

## Analyse en composantes principales.

Exemple d'une ACP non normée (statistiques exploratoires).

Le tableau suivant est composé de 9 élèves et de 5 notes. La signification des abréviations des notes est la suivante : **math**-mathématiques, **scie**-sciences, **fran**-français, **lati**-latin et **d.m**-dessin-musique

	math	scie	fran	lati	dm
jean	6	6.0	5	5.5	8.0
aline	8.0	8.0	8.0	8.0	9.0
annie	6	7	11	9.5	11.0
monique	14.5	14.5	15.5	15	8.0
didier	14	14	12	12.5	10.0
andre	11	10	5.5	7	13.0
pierre	5.5	7	14	11.5	10.0
brigitte	13	12.5	8.5	9.5	12.0
evelyne	9	9.5	12.5	12	18.0

TABLE 1 – Fichier notes.txt

En utilisant la métrique identité et en donnant la même importance à chaque individu, on obtient les résultats de l'ACP avec le package **FactoMineR**. Il s'agit maintenant d'interpréter les résultats. Les réponses aux questions posées demandent généralement très peu de calcul.

1. Quelles sont les coordonnées de l'individu **jean** dans l'ancien premier axe ? dans le nouveau ? Que représente la première composante principale ?
2. Quelle est la valeur de l'inertie du nuage projeté sur le premier axe principal ? sur le deuxième ? sur le troisième ?  
Quelle est l'inertie du nuage projeté sur le premier plan principal ? sur les trois premiers axes principaux réunis ?  
Traduire ces inerties en pourcentage d'inertie. Combien d'axes cela incite-il à retenir ?
3. Quel est le coefficient de corrélation linéaire de la première composante principale avec la variables **math** ? avec les autres variables ? Mêmes questions avec la second puis la troisième composante principale ? Quel est le nom du graphique qui représente ces corrélations ?  
À partir de ces éléments, interpréter les trois premiers nouveaux axes.
4. Quels sont les individus qui contribuent le plus à l'inertie du premier axe principal ? du second ?
5. Quels sont les individus bien représentés sur le premier axe principal ? le second ?  
Quels sont les individus bien représentés dans le premier plan principal ?

Voici le code R et les sorties obtenues.

```
> library(FactoMineR)
> x <- read.table("notes.txt")
> res.pca <- PCA(x,scale.unit=FALSE,graph=FALSE)
> res.pca$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	28.253249801	57.68876314	57.68876
comp 2	12.074723274	24.65471604	82.34348
comp 3	8.615733579	17.59199445	99.93547
comp 4	0.021732182	0.04437375	99.97985
comp 5	0.009869805	0.02015261	100.00000

```
> res.pca$var$coord
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
math	2.7349502	-1.9700748	0.15064642	-0.042566843	-0.056880954
scie	2.6948360	-1.2926370	0.04242315	0.081530696	0.054278457
fran	2.6170525	2.2598942	-0.31720044	0.058043833	-0.040710559
lati	2.5759832	1.1232120	-0.06617039	-0.099388800	0.045047500
dm	0.1627982	0.3922897	2.91311601	0.005076587	-0.001258566

```
> res.pca$var$cor
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
math	0.81041737	-0.5837703	0.04463938	-0.012613359	-0.0168549005
scie	0.90106403	-0.4322151	0.01418490	0.027261168	0.0181489210
fran	0.75354235	0.6507038	-0.09133327	0.016712881	-0.0117220157
lati	0.91570696	0.3992778	-0.02352216	-0.035330594	0.0160134233
dm	0.05529981	0.1332542	0.98953631	0.001724431	-0.0004275134

