

L'analyse factorielle multiple duale pour l'analyse de la collecte multimode

Marine GUILLERM et Tiaray RAZAFINDRANOVA

Insee, DMCSI

1^{er} avril 2015

Sommaire

- 1 Problématique
- 2 L'analyse factorielle multiple duale
- 3 Application à l'analyse du multimode

Les effets de mode

- Contexte de développement de la collecte par internet.
- Les modes de collecte ne sont en général pas neutres sur les réponses des enquêtés : biais de désirabilité sociale, satisficing...
- Un dispositif expérimental en 2013 : “Vols, violence et sécurité” (VVS) conçu à des fins de comparaison avec l'enquête de victimation “Cadre de vie et sécurité” (CVS).
 - VVS : interrogation en auto-administré, sur internet ou par questionnaire papier.
 - CVS : via un enquêteur en face à face.

Le dispositif expérimental “Vols, violence et sécurité” (VVS)

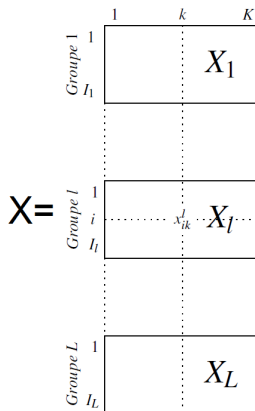
- Des premiers résultats : Des taux de victimation plus élevés dans VVS par rapport à CVS (Razafindranovona et al., 2014).
- Une autre optique : les liaisons entre variables :
 - Quelles sont les modalités prises simultanément par un grand nombre d'individus ?
 - Ces liaisons diffèrent-elles d'un mode de collecte à un autre ?
→ susceptible de révéler des différences de comportement de réponse.
- Variables d'intérêt ici : les questions d'opinion.

Sommaire

- 1 Problématique
- 2 L'analyse factorielle multiple duale
- 3 Application à l'analyse du multimode

L'AFMD : pour quelles données et quelle problématique ? (1/2)

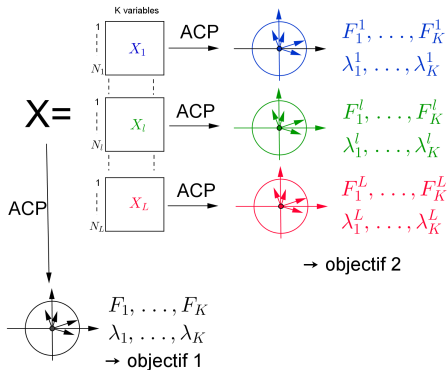
- Analyse de multi-tableaux de données individus \times variables, où les individus sont structurés en groupes \rightarrow groupés par dates d'observation, par pays, par mode de collecte...
- Méthode proposée par Lê et Pagès (2010). Extension aux variables catégorielles : Abascal et al. (2012)



L'AFMD : pour quelles données et quelle problématique ? (2/2)

- Ces multi-tableaux offrent plusieurs points de vue.
- Les objectifs de l'AFMD :
 - Objectif 1 : Analyses des liaisons entre variables au niveau global (sur l'ensemble des individus).
 - Objectif 2 : Analyses des liaisons entre variables aux niveaux partiels, ie au sein de chaque groupe d'individus.
 - Objectif 3 : Comparaison des groupes d'individus entre eux.
→ les corrélations entre variables sont-elles différentes d'un groupe d'individus à un autre ?

Les limites des analyses factorielles classiques



- Mener une ACP sur chaque sous tableaux de données
- Les analyses sont indépendantes et donc difficilement comparables entre elles.
→ Ne répond pas à l'objectif 3 (comparaison des groupes d'individus entre eux).

Principe de l'AFMD

Mener une ACP sur le tableau, noté Z , des **données centrées et réduites par groupe** :

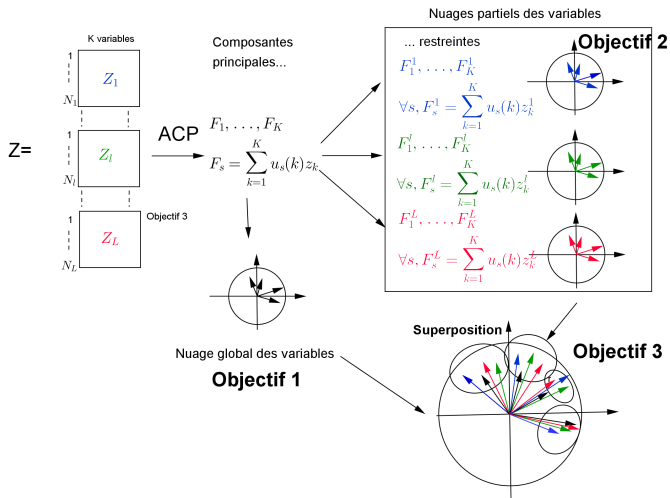
Les caractéristiques de cette transformation :

