Fiche TD avec le logiciel R : tdr620

Initiation à l'analyse factorielle des correspondances

A.B. Dufour & M. Royer & J.R. Lobry

Dans cette fiche, on étudie l'Analyse Factorielle des Correspondances. Cette technique statistique permet de réduire le nombre de variables, afin d'obtenir une représentation graphique des tableaux de contingence. Elle vise à y rassembler la quasi-totalité de l'information initiale, en s'attachant aux correspondances entre les caractères.

Table des matières

1 Exemple introductif

1.1 Les données

L'exemple porte sur la couleur des yeux et la couleur des cheveux de 592 étudiants. Les données ont été collectées dans le cadre d'un projet de classe par les étudiants d'un cours de statistique élémentaire à l'Université de Delaware [2].

```
snee74 <- read.table("http://pbil.univ-lyon1.fr/R/donnees/snee74.txt",
    header = TRUE)
names(snee74)
```

```
[1] "cheveux" "yeux"    "sexe"
```

```
head(snee74)
```

```
  cheveux   yeux    sexe
1    Noir Marron    Male
2   Blond   Bleu Femelle
3    Noir   Bleu    Male
4  Marron Marron Femelle
5    Roux Marron    Male
6  Marron   Bleu    Male
```

La couleur des cheveux est définie par 4 modalités : blond, marron, noir et roux.

```
cheveux <- snee74$cheveux
summary(cheveux)
```

```
 Blond Marron   Noir   Roux
   127    286    108     71
```

La couleur des yeux est définie par 4 modalités : bleu, marron, noisette et vert.

```
yeux <- snee74$yeux
summary(yeux)
```

```
    Bleu   Marron Noisette     Vert
     215      220       93       64
```

Le lien entre les deux couleurs s'obtient à l'aide d'un tableau croisé qui ventile la population entre les modalités de ces deux variables qualitatives. C'est une table de contingence.

```
(couleurs <- table(yeux, cheveux))
```

```
          cheveux
yeux       Blond Marron Noir Roux
  Bleu        94     84   20   17
  Marron       7    119   68   26
  Noisette    10     54   15   14
  Vert        16     29    5   14
```

Par commodité, on transforme cet objet en un `data.frame` :

```
(dfcouleurs <- data.frame(unclass(couleurs)))
```

```
         Blond Marron Noir Roux
Bleu        94     84   20   17
Marron       7    119   68   26
Noisette    10     54   15   14
Vert        16     29    5   14
```

1.2 Définition d'un score *a priori*

On va affecter *a priori* un score à chacune des colonnes (*couleur des cheveux*), par exemple (1,-1,-1,1), qui opère une opposition entre cheveux foncés (`Marron`, `Noir`) et clairs (`Blond`, `Roux`).

```
scorecheveux <- c(1, -1, -1, 1)
names(scorecheveux) <- colnames(couleurs)
scorecheveux
```

```
Blond Marron   Noir   Roux 
    1     -1     -1      1 
```

Pour chaque ligne de la table de contingence (*couleur des yeux*), une fréquence observée correspond à chaque couleur de cheveux. Ainsi, pour la modalité yeux `Bleu` on obtient :

```
dfcouleurs <- data.frame(unclass(couleurs))
dfcouleurs["Bleu", ]/sum(dfcouleurs["Bleu", ])
```

```
         Blond    Marron       Noir       Roux
Bleu 0.4372093 0.3906977 0.09302326 0.07906977
```

Il est alors possible de calculer le score moyen pour la modalité yeux `Bleu` :

```
yeux.bleu <- dfcouleurs["Bleu", ]/sum(dfcouleurs["Bleu", ])
yeux.bleu * scorecheveux
```

```
         Blond     Marron        Noir       Roux
Bleu 0.4372093 -0.3906977 -0.09302326 0.07906977
```

```
sum(yeux.bleu * scorecheveux)
```

```
[1] 0.03255814
```

