

# Rapport de l'exercice 2 du TP1 de statistique exploratoire : Analyse en Composane Principales

Rémy Gaudré

23/09/2019

Question 1 : Quels renseignements peut-on déduire du tableau des statistique sommaires ?

```
donnees = read.csv("notes_eleves.csv",header = TRUE,sep=";", dec="," )
```

```
# Statistiques Descriptives  
describe(donnees[-1])
```

```
##      vars  n mean  sd median trimmed  mad min  max range  skew kurtosis  
## ORTH    1 27 10.02 4.82   11.0   10.00 6.67 0.0 20.0  20.0  0.02   -0.98  
## GRAM    2 27  7.56 2.88    7.5    7.46 1.48 2.0 14.5  12.5  0.35    0.27  
## EXPR    3 27  8.89 3.05    9.0    8.80 2.22 2.0 16.5  14.5  0.27    0.29  
## RECI    4 27 10.07 3.58   10.0   10.04 5.19 4.0 16.0  12.0  0.00   -1.22  
## MATH    5 27 12.57 2.82   12.0   12.46 2.97 8.0 18.0  10.0  0.39   -0.96  
## ANGL    6 27 10.93 3.33   11.0   10.93 2.97 3.0 17.0  14.0 -0.06   -0.31  
## HIST    7 27  9.20 2.13    9.0    9.11 2.97 5.0 15.0  10.0  0.52    0.20  
## BIOL    8 27  9.59 3.70   10.0    9.54 4.45 2.0 17.0  15.0  0.11   -0.75  
## EDMU    9 27 15.20 3.24   16.0   15.59 2.97 7.0 19.0  12.0 -1.14    0.38  
## ARTS   10 27  9.02 3.93    9.0    9.22 5.19 1.5 14.5  13.0 -0.30   -1.09  
## TECH   11 27 12.63 4.86   14.0   13.26 2.97 0.0 18.0  18.0 -1.27    0.90  
## EPS    12 27 13.07 2.98   13.5   13.15 3.71 5.0 18.5  13.5 -0.37    0.16  
## GEO    13 27 13.80 3.25   14.0   14.20 1.48 0.0 18.0  18.0 -2.65    9.32  
## EXPO   14 27 14.48 4.50   15.0   15.43 2.97 0.0 18.0  18.0 -2.35    5.00  
##      se  
## ORTH 0.93  
## GRAM 0.55  
## EXPR 0.59  
## RECI 0.69  
## MATH 0.54  
## ANGL 0.64  
## HIST 0.41  
## BIOL 0.71  
## EDMU 0.62  
## ARTS 0.76  
## TECH 0.94  
## EPS  0.57  
## GEO  0.63  
## EXPO 0.87
```

On note que toutes les données sont de la même nature. Les variables sont quantitatives : ce sont des notes de contrôle. Elles sont comprises entre 0 et 20.

