

Analyse en composantes principales avec R

Plusieurs packages fournissent des outils permettant de réaliser une analyse en composantes principales. On peut citer :

- anciennement dans le package mva et désormais dans le package stats : prcomp, princomp
- dans le package ade4 : dudi.pca
- dans le package FactoMineR : PCA

Les résultats les plus complets semblent être ceux fournis par la procédure PCA.

Exemple :

Reprenons l'exemple "Budget-Temps-ONU" déjà décrit dans les exemples traités sous Statistica.

Présentation des données :

Il s'agit d'une enquête (ONU 1967) sur les budgets-temps (temps passé dans différentes activités au cours de la journée). Le tableau suivant comprend 10 variables numériques et 4 variables catégorisées. Les 10 variables numériques sont: le temps passé en: Profession, Transport, Ménage, Enfants, Courses, Toilette, Repas, Sommeil, Télé, Loisirs.

Les 4 variables catégorisées sont: Le sexe (1=Hommes 2=Femmes), l'activité (1=Actifs 2=Non Act. 9=Non précisé), l'état civil (1=Célibataires 2=Mariés 9=Non précisé), le Pays (1=USA 2=Pays de l'Ouest 3=Pays de l'Est 4=Yougoslavie).

Le code suivant est utilisé pour identifier les lignes: H: Hommes, F: Femmes, A: Actifs, N: Non Actifs(ves), M: Mariés, C: Célibataires, U: USA, W: Pays de l'Ouest sauf USA, E : Est sauf Yougoslavie, Y: Yougoslavie

Les temps sont notés en centièmes d'heures. La première case en haut à gauche du tableau (HAU) indique que les Hommes Actifs des USA passent en moyenne 6 heures et 6 minutes (6 heures + 10/100 d'heure, soit 6 heures et 6mn) en activité PROFessionnelle. Le total d'une ligne (sur ces 10 variables numériques) est 2400 (24 heures).

Données :

```
> budget
PROF TRAN MENA ENFA COUR TOIL REPA SOMM TELE LOIS
HAU 610 140 140 60 10 120 95 115 760 175 315
FAU 475 90 250 30 140 120 100 775 115 305
FNU 10 0 495 110 170 110 130 785 160 430
HMU 615 140 65 10 115 90 115 765 180 305
FMU 179 29 421 87 161 112 119 776 143 373
HCU 585 115 50 0 150 105 100 760 150 385
FCU 482 94 196 18 141 130 96 775 132 336
```

```

HAW 653 100 95 7 57 85 150 808 115 330
FAW 511 70 307 30 80 95 142 816 87 262
FNW 20 7 568 87 112 90 180 843 125 368
HMY 656 97 97 10 52 85 152 808 122 321
FMW 168 22 528 69 102 83 174 824 119 311
HCW 643 105 72 0 62 77 140 813 100 388
FCW 429 34 262 14 92 97 147 849 84 392
HAY 650 140 120 15 85 90 105 760 70 365
FAY 560 105 375 45 90 90 95 745 60 235
FNY 10 10 710 55 145 85 130 815 60 380
HMY 650 145 112 15 85 90 105 760 80 358
FMY 260 52 576 59 116 85 117 775 65 295
HCY 615 125 95 0 115 90 85 760 40 475
FCY 433 89 318 23 112 96 102 774 45 408
HAE 650 142 122 22 76 94 100 764 96 334
FAE 578 106 338 42 106 94 92 752 64 228
FNE 24 8 594 72 158 92 128 840 86 398
HME 652 133 134 22 68 94 102 763 122 310
FME 436 79 433 60 119 90 107 772 73 231
HCE 627 148 68 0 88 92 86 770 58 463
FCE 434 86 297 21 129 102 94 799 58 380
>

```

Le fichier correspondant peut être téléchargé [ici](#) et chargé sous R à l'aide de la commande :

```
load(file.choose()) # Sélectionner le fichier Budget-Temps-ONU.Rdata
```

Les résultats de l'ACP peuvent alors être obtenus par :

```

> library(FactoMineR)
> budget.pca <- PCA(budget)
> summary(budget.pca)
      Length Class      Mode
eig   3      data.frame list
var   4      -none-      list
ind   3      -none-      list
svd   3      -none-      list
call  7      -none-      list

