

Une nouvelle approche statistique pour la construction d'indices composites de qualité de vie à l'échelle communale

Marie Chavent^{a,b}, Vanessa Kuentz-Simonet^c, Amaury Labenne^c,
Jérôme Saracco^{a,b}

^aUniv. Bordeaux, IMB, UMR 5251

^bINRIA Bordeaux Sud-Ouest, CQFD

^cIRSTEA, UR ADBX

3èmes Rencontres R, Montpellier, 27 juin 2014

Introduction

- Mesure de la qualité de vie via deux approches différentes et complémentaires
 - ▶ Enquête auprès des citoyens : mesure des niveaux de satisfaction de la vie
 - ▶ Analyse des bases de données nationales : création d'indices composites de condition de vie à l'échelle communale

Introduction

- Mesure de la qualité de vie via deux approches différentes et complémentaires
 - ▶ Enquête auprès des citoyens : mesure des niveaux de satisfaction de la vie
 - ▶ Analyse des bases de données nationales : création d'indices composites de condition de vie à l'échelle communale
- Nécessité de créer des indices composites résumant au mieux l'information des données nationales
 - ▶ Méthode d'analyse factorielle pour des variables structurées en groupes
 - ▶ Composantes principales = Combinaison linéaire des variables = INDICE COMPOSITE

Introduction

- Mesure de la qualité de vie via deux approches différentes et complémentaires
 - ▶ Enquête auprès des citoyens : mesure des niveaux de satisfaction de la vie
 - ▶ Analyse des bases de données nationales : création d'indices composites de condition de vie à l'échelle communale
- Nécessité de créer des indices composites résumant au mieux l'information des données nationales
 - ▶ Méthode d'analyse factorielle pour des variables structurées en groupes
 - ▶ Composantes principales = Combinaison linéaire des variables = INDICE COMPOSITE
- Ces méthodes soulèvent plusieurs questions :
 - ▶ Combien de composantes principales retenir ?
 - ▶ Peut-on à l'aide d'une méthode de sélection d'un nombre restreint de variables obtenir des indices qui soient fortement corrélés aux indices calculés sur toutes les variables (facilité d'interprétation)

Présentation des données

Les données réparties en groupes :

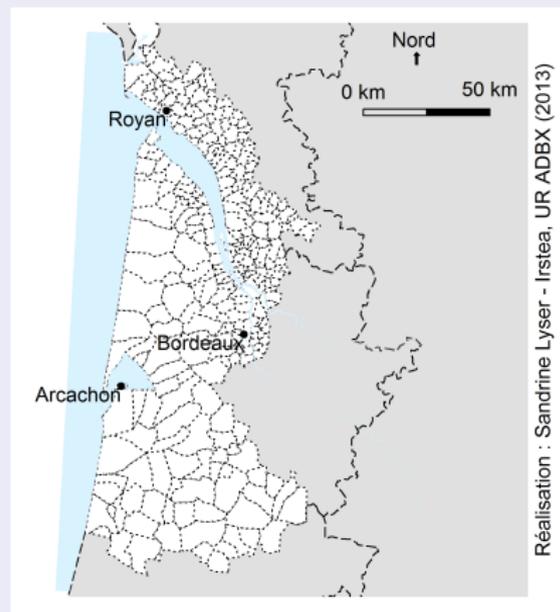
- L'analyse porte sur $n=303$ communes décrites par $p=44$ variables quantitatives et qualitatives.
- Les $p=44$ variables peuvent être réparties en $G=5$ groupes de variables :
 - ▶ Economic conditions,
 - ▶ Living conditions,
 - ▶ Family situations,
 - ▶ Services access,
 - ▶ Natural Environment.

Présentation des données

Les données réparties en groupes :

- L'analyse porte sur $n=303$ communes décrites par $p=44$ variables quantitatives et qualitatives.
- Les $p=44$ variables peuvent être réparties en $G=5$ groupes de variables :
 - ▶ Economic conditions,
 - ▶ Living conditions,
 - ▶ Family situations,
 - ▶ Services access,
 - ▶ Natural Environment.

Carte du territoire étudié :



Sommaire

- 1 La méthode MFAMix pour la création d'indices composites
- 2 Choix du nombre de composantes principales en fonction de la stabilité
- 3 Choix des variables les plus importantes dans la création des indices

La méthode MFAmix (1/5)

L'analyse factorielle multiple

- Analyse factorielle multiple (AFM) pour variables quantitatives : Escofier et Pages (1983).
- AFM pour données quantitatives et qualitatives par groupes : Pages J (2002).
Package FactoMineR : fonction MFA
- Une extension de l'analyse factorielle multiple pour des groupes de variables mixtes : MFAmix. Chavent et al. (2013) 2èmes rencontres R, Lyon
→ Package PCAmixdata

La méthode MFAmix (1/5)

L'analyse factorielle multiple

- Analyse factorielle multiple (AFM) pour variables quantitatives : Escofier et Pages (1983).
- AFM pour données quantitatives et qualitatives par groupes : Pages J (2002).
Package FactoMineR : fonction MFA
- Une extension de l'analyse factorielle multiple pour des groupes de variables mixtes : MFAmix. Chavent et al. (2013) 2èmes rencontres R, Lyon
→ Package PCAmixdata

L'analyse factorielle de données mixtes

- Escofier B, (1979), Traitement simultané de variables qualitatives et quantitatives en analyse factorielle
- Kiers, (1991), Simple structure in component analysis techniques for mixtures of qualitative and quantitative variables
- Pages, (2004), Analyse factorielle de données mixtes
Husson et al. FactoMineR (fonction AFDM)
- Chavent et al. (2012), Orthogonal rotation in PCAMIX

La méthode MFAMix (2/5)

Principe de la méthode :

La méthode MFAMix est basée sur la DVSG de \mathbf{Z} (la matrice des données brutes précédemment recodées) avec les métriques \mathbf{D} pour les individus et \mathbf{M} pour les variables. On a ainsi :

$$\mathbf{Z} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^t, \text{ avec :}$$

- $\mathbf{\Lambda} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})$ la matrice des valeurs singulières de $\mathbf{Z}\mathbf{D}\mathbf{Z}^t\mathbf{M}$ et $\mathbf{Z}^t\mathbf{D}\mathbf{Z}\mathbf{M}$ où r est le rang de \mathbf{Z} ;
- \mathbf{U} la matrice $n \times r$ des vecteurs propres de $\mathbf{Z}\mathbf{M}\mathbf{Z}^t\mathbf{D}$ et $\mathbf{U}^t\mathbf{D}\mathbf{U} = \mathbb{I}_r$;
- \mathbf{V} la matrice $p \times r$ des vecteurs propres de $\mathbf{Z}^t\mathbf{D}\mathbf{Z}\mathbf{M}$ et $\mathbf{V}^t\mathbf{M}\mathbf{V} = \mathbb{I}_r$.