Ce score moyen positif montre que les individus aux yeux `Bleu` ont des cheveux plutôt clairs.
Ce score moyen peut être calculé pour toutes les couleurs de yeux.

```
freqyeux <- apply(dfcouleurs, 1, function(x) x/sum(x))
freqyeux
```

```
             Bleu     Marron  Noisette     Vert
Blond  0.43720930 0.03181818 0.1075269 0.250000
Marron 0.39069767 0.54090909 0.5806452 0.453125
Noir   0.09302326 0.30909091 0.1612903 0.078125
Roux   0.07906977 0.11818182 0.1505376 0.218750
```

```
t(freqyeux)
```

```
              Blond    Marron       Noir       Roux
Bleu     0.43720930 0.3906977 0.09302326 0.07906977
Marron   0.03181818 0.5409091 0.30909091 0.11818182
Noisette 0.10752688 0.5806452 0.16129032 0.15053763
Vert     0.25000000 0.4531250 0.07812500 0.21875000
```

```
scoreyeux <- apply(t(freqyeux), 1, function(x) sum(x * scorecheveux))
scoreyeux
```

```
      Bleu      Marron    Noisette        Vert 
0.03255814 -0.70000000 -0.48387097 -0.06250000 
```

Pour les yeux marrons, on obtient un score moyen égal à -0.7 qui est négatif et indique donc que les cheveux foncés dominent dans cette sous-population.

On pourrait assez bien séparer les 4 couleurs des yeux sur la base du score proposé pour la couleur des cheveux. Cependant, deux questions se posent :

- ⋆ Existe-t-il un score des cheveux qui permet de discriminer encore mieux la couleur des yeux ?
- ⋆ Lorsqu'on connaît moins bien le sujet, (ici, l'opposition *clair/foncé* est naturelle), comment définir un score qui permette de mieux comprendre la structure du tableau de données ?

1.3 Notion de score optimum

L'Analyse Factorielle des Correspondances (AFC) est la méthode permettant de définir pour une table de contingence un score sur les colonnes tel que les scores moyens des lignes (obtenus en utilisant les fréquences des tableaux de profils) soient les plus séparés possibles, au sens de la variance de ces scores moyens. Et inversement.
Cette méthode choisit comme score optimal pour les colonnes (cheveux) les valeurs :

```
library(ade4)
ac <- dudi.coa(dfcouleurs, scannf = F, nf = 3)
rownames(ac$c1)
[1] "Blond"  "Marron" "Noir"   "Roux"
ac$c1[, 1]
[1]  1.8282287 -0.3244635 -1.1042772 -0.2834725
```

On vérifie que les valeurs extrêmes sont obtenues pour les modalités `Blond` et `Noir`, ce qui reflète que la structure majeure de ce jeu de données est l'opposition *clair/foncé*.

Exercice. Retrouver le score moyen des lignes (couleur des yeux) à partir des scores optimaux de la couleur des cheveux obtenus par l'AFC.
Réponse :

```
     Bleu     Marron  Noisette       Vert
0.5474139 -0.4921577 -0.2125969  0.1617534
```

Il est important de noter que si on cherche d'abord les scores optimaux pour le critère *couleur des yeux* par la méthode AFC (coordonnées des lignes, `li` sous `ade4`), on obtient le même résultat :

```
[1] "Bleu"      "Marron"    "Noisette" "Vert"
[1]  0.5474139 -0.4921577 -0.2125969  0.1617534
```

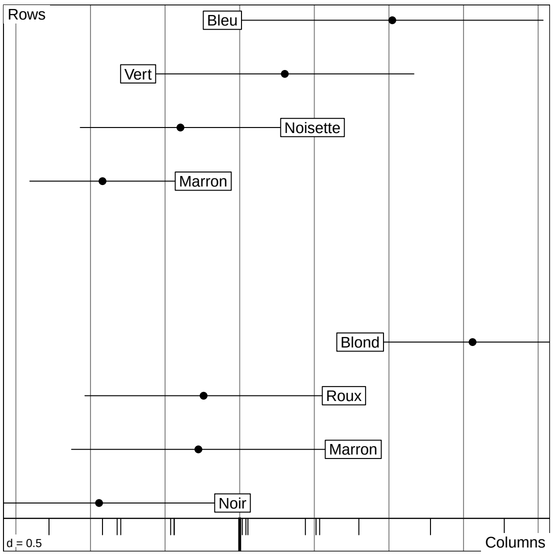
Le raisonnement que l'on vient de tenir peut se reproduire dans la recherche de score moyen des couleurs de cheveux à partir des scores optimaux de la couleur des yeux.

```
rownames(ac$co)
[1] "Blond"  "Marron" "Noir"   "Roux"
ac$co[, 1]
[1]  0.8353478 -0.1482527 -0.5045624 -0.1295233
```

