

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 Legleye¹, Aida Eslami², Stéphanie Bougeard³

1. INED; 2. Inserm, U1178; Univ Paris-Sud; Univ Paris Descartes, UMR-S1178

2. Lunam University

3. Anses

Journées de méthodologie statistique, 2015



Problématique – contexte (1)

1. Enquêtes internationales /dans des sous-populations différentes

→ Comment assurer la comparabilité des mesures d'une même échelle?

Traditionnellement, analyses factorielles complexes (exploratoire, confirmatoire)

Premier stade : invariance structurelle ou configurale

→ Le test a-t-il la même structure ?

Puis : invariance métrique et scalaire

Problématique – contexte (2)

2. Le phénomène étudié n'est pas réductible à un unique indicateur:

→ Ses différentes dimensions sont à considérer simultanément, par ex. pour l'alcool: fréquence de consommation, nombre de verres bus un jour moyen, fréquence des ivresses

3. Les données sont groupées ou hiérarchisées

→ on cherche à expliquer le phénomène

→ l'analyse de régression classique est inadaptée: modèles hiérarchiques ou mixtes, multiniveaux

Présentation des données

- Base de N individus répartis en M groupes (lignes)
- Tableau X de Q variables
 - Centrées par colonne par groupe ou sur l'ensemble
- Tableau Y de P variables
 - Centrées par colonne par groupe ou sur l'ensemble
- On cherche à décrire Y (mgACP)
- On cherche à décrire X et Y et expliquer Y par X (mgPLS)

Contraintes:

- tenir compte de la structure en M groupe
- s'affranchir de l'effet groupe (analyse intra-groupe et pas inter-groupe)

L'ACP multigroupes



1. Objectifs

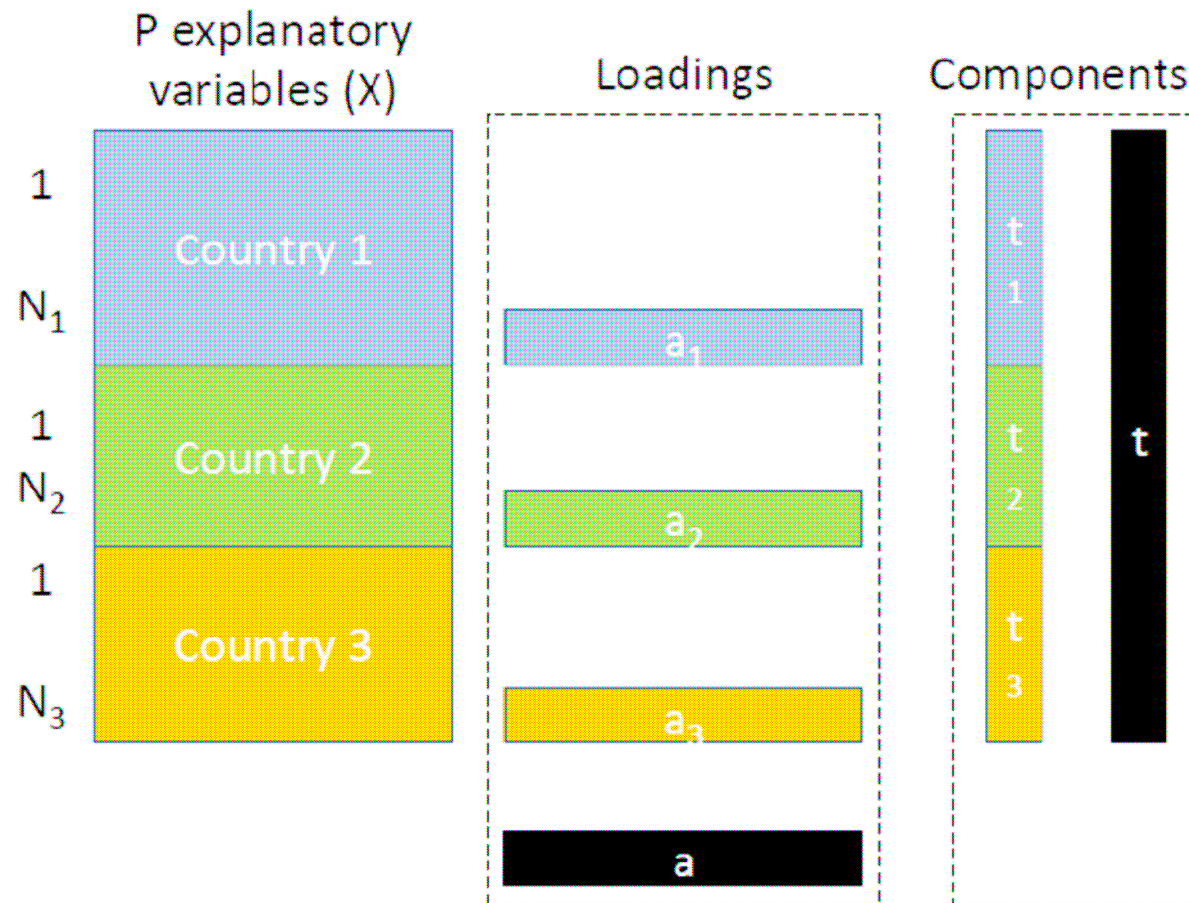
BUT: description de Y

→ trouver un vecteur de loadings communs expliquant les P variables, commun à tous les individus, \mathbf{a} , tenant compte du caractère groupé

\mathbf{a} est la concaténation des vecteurs de loadings partiels propres à chaque groupe m , \mathbf{a}_m , expliquant les P variables pour le groupe m .

