d'insatisfaction alors que le deuxième est plutôt un axe de la non réponse.

Le premier axe du nuage global oppose deux profils de répondants : les répondants exprimant une certaine insatisfaction et une attente en terme de sécurité se situent à la droite de l'axe, ceux qui, en revanche, ne trouvent guère à se plaindre sur cette question se situent plutôt à gauche.

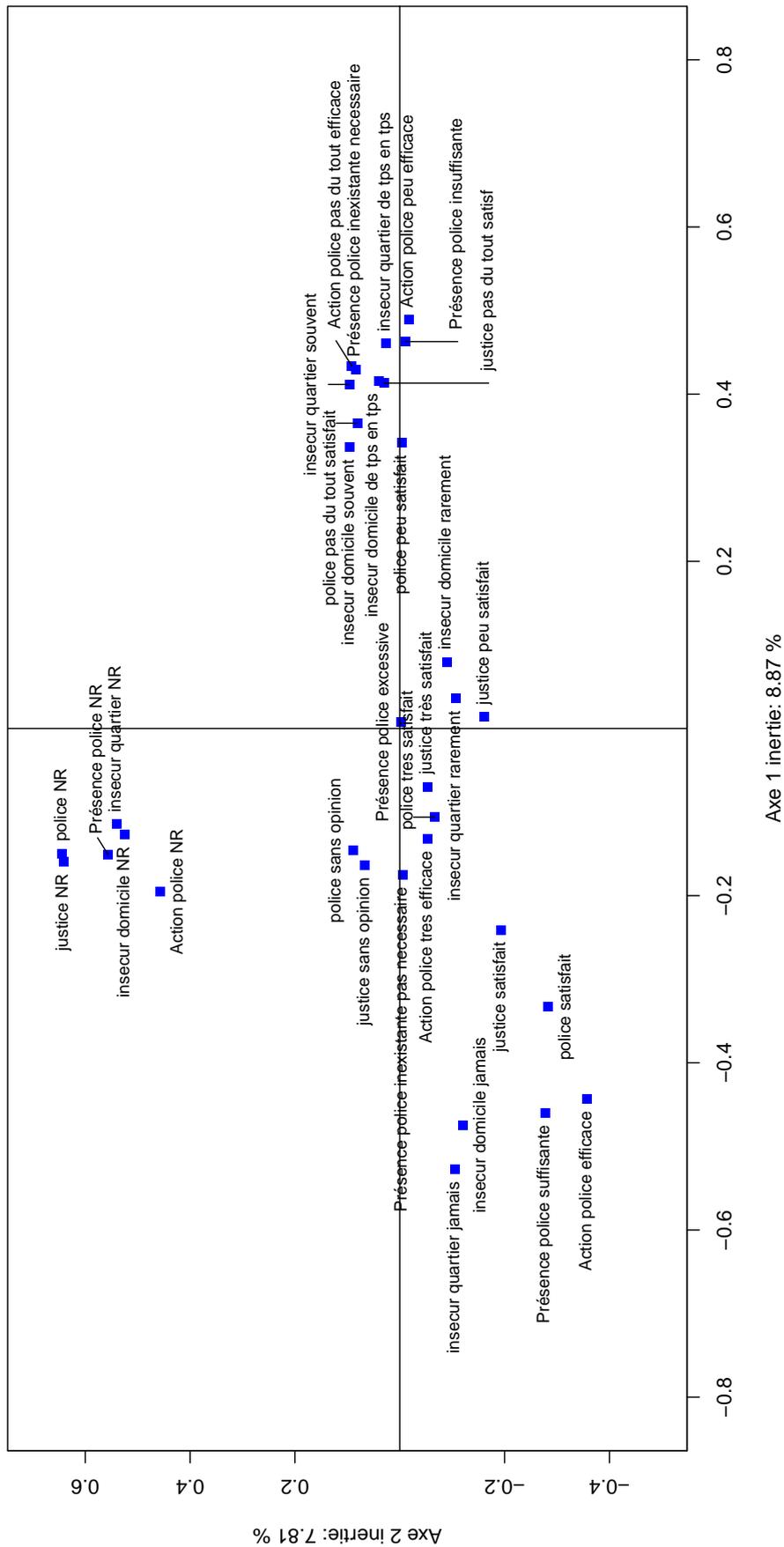
On retrouve ainsi à droite des modalités d'insatisfaction générale vis-à-vis des institutions que sont la police, la gendarmerie ou la justice. On retrouve également à droite de l'axe des modalités exprimant une critique quant à l'efficacité ou la présence (insuffisante) des forces de l'ordre au quotidien dans le quartier ou le village. Et cette insatisfaction va de pair avec un sentiment d'insécurité à domicile ou dans le quartier puisque les modalités relatives à ce sentiment se situent également à droite de l'axe.

En revanche, à gauche de ce premier axe, on retrouve le profil de personnes qui déclarent être simultanément satisfaites des actions de la police, de la gendarmerie et de la justice dans la société française comme au quotidien, dans leur quartier. Ce satisfecit va de pair avec le fait de se sentir en sécurité à domicile ou dans le quartier.

Le deuxième axe correspond quant à lui à la non réponse. Les modalités qui se trouvent du côté positif de cet axe sont ainsi les modalités de non réponse aux différentes questions traitées ici : sentiment général sur l'action des institutions, avis sur l'efficacité ou la présence des forces de l'ordre dans le quartier et sentiment d'insécurité. On retrouve également de ce côté de l'axe des modalités proposées uniquement dans les questions générales sur les institutions où le répondant peut exprimer qu'il n'a pas d'opinion sur le sujet.

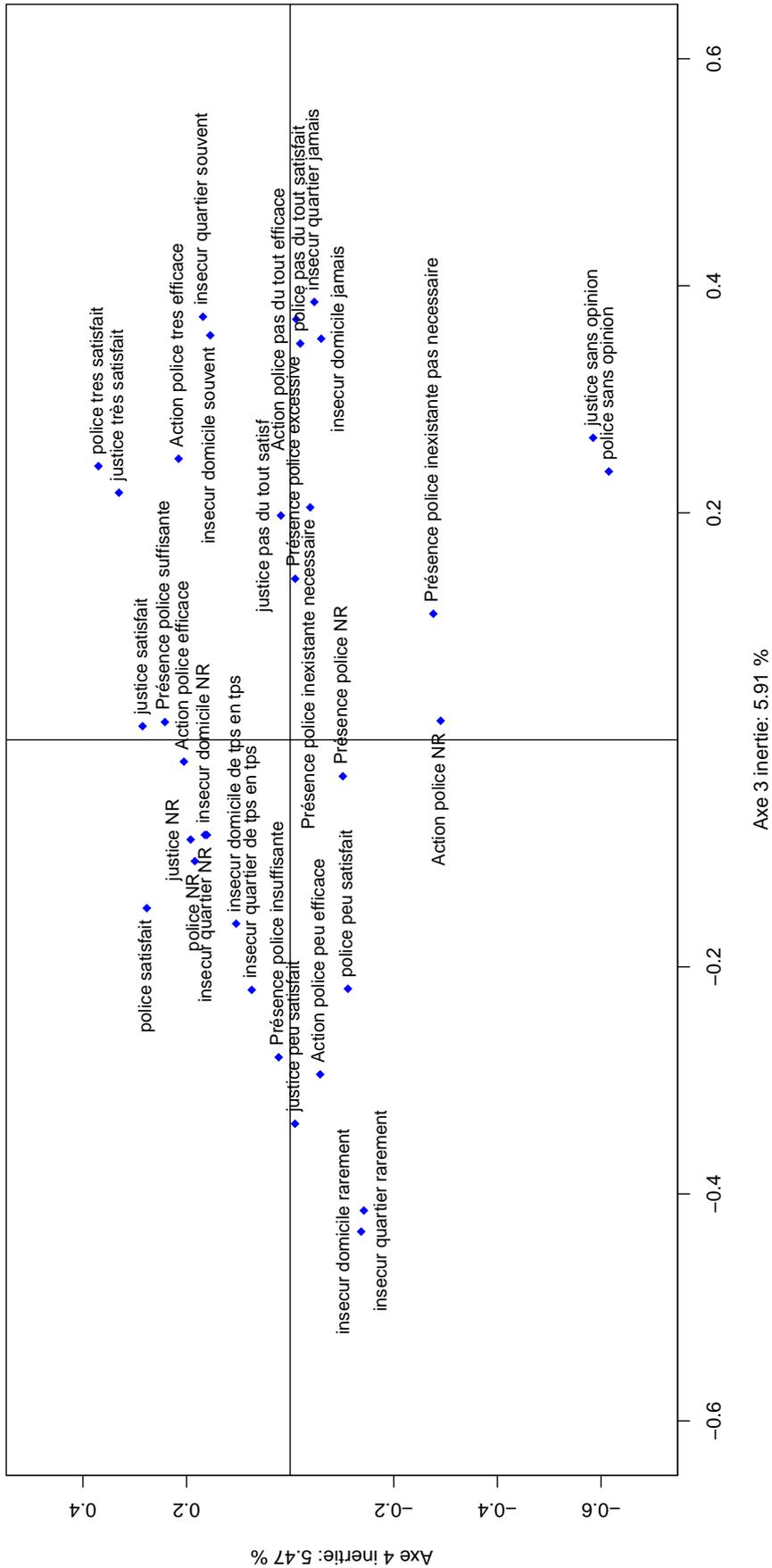
L'essentiel de l'analyse effectuée ici portera sur le plan factoriel formé de ces deux axes qui représentent environ 16 % de l'inertie totale (tableau 3). L'interprétation des axes suivants, moins informatifs par définition, se révèle en effet moins évidente. On peut

néanmoins dire un mot sur le troisième axe : les modalités qui se projettent nettement du côté positif de l'axe sont les modalités associées à des opinions très tranchées, à l'inverse des modalités qui se projettent à gauche associées à des opinions plus modérées (graphique 6).



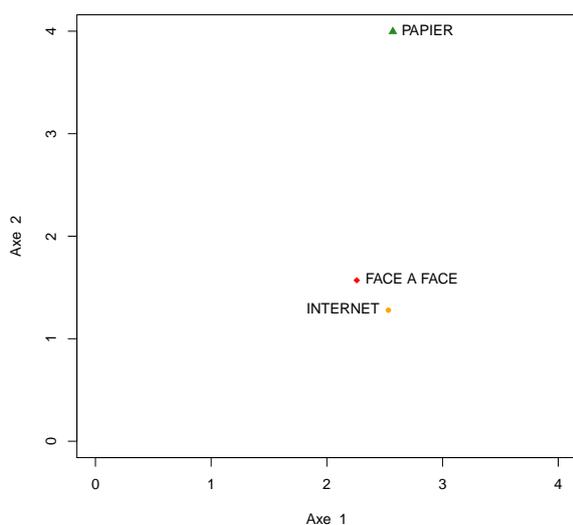
Graphique 5 – 1^{re} AFMMD : Nuage global des modalités, premier plan factoriel

NR = Non réponse.

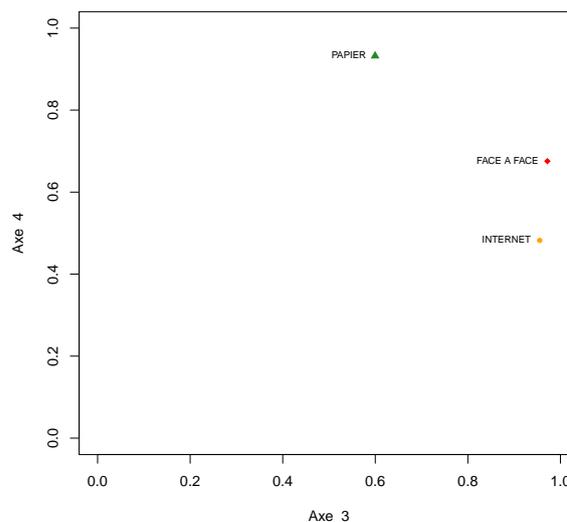


Graphique 6 – 1^{re} AFMMD : Nuage global des modalités, plan factoriel des axes 3 et 4

NR = Non réponse.



Graphique 7 – Nuage des groupes d’individus



Graphique 8 – Nuage des groupes d’individus, coordonnées normalisées

2.2.2 Les nuages partiels : corrélations intra-groupe

Les nuages partiels des modalités représentent les liaisons entre modalités pour chaque sous-groupe de répondants et la coordonnée d’une modalité partielle sur un axe correspond à sa corrélation avec la composante principale restreinte au groupe d’individus. Les composantes principales construites dans le cadre de l’AFMD ont la même interprétation d’un groupe d’individus à un autre.

Comme expliqué à la section 1.3.1, ce repère commun est un compromis entre les trois groupes d’individus partitionnés selon le mode de collecte. Les coordonnées des groupes (voir section 1.4.3, graphiques 7 et 8 et tableau 4 en annexe) correspondent à l’inertie projetée du groupe d’individus sur l’axe et offrent donc pour chaque groupe une mesure de leur qualité de représentation. Sur le premier axe factoriel, pour les groupes d’individus interrogés en face à face et sur internet, les coordonnées sont proches de l’inertie maximale de la projection du groupe d’individus sur un axe (soit la valeur de la première valeur propre, notée λ_1^l , dans l’analyse factorielle sur ce seul groupe d’individus). Pour ces deux groupes d’individus, le premier axe construit dans le cadre de l’AFMD diffère donc peu de celui qu’on obtiendrait en menant une analyse factorielle sur ce seul groupe d’individus. Pour le groupe des individus interrogés par questionnaire papier, l’inertie sur le premier axe ne représente que 60 % de l’inertie maximale. Ce groupe est mieux représenté sur le deuxième axe factoriel (93% de l’inertie maximale projetée sur un axe). L’axe de la non-réponse ressortirait en premier dans une analyse factorielle menée sur ce seul sous-tableau.

Les nuages partiels des variables (graphiques 21, 22 et 23 en annexe page 47) font apparaître des structures de corrélation assez proches d’un groupe d’individus à un autre. L’axe de la non réponse pour les individus interrogés par internet doit cependant être interprété avec précaution en raison du très faible nombre de non réponses dans ce mode de collecte.

