

Méthodes d'analyse des données multidimensionnelles avec R Commander et Factominer

1 Installer les composants logiciels nécessaires

1.1 R Commander et FactoMineR sur les appareils de la salle de TD

Le package FactoMineR et le menu FactoMineR de R Commander ont été installés sur (presque tous) les appareils de la salle A206. Il suffit donc de charger R (de préférence à l'aide du raccourci disponible sur le site Web) puis d'utiliser le menu Packages > Charger le package ... > Rcmdr.

1.2 R Commander et FactoMineR sur un appareil personnel

On se reportera au polycopié de L3 (EC PSY54A) pour l'installation de R Commander.

1.2.1 Installation automatique

On peut ensuite installer FactoMineR en se rendant sur le site : <http://factominer.free.fr/> et en suivant les instructions données dans le paragraphe : "Installing FactoMineR and its Graphical User Interface". Mais ce procédé d'installation peut échouer, car il ne teste pas suffisamment la configuration de la machine sur laquelle le package doit être installé.

Notamment, cette méthode est à proscrire si votre appareil ne fonctionne pas sous Windows.

1.2.2 Installation en partie manuelle

On peut aussi utiliser le menu Packages > Installer le(s) Package(s) de la console de R pour installer FactoMineR et les packages qu'il utilise (au moins "ellipse", "scatterplot3d").

Ensuite, on télécharge les deux fichiers :

<http://factominer.free.fr/Rcmdr-facto-fr.r> et <http://factominer.free.fr/add-menu-facto-fr.txt>

Pour installer l'interface utilisateur (c'est-à-dire le menu), affichez le contenu du répertoire \etc relatif à R Commander (par exemple : C:\Program Files\R\R-2.15.1\library\Rcmdr\etc, plus généralement :

<Répertoire d'installation de R>\library\Rcmdr\etc) et faites les opérations suivantes :

- Faites une copie de sauvegarde de <Répertoire d'installation de R>\library\Rcmdr\etc\ cmdr-menus.txt
- Ouvrez le fichier <Répertoire d'installation de R>\library\Rcmdr\etc\Rcmdr-menus.txt à l'aide d'un éditeur de texte et ajoutez en fin de fichier le contenu du fichier add-menu-facto-fr.txt qui a été téléchargé. Enregistrez la nouvelle version. N.B. Si vous utilisez un système autre que Windows, il est préférable de commencer par modifier le fichier, en supprimant notamment tous les caractères accentués dans les intitulés des menus.
- Copiez dans le répertoire <Répertoire d'installation de R>\library\Rcmdr\etc\ le fichier Rcmdr-facto-fr.r

En cas de messages d'erreur au chargement de R Commander :

- vérifiez dans la console si l'erreur n'est pas due à des packages manquants. Si c'est le cas, installez ces packages ;

- si l'erreur est d'une autre nature, il est possible de revenir en arrière en supprimant le fichier Rcmdr-facto-fr.r qui a été copié dans <Répertoire d'installation de R>\library\Rcmdr\etc\ et en remplaçant le fichier <Répertoire d'installation de R>\library\Rcmdr\etc\Rcmdr.txt par la copie de sauvegarde qui avait été faite.

En dernier recours, on peut aussi supprimer le répertoire <Répertoire d'installation de R>\library\Rcmdr\ et réinstaller R Commander.

2 ACP : résultats fournis par Factominer

2.1 Exemple

Source : Philippe TISON. La dépression du sujet âgé. Hypothèse de schémas cognitifs spécifiques. Thèse de doctorat de psychologie soutenue le 23 juin 2003. Accessible en ligne à l'adresse : http://documents.univ-lille3.fr/files/pub/www/recherche/theses/tison-philippe/html/these_body.html

Dans le cadre de l'étude citée ci-dessus, l'auteur considère trois instruments d'évaluation utilisés en ce qui concerne les pathologies dépressives : GDS, réellement spécifique aux sujets âgés, HDRS et MADRS qui sont d'une spécificité acceptable.

- La GDS (Geriatric Depression Scale) de Yesavage et Brink (1983) est un questionnaire d'évaluation en 30 items, à réponse oui-non avec possibilité d'auto et d'hétéro-évaluation. On la retrouve parfois traduite sous l'expression d'échelle de dépression gériatrique. Elle permet le dépistage de la dépression chez la personne âgée.
- L'HDRS (Hamilton Depression Rating Scale) de Hamilton (1967), est une des échelles les plus utilisées dans le monde pour l'hétéro-évaluation de la dépression. Cette échelle est un bon indicateur de l'intensité globale du syndrome dépressif, suffisamment sensible pour apprécier le changement sous traitement antidépresseur. Ce n'est pas un instrument de diagnostic.
- La MADRS (Montgomery and Asberg Depression Rating Scale) de Montgomery et Asberg (1979), est également un instrument souvent utilisé. Cette échelle fournit un bon indice de gravité globale de la dépression du fait de sa grande sensibilité de changements aux traitements. Cette échelle semble plus sensible et plus fidèle que l'HDRS.

