**Multigroup analyses to study the CAST(Cannabis Abuse Screening Test)in thirteen European countries**

**Résumé:**

Nous proposons deux nouvelles méthodes pour explorer l'invariance configurale d'une échelle de repérage administrée dans M groupes différents, fondées sur l'analyse en composantes principales (ACP) et les moindres carrés partiels (PLS). L'analyse multigroupes en composantes principales (mgACP) cherche le vecteur de loadings communs *a* pour les P variables d'intérêt, qui soit communs aux M groupes étudiés et également le plus proche des loadings propres aux M groupes. L'analyse PLS multigroupes (mgPLS) cherche les vecteurs de loadings *a* (pour les P variables de l'espace Y) et *b* (pour les Q variables explicatives de l'espace X) communs à tous les groupes tels que leurs variables latentes $u_m=\mathbf{Y_m}b$, $t_m=\mathbf{X_m}a$ aient une covariance maximale.

Les données proviennent de l'enquête ESPAD (European survey project on alcohol and other drugs) 2011, une enquête scolaire menée dans 36 pays. L'échelle étudiée (P=6 variables représentant Y dans mgPLS, X dans mgPCA) est le Cannabis abuse screening test (CAST) ; 11 variables additionnelles (usages de substances psychoactives, perception de l'accessibilité du cannabis ,des risques liés à sa consommation) forment les Q variables explicatives de l'espae X dans la mgPLS.

La mgPCA montrent la quasi unidimensionnalité du CAST (la première dimension commune représentant 46,6% de l'inertie, la seconde 16.1%) : tous les pays sont en parfaite concordance avec la structure commune (similarité >0.98 pour la première dimension, >0.95 pour la seconde) sauf le Kosovo qui dévie quelque peu sur la seconde (0.88).

La mgPLS montre une très bonne concordance avec la structure moyenne sur les deux premières dimensions de X (similarité minimale sur la première dimension=0.81 pour Chypre, 0.84 pour la Roumanie ; l'Ukraine, le Kosovo et le Liechtenstein étant à part pour la seconde : similarités=0.77, 0.61 et 0.25). La concordance dans Y est également très bonne, comme dans la mgPCA sauf pour la Roumanie (similarités=0.65 sur la première dimension, 0.76 sur la seconde) et le Liechtenstein sur la seconde (0.39). Les graphiques et les mesures proposées notamment les coefficients de la régression PLS entre X et Y permettent de décrire et d'interpréter les liens et les spécificités nationales entre X et Y.

Le package R multigoup est développé par Aida Eslami et ses collègues.

**Abstract:**

We propose two new methods, respectively based on classical principal component analysis (PCA) and partial least square (PLS), to explore the configural (structural) invariance of a scale administered in M different groups.

Multigroup PCA (mgPCA) seeks a common vector of loadings *a* to study the relationships between the P variables of interest, these links being common to all the individuals clustered in M groups and being viewed through their group vector of loadings.

Multigroup PLS (mgPLS) seeks vectors of loadings *a* (for the P variables –Y space-) and *b* (for the Q explicative variables of Y, the X-space) common to all the groups such as their associated latent variables $u_m = \mathbf{Y_m} b$, $t_m = \mathbf{X_m} a$ are tightly linked following a criterion that maximizes their covariance.

Data come from the last 2011 European survey school project on alcohol and other drugs (ESPAD), a school survey among the pupils aged 15-16 in 36 European countries: the Cannabis abuse screening test (CAST) has been implemented in 13 countries (n=5204). The P=6 variables of the CAST form the Y space in the mgPLS (and the X space in the mgPCA), while Q=11 additional explanative variables (patterns of substance use, availability and perception of risks associated with substance use) form the X space in the mgPLS.

mgPCA shows that the CAST is quasi-unidimensional (1st common dimension explaining 46.6% of the variance, the second 16.1%): all countries present almost perfect concordance with the common structure (similarities>0.98 for the 1st dimension and >0.95 on the second) but Kosovo is somewhat deviant on the second dimension (similarity=0.88).

mgPLS shows a very high concordance of the countries in the two first dimensions of the X space (minimal similarity with the common structure=0.81 for Cyprus, 0.84 for Romania on the first dimension; Ukraine, Kosovo and Liechtenstein being apart for the second: similarities=0.77, 0.61 and 0.02). As in mgPA, the concordance in Y is very high except for Romania (similarity=0.65 on the first dimension, 0.76 on the second) and Liechtenstein on the second dimension (similarity=0.39). Graphics and measures explaining the common and specific relations between X and Y variables across groups are produced.

The R package multigroup has been developed by Eslami, Bougeard and colleagues.

## 1 Introduction

Many screening scales assessing cannabis-related problems have been developed and tested in recent years (Beck and Legleye 2008), although very few have been validated in Europe (Piontek, Kraus et al. 2008). One of the most-used is the Cannabis Abuse Screening Test (CAST) (Legleye, Karila et al. 2007). Originally designed for adolescents, it was adopted in the European School Survey Project on Alcohol and other Drugs (ESPAD) (Hibell, Guttormsson et al. 2009; Piontek, Kraus et al. 2009). Its psychometric properties have been assessed in representative sample of adolescents in France (Legleye, Piontek et al. 2011; Legleye, Piontek et al. 2013) and in Italy (Bastiani, Siciliano et al. 2013), among Hungarian students (Gyepesi, Urban et al. 2014) and among French adults (Legleye et al. 2014 to be published). Good internal, psychometric and screening properties were also found in small samples of young adults in Spain (Cuenca-Royo, Sánchez-Niubó et al. 2012; Fernandez-Artamendi, Fernández-Hermida et al. 2012).

The patterns of cannabis use are not uniform in Europe (EMCDDA 2011). Some countries like France, UK or Czech Republic show very high prevalences compared to the Scandinavian countries or compared to some southern countries like Malta or Latvia. The validity of cross-country or cross-cultural depends on the fact that the instrument that is used measures the same concepts in all countries or cultures. Although the CAST was added as an optional module in the ESPAD questionnaire since 2007 and is widely used since, no study assessed its cross-cultural validity in the European context. This article is divided into two sections.

We address first the problem of describing the multivariate datasets that is divided into 13 groups (ie. the countries), all individuals of the countries answering to the same questions. The aim is to seek common parameters, common loadings across groups as well as group parameters to understand the group specificity in comparison with the common structure. A multigroup Principal Component Analysis (mgPCA) will be used here for this purpose (Eslami et al., 2013). This analysis is followed by a multigroup partial least-squares analysis (mgPLS), aiming to explain the CAST by nine variables which describe some aspects of the drug use of the respondents and of their consumption context. The main purpose is to investigate the links that are common to all the countries, but differences and similarities between the thirteen countries are also of paramount interest.