```

Les résultats relatifs aux valeurs propres

```

> budget.pca$eig
      eigenvalue  inertia cumulative inertia
comp 1  4.589e+00 4.589e+01          45.89
comp 2  2.120e+00 2.120e+01          67.09
...
comp 10 8.892e-33 8.892e-32          100.00

```

Les résultats relatifs aux variables

```

> budget.pca$var
$coord
      Dim.1  Dim.2  Dim.3  Dim.4  Dim.5
PROF -0.97717 -0.12105 0.084576 -0.06694 0.096041
TRAN -0.97978 0.05812 0.008394 -0.04555 -0.111058
...
LOIS 0.09218 0.11026 -0.030219 0.95745 -0.215238

```

```

$cor
      Dim.1    Dim.2    Dim.3    Dim.4    Dim.5
PROF -0.97717 -0.12105  0.084576 -0.06694  0.096041
TRAN -0.97978  0.05812  0.008394 -0.04555 -0.111058
....
LOIS  0.09218  0.11026 -0.030219  0.95745 -0.215238

$cos2
      Dim.1    Dim.2    Dim.3    Dim.4    Dim.5
PROF 0.954868 0.0146521 7.153e-03 0.0044816 9.224e-03
TRAN 0.959968 0.0033774 7.046e-05 0.0020752 1.233e-02
...
LOIS 0.008496 0.0121563 9.132e-04 0.9167161 4.633e-02

$contrib
      Dim.1    Dim.2    Dim.3    Dim.4    Dim.5
PROF 20.8093  0.69119  0.541505  0.37495  1.969202
TRAN 20.9204  0.15932  0.005334  0.17362  2.633131
...
LOIS  0.1852  0.57345  0.069132 76.69627  9.890316

```

Les résultats relatifs aux individus

```

> budget.pca$ind
$coord
      Dim.1    Dim.2    Dim.3    Dim.4    Dim.5
HAU -1.77294  0.68605  1.8713 -0.5752 -0.854389
FAU -0.17159  2.21532  0.6608 -0.4376  1.251659
...
FCE -0.05515  0.80353 -1.0024  0.9678  0.842224

$cos2
      Dim.1    Dim.2    Dim.3    Dim.4    Dim.5
HAU 0.3637027 0.0544596 0.405163 0.0382761 8.446e-02
FAU 0.0041132 0.6856053 0.060995 0.0267467 2.189e-01
...
FCE 0.0008598 0.1825350 0.284079 0.2648023 2.005e-01

$contrib
      Dim.1    Dim.2    Dim.3    Dim.4    Dim.5
HAU 2.446490  0.792966 9.46713  0.988441 5.566e+00
FAU 0.022916  8.268202 1.18043  0.572071 1.195e+01
...
FCE 0.002367  1.087789 2.71672  2.798746 5.408e+00

```

Les valeurs propres et vecteurs propres

```

> budget.pca$svd
$vs
[1] 2.142e+00 1.456e+00 1.149e+00 1.093e+00 6.844e-01 4.461e-01
2.164e-01 1.925e-01 1.547e-01 9.430e-17

$U
      [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    ....
[1,] -0.82766  0.47120  1.6281 -0.52608 -1.248368 -1.49193 -0.05026
...

$V
      [,1]    [,2]    [,3]    [,4]    [,5]    [,6] ...