2.2.3 Les nuages partiels : étude de l'inter-groupe

A ce stade de l'analyse, des groupes de modalités prises simultanément par un grand nombre d'individus (aux niveaux global et partiels) ont été mis en évidence. Si l'on s'en tient là, l'AFMD n'apporte aucun résultat supplémentaire par rapport à des analyses factorielles indépendantes (une analyse factorielle sur l'ensemble du tableau de données et une sur chacun des sous-tableaux)².

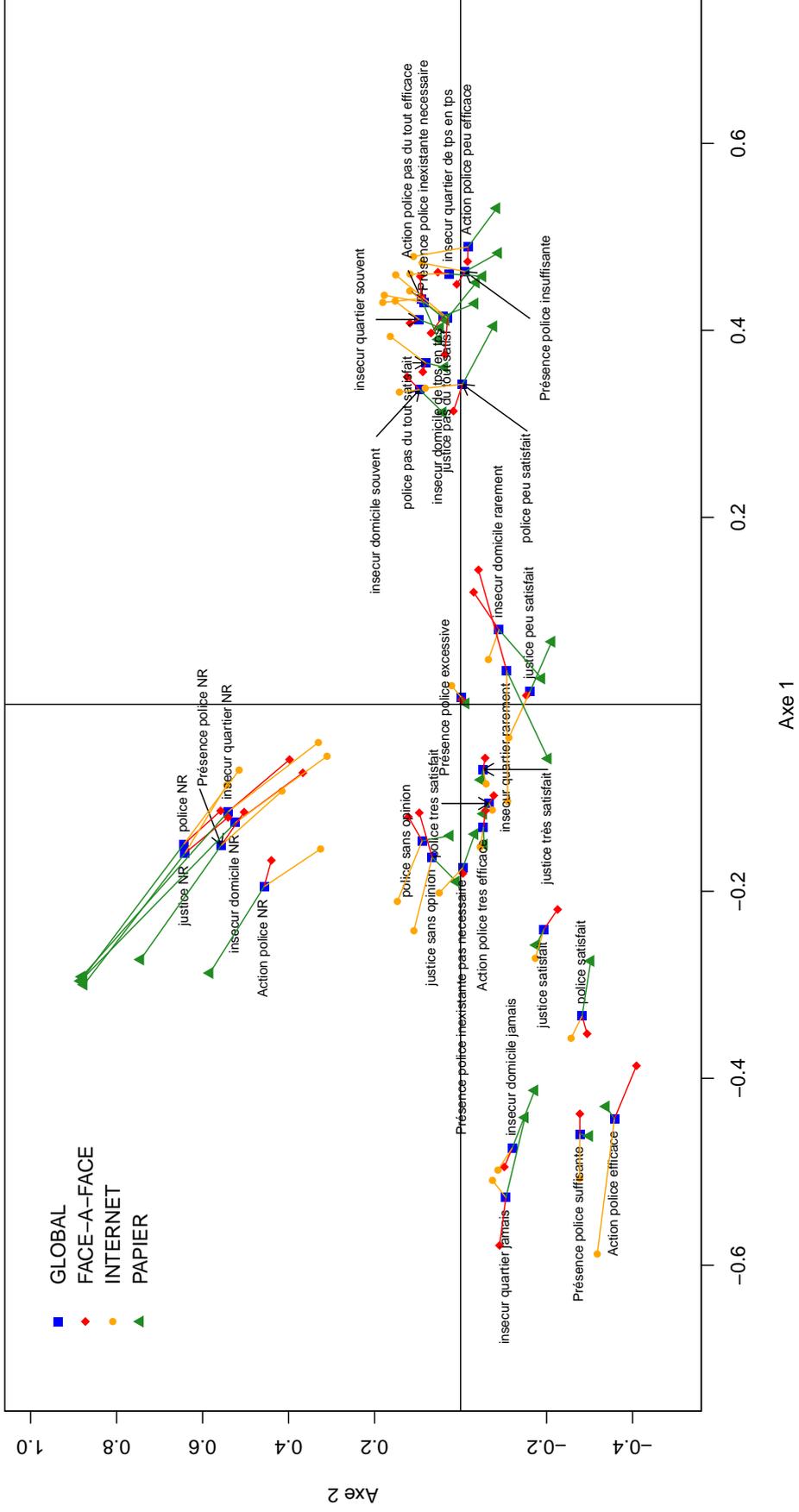
L'intérêt de l'AFMD réside dans l'analyse de l'inter-groupe, c'est-à-dire dans la comparaison des liaisons entre variables d'un groupe d'individus à un autre. L'AFMD fournit ainsi un repère particulièrement adapté à cette analyse. Les nuages global et partiels peuvent en effet être superposés. Chaque modalité est alors représentée par quatre points : trois points partiels (un pour chaque mode de collecte) et un point global.

Dans ce repère, la distance entre les points homologues (les points associés à une même modalité mais à des modes de collecte différents) a une interprétation. Ils sont d'autant plus éloignés que les liaisons de cette modalité avec les autres modalités sont différentes d'un mode de collecte à un autre. Sur un axe donné, les coordonnées de deux points homologues sont d'autant plus éloignées que la corrélation de la modalité concernée avec les autres modalités et en particulier celles les plus liées à l'axe considéré sont différentes (voir section 1.4.4). Les graphiques 9 et 13 représentent ainsi la superposition des nuages global et partiels des modalités pour les quatre premiers axes.

Les distances entre les groupes d'individus sont aussi une mesure globale des différences de corrélations entre variables d'un groupe d'individus à un autre. Le nuage des groupes d'individus (graphique 7) fait apparaître des différences importantes entre le groupe des individus interrogés par questionnaire papier d'une part et ceux interrogés par internet ou en face à face d'autre part. Ils se distinguent particulièrement sur le deuxième axe qui est celui de la non réponse. Cela traduit le fait que le groupe des individus répondant sur questionnaire papier se distinguent des autres groupes par les corrélations entre les modalités de non réponse aux questions.

Les différences entre les coordonnées des groupes d'individus sur le premier axe sont moins importantes. On note néanmoins que la coordonnée du groupe interrogé par internet est la plus élevée, suivie de celle du groupe interrogé par questionnaire papier. Les corrélations entre modalités les plus liées au premier axe sont donc plus élevées pour le sous-groupe "internet" et "papier" par rapport au groupe "face à face". Cet axe représentant un axe d'insatisfaction, on en déduit que les corrélations entre modalités d'insatisfaction d'une part et de satisfaction d'autre part sont plus importantes en moyenne parmi les individus interrogés par questionnaire papier et par internet. Les opinions exprimées par papier sont plus tranchées que celles exprimées par internet, elle-mêmes plus tranchées que celle exprimées en face à face.

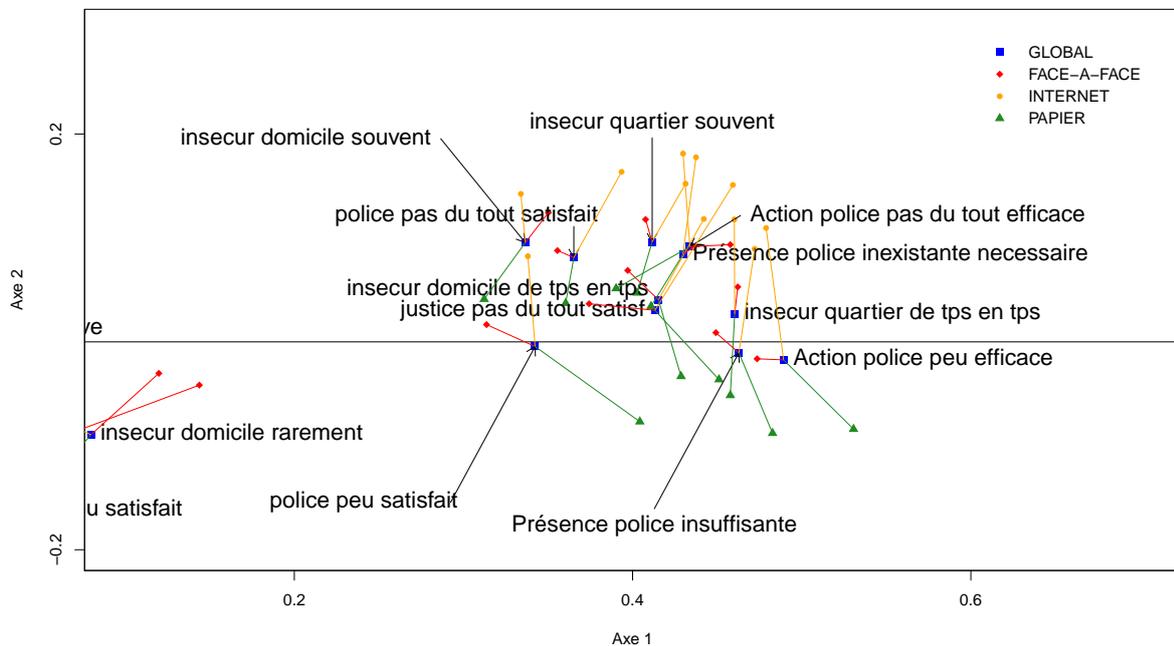
2. Il est même préférable dans ce cas de mener des analyses séparées : le repère construit dans le cadre de l'AFMD est un compromis entre les différents groupes d'individus et en général différent des repères optimaux de chaque groupe.



Graphique 9 – 1^{re} AFMD : superposition des nuages global et partiels des modalités (axes 1 et 2)

Modalités à droite du premier axe (insatisfaction)

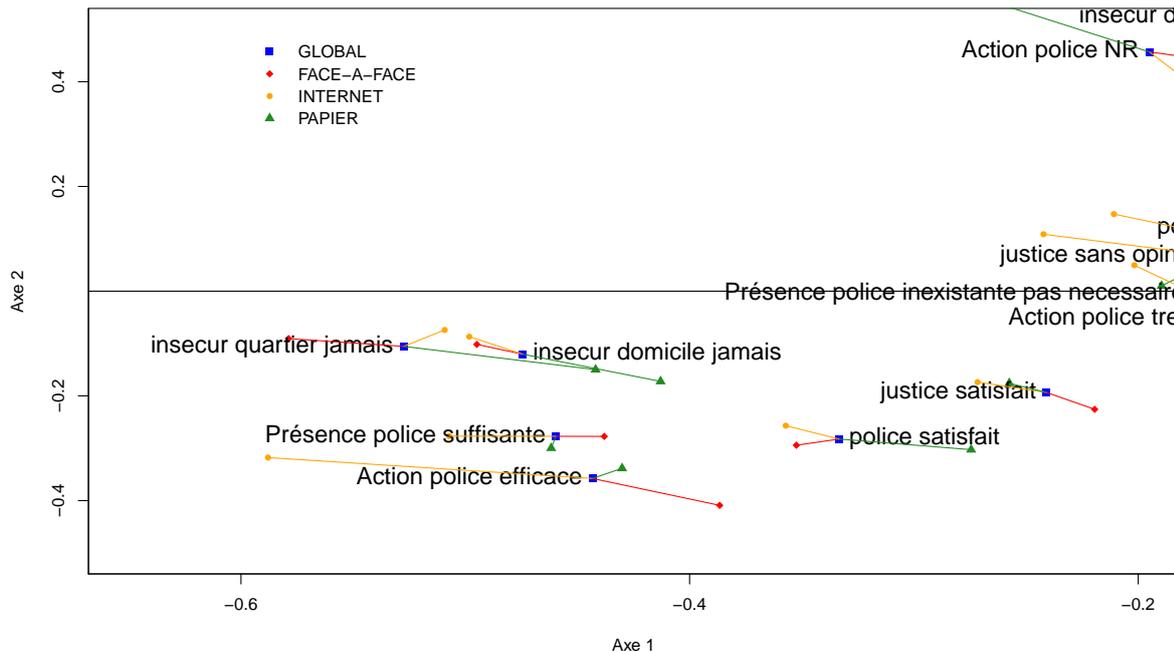
Le graphique 10 est un zoom du graphique 9 à droite du premier axe. Les points partiels homologues sont reliés au point global, permettant de visualiser la trajectoire quand on passe d'un mode de collecte à un autre. À titre d'exemple, on remarque que la coordonnée de la modalité "Pas du tout satisfait de l'action de la justice et des tribunaux dans le traitement de la délinquance" (abrégé *justice pas du tout satisf* sur les graphiques) est plus élevée pour le point partiel associé à la collecte par internet par rapport à celle en face à face. Cela traduit qu'en moyenne cette modalité est plus corrélée avec les modalités les plus liées à l'axe. Pour la plupart des modalités représentées sur le graphique 10, les points associés aux modes de collecte internet et papier ont en général une coordonnée plus élevée que celui du mode de collecte face à face. Les individus interrogés en face à face qui expriment une insatisfaction à une question modèrent ensuite plus leur réponse aux autres questions que ceux qui répondent par internet ou par papier dont le comportement de réponse semble plus systématique. On remarque en effet que les modalités les plus extrêmes sont plus liées dans une interrogation par internet et sur questionnaire papier par rapport à une interrogation en face à face (voir matrice de corrélations 6). Ce phénomène peut s'interpréter comme une conjonction des effets désirabilité sociale et *satisficing* : en auto-administré, donc sans l'intermédiation d'un enquêteur, les répondants ont tendance à fournir des réponses plus tranchées et ces réponses tranchées sont plus systématiquement reportées simultanément car l'effort fourni pour reporter la réponse optimale est moindre.



Graphique 10 – 1^{re} AFMD : superposition des nuages global et partiels : zoom à la droite du premier axe

Modalités à gauche du premier axe (satisfaction)

Le même type d'interprétation que précédemment s'applique aux modalités liées à la satisfaction de la partie gauche du premier axe factoriel (graphique 11) : les opinions sont plus tranchées et plus simultanément reportées en auto-administré. Les points associés aux modes de collecte internet et papier sont toutefois moins systématiquement orientés vers les extrêmes par rapport à ce qui est observé à droite de l'axe.



