25/02/2015

Projet de modélisation

Analyse en composantes principales

Corentin BERTHELET

Elisa LESCARRET

Racel MOSHIUR

Jérôme PIPINO

Table des matières

[I – Tableau de données 2](#_Toc412550727)

[II – Le Code : Répartition des tâches 3](#_Toc412550728)

[III – Présentation des résultats de la première ACP 4](#_Toc412550729)

[1. Le tableau de données et le tableau centré réduit 4](#_Toc412550730)

[2. Matrice de corrélations 6](#_Toc412550731)

[3. Valeurs propres, vecteurs propres associés et base du plan principal 6](#_Toc412550732)

[4. Qualité de la représentation dans le plan principal 7](#_Toc412550733)

[5. Coordonnées de deux premières composantes principales C1 et C2 8](#_Toc412550734)

[6. Représentation graphique du cercle des corrélations et des projections des caractères 9](#_Toc412550735)

[7. Représentation graphique de la qualité de projection des individus 9](#_Toc412550736)

[8. Tableau donnant la qualité de projection des individus sur les deux premiers axe principaux 10](#_Toc412550737)

[9. Tableau des contributions (sur chaque axe principal et globale) des individus (facultatif) 11](#_Toc412550738)

[10. Le nuage de points 12](#_Toc412550739)

# I – Le tableau de données

Nous avons choisi de comparer le niveau de développement des pays de l’Union Européenne par le biais de différentes mesures quantitatives. Nous pouvons ainsi produire un comparatif cohérent tenant compte de variables économiques et démographiques notamment.

Ce sujet nous est apparu pertinent dans la mesure où il nous permet d’avoir une vue d’ensemble de l’avancement des états européens en 2005. Par ailleurs, un des prolongements à ce projet serait de reproduire l’ACP en prenant des variables plus actuelles et ainsi constater les évolutions des pays d’après ces divers indicateurs.

Notre matrice contient 27 individus et 7 caractères. Ces derniers sont :

* espérance de vie
* mortalité infantile
* taux d’activité
* taux de chômage
* pnb par habitant
* pourcentage du budget consacré à l’éducation
* pourcentage du budget consacré à la santé

Individus : 27 pays européens

Ces données sont issues d’une étude menée par l’ONU. Pour plus d’informations, veuillez-vous référer à ce lien :

<http://wanko.free.fr/COURS/COURS%20de%20STAT/stat.pdf>

# II – Le Code : Répartition des tâches

|  |  |
| --- | --- |
| Personne effectuant la tâche | Tâches attribuées |
| Corentin | * Centrer et réduire le tableau : * calcul des caractères centrés * calcul des caractères centrés réduits * Axes principaux et composantes principales   (Longueurs et Orthogonalité)   * Qualité de représentation des individus avec sa représentation graphique |
| Elisa | * Déterminer les valeurs propres et les vecteurs propres associés * Déterminer la base du plan principale * Cercle de corrélation et affichage du cercle de corrélation et des projections des caractères |
| Racel | * Déterminer la matrice de corrélations * Coordonnées des composantes principales avec affichage des coordonnées des deux premières composantes principales * Tableau des contributions |
| Jérôme | * Lire un fichier .xls ou .txt pour le convertir en matrice * Qualité de la représentation * Nuage de points et son affichage |

# III – Présentation des résultats de la première ACP

## Le tableau de données et le tableau centré réduit

Tableau de données :

74.8 4.4 48.8 8.2 26768. 4.3 10.6

75.4 4.8 49. 4.1 29075. 4.9 8.

75.1 5. 42.3 7.3 27952. 5.8 8.8

75.3 5.6 50.9 3.8 12724. 5.8 6.

74.5 5.3 73.8 4.5 30096. 8.1 8.4

75.6 3.9 40.3 11.4 22538. 5.6 7.7

64.9 8.4 52.2 6.8 10201. 6.8 5.5

74.6 3.8 56.8 9.1 27215. 5.6 6.6

75.5 4.6 47.8 8.7 27560. 5.6 9.4

75.4 6.1 37.6 10.3 17670. 2.3 9.2

68.4 9.2 45.6 8.4 12733. 5.2 5.7

73. 5.9 47.5 4.4 32549. 4.5 7.2

76.7 4.5 36. 9.1 26946. 4.6 8.