L'ACP multigroupes

2. Schéma d'analyse



L'ACP multigroupes



3. Algorithme et contraintes

- Le vecteur $a^{(1)}$ est trouvé sous la contrainte:

$$\text{Max } \sum_{m=1}^M \langle a_m^{(1)}, a^{(1)} \rangle^2, \quad a_m^{(1)} = X'_m t_m^{(1)}$$

Avec : $\| t_m^{(1)} \| = \| a^{(1)} \| = 1$.

- $a^{(1)}$ est le vecteur propre de $\sum X'_m X_m$.
- $t_m^{(1)}$ est la composante pour chaque groupe : $t_m^{(1)} = X_m a^{(1)}$
- $a_m^{(1)}$ est le loading du groupe m : $a_m^{(1)} = X'_m t_m^{(1)}$
- $t^{(1)}$ est la concaténation des vecteurs $t_1^{(1)}$ à $t_M^{(1)}$ (mêmes loadings)

L'ACP multigroupes (3)



3. Algorithme et contraintes

- **Processus itératif** : on procède d'abord pour la première dimension $a^{(1)}$
 - On trouve les autres paramètres associés aux dimensions de rang supérieur (jusqu'à $P=6$ ici) par déflation avec les contraintes :
 - Les loadings communs $a^{(1)}, a^{(2)}, \dots, a^{(6)}$ sont orthogonaux
 - Les composantes communes $t^{(1)}, t^{(2)}, \dots, t^{(6)}$ sont orthogonales
- Indice de similarité du groupe/structure commune en (i):