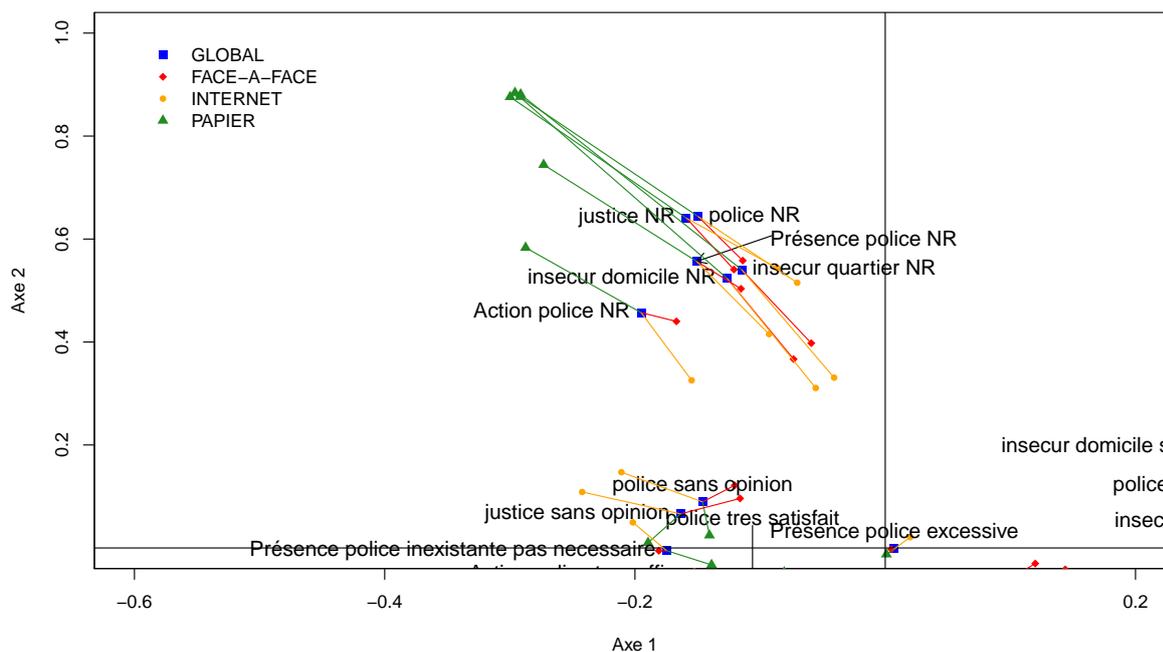
Graphique 11 – Superposition : zoom à la gauche du premier axe

Modalités à droite du deuxième axe (non réponse)

Sur le second axe factoriel, les modalités de non réponses associées au mode de collecte papier ont des coordonnées toujours plus élevées que celles des modes de collecte face à face et internet (graphique 12).

Précisons d'emblée qu'il semble difficile d'interpréter les résultats associés au mode de collecte par internet. En effet, les effectifs de non répondants sont très faibles (quelques unités), la non réponse étant semi-bloquante dans ce mode de collecte. La modalité de non réponse n'est offerte que dans un deuxième temps, lorsque l'individu essaie de passer la question si bien que certains individus ne sont pas forcément conscients que cette possibilité existe. Il est ainsi difficile dans ce cas d'interpréter des coefficients de corrélation. Si l'on compare les points partiels associés au questionnaire papier et à l'interrogation en face à face, les différences de coordonnées mettent en avant des différences de corrélation dans les modalités de non réponse entre ces deux modes de collecte. Les non répondants à une question en face à face répondent plus souvent aux autres questions que ceux répondants sur questionnaire papier. Il y a moins de répétition de non réponse en face à face.

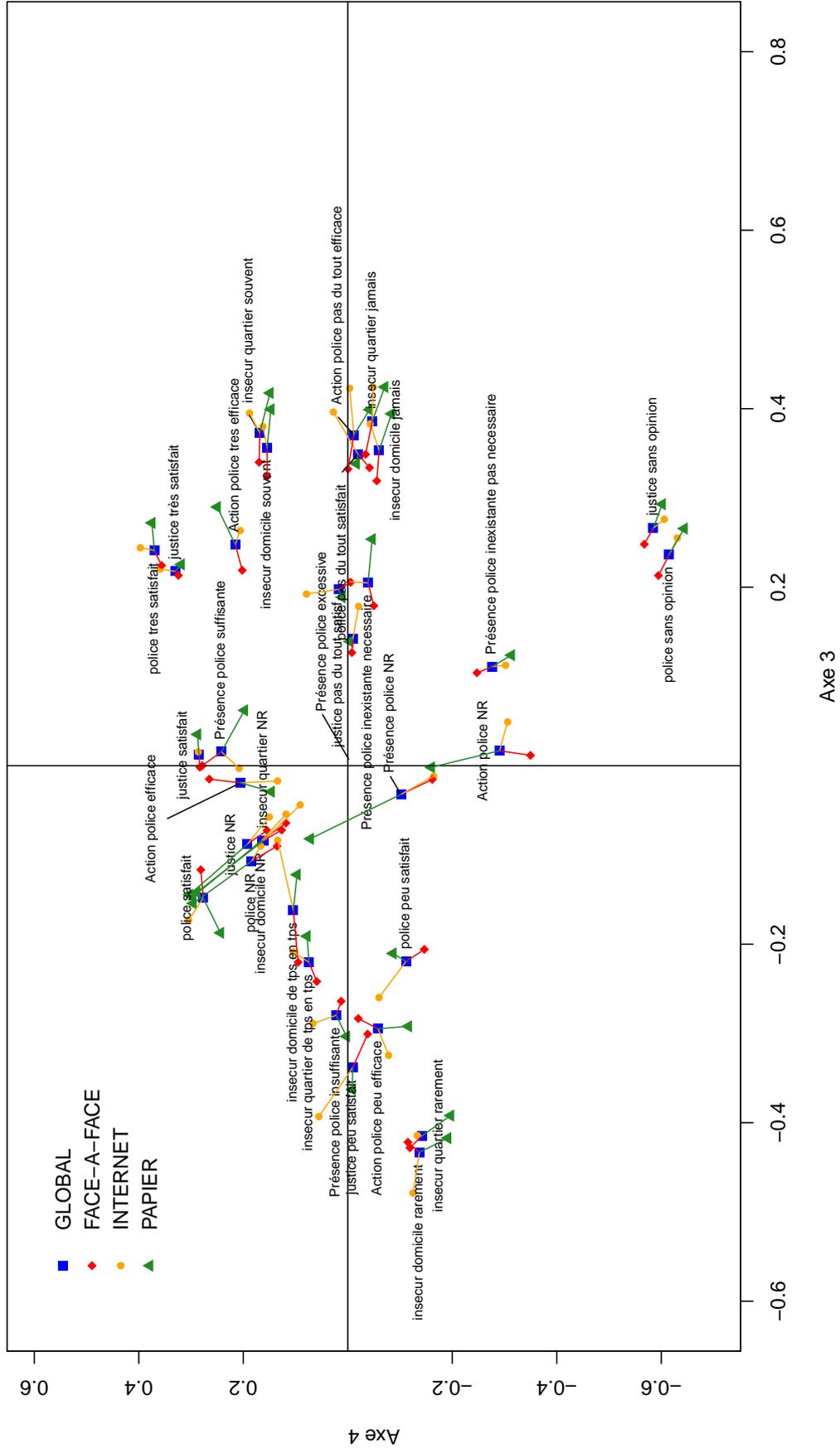
Cette tendance relève vraisemblablement du *satisficing* : sans l'intermédiation d'un enquêteur, le répondant est moins impliqué et en particulier, les efforts cognitifs entrepris sont moindres. L'une des manières de minimiser l'effort est d'utiliser assez fréquemment l'échappatoire que constitue le fait de ne pas répondre à la question. En revanche, en face à face, la présence d'un enquêteur fournit une motivation à répondre : pour des raisons qui tiennent en partie à la bienveillance, la tendance systématique à ne pas répondre est moindre qu'en auto-administré.



Graphique 12 – 1^{re} AFMD : superposition des nuages global et partiels : zoom à la droite du deuxième axe

L'examen du troisième axe

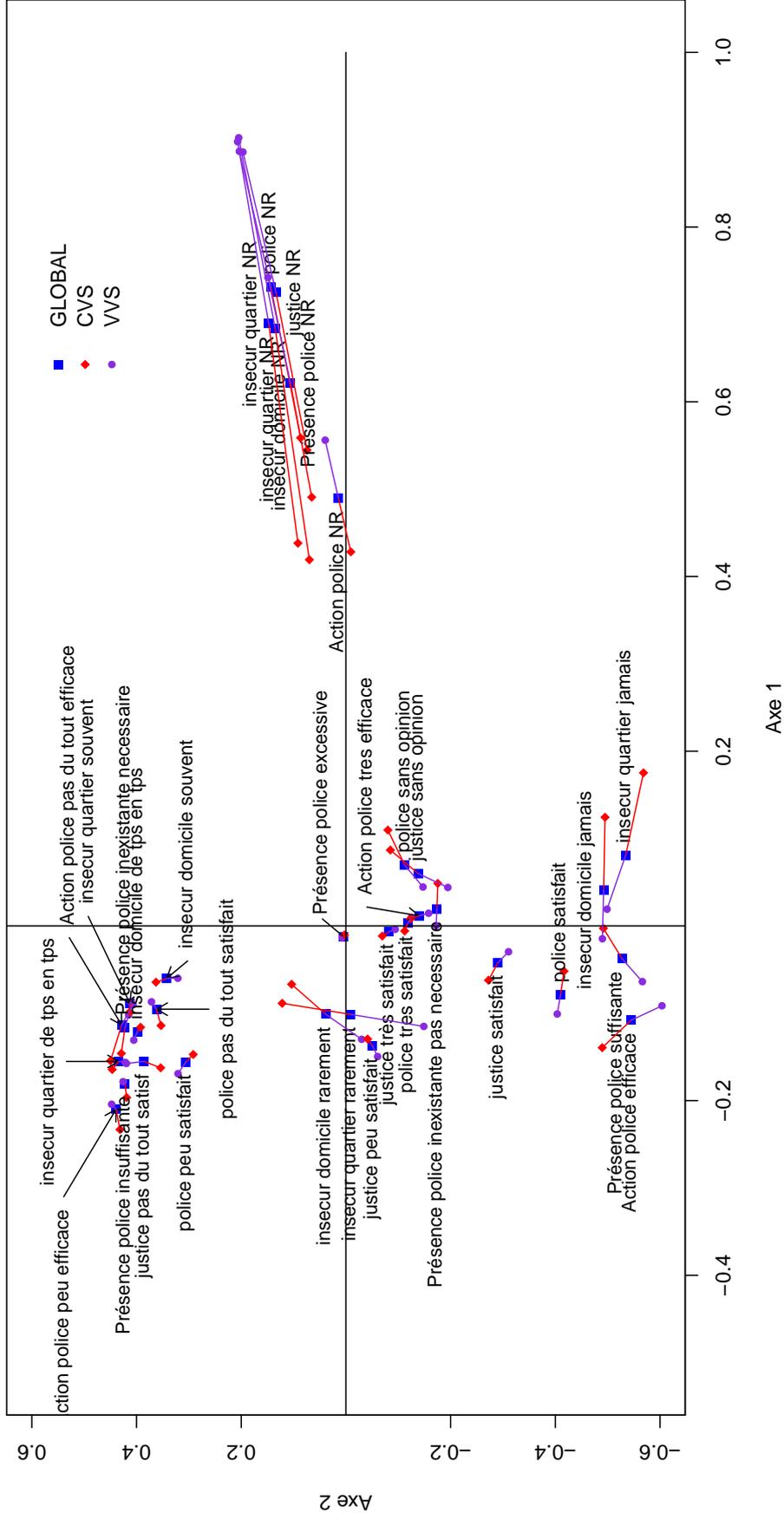
L'examen de l'axe 3 confirme cette tendance (graphique 13). Cet axe oppose deux types de répondants : ceux qui renseignent des modalités plutôt modérées (peu satisfait, de temps en temps), à ceux qui renseignent des modalités extrêmes (très satisfait, pas du tout satisfait...). Sur cet axe, les projections des points partiels associés aux modes de collecte internet et papier sont systématiquement plus éloignées du centre de l'axe que celles des points associés au mode de collecte en face à face. La partie gauche de l'axe indique ainsi que les individus renseignant des modalités associées à des opinions modérées à une question le font aussi aux autres questions et ce comportement est plus marqué dans un mode d'interrogation par internet et par papier.