1.4 Représentations graphiques

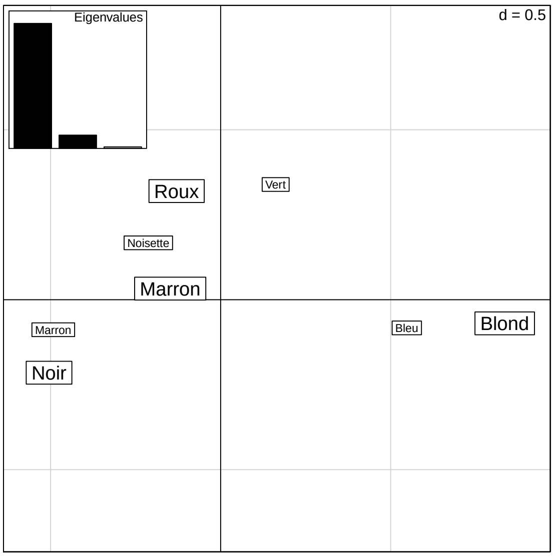
On peut alors donner une représentation graphique des valeurs obtenues pour les scores des lignes (resp. des colonnes) pour le premier score optimal des colonnes (resp. des lignes).

```
score(ac)
```



Et pour finir, on donne une représentation graphique des résultats obtenus sur les deux premiers scores optimaux.

```
scatter(ac)
```



2 Table de Contingence

2.1 Tableau des données

La table engendrée par le croisement de deux variables qualitatives s'appelle une *table de contingence* observée. Il est important de rappeler que :

i) tout individu présente une modalité et une seule de chaque variable ;

ii) chaque modalité doit avoir été observée au moins une fois, sinon elle est supprimée.

Les données proviennent d'une société d'assurance automobile. Les deux variables retenues pour l'analyse sont :

1. le mode de règlement : `annuel`, `mensuel`, `semestriel` ou `trimestriel` ;
2. la situation maritale : `célibataire`, `concubin`, `divorcé`, `marié` ou `veuf`.

On construit la table de contingence liée à ces variables.

```
sitpay <- matrix(c(209, 1483, 41, 320, 60, 34, 151, 1, 70, 10, 535,
    2448, 33, 897, 135, 77, 245, 4, 139, 9), byrow = T, ncol = 5)
colnames(sitpay) <- c("célibataire", "concubin", "divorcé", "marié",
    "veuf")
rownames(sitpay) <- c("annuel", "mensuel", "semestriel", "trimestriel")
sitpay
```

```
            célibataire concubin divorcé marié veuf
annuel              209     1483      41   320   60
mensuel              34      151       1    70   10
semestriel          535     2448      33   897  135
trimestriel          77      245       4   139    9
```

Les informations de base sont le nombre total d'individus (`n`), le nombre de modalités pour la variable 'mode de règlement' (`I`) et le nombre de modalités pour la variable 'situation maritale' (`J`).

```
n <- sum(sitpay)
I <- nrow(sitpay)
J <- ncol(sitpay)
```

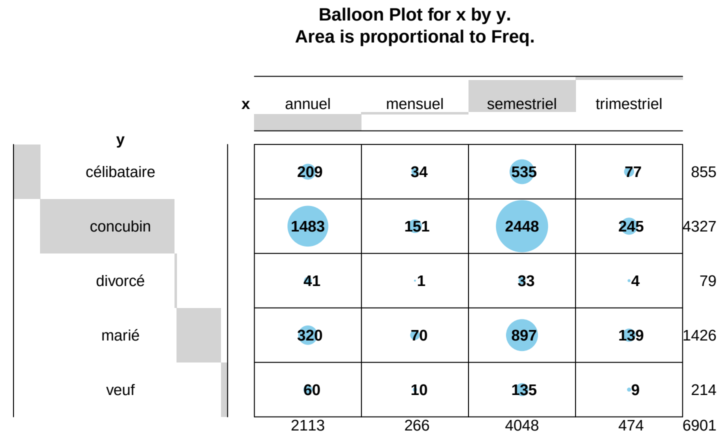
On peut construire le tableau des fréquences relatives où chaque terme est de la forme $f_{ij} = \frac{n_{ij}}{n}$.

```
freqsitpay <- sitpay/n
round(freqsitpay, digits = 4)
```

```
            célibataire concubin divorcé  marié   veuf
annuel           0.0303   0.2149  0.0059 0.0464 0.0087
mensuel          0.0049   0.0219  0.0001 0.0101 0.0014
semestriel       0.0775   0.3547  0.0048 0.1300 0.0196
trimestriel      0.0112   0.0355  0.0006 0.0201 0.0013
```