La méthode MFAmix (2/5)

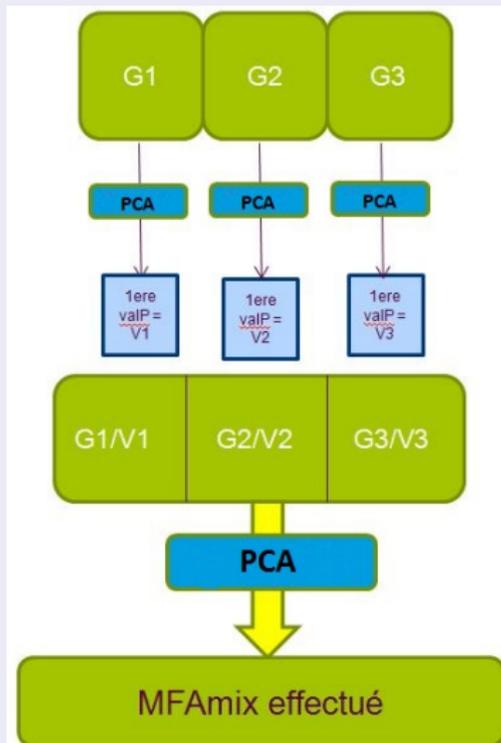
Principe de la méthode :

La méthode MFAmix est basée sur la DVSG de \mathbf{Z} (la matrice des données brutes précédemment recodées) avec les métriques \mathbf{D} pour les individus et \mathbf{M} pour les variables. On a ainsi :

$$\mathbf{Z} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^t, \text{ avec :}$$

- $\mathbf{\Lambda} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})$ la matrice des valeurs singulières de $\mathbf{Z}\mathbf{D}\mathbf{Z}^t\mathbf{M}$ et $\mathbf{Z}^t\mathbf{D}\mathbf{Z}\mathbf{M}$ où r est le rang de \mathbf{Z} ;
- \mathbf{U} la matrice $n \times r$ des vecteurs propres de $\mathbf{Z}\mathbf{M}\mathbf{Z}^t\mathbf{D}$ et $\mathbf{U}^t\mathbf{D}\mathbf{U} = \mathbb{I}_r$;
- \mathbf{V} la matrice $p \times r$ des vecteurs propres de $\mathbf{Z}^t\mathbf{D}\mathbf{Z}\mathbf{M}$ et $\mathbf{V}^t\mathbf{M}\mathbf{V} = \mathbb{I}_r$.

Schéma explicatif



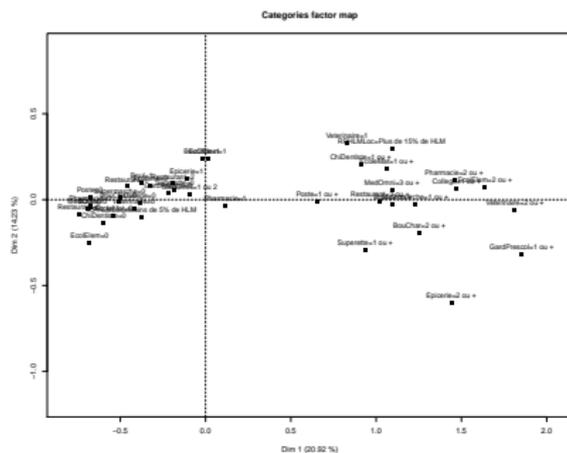
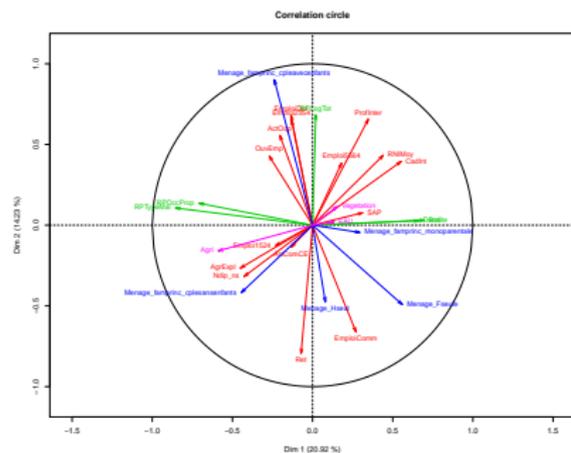
La méthode MFAmix (3/5)

```
library(PCAmixdata)
res.MFA<-MFAmix(data=data_littoral,
                group=vect.group,name.group=name.group,ndim=10,rename.categ=TRUE)

## **Results of the Multiple Factor Analysis for mixed data (MFAmix)**
## The analysis was performed on individuals, described by variables
## *Results are available in the following objects :
##
##   name                description
## 1 "$eig"              "eigenvalues"
## 2 "$separate.analyses" "separate analyses for each group of variables"
## 3 "$group"           "results for all the groups"
## 4 "$partial.axes"    "results for the partial axes"
## 5 "$ind"             "results for the individuals"
## 6 "$quanti.var"      "results for the quantitatives variables"
## 7 "$quali.var"       "results for the categorials variables"
## 8 "$global.pca"      "results for the global PCA"
## 9 "$recap.eig.separate" "ndim first eigenvalues of the separate analyses"
```

La méthode MFAmix (4/5)

```
par(mfrow=c(1,2))  
plot(res.MFA,choice="var",habillage="group",cex=0.9,leg=FALSE,axes=c(1,2))  
plot(res.MFA,choice="ind",invisible="ind",habillage="group",cex=0.9,leg=FALSE,axes=c(1,2))
```



Beaucoup de variables, Difficulté d'interprétation

La méthode MFAMix (5/5)

Conclusions sur MFAMix

- La méthode MFAMix a permis de créer des indices composites (les composantes principales).
- Combien de composantes principales retenir ?
- Ces CP sont des combinaisons linéaires des 44 variables. Il est nécessaire de restreindre le nombre de variables pour faciliter l'interprétation.