```
> res.pca$ind$coord
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
jean	-8.700907	-1.7027046	-2.5539182	-0.14945398	0.11731596
aline	-3.938596	-0.7085441	-1.8104644	-0.09068389	-0.04349922
annie	-3.209392	3.4590552	-0.3006617	0.17254286	-0.01928215
monique	9.755741	0.2157421	-3.3436726	-0.17347137	-0.10041455
didier	6.371422	-2.1733326	-0.9570588	0.07066256	0.18799232
andre	-2.974017	-4.6509322	2.6349457	-0.02321315	-0.14809545
pierre	-1.050967	6.2271742	-1.6880636	0.11529582	-0.04281219
brigitte	1.980533	-4.0685562	1.4007122	0.24321198	-0.01039742
evelyne	1.766183	3.4020982	6.6181814	-0.16489082	0.05919270

```
> res.pca$ind$cos2
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
jean	0.88894548	0.0340427538	0.076587880	2.622775e-04	1.616070e-04
aline	0.80365559	0.0260088320	0.169811511	4.260371e-04	9.802799e-05
annie	0.46012197	0.5344933508	0.004038159	1.329907e-03	1.660880e-05
monique	0.89414954	0.0004372797	0.105035734	2.827126e-04	9.472899e-05
didier	0.87726184	0.1020725256	0.019794006	1.079034e-04	7.637245e-04
andre	0.23623016	0.5777346580	0.185435015	1.439185e-05	5.857764e-04
pierre	0.02583884	0.9071461473	0.066661161	3.109728e-04	4.287757e-05
brigitte	0.17435766	0.7357964730	0.087211720	2.629342e-03	4.805383e-06
evelyne	0.05330060	0.1977673369	0.748407627	4.645728e-04	5.986834e-05

```
> res.pca$ind$contrib
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
jean	29.7726930	2.66783459	8.4116112	11.4200796	15.4939841
aline	6.1005911	0.46197026	4.2271263	4.2045012	2.1301584
annie	4.0507433	11.01020178	0.1165793	15.2211729	0.4185620
monique	37.4291227	0.04283019	14.4182553	15.3854355	11.3512117
didier	15.9647400	4.34643006	1.1812517	2.5528947	39.7858978
andre	3.4783719	19.90491457	8.9538242	0.2755005	24.6906420
pierre	0.4343771	35.68307951	3.6748775	6.7964365	2.0634020
brigitte	1.5425990	15.23214081	2.5302476	30.2429457	0.1217026
evelyne	1.2267619	10.65059822	56.4862269	13.9010335	3.9444393

```
> par(mfrow=c(1,2))
```

```
> barplot(res.pca$eig[,1],main="pourcentage d'inertie",names.arg=1:nrow(res.pca$eig))
```

```
> barplot(cumsum(res.pca$eig[,1]),main="pourcentage d'inertie cumulée",names.arg=1:nrow(res.pca$eig))
```