Question 2 : Relevez les coefficient de corrélation les plus importants.

```
# Matrice des corrélation et P-values
rcorr(as.matrix(donnees[-1]))
```

```
##      ORTH  GRAM EXPR  RECI  MATH  ANGL  HIST  BIOL EDMU  ARTS  TECH  EPS
## ORTH  1.00  0.53 0.51 -0.07  0.38  0.65  0.54  0.35 0.18  0.31 -0.29 -0.04
## GRAM  0.53  1.00 0.35 -0.10  0.34  0.40  0.30 -0.02 0.21  0.15 -0.17 -0.09
## EXPR  0.51  0.35 1.00  0.24  0.59  0.54  0.59  0.40 0.63  0.72  0.00  0.03
## RECI -0.07 -0.10 0.24  1.00  0.18 -0.04 -0.06  0.13 0.41  0.32  0.34  0.24
## MATH  0.38  0.34 0.59  0.18  1.00  0.70  0.64  0.12 0.28  0.37  0.21 -0.06
## ANGL  0.65  0.40 0.54 -0.04  0.70  1.00  0.59  0.17 0.00  0.29 -0.23 -0.27
## HIST  0.54  0.30 0.59 -0.06  0.64  0.59  1.00  0.26 0.22  0.26 -0.01  0.07
## BIOL  0.35 -0.02 0.40  0.13  0.12  0.17  0.26  1.00 0.46  0.58 -0.17  0.10
## EDMU  0.18  0.21 0.63  0.41  0.28  0.00  0.22  0.46 1.00  0.70  0.18  0.29
## ARTS  0.31  0.15 0.72  0.32  0.37  0.29  0.26  0.58 0.70  1.00 -0.05  0.07
## TECH -0.29 -0.17 0.00  0.34  0.21 -0.23 -0.01 -0.17 0.18 -0.05  1.00  0.07
## EPS -0.04 -0.09 0.03  0.24 -0.06 -0.27  0.07  0.10 0.29  0.07  0.07  1.00
## GEO  0.16 -0.07 0.39 -0.11  0.33  0.13  0.33  0.46 0.32  0.24  0.15  0.35
## EXPO 0.14  0.05 0.31  0.33  0.41  0.25  0.26  0.56 0.43  0.57  0.29 -0.22
##      GEO  EXPO
## ORTH  0.16  0.14
## GRAM -0.07  0.05
## EXPR  0.39  0.31
## RECI -0.11  0.33
## MATH  0.33  0.41
## ANGL  0.13  0.25
## HIST  0.33  0.26
## BIOL  0.46  0.56
## EDMU  0.32  0.43
## ARTS  0.24  0.57
## TECH  0.15  0.29
## EPS   0.35 -0.22
## GEO   1.00  0.09
## EXPO  0.09  1.00
##
## n= 27
##
##
## P
##      ORTH  GRAM  EXPR  RECI  MATH  ANGL  HIST  BIOL  EDMU  ARTS
## ORTH           0.0043 0.0063 0.7442 0.0523 0.0002 0.0033 0.0732 0.3599 0.1188
## GRAM 0.0043           0.0700 0.6056 0.0816 0.0368 0.1351 0.9195 0.3046 0.4699
## EXPR 0.0063 0.0700           0.2356 0.0012 0.0038 0.0012 0.0378 0.0004 0.0000
## RECI 0.7442 0.6056 0.2356           0.3675 0.8375 0.7753 0.5032 0.0329 0.0982
## MATH 0.0523 0.0816 0.0012 0.3675           0.0000 0.0003 0.5562 0.1643 0.0566
## ANGL 0.0002 0.0368 0.0038 0.8375 0.0000           0.0011 0.3922 0.9837 0.1468
## HIST 0.0033 0.1351 0.0012 0.7753 0.0003 0.0011           0.1915 0.2673 0.1875
## BIOL 0.0732 0.9195 0.0378 0.5032 0.5562 0.3922 0.1915           0.0169 0.0014
## EDMU 0.3599 0.3046 0.0004 0.0329 0.1643 0.9837 0.2673 0.0169           0.0000
## ARTS 0.1188 0.4699 0.0000 0.0982 0.0566 0.1468 0.1875 0.0014 0.0000
## TECH 0.1371 0.4072 0.9857 0.0819 0.2933 0.2568 0.9418 0.4107 0.3798 0.7893
```

```
## EPS 0.8445 0.6384 0.8961 0.2350 0.7653 0.1676 0.7387 0.6266 0.1389 0.7454
## GEO 0.4182 0.7221 0.0415 0.5819 0.0968 0.5041 0.0939 0.0151 0.1014 0.2287
## EXPO 0.4777 0.7937 0.1099 0.0929 0.0332 0.2018 0.1931 0.0023 0.0241 0.0019
##      TECH    EPS    GEO    EXPO
## ORTH 0.1371 0.8445 0.4182 0.4777
## GRAM 0.4072 0.6384 0.7221 0.7937
## EXPR 0.9857 0.8961 0.0415 0.1099
## RECI 0.0819 0.2350 0.5819 0.0929
## MATH 0.2933 0.7653 0.0968 0.0332
## ANGL 0.2568 0.1676 0.5041 0.2018
## HIST 0.9418 0.7387 0.0939 0.1931
## BIOL 0.4107 0.6266 0.0151 0.0023
## EDMU 0.3798 0.1389 0.1014 0.0241
## ARTS 0.7893 0.7454 0.2287 0.0019
## TECH          0.7298 0.4414 0.1441
## EPS 0.7298          0.0761 0.2737
## GEO 0.4414 0.0761          0.6395
## EXPO 0.1441 0.2737 0.6395
```

Celon la p-value (si la p-value < 0.05, la corrélation est considérée significative) :