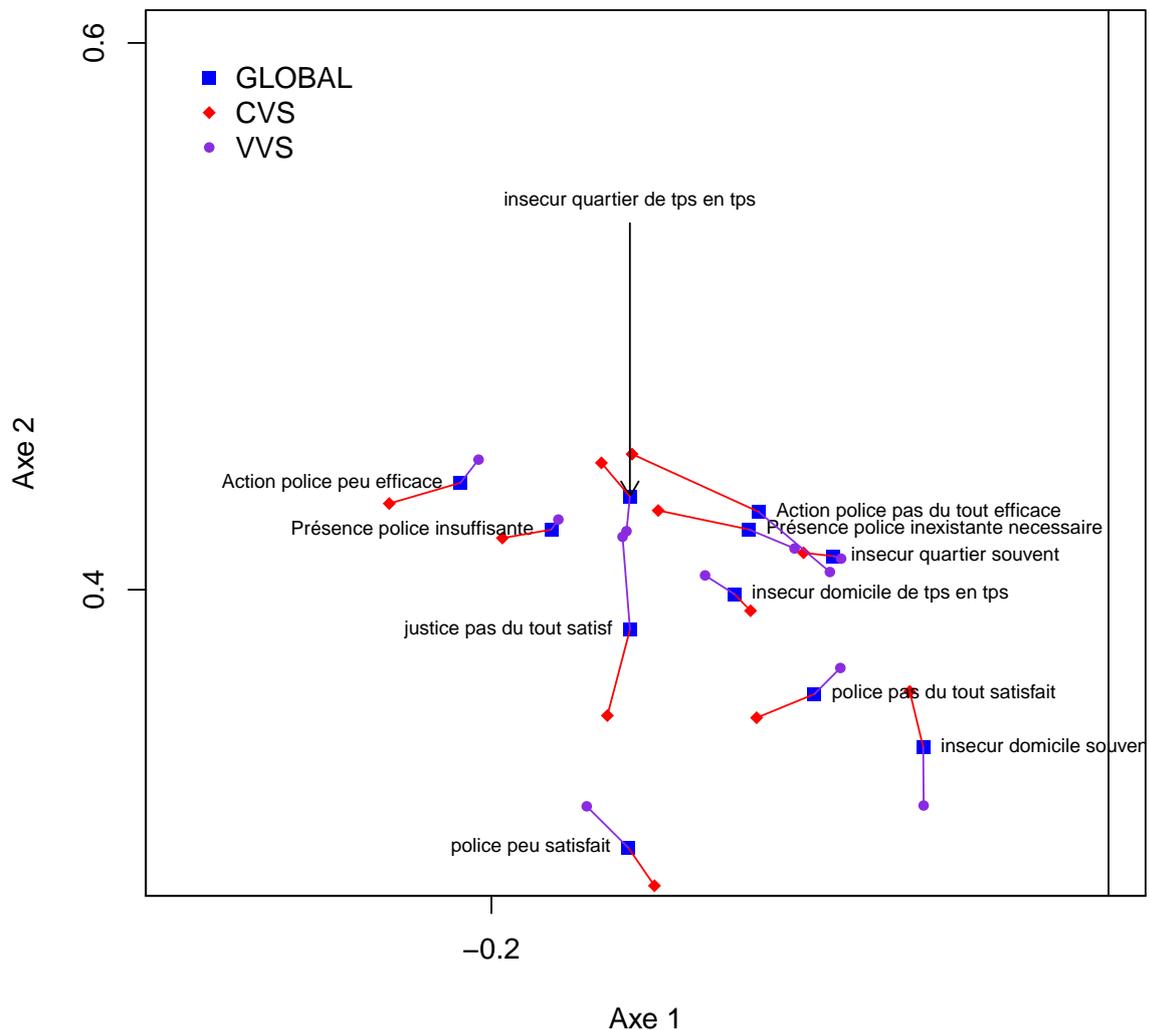
Graphique 13 – 1^{re} AFMD : superposition des nuages global et partiels des modalités (axes 3 et 4)

2.2.4 D'une analyse de données groupées par mode... à une analyse de données groupées par enquête

Jusqu'ici l'analyse a porté sur des données groupées par mode de collecte : face à face, internet et papier. Rappelons que la possibilité de répondre par questionnaire papier n'est offerte aux enquêtés du dispositif VVS qu'après la première relance : on peut souhaiter considérer le protocole de VVS comme un tout et non comme une juxtaposition de deux modes de collecte bien distincts. De plus, la littérature sur le sujet indique que la dichotomie essentielle se situe entre les modes avec intermédiation d'enquêteur et les modes auto-administrés. Pour ces raisons, les modes de collecte internet et papier ont été regroupés dans une deuxième analyse. Les représentations graphiques des nuages global et partiels sont globalement très similaires à ce qui a été présenté précédemment, hormis le fait que l'axe de la non réponse ressort désormais en premier. Deux illustrations (graphiques 14 et 15) de cette nouvelle configuration de regroupement sont fournies : la superposition des nuages global et partiels ainsi qu'un zoom sur les modalités situées à droite du deuxième axe.



Graphique 14 – 2^e AFMD : superposition des nuages global et partiels des modalités (axes 1 et 2)



Graphique 15 – 2^e AFMD : superposition des nuages global et partiels (zoom à droite du deuxième axe)

Les coordonnées des groupes d'individus (tableau 4 en annexe) confirment des comportements de non réponse plus systématiquement répétés dans un mode de collecte auto-administré par rapport à une collecte par l'intermédiation d'un enquêteur. Les corrélations entre modalités d'insatisfaction d'une part et de satisfaction d'autre part sont plus élevées dans la collecte sur questionnaire papier ou par internet.

2.2.5 Projection de variables supplémentaires

Les variables d'intérêt de notre analyse sont les questions d'opinion de l'enquête : on souhaite connaître la structure sous-jacente de ces données groupées par enquête. Mais, au delà de cet aspect, l'analyse peut être enrichie en introduisant quelques variables illustra-

tives qui seront projetées sur les axes factoriels construits à partir des questions d’opinion (tableau 5).

Variables supplémentaires : victimations

Comme attendu, les individus ayant subi une atteinte (à la personne ou aux biens) se projettent nettement à droite du premier axe factoriel (voir tableau 5 en annexe). Il n’est en effet guère surprenant que ces personnes reportent dans leur différentes réponses une certaine insatisfaction quant à la sécurité ou leur cadre de vie. La projection se situe, globalement, d’autant plus à l’extrémité droite de l’axe que l’atteinte subie est grave : ainsi, l’insatisfaction des personnes victimes de violences physiques ou de vols avec violence est plus forte que celles des victimes de vols de véhicule ou de vols sans violence.

Variables supplémentaires : caractéristiques sociodémographiques

La projection des caractéristiques socio-démographiques sur les axes factoriels n’est pas aussi marquée que celle des variables de victimation. Néanmoins, certaines tendances peuvent être observées. Par exemple, sur l’axe de l’insatisfaction, il existe une certaine opposition entre les ruraux et les urbains : alors que les personnes vivant en milieu rural se situent plutôt à gauche de l’axe, celles qui habitent dans les grandes unités urbaines se projettent plutôt du côté de l’insatisfaction. Deux autres constats sur cet axe : les personnes ayant un diplôme de l’enseignement supérieur plus élevé que Bac+2 sont du côté gauche de l’axe et en revanche, celles qui vivent au sein de familles monoparentales sont plutôt à droite de l’axe. En ce qui concerne le deuxième axe, correspondant à la non réponse, le résultat le plus marquant est que les personnes très âgées se projettent assez nettement du côté de la non réponse aux différentes questions d’opinion. Ce diagnostic pourrait encore être affiné en analysant comment ces coordonnées évoluent d’un mode de collecte à un autre.

Ces différents résultats montrent qu’il existe une corrélation entre certaines des caractéristiques sociodémographiques des individus et les axes factoriels issus de l’AFMD sur les questions d’opinion. Or, la population qui répond à l’enquête CVS n’est pas identique à celle qui répond à VVS : il existe des différences entre ces populations qui relèvent de la sélection. Ainsi, les effets que l’AFMD nous permet de détecter peuvent ne pas être strictement liés à la mesure mais être en partie confondus avec des effets de sélection. Afin de voir si les résultats trouvés à partir de l’AFMD sur les données groupées par enquête sont robustes à un contrôle de la sélection (sur caractéristiques observables), des techniques de *matching* ont été mises en œuvre.

2.2.6 Robustesse des résultats après appariement (ou *matching*)

Les techniques d’appariement (ou *matching*) sont aujourd’hui très fréquemment utilisées avec de nombreux exemples d’application dans des domaines aussi variés que la statistique (traitement de données manquantes), l’économie (évaluation d’une politique publique), l’épidémiologie (contrôle des facteurs de confusion), la sociologie ou encore les sciences politiques (Stuart, 2010). Ces techniques peuvent être mobilisées dès lors que l’on souhaite rapprocher des groupes d’observations, pour lesquels on dispose d’un ensemble de variables communes. Par exemple, le *matching* peut être utilisé pour comparer deux populations, en contrôlant les différences sur certaines caractéristiques observables, afin

de purger la comparaison de certains effets de composition.

En effet, dans le cadre de notre analyse, la comparaison (via AFMD) des réponses collectées pour CVS et VVS n'est pas neutre face à la sélection. Les différences constatées peuvent ainsi relever, pour partie, de la sélection : les personnes qui répondent à ces enquêtes ne sont pas les mêmes. En effet, si les échantillons d'individus sélectionnés pour ces deux dispositifs sont comparables, l'ampleur ainsi que les déterminants de la non réponse ne sont pas du même ordre d'où des différences, *in fine*, entre les populations de répondants. Afin de contrôler ces différences de structure entre les populations, des techniques d'appariement sont utilisées de manière à ce que certaines caractéristiques observables bien choisies se distribuent de la même manière au sein des deux populations de répondants.

Principes généraux du *matching*

Le contexte est ici proche de celui de l'évaluation de l'effet causal d'un traitement Givord (2010) où le traitement serait le fait de répondre par l'un des deux types de modes de collecte considérés (avec enquêteur *vs* auto-administré).

Le *matching* est basé sur l'idée intuitive de comparer des individus traités avec des individus comparables (par leurs caractéristiques observables) qui eux ne sont pas traités. La différence est alors attribuée au traitement.

La méthode initialement proposée par Rubin (1977) consiste à appairer chaque individu i à un individu non traité comparable $\tilde{i}(i)$, au sens des caractéristiques observables X , qui est son "plus proche voisin". Par exemple, on peut prendre pour contrefactuel l'individu non traité dont l'indice vérifie :

$$\tilde{i}(i) = \operatorname{argmin}_{j \in I_0} \|X_i - X_j\|$$

où $\| \cdot \|$ est une distance (par exemple, la distance de Mahalanobis), I_0 l'ensemble des individus non traités.

L'hypothèse principale, dite d'indépendance conditionnelle, sur laquelle repose l'estimation par *matching* est que, conditionnellement aux variables observables de contrôle X , l'affectation au traitement est indépendante des variables d'intérêt. Afin que cette hypothèse soit vérifiée, on peut souhaiter effectuer l'appariement sur un grand nombre de variables. Mais les proches voisins peuvent alors être difficiles à trouver. Un moyen de résoudre ce problème de dimensionnalité est d'utiliser le score de propension. Rosenbaum et Rubin (1983) ont en effet montré que si la propriété d'indépendance conditionnelle est vérifiée, alors l'estimation de l'effet sur les traités peut s'effectuer en appariant sur le score de propension, à savoir la probabilité d'être traité conditionnellement aux observables X .

Ainsi la technique d'appariement qui est mise en œuvre ici est celle du *matching* selon la méthode du plus proche voisin, utilisant la distance sur les scores de propension estimés (à partir d'une régression logistique). Et pour répondre à notre problématique, les variables de contrôle X doivent être à la fois liées au processus de sélection (le fait de répondre via un certain mode de collecte) ainsi qu'aux variables d'intérêt.

Mise en œuvre pratique sur les données AFMD

Notre stratégie consiste à effectuer une AFMD où les trois groupes de données sont :

- les données des répondants CVS (face à face) ;
- les données des répondants VVS (internet ou papier) ;
- les données des répondants VVS dont les caractéristiques observables sont contrôlées par *matching* en utilisant le package MatchIt de R avec comme variables de contrôle X, l'âge, le sexe, le diplôme, la catégorie socioprofessionnelle, la taille d'unité urbaine et le type de ménages : pour chaque individu de CVS, l'algorithme de traitement trouve un individu de VVS comparable par ses caractéristiques observables et cet ensemble d'individus trouvés constitue le troisième groupe de nos données.

Mettre en œuvre une telle stratégie se justifie car des différences existent dans les structures des populations qui répondent à CVS et VVS. Il est important de rappeler que les personnes qui répondent aux enquêtes par internet ont des caractéristiques spécifiques : ces personnes sont en général plus jeunes et plus diplômées que la population cible ou que celle qui répond aux enquêtes en face à face, en partie parce que l'accès à internet n'est pas uniforme au sein de la population. Ici, cet écueil est moindre car les personnes peuvent également répondre en retournant la version papier du questionnaire par courrier postal. Les différences dans les structures des populations de CVS et VVS sont ainsi moins marquées que si la réponse était uniquement possible par internet pour VVS³.

L'idée est de pouvoir représenter dans un même repère factoriel, par superposition, les nuages partiels de CVS et ceux de VVS sans ou avec contrôle par *matching*. Ceci nous permettra de constater si les différences entre les modes repérées avec le premier jeu d'AFMD persistent lorsque l'on contrôle de quelques caractéristiques observables.

Résultats du matching

Parmi les principales différences de structures de population que l'on peut constater :

- les femmes sont, en proportion, un peu plus nombreuses dans CVS que dans VVS ;
- la proportion de personnes très âgées est plus importante dans CVS ;
- les personnes sans diplôme sont plus nombreuses dans CVS et, en revanche, la proportion des diplômés du supérieur est moindre que dans VVS ;
- plus de réponses proviennent d'habitants de l'agglomération parisienne dans VVS que dans CVS.

Le *matching* permet de disposer, par construction, d'un troisième groupe de répondants : une population de répondants VVS dont la structure, sur les caractéristiques observables retenues, est très proche de celle de CVS. Sans détailler l'ensemble des fréquences, le tableau 2 indique, pour les trois groupes de répondants (CVS, VVS et répondants VVS appariés) les proportions de répondants appartenant aux catégories recensées plus haut. Ainsi, par exemple, la proportion de personnes sans diplôme ou avec un certificat d'études du groupe de répondants VVS appariés (26,5 %) est très proche de celle de CVS (26,3 %) alors qu'elle est bien inférieure dans la population "non contrôlée" de répondants VVS

3. Par exemple, les 25-39 ans sont en proportion plus nombreux dans les répondants VVS par internet (28,9 %) que dans les répondants CVS (20,0 %). Mais en tenant compte des réponses papier, l'écart se réduit substantiellement et la proportion globale de 25-39 ans parmi les répondants VVS est de 20,5 %.

(18,6 %).