Par ailleurs, l'auteur considère la triade cognitive de Beck. Pour Beck et coll. (1979), la dépression s'explique par la présence et l'activité constante de schémas cognitifs négatifs, qui orientent le traitement des informations externes dans un sens toujours négatif. En résultent donc des pensées négatives sur soi, sur le monde et sur le futur (triade cognitive de Beck), et par conséquent des émotions de tristesse. Selon l'auteur, de manière empirique, nous retrouvons en effet assez facilement la triade cognitive de Beck dans le discours des sujets âgés déprimés, mais nous trouvons également 2 cognitions qui n'entrent pas clairement dans cette classification : il s'agit de l'idée d'un changement impossible avec l'âge (CIA) « je suis trop vieux pour changer », et l'idée d'une banalisation des symptômes dépressifs avec l'âge (BSD) « c'est normal d'être déprimé à mon âge ». Ces 2 cognitions semblent fréquentes et paraissent dépasser un cadre anecdotique. Ces productions mentales étant bien issues de raisonnements, d'interprétations établies à partir de faits de l'environnement, donc correspondant bien à des pensées négatives de patients déprimés, il semblait alors logique qu'elles puissent être correctement réparties dans le système de la triade cognitive de Beck (TCB). Or, CIA et BSD sont deux cognitions qui n'entrent pas clairement dans la TCB, et qui semblent se différencier légèrement des pensées typiques et habituelles de la TCB.

L'auteur a mené une étude sur un échantillon de 100 sujets âgés déprimés. Ces sujets sont évalués d'une part sur les 3 échelles GDS, HDRS, MADRS et d'autre part sur des échelles évaluant la TCB, le CIA et la

BSD, à l'aide d'un questionnaire de cognitions spécifiques. Pour chacune de ces échelles, un score est d'autant plus élevé qu'il correspond à une dépression plus profonde ou à des cognitions plus affirmées.

L'auteur analyse ensuite les liens entre les 7 variables observées sur les sujets (les 6 échelles précédentes et la variable Age) à l'aide d'une analyse en composantes principales normée.

2.2 Traitement sous Factominer

Chargez le jeu de données Depression.Age.RData

Vérifiez que les paramètres descriptifs et les corrélations entre les 7 variables sont les suivants :

- Moyennes et écarts types pour les 7 variables :

	N Actifs	Moyenne	Ecart-type
Age	100	79,46	8,25
GDS	100	18,56	4,19
HDRS	100	16,91	5,12
MADRS	100	22,58	5,92
CIA	100	4,41	1,54
BSD	100	3,36	1,63
TCB	100	3,59	1,57

- Corrélations entre les variables :

	Age	GDS	HDRS	MADRS	CIA	BSD	TCB
Age	1,00	-0,03	-0,13	-0,16	0,20	0,15	0,10
GDS	-0,03	1,00	0,14	0,40	0,32	0,44	0,39
HDRS	-0,13	0,14	1,00	0,60	-0,11	-0,18	-0,01
MADRS	-0,16	0,40	0,60	1,00	-0,02	-0,01	0,20
CIA	0,20	0,32	-0,11	-0,02	1,00	0,21	0,33
BSD	0,15	0,44	-0,18	-0,01	0,21	1,00	0,29
TCB	0,10	0,39	-0,01	0,20	0,33	0,29	1,00

Utilisez ces résultats pour répondre aux questions suivantes :

- Examiner et commenter le tableau des corrélations entre les variables. Quelles sont les corrélations qui sont significatives, sachant que, pour un échantillon de taille 100, la valeur critique du coefficient de corrélation au seuil de 5% bilatéral est $r_{crit} = 0,1946$?

- Préalablement à son étude, l'auteur a fait les hypothèses de recherche suivantes :