PLS multigroupes



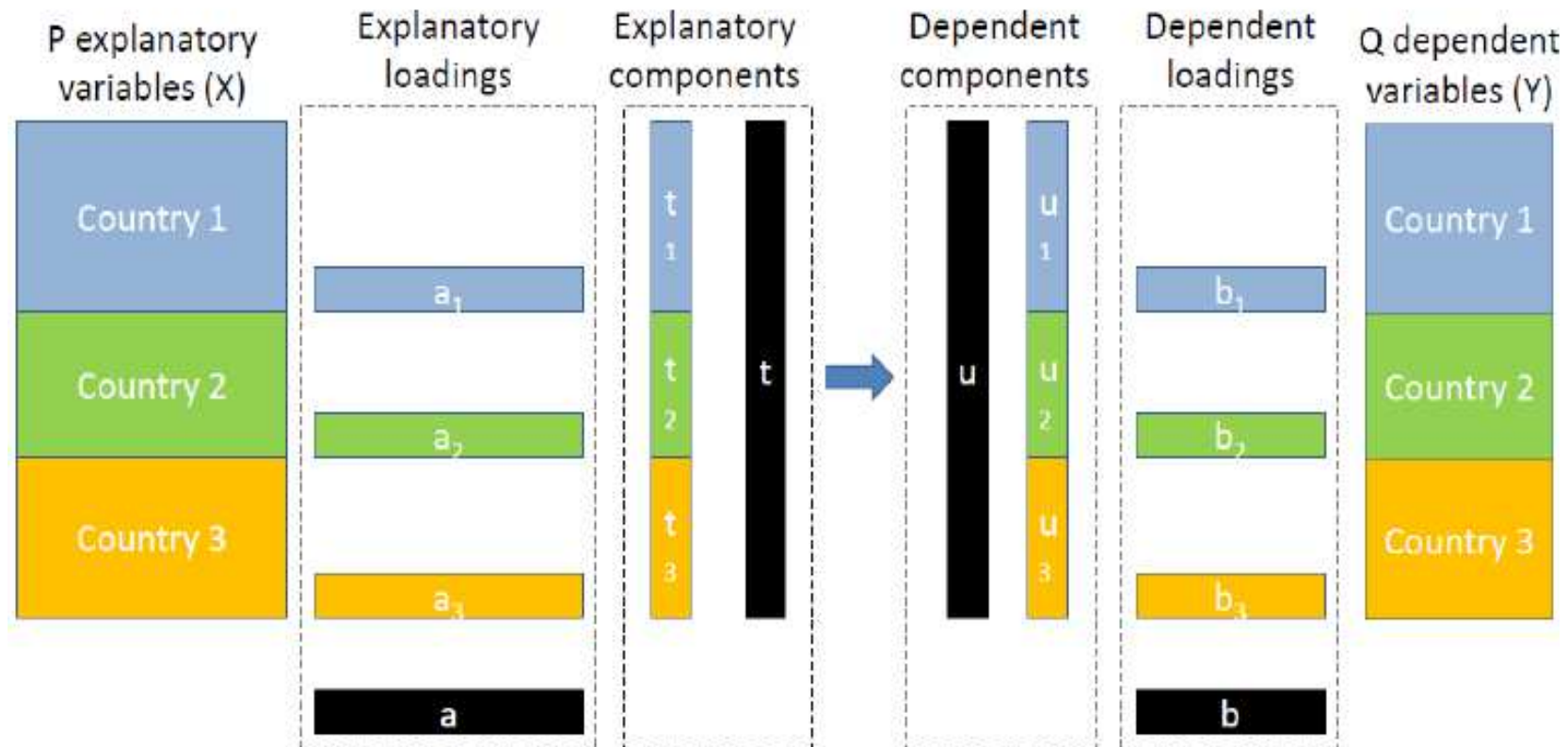
1. Objectifs

Etudier la communauté de structure entre X et Y et prédire Y par X en tenant compte des groupes

→ Trouver des vecteurs de loadings a dans X et b dans Y, tels que leurs variables latentes $t_m = X_m a$, $u_m = Y_m b$ soient les plus liées:

PLS multigroupes

2. Schéma d'analyse



PLS multigroupes

3. Critère de liaison (1)

● Critère de liaison pour la 1^{ère} composante:

$$\text{Max } \sum_{m=1}^M \text{Cov}(X_m a^{(1)}, Y_m b^{(1)}) ,$$

Avec : $\| a^{(1)} \| = \| b^{(1)} \| = 1$

$$t_m^{(1)} = X_m a^{(1)} \text{ et } u_m^{(1)} = Y_m b^{(1)}$$

Les composantes communes sont:

- $t^{(1)}$, concaténation verticale des $t_m^{(1)}$
- $u^{(1)}$, concaténation verticale des $u_m^{(1)}$

PLS multigroupes



3. Critère de liaison (2): extension de la mgACP

Le critère est une extension de celui de l'ACP multigroupes avec :

- t_m composantes des groupes dans l'espace Y;
- $t^{(i)}$ orthogonales entre elles
- Indice de similarité du groupe/structure commune en (i):
 - En X: $|\text{Cos}(a^{(i)}, a_m^{(i)})|$: on peut cumuler sur plusieurs (i)
 - En Y: $|\text{Cos}(b^{(i)}, b_m^{(i)})|$: on peut cumuler sur plusieurs (i)

Les données



1. L'enquête ESPAD

- European survey project on alcohol and other drugs (ESPAD)
- Enquête scolaire quadriennale sur les drogues (licites ou non)
- 15-16 ans, 36 pays en 2011
- 13 pays ont passé le questionnaire CAST (Cannabis abuse screening test) en 2011

- N=5204 élèves de 13 pays, fumeurs de cannabis dans l'année.

Les données

2. CAST



6 questions (jamais, rarement, de temps en temps, assez souvent, très souvent)

- Avez-vous fumé du cannabis avant midi ?
- avez-vous fumé du cannabis lorsque vous étiez seul ?
- Avez-vous eu des problèmes de mémoire quand vous fumiez du cannabis ?
- Des amis, des proches ou des membres de votre famille vous ont-ils dit que vous devriez réduire votre consommation de cannabis ?
- Avez-vous essayé de réduire ou d'arrêter votre consommation de cannabis sans y arriver ?
- Avez-vous eu des problèmes à cause de votre consommation de cannabis (dispute, bagarre, accident, mauvais résultat à l'école...)?