Choix du nombre de composantes à retenir (1/3)

- Le nombre de composantes principales (q) à retenir est un problème très souvent abordé en analyse factorielle.

Choix du nombre de composantes à retenir (1/3)

- Le nombre de composantes principales (q) à retenir est un problème très souvent abordé en analyse factorielle.
- Plusieurs méthodes existent dans la littérature :
 - ▶ Besse, P. (1992). PCA stability and choice of dimensionality. *Statistics and Probability Letters*, **13**, 405-410.
 - ▶ Josse, J., Husson, F. (2012). Selecting the number of components in PCA using cross-validation approximations. *Computational Statistics and Data Analysis*, **56**, pp. 1869-1879.

Choix du nombre de composantes à retenir (1/3)

- Le nombre de composantes principales (q) à retenir est un problème très souvent abordé en analyse factorielle.
- Plusieurs méthodes existent dans la littérature :
 - ▶ Besse, P. (1992). PCA stability and choice of dimensionality. *Statistics and Probability Letters*, **13**, 405-410.
 - ▶ Josse, J., Husson, F. (2012). Selecting the number of components in PCA using cross-validation approximations. *Computational Statistics and Data Analysis*, **56**, pp. 1869-1879.
- Ici nous avons choisi d'utiliser la méthode proposée par Besse utilisant une approche Bootstrap.

Choix du nombre de CP à retenir (2/3)

On définit la matrice de projection \mathbf{M} -orthogonale des lignes de \mathbf{Z} sur $E_q = \text{Im}(\mathbf{V}_q)$ (q est le nombre de CP retenu) comme suit :

$$\mathbf{P}_q = \mathbf{V}_q \mathbf{V}_q^t \mathbf{M}$$

Choix du nombre de CP à retenir (2/3)

On définit la matrice de projection \mathbf{M} -orthogonale des lignes de \mathbf{Z} sur $E_q = \text{Im}(\mathbf{V}_q)$ (q est le nombre de CP retenu) comme suit :

$$\mathbf{P}_q = \mathbf{V}_q \mathbf{V}_q^t \mathbf{M}$$

La fonction de perte reposant sur la distance euclidienne entre deux projecteurs orthogonaux est donnée par :

$$\mathcal{L}_q = \frac{1}{2} \|\mathbf{P}_q - \mathbf{P}_q^*\|^2 = q - \text{Tr}(\mathbf{P}_q \mathbf{P}_q^*).$$

Choix du nombre de CP à retenir (2/3)

On définit la matrice de projection \mathbf{M} -orthogonale des lignes de \mathbf{Z} sur $E_q = \text{Im}(\mathbf{V}_q)$ (q est le nombre de CP retenu) comme suit :

$$\mathbf{P}_q = \mathbf{V}_q \mathbf{V}_q^t \mathbf{M}$$

La fonction de perte reposant sur la distance euclidienne entre deux projecteurs orthogonaux est donnée par :

$$\mathcal{L}_q = \frac{1}{2} \|\mathbf{P}_q - \mathbf{P}_q^*\|^2 = q - \text{Tr}(\mathbf{P}_q \mathbf{P}_q^*).$$

Finalement, le risque est défini comme l'espérance de la fonction de perte:

$$R_q = E[\mathcal{L}_q].$$

Choix du nombre de CP à retenir (2/3)

On définit la matrice de projection \mathbf{M} -orthogonale des lignes de \mathbf{Z} sur $E_q = \text{Im}(\mathbf{V}_q)$ (q est le nombre de CP retenu) comme suit :

$$\mathbf{P}_q = \mathbf{V}_q \mathbf{V}_q^t \mathbf{M}$$

La fonction de perte reposant sur la distance euclidienne entre deux projecteurs orthogonaux est donnée par :

$$\mathcal{L}_q = \frac{1}{2} \|\mathbf{P}_q - \mathbf{P}_q^*\|^2 = q - \text{Tr}(\mathbf{P}_q \mathbf{P}_q^*).$$

Finalement, le risque est défini comme l'espérance de la fonction de perte:

$$R_q = E[\mathcal{L}_q].$$

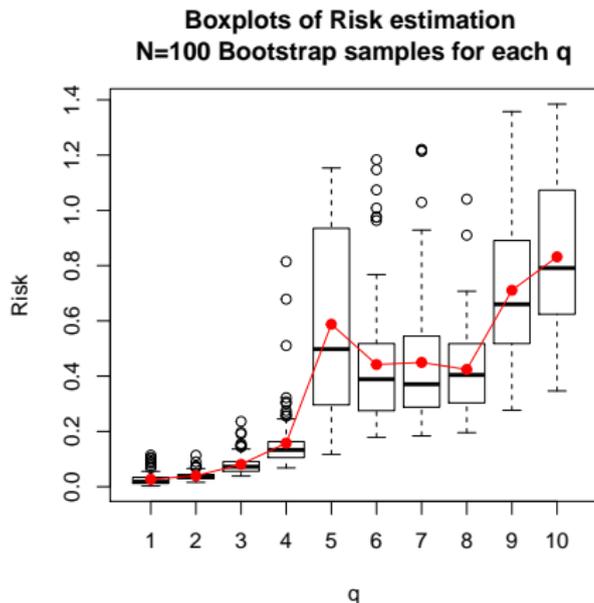
L'idée est d'estimer R_q par un estimateur bootstrap :

$$\widehat{R}_{Bq} = \frac{1}{B} \sum_{b=1}^B \left(q - \text{Tr}(P_q^{*b} P_q) \right) = q - \text{Tr}(P_q^{*(\cdot)} P_q).$$

Où B est le nombre d'échantillons bootstrap, P_q^{*b} est la matrice de projection obtenue avec MFAMix sur le b -ème échantillon bootstrap et $P_q^{*(\cdot)} = \frac{1}{B} \sum_{b=1}^B P_q^{*b}$.