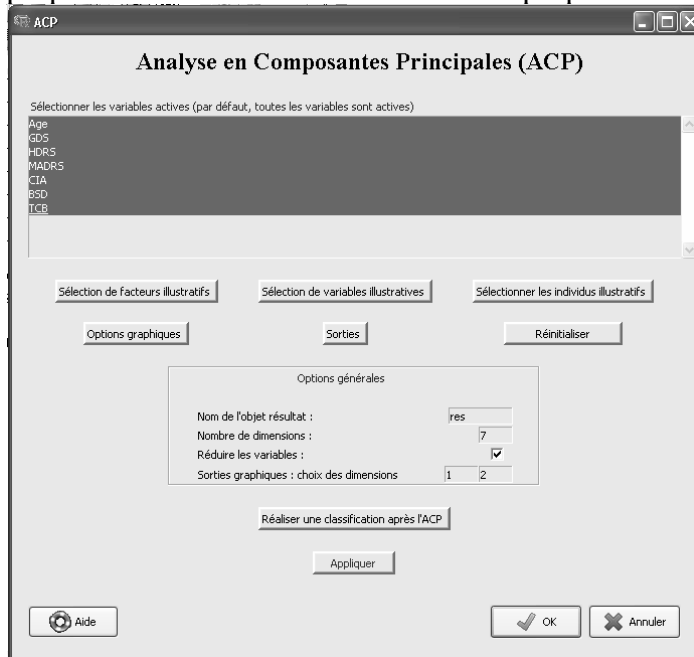
- L'adhésion aux cognitions de la Triade Cognitive de Beck (TCB), du « changement impossible avec l'âge » (CIA), et de la « banalisation des symptômes dépressifs avec l'âge » (BSD) est différente selon l'état dépressif des sujets âgés.

- Le niveau d'adhésion aux cognitions est fonction de la sévérité de la dépression estimée à l'HDRS.

Dans quelle mesure l'étude des corrélations confirme-t-elle ces hypothèses de recherche ?

Effectuez ensuite une analyse en composantes principales normée sur les 7 variables. On peut commencer par réaliser l'étude sur 7 axes, en n'enregistrant pas les résultats numériques :

Pour cela, utilisez le menu Factominer > Analyse en composantes principales
Sélectionnez les 7 variables numériques. Utilisez le bouton "Options graphiques" pour supprimer les sorties graphiques et le bouton "Sorties" pour demander l'écriture des données relatives aux valeurs propres dans un fichier nommé Valeurs-propres-ACP.csv :



On obtient le résultat suivant, qui nous conduit à étudier 2 ou 3 axes :

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	2,120921888	30,29888411	30,29888411
comp 2	1,78378799	25,48268557	55,78156968
comp 3	0,931848987	13,31212838	69,09369806
comp 4	0,784427575	11,20610821	80,29980627
comp 5	0,63993645	9,141949286	89,44175556
comp 6	0,419604357	5,99434796	95,43610352
comp 7	0,319472754	4,563896481	100

Refaites ensuite l'analyse, en indiquant 3 axes.

Ouvrez le fichier Depression.Age.ACP.csv à l'aide d'Excel ou de LibreOffice Calc et utilisez le menu Recherche/Remplacement pour remplacer les points par des virgules, afin que les données soient reconnues comme données numériques.

Outre les graphiques des variables et des individus, et les résultats relatifs aux valeurs propres, on obtient les résultats suivants :

Pour les variables :

Les coordonnées :

coord	Dim.1	Dim.2	Dim.3
Age	0,1219	-0,4836	0,7569
...

Les corrélations entre les variables et les composantes principales, identiques aux coordonnées :

cor	Dim.1	Dim.2	Dim.3

Age	0,1219	-0,4836	0,7569
...

Les qualités de représentation (cosinus²)

cos2			
	Dim.1	Dim.2	Dim.3
Age	0,0149	0,2339	0,5729
...

Les contributions à l'inertie de chacun des axes

contrib			
	Dim.1	Dim.2	Dim.3
Age	0,7009	13,11	61,4845
...

Des résultats analogues (coordonnées, cosinus2 et contributions) pour les individus :

coord			
	Dim.1	Dim.2	Dim.3
1	1,7086	2,4971	0,4625
2

cos2			
	Dim.1	Dim.2	Dim.3
1	0,2861	0,6111	0,021
2

contrib			
	Dim.1	Dim.2	Dim.3
1	1,3765	3,4955	0,2295
2

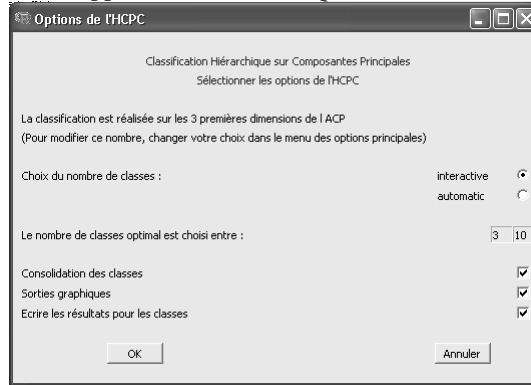
Les distances à l'origine de chacun des individus :

dist				
3,194305	2,061786	2,485538	2,546475	...