- Les notes d'Orthographe sont corrélées avec les notes de Grammaire, d'Expression écrite, d'Anglais, et d'Histoire.
- Les notes de Grammaire sont corrélées avec les notes de Grammaire et d'Anglais.
- Les notes d'Expression écrite sont aussi corrélées avec les notes d'Orthographe, de Mathématique, d'Anglais, d'Histoire, de Biologie, d'éducation musicale, et d'art plastique.
- Les notes de Récitation sont corrélées avec les notes d'éducation musicale.
- Les notes de Mathématique sont corrélées avec les notes d'Expression écrite, d'Anglais, d'Histoire, et d'Exposé.
- Les notes d'Anglais sont corrélées avec les notes d'Orthographe, de Grammaire, d'Expression écrite, de Mathématique et d'Histoire.
- Les notes d'Histoire sont corrélées avec les notes d'Orthographe, d'Expression écrite, de Mathématique et d'Anglais.
- Les notes de Biologie sont corrélées avec les notes d'Expression écrite, d'éducation musicales, d'Art plastique et de Géographie.
- Les notes d'éducation musicale sont corrélées avec les notes d'Expression écrite, de Récitation, de Biologie, d'Art Plastique, et d'Exposé.
- Les notes d'Art Plastique sont corrélées avec les notes d'Expression écrite, de Biologie, d'éducation musicale, et d'Exposé.
- Les notes de Géographie sont corrélées avec les notes d'Expression écrite, de Biologie.
- Les notes d'Exposé sont corrélées avec les notes de Mathématique, de Biologie, d'éducation musicale et d'Art plastique.
- Les notes de technologie et d'EPS ne sont pas corrélées significativement avec d'autres variables.

### Question 3 : Réaliser l'ACP

```
# ACP
donnees.pca = PCA(donnees[-1],scale.unit=TRUE,ncp=4,graph=FALSE)
```

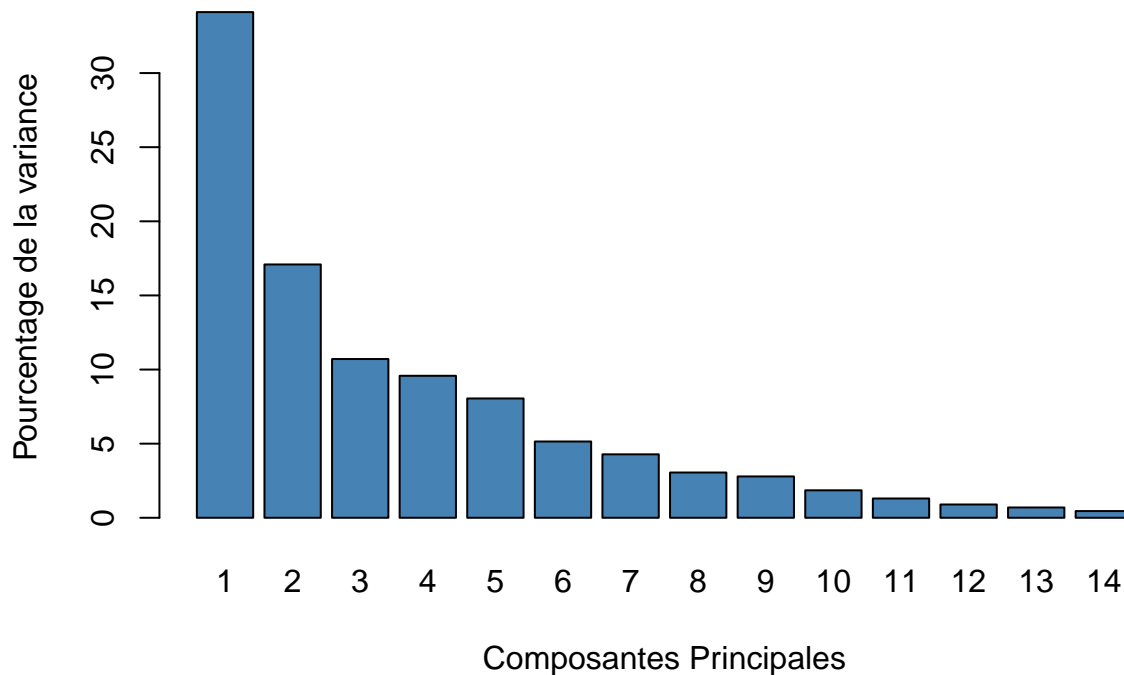
## Question 4 : Combien de composantes principales peut-on retenir ?

```
# Nombre de composantes principales  
donnees.pca$eig
```

```
##          eigenvalue percentage of variance  
## comp 1  4.77570045          34.1121461  
## comp 2  2.39238450          17.0884607  
## comp 3  1.49934217          10.7095870  
## comp 4  1.34084173           9.5774409  
## comp 5  1.12670192           8.0478709  
## comp 6  0.72038480           5.1456057  
## comp 7  0.59919071           4.2799337  
## comp 8  0.42748490           3.0534636  
## comp 9  0.39040787           2.7886277  
## comp 10 0.25931678           1.8522627  
## comp 11 0.18210022           1.3007158  
## comp 12 0.12552548           0.8966105  
## comp 13 0.09728799           0.6949142  
## comp 14 0.06333049           0.4523606  
##          cumulative percentage of variance  
## comp 1          34.11215  
## comp 2          51.20061  
## comp 3          61.91019  
## comp 4          71.48763  
## comp 5          79.53551  
## comp 6          84.68111  
## comp 7          88.96104  
## comp 8          92.01451  
## comp 9          94.80314  
## comp 10         96.65540  
## comp 11         97.95611  
## comp 12         98.85273  
## comp 13         99.54764  
## comp 14        100.00000
```

```
barplot(donnees.pca$eig[, 2],  
        names.arg = 1:nrow(donnees.pca$eig),  
        main = "La variances expliquée pas les composantes principales (%)",  
        xlab = "Composantes Principales",  
        ylab = "Pourcentage de la variance",  
        col = "steelblue")
```