64.5 10.4 49.7 8.5 7809. 6.2 4.8

65.9 8.6 54.6 10.9 8359. 5.2 5.7

75.2 5.8 42.5 2.4 50410. 4. 6.1

76.2 6. 22.9 7.4 9875. 5.7 8.9

76.2 3.8 69.2 3.9 37070. 7.4 8.5

75.5 5.1 52.9 2.7 29614. 5.2 8.2

70.3 8.1 49.5 18.1 9852. 5.1 4.2

72.4 5.5 54.1 5. 18500. 5.5 8.2

75. 5.6 53. 5.1 26756. 4.7 7.3

69.7 8.6 52.9 17.4 12314. 4.3 6.4

72.3 4.9 51.3 11.3 17762. 5.2 8.2

77.5 3.4 76.2 4.9 26849. 7.3 7.9

Tableau centré réduit :

0.4110657 - 0.7602754 - 0.1715495 0.1393985 0.4346060 - 0.9251686 1.9141819

0.5763080 - 0.5413161 - 0.1529850 - 0.9032647 0.6631125 - 0.3992833 0.2873590

0.4936869 - 0.4318364 - 0.7748949 - 0.0894788 0.5518802 0.3895447 0.7879199

0.5487676 - 0.1033975 0.0233775 - 0.9795571 - 0.9564412 0.3895447 - 0.9640432

0.3284446 - 0.2676169 2.1490098 - 0.8015415 0.7642417 2.4054385 0.5376395

0.6313888 - 1.0339745 - 0.9605397 0.9531844 0.0156278 0.2142496 0.0996487

- 2.3154323 1.4293177 0.1440466 - 0.2166328 - 1.2063424 1.2660203 - 1.2768938

0.3559850 - 1.0887143 0.5710295 0.3682758 0.4788810 0.2142496 - 0.5886225

0.6038484 - 0.6507957 - 0.2643719 0.2665526 0.5130529 0.2142496 1.1633406

0.5763080 0.1703017 - 1.2111601 0.6734455 - 0.4665438 - 2.6781198 1.0382004

- 1.3515188 1.8672363 - 0.4685811 0.1902601 - 0.9555498 - 0.1363406 - 1.1517535

- 0.0846612 0.0608220 - 0.2922186 - 0.8269723 1.0072095 - 0.7498735 - 0.2132019

0.9343330 - 0.7055355 - 1.3596758 0.3682758 0.4522367 - 0.6622260 0.2873590

- 2.4255938 2.5241142 - 0.0880094 0.2156909 - 1.4432682 0.7401349 - 1.7148845

- 2.0400284 1.5387974 0.3668203 0.8260304 - 1.3887911 - 0.1363406 - 1.1517535

0.5212273 0.0060822 - 0.7563304 - 1.3355885 2.7763275 - 1.1881113 - 0.9014731

0.7966311 0.1155619 - 2.5756489 - 0.0640480 - 1.2386325 0.3018971 0.8504900

0.7966311 - 1.0887143 1.7220269 - 0.9541263 1.455011 1.7919056 0.6002096

0.6038484 - 0.3770966 0.2090222 - 1.259296 0.7165000 - 0.1363406 0.4124993

- 0.8282515 1.2650982 - 0.1065738 2.6570487 - 1.2409106 - 0.2239882 - 2.0903052

- 0.2499035 - 0.1581373 0.3204091 - 0.6743874 - 0.3843330 0.1266020 0.4124993

0.4661465 - 0.1033975 0.2183045 - 0.6489566 0.4334174 - 0.5745784 - 0.1506317

- 0.9934938 1.5387974 0.2090222 2.479033 - 0.9970514 - 0.9251686 - 0.7137628

- 0.2774439 - 0.4865762 0.0605064 0.9277536 - 0.4574313 - 0.1363406 0.4124993

Lignes 215 à 232. Fonction utilisée : **importFile** ligne 5 et **calculCarCentreRed** ligne 115.

## Matrice de corrélations

Matrice de corrélation :

1. - 0.8620862 - 0.0035312 - 0.4116394 0.6511477 - 0.0628594 0.7016753

- 0.8620862 1. - 0.1768509 0.3782614 - 0.6166926 - 0.0862422 - 0.6745851

- 0.0035312 - 0.1768509 1. - 0.2425572 0.2178180 0.5999526 - 0.0095787

- 0.4116394 0.3782614 - 0.2425572 1. - 0.5932732 - 0.2757934 - 0.3407958

0.6511477 - 0.6166926 0.2178180 - 0.5932732 1. 0.0006703 0.4231797

- 0.0628594 - 0.0862422 0.5999526 - 0.2757934 0.0006703 1. - 0.0867529

0.7016753 - 0.6745851 - 0.0095787 - 0.3407958 0.4231797 - 0.0867529 1.