## 2- Data and measurement

The 2011 ESPAD European survey is a pen and paper self-administered school survey which aims at collecting data on alcohol and other drugs following the same protocol in various countries (www.espad.org). In 2011, one optional module in the questionnaire was the Cannabis Abuse Screening Test (CAST) (Legleye, Piontek et al. 2011), that was chosen by 13 countries (). The database consists of 5204 pupils aged 15-16 originated from the following countries who reported having smoked cannabis in the last 12 months and who answered all CAST questions: Belgium (n=331), Cyprus (n=177), Czech Republic (n=1013), France (n=723), Germany (n=365), Italy (n=617), Kosovo (n=55), Latvia (n=292), Lichtenstein (n=52), Poland (n=1,113), Romania (n=93), Slovak Republic (n=246) and Ukraine (n=127). Only the questionnaires with no missing data on the CAST questions were kept for the analysis (231 were discarded).

The CAST assesses the following aspects of cannabis consumption in the past 12 months: non-recreational use (CAST 1 “Have you smoked cannabis before midday?”, CAST 2 “Have you smoked cannabis when you were alone?”), memory disorders (CAST 3 “Have you had memory problems when you smoked cannabis?”), reproaches from family or friends (CAST 4 “Have friends or family members told you that you should reduce or stop your cannabis consumption?”), unsuccessful attempts to quit (CAST 5 “Have you tried to reduce or stop you cannabis use without succeeding?”), and problems linked to cannabis consumption (CAST 6 “Have you had problems because of your cannabis use (argument, fight, accident, poor results at school, etc.)?”). All items are answered on a five-point scale (0 “never”, 1 “rarely”, 2 “from time to time”, 3 “fairly often”, 4 “very often”).

*Additional variables for mgPLS*

The 11 following variables were also considered for the multilevel analysis: frequency of cannabis use and of alcohol use in the last 12 months and in the last 30 days (0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+), frequency of drunkenness during life and in the last 12 months (0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+), number of tobacco cigarettes in the last 30 days, age at first cannabis use (9 or less, 10, 11, 12, 13, 14, 15, 16+), proportion of friends smoking cannabis (none, some, most, almost all, all), the perception of risks associated to regular cannabis use (no risk, light risk, moderate risk, important risk) and perceived availability of cannabis (getting cannabis seems: impossible, very difficult, fairly difficult, fairly easy, very easy).

## 3- Statistical analysis

### *3.1- Multigroup principal component analysis*

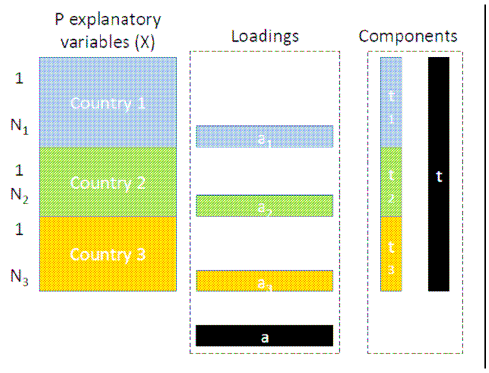



Figure 1: Schema of the mgPCA

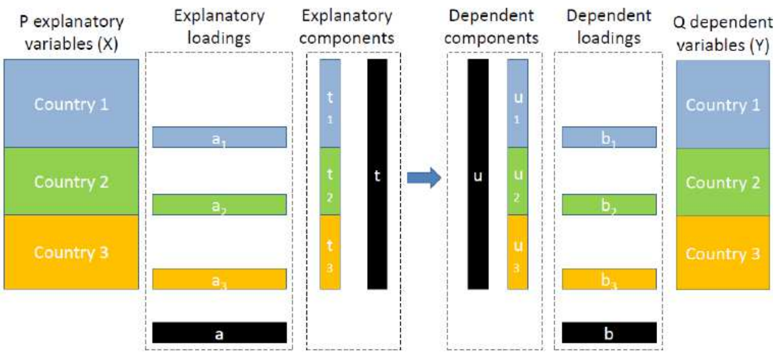
Consider the multigroup setting where we have a single dataset X involving P variables and N individuals a priori divided into M groups $X_m$ with m = (1, …M). We assume that each dataset $X_m$ of dimensions ($N_m$*P) is column-centered. As the six variables have different variances, data are globally centered and scaled to give the same weight to all the variables in the analysis.The aim of the analysis is to seek a common vector of loadings *a* to study the relationships between the P variables, these links being common to all the individuals. To better understand the group specificity in comparison with the common structure, the P variables may also be viewed through their group vector of loadings $a_m$. Multigroup PCA consists in seeking a vector of loading *a* common to all the groups, tightly linked to the M vectors of group loadings ($a_1$, …$a_M$): so as to maximize the following criterion (1) for the first dimension h=1 (Eslami, Qannari et al. 2013).

$$\text{Max} \sum_{[m=1,M]} \langle a_m^{(1)}, a^{(1)} \rangle^2, \text{ with } a_m^{(1)} = X'_m t_m^{(1)} \text{ and } \|t_m^{(1)}\| = \|a^{(1)}\| = 1 \qquad (1)$$

Subsequent vectors of loadings ($a^{(2)}$, … $a^{(H)}$) are sought by considering the same maximization problem and adding constraints of orthogonality of the vector of loadings to be determined at the current stage with those determined at previous stages, where H = rank(X) is the maximum dimension of the analysis. The group component $t_m^{(h)}=\mathbf{X}_m a^{(h)}$ is the principal component in group m associated with the common vector of loadings $a^{(h)}$. We denote by $t^{(h)}$ the common component associated with dimension h and defined by the vertical concatenation of the group components ($t_1^{(h)}$, … $t_M^{(h)}$). This concatenation is possible since all group components share the same loadings. These latter common components ($t^{(1)}$, … $t^{(H)}$) are orthogonal with each other.

We will apply this analysis to the P=6 CAST variables and the M=13 countries. The aim of the mgPCA is to investigate the relationships between the CAST variables: it exhibits the factorial structure that is common to all countries as well as the factorial structures that are proper to each country and the similarities between the group structures and the common structure. .

### *3.2- Multigroup partial least squares analysis*




Pareil pour le titre, c'est toujours mieux

Multigroup PLS seeks vectors of loadings *a* and *b* common to all the groups respectively in explanatory (X) and dependent (Y) datasets, such as their associated latent variables $t_m=\mathbf{X_m}a$, $u_m=\mathbf{Y_m}b$are tightly linked following the criterion (2) for the first dimension:

$$\text{Max} \sum \text{Cov}(\mathbf{X_m}a^{(1)}, \mathbf{Y_m}b^{(1)}) \text{ with } ||a^{(1)}||=||b^{(1)}||=1 \qquad (2)$$

[m=1…M]

While considering that $\sum_m<a_m,a>^2=\sum_m<\mathbf{X_m}', t_m>^2=\sum_m<\mathbf{X_m}a, t_m>^2=\sum_m N_m^2\text{cov}^2(\mathbf{X_m}a, t_m)$, the criterion (2) can be viewed as an extension of the mgPCA criterion (1) with the constraints that the group components $t_m$ are now constrained to be in the dependent space. The common components $t^{(1)}$ and $u^{(1)}$, respectively associated with X and Y, come from the vertical concatenation of the group components ($t_1^{(1)}$, … $t_M^{(1)}$) and ($u_1^{(1)}$, … $u_M^{(1)}$). As for multigroup PCA, this concatenation is possible since the group components respectively share the same loadings. To better understand the group specificity in comparison with the common structure, let $a_m^{(1)}$ and $b_m^{(1)}$ be the specific group vectors of loadings associated with $\mathbf{X}_m$ and $\mathbf{Y}_m$.