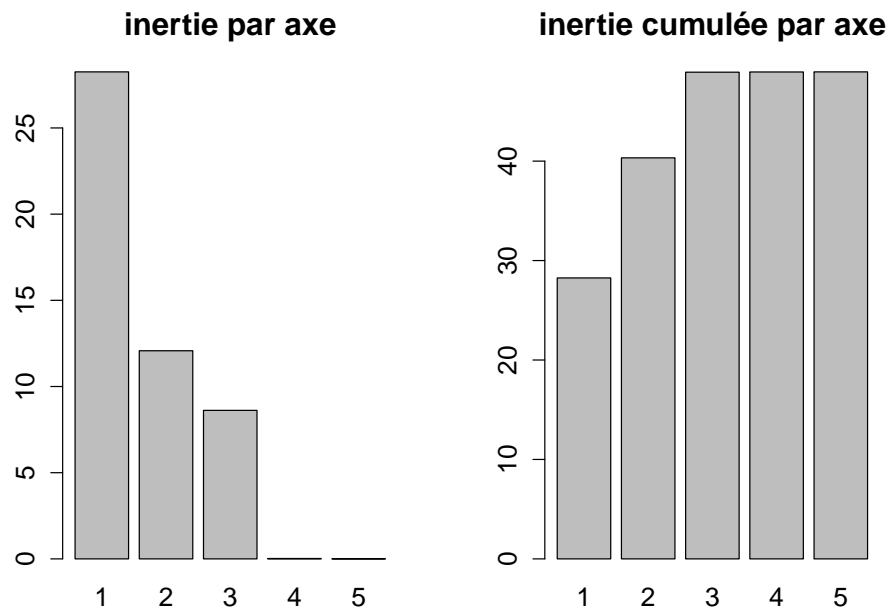


FIGURE 1 – Valeurs des inerties et des inerties cumulées

```
> par(mfrow=c(1,2))
> plot(res.pca,choix="ind",axes=1:2)
> plot(res.pca,choix="ind",axes=c(1,3))
```

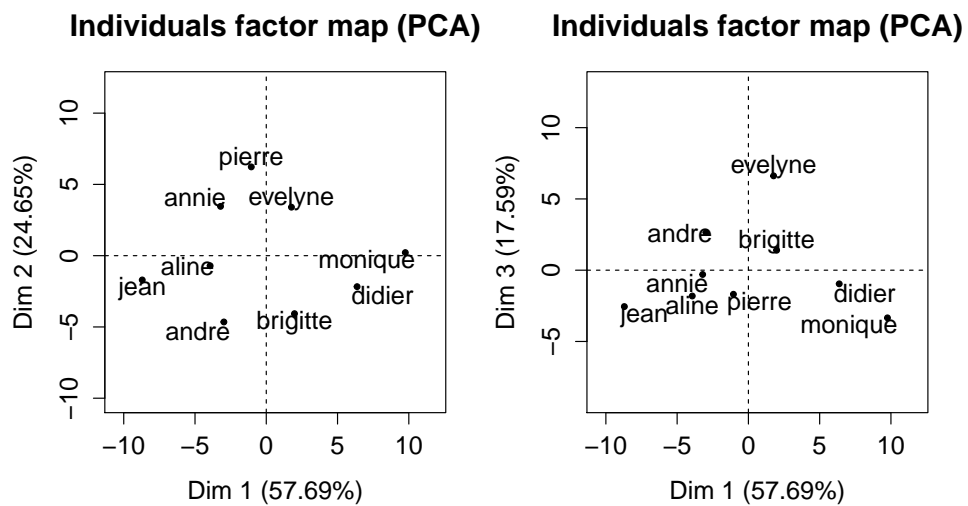


FIGURE 2 – Individus sur les deux premier plans factoriels

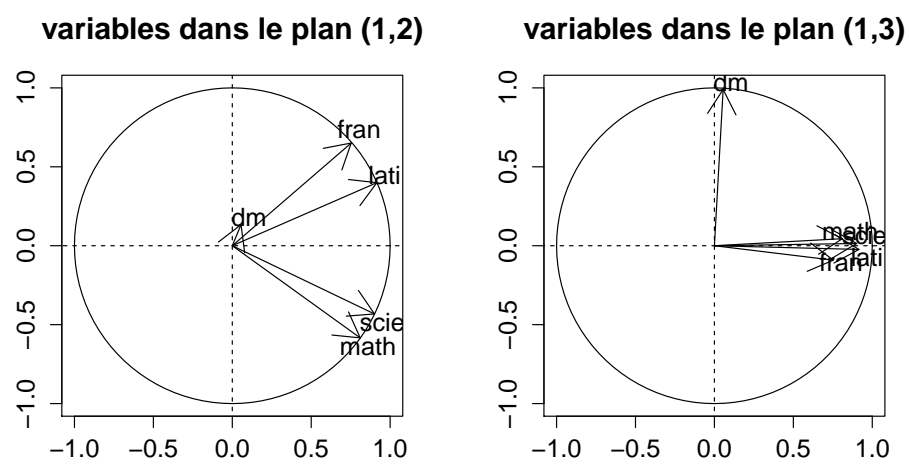


FIGURE 3 – Variables sur les deux premier plans factoriels