Le logiciel donne ensuite des résultats relatifs aux dimensions en donnant pour chacune d'entre elles, la liste des variables significativement corrélées avec la composante concernée, ainsi que la p-value correspondante :

Dim.1		
quanti		
	correlation	p.value
GDS	0,8288	0
TCB	0,6991	0
BSD	0,5879	0
CIA	0,5456	0
MADRS	0,4963	0
HDRS	0,2021	0,0438

Le logiciel permet également d'effectuer une classification des individus à partir des résultats de l'ACP. Reprenez le même menu et cliquez sur le bouton "Réaliser une classification après l'ACP" :



Demandez l'écriture des résultats numériques dans le fichier Depression.Age.ACP2.csv. Dans le graphique "Hierarchical Clustering", cliquez de façon à indiquer le nombre de classes souhaité (par exemple 4 classes).

On obtient alors de nombreux résultats complémentaires dans la fenêtre de sortie :

- La classe dans laquelle est rangé chaque individu :

```

    clust
54      4
85      4
...

```

Le coefficient eta-2 est calculé entre la variable qualitative "classe d'appartenance" et chacune des variables numériques qui ont participé à l'ACP. La p-value indiquée est celle que l'on obtiendrait si l'on faisait une analyse de variance à un facteur en utilisant l'appartenance aux classes comme variable indépendante et la variable numérique concernée comme variable dépendante :

```

> res.hcpc$desc.var
$quanti.var
      Eta2      P-value
MADRS 0.5768872 7.081087e-18
GDS    0.5360300 5.706143e-16
TCB    0.3995200 1.179401e-10
HDRS   0.3549011 3.478603e-09
BSD    0.3069478 1.014909e-07
CIA    0.2800726 6.038341e-07
Age    0.2217425 2.280821e-05

```

Les valeurs tests pour chaque variable et chaque classe.

```

$quanti
$quanti$`1`
      v.test Mean in category Overall mean sd in category Overall sd      p.value
GDS  5.662286      23.625018      18.56      3.339481      4.17  1.493692e-08
MADRS 5.397334      29.399423      22.58      4.167307      5.89  6.763858e-08
TCB   4.462470       5.083321       3.59      1.425367      1.56  8.102011e-06
HDRS  3.413416      20.637007      16.91      4.192585      5.09  6.415382e-04
CIA   2.416589       5.203135       4.41      1.265161      1.53  1.566669e-02
BSD   2.343868       4.174519       3.36      1.258381      1.62  1.908493e-02
...

```

Les tests menés correspondent à l'idée suivante : dans une population d'effectif N (ici N=100), on tire au hasard sans remise un échantillon de taille n (ici, n est l'effectif de la classe, n=18). Quelle est la probabilité d'observer un écart au moins aussi élevé entre la moyenne d'échantillon et la moyenne de la population ? Autrement dit, on peut énoncer H0 et H1 sous la forme :

H0 : L'échantillon est tiré au hasard dans la population observée.

H1 : L'échantillon n'est pas tiré au hasard dans la population observée.

Le principe du calcul est le suivant : pour chaque variable et chaque classe, on calcule la moyenne \bar{x}_c des valeurs de la variable sur les individus constituant la classe, la moyenne sur l'ensemble des individus \bar{x} , l'écart type (non corrigé) s_c des valeurs de la variable sur les individus constituant la classe et l'écart type (non corrigé) s de l'ensemble des valeurs sur tous les individus des observations. On calcule ensuite la valeur test selon la formule donnée par Escoffier et Pagès :

$$v.test = \sqrt{\frac{N_c(N-1)}{N-N_c} \frac{\bar{x}_c - \bar{x}}{s}}$$

On calcule ensuite la p-value correspondante pour un test bilatéral, en utilisant une loi normale centrée réduite. Cette p-value indique si la moyenne de la classe est significativement différente de celle de l'ensemble, pour la variable considérée.

Des calculs analogues sont menés sur les composantes principales (coordonnées des individus sur les composantes principales) :

```
attr(,"class")
[1] "catdes" "list "
```

```
> res.hcpc$desc.axes
$quanti.var
      Eta2      P-value
Dim.1 0.7500547 8.551039e-29
Dim.2 0.5856780 2.603760e-18
```

```
$quanti
$quanti$`1`
      v.test Mean in category Overall mean sd in category Overall sd p.value
Dim.1 6.495502      2.0292154 -3.609960e-16      0.8351718      1.456339 8.275688e-11
Dim.2 2.895480      0.8295552 -3.199006e-16      1.0246548      1.335585 3.785791e-03
...
```