These group loadings are retrieved from $a_m^{(1)}=\mathbf{X}_m'u_m^{(1)}/||\mathbf{X}_m'u_m^{(1)}||$, $b^{(1)}_m=\mathbf{Y}_m't_m^{(1)}/||\mathbf{Y}_m't_m^{(1)}||$. It follows that the group loadings come from the common loadings weighted by the covariances between $\mathbf{X}_m$ and $\mathbf{Y}_m$. Subsequent loadings and components are sought by considering the same maximization problem and using the deflated **X** (and eventually **Y**) with respect to the first common component $t^{(1)}$. Thereafter, the common components ($t^{(1)}$, … $t^{(h)}$) are orthogonal with each other's. From this property, a common regression model to explain Y from X can be set up by means of the optimal number of common components ($t^{(1)}$, … $t^{(hopt)}$), as described in (Eslami, Qannari et al. 2014).

Thereafter, the group specificity in comparison with the common structure may be studied. Actually, we are interested in understanding if the relationships between the variables under study are the same for all the groups. This can be achieved by the comparison of the graphical displays of the group and the common loadings.

The relative importance of each group dimension h = (1, … H) is given by the percentage of the part of variance of Xm restored by $t^{(h)}$, i.e var($t_m^{(h)}$) = $a^{(h)\prime}\mathbf{X_m}'\mathbf{X_m}a^{(h)}$. This comparison is easily performed for a limited group number and a reduced-dimension space but for most cases, an overall index is first needed. For this purpose, we consider the similarity indices processed for the optimal number of dimensions hopt under study and given by Eq. (3).

$$S_m^{(hopt)}=1/h_{opt} \sum|a^{(h)\prime}\ a_m^{(h)}|=1/h_{opt} \sum|\cos(a^{(h)}a_m^{(h)}|\ \text{for } m=1, \dots M \qquad (3)$$

For each group, this index varies between 0 (total dissimilarity, i.e., orthogonal common and group vectors) and 1 (perfect agreement up to dimension $h_{opt}$ ). For multigroup PLS, these indices may also been processed for the comparison of the dependent loadings b and ($b_1$, … $b_m$). This leads to the common and group loadings comparison in the explanatory space on the one hand and in the dependent one on the other hand.

mgPCA and mgPLS were computed in R (package multigroup: www.cran.r-project.org/web/packages/multigroup/index.html).

## 4- Results

### *4.1- Data pre-treatment*

The box-plots of the CAST variables are shown in Figure 1. Means and standard deviations are different across countries (table 1): for each variable, the ratio of the maximum and minimum is above 1.50 and above 1.70 for the standard deviation.[et alors … si tu ne fais rien de cette info, ne pas la mettre ; mais tu t’en sers pour centrer et réduire, non ? OK, vu après, mais trop loin pour faire le lien]

ESPAD data show very large differences in CAST scores in the 13 countries that used the test in their questionnaire (Pabst, Kraus et al. 2012): Cyprus had the highest mean (2.5) followed by Italy (1.6) and Latvia (1.0). However, France has the highest mean frequency of use in the last 12 months (15.9) while Cyprus lies in 8$^{th}$ position while Latvia stands in 10$^{th}$ position (7.0) and Kosovo in last position (2.5). There are thus large discrepancies between cannabis frequency of use and CAST scores across countries.

We then choose to center and scale all the variables in the whole sample. The between group variance (i.e. the country effect) represents 2.59% of the total. It will be removed hereafter. That this proportion of explained variance is small does not mean that the structure of the CAST is the same in all countries.

*4.2- Study of the links between the CAST variables: Results from Mg-PCA*

Data are centered and scaled by group to give the same weight to all the countries in the analysis. The first result of the mgPCA ran on the CAST variables is that two or three dimensions could be retained since they explain 74.3% of the total variance (the first component comprises 46.4%, the second 16.1% and the third 11.8%). The graphical display in Figure 2a depicts the common loadings as projected in the first factorial plan. It allows us to investigate the relationships between the six variables from CAST which are common to all the countries. It follows that the two variables related to the non-recreational use (CAST1, CAST2) are linked and quite independent from reproaches from family or friends (CAST4) and unsuccessful quite attempts (CAST5). The loadings of the CAST variables on the first global component are quite equal, ranging between 0.40 (CAST4) and 0.45 (CAST2) except for CAST5 (0.32): this component can be interpreted as the frequency of problems and use. The loadings on the second component oppose CAST1 and CAST2 (-0.49, -0.38) on one side and CAST4 and CAST5 on the other (0.36, 0.68): the second component can be interpreted as an opposition of non-recreational use (frequency of use) and dependency symptoms. The loadings on the third component oppose CAST3 and CAST6 (-0.65, -0.46) on one side and CAST1, 2 and 5 (0.33, 0.33, 0.38) on the other: the third component can be interpreted as an opposition of memory and social problems and non-recreational use and dependency symptoms.

Figure 3a shows the similarity between the countries on the first axis of the common structure. It shows an almost perfect coherence with the common structure as viewed from the first common axis: no country shows deviation, except Liechtenstein, Romania and to some extent, Cyprus and Kosovo (although their indexis above 0.98). Figure 3b combines the two first common axes. Again Liechtenstein, Romania and Cyprus deviate in the first axis, while Ukraine, Liechtenstein and especially Kosovo deviate in the second axis (for this latter, the similarity index is below 0.9 whereas it is above 0.96 for the others). Finally, Figure 3c offers a three dimensional plot of the similarities from the point of view of the first three axes: while the fit of Romania and Liechtenstein with the common structure is equivalent along the third axis, Kosovo once again deviates with a similarity index below 0.85. These three graphs show us that the CAST structure of the countries is almost the same as viewed from the first common axis (with a relative exception for Liechtenstein and Romania), while the CAST structure of Kosovo is very singular on the second and third axis. These countries need a more in-depth examination.

