

## L'invariance structurale du test de repérage de l'usage problématique de cannabis CAST dans 13 pays européens: application des méthodes d'ACP et PLS multigroupes

Stéphane Legleye<sup>1,2,3</sup>, Aida Eslami<sup>4</sup>, Stéphanie Bougeard<sup>5</sup>

1. Institut national des études démographiques (INED), Paris
2. Inserm, U669, Paris
3. Univ Paris-Sud and Univ Paris Descartes, UMR-S0669, Paris)
4. LUNAM University, ONIRIS, USC Sensometrics and Chemometrics Laboratory, Nantes
5. Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (Anses), Ploufragan

**Domaine** : Analyse des données

### Résumé

Le recours à des données d'enquêtes menées au sein de plusieurs populations est courant en épidémiologie et en statistique publique. Toutefois, comparer les résultats obtenus sur différentes populations nécessite de prendre quelques précautions. L'analyse des indicateurs simples peut se faire par le recours à des modélisations multiniveaux, qui tiennent compte du caractère groupé de la collecte des données. Mais dans de nombreux cas, il est nécessaire de considérer non pas une mais un ensemble de mesures liées, comme par exemple les questions d'un test décrivant la qualité de vie. Ces données multidimensionnelles ne peuvent être traitées par des analyses de régressions classiques ou multiniveaux. Il importe par exemple de s'assurer que le test mesure bien la même chose de la même façon dans toutes les populations : on parle d'invariance structurale, métrique et scalaire [1]. Toutefois, les méthodes de modélisation et de test sont complexes et ne permettent pas aisément d'expliquer les différences observées d'une population à l'autre. Nous présentons ici deux méthodes d'analyse des données multitableaux qui tiennent compte de la structure en groupes (c'est-à-dire ici en populations définies a priori) des individus statistiques. Il s'agit de l'ACP multigroupe et de la PLS multigroupe. Ces deux méthodes offrent un cadre descriptif et explicatif intuitif et graphique des données fondé sur des considérations géométriques.

L'ACP multigroupe (mgACP) vise à comparer la structure d'un ensemble de variables X mesurées dans M groupes d'individus définis a priori. Elle établit un compromis entre une structure globale commune à tous les groupes et les structures propres à chacun. Il s'agit d'explorer et de décrire l'invariance structurale du tableau X au sein des M groupes. Les déviations des groupes à la structure commune sont quantifiées et peuvent être représentées graphiquement.

La PLS multigroupe (mgPLS) est une méthode explicative de la structure d'un tableau de variables Y par un bloc (ou tableau) de variables X mesurées dans les mêmes M groupes d'individus définis a priori. Cette méthode établit également un compromis entre une structure explicative globale de Y par les X et les structures équivalentes propres à chaque groupe. Les parts de variances expliquées par les variables X sont quantifiables et les écarts à la structure commune explicative sont quantifiables et représentables graphiquement.

Nous présentons chaque méthode puis leur application à un jeu de données issu de l'enquête européenne ESPAD (European school project on alcohol and other drugs), enquête scolaire auprès des 15-16 ans réalisée tous les 4 ans dans une quarantaine de pays européens ([www.espad.org](http://www.espad.org)). Pour notre propos, nous utilisons les données relatives à un test de repérage de l'usage problématique de cannabis, le CAST -Cannabis abuse screening test-[2-4] proposé comme module optionnel depuis 2007 mais pour lequel aucune étude de validité internationale n'a été faite. Le CAST comporte 6 questions simples dont la réponse est demandée sur une échelle de fréquence de Likert (de « jamais » à « très souvent »). Notre analyse porte sur 5204 élèves de 15-16 ans des treize pays ayant administré le CAST en 2011.

La mgACP et la mgPLS ont été mis au point par Aida Eslamiet ses collègues [5-7] et sont accessibles sur le logiciel R.

Les résultats montrent une très bonne invariance structurelle du CAST sauf dans trois pays dont la mgPLS permet de rendre compte de la situation particulière.

## Références

1. Matsumoto, D. and F.J. Van de Vijver, eds. *Cross-Cultural Research Methods in Psychology*. 2011, Cambridge University Press: New York. 392.
2. Legleye, S., et al., *A validation of the Cannabis Abuse Screening Test (CAST) using a latent class analysis of the DSM-IV among adolescents*. *Int J Methods Psychiatr Res*, 2013. **22**(1): p. 16-26.
3. Pabst, A., L. Kraus, and D. Piontek, *The Cannabis Abuse Screening Test (CAST): Examining the prevalence of cannabis-related problems among adolescents in 13 European countries*, in *The 2011 ESPAD Report. Substance Use Among Students in 36 European Countries*, B. Hibell, et al., Editors. 2012, The Swedish Council for Information on Alcohol and Other Drugs (CAN): Stockholm, Sweden. p. 158-163.
4. Legleye, S., D. Piontek, and L. Kraus, *Psychometric properties of the Cannabis Abuse Screening Test (CAST) in a French sample of adolescents*. *Drug Alcohol Depend*, 2011. **113**(2-3): p. 229-35.
5. Eslami, A., et al., *General overview of methods of analysis of multi-group datasets*. *Revue des Nouvelles Technologies de l'Information*, 2013(25): p. 108-23.
6. Eslami, A., et al., *Two-block multi-group data analysis. Application to epidemiology*, in *New perspectives in Partial Least Squares and Related Methods*, G. Russolillo, Editor. 2014, Springer Verlag.
7. Eslami, A., et al., *Algorithms for multi-group PLS*. *Journal of Chemometrics*, 2014. **28**(3): p. 192-201.