Lignes 233 à 237. Fonction utilisée : **correlation** ligne 53.

## Valeurs propres, vecteurs propres associés et base du plan principal

Valeurs propres retenues:

3.3396243

1.7119496

Base du plan:

0.4899267 0.1998996

- 0.4882145 - 0.0655521

0.1297639 - 0.6372324

- 0.3596099 0.2383672

0.4438272 - 0.0233121

0.0651490 - 0.6659745

0.4174128 0.2208926

Lignes 238 à 240. Fonction utilisée : **calculValeurVecteurBase** ligne 160.

## Qualité de la représentation dans le plan principal

Qualité de la représentation :

95.417838 72.165341 58.269668 55.120143 53.083978 50.593595 49.311868

72.165341 48.912845 35.017172 31.867647 29.831481 27.341099 26.059372

58.269668 35.017172 21.121499 17.971974 15.935808 13.445426 12.163699

55.120143 31.867647 17.971974 14.822449 12.786283 10.295901 9.0141742

53.083978 29.831481 15.935808 12.786283 10.750118 8.2597349 6.9780085

50.593595 27.341099 13.445426 10.295901 8.2597349 5.769352 4.4876257

49.311868 26.059372 12.163699 9.0141742 6.9780085 4.4876257 3.2058993

Lignes 242 à 246. Fonction utilisée : **qualiteRepresentation** ligne 14

## Coordonnées des deux premières composantes principales C1 et C2

Composantes principales:

1.4317998 1.3033902

1.2398399 0.3467959

0.9835291 0.5012083

- 0.1268922 - 0.5819949

1.5789951 - 2.9782973

0.4092065 0.9122511

- 2.72153 - 1.7970492

0.6283960 - 0.4174357

1.2106668 0.4977020

- 0.1483210 3.0601242

- 2.6167195 - 0.1899611

0.4974764 0.3969991

0.7708540 1.6811951

- 3.8178149 - 1.3809108

- 3.1061988 - 0.6767623

1.4130594 0.7947932

- 0.1523871 1.7933773

2.5014366 - 2.1888509

1.4412266 - 0.1227278

- 3.430615 0.1691343

0.2487174 - 0.3887523

0.6326106 0.1454370

- 2.9030897 0.6399711

- 0.2638719 0.3516057

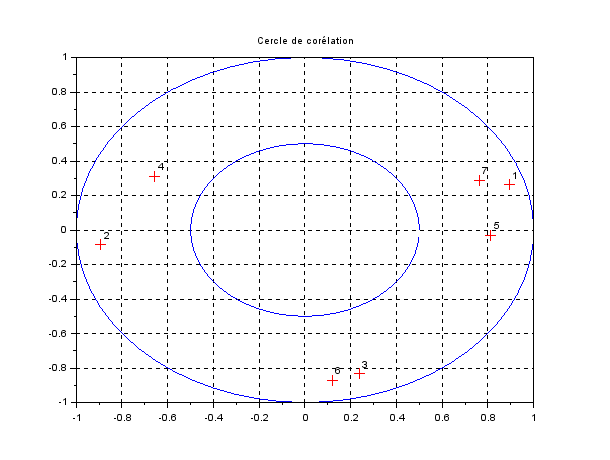
2.1648667 - 2.4573117

2.4252856 0.0561827

- 0.2905263 0.529887

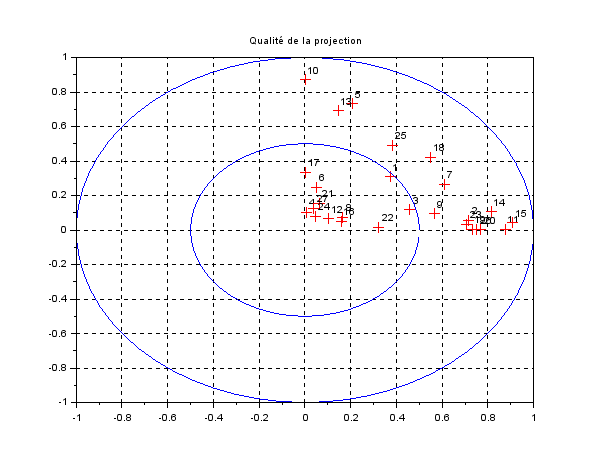
Lignes 247 à 258. Fonction utilisée : **composante** ligne 65.