Figures4shows the first factorial plan obtained by mgPCA for all countries and the factorial for each of the three most deviating countries, Liechtenstein and Romania that deviates on the first axis and Kosovo that deviates on the next two axes. In Kosovo, the first axis is very strong (79.0% of the variance compared to 6.3% for the second), meaning that the CAST is highly unidimensional.CAST2 (smoking when alone) and CAST6 (problems due to consumption) are highly correlated, while the correlation between CAST1 and CAST2 is lower, which is unique in our study. In Liechtenstein, thepercentages of the variances explained by the axes are more balanced (34.2% and 21.1%), highlighting that the CAST tends to be less unidimensional. CAST3 (memory problems) and CAST6 on one side and CAST4 (reproaches) and CAST5 (unsuccessful quite attempts) on the other side, are highly correlated, which is also unique.In Romania, the percentages of variances explained by the axes are closer to the common structure (40.2% and 14.8%), but these shares suggest that the CAST is less unidimensional than in the common structure. CAST1, 2 and 3 are highly correlated, which is also unique.

*4.3- Study of the links between CAST and the context: results from MgPLS*

The previous analyses showed that although the CAST seems to have a strong structure that is robust and common to the majority of countries, some exceptions are noticeable. Kosovo and Liechtenstein, but also Romania differ from the global structure. The purpose of the mgPLS is to try to find some determinants of these differences by looking at the relations between the CAST and some elements of the context of drug use as reported by the respondents. The 11 additional variables we consider refer to licit drug use (alcohol consumption and frequency of drunkenness, tobacco), as well as some other features that relate to cannabis use: the age at the first use, the perception of its availability, the perception of risks associated with a regular use of cannabis and the proportion of friends smoking cannabis. These variables compose the X-space while the CAST variables compose the Y-space.

The mgPLS exhibit two important dimensions, the first explaining 33.1% of the variance and the second 16.4% (the third accounting for 9.1%).The two first dimensions account thus for 49.5% of the variance in the X space while they account for 15.9% on the Y space (14.5% for the first dimension). Two dimensions may thus be retained for the description of the results.

Figure 5 displays the similarities of the countries compared to the first two common PLS dimensions. As for mgPCA, all countries show high fit for the CAST (Y-space) with the common structure on the first dimension and second dimension, with the exceptions of Romania (0.68 on the first dimension) and Romania and Liechtenstein (0.76 and 0.39)on the first dimension. Compared to the other countries, Kosovo is apart on the second dimension although the similarity is very high (0.89).

For the explanatory variables (X-space), all countries show high fit with the common first dimension, with a relative exception of Cyprus (index=0.81) and Romania (index =0.84) while Latvia, Ukraine (0.77), Liechtenstein (0.60) and especially Kosovo (0.25) show poor fit on the second common dimension.

The loadings plots of the common structure and of the deviating countries are shown figure 6. The panel a represents the common structure on the first factorial plan. All the CAST variables appear oriented the same way, along with C25b and C25c (frequency of cannabis use in the last 12 months as well as in the last 30 days) and C26 (age at first cannabis use). The variables CAST1 and CAST2 appear highly correlated and well represented in the plan, while the other variables appear correlated but less well represented. CAST4, CAST5and CAST6 are better represented by the third dimension while CAST3 is moderately linked to all the six common dimensions.

The regression coefficients obtained for the two common dimensions are depicted in Table 2. The variables that show the highest coefficients with the CAST variables are C25b and C25c (frequency of cannabis use in the last 12 months and in the last 30 days). They present very similar coefficients on each CAST variable: the highest are for CAST1and CAST2 (0.21 and 0.19) and the lowest on CASST5. The coefficients for CAST3, CAST 4 and CAST 6 are also similar (#0.12, #0.10 and #0.10). That the coefficients for CAST5 are low (0.05) suggest that “unsuccessful quit attempts” is somewhat apart in the CAST scale and is not well explained by the frequency of use. The other context variables show lowinfluence on the CAST variables. This is noticeable for the frequency of alcohol use in the last 12 months (C12b) and in the last 30 days (C12c), the frequency of drunkenness in the lifetime (C19a) and in the last 12 months (C19b), as well as with the tobacco use (C09).This suggests that the problematic use of cannabis is a special pattern of use that is not linked to the use of licit drugs. Also surprisingly, the perceived availability of cannabis (C24), the proportion of friends that smoke cannabis (C34d) and the perception of the risk of regular cannabis smoking (C36h) show low correlations with the CAST variables.

As already shown, Kosovo appears singular on the X and Y spaces: in Figure 6 panel b, we see that the first component explains a higher share of the variance than in the common analysis (49.4% vs 33.1%). It means that the X-space is more unidimensional, meaning that the context variables are more correlated than in the other countries. CAST variables are better represented on the plan and more highly correlated, meaning that the CAST is clearly unidimensional in this country (as previously seen with the mgPCA). The CAST variables are linked with the frequency of cannabis use (C25b and C25c), the perceived availability of cannabis the perception of risk associated with regular cannabis use (C36h), the use of tobacco and the frequency of drunkenness (C09 and C19a C19b) but also negatively with the frequency of alcohol use in the last 30 days (C12c). There is almost no link with the age at first use (C26). In this country, cannabis is smoked by people who drink less but who perceive cannabis as dangerous. Kosovo is also the country with the one of the lowest CAST score and the lowest frequency of use (see Table 1).

Figure 6 panel c shows that in Liechtenstein, the CAST variables are correlated to tobacco smoking (C09) and frequency of alcohol use (C12c), and also, to some extent, with the frequency of drunkenness (C19b). The link with the frequency of cannabis use in the last 30 days (C25c) is close to 0, as is the link with the age at the first cannabis use (C26), and the one with the frequency of cannabis use in the last 12 months (C25b) is also low. Perceiving regular cannabis use as dangerous (C36h) is weakly and negatively linked with the CAST variables. In Liechtenstein, problematic cannabis use as described with the CAST is mainly associated with licit drug use and neither with the precocity of use nor with the frequency of use. This country is also characterized by rather high mean CAST score and a high frequency of use.

Figure 6 panel d shows the case of Ukraine, a country in which the CAST structure as exhibited by the mgPCA is very close to the common structure. The percentages of variance explained by the first components are close to those measured for Liechtenstein and higher than those measured in average, showing that the X-space less unidimensional than in the average. However, Ukraine stands moderately apart for the X-space (similarity index=0.77 on the second axis). This particularity is partly due to the precocity of the first cannabis use (C26), which is more strongly and negatively linked to the CAST variables. Plays also a role the stronger effect of the frequency of alcohol use in the last 12 months (C12b), and the weaker effect of the frequency of cannabis use (C25b and C25c). The proportion of friends smoking cannabis (C34d) is also more influent than in the average. In Ukraine, problematic cannabis use is strongly linked to precocity of use and alcohol use and is a more collective behavior than in the average. This country is also characterized by a moderate frequency of use but a rather high CAST score.