Par ailleurs, le contrôle de la sélection ne peut s’effectuer que sur des variables observables. Les taux de réponse moindres pour l’enquête expérimentale VVS peuvent laisser penser que les répondants sur internet et papier ont des caractéristiques particulières non contrôlées par le *matching*. Une hypothèse plausible (Razafindranovona et al., 2014) serait que ces répondants sont en moyenne plus intéressés, plus motivés par la thématique de l’enquête. Cela pourrait éventuellement nuancer les conclusions obtenues en termes d’effets de mode. Et si l’on pense que la motivation pour la thématique est positivement liée au fait d’avoir été victime d’une atteinte, une manière de contrôler partiellement ces effets de sélection serait d’introduire les variables de victimation dans l’appariement. Cette analyse complémentaire a été réalisée et les résultats (non présentés ici mais disponibles auprès des auteurs) sont très proches de ceux exposés ci-après.

| | CVS (%) | VVS (%) | VVS après <i>matching</i> (%) |
|-------------------------------------|---------|---------|-------------------------------|
| Sexe | | | |
| Femmes | 56,9 | 53,2 | 56,8 |
| Âge | | | |
| 70 ans et plus | 20,4 | 17,5 | 20,1 |
| Diplôme | | | |
| Aucun diplôme ou CEP | 26,3 | 18,6 | 26,5 |
| Diplôme de niveau Bac+2 | 11,5 | 14,6 | 11,5 |
| Diplôme de niveau supérieur à Bac+2 | 14,2 | 16,7 | 14,3 |
| Unité urbaine | | | |
| Paris | 10,4 | 14,8 | 10,2 |

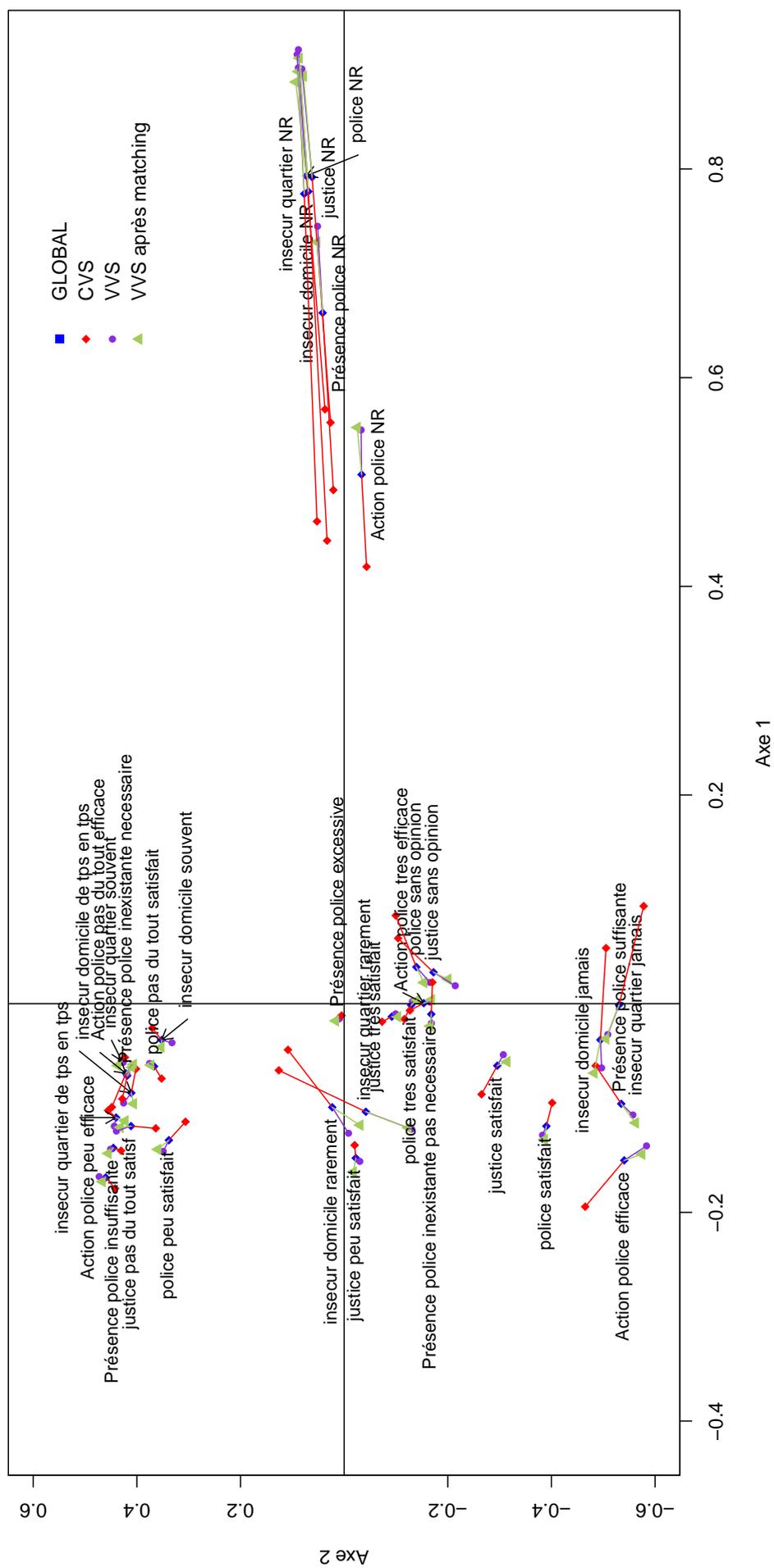
Tableau 2 – Comparaison sur quelques caractéristiques observables avant et après *matching*

Une AFMD est alors appliquée sur le tableau de données dont les trois groupes sont les répondants CVS, les répondants VVS et le groupe constitué via *matching*, les répondants VVS appariés. Le principal résultat de cette analyse est que les répondants VVS appariés se projettent sur les axes factoriels de la même manière que les répondants VVS non contrôlés. Ce résultat peut ainsi s’illustrer par la représentation graphique de la superposition des nuages global et partiels sur le premier plan factoriel (graphique 16) : les points VVS et VVS appariés sont pratiquement confondus. Le constat est similaire lorsque l’on s’intéresse à certaines modalités bien spécifiques qui contribuent à la construction des axes, comme par exemple, celles qui se situent à la droite du premier axe factoriel (graphique 19). Plus précisément, la position du groupe des VVS appariés se rapproche tout de même très légèrement de celle des répondants CVS⁴ mais le mouvement n’est que minime : globalement, la position des deux groupes de répondants VVS, qu’ils soient contrôlés ou non, est très proche.

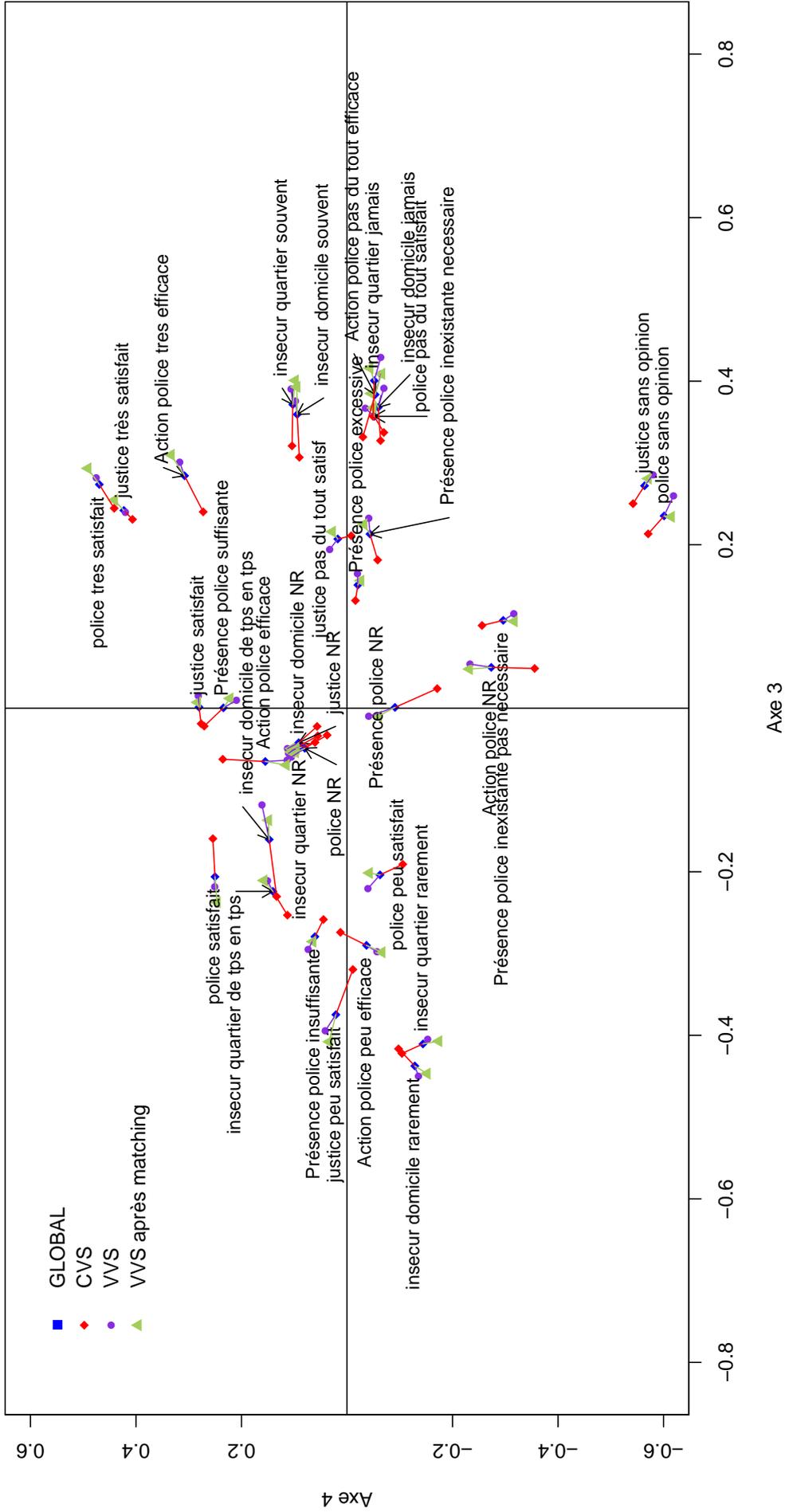
Un tel résultat semble indiquer une certaine robustesse de nos conclusions en termes d’effets de mode. Que l’on contrôle ou non par certaines caractéristiques observables,

4. Un rapprochement est attendu puisque les différences de structures de population sont moindres entre ces populations qu’entre CVS et VVS.

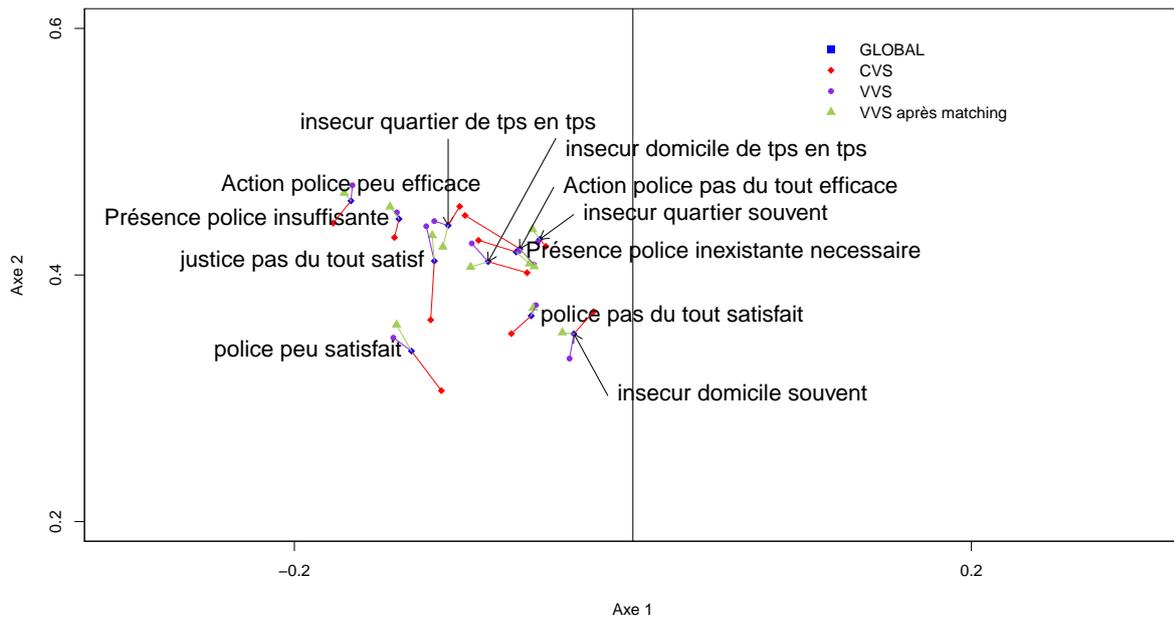
les différences CVS/VVS dans les associations entre modalités sont du même ordre : les opinions sont plus tranchées et plus souvent simultanément répétées en auto-administré et le recours systématique à la non réponse est une stratégie moins utilisée lorsque l'on fait face à un enquêteur. Ces résultats ne sont pas en soi surprenants et plutôt conformes à la théorie. Mais rappelons le caractère exploratoire de la démarche : l'analyse est effectuée sans *a priori* et les conclusions présentées ici reflètent véritablement la structure sous-jacente des réponses aux questions d'opinion de ces enquêtes de victimation.



