

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

Les organismes producteurs d'information statistique envisagent de développer la collecte de données auprès des ménages par internet notamment en raison de ses avantages en termes de coût financier et temporel. Que ce mode de collecte soit utilisé en complément d'autres modes (en face-à-face avec un enquêteur, de manière auto-administrée par questionnaire papier) ou comme mode de collecte principal, la question de la qualité des données ainsi recueillies se pose. En effet, on observe en général des différences entre les modes de collecte. Elles peuvent s'expliquer par des effets de mode et/ou des effets de sélection, parfois confondus (Vannieuwenhuyze et al., 2014). Les effets de sélection sont liés aux différences de profils entre répondants. On parle d'effets de mode quand un individu répond différemment à une question selon le mode de collecte : biais de désirabilité sociale en présence d'un enquêteur, *satisficing*, etc.

Le dispositif "Vols, violences et sécurité" (VVS) mené à titre expérimental par l'Insee interroge les individus par internet ou par questionnaire papier sur le thème de la victimation. Il offre la possibilité de mener une analyse de la collecte multimode, en comparant les données ainsi recueillies à celles de l'enquête "Cadre de vie et sécurité" (CVS) réalisée chaque année en face à face. On s'intéresse ici aux réponses aux questions d'opinion en matière de sécurité (sentiment d'insécurité à domicile, dans le quartier, rôle de la police, de la justice...). Ce type de questions est le plus susceptible d'être affecté par des effets de mode.

Il est classique dans une analyse du multimode de s'intéresser aux différentes variables de manière séparée. On montre alors que l'occurrence d'une certaine modalité est plus fréquente dans un mode plutôt qu'un autre. Ici, une autre optique est proposée en s'intéressant aux liaisons entre variables et en analysant comment elles évoluent quand on passe d'un mode de collecte à un autre. Pour des variables catégorielles, il s'agit donc de mettre en évidence des ensembles de modalités renseignées simultanément par un grand nombre d'individus ou au contraire rarement renseignées conjointement, et à analyser les différences de liaisons constatées selon le mode de collecte.

L'analyse factorielle multiple duale (AFMD) proposée par Lê et Pagès (2010) est particulièrement adaptée à cette analyse. Elle répond à l'analyse factorielle de multi-tableaux se présentant sous la forme d'un empilement de tableaux de données. Cette configuration se retrouve quand on dispose de plusieurs tableaux de données individus x variables (ici un tableau pour chaque mode de collecte) dont les variables sont communes. L'AFMD construit un unique jeu de composantes principales et assure que la contribution des différents groupes d'individus dans sa construction est équilibrée. Comme dans une analyse factorielle classique, ces composantes principales permettent de représenter graphiquement les liaisons entre variables (nuage des variables). 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. Ces nuages peuvent être superposés. Chaque variable est ainsi représentée par quatre points (un point global et trois points partiels associés chacun à un mode de collecte). La trajectoire de ces points homologues permet d'analyser les différences de corrélations entre variables d'un groupe d'individus à un autre (étude de l'inter-groupe).

Une AFMD menée sur les données du dispositif CVS/VVS permet de mettre en évidence des différences de comportements de réponse selon le mode de collecte. Les premiers résultats montrent un premier axe factoriel opposant les individus globalement satisfaits des institutions (police, gendarmerie, justice) et se sentant en sécurité à leur domicile, aux individus globalement insatisfaits et exprimant un sentiment d'insécurité. Les opinions exprimées apparaissent plus tranchées quand l'individu répond de manière auto-administrée (par internet ou par papier) par rapport à une interrogation en face à face. Sur le deuxième axe, la non réponse aux différentes questions apparaît

liée : ne pas répondre à une question va de pair avec la non réponse aux autres questions. Cette liaison est plus forte dans une interrogation sur questionnaire papier par rapport à une interrogation en face à face.

Ces résultats seront complétés en les croisant avec des variables sociodémographiques et des informations sur les éventuels types de violence ou vol dont les individus auraient pu être victimes.

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, no 2, p. 415–428.

Lê, S. et J. Pagès. 2010, «DMFA : dual multiple factor analysis», *Communications in statistics. Theory and methods*, vol. 39, no 3, p. 483–492.

Vannieuwenhuyze, J., G. Loosveldt et G.Molenberghs. 2014, «Evaluating mode effects in mixed-mode survey data using covariate adjustment models», *Journal of Official Statistics*, vol. 30, no 1, p. 1-21.