Finally, Figure 6 panel e shows the case of Romania, a country that is close to the average in the X-space (indexes=0.84 and 0.81 on the two first components) but somewhat apart on the Y-space (similarity indexes=0.68 on the first common component, 0.76 on the second). The percentages of variance explained by the first components are higher than in the average (48.9% and 19.0%) and cumulate to a very high share of the total. The particularity of Romania is that the CAST variables are mostly linked to the frequency of use in the last 12 months (C25b) and, to a smaller extent, to the proportion of friends who smoke cannabis (C34d) and the frequency of drunkenness in the last 12 months (C19b). Licit drug use, precocity of cannabis use, perception of risks associated with cannabis, perceived availability of cannabis play a very small role. In Romania, problematic cannabis use is essentially linked to the frequency of use. Romania has a very low frequency of use and a rather high mean CAST score.

## 5- Discussion&conclusion

Both methods MgPCA and mgPLS offer simple and efficient ways to describe (mgPCA) and explain multigroup and multitable data (mgPLS). One main interest is to provide intuitive graphics such as loadings plots as well as similarity indexes with the common structure that allows estimating the differences between groups and identifying the most deviant ones.

In our case, the CAST factorial structure as observed in mgPCA appears stable across the 13 countries with some exceptions that are very specific: Kosovo, Liechtenstein and Romania. In these countries, the mgPCA exhibits very peculiar structures that can also be studied with mgPLS: this comprehension is useful to the global judgment of the invariance of the CAST structure.

The method mgPCA has a descriptive aim. It is based on similar grounds than dual multiple factorial analysis (DMFA) as implemented in the popular FactoMineR R package (Eslami, Qannari et al. 2013; Eslami, Qannari et al. 2013). MgPCA and DMFA lead to the same common loadings, but the advantages of mgPCA are: it leads to group loadings as well as common loadings and it is based on a stratforward maximization criterion which can be developed to more complicated data structure such as two block and multiblock analysis. Additionally, mgPCA as implemented in the multigroup package produces a similarity index that is very useful.

The mgPLS is a compromise between description and explanation of a table Y by a table X while taking account for the group structure. In complement to mgPCA, it can reveal differences in the association between X and Y between groups and show the most deviant ones from this point of view. In our case, we found that most of the countries who exhibit particular CAST structures also presented singular context structures.

One strength of the mgPLS approach is that it can adequately handle clustered data when the number of groups is too limited to conduct classical multilevel regressions. By definition, it also takes into account the fact that the explained phenomenon is not simply reducible to a single indicator. In our case, the problematic cannabis use is a concept that is more adequately described by several variables instead of one: the two main dimensions of the CAST are thus studied at the same time.

One limit of our analysis is that we do not consider socioeconomic variables that can play a role (wealth of the countries, unemployment rates, level of education, legislations towards drug use, etc.) For example, Kosovo and Liechtenstein are very small and peculiar countries, one being one of the wealthiest countries in the world with only 37000 inhabitants (Liechtenstein), the other one being a very recent independent state that endured war and social troubles in the past years. Romania is a recent member of the European Union, relatively poor with an important agricultural sector. Our interpretations of the differences is thus very limited.

Potential application in survey methodology relates to the measurement of data collection mode effects (when using a mix-mode survey combining internet and telephone interviews for example). The more classical DMFA has been used recently for an experiment in a survey on violence (Guillerm and Razafindranovona 2015): the same analysis can be ran with mgPCA and completed with mgPLS in order to explain the differences in the various data collection modes, for examples those that are linked to sociodemographic variables.