- Ne modifie pas les corrélations entre les variables à l'intérieur des groupes d'individus.
- Équilibre la contribution des différents groupes à la construction des axes factoriels.
→ Compromis entre les différents groupes d'individus :
Le programme de recherche des axes factoriels :

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

- Construit des composantes principales qui ont la même interprétation d'un groupe d'individus à un autre.

Schéma général de l'AFMD



Analyse de l'inter-groupe

Les corrélations entre variables différentes d'un groupe d'individus à un autre (objectif 3) ?

- Apport essentiel de l'AFMD.
- Une analyse plus fine : superposition des nuages global et partiels
→ analyse de la trajectoire des points homologues (associés à la même variable mais à des groupes d'individus différents).
- Une représentation globale de ces différences : le nuage des groupes d'individus.
 - Deux groupes sont d'autant plus proches que les corrélations entre variables sont similaires d'un groupe à un autre.
 - Analyse axe par axe. Des coordonnées différentes révèlent des différences de corrélations, en particulier pour les variables les plus corrélées à l'axe.

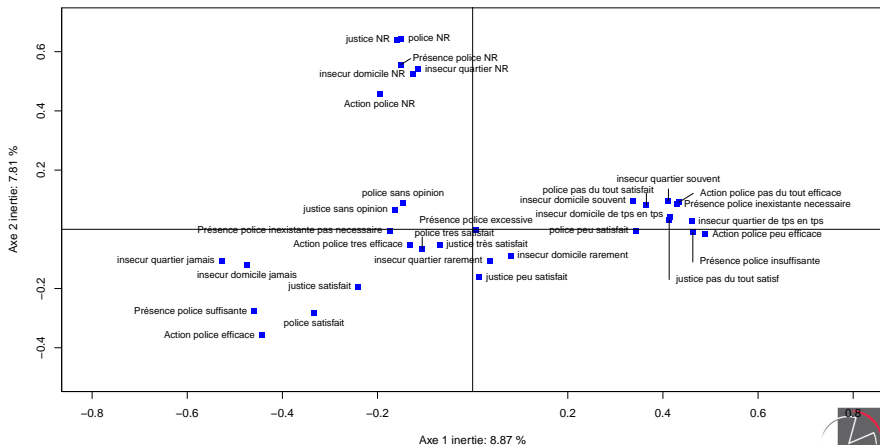
Sommaire

- 1 Problématique
- 2 L'analyse factorielle multiple duale
- 3 Application à l'analyse du multimode