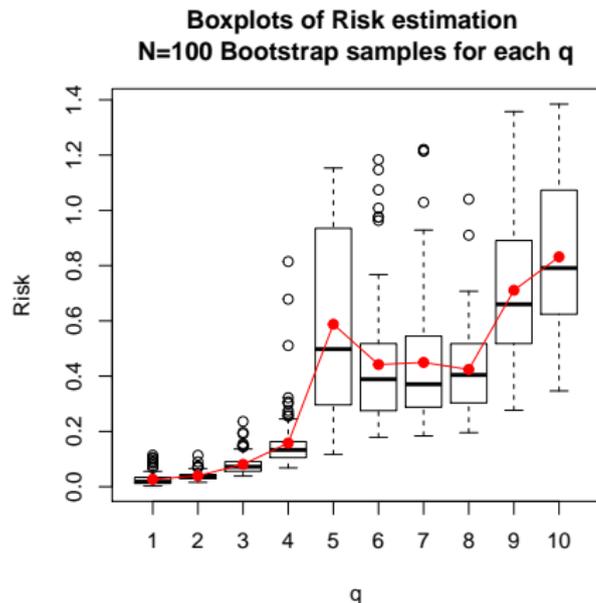
Choix du nombre de CP à retenir (3/3)

```
library(boot)
res.stab<-stability.CP.bootstrap(res.mfa=res.MFA,i,q.max=10,NB=100,graph=TRUE)
```



Choix du nombre de CP à retenir (3/3)

```
library(boot)
res.stab<-stability.CP.bootstrap(res.mfa=res.MFA,i,q.max=10,NB=100,graph=TRUE)
```



On choisit de retenir $q = 3$ composantes.

Reduction du nombre de variables (1/6)

Reduction du nombre de variables (1/6)

Objectif :

Sélectionner un sous-ensemble de $p^* < p$ variables tel que les CP (notées \mathbf{F}^*) calculées sur ce sous-ensemble soient le plus proche des CP de base (notées \mathbf{F}) au sens d'une mesure de liaison.

Reduction du nombre de variables (1/6)

Objectif :

Sélectionner un sous-ensemble de $p^* < p$ variables tel que les CP (notées \mathbf{F}^*) calculées sur ce sous-ensemble soient le plus proche des CP de base (notées \mathbf{F}) au sens d'une mesure de liaison.

- Méthode "Closest Submodel Selection" (CSS) développée par Coudret et al. (2014) dans le cadre de la régression SIR.

Reduction du nombre de variables (1/6)

Objectif :

Sélectionner un sous-ensemble de $p^* < p$ variables tel que les CP (notées \mathbf{F}^*) calculées sur ce sous-ensemble soient le plus proche des CP de base (notées \mathbf{F}) au sens d'une mesure de liaison.

- Méthode "Closest Submodel Selection" (CSS) développée par Coudret et al. (2014) dans le cadre de la régression SIR.
- Nécessité de définir une mesure de liaison entre deux groupes de 3 variables (les CP \mathbf{F} et \mathbf{F}^*)

Reduction du nombre de variables (2/6)

On note :

- $\mathbf{F} = \mathbf{U}\mathbf{\Lambda}$ les CP obtenues grâce à MFAMix sur les p variables,
- $\mathbf{F}^* = \mathbf{U}^*\mathbf{\Lambda}^*$ les CP obtenues grâce à MFAMix réalisé sur les p^* variables.

Reduction du nombre de variables (2/6)

On note :

- $\mathbf{F} = \mathbf{U}\mathbf{\Lambda}$ les CP obtenues grâce à MFAMix sur les p variables,
- $\mathbf{F}^* = \mathbf{U}^*\mathbf{\Lambda}^*$ les CP obtenues grâce à MFAMix réalisé sur les p^* variables.

La mesure de liaison entre \mathbf{F} et \mathbf{F}^* est définie comme suit :

$$\mathcal{D}(\mathbf{F}\mathbf{F}^*) = \frac{1}{q} \text{Tr}(P_F P_{F^*}).$$

Où :

- $P_F = \mathbf{F}\mathbf{F}^T\mathbf{D}$ est la matrice de projection \mathbf{D} -orthogonale sur \mathbf{F} ,
- $P_{F^*} = \mathbf{F}^*\mathbf{F}^{*T}\mathbf{D}$ est la matrice de projection \mathbf{D} -orthogonale sur \mathbf{F}^* .

Reduction du nombre de variables (2/6)

On note :

- $\mathbf{F} = \mathbf{U}\mathbf{\Lambda}$ les CP obtenues grâce à MFAMix sur les p variables,
- $\mathbf{F}^* = \mathbf{U}^*\mathbf{\Lambda}^*$ les CP obtenues grâce à MFAMix réalisé sur les p^* variables.

La mesure de liaison entre \mathbf{F} et \mathbf{F}^* est définie comme suit :

$$\mathcal{D}(\mathbf{F}\mathbf{F}^*) = \frac{1}{q} \text{Tr}(P_{\mathbf{F}}P_{\mathbf{F}^*}).$$

Où :

- $P_{\mathbf{F}} = \mathbf{F}\mathbf{F}^T\mathbf{D}$ est la matrice de projection \mathbf{D} -orthogonale sur \mathbf{F} ,
- $P_{\mathbf{F}^*} = \mathbf{F}^*\mathbf{F}^{*T}\mathbf{D}$ est la matrice de projection \mathbf{D} -orthogonale sur \mathbf{F}^* .

Cette mesure de liaison va nous permettre via la méthode (CSS) de choisir un sous-ensemble de variable permettant d'obtenir des CP \mathbf{F}^* le plus liées aux CP de base \mathbf{F} .

Reduction du nombre de variables (3/6)

La méthode CSS en AFM

Reduction du nombre de variables (3/6)

La méthode CSS en AFM

- Step 1 : Choisir N_0 , le nombre de sous-espaces a évaluer et ζ le pourcentage de sous-espaces retenus parmi les N_0 évalués.