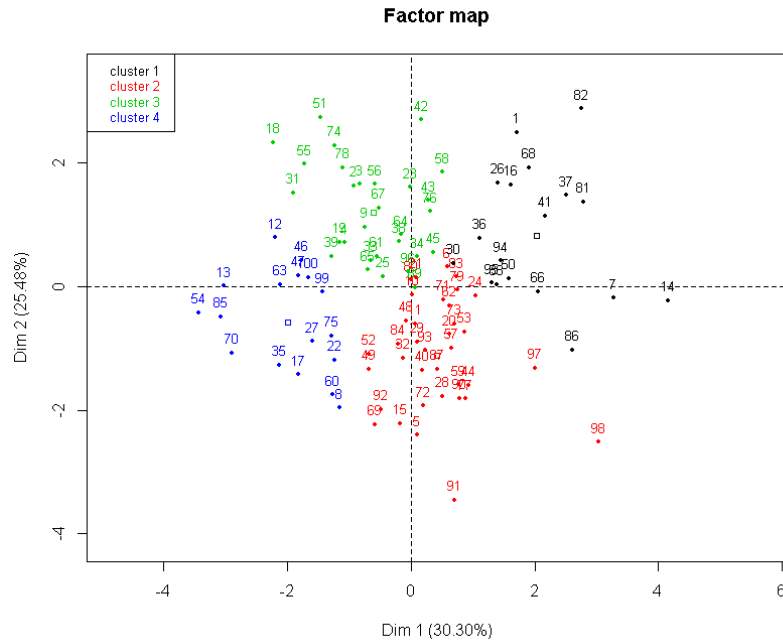
Pour chaque classe et les individus qui lui appartiennent, on indique la distance de l'individu au centre de classe, distance calculée à partir des coordonnées factorielles de l'individu :

```
attr(,"class")
[1] "catdes" "list "
```

```
> res.hcpc$desc.ind
$para
cluster: 1
      41      36      26      66      16
0.5902048 1.0271860 1.0941814 1.1584004 1.1913318
-----
...
```

Enfin, pour chaque classe et les individus qui lui appartiennent, on indique le minimum des distances de cet individu aux centres des autres classes :

```
$dist
cluster: 1
      37      14      82      81      7
4.020251 3.962107 3.801437 3.566495 3.144243
-----
...
```

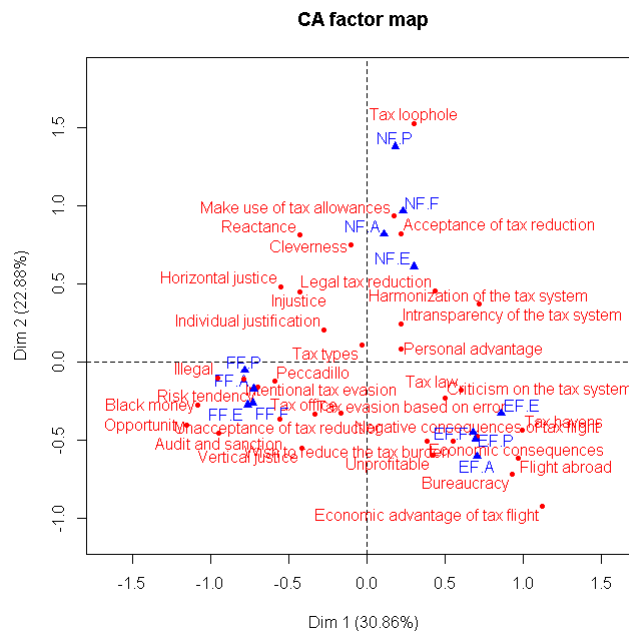


3 Analyse Factorielle des Correspondances : résultats fournis par Factominer

Le menu Factominer de R Commander permet de réaliser une AFC à partir d'un tableau de contingence. Pour explorer les possibilités fournies par le logiciel, nous allons utiliser l'exercice 2.4.5 p. 71 des feuilles de TD.

Chargez le jeu de données Tax.Avoidance.RData et observez la façon dont les données ont été structurées. Les données ont été importées depuis un fichier Excel. Le menu Données > Jeu de données actif > Nom des cas... a été utilisé pour nommer les modalités lignes d'après les libellés des 34 catégories sémantiques.

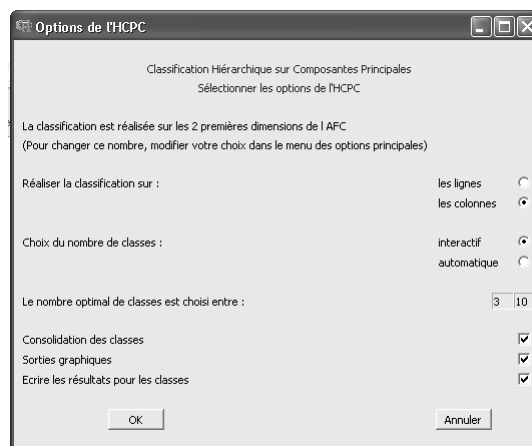
Réalisez une AFC à l'aide du menu FactoMineR > Analyse Factorielle des Correspondances (AFC). On garde toutes les lignes et toutes les colonnes comme individus actifs et on choisit de mener l'étude sur les deux premiers axes factoriels. Comme précédemment, on peut demander (bouton "Sorties") que les résultats numériques soient écrits dans un fichier .csv nommé par exemple Tax.Avoidance.AFC.csv.