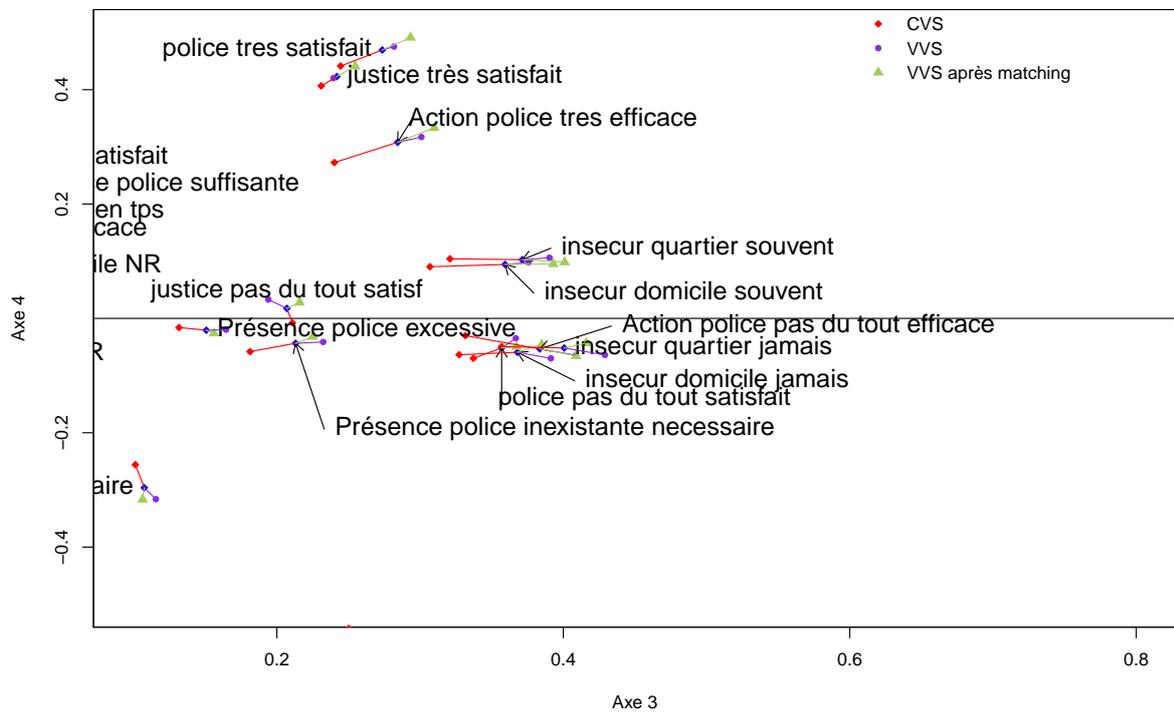
Graphique 16 – 3^e AFMD : superposition des nuages global et partiels des modalités, axes 1 et 2



Graphique 17 – 3^e AFMD : superposition des nuages global et partiels des modalités, axes 3 et 4



Graphique 18 – 3^e AFMD : superposition des nuages global et partiels des modalités, zoom à la droite du deuxième axe



Graphique 19 – 3^e AFMD : superposition des nuages global et partiels des modalités, zoom à la droite du troisième axe

Références

- Abascal, E., V. Díaz de Rada, I. García Lautre et I. Landaluce. 2013, «Extending dual multiple factor analysis to categorical tables», *Journal of Applied Statistics*, vol. 40, n° 2, p. 415–428.
- Couper, M. 2011, «The future of modes of data collection», *Public opinion quarterly*, vol. 75, n° 5, p. 889–908.
- Escofier, B. et J. Pagès. 1998, *Analyses factorielles simples et multiples ; objectifs, méthodes et interprétation*, DUNOD.
- Gelein, B. et O. Sautory. 2002, «De nouvelles macros sas d'analyse de données à l'insee : comment réaliser une analyse factorielle multiple?», *Huitièmes Journées de Méthodologie Statistique*. URL http://jms.insee.fr/files/documents/2002/361_1-JMS2002_SESSION8_GELEIN-SAUTORY_MACRO-SAS-ANALYSE-FACTORIELLE-MULTIPLE_ACTES.PDF.
- Givord, P. 2010, «Méthodes économétriques pour l'évaluation de politiques publiques», *INSEE, Document de travail de la DESE*.
- Lê, S. et J. Pagès. 2010, «DMFA : dual multiple factor analysis», *Communications in statistics. Theory and methods*, vol. 39, n° 3, p. 483–492.
- de Peretti, G. et T. Razafindranovona. 2014, «Les enquêtes multimode : attention aux effets de mode», *Statistique et société*, vol. 2.
- Razafindranovona, T., B. Dietsch, C. Burricand et G. de Peretti. 2014, «Le multimode pour mesurer la victimation : est-on dans la zone de sécurité?», dans *8ème colloque francophone sur les sondages*.
- Stuart, E. A. 2010, «Matching methods for causal inference : A review and a look forward», *Statistical science : a review journal of the Institute of Mathematical Statistics*, vol. 25, n° 1, p. 1.
- Zauberman, R. et P. Robert. 2011, «Les enquêtes de victimation en europe», *Economie et statistique*, vol. 448, n° 1, p. 89–105.

1 Annexe : le nuage des groupes d'individus

1.1 Le produit scalaire de Hilbert-Schmidt

Chaque groupe d'individus l est représenté par sa matrice d'inertie $C_l = 1/N^l Z'_l Z_l$. On définit sur l'espace des matrices d'inertie le produit scalaire de Hilbert-Schmidt entre les matrices d'inertie C_{l_1} et C_{l_2} de dimension K :

$$\langle C_{l_1}, C_{l_2} \rangle_{R^{K^2}} = \text{Trace}(C_{l_1} C_{l_2})$$

Quelques propriétés utiles de ce produit scalaire :

- Le produit scalaire entre deux matrices d'inertie est toujours positif

On montre que :

$$\begin{aligned} \langle C_{l_1}, C_{l_2} \rangle_{R^{K^2}} &= \frac{1}{N_{l_1} N_{l_2}} \text{Trace}(Z'_{l_1} Z_{l_1} Z'_{l_2} Z_{l_2}) = \frac{1}{N_{l_1} N_{l_2}} \text{Trace}(Z_{l_1} Z'_{l_2} Z_{l_2} Z'_{l_1}) \\ &= \frac{1}{N_{l_1} N_{l_2}} \text{Trace}(BB') \end{aligned}$$

avec $B = Z_{l_1} Z'_{l_2}$ la matrice dont le terme général est noté B_{ij} ($1 \leq i, j \leq K$).
D'où

$$\langle C_{l_1}, C_{l_2} \rangle_{R^{K^2}} = \frac{1}{N_{l_1} N_{l_2}} \text{Trace}(BB') = \frac{1}{N_{l_1} N_{l_2}} \sum_{i=1}^K \sum_{j=1}^K B_{ij}^2 \geq 0$$

- La distance au sens du produit scalaire de Hilbert-Schmidt entre deux groupes d'individus l_1 et l_2 est d'autant plus grande que les corrélations entre les variables sont différentes d'un groupe à l'autre, et d'autant plus pour les variables les plus corrélées entre elles. La distance entre deux groupes vaut en effet :

$$\| C_{l_1} - C_{l_2} \|_{R^{K^2}}^2 = \sum_{h=1}^K \sum_{k=1}^K r(k, h) [r(k^{l_1}, h^{l_1}) - r(k^{l_2}, h^{l_2})]^2 \quad (2)$$

- La norme d'une matrice d'inertie vaut :

$$\| C_l \|_{R^{K^2}}^2 = \sum_{k=1}^K [\lambda_k^l]^2$$

avec λ_k^l l'inertie du k^e axe factoriel dans l'ACP normée sur les individus du groupe l . Cette somme divisée par $[\lambda_1^l]^2$, soit $\sum_{k=1}^K [\lambda_k^l / \lambda_1^l]^2$ (où λ_1^l est l'inertie maximale du groupe d'individus l projeté sur un axe) correspond à la dimensionnalité du nuage du groupe l , c'est-à-dire au nombre de directions orthogonales équivalentes en terme d'inertie à celui maximisant l'inertie. Pour plus de détails sur cet indicateur voir (Escofier et Pagès, 1998).

1.2 La base des matrices U_s

On montre que les matrices $U_s = u_s u'_s$ associé aux composantes principales issues de l'AFMD $u_s = (u_s(1), \dots, u_s(K))'$, forment une base orthonormée de l'espace R^{K^2} pour le produit scalaire de Hilbert-Schmidt :

– Les U_s sont de norme 1.

$$\begin{aligned}\|U_s\|^2 &= \text{Trace}(u_s u_s' u_s u_s') = \text{Trace}(u_s' u_s u_s' u_s) = (u_s' u_s)^2 \\ &= \left[\sum_{k=1}^K u_s(k)^2 \right]^2 = \|u_s\|_{\mathbb{R}^K}^4 = 1\end{aligned}$$

– Les matrices U_{s_1} et U_{s_2} ($s_1 \neq s_2$) sont deux à deux orthogonales :

$$\langle U_{s_1}, U_{s_2} \rangle_{\mathbb{R}^{K^2}} = \langle u_{s_1} u_{s_1}', u_{s_2} u_{s_2}' \rangle_{\mathbb{R}^{K^2}} = 0$$

car $u_{s_1}' u_{s_2} = 0$, u_{s_1} et u_{s_2} étant orthogonaux par construction.

On montre que cette base orthonormée de matrices symétriques permet de représenter au mieux le nuage des groupes d'individus, en imposant que les axes soient engendrés par des matrices symétriques de norme 1 et que ce soit la somme (pondérée) des projections qui soit maximiser au lieu comme c'est le cas classiquement de la somme des carrés (ceci conduit à des calculs plus simples et peut se justifier par le fait que les projections sont nécessairement positives). Ce choix est analogue à celui qui est fait en analyse factoriel multiple. Le lecteur intéressé par ce point peut trouver plus de détails dans Escofier et Pagès (1998). Formellement, le programme de maximisation de recherche du s^e axe s'écrit ainsi :

$$\underset{\substack{v \in \mathbb{R}^K / \|vv'\|_{\mathbb{R}^{K^2}}^2 = 1 \\ vv' \perp v_k v_k', k=1, \dots, s-1}}{\text{argmax}} \sum_{l=1}^L \frac{N^l}{N} \langle C_l, vv' \rangle_{\mathbb{R}^{K^2}}$$

Preuve :

$$\begin{aligned}\underset{\substack{v \in \mathbb{R}^K / \|vv'\|_{\mathbb{R}^{K^2}}^2 = 1 \\ vv' \perp v_k v_k', k=1, \dots, s-1}}{\text{argmax}} \sum_{l=1}^L \frac{N^l}{N} \langle C_l, vv' \rangle_{\mathbb{R}^{K^2}} &= \underset{\substack{v \in \mathbb{R}^K / \|vv'\|_{\mathbb{R}^{K^2}}^2 = 1 \\ vv' \perp v_k v_k', k=1, \dots, s-1}}{\text{argmax}} \sum_{l=1}^L \frac{N^l}{N} \text{Trace}(C_l vv') \\ &= \underset{\substack{v \in \mathbb{R}^K / \|vv'\|_{\mathbb{R}^{K^2}}^2 = 1 \\ vv' \perp v_k v_k', k=1, \dots, s-1}}{\text{argmax}} \sum_{l=1}^L \frac{N^l}{N} \text{Trace}(v' C_l v) = \underset{\substack{v \in \mathbb{R}^K / \|vv'\|_{\mathbb{R}^{K^2}}^2 = 1 \\ vv' \perp v_k v_k', k=1, \dots, s-1}}{\text{argmax}} \sum_{l=1}^L \frac{N^l}{N} \times v' \left[\frac{1}{N^l} Z_l' Z_l \right] v\end{aligned}$$

On retrouve ainsi le même programme de maximisation (1) sous-jacent à la construction des axes de l'AFMD.

2 Questionnaire (extrait)

| | |
|---|----------------------------|
| C2 A propos de l'action en général de la police ou de la gendarmerie nationale dans la société française actuelle, vous diriez personnellement qu'elle est | |
| 1. Très satisfaisante..... | <input type="checkbox"/> 1 |
| 2. Satisfaisante..... | <input type="checkbox"/> 2 |
| 3. Peu satisfaisante..... | <input type="checkbox"/> 3 |
| 4. Pas du tout satisfaisante..... | <input type="checkbox"/> 4 |
| 5. Pas d'opinion..... | <input type="checkbox"/> 5 |
| Sans réponse..... | <input type="checkbox"/> |
| | POLICE |

| | |
|---|----------------------------|
| C3 A propos de l'action de la Justice et des tribunaux dans le traitement de la délinquance, vous diriez personnellement qu'elle est | |
| 1. Très satisfaisante..... | <input type="checkbox"/> 1 |
| 2. Satisfaisante..... | <input type="checkbox"/> 2 |
| 3. Peu satisfaisante..... | <input type="checkbox"/> 3 |
| 4. Pas du tout satisfaisante..... | <input type="checkbox"/> 4 |
| 5. Pas d'opinion..... | <input type="checkbox"/> 5 |
| Sans réponse..... | <input type="checkbox"/> |
| | JUSTICE |

| | |
|--|----------------------------|
| C4 Vous arrive-t-il personnellement de vous sentir en insécurité à votre domicile ? | |
| 1. Souvent..... | <input type="checkbox"/> 1 |
| 2. De temps en temps..... | <input type="checkbox"/> 2 |
| 3. Rarement..... | <input type="checkbox"/> 3 |
| 4. Jamais..... | <input type="checkbox"/> 4 |
| Sans réponse..... | <input type="checkbox"/> |
| | INSECURD |

| | |
|---|----------------------------|
| C5 Vous arrive-t-il personnellement de vous sentir en insécurité dans votre quartier ou dans votre village ? | |
| 1. Souvent..... | <input type="checkbox"/> 1 |
| 2. De temps en temps..... | <input type="checkbox"/> 2 |
| 3. Rarement..... | <input type="checkbox"/> 3 |
| 4. Jamais..... | <input type="checkbox"/> 4 |
| Sans réponse..... | <input type="checkbox"/> |
| | INSECURQ |

| | |
|--|----------------------------|
| H4 De la présence de la police ou de la gendarmerie dans votre quartier (ou votre village), vous diriez ? | |
| 1. Qu'elle est suffisante..... | <input type="checkbox"/> 1 |
| 2. Qu'elle est excessive..... | <input type="checkbox"/> 2 |
| 3. Qu'elle est insuffisante..... | <input type="checkbox"/> 3 |
| 4. Qu'elle est inexistante alors qu'elle serait nécessaire..... | <input type="checkbox"/> 4 |
| 5. Qu'elle est inexistante mais qu'elle n'est pas nécessaire..... | <input type="checkbox"/> 5 |
| Sans réponse..... | <input type="checkbox"/> |
| | POLPRES |

| | |
|---|----------------------------|
| H5 De l'action de la police ou de la gendarmerie en matière de lutte contre la délinquance dans votre quartier (ou votre village), vous diriez ? | |
| 1. Qu'elle est très efficace..... | <input type="checkbox"/> 1 |
| 2. Qu'elle est efficace..... | <input type="checkbox"/> 2 |
| 3. Qu'elle est peu efficace..... | <input type="checkbox"/> 3 |
| 4. Qu'elle n'est pas efficace du tout..... | <input type="checkbox"/> 4 |
| Sans réponse..... | <input type="checkbox"/> |
| | POLACT |

| | | |
|--|----------------------------|-----------------|
| C1 Dans la liste des problèmes suivants, quels sont d'après vous les trois les plus préoccupants dans la société française actuelle ? | | SET OF 3 |
| 1. La santé (alimentation, SIDA, grippe et autres épidémies,....) | <input type="checkbox"/> 1 | |
| 2. L'environnement (pollution de l'air, pollution des sols, qualité de l'eau, ...) | <input type="checkbox"/> 2 | |
| 3. Le terrorisme, les attentats..... | <input type="checkbox"/> 3 | |
| 4. Le chômage, la précarité de l'emploi..... | <input type="checkbox"/> 4 | |
| 5. La délinquance..... | <input type="checkbox"/> 5 | |
| 6. La pauvreté..... | <input type="checkbox"/> 6 | |
| 7. La sécurité routière..... | <input type="checkbox"/> 7 | |
| 8. Le racisme, la discrimination..... | <input type="checkbox"/> 8 | |
| Sans réponse..... | <input type="checkbox"/> | |
| | | PROBSOC |

Contrôle : vérifier que pas plus de 3 modalités cochées. Si 4 modalités ou plus cochées, faire apparaître : « vous ne pouvez pas cocher plus de 3 choix ».