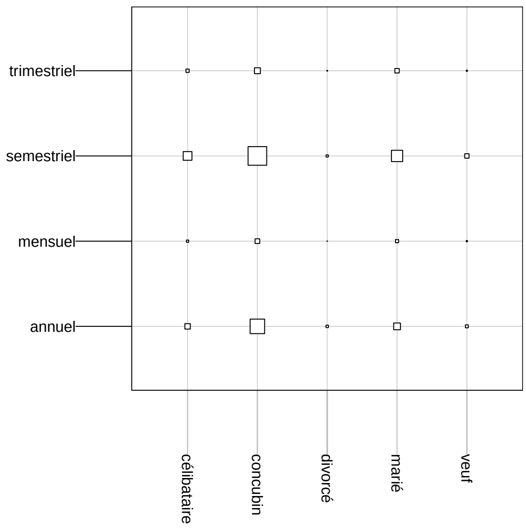
On peut obtenir différentes représentations graphiques de la table de contingence. Le principe est d'utiliser des symboles dont la surface est proportionelle aux effectifs :

```
library(gplots)
balloonplot(as.table(sitpay))
```



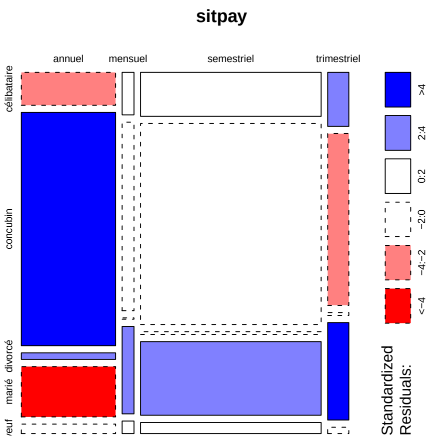
On retrouve ce principe dans la fonction `table.cont` du paquet ade4 :

```
table.cont(sitpay, row.labels = rownames(sitpay), col.labels = colnames(sitpay),
    csize = 1)
```



La fonction `mosaicplot` permet de mettre en évidence les liens les plus importants :

```
mosaicplot(sitpay, shade = TRUE)
```



2.2 Tableaux des profils lignes et colonnes

On calcule maintenant les fréquences conditionnelles. Pour ce faire, on note V_1 et V_2, les deux variables qualitatives étudiées.

Profils lignes

Les fréquences conditionnelles associées aux profils lignes sont notées $f_{i|j}$ et définies par

$$f_{j|i} = P(V_2 = j | V_1 = i) = \frac{P(V_2 = j \cap V_1 = i)}{P(V_1 = i)}$$

$$f_{j|i} = \frac{\frac{n_{ij}}{n}}{\frac{n_{i\bullet}}{n}} = \frac{n_{ij}}{n_{i\bullet}}$$

```
profLignes <- prop.table(sitpay, 1)
profLignes
             célibataire  concubin     divorcé     marié       veuf
annuel         0.0989115 0.7018457 0.019403691 0.1514434 0.02839565
mensuel        0.1278195 0.5676692 0.003759398 0.2631579 0.03759398
semestriel     0.1321640 0.6047431 0.008152174 0.2215909 0.03334980
trimestriel    0.1624473 0.5168776 0.008438819 0.2932489 0.01898734
```

On vérifie que les sommes en lignes sont toutes égales à 1.

```
rowSums(profLignes)
     annuel     mensuel  semestriel trimestriel
          1           1           1           1
```