Examinons les résultats numériques produits par le logiciel. Nous retrouvons les valeurs propres, les coordonnées, les contributions à l'inertie et les cos2 des modalités lignes et colonnes ainsi que l'inertie (absolue) de ces modalités dans l'espace multidimensionnel. On pourra vérifier par exemple que la somme des inerties des modalités colonnes est égale à l'inertie totale (Φ^2), c'est-à-dire à la somme des valeurs propres.

A la suite de ces résultats, le logiciel classe les modalités lignes et les modalités colonnes par coordonnée croissante, sur chaque dimension.

On peut également demander que soit effectuée une classification sur les lignes ou sur les colonnes, à partir des coordonnées dans le premier plan factoriel :



Nous obtiendrons des résultats sans surprise en choisissant de faire une classification des colonnes et en choisissant 3 classes dans le graphique interactif. Les résultats numériques sont fournis dans la fenêtre de sortie :

Appartenance des 12 modalités colonnes aux différentes classes

```
> res.hcpc$data.clust[,ncol(res.hcpc$data.clust),drop=F]
      clust
FF.P      1
FF.E      1
FF.F      1
FF.A      1
```

```
NF.A 3
NF.P 3
NF.F 3
NF.E 3
EF.F 2
EF.P 2
EF.A 2
EF.E 2
```

Coefficients eta-2 entre les classes et les coordonnées des modalités sur les deux dimensions

```
> res.hcpc$desc.var
$quanti.var
      Eta2      P-value
Dim.1 0.9906833 7.272556e-10
Dim.2 0.9203743 1.134331e-05
```

Pour la première classe, seule la première dimension fournit une valeur test significative

```
$quanti
$quanti$`1`
      v.test Mean in category Overall mean sd in category Overall sd
Dim.1 -3.089931      -0.7526344  0.06041878  0.02339161  0.6170943
      p.value
Dim.1 0.00200203
```

Pour la deuxième classe, les deux dimensions fournissent des valeurs tests significatives

```
$quanti$`2`
      v.test Mean in category Overall mean sd in category Overall sd
Dim.1  2.551137      0.7316993  0.06041878  0.07231953  0.6170943
Dim.2 -2.076352     -0.4716526  0.09225827  0.09866733  0.6369286
      p.value
Dim.1 0.01073720
Dim.2 0.03786138
```

Pour la troisième classe, seule la deuxième dimension fournit une valeur test significative

```
$quanti$`3`
      v.test Mean in category Overall mean sd in category Overall sd
Dim.2  3.126156      0.9412824  0.09225827  0.2816268  0.6369286
      p.value
Dim.2 0.001771077
```

```
attr(,"class")
[1] "catdes" "list "
```

La description des axes fournit des résultats identiques aux précédents

Les distances au centre de classe pour les modalités faisant partie de la classe considérée

```
attr(,"class")
[1] "catdes" "list "
```

```
> res.hcpc$desc.ind
$para
cluster: 1
      FF.A      FF.F      FF.E      FF.P
0.02987127 0.07545981 0.08491899 0.14079825
-----
cluster: 2
      EF.P      EF.F      EF.A      EF.E
0.04665165 0.05971005 0.13706683 0.18796927
-----
cluster: 3
      NF.F      NF.A      NF.E      NF.P
0.03346484 0.15767943 0.34689981 0.43634580
```

Le minimum des distances aux centres des autres classes pour les modalités faisant partie de la classe considérée

```
$dist
cluster: 1
```