## Représentation graphique du cercle des corrélations et des projections des caractères



Lignes 260 à 263. Fonctions utilisées : **coordonneCaractere** ligne 76, **afficherCercleCor** ligne 178

## Représentation graphique de la qualité de projection des individus



Lignes 264 à 267. Fonctions utilisées : **QualiteRepresentationIndividu** ligne 198, **afficherCercleCor** ligne 178

## Tableau donnant la qualité de projection des individus sur les deux premiers axes principaux

Qualité de représentation des individus:

0.5673518 0.2179407

0.4977040 0.4706697

0.5174837 0.3265622

0.5192350 0.2787341

0.0095827 0.9424071

0.8589865 0.0024444

0.0853510 0.7067849

0.7701715 0.1490026

0.8827925 0.0583729

0.8083325 0.1596240

0.7226521 0.0532386

0.7478881 0.0005016

0.6926562 0.0401786

0.6153886 0.2511293

0.7974122 0.0641729

Lignes 268 à 271. Fonction utilisée : **QualiteRepresentationIndividu** ligne 198

## Tableau des contributions (sur chaque axe principal et globale) des individus

Contributions:

2.2735431 3.6753115 5.9488546

1.7047859 0.2601916 1.9649774

1.0727858 0.5434778 1.6162636

0.0178570 0.7327970 0.750654

2.765032 19.190282 21.955314

0.1857050 1.800421 1.986126

8.2141922 6.9865892 15.200781

0.4379308 0.3769853 0.8149160

1.6255034 0.5359004 2.1614038

0.0243974 20.25925 20.283647

7.5936919 0.0780683 7.6717602

0.2744628 0.3409764 0.6154392

0.6589961 6.114789 6.773785

16.164726 4.1254994 20.290226

10.700328 0.9908721 11.6912

2.2144172 1.3666383 3.5810556

0.0257534 6.9580671 6.9838205

6.9393314 10.365199 17.30453

2.3035793 0.0325860 2.3361653

13.052166 0.0618883 13.114054

0.0686043 0.3269575 0.3955617

0.4438248 0.0457610 0.4895858

9.3467267 0.8860658 10.232792

0.0772191 0.2674589 0.3446780

5.197577 13.063687 18.261264

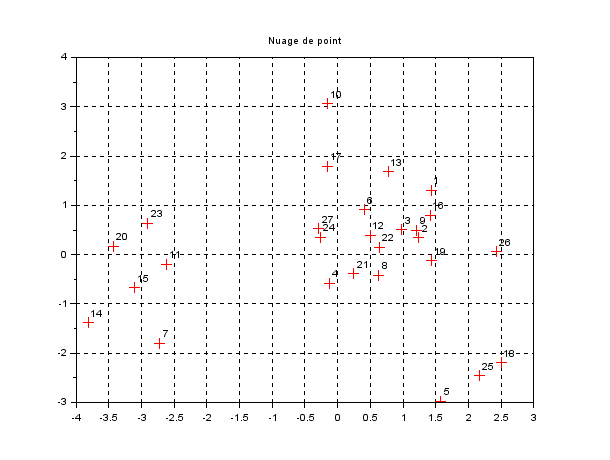
6.5232554 0.0068289 6.5300843

0.0936072 0.6074518 0.7010590

100. 100. 200.

Ligne 273 à 276. Fonction utilisée : **contributionIndividu** ligne 84.

## Le nuage de points



Ligne 278. Fonction utilisée : **tracerNuage** ligne 35

# III – Interprétation des résultats

## Les caractères de la première ACP

## Les individus de la première ACP

1. Les individus bien représentés par leur projection sur le plan sont les individus 2, 5, 7, 10, 11, 13,14, 15, 18, 19, 20, 23, 25 et 26. En effet, la représentation graphique de la qualité de projection des individus nous montrent que ceux-ci sont représentés entre le cercle unitaire et le cercle d’incertitude. Nous pouvons donc les considérer comme étant bien représentés par leur projection sur le plan principal.
2. A l’inverse, les individus moins bien représentés sont les individus 1, 3, 4, 6, 8, 9, 12, 16, 17, 21, 22, 24 et 27. En effet, ceux-ci se trouvent soit à l’intérieur du cercle d’incertitude, soit très proche de celui-ci.

Trop éloignés du cercle unitaire, ces individus ne sont donc pas les plus intéressants à interpréter.

1. Au vu des résultats obtenus par le biais de la représentation graphique de la qualité de projection des individus, nous constatons que plusieurs individus sont mal représentés.

## Conclusion

En définitive, cette ACP nous montre que certaines de nos données de départ ne se sont pas révélées pertinentes du fait de leur manque de représentativité.

# IV- La deuxième ACP