## La variance expliquée par les composantes principales (%)



D'après le critère des 80% de variance expliquée : nous avons 6 composantes principales à retenir.  
 Le critère du coude n'est pas vraiment applicable ici, car la pente la plus importante est juste après le premier axe, donc on garderait qu'un seul axe.  
 D'après le critère de Kaiser, nous avons 5 composantes principales à retenir.

## Question 5 : Interprétez les trois premiers axes par rapport aux disciplines enseignées.

```
## Contributions
donnees.pca$var$contrib
```

	Dim.1	Dim.2	Dim.3	Dim.4
ORTH	8.857178434	10.147700841	2.011115774	0.819262759
GRAM	3.647913177	9.862949216	0.343291721	0.008243911
EXPR	16.001163011	0.009919583	0.146062754	0.142529916
RECI	1.280583878	14.800939040	9.201063052	0.030321946
MATH	11.258878062	1.249038115	6.282449446	13.644601594
ANGL	9.147755175	14.164029878	2.131100334	0.304427329
HIST	10.205943479	3.985966555	0.352172894	8.812289759
BIOL	7.493655757	3.591472838	7.035864125	14.725203802
EDMU	8.707419780	11.968896845	0.632205474	0.643621144
ARTS	12.033404324	4.583971428	0.002441541	9.735102286
TECH	0.001890561	11.091355234	14.496417584	23.845091334
EPS	0.039310859	7.801449997	24.155785266	10.697922436
GEO	4.189905868	2.037057095	19.492996119	10.971312784
EXPO	7.134997636	4.705253336	13.717033917	5.620069002

D'après le critère de la contribution moyenne (ici 7.14%), une variable est considérée comme contributive significativement à un axe lorsque sa contribution est supérieure à 7.14%.

- Sur la première composante principale, les notes d'Orthographe, d'Expression écrite, de Mathématique, d'Anglais, d'Histoire, de Biologie, d'éducation musicale, et d'Art Plastique ont une contribution supérieure à la moyenne.
- Sur la seconde composante principale, les notes d'Orthographe, de Grammaire, de récitation, d'Anglais, d'éducation musicale, de Technologie, et d'EPS ont une contribution supérieur à la moyenne.
- Sur la troisième composante principale, les notes de récitation, de Technologie, d'EPS, de Géographie et d'Exposé ont une contribution significative.

```
## Qualité de représentation (cos²)
donnees.pca$var$cos2
```