Les données

3. Les données explicatives



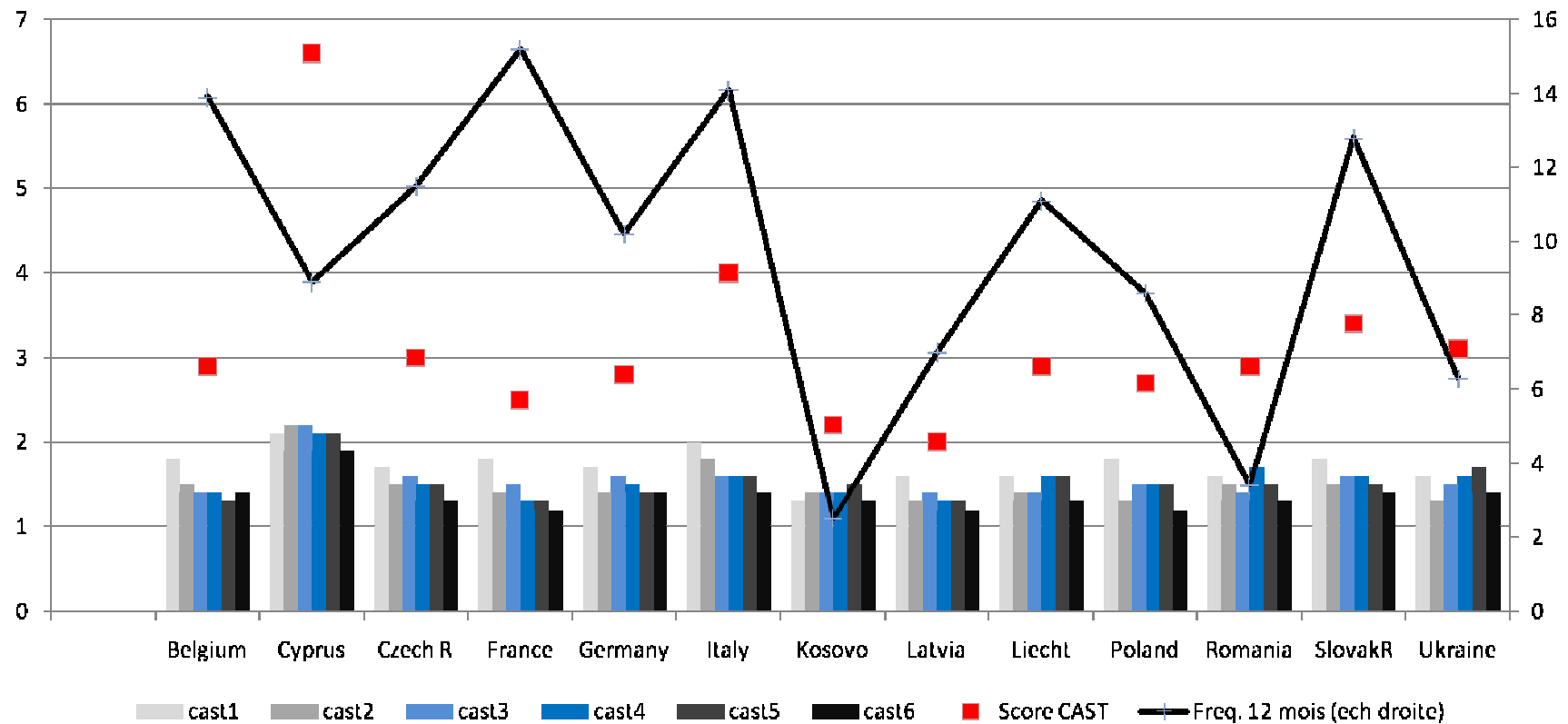
11 variables ordonnées (4 ou 7 valeurs)

- Fréquence de consommation de cannabis, alcool, 12 mois et 30 derniers jours
- Fréquence des ivresses (vie et 30 derniers jours)
- Nombre de cigarettes de tabac / jour
- Âge au premier usage de cannabis
- Proportion d'amis fumant du cannabis
- Risque perçu associé à la consommation régulière de cannabis
- Accessibilité perçue du cannabis

Résultats



1. Moy. items & du score CAST, fréq. d'usage de cannabis

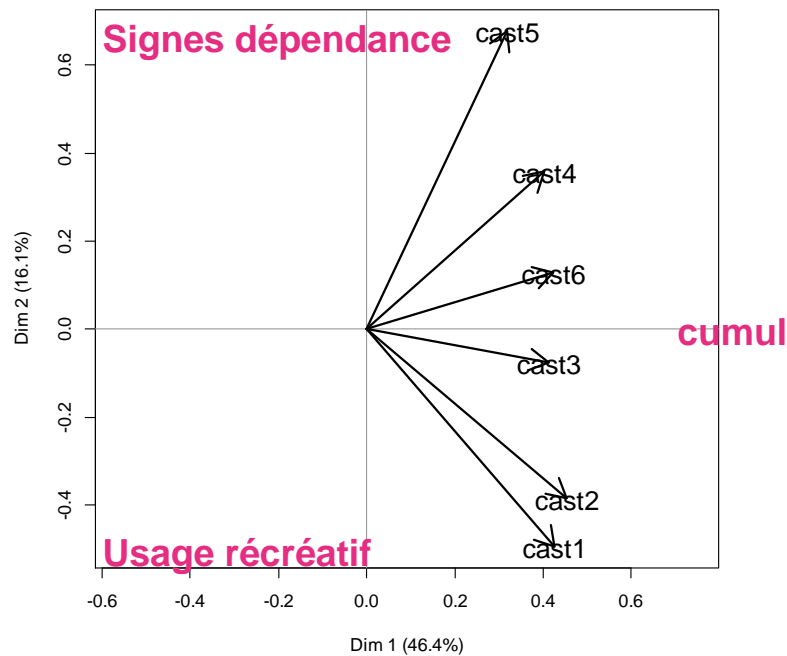
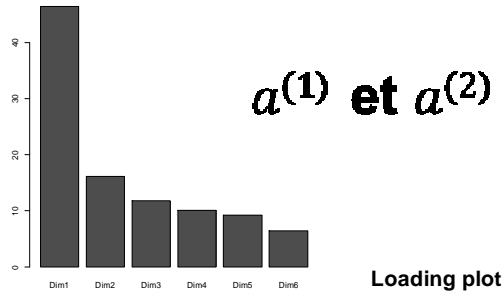


Variance intergroupes : 2,5%: sera éliminée par la suite.