```

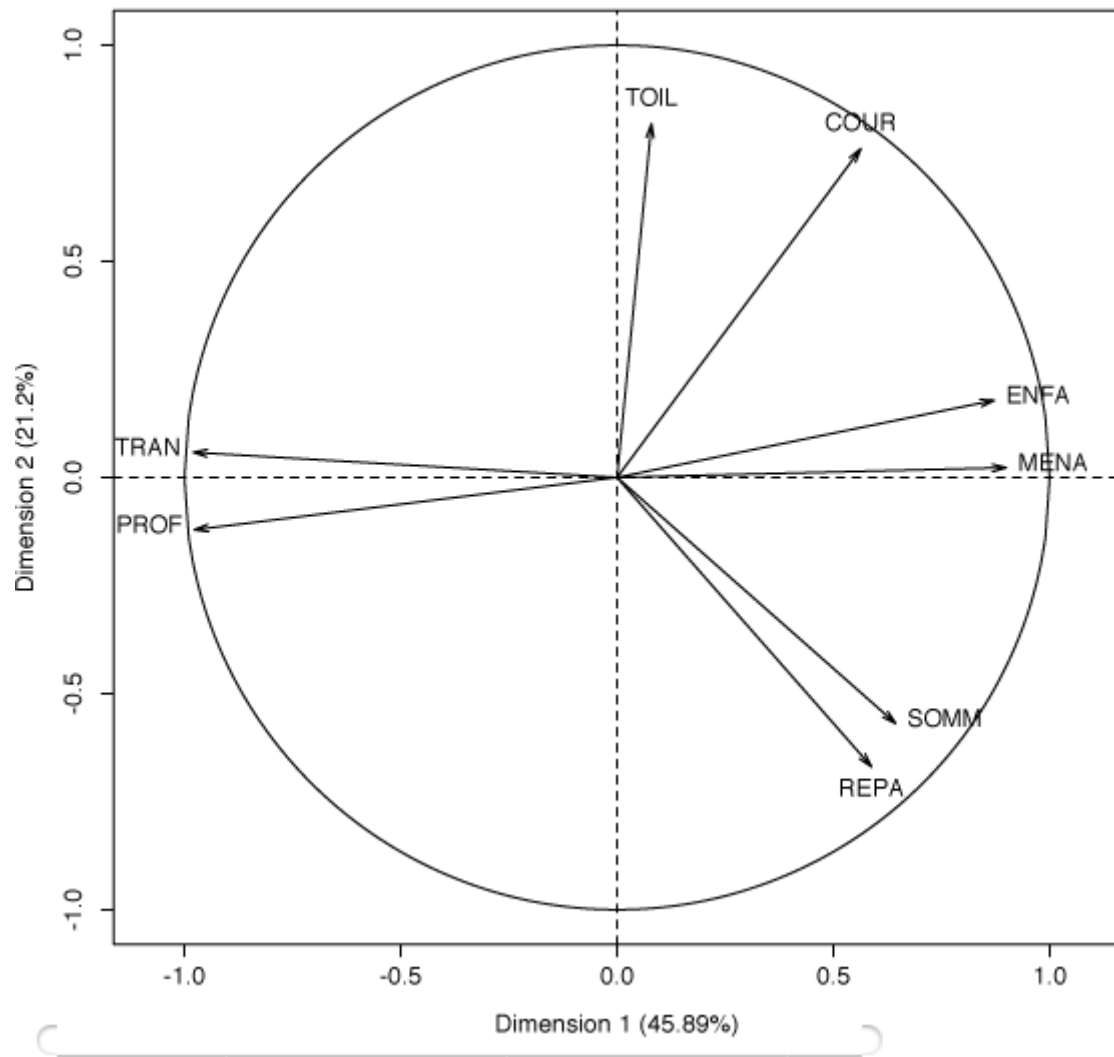
```
[1,] -0.45617 -0.08314 0.073587 -0.06123 0.140328 0.04065 ...  
...
```

Les paramètres d'appel

```
> budget.pca$call  
$row.w  
[1] 0.03571 0.03571 0.03571 0.03571 0.03571 0.03571 0.03571 ...  
...  
$col.w  
[1] 1 1 1 1 1 1 1 1 1 1  
$scale.unit  
[1] TRUE  
$ncp  
[1] 5  
$centre  
  PROF  TRAN  MENA  ENFA  COUR  TOIL  REPA  SOMM  TELE  LOIS  
450.54  86.11 277.07 33.32 108.79 94.93 118.14 785.93 99.43 345.75  
$ecart.type  
  PROF  TRAN  MENA  ENFA  COUR  TOIL  REPA  SOMM  TELE  LOIS  
222.72  47.18 195.10 29.91 31.97 11.31 25.21 29.23 38.70 62.81  
$X  
  PROF TRAN MENA ENFA COUR TOIL REPA SOMM TELE LOIS  
HAU  610  140   60   10  120   95  115  760  175  315  
...
```

Les graphiques :

Variables factor map (PCA)



[illegible]