FF.E	FF.F	FF.A	FF.P
1.512182	1.476959	1.453368	1.401372

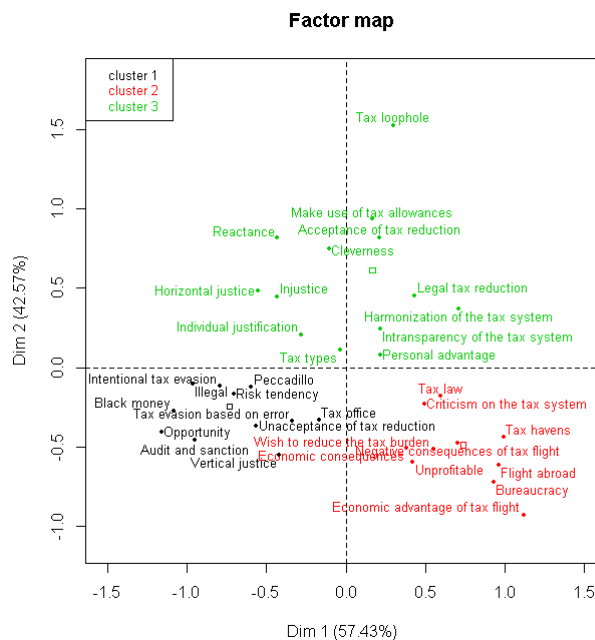
cluster: 2

EF.A	EF.P	EF.F	EF.E
1.514131	1.476553	1.451026	1.429699

cluster: 3

NF.P	NF.F	NF.A	NF.E
1.825761	1.515716	1.325006	1.163512

Les résultats sont ici ceux attendus. Il pourrait être plus intéressant de faire la classification sur les modalités lignes, en conservant 3 classes, pour caractériser les catégories sémantiques par rapport aux 3 scénarios :



4 Analyse des correspondances multiples : résultats fournis par Factominer

Le menu FactoMineR de R Commander permet de réaliser une ACM à partir d'un tableau protocole. Nous allons l'essayer sur l'exemple suivant :

Source : Site de B. Le Roux <http://www.math-info.univ-paris5.fr/~lerb/> consulté le 27 février 2013.

On a soumis à un échantillon de 1215 sujets (britanniques) un questionnaire relatif à leurs préférences en matière de sorties et de loisirs culturels. On s'intéresse ici aux quatre questions suivantes :

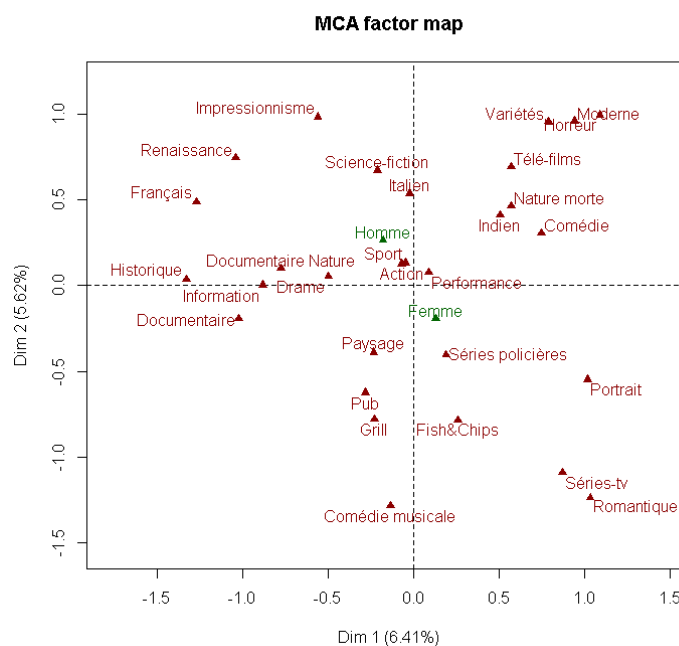
- TV : préférences en matière de programme de télévision ; 8 modalités : Séries-tv, Documentaire Nature, Variétés, Information, Sport, Drame, Séries policières, Télé-films
- Film : préférences en matière de films ; 8 modalités : Action, Horreur, Historique, Documentaire, Comédie, Comédie musicale, Science-fiction, Romantique
- Art : préférences en matière d'art ; 7 modalités : Paysage, Nature morte, Portrait, Impressionnisme, Performance, Moderne, Renaissance
- Restaurant : préférences en matière de sorties au restaurant ; 6 modalités : Grill, Indien, Pub, Italien, Fish&Chips, Français ; "Indien" regroupe en fait différents types de restaurants orientaux : Indien, Chinois, Thaï.

On a aussi relevé le sexe de la personne interrogée (deux modalités), qui sera utilisée comme variable supplémentaire.

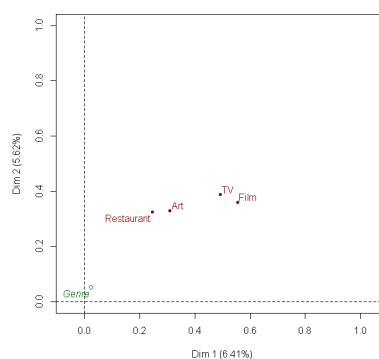
Chargez le jeu de données Taste.Example.RData. Explorez les données au niveau descriptif puis réalisez une ACM avec les spécifications suivantes :

- Variables actives : TV, Film, Art, Restaurant
- Variable supplémentaire : Genre
- Individus supplémentaires : observations de numéro supérieur ou égal à 1216.
- Nombre d'axes à étudier : 2
- Spécifications pour le graphique : ne pas représenter les individus (ni actifs, ni supplémentaires).