Reduction du nombre de variables (3/6)

La méthode CSS en AFM

- Step 1 : Choisir N_0 , le nombre de sous-espaces a évaluer et ζ le pourcentage de sous-espaces retenus parmi les N_0 évalués.
- Step 2 : Pour $a = 1 \dots N_0$, répéter :

Reduction du nombre de variables (3/6)

La méthode CSS en AFM

- Step 1 : Choisir N_0 , le nombre de sous-espaces à évaluer et ζ le pourcentage de sous-espaces retenus parmi les N_0 évalués.
- Step 2 : Pour $a = 1 \dots N_0$, répéter :
 - ▶ Step 2.1 : Sélectionner aléatoirement p_0 variables parmi les p et construire la matrice $X^{(a)}$ contenant les variables sélectionnées.

Reduction du nombre de variables (3/6)

La méthode CSS en AFM

- Step 1 : Choisir N_0 , le nombre de sous-espaces à évaluer et ζ le pourcentage de sous-espaces retenus parmi les N_0 évalués.
- Step 2 : Pour $a = 1 \dots N_0$, répéter :
 - ▶ Step 2.1 : Sélectionner aléatoirement p_0 variables parmi les p et construire la matrice $X^{(a)}$ contenant les variables sélectionnées.
 - ▶ Step 2.2 : Réaliser MFAMix sur $X^{(a)}$ et calculer $\mathcal{D}(\mathbf{FF}^{(a)})$.

Reduction du nombre de variables (3/6)

La méthode CSS en AFM

- Step 1 : Choisir N_0 , le nombre de sous-espaces à évaluer et ζ le pourcentage de sous-espaces retenus parmi les N_0 évalués.
- Step 2 : Pour $a = 1 \dots N_0$, répéter :
 - ▶ Step 2.1 : Sélectionner aléatoirement p_0 variables parmi les p et construire la matrice $X^{(a)}$ contenant les variables sélectionnées.
 - ▶ Step 2.2 : Réaliser MFAMix sur $X^{(a)}$ et calculer $\mathcal{D}(\mathbf{F}^{(a)})$.
- Step 3 : Retenir les $N_1 = \zeta N_0$ sous-espaces ayant la plus grande mesure de liaison avec \mathbf{F} .

Reduction du nombre de variables (3/6)

La méthode CSS en AFM

- Step 1 : Choisir N_0 , le nombre de sous-espaces à évaluer et ζ le pourcentage de sous-espaces retenus parmi les N_0 évalués.
- Step 2 : Pour $a = 1 \dots N_0$, répéter :
 - ▶ Step 2.1 : Sélectionner aléatoirement p_0 variables parmi les p et construire la matrice $X^{(a)}$ contenant les variables sélectionnées.
 - ▶ Step 2.2 : Réaliser MFAMix sur $X^{(a)}$ et calculer $\mathcal{D}(\mathbf{F}^{(a)})$.
- Step 3 : Retenir les $N_1 = \zeta N_0$ sous-espaces ayant la plus grande mesure de liaison avec \mathbf{F} .
- Step 4 : Compter le nombre de fois où apparaît chaque variable dans les N_1 meilleurs sous-espaces. Ces variables sont retenues pour réaliser MFAMix par la suite.

Reduction du nombre de variables (4/6)

```
reduc.CSS <- reduc.MFAMix.CSS(object = res.MFA, NO = 25000, p0 = 20, zeta = 5/100,  
  ndim = 3, q = 3, graph = FALSE, p.min.group = c(0, 0, 0, 0, 0))
```

Reduction du nombre de variables (4/6)

```
reduc.CSS <- reduc.MFAMix.CSS(object = res.MFA, NO = 25000, p0 = 20, zeta = 5/100,  
  ndim = 3, q = 3, graph = FALSE, p.min.group = c(0, 0, 0, 0, 0))
```

On affiche ici (par ordre décroissant) le nombre de fois où chaque variable est apparue dans les N_1 meilleurs sous-espaces.

Reduction du nombre de variables (4/6)

```
reduc.CSS <- reduc.MFAMix.CSS(object = res.MFA, NO = 25000, p0 = 20, zeta = 5/100,  
  ndim = 3, q = 3, graph = FALSE, p.min.group = c(0, 0, 0, 0, 0))
```

On affiche ici (par ordre décroissant) le nombre de fois où chaque variable est apparue dans les N_1 meilleurs sous-espaces.

```
reduc.CSS$var.chosen[1:20]
```

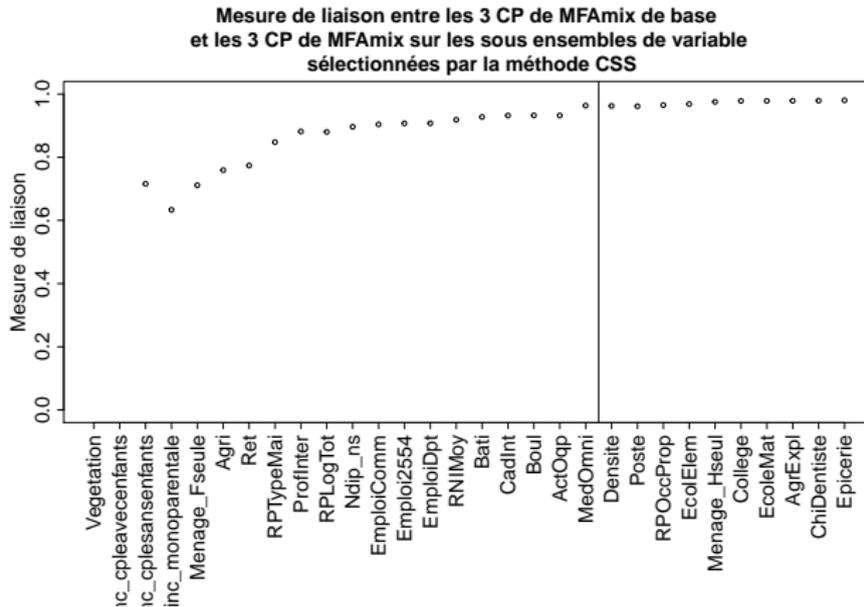
```
##           Vegetation Menage_famprinc_cpleaveenfants  
##           1221                               996  
## Menage_famprinc_cplesansenfants Menage_famprinc_monoparentale  
##           993                               922  
##           Menage_Fseule                               Agri  
##           709                               658  
##           Ret                               RPTypMai  
##           653                               648  
##           ProfInter                               RPLogTot  
##           638                               624  
##           Ndip_ns                               EmploiComm  
##           623                               596  
##           Emploi2554                               EmploiDpt  
##           594                               584  
##           RNIMoy                               Bati  
##           572                               542  
##           CadInt                               Boul  
##           541                               534  
##           ActOqp                               MedOmni  
##           526                               519
```

Reduction du nombre de variables (5/6)

On va regarder plus précisément quelle est la mesure de liaison en fonction du sous-ensemble de variables sélectionnées.