	Dim.1	Dim.2	Dim.3	Dim.4
ORTH	4.229923e-01	0.2427720217	3.015351e-02	0.0109850169
GRAM	1.742134e-01	0.2359596680	5.147118e-03	0.0001105378
EXPR	7.641676e-01	0.0002373146	2.189980e-03	0.0019111006
RECI	6.115685e-02	0.3540953710	1.379554e-01	0.0004065693
MATH	5.376903e-01	0.0298817942	9.419541e-02	0.1829525118
ANGL	4.368694e-01	0.3388580549	3.195249e-02	0.0040818887
HIST	4.874053e-01	0.0953596459	5.280277e-03	0.1181588583
BIOL	3.578746e-01	0.0859218394	1.054917e-01	0.1974416771
EDMU	4.158403e-01	0.2863420326	9.478923e-03	0.0086299409
ARTS	5.746793e-01	0.1096662218	3.660706e-05	0.1305323137
TECH	9.028755e-05	0.2653478631	2.173509e-01	0.3197249347
EPS	1.877369e-03	0.1866406803	3.621779e-01	0.1434422081
GEO	2.000974e-01	0.0487342381	2.922667e-01	0.1471079399
EXPO	3.407461e-01	0.1125677514	2.056653e-01	0.0753562303

Une variable est considéré comme bien représenté sur un axe à partir d'un  $\cos^2 > 0.5$  et très bien représenté à partir d'un  $\cos^2 > 0.7$ .

- Sur la première composante principale, les notes d'Expression sont très bien représentés. Les notes de Mathématique et d'Art plastique sont bien représentés.
- Sur la seconde composante principale, les notes de recital sont les mieux représentées avec 35%.
- Sur la troisième composante principale, les notes d'EPS sont les mieux représentées avec 36%.

```
## Coordonnées
donnees.pca$var$coord
```

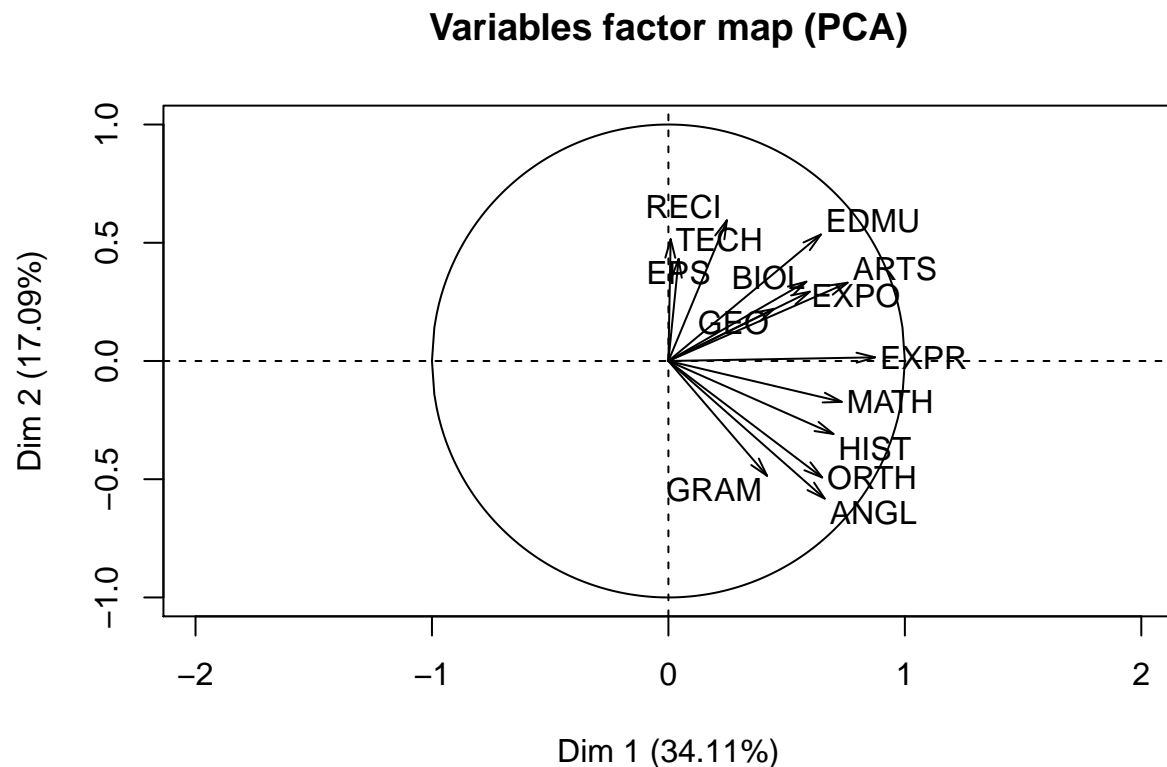
	Dim.1	Dim.2	Dim.3	Dim.4
ORTH	0.650378590	-0.49271901	0.173647652	-0.10480943
GRAM	0.417388795	-0.48575680	-0.071743415	-0.01051370
EXPR	0.874166811	0.01540502	0.046797227	0.04371614
RECI	0.247299110	0.59505913	-0.371423503	0.02016356
MATH	0.733273680	-0.17286351	-0.306912714	0.42772948
ANGL	0.660960956	-0.58211516	-0.178752583	0.06388966
HIST	0.698144175	-0.30880357	0.072665513	0.34374243
BIOL	0.598226171	0.29312427	0.324794825	-0.44434410
EDMU	0.644856795	0.53510937	0.097359762	-0.09289748
ARTS	0.758076081	0.33115891	0.006050377	-0.36129256
TECH	0.009501976	0.51511927	-0.466209076	0.56544225
EPS	0.043328615	0.43201931	0.601812160	0.37873765
GEO	0.447322427	0.22075833	0.540616973	0.38354653
EXPO	0.583734627	0.33551118	-0.453503335	-0.27451089