Les résultats produits devraient vous conduire au graphique suivant :



A noter : le graphique relatif aux variables positionne celles-ci selon les valeurs des coefficients eta-2 sur chacun des axes :



Les résultats numériques permettent de répondre aux questions suivantes :

- 1) a) L'inertie totale du nuage de points est $I = 6,25$. Comment peut-on retrouver cette valeur ? On sait que l'inertie relative des questions dépend seulement de leur nombre de modalités. Calculer l'inertie relative de chacune des quatre questions.

b) Examiner les tableaux donnant les tris à plat des cinq variables étudiées. Pour la méthode utilisée, quelles sont les recommandations généralement indiquées en ce qui concerne le nombre de modalités des différentes questions et les fréquences des modalités ? Dans quelle mesure ces conditions sont-elles vérifiées sur l'exemple traité ici ? Quelle est la variable pour laquelle ces recommandations sont les moins bien respectées ?

2) On donne ci-dessous le tableau des taux d'inertie modifiés :

	ValProp.	$\lambda' = (\lambda - \lambda_m)^2$	Taux d'inertie modifiés	Cumuls
1	0,4004	0,02261	47,59%	47,59%
2	0,3512	0,01023	21,54%	69,13%
3	0,3250	0,00563	11,84%	80,97%
4	0,3081	0,00337	7,10%	88,07%
5	0,2989	0,00239	5,03%	93,10%
6	0,2876	0,00141	2,98%	96,07%
7	0,2782	0,00080	1,68%	97,75%
8	0,2739	0,00057	1,20%	98,95%
9	0,2682	0,00033	0,70%	99,65%
10	0,2600	0,00010	0,21%	99,86%
11	0,2582	0,00007	0,14%	100,00%
12	0,2512	0,00000	0,00%	100,00%
	Total	0,04751		

a) Pourquoi ce tableau ne comporte-t-il que 12 valeurs propres, alors que le tableau complet en indique 25 ? Indiquer comment a été calculée la première valeur propre modifiée (0,02261) et le premier taux d'inertie modifié (47,59%).

b) Interpréter les tableaux relatifs aux valeurs propres. Combien d'axes serait-il pertinent d'étudier (on n'étudiera cependant que les deux premiers axes) ?

3) Etude des deux premiers axes.

a) Quelles sont les modalités dont la contribution est supérieure à la moyenne sur le premier axe ? Pour chacune d'elles, préciser le signe de la coordonnée correspondante. Comment peut-on interpréter cet axe en termes d'opposition entre modalités ?

b) Même question pour le deuxième axe.

4) Le tableau "Synthèse par question sur les deux premiers axes" indique l'inertie de chacune des quatre questions actives sur les deux premiers axes.

a) Indiquer comment a été calculée l'inertie de la question "TV" sur le premier axe (0,3075).

b) Quels commentaires peut-on faire concernant les inerties des quatre questions sur chacun des deux axes ?

c) Calculer le rapport de corrélation (coefficient η^2) entre la première variable factorielle et la question "Film", puis de même entre la première variable factorielle et la question "Restaurant". Commenter les résultats obtenus.

5) Des chercheurs ont proposé d'expliquer les résultats de cette étude en termes de goûts du public : goûts plutôt populaires ou au contraire plutôt sophistiqués, goûts pour les fictions ou au contraire pour les

produits culturels en prise avec la réalité. Dans quelle mesure ces éléments permettent-ils de décrire les oppositions observées sur les deux axes étudiés ?

6) La variable "Sexe" a été prise comme variable supplémentaire dans cette étude. Commenter la position des modalités de cette variable sur le graphique. Peut-on en déduire des résultats concernant le lien entre les goûts des sujets et leur sexe ? Lesquels ?

Remarque. Il est également possible de demander une classification. Celle-ci porte sur les lignes (ici les personnes interrogées). Outre les résultats observés dans les méthodes précédentes, on obtient :

- Les résultats des tests du khi-2 sur les tableaux de contingence obtenus en croisant l'appartenance à une classe et la modalité choisie pour chaque question :

```
> res.hcpc$desc.var
$test.chi2
              p.value df
TV           7.701093e-207 14
Film        5.387389e-140 14
...
```

- La composition des classes en termes de fréquence des différentes modalités (fréquence de la modalité dans la classe et fréquence de la classe dans la modalité) :

```
$category
$category$`1`
              Cla/Mod Mod/Cla      Global      p.value      v.test
TV=Information  83.181818    36.6 18.106996 1.834020e-45 14.151275
Film=Historique 87.857143    24.6 11.522634 4.202302e-34 12.175419
...
```