Profils colonnes

Les fréquences conditionnelles associées aux profils colonnes sont notées $f_{i|j}$ et définies par

$$f_{i|j} = P(V_1 = i | V_2 = j) = \frac{P(V_1 = i \cap V_2 = j)}{P(V_2 = j)}$$

$$f_{i|j} = \frac{\frac{n_{ij}}{n}}{\frac{n_{\bullet j}}{n}} = \frac{n_{ij}}{n_{\bullet j}}$$

```
profColonnes <- prop.table(sitpay, 2)
profColonnes
            célibataire    concubin    divorcé       marié        veuf
annuel       0.24444444 0.34273168 0.51898734 0.22440393 0.28037383
mensuel      0.03976608 0.03489716 0.01265823 0.04908836 0.04672897
semestriel   0.62573099 0.56574994 0.41772152 0.62903226 0.63084112
trimestriel  0.09005848 0.05662122 0.05063291 0.09747546 0.04205607
```

On vérifie également que les sommes en colonnes sont toutes égales à 1.

```
colSums(profColonnes)
célibataire    concubin     divorcé       marié        veuf
          1           1           1           1           1
```

2.3 Lien avec le test du χ^2 d'indépendance

Le test du Khi-Deux d'indépendance entre deux variables est caractérisé par les deux hypothèses :

⋆ H_0 : les deux variables sont indépendantes

⋆ H_1 : les deux variables sont liées.

Sous l'hypothèse nulle H_0, $P(V_2 = j \cap V_1 = i) = P(V_1 = i) \times P(V_2 = j)$.
Ainsi, sous H_0, la fréquence théorique est égale à $\frac{n_{i\bullet}}{n} \times \frac{n_{\bullet j}}{n}$.

On en déduit la table des **effectifs théoriques** (qui serait observée sous H_0), en conservant les effectifs marginaux observés.

$$\frac{n_{i\bullet} \times n_{\bullet j}}{n}$$

```
reschi <- chisq.test(sitpay)
reschi$expected
            célibataire   concubin    divorcé      marié       veuf
annuel        261.79032 1324.8734 24.188813 436.62339  65.52413
mensuel        32.95609  166.7848  3.045066  54.96537   8.24866
semestriel    501.52731 2538.1388 46.339951 836.46544 125.52847
trimestriel    58.72627  297.2030  5.426170  97.94580  14.69874
```

La statistique du test est la suivante :

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n})^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}}$$

Elle tend vers une loi du χ^2 à $(I-1) \times (J-1)$ degrés de liberté. Dans l'étude de la relation entre le mode de payement de l'assurance automobile et la situation maritale, le résultat au test du Khi-Deux est

```
reschi
        Pearson's Chi-squared test
data:  sitpay
X-squared = 129.2212, df = 12, p-value < 2.2e-16
```

Comme la p-value est très faible, on rejette l'hypothèse nulle. Les variables sont liées. Il est alors intéressant d'explorer la structure de cette relation.

Définition
On appelle **lien** entre la modalité i de la variable V_1 et la modalité j de la variable V_2 la quantité :

$$\frac{1}{n} \frac{\left(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n})\right)^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}}$$

Les couples de modalités (i, j) qui correspondent aux liens les plus importants sont les plus responsables de la dépendance entre la variable V_1 et la variable V_2.

Conclusion. Que la liaison entre les 2 variables soit statistiquement significative ou non, on peut explorer la structure du tableau plus en détail. Lorsque les variables présentent de nombreuses modalités, il est difficile d'extraire une information pertinente si on se contente d'observer le tableau de données. La technique de l'Analyse Factorielle des Correspondances (AFC) est là pour pallier cette déficience.

3 Compréhension des résultats d'une AFC

Les résultats de l'AFC de la table de contingence permettant d'étudier le lien entre le mode de payement et la situation maritale.

```
dfsitpay <- as.data.frame(sitpay)
afc <- dudi.coa(dfsitpay, scannf = F, nf = 3)
names(afc)
```

```
 [1] "tab"  "cw"   "lw"   "eig"  "rank" "nf"   "l1"   "co"   "li"   "c1"   "call"
[12] "N"
```

3.1 Le tableau analysé

Le tableau analysé est :

```
afc$tab
```

```
            célibataire    concubin    divorcé       marié        veuf
annuel      -0.20165115  0.11935227  0.6949984 -0.26710293 -0.08430676
mensuel      0.03167568 -0.09464179 -0.6715999  0.27352919  0.21231818
semestriel   0.06674150 -0.03551375 -0.2878715  0.07236947  0.07545321
trimestriel  0.31116786 -0.17564766 -0.2628318  0.41915215 -0.38770259
```