Graphique 20 – Extrait du questionnaire CVS/VVS : variables d'intérêt

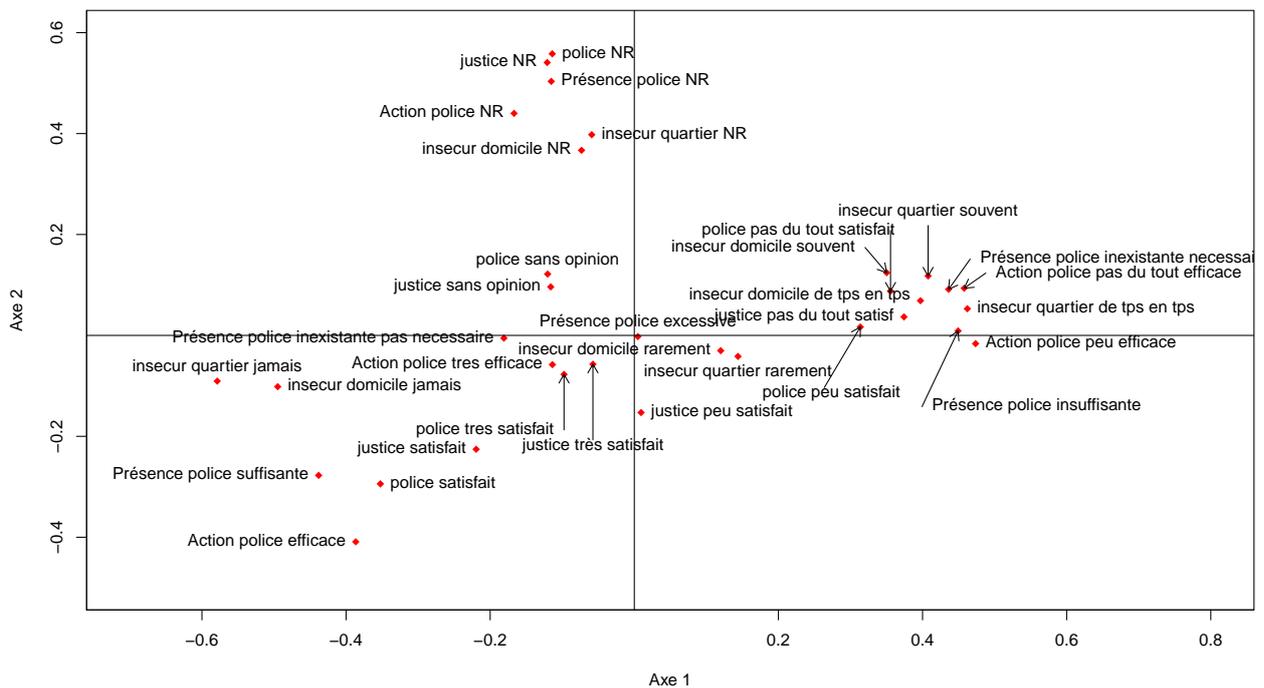
3 Principaux résultats des AFMD

3.1 Les valeurs propres

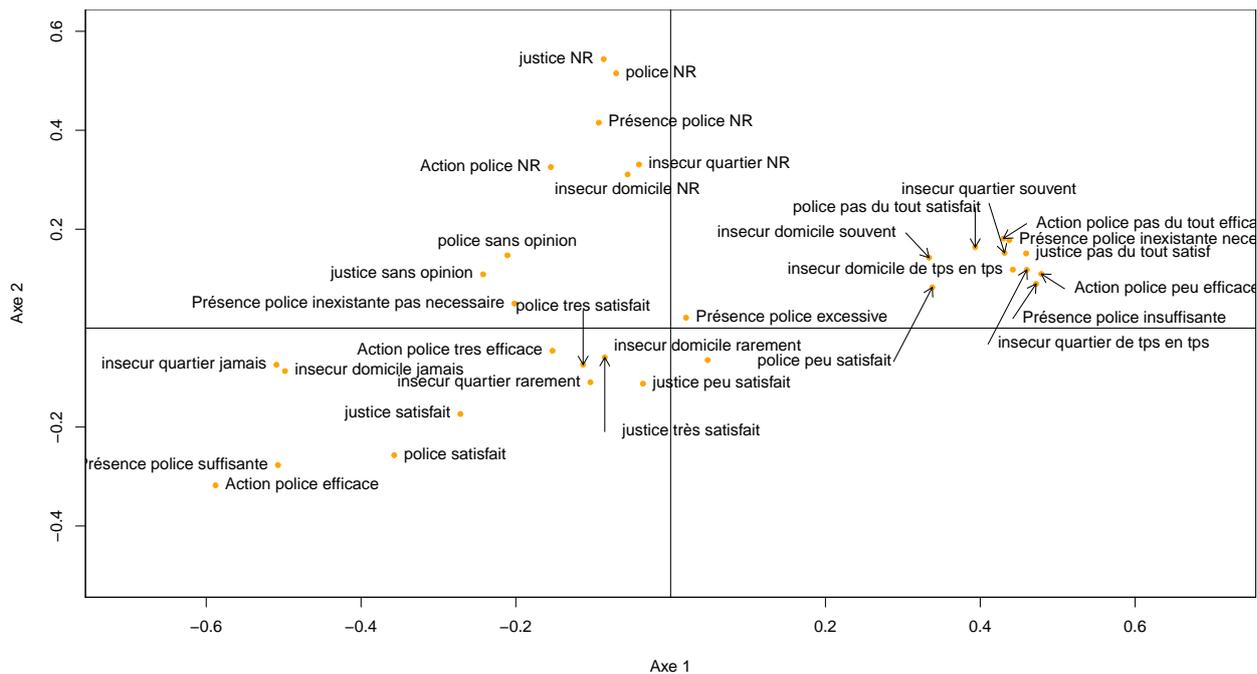
Tableau 3 – Les valeurs propres des 10 premiers axes

| Axe | 1 ^{er} AFMD | | | 2 ^e AFMD | | | 3 ^e AFMD | | |
|--------|----------------------|-------------|---------------------|---------------------|-------------|---------------------|---------------------|-------------|---------------------|
| | Inertie | % d'inertie | % d'inertie cumulée | Inertie | % d'inertie | % d'inertie cumulée | Inertie | % d'inertie | % d'inertie cumulée |
| Axe 1 | 2,4 | 8,87 | 8,87 | 2,74 | 10,13 | 10,13 | 3,01 | 11,14 | 11,14 |
| Axe 2 | 2,11 | 7,81 | 16,68 | 2,36 | 8,75 | 18,89 | 2,4 | 8,91 | 20,04 |
| Axe 3 | 1,6 | 5,91 | 22,59 | 1,6 | 5,93 | 24,82 | 1,64 | 6,09 | 26,13 |
| Axe 4 | 1,48 | 5,47 | 28,06 | 1,48 | 5,48 | 30,29 | 1,5 | 5,54 | 31,67 |
| Axe 5 | 1,36 | 5,05 | 33,11 | 1,36 | 5,04 | 35,33 | 1,38 | 5,11 | 36,78 |
| Axe 6 | 1,31 | 4,86 | 37,97 | 1,31 | 4,87 | 40,2 | 1,31 | 4,86 | 41,64 |
| Axe 7 | 1,26 | 4,67 | 42,64 | 1,26 | 4,68 | 44,88 | 1,27 | 4,69 | 46,33 |
| Axe 8 | 1,2 | 4,44 | 47,08 | 1,2 | 4,43 | 49,32 | 1,2 | 4,44 | 50,77 |
| Axe 9 | 1,07 | 3,95 | 51,03 | 1,05 | 3,9 | 53,21 | 1,04 | 3,87 | 54,64 |
| Axe 10 | 1,02 | 3,78 | 54,81 | 1,02 | 3,78 | 56,99 | 1,01 | 3,75 | 58,39 |

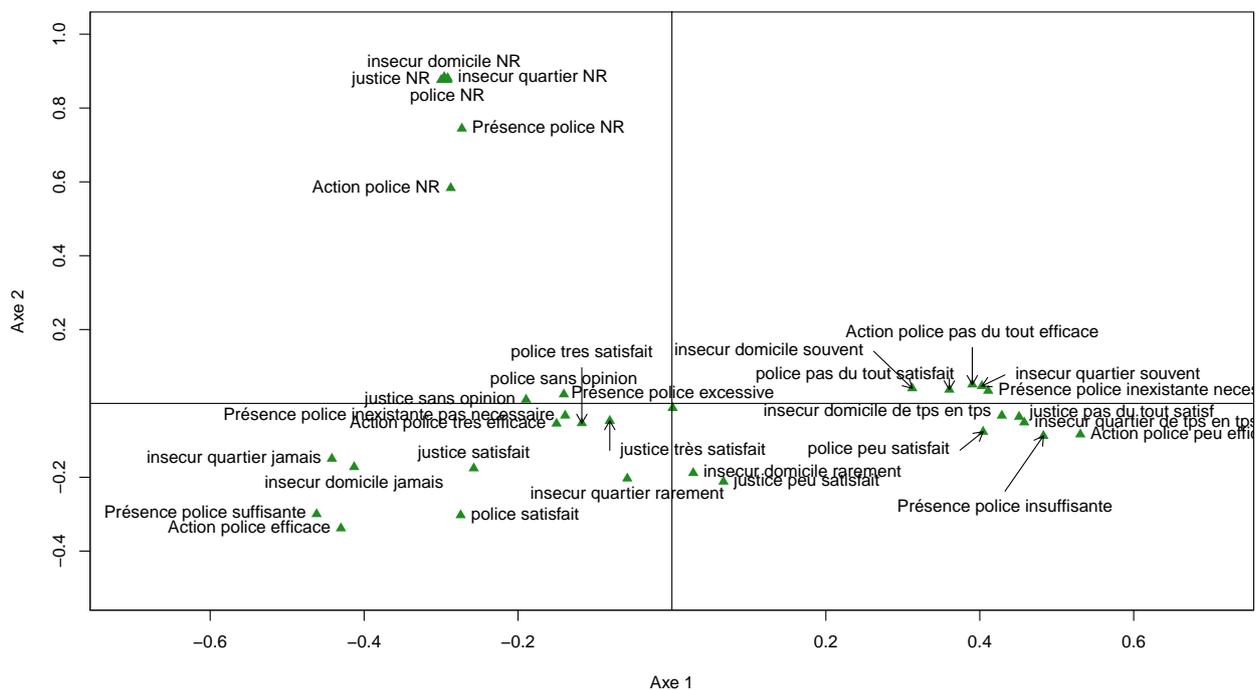
3.2 Les nuages partiels dans la première AFMD



Graphique 21 – Nuage partiel des variables, groupe des individus interrogés en face à face (axes 1 et 2)



Graphique 22 – Nuage partiel des variables, groupe des individus interrogés par internet (axes 1 et 2)



Graphique 23 – Nuage partiel des variables, groupe des individus interrogés par questionnaire papier (axes 1 et 2)

3.3 Les coordonnées des groupes d'individus

Tableau 4 – Coordonnées des groupes d'individus

| | Axe 1 | | | Axe 2 | | | Axe 3 | | | Axe 4 | | |
|----------------------|-------|----------------------|-------|-------|----------------------|-------|-------|----------------------|------|-------|----------------------|------|
| | Coord | Coord/ λ_1^l | cos2 | Coord | Coord/ λ_1^l | cos2 | Coord | Coord/ λ_1^l | cos2 | Coord | Coord/ λ_1^l | cos2 |
| 1 ^{er} AFMD | | | | | | | | | | | | |
| FAF | 2,26 | 0,97 | 16,28 | 1,57 | 0,68 | 7,87 | 1,42 | 0,61 | 6,44 | 1,4 | 0,6 | 6,24 |
| INTERNET | 2,53 | 0,96 | 19,23 | 1,28 | 0,48 | 4,9 | 1,79 | 0,68 | 9,63 | 1,51 | 0,57 | 6,86 |
| PAPIER | 2,57 | 0,6 | 14,46 | 4 | 0,93 | 34,95 | 1,8 | 0,42 | 7,08 | 1,61 | 0,38 | 5,68 |
| 2 ^e AFMD | | | | | | | | | | | | |
| CVS | 1,55 | 0,67 | 7,65 | 2,23 | 0,96 | 15,84 | 1,43 | 0,61 | 6,48 | 1,42 | 0,61 | 6,44 |
| VVS | 4,08 | 0,99 | 37,17 | 2,52 | 0,61 | 14,16 | 1,8 | 0,44 | 7,23 | 1,55 | 0,37 | 5,33 |
| 3 ^e AFMD | | | | | | | | | | | | |
| CVS | 1,51 | 0,64 | 7,3 | 2,26 | 0,96 | 16,27 | 1,42 | 0,6 | 6,41 | 1,41 | 0,6 | 6,32 |
| VVS | 4,12 | 1 | 37,88 | 2,52 | 0,61 | 14,17 | 1,8 | 0,44 | 7,24 | 1,55 | 0,37 | 5,35 |
| VVSmatch | 4,03 | 1 | 36,73 | 2,49 | 0,62 | 14,04 | 1,82 | 0,45 | 7,51 | 1,58 | 0,39 | 5,63 |