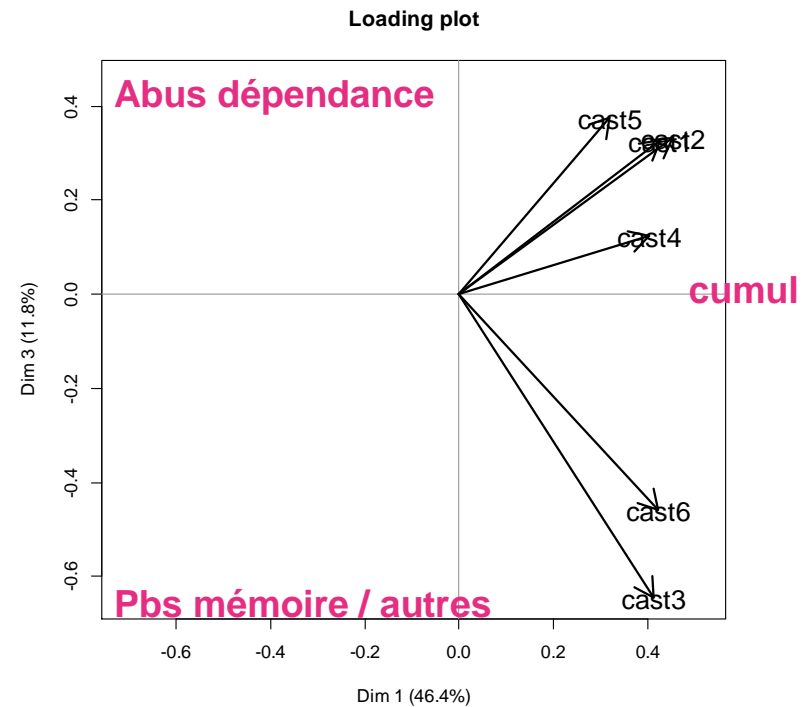
Résultats

2. ACP multigroupes : loadings plans 1-2, 1-3

$a^{(1)}$ et $a^{(2)}$ représentent 46,4% et 16,1% de la variance totale



a : mgACP axes 1-2

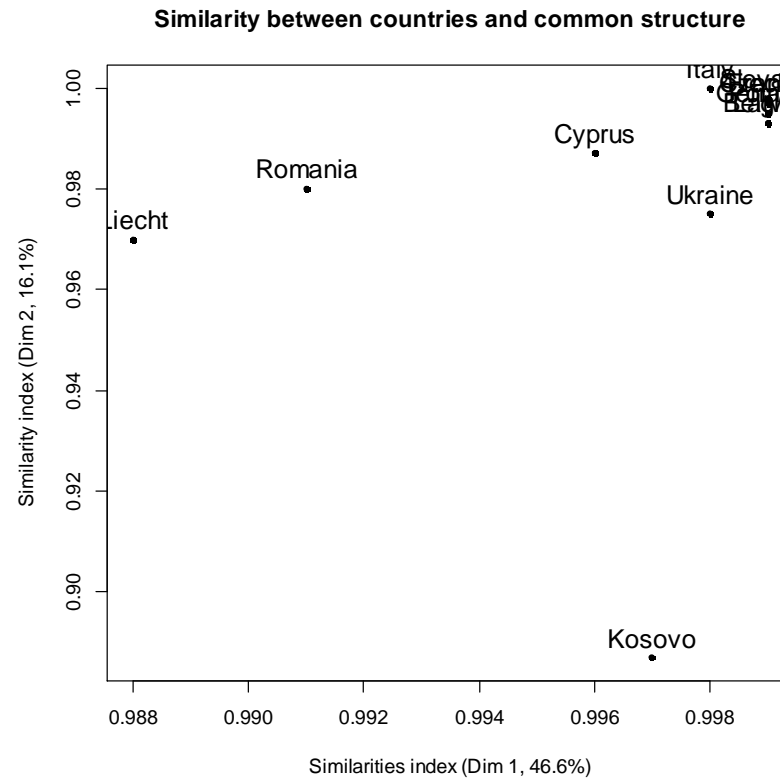


b : mgACP axes 1-3

Résultats



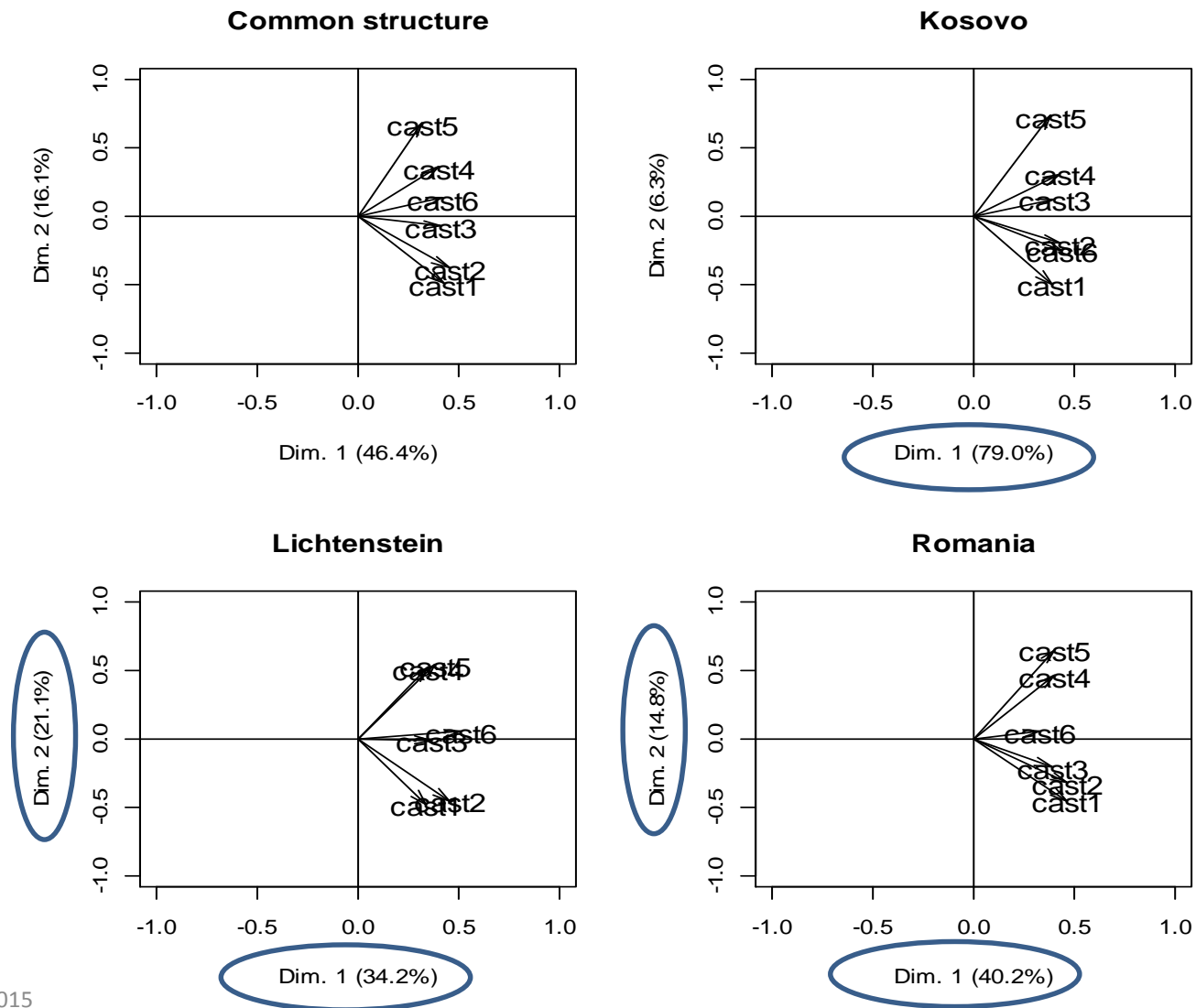
3. ACP multigroupes: similarités / $a^{(1)}$ et $a^{(2)}$



Similarité / $a^{(1)}$ et $a^{(2)}$: $S_i = |\cos(a^{(i)}, a_m^{(i)})|$
 $S_1 > 0,98$; $S_2 > 0,88$

Résultats

4. ACP multigroupes: loadings des pays « déviants »



Résultats



5. ACP multigroupes: bilan

- Faible variance inter-groupes
- Structure bidimensionnelle avec forte première composante
 - Comparable à ce qui a été trouvé dans d'autres enquêtes

- Très forte stabilité de la structure pour les 13 pays
 - Similarités / structure commune très fortes