C'est le lien entre les effectifs théoriques et les effectifs observés.

```
(dfsitpay - reschi$expected)/reschi$expected
```

```
            célibataire    concubin    divorcé       marié        veuf
annuel      -0.20165115  0.11935227  0.6949984 -0.26710293 -0.08430676
mensuel      0.03167568 -0.09464179 -0.6715999  0.27352919  0.21231818
semestriel   0.06674150 -0.03551375 -0.2878715  0.07236947  0.07545321
trimestriel  0.31116786 -0.17564766 -0.2628318  0.41915215 -0.38770259
```

3.2 Les pondérations

Les pondérations des lignes et des colonnes sont les fréquences marginales de la table de contingence observée.

```
afc$cw
célibataire    concubin     divorcé       marié        veuf
 0.12389509  0.62701058  0.01144762  0.20663672  0.03101000
apply(dfsitpay, 2, function(x) sum(x)/n)
célibataire    concubin     divorcé       marié        veuf
 0.12389509  0.62701058  0.01144762  0.20663672  0.03101000
afc$lw
     annuel     mensuel  semestriel trimestriel
 0.30618751  0.03854514  0.58658165  0.06868570
apply(dfsitpay, 1, function(x) sum(x)/n)
     annuel     mensuel  semestriel trimestriel
 0.30618751  0.03854514  0.58658165  0.06868570
```

3.3 La matrice diagonalisée

La matrice diagonalisée est **H**.

```
matZ <- as.matrix(afc$tab)
DI <- diag(afc$lw)
DrJ <- diag(sqrt(afc$cw))
matH <- DrJ %*% t(matZ) %*% DI %*% matZ %*% DrJ
```

Le rang de la matrice analysée est donné par $min(I-1, J-1)$ soit

```
min(I - 1, J - 1)
[1] 3
afc$rank
[1] 3
```

Les valeurs propres et les vecteurs propres issus de cette diagonalisation sont :

```
eigen(matH)
$values
[1]  1.765310e-02  9.560696e-04  1.158312e-04 -1.528047e-18 -2.520770e-18
$vectors
            [,1]        [,2]        [,3]       [,4]       [,5]
[1,]  0.38337054 -0.15133044  0.82387057  0.0000000 -0.3890545
[2,] -0.51798828  0.08710879 -0.01869848 -0.6187179 -0.5838996
[3,] -0.34849169 -0.61094688 -0.13452330  0.5784541 -0.3906300
[4,]  0.68039274 -0.18151055 -0.54538573 -0.1881184 -0.4138665
[5,]  0.01828753  0.75053913 -0.07311325  0.4971826 -0.4287429
reseigen <- eigen(matH)
```

On retrouve bien les valeurs propres de l'analyse.

```
reseigen$values
[1]  1.765310e-02  9.560696e-04  1.158312e-04 -1.528047e-18 -2.520770e-18
afc$eig
[1] 0.0176531021 0.0009560696 0.0001158312
```

3.4 Les coordonnées des lignes

Les coordonnées des lignes dites axes principaux s'obtiennent par $\mathbf{ZD}_J^{1/2}\mathbf{U}$ (*cf* cours[1]). Elles sont centrées, de variances λ et de covariances nulles.

```
matZ %*% DrJ %*% reseigen$vectors[, 1:3]
```

```
                  [,1]        [,2]         [,3]
annuel      -0.18496237 -0.01556050 -0.002942584
mensuel      0.15341763  0.04117812 -0.050292921
semestriel   0.05693235  0.01681398  0.005110564
trimestriel  0.25222413 -0.09733548 -0.002303726
```

```
afc$li
```

```
                 Axis1       Axis2        Axis3
annuel       0.18496237  0.01556050 -0.002942584
mensuel     -0.15341763 -0.04117812 -0.050292921
semestriel  -0.05693235 -0.01681398  0.005110564
trimestriel -0.25222413  0.09733548 -0.002303726
```

```
sum(afc$li$Axis1 * afc$lw)
```

```
[1] -2.688821e-17
```

```
sum(afc$li$Axis2 * afc$lw)
```

```
[1] 4.336809e-18
```

```
sum(afc$li$Axis1 * afc$li$Axis1 * afc$lw)
```

```
[1] 0.01765310
```

```
afc$eig[1]
```

```
[1] 0.01765310
```

```
sum(afc$li$Axis2 * afc$li$Axis2 * afc$lw)
```

```
[1] 0.0009560696
```

```
afc$eig[2]
```

```
[1] 0.0009560696
```

```
sum(afc$li$Axis1 * afc$li$Axis2 * afc$lw)
```

```
[1] 4.065758e-19
```