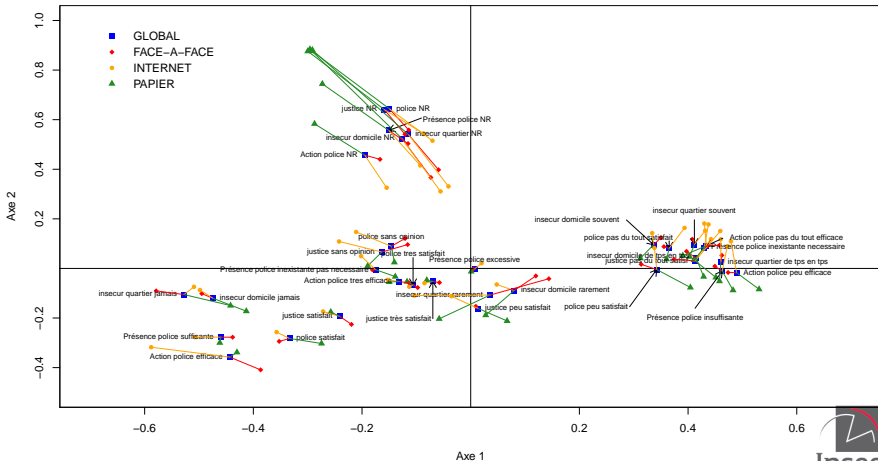
Les données

- Les individus groupés selon le mode de collecte
- Les variables : les questions d'opinion
 - opinion générale sur l'action de la police ou de la gendarmerie ;
→ très satisfaisante / satisfaisante / peu satisfaisante / pas du tout satisfaisante / pas d'opinion / NR
 - opinion générale sur l'action de la justice et des tribunaux dans le traitement de la délinquance ;
→ très satisfaisante / satisfaisante / peu satisfaisante / pas du tout satisfaisante / pas d'opinion / NR
 - sentiment d'insécurité au domicile ;
→ souvent / de temps en temps / rarement / jamais / NR
 - sentiment d'insécurité dans le quartier (ou village) ;
→ souvent / de temps en temps / rarement / jamais / NR
 - avis sur la présence de la police ou de la gendarmerie dans le quartier (ou village) ;
→ suffisante / excessive / insuffisante / inexistante alors qu'elle serait nécessaire / inexistante mais elle n'est pas nécessaire / NR
 - avis sur l'efficacité de la police ou de la gendarmerie dans le quartier (ou village) ;
→ très efficace / efficace / peu efficace / pas efficace du tout / NR

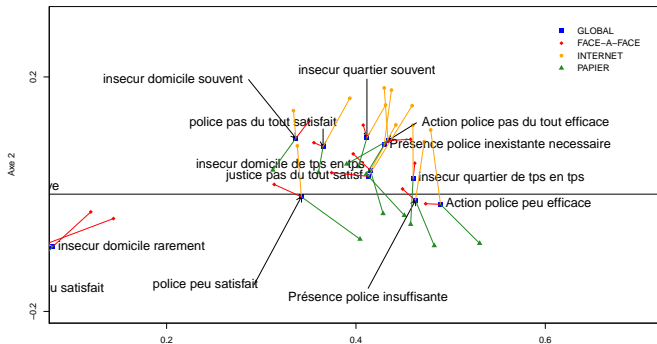
Les résultats, nuage global des variables



La superposition des nuages global et partiels - axe 1 et 2

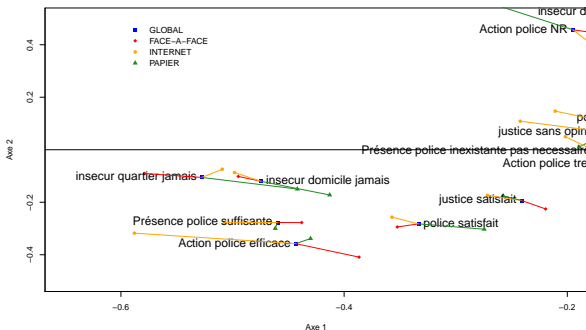


La superposition des nuages global et partiels - axe 1 et 2 - zoom à droite



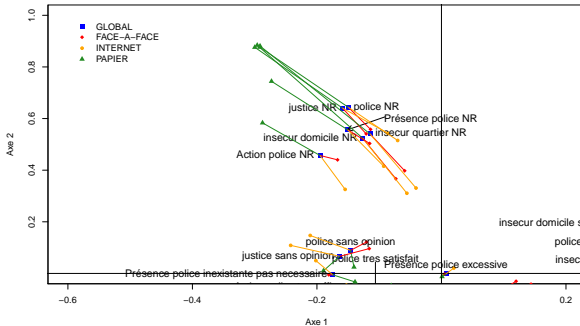
Graphique: Superposition des nuages global et partiels des modalités

La superposition des nuages global et partiels - axe 1 et 2 - zoom à gauche



Graphique: Superposition des nuages global et partiels des modalités

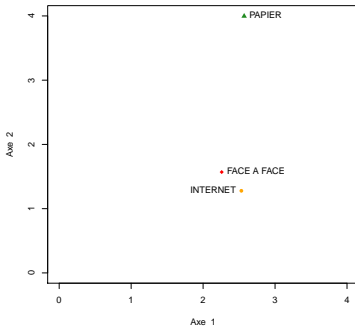
La superposition des nuages global et partiels - axe 1 et 2 - zoom en haut