Résultats

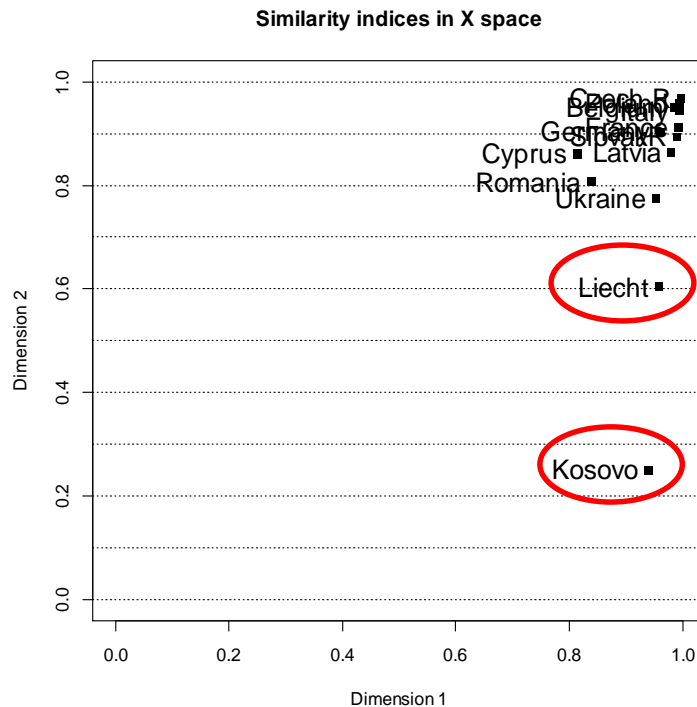


6. PLS multigroupes: inerties expliquées

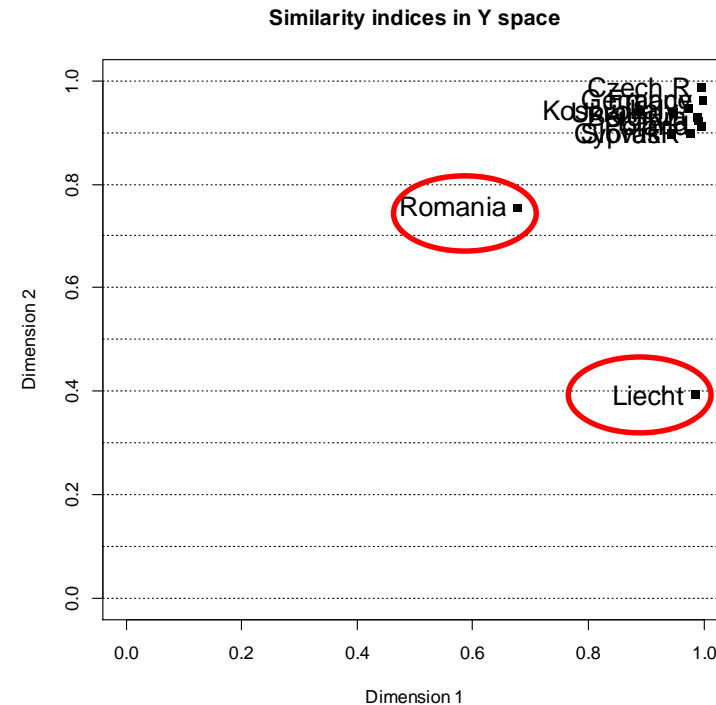
- Variance inter-groupes en X : 5%. Elle est ôtée par la suite.
- Deux importantes dimensions :
 - 33,1% et 16,4% de X
 - 14,5% et 1,4% pour Y (CAST)
 - Y (CAST) relativement indépendant des X
 - Explication : 11 variables en X, assez disparates, vs 6 variables en Y, formant une échelle cohérente

Résultats

7. PLS multigroupes: similarités / $t^{(1)}$ et $t^{(2)}$



Liechtenstein et Kosovo sur $t^{(2)}$

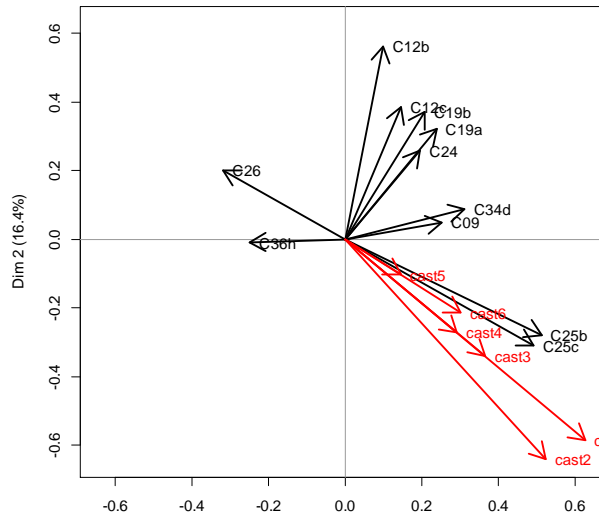


Roumanie ($t^{(1)}$ et $t^{(2)}$)
Liechtenstein ($t^{(2)}$)

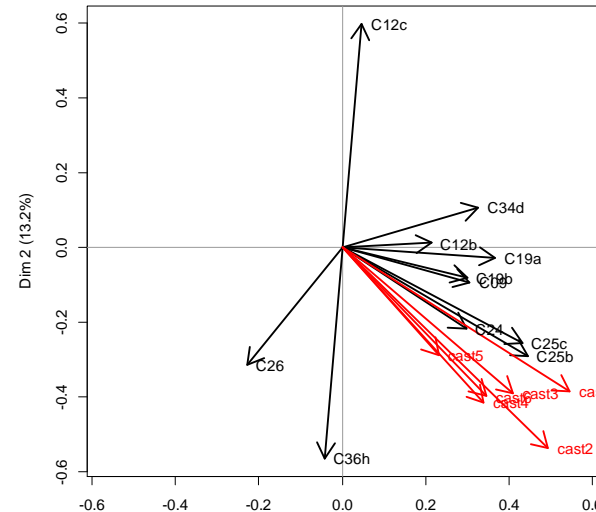
(comme dans mgACP)