3.5 Les coordonnées des colonnes

Les coordonnées des colonnes dites composantes principales s'obtiennent par $\mathbf{D}_J^{-1/2}\mathbf{U}\boldsymbol{\Lambda}^{\mathbf{1/2}}$ (cf cours). Elles sont centrées, de variances λ et de covariances nulles.

```
diag(1/sqrt(afc$cw)) %*% reseigen$vectors[, 1:3] %*% diag(sqrt(afc$eig))
```

```
            [,1]         [,2]          [,3]
[1,]  0.14471122 -0.013293643  0.0251909629
[2,] -0.08691466  0.003401491 -0.0002541451
[3,] -0.43275831 -0.176559333 -0.0135317134
[4,]  0.19886870 -0.012346472 -0.0129125746
[5,]  0.01379796  0.131785378 -0.0044684615
```

```
afc$co
```

```
                 Comp1        Comp2         Comp3
célibataire -0.14471122  0.013293643  0.0251909629
concubin     0.08691466 -0.003401491 -0.0002541451
divorcé      0.43275831  0.176559333 -0.0135317134
marié       -0.19886870  0.012346472 -0.0129125746
veuf        -0.01379796 -0.131785378 -0.0044684615
```

```
sum(afc$co$Comp1 * afc$cw)
```

```
[1] -5.795061e-17
```

```
sum(afc$co$Comp2 * afc$cw)
```

```
[1] -3.729655e-17
```

```
sum(afc$co$Comp1 * afc$co$Comp1 * afc$cw)
```

```
[1] 0.01765310
```

[1] http://pbil.univ-lyon1.fr/R/pdf/add2.pdf.

```
afc$eig[1]
```
```
[1] 0.01765310
```
```
sum(afc$co$Comp2 * afc$co$Comp2 * afc$cw)
```
```
[1] 0.0009560696
```
```
afc$eig[2]
```
```
[1] 0.0009560696
```
```
sum(afc$co$Comp1 * afc$co$Comp2 * afc$cw)
```
```
[1] -2.913793e-19
```

3.6 Lien avec le Khi-Deux

Lien entre l'inertie totale et la valeur de la statistique du Khi-Deux :

$$I_T = \frac{\chi^2}{n}$$

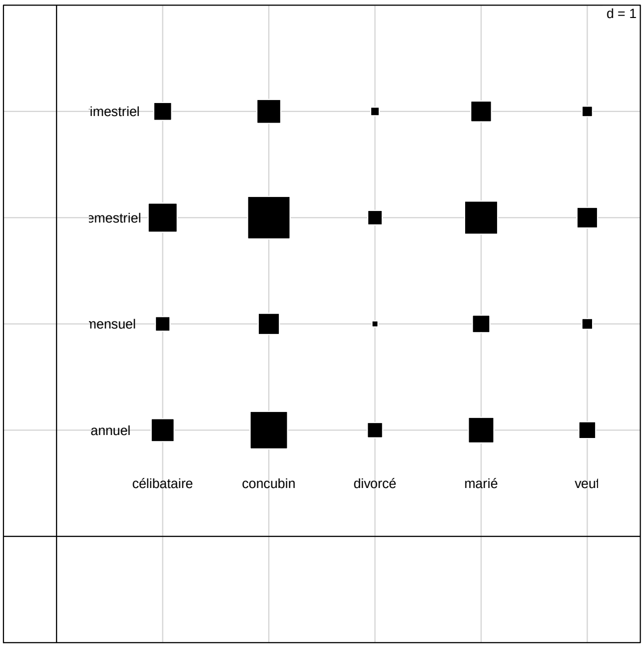
.

```
reschi$statistic
```
```
X-squared
 129.2212
```
```
reschi$statistic/sum(sitpay)
```
```
 X-squared
0.01872500
```
```
sum(afc$eig)
```
```
[1] 0.01872500
```