- Les notes d'Orthographe, d'Expression écrite, de Mathématique, d'Anglais, d'Histoire, de Biologie, d'éducation musicale, et d'Art Plastique ont sont toutes corrélées positivement sur la première composante principale.

- Sur la seconde composante principale, les notes d'Orthographe, de Grammaire, d'Anglais, sont inversement corrélées aux notes de Récitation, d'éducation musicale, de Technologie, et d'EPS.
- Sur la troisième composante principale, les notes de Récitation, de Technologie et d'Exposé sont inversement corrélées aux notes d'EPS et de Géographie.

**Question 6 : Quel phénomène peut-on déduire du cercle des corrélations formé par les deux premiers facteurs. Quelles sont les disciplines les mieux représentées dans ce plan ?**

```
## Cercle des corrélations formé par les deux premiers facteurs
plot(donnees.pca, axes=c(1,2), choix = "var")
```



On remarque au premier coup d'oeil que les variables sont toutes positivement corrélées sur la première composante. On remarque aussi qu'une bonne majorité de l'information est représentée par ces 2 axes comme les flèches atteignent au moins ou presque le demi-rayon. On observe aussi 2 groupes de variables se comportant de la même façon sur ces 2 premiers axes : éducation musicale, Exposé, Géographie, Biologie et Art plastique sont corrélées positivement sur les 2 axes, tandis que Mathématique, Histoire, Orthographe, Grammaire et Anglais sont corrélées positivement sur le premier axe et négativement sur le second. Les disciplines les mieux représentées dans ce plan sont Education Musicale, Art Plastique, Expression écrite, Mathématique, Histoire, Orthographe et Anglais.

**Question 7 : Que peut-on déduire de ce phénomène pour les 27 élèves de cette classe de 6ème ?**

D'après ces 2 groupes de matières en général, plus un élève est bon en Math, en Histoire en Orthographe et en Anglais, moins il est bon en éducation musicale, en Exposé, en Géographie, en Art Plastique et en Biologie et inversement.

## Question 8 : Interprétez les contributions des élèves EL10 et EL12 par rapports aux axes factoriels 1 et 2.

```
## Contribution des individus
donnees.pca$ind$contrib
```