Reduction du nombre de variables (5/6)

On va regarder plus précisément quelle est la mesure de liaison en fonction du sous-ensemble de variables sélectionnées.



Mesure de liaison entre les CP en fonction de l'ensemble de variable (rangées par ordre décroissant d'apparition dans CSS) choisies.

Reduction du nombre de variables (6/6)

Ici, on va relancer MFAMix sur le sous-ensemble des 20 variables qui sont apparues le plus souvent dans la méthode CSS puis interpréter brièvement les résultats.

Reduction du nombre de variables (6/6)

Ici, on va relancer MFAmix sur le sous-ensemble des 20 variables qui sont apparues le plus souvent dans la méthode CSS puis interpréter brièvement les résultats.

```
base.CSS <- data_littoral[, names(CSS.chosen[[2]][1:20])]
MFAmix.CSS <- MFAmix.sub.var(object = res.MFA, data.sub.var = base.CSS, ndim = 3,
  rename.categ = TRUE)
```

Reduction du nombre de variables (6/6)

Ici, on va relancer MFAMix sur le sous-ensemble des 20 variables qui sont apparues le plus souvent dans la méthode CSS puis interpréter brièvement les résultats.

```
base.CSS <- data_littoral[, names(CSS.chosen[[2]][1:20])]
MFAMix.CSS <- MFAMix.sub.var(object = res.MFA, data.sub.var = base.CSS, ndim = 3,
  rename.categ = TRUE)
```

On regarde les corrélations (2 à 2) entre les CP de MFAMix de base et les CP obtenues grâce à la méthode CSS.

```
indice.base<-res.MFA$ind$coord[,1:3]
indice.CSS<-MFAMix.CSS$ind$coord[,1:3]
cor(indice.CSS,indice.base)
```

```
##           dim 1    dim 2    dim 3
## dim 1 -0.97370  0.12667 -0.06917
## dim 2  0.12105  0.97457  0.04265
## dim 3 -0.07968 -0.05191  0.97271
```

On voit que les CP sont très liées, on peut donc relancer MFAMix sur les 20 variables choisies sans perdre trop d'information.

Interprétation sommaire des indices synthétiques (1/5)

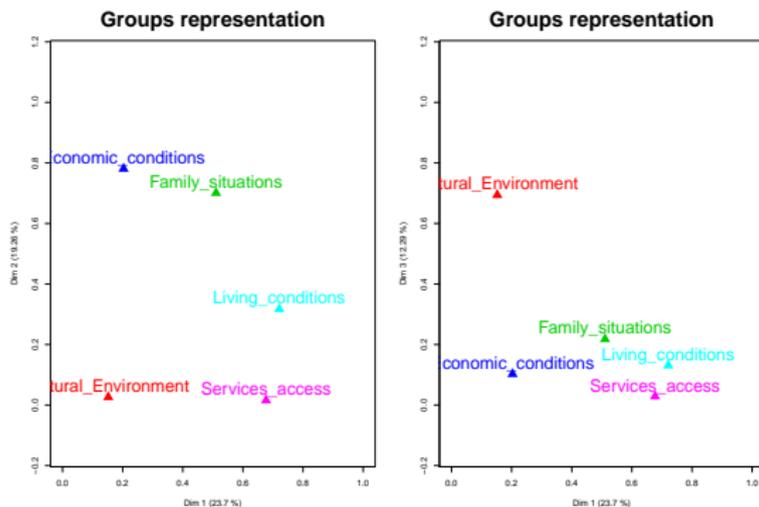


Figure : Représentation des groupes sur les deux premiers plans factoriels

Interprétation sommaire des indices synthétiques (1/5)

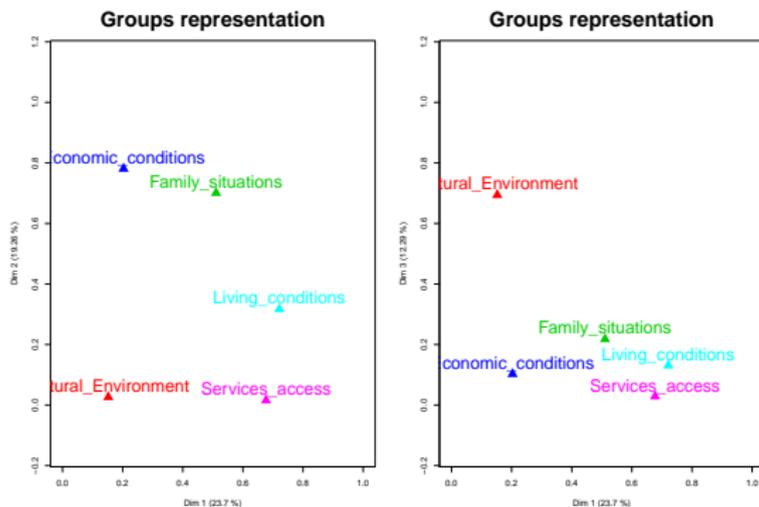


Figure : Représentation des groupes sur les deux premiers plans factoriels

Axe 1 : Conditions de logement et accès aux services

Axe 2 : Conditions économiques et familiales

Axe 3 : Environnement

Interprétation sommaire des indices synthétiques (2/5)