Figure 1 : Distribution of the CAST variables across countries

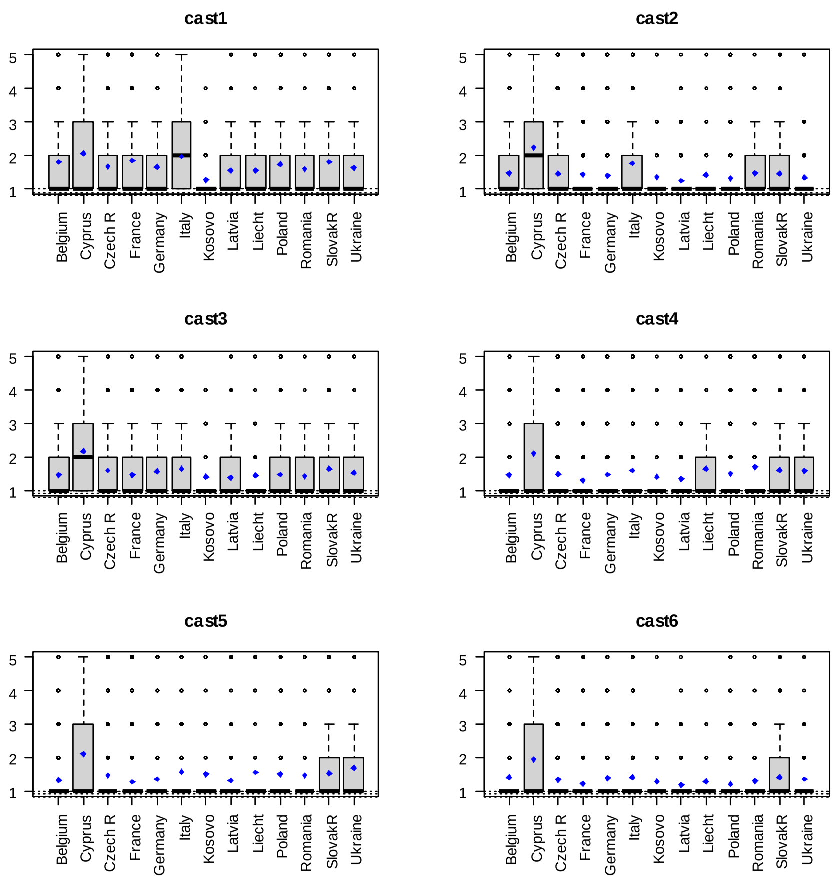


**Figure**2 : global loadings plots of the multigroup PCA and 3D plot on the first 3 components

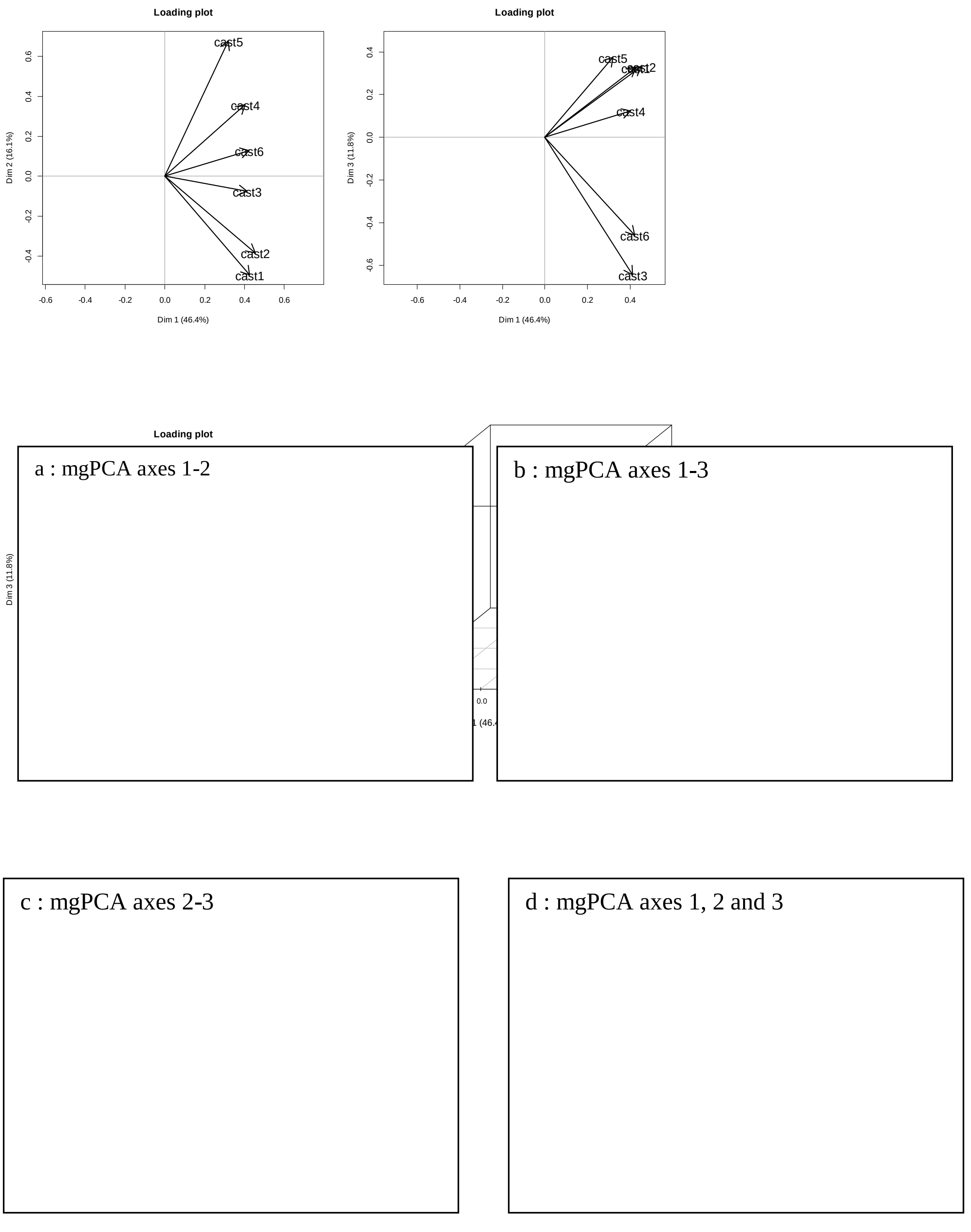
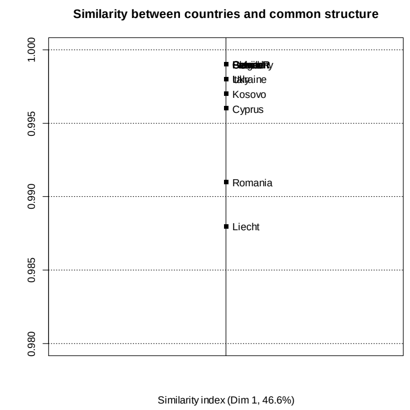
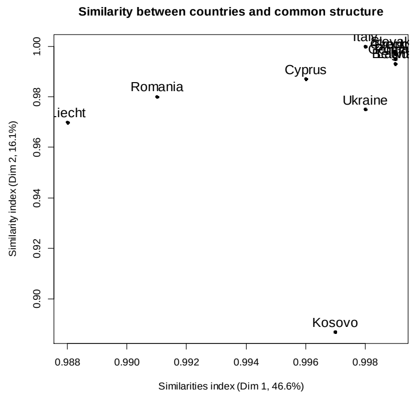
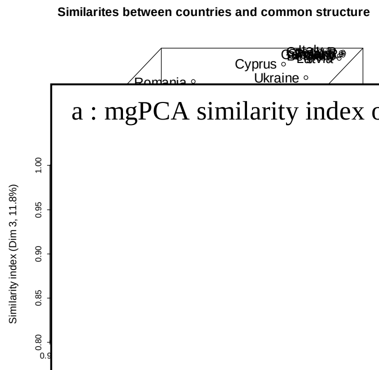


Figure 3 : Similarities between countries and the common structure (first axis, first plan, and 3D plot of the first three axes)




a : mgPCA similarity index on axis 1

b : mgPCA similarity indexes on axes 1-2

c : mgPCA similarity indexes on axes 1, 2, 3

Figure 4 : PCA for the most deviating countries.