3.7 Visualisation finale

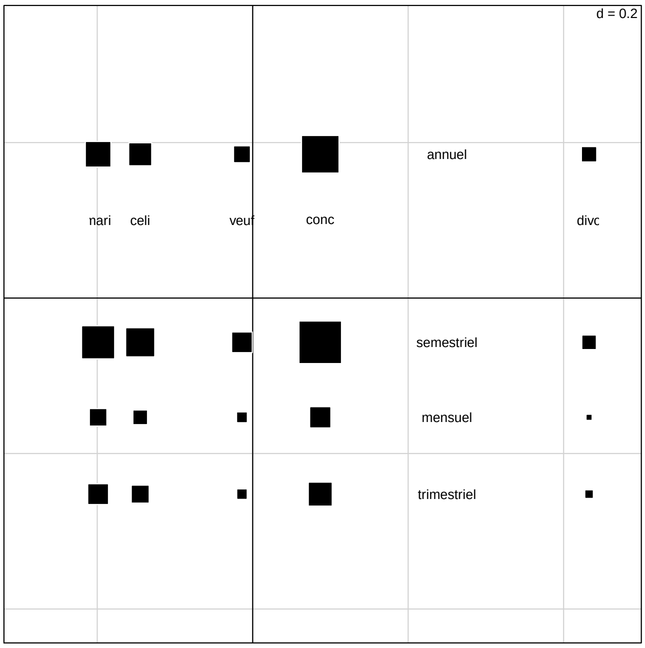
La représentation des modalités de la table de contingence :

```
k <- 0
effectif <- 0
for (i in 1:4) {
    for (j in 1:5) {
        k = k + 1
        effectif[k] <- sitpay[i, j]
    }
}
xdep <- rep(1:5, 4)
ydep <- rep(1:4, rep(5, 4))
s.value(cbind(xdep, ydep), sqrt(effectif), clegend = 0)
text(1:5, rep(0.5, 5), lab = colnames(sitpay), las = 2, cex = 0.75)
text(rep(0.51, 4), 1:4, lab = rownames(sitpay), las = 2, cex = 0.75)
```



La représentation des modalités suite à l'analyse :

```
xarr <- rep(afc$co[, 1], 4)
yarr <- rep(afc$li[, 1], rep(5, 4))
s.value(cbind(xarr, yarr), sqrt(effectif), xlim = c(-0.32, 0.5),
    clegend = 0)
labcol <- c("celi", "conc", "divo", "mari", "veuf")
text(afc$co[, 1], rep(0.1, 5), lab = labcol, cex = 0.75)
text(rep(0.25, 4), afc$li[, 1], lab = rownames(sitpay), cex = 0.75)
```



4 Application

La recherche scientifique ne soulève que peu de controverse ou de résistance de la part du grand public. Dans quelques cas, rares, le débat entre science, morale et religion resurgit. C'est le cas de la recherche sur les embryons humains.

L'analyse que nous nous proposons de réaliser est adaptée de l'article *Attitudes towards Embryo research, worldviews and the moral status of the Embryo Frame* [1]. Les données ont été recueillies au près du grand public dans 9 pays européens et elles concernent le statut accordé à l'embryon.

> A human embryo that is a few days old . . . (1) "is a mere cluster of cells, and it makes no sense to discuss its moral condition"; (2) "has a moral condition halfway between that of a cluster of cells and that of a human being"; (3) "is closer in its moral condition to a human being than to a mere cluster of cells"; (4) "has the same moral condition as a human being."

Les données sont rangées dans la table de contingence ci-dessous. Répondre à la question "Peut-on faire une typologie des pays ?".

Pays	(1)	(2)	(3)	(4)	non réponses
Autriche	64	223	243	326	144
Danemark	373	234	112	219	62
France	227	224	166	282	100
Allemagne	88	218	259	289	146
Italie	203	157	137	373	130
Pays.Bas	207	329	154	223	88
Pologne	138	154	98	382	229
Royaume.Uni	255	168	117	236	224
Espagne	215	188	125	298	174

Pour entrer les données, utilisez :

```
res <- data.frame()
fix(res)
```

Références

[1] R Pardo and F. Calvo. Attitudes towards embryo research, worldviews and the moral status of the embryo frame. *Science Communication*, 30, 1 :8–47, 2008.

[2] R.D. Snee. Graphical display of two-way contingency tables. *The American Statistician*, 28 :9–12, 1974.