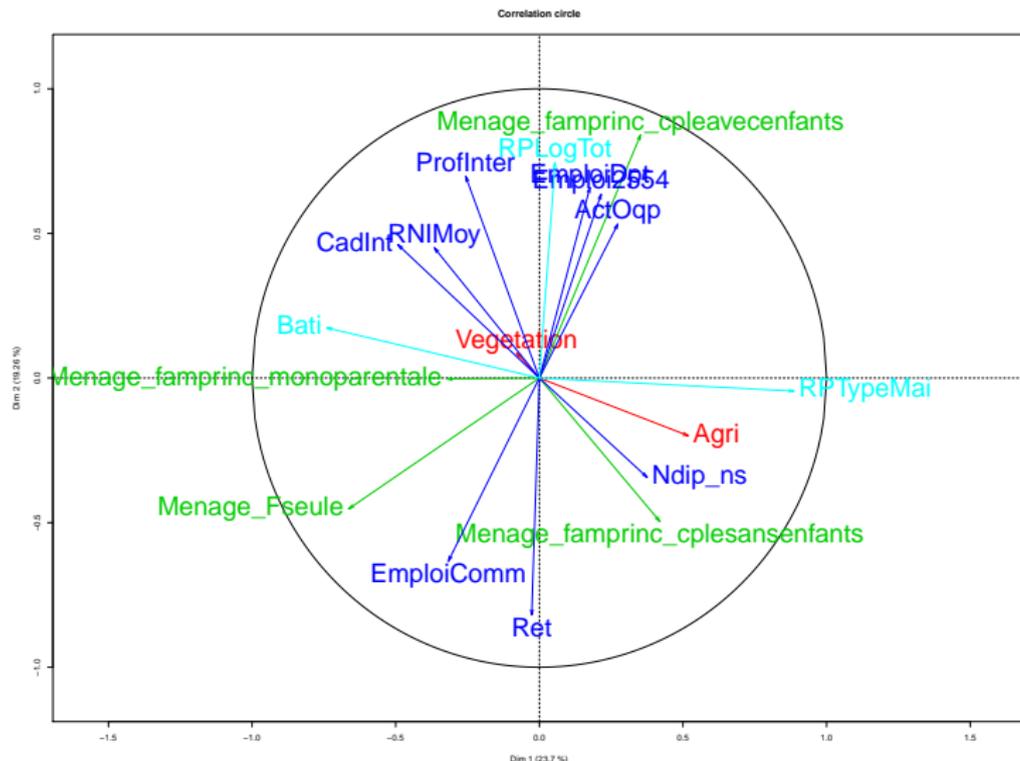


Figure : Cercle des corrélations des variables quantitatives sur le plan (1,2)

Interprétation sommaire des indices synthétiques (3/5)

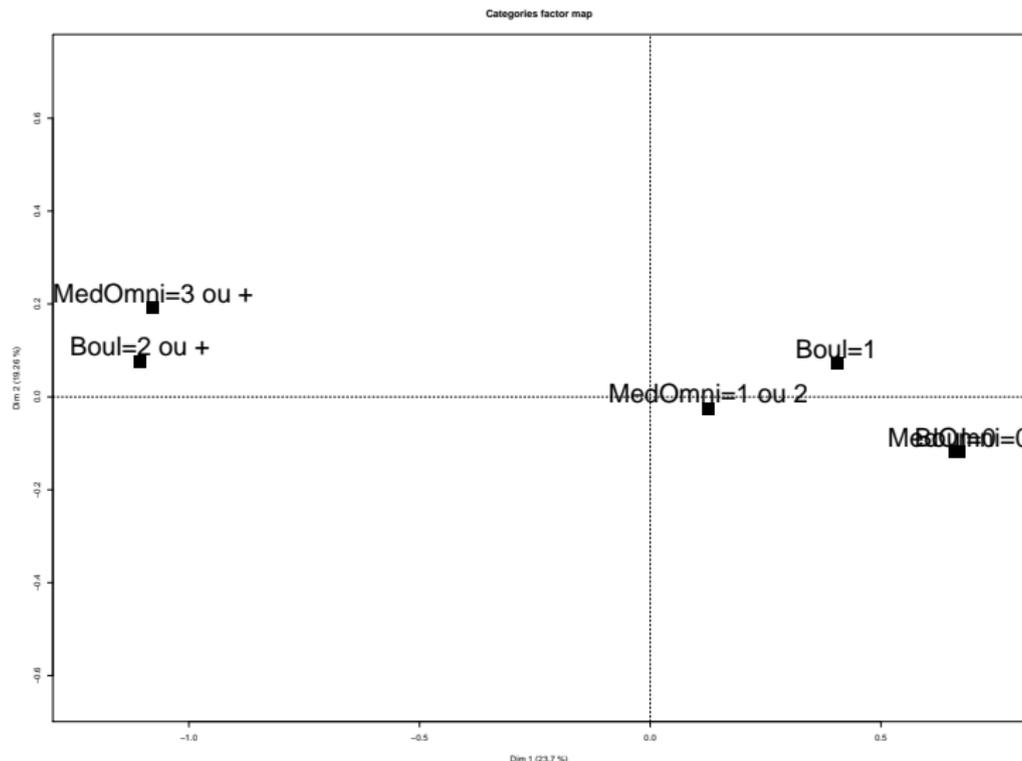


Figure : Représentation des modalités des variables qualitatives sur le plan (1,2)

Interprétation sommaire des indices synthétiques (4/5)