3.4 Les coordonnées des variables supplémentaires

Tableau 5: Coordonnées des variables supplémentaires

| | 1 ^{re} AFMD | | 2 ^e AFMD | | 3 ^e AFMD | | | | | |
|---------------------------------|----------------------|--------|---------------------|--------|---------------------|--------|-------|--------|--------|--------|
| | Axe 1 | Axe 2 | Axe 1 | Axe 2 | Axe 1 | Axe 2 | | | | |
| | Coord | V test | Coord | V test | Coord | V test | | | | |
| Prob. dans la société... | | | | | | | | | | |
| Santé | -0,05 | -3,88 | -0,1 | -9,15 | -0,09 | -6,75 | -0,09 | -14,78 | -11,99 | -0,09 |
| Environnement | -0,15 | -9,41 | -0,13 | -8,49 | -0,11 | -6,39 | -0,21 | -12,69 | -20,34 | -0,2 |
| Terrorisme | 0,22 | 11,06 | -0,09 | -4,63 | -0,17 | -8,16 | 0,16 | 8,33 | 15,95 | 0,2 |
| Chômage | -0,02 | -3,57 | -0,1 | -22,68 | -0,12 | -24,5 | -0,06 | -13,94 | -20,41 | -0,06 |
| Délinquance | 0,33 | 36,03 | -0,07 | -7,69 | -0,23 | -23,25 | 0,27 | 29,26 | -35,6 | 0,31 |
| pauvreté | -0,05 | -5,37 | -0,09 | -10,34 | -0,1 | -10,46 | -0,09 | -9,93 | -20,86 | -0,09 |
| Sécurité routière | 0,02 | 0,62 | -0,09 | -3,19 | -0,1 | -2,94 | -0,02 | -0,7 | -5,24 | -1,1 |
| Racisme | -0,14 | -7,22 | -0,05 | -2,79 | -0,04 | -2,06 | -0,17 | -8,36 | -14,66 | -0,19 |
| Sexe | | | | | | | | | | |
| Femme | 0,05 | 6,35 | 0,07 | 9,31 | 0,04 | 4,79 | 0,08 | 9,48 | 8,56 | 13,01 |
| Homme | -0,07 | -6,35 | -0,09 | -9,31 | -0,05 | -4,79 | -0,1 | -9,48 | -8,56 | -13,01 |
| CSP | | | | | | | | | | |
| Agriculteur | -0,39 | -4,58 | -0,13 | -1,65 | 0,03 | 0,3 | -0,41 | -4,86 | -0,02 | -0,32 |
| Commerçant, artisan | 0,13 | 2,6 | -0,16 | -3,31 | -0,25 | -4,56 | 0,04 | 0,83 | -0,26 | -7,18 |
| Cadre, profession libérale | -0,13 | -4,44 | -0,12 | -4,33 | -0,15 | -4,81 | -0,2 | -6,58 | -0,2 | -9,22 |
| Profession intermédiaire | 0,03 | 1,35 | -0,13 | -5,81 | -0,22 | -8,9 | -0,05 | -2,04 | -0,25 | -14,8 |
| Employé | 0,21 | 9,57 | -0,09 | -4,19 | -0,22 | -9,47 | 0,14 | 6,53 | -0,23 | -15,42 |
| Ouvrier | 0,11 | 3,94 | -0,15 | -5,99 | -0,21 | -7,41 | 0,03 | 0,99 | -0,23 | -12,3 |
| Retraité | -0,06 | -4,48 | -0,03 | -2,6 | 0,02 | 1,03 | -0,06 | -4,62 | -0,03 | -2,77 |
| Inactif | -0,01 | -0,38 | 0,08 | 3,28 | 0,07 | 2,66 | 0,02 | 0,98 | 0,07 | 4,16 |
| Victime... | | | | | | | | | | |
| Cambrilage | 0,88 | 21,23 | 0,04 | 0,97 | -0,37 | -8,36 | 0,8 | 19,58 | -0,31 | -10,92 |
| Vol de véhicule | 0,63 | 14,77 | 0,06 | 1,43 | -0,28 | -6,12 | 0,58 | 13,7 | -0,24 | -8,13 |
| Vol avec violence | 1,12 | 15,3 | 0,15 | 2,18 | -0,39 | -4,96 | 1,07 | 14,7 | -0,29 | -5,98 |
| Vol sans violence | 0,64 | 14,42 | 0,03 | 0,78 | -0,31 | -6,59 | 0,58 | 13,09 | -0,3 | -9,9 |
| Violence physique | 1 | 16,31 | 0,09 | 1,49 | -0,4 | -6,04 | 0,93 | 15,31 | -0,29 | -6,77 |
| Menace | 1,01 | 26,37 | 0,07 | 2,01 | -0,43 | -10,52 | 0,93 | 24,45 | -0,36 | -13,81 |
| Age | | | | | | | | | | |
| 14-24 ans | -0,02 | -0,54 | 0,03 | 1,16 | -0,05 | -1,59 | -0,02 | -0,79 | -0,11 | -5,11 |
| | | | | | | | | | | -2,92 |
| | | | | | | | | | | -0,06 |

| | 1 ^{re} AFMD | | | | 2 ^e AFMD | | | | 3 ^e AFMD | | | |
|----------------------------|----------------------|-----------------|----------------|-----------------|---------------------|-----------------|----------------|-----------------|---------------------|-----------------|----------------|-----------------|
| | Axe 1 Coord | Axe 1 V test | Axe 2 Coord | Axe 2 V test | Axe 1 Coord | Axe 1 V test | Axe 2 Coord | Axe 2 V test | Axe 1 Coord | Axe 1 V test | Axe 2 Coord | Axe 2 V test |
| 25-39 ans | 0,12 | 6,77 | -0,05 | -3,12 | -0,17 | -8,78 | 0,07 | 3,97 | -0,18 | -13,74 | 7,19 | 0,08 |
| 40-49 ans | 0,06 | 2,94 | -0,05 | -2,62 | -0,12 | -5,21 | 0,02 | 1,14 | -0,12 | -8,25 | 2,2 | 0,03 |
| 50-59 ans | 0,03 | 1,3 | -0,07 | -3,68 | -0,08 | -3,47 | -0,01 | -0,26 | -0,09 | -6,56 | 0,47 | 0,01 |
| 60-69 ans | 0,01 | 0,34 | -0,07 | -3,78 | -0,04 | -1,69 | -0,02 | -0,75 | -0,01 | -1,02 | -1,17 | -0,01 |
| 70ans et plus | -0,21 | -10,97 | 0,22 | 12,09 | 0,41 | 20,14 | -0,07 | -3,63 | 0,44 | 32,79 | -6,68 | -0,08 |
| Type de ménage | | | | | | | | | | | | |
| couple avec enfant | 0,04 | 3,01 | -0,1 | -7,51 | -0,18 | -12,65 | -0,02 | -1,63 | -0,2 | -20,7 | -1,77 | -0,02 |
| couple sans enfant | -0,04 | -2,62 | -0,07 | -4,74 | -0,05 | -3,18 | -0,07 | -4,35 | -0,04 | -3,36 | -6,44 | -0,06 |
| Famille monoparentale | 0,28 | 7,57 | 0,07 | 2,06 | -0,04 | -1,11 | 0,28 | 7,79 | -0,08 | -3,13 | 11,02 | 0,24 |
| Personne seule | -0,06 | -4,35 | 0,13 | 10,85 | 0,22 | 16,07 | 0,02 | 1,75 | 0,23 | 25,25 | 2,09 | 0,02 |
| Lieu de résidence | | | | | | | | | | | | |
| Paris | 0,24 | 9,59 | 0,1 | 4,35 | -0,03 | -1,28 | 0,25 | 10,2 | -0,03 | -1,58 | 14,99 | 0,24 |
| Rural | -0,27 | -16,65 | -0,03 | -2,02 | 0,07 | 4,16 | -0,26 | -16,19 | 0,05 | 4,53 | -25,95 | -0,27 |
| UU de moins de 100000 hbts | -0,06 | -4,09 | -0,03 | -2,66 | 0,02 | 1,18 | -0,06 | -4,36 | 0,02 | 2,47 | -6,64 | -0,06 |
| UU de plus de 100000 hbts | 0,18 | 12,85 | 0,02 | 1,47 | -0,06 | -4,17 | 0,17 | 12,25 | -0,06 | -5,63 | 20,42 | 0,18 |
| Niveau de diplôme | | | | | | | | | | | | |
| Aucun diplôme ou CEP | -0,06 | -3,5 | 0,18 | 11,17 | 0,27 | 14,55 | 0,04 | 2,34 | 0,28 | 23,7 | 1,85 | 0,02 |
| Bac | 0,06 | 3,09 | -0,03 | -1,56 | -0,07 | -3,05 | 0,04 | 2,03 | -0,07 | -4,9 | 4,07 | 0,05 |
| Brevet | -0,05 | -1,5 | -0,07 | -2,36 | -0,08 | -2,38 | -0,08 | -2,56 | -0,11 | -5,02 | -5,92 | -0,12 |
| CAP BEP | 0,09 | 5,43 | -0,03 | -1,65 | -0,01 | -0,61 | 0,08 | 5,07 | 0,01 | 0,91 | 8,84 | 0,09 |
| Bac+2 | 0,08 | 3,34 | -0,07 | -2,96 | -0,17 | -6,44 | 0,03 | 1,12 | -0,18 | -10,26 | 3,15 | 0,05 |
| Diplômé du supérieur | -0,17 | -7,58 | -0,1 | -4,88 | -0,12 | -5,21 | -0,22 | -9,97 | -0,17 | -10,71 | -15,45 | -0,22 |

4 Quelques matrices de corrélations

Tableau 6: Modalités d'insatisfaction

| | Global | | | | | | | |
|--------------------------------------|-----------------------|--------------------------|------------------|--------------------------|-------------------|---------|----------------------|--|
| | Pol. pas du tt satisf | Just. pas du tout satisf | Inséc au dom svt | Inséc ds le quartier svt | Prés. inex nécess | pol. et | Action pas eff du tt | |
| Police pas du tout satisf | 1 | 0,2559 | 0,0732 | 0,0958 | 0,155 | | 0,2732 | |
| Justice pas du tout satisf | 0,2559 | 1 | 0,0858 | 0,1051 | 0,1121 | | 0,1609 | |
| Insécurité au domicile souvent | 0,0732 | 0,0858 | 1 | 0,4674 | 0,0797 | | 0,1057 | |
| Insécurité dans le quartier souvent | 0,0958 | 0,1051 | 0,4674 | 1 | 0,1314 | | 0,1759 | |
| Présence de la police et necess | 0,155 | 0,1121 | 0,0797 | 0,1314 | 1 | | 0,2818 | |
| Action de la police efficace du tout | 0,2732 | 0,1609 | 0,1057 | 0,1759 | 0,2818 | | 1 | |
| | | | | | | | | |
| | FAF | | | | | | | |
| | Pol. pas du tt satisf | Just. pas du tout satisf | Inséc au dom svt | Inséc ds le quartier svt | Prés. inex nécess | pol. et | Action pas eff du tt | |
| Police pas du tout satisf | 1 | 0,252 | 0,053 | 0,0613 | 0,1352 | | 0,2273 | |
| Justice pas du tout satisf | 0,252 | 1 | 0,058 | 0,0475 | 0,0941 | | 0,1461 | |
| Insécurité au domicile souvent | 0,053 | 0,058 | 1 | 0,4667 | 0,0858 | | 0,0899 | |
| Insécurité dans le quartier souvent | 0,0613 | 0,0475 | 0,4667 | 1 | 0,1016 | | 0,1661 | |
| Présence de la police et necess | 0,1352 | 0,0941 | 0,0858 | 0,1016 | 1 | | 0,2704 | |
| Action de la police efficace du tout | 0,2273 | 0,1461 | 0,0899 | 0,1661 | 0,2704 | | 1 | |

| | Internet | | | | | | |
|--|-------------------------|--------------------------|------------------|--------------------------|-------------------|---------|---------------------------|
| | Pol. pas du tout satisf | Just. pas du tout satisf | Inséc au dom svt | Inséc ds le quartier svt | Prés. inex nécess | pol. et | Action pol. pas eff du tt |
| Police pas du tout satisf | 1 | 0,2766 | 0,1438 | 0,1621 | 0,1836 | | 0,3654 |
| Justice pas du tout satisf | 0,2766 | 1 | 0,1234 | 0,1619 | 0,1516 | | 0,1883 |
| Insécurité au domicile sou-vent | 0,1438 | 0,1234 | 1 | 0,4356 | 0,077 | | 0,1413 |
| Insécurité dans le quartier souvent | 0,1621 | 0,1619 | 0,4356 | 1 | 0,1671 | | 0,1892 |
| Présence de la police inex et necess | 0,1836 | 0,1516 | 0,077 | 0,1671 | 1 | | 0,2821 |
| Action de la police pas efficace du tout | 0,3654 | 0,1883 | 0,1413 | 0,1892 | 0,2821 | | 1 |

| | Papier | | | | | | |
|--|-------------------------|--------------------------|------------------|--------------------------|-------------------|---------|---------------------------|
| | Pol. pas du tout satisf | Just. pas du tout satisf | Inséc au dom svt | Inséc ds le quartier svt | Prés. inex nécess | pol. et | Action pol. pas eff du tt |
| Police pas du tout satisf | 1 | 0,2544 | 0,0625 | 0,1076 | 0,1763 | | 0,2849 |
| Justice pas du tout satisf | 0,2544 | 1 | 0,0909 | 0,1237 | 0,1208 | | 0,1522 |
| Insécurité au domicile sou-vent | 0,0625 | 0,0909 | 1 | 0,4867 | 0,0809 | | 0,0995 |
| Insécurité dans le quartier souvent | 0,1076 | 0,1237 | 0,4867 | 1 | 0,1638 | | 0,1766 |
| Présence de la police inex et necess | 0,1763 | 0,1208 | 0,0809 | 0,1638 | 1 | | 0,3154 |
| Action de la police pas efficace du tout | 0,2849 | 0,1522 | 0,0995 | 0,1766 | 0,3154 | | 1 |