Graphique: Superposition des nuages global et partiels des modalités

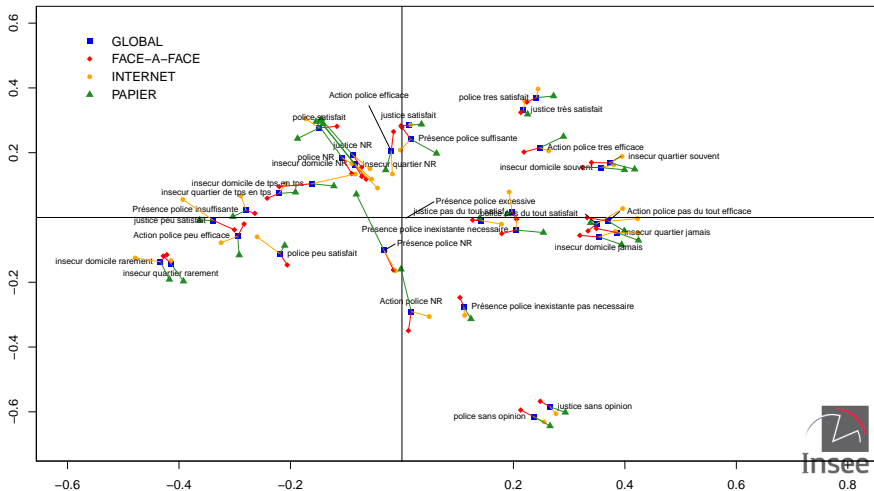


Les groupes d'individus

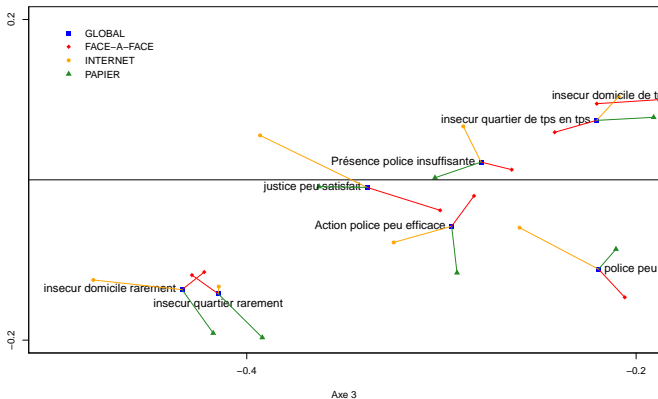


Graphique: Nuage des groupes d'individus

La superposition des nuages global et partiels - axe 3 et 4

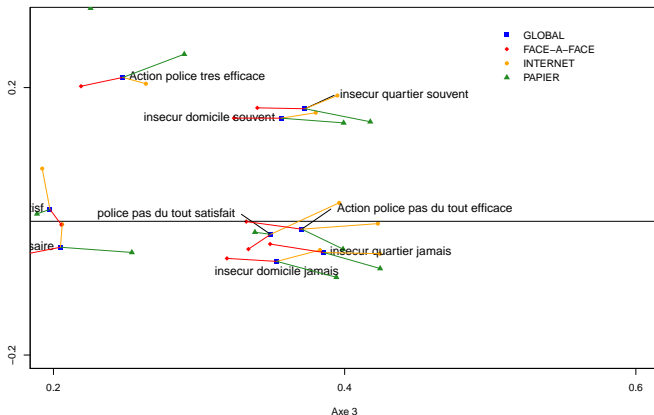


La superposition des nuages global et partiels - axe 3 et 4 - zoom à gauche



Graphique: Superposition des nuages global et partiels des modalités

La superposition des nuages global et partiels - axe 3 et 4 - zoom à droite



Graphique: Superposition des nuages global et partiels des modalités

Conclusion

- Effets de mode mais aussi sûrement des effets de sélection
- Contrôle de la sélection :
 - Regroupement des modes de collecte internet et papier
 - Matching pour contrôler de la sélection (sur variables socio démographiques)
- Modifie peu les conclusions.
- Sélection sur variables inobservables

Merci de votre attention !

Bibliographie

- Abascal, E., V. Díaz de Rada, I. García Lautre et I. Landaluce. 2012, «A comparison of two modes of data collection using multidimensional analysis», *Revista Internacional de Sociología (RIS)*, vol. 70, n° 3, p. 511–532.
- Lê, S. et J. Pagès. 2010, «DMFA : dual multiple factor analysis», *Communications in statistics. Theory and methods*, vol. 39, n° 3, p. 483–492.
- Razafindranovona, T., B. Dietsch, C. Burrigand 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*.