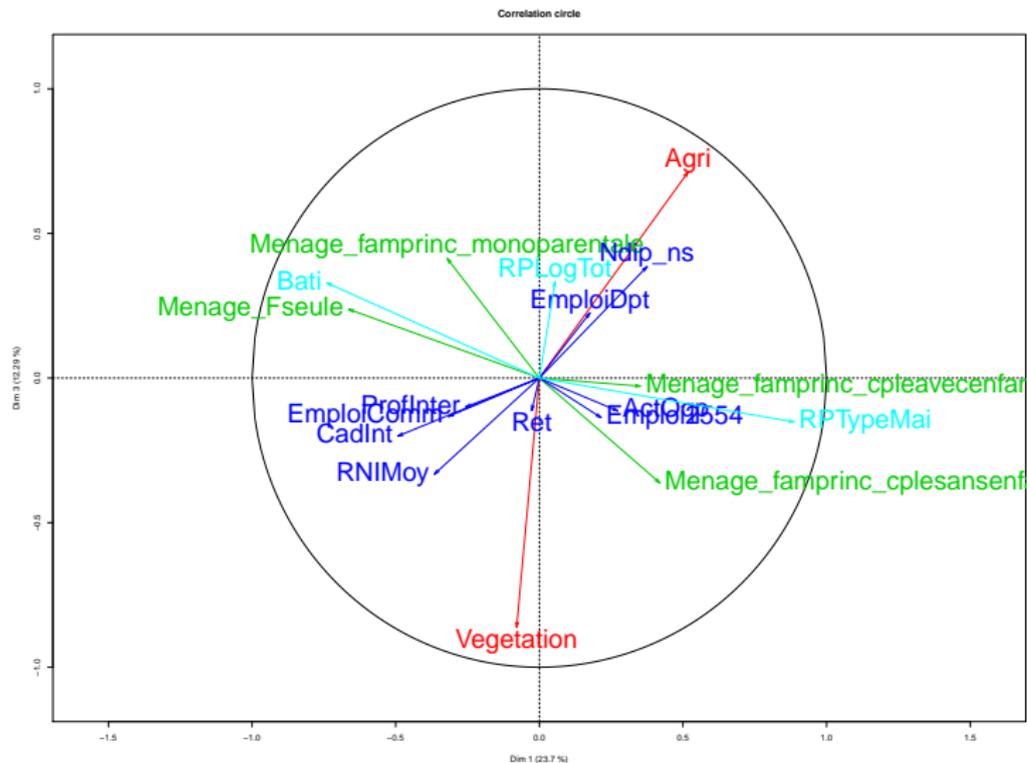


Figure : Cercle des corrélations des variables quantitatives sur le plan (1,3)

Interprétation sommaire des indices synthétiques (5/5)

Interprétation sommaire des indices synthétiques (5/5)

- **L'indice 1** oppose les communes où l'on trouve principalement des habitations de type maison avec peu de services aux communes fortement bâties avec une forte proportion d'appartements et un bon accès aux services et une proportion supérieure à la moyenne de femmes seules. Cet indice caractérise principalement l'accès aux services et les conditions de logement.

Interprétation sommaire des indices synthétiques (5/5)

- **L'indice 1** oppose les communes où l'on trouve principalement des habitations de type maison avec peu de services aux communes fortement bâties avec une forte proportion d'appartements et un bon accès aux services et une proportion supérieure à la moyenne de femmes seules. Cet indice caractérise principalement l'accès aux services et les conditions de logement.
- **L'indice 2** oppose d'une part, les communes avec une forte proportion de logements occupés par les propriétaires qui sont principalement des familles avec enfants occupant des emplois intermédiaires à l'extérieur de leurs communes. Et d'autre part les communes majoritairement occupées par des retraitées et par des gens travaillant au sein de leur commune de résidence.

Interprétation sommaire des indices synthétiques (5/5)

- **L'indice 1** oppose les communes où l'on trouve principalement des habitations de type maison avec peu de services aux communes fortement bâties avec une forte proportion d'appartements et un bon accès aux services et une proportion supérieure à la moyenne de femmes seules. Cet indice caractérise principalement l'accès aux services et les conditions de logement.
- **L'indice 2** oppose d'une part, les communes avec une forte proportion de logements occupés par les propriétaires qui sont principalement des familles avec enfants occupant des emplois intermédiaires à l'extérieur de leurs communes. Et d'autre part les communes majoritairement occupées par des retraitées et par des gens travaillant au sein de leur commune de résidence.
- **L'indice 3** oppose les communes de type agricole aux communes plus végétalisées (territoires forestiers).

Conclusion

Conclusion

- La méthode MFAMix permet de réaliser une AFM lorsque les variables sont mixtes à l'intérieur des groupes. Cette méthode permet de créer des indices synthétiques (les composantes principales).

Conclusion

- La méthode MFAMix permet de réaliser une AFM lorsque les variables sont mixtes à l'intérieur des groupes. Cette méthode permet de créer des indices synthétiques (les composantes principales).
- L'utilisation d'une autre approche que l'ébouilissement des valeurs propres nous a permis de choisir un nombre de composantes principales à interpréter.

Conclusion

- La méthode MFAMix permet de réaliser une AFM lorsque les variables sont mixtes à l'intérieur des groupes. Cette méthode permet de créer des indices synthétiques (les composantes principales).
- L'utilisation d'une autre approche que l'ébouilissement des valeurs propres nous a permis de choisir un nombre de composantes principales à interpréter.
- L'utilisation d'une méthode de sélection de variables nous a permis de créer des indices fortement corrélés aux indices de base mais plus simple à interpréter et à diffuser.

Conclusion

- La méthode MFAMix permet de réaliser une AFM lorsque les variables sont mixtes à l'intérieur des groupes. Cette méthode permet de créer des indices synthétiques (les composantes principales).
- L'utilisation d'une autre approche que l'ébouilisé des valeurs propres nous a permis de choisir un nombre de composantes principales à interpréter.
- L'utilisation d'une méthode de sélection de variables nous a permis de créer des indices fortement corrélés aux indices de base mais plus simple à interpréter et à diffuser.
- Les fonctions MFAMix et PCAMix sont disponibles dans le Package PCAMixdata disponible sur le CRAN.
Les autres fonctions de sélection de variables seront intégrées ultérieurement mais sont disponibles auprès des auteurs.

Merci de votre attention

Quelques références

- 1 Noll, Heinz-Herbert (2006). Towards a European System of Social Indicators: Theoretical Framework and System Architecture. *Social Indicators Research*.
- 2 Escofier B et Pagès J (1983), Méthode pour l'analyse de plusieurs groupes de variables. Application à la caractérisation des vins rouges du Val de Loire, *Revue de statistique appliquée*, 31(2) : 43-59.
- 3 Chavent, M., Kuentz-Simonet, V., Labenne, A., Saracco, J. (2013). Une extension de l'analyse factorielle multiple pour des groupes de variables mixtes : MFAMix. *45èmes Journées de la Statistique, Toulouse*.
- 4 Besse, P. (1992). PCA stability and choice of dimensionality. *Statistics and Probability Letters*, **13**, 405-410.
- 5 Josse, J., Husson, F. (2012). Selecting the number of components in PCA using cross-validation approximations. *Computational Statistics and Data Analysis*, **56**, pp. 1869-1879.
- 6 Coudret, R., Liquet, B., Saracco, J. (2014). Comparison of sliced inverse regression method approaches for undetermined cases. *Journal de la Société française de Statistique*, in press.