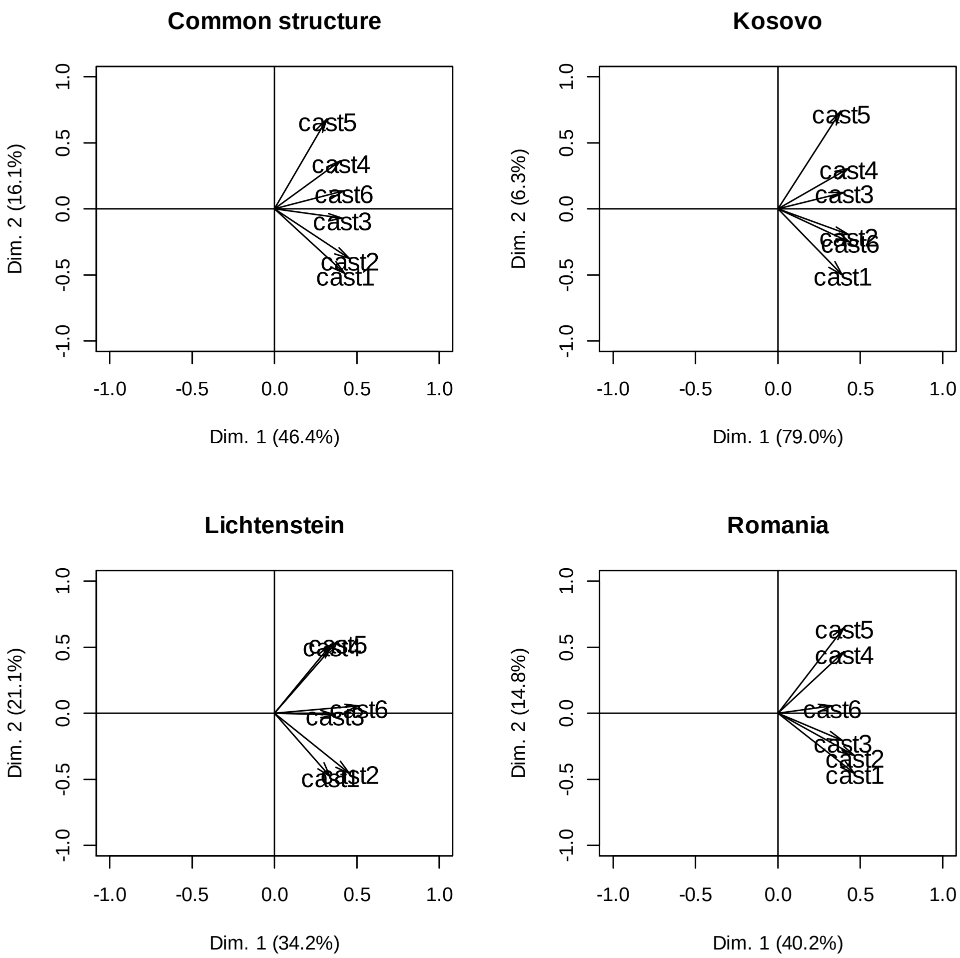


Figure 5 : Similarities between countries and common PLS dimensions

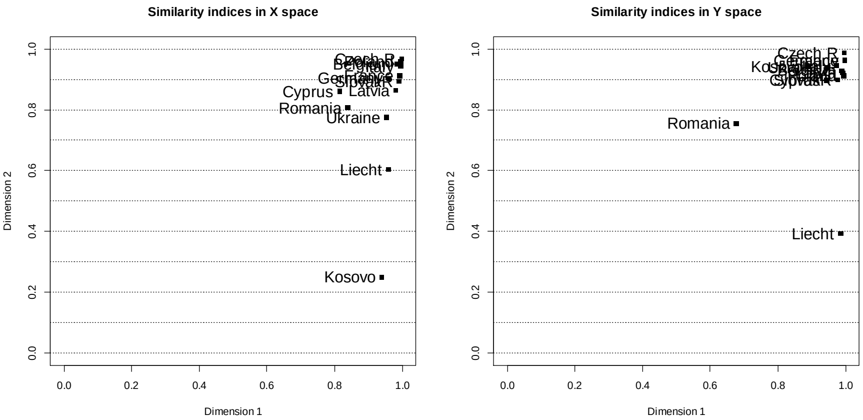
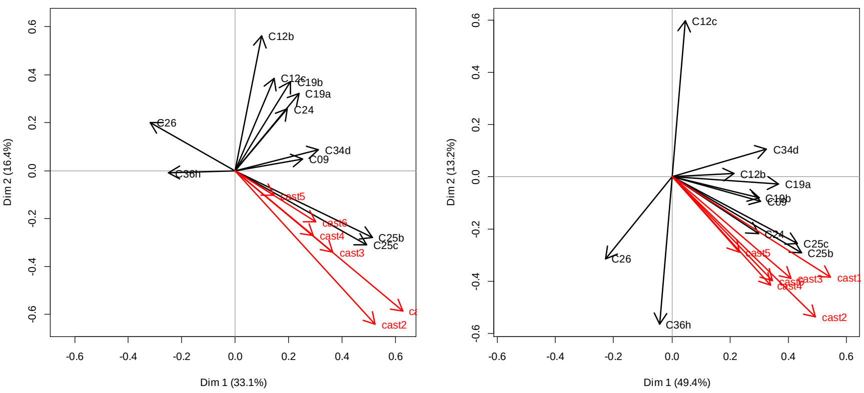


Figure 6 : Common loadings plot, Kosovo, Liechtenstein, Ukraine and Romania




a : mgPLS common loadings plot

b : mgPLS loadings plot in Kosovo

c : mgPLS loadings plot in Liechtenstein

d : mgPLS loadings plot in Ukraine

e : mgPLS loadings plot in Romania

Table 1 : means and standard deviations of the CAST in 13 countries

| | | Means | | | | | | | | Standard deviations | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| country | N= | Freq. of use | cast1 | cast2 | cast3 | cast4 | cast5 | cast6 | CAST score | Freq. of use | cast1 | cast2 | cast3 | cast4 | cast5 | cast6 | CAST score |
| Belgium | **331** | 13.9 | 1.8 | 1.5 | 1.4 | 1.4 | 1.3 | 1.4 | 2.9 | 16.9 | 1.1 | 0.9 | 0.9 | 1.0 | 0.9 | 0.9 | 4.0 |
| Cyprus | **177** | 8.9 | 2.1 | 2.2 | 2.2 | 2.1 | 2.1 | 1.9 | 6.6 | 13.5 | 1.4 | 1.4 | 1.4 | 1.5 | 1.5 | 1.4 | 7.1 |
| Czech R | **1013** | 11.5 | 1.7 | 1.5 | 1.6 | 1.5 | 1.5 | 1.3 | 3.0 | 15.8 | 1.0 | 0.9 | 1.0 | 1.0 | 1.0 | 0.8 | 3.9 |
| France | **723** | 15.2 | 1.8 | 1.4 | 1.5 | 1.3 | 1.3 | 1.2 | 2.5 | 17.2 | 1.1 | 0.9 | 0.9 | 0.8 | 0.9 | 0.7 | 3.4 |
| Germany | **365** | 10.2 | 1.7 | 1.4 | 1.6 | 1.5 | 1.4 | 1.4 | 2.8 | 14.8 | 1.1 | 0.9 | 1.0 | 1.0 | 1.0 | 0.8 | 3.9 |
| Italy | **617** | 14.1 | 2.0 | 1.8 | 1.6 | 1.6 | 1.6 | 1.4 | 4.0 | 16.9 | 1.2 | 1.2 | 1.1 | 1.2 | 1.2 | 0.9 | 4.4 |
| Kosovo | **55** | 2.5 | 1.3 | 1.4 | 1.4 | 1.4 | 1.5 | 1.3 | 2.2 | 4.5 | 0.7 | 0.9 | 0.9 | 0.9 | 1.1 | 0.8 | 4.7 |
| Latvia | **292** | 7.0 | 1.6 | 1.3 | 1.4 | 1.3 | 1.3 | 1.2 | 2.0 | 12.0 | 0.9 | 0.6 | 0.7 | 0.8 | 0.8 | 0.6 | 2.9 |
| Liecht | **52** | 11.1 | 1.6 | 1.4 | 1.4 | 1.6 | 1.6 | 1.3 | 2.9 | 15.6 | 1.1 | 1.0 | 1.0 | 1.2 | 1.3 | 0.7 | 3.6 |
| Poland | **1113** | 8.6 | 1.8 | 1.3 | 1.5 | 1.5 | 1.5 | 1.2 | 2.7 | 13.2 | 1.0 | 0.8 | 0.9 | 1.0 | 1.2 | 0.7 | 3.7 |
| Romania | **93** | 3.4 | 1.6 | 1.5 | 1.4 | 1.7 | 1.5 | 1.3 | 2.9 | 6.7 | 1.0 | 0.8 | 0.8 | 1.4 | 1.1 | 0.8 | 3.8 |
| SlovakR | **246** | 12.8 | 1.8 | 1.5 | 1.6 | 1.6 | 1.5 | 1.4 | 3.4 | 17.2 | 1.1 | 0.9 | 1.0 | 1.0 | 1.1 | 0.8 | 4.0 |
| Ukraine | **127** | 6.3 | 1.6 | 1.3 | 1.5 | 1.6 | 1.7 | 1.4 | 3.1 | 10.2 | 0.9 | 0.8 | 1.0 | 1.0 | 1.2 | 0.8 | 4.0 |
| Global | | 11.1 | 1.8 | 1.5 | 1.5 | 1.5 | 1.5 | 1.3 | | 15.4 | 1.1 | 1.0 | 1.0 | 1.0 | 1.1 | 0.8 | |
| Max | | Fra | Cyp | Cyp | Cyp | Cyp | Cyp | Cyp | Cyp | 17.2 | Cyp | Cyp | Cyp | Cyp | Cyp | Cyp | Cyp |
| Min | | Kos | Kos | Lat | Kos | Fra | Fra | Lat | Lat | 4.5 | Kos | Lat | Lat | Fra | Lat | Lat | Lat |
| Max/Min | | 6.0 | 1.6 | 1.8 | 1.6 | 1.6 | 1.6 | 1.6 | 3.0 | 3.9 | 2.1 | 2.3 | 1.9 | 1.8 | 1.8 | 2.4 | 2.4 |