Résultats

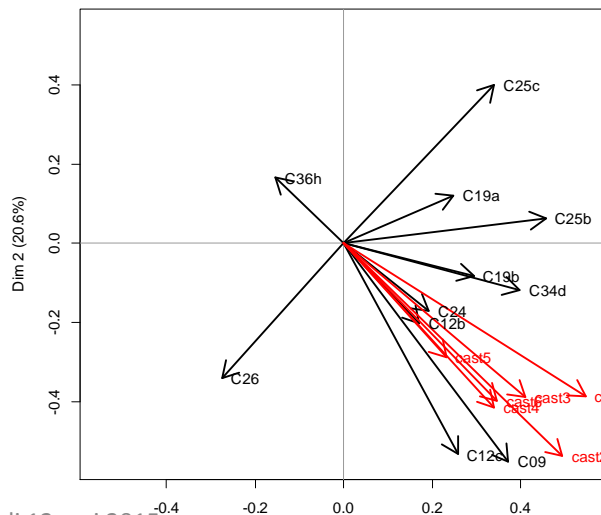
8. PLS multigroupes: loadings communs et déviants



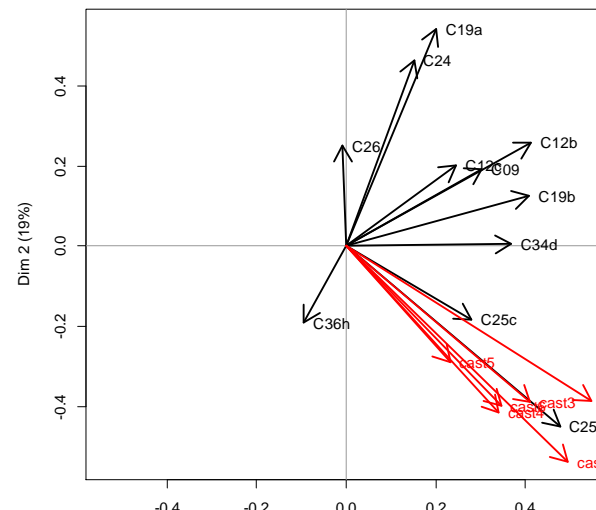
Loadings communs Dim 1 (33.1%)



Kosovo (X) Dim 1 (49.4%)



Liechtenstein (X et Y) Dim 1 (41.1%)



Roumanie (Y) Dim 1 (48.9%)

Résultats

9. PLS multigroupes: bilan



- Faible variance inter-groupes
- Le CAST apparaît assez peu lié aux variables explicatives: X explique 15% de sa variance (sur $t^{(1)}$ et $t^{(2)}$), 16% au total.
- On retrouve les mêmes pays déviants en Y qu'avec la mgPCA : Roumanie et Liechtenstein.

→ Le CAST est toujours assez invariant

- Les loadings et les coefficients de régression permettent de montrer que (sur $t^{(1)}$ et $t^{(2)}$): le Liechtenstein dévie en Y et X ; la Roumanie dévie en X ; le Kosovo dévie en X.

mgACP et mgPLS : multidimensionnalité du phénomène étudié
prise en compte

- mgACP est facile à utiliser: descriptif
 - Proche de l'AFM duale mais : produit des loadings communs et de groupes et un indice de similarité
 - Evaluer l'invariance de questionnaires / sous-populations
- mgPLS facile à utiliser : descriptif et explicatif
 - Permet d'enrichir la mgACP
 - Loadings communs et de groupes, indice de similarité
 - Utilisation de type multiniveau même lorsque le nombre de groupes est réduit

Références

- Eslami, A., Qannari, E. M., Kohler, A., & Bougeard, S. (2013a). Analyses factorielles de données structurées en groupes d'individus. *Journal de la Société Française de Statistique*, 154(3), 44-57.
- Eslami, A., Qannari, E. M., Kohler, A., & Bougeard, S. (2013b). General overview of methods of analysis of multi-group datasets. *Revue des Nouvelles Technologies de l'Information*(25), 108-123.
- Eslami, A., Qannari, E. M., Kohler, A., & Bougeard, S. (2014). Two-block multi-group data analysis. Application to epidemiology. In G. Russolillo (Ed.), *New perspectives in Partial Least Squares and Related Methods*: Springer Verlag.
- Guillerme, M., & Razafindranovona, T. (2015). *L'apport de l'analyse factorielle multiple duale pour l'analyse de la collecte multimode* Paper presented at the Journées de méthodologie statistique, Paris.
- Legleye, S., Piontek, D., Kraus, L., Morand, E., & Falissard, B. (2013). 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*, 22(1), 16-26. doi: 10.1002/mpr.1378
- PACKAGE R : multigroup, développé par Aida Eslami