##	Dim.1	Dim.2	Dim.3	Dim.4
## 1	9.328905241	9.59126708	0.616515806	0.17696256
## 2	0.125275069	3.96118861	0.258445075	12.52737590
## 3	0.090900704	0.16283004	5.087265764	0.32642387
## 4	0.565357272	1.30848854	9.031255314	20.89056129
## 5	1.487781989	1.64284957	0.148851085	0.30166113
## 6	0.707432497	0.57129277	3.707371169	3.71293267
## 7	1.853709107	0.76715711	5.398964876	1.60627596
## 8	0.112046826	2.41932985	0.013041174	0.11265310
## 9	7.407644214	1.52394402	0.009609208	0.49971364
## 10	18.588377979	7.22537004	0.003702659	2.66022614
## 11	0.306502245	1.89876424	1.487091590	1.53197702
## 12	20.985890664	0.00230761	2.387411364	7.84839701
## 13	3.231582745	14.07178676	14.446330034	0.20323916
## 14	0.043589696	0.38069178	0.008154061	0.04808136
## 15	0.158993355	1.06053122	0.105021975	0.57298895
## 16	0.008219267	6.63073046	0.008769846	5.07773785
## 17	1.187170818	5.85814498	1.849005639	1.56201943
## 18	0.293940859	1.67494396	3.919286109	0.04749646
## 19	0.002322024	5.62257996	0.012798694	1.96367488
## 20	2.054270051	0.03005751	4.669204845	0.47070515
## 21	12.582779600	5.23836937	0.273267018	1.38426323
## 22	1.781543769	5.83680752	1.427638803	0.02488253
## 23	4.620411054	16.67349157	22.144458992	1.15307352
## 24	11.219745959	0.10732552	15.428296052	26.35988703
## 25	0.484483948	0.01002931	5.460474394	7.68881226
## 26	0.057921681	5.00334513	1.027681671	1.23435008
## 27	0.713201366	0.72637544	1.070086782	0.01362780

La contribution moyenne des individus est de  $100/27 = 3.704$ . La contribution de EL10 à l'axe 1 est de 18.5884 soit plus de 5 fois la contribution moyenne. La contribution de EL12 à l'axe 1 est de 20.9859 soit plus de 5 fois la contribution moyenne aussi. Les élèves EL10 et EL12 contribuent donc beaucoup à l'axe 1. La contribution de EL10 à l'axe 2 est de 7.2253 soit presque 2 fois la contribution moyenne. La contribution de EL12 à l'axe 2 est de 0.0023 soit très inférieur à la moyenne. L'élève EL10 a donc plus contribué à la construction de l'axe 2 de l'ACP que l'élève EL12.

## Question 9 : Quels sont les élèves parmi les 27 qui expliquent les axes 1,2 et 3 ?

Les élèves EL01, EL09, EL10, EL12, EL21 et EL21 expliquent le premier axe. Les élèves EL01, EL02, EL10, EL13, EL16, EL17, EL19, EL21, EL22, EL23 et EL26 expliquent le second axe. Les élèves EL03, EL04, EL07, EL13, EL20, EL23, EL24 et EL25 explique le troisième axe.



### **Question 10 : Faites un résumé des résultats de l'ACP. Peut-on dégager des tendances quant aux matières enseignées dans cette classe.**

On peut distinguer 2 groupes de disciplines enseignées dans cette école selon les résultats qu'ont obtenu les élèves. Le premier groupe est composé de : Technologie, EPS, Récitation, Exposé, Géographie, Education Musicale, Art Plastique, et Biologie. Le second groupe est quant à lui composé de : Mathématique, Histoire, Orthographe, Anglais et Grammaire. L'expression écrite est une discipline qui se trouve entre les deux groupes. Les deux groupes sont corrélés positivement selon la première composante principale. C'est à dire que si un élève a eu quelques bonnes notes, le reste des notes a tendance à être bon aussi et inversement. Maintenant la seconde composante principale nous indique 2 groupes de notes principaux. Cela nous montre que certaines disciplines ont plus de chance d'avoir des notes similaires. L'élève EL10 qui a beaucoup contribué à ces deux premiers axes confirme ces tendances : il a de très bonnes notes dans un premier groupe de discipline : 20 en Orthographe, 14.5 en Grammaire, 18 en Mathématique, 15 en Histoire, et 16.5 en Anglais. Il a aussi un très bon résultat en Expression écrite : 16.5. Les notes du second groupe de discipline reste au dessus de la moyenne donc bonnes mais pas aussi biens : 10 en Récitation, 13.5 en Art plastique, 10.5 en Biologie, 13 en Géographie, 12 en Technologie et 15 en Exposé. Cet exemple a beaucoup contribué aux 2 premiers axes parce qu'il représente bien ce que les axes décrivent.

On note aussi que les notes de Géographie et d'EPS ont leur meilleure représentation sur la troisième composante principale, et ces disciplines sont corrélées positivement entre elles sur cette axe. Elles sont cependant inversement corrélées aux notes de Récitation, de Technologie et d'Exposé. Ce comportement reste moins important que ceux décrits par les deux premiers axes.