Table 2 : regression coefficients for the two first PLS dimensions [evident que c'est cumulé pour 2 dimensions de pls]

| | cast1 | cast2 | cast3 | cast4 | cast5 | cast6 |
|---|---|---|---|---|---|---|
| C25b | **0.212** | **0.191** | **0.124** | **0.099** | 0.046 | **0.096** |
| C25c | **0.209** | **0.189** | **0.122** | **0.097** | 0.045 | **0.094** |
| C26 | **-0.135** | **-0.122** | -0.079 | -0.063 | -0.029 | -0.061 |
| C09 | 0.080 | 0.067 | 0.047 | 0.037 | 0.018 | 0.038 |
| C12b | -0.040 | -0.052 | -0.023 | -0.018 | -0.005 | -0.011 |
| C12c | -0.001 | -0.013 | 0.000 | 0.000 | 0.002 | 0.005 |
| C19a | 0.039 | 0.023 | 0.023 | 0.018 | 0.011 | 0.023 |
| C19b | 0.022 | 0.007 | 0.013 | 0.010 | 0.007 | 0.016 |
| C24 | 0.033 | 0.020 | 0.019 | 0.015 | 0.009 | 0.019 |
| C34d | **0.095** | 0.079 | 0.055 | 0.044 | 0.022 | 0.046 |
| C36h | -0.084 | -0.072 | -0.049 | -0.039 | -0.019 | -0.040 |

In bold type the coefficients close to 0.1 or above in absolute value.

**References**


Bastiani, L., V. Siciliano, et al. (2013). "Optimal scaling of the CAST and of SDS Scale in a national sample of adolescents." Addictive Behaviors(38): 2060-2067.

Beck, F. and S. Legleye (2008). Measuring cannabis related problems and dependence at the population level A Cannabis reader: global issues and local experiences S. Rödner Sznitman, B. Olsson and R. Room. Lisbon, European Monitoring Center on Drug and Drug Addiction (EMCDDA). **2:** 29-57.

Cuenca-Royo, A. M., A. Sánchez-Niubó, et al. (2012). "Psychometric properties of the CAST and SDS scales in young adult cannabis users." Addictive Behaviors(37).

EMCDDA (2011). 2011 Annual report on the state of the drugs problem in Europe. Lisbon, EMCDDA**:** 105.

Eslami, A., E. M. Qannari, et al. (2013). "Analyses factorielles de données structurées en groupes d'individus." Journal de la Société Française de Statistique**154**(3): 44-57.

Eslami, A., E. M. Qannari, et al. (2013). "General overview of methods of analysis of multi-group datasets." Revue des Nouvelles Technologies de l'Information(25): 108-123.

Eslami, A., E. M. Qannari, et al. (2014). Two-block multi-group data analysis. Application to epidemiology. New perspectives in Partial Least Squares and Related Methods. G. Russolillo, Springer Verlag.

Fernandez-Artamendi, S., J. R. Fernández-Hermida, et al. (2012). "Screening of cannabis-related problems among youth: the CPQ-A-S and CAST questionnaires." Substance Abuse Treatment Prevention and Policy**7**(13): 2-9.

Guillerm, M. and T. Razafindranovona (2015). L'apport de l'analyse factorielle multiple duale pour l'analyse de la collecte multimode Journées de méthodologie statistique. Paris, INSEE.

Gyepesi, A., R. Urban, et al. (2014). "Psychometric properties of the cannabis abuse screening test in Hungarian samples of adolescents and young adults." Eur Addict Res**20**(3): 119-128.

Hibell, B., U. Guttormsson, et al. (2009). The 2007 ESPAD Report - Substance Use Among Students in 35 European Countries. Stockholm, The Swedish Council for Information on Alcohol and Other Drugs (CAN)**:** 408.

Legleye, S., L. Karila, et al. (2007). "Validation of the CAST, a general population Cannabis abuse screening test." Journal of Substance Use**12**(4): 233-242.

Legleye, S., D. Piontek, et al. (2011). "Psychometric properties of the Cannabis Abuse Screening Test (CAST) in a French sample of adolescents." Drug Alcohol Depend**113**(2-3): 229-235.

Legleye, S., D. Piontek, et al. (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.

Pabst, A., L. Kraus, et al. (2012). The Cannabis Abuse Screening Test (CAST): Examining the prevalence of cannabis-related problems among adolescents in 13 European countries. The 2011 ESPAD Report. Substance Use Among Students in 36 European Countries. B. Hibell, U. Guttormsson, A. Salmeet al. Stockholm, Sweden, The Swedish Council for Information on Alcohol and Other Drugs (CAN)**:** 158-163.

Piontek, D., L. Kraus, et al. (2008). "Short scales to assess cannabis-related problems: a review of psychometric properties." Subst Abuse Treat Prev Policy**3**(1): 25.

Piontek, D., L. Kraus, et al. (2009). The ESPAD cannabis module. Cannabis-related problems among adolescents in 17 ESPAD countries. . The 2007 ESPAD report. Substance use among students in 35 European countries. U. G. In B. Hibell, S. Ahlström, O. Balakireva, T. Bjarnason, A. Kokkevi & L. Kraus. Stockholm, The Swedish Council for Information on Alcohol and other Drugs (CAN)**:** 164-170.