

Université Claude Bernard - Lyon 1

MADG

Année 2017-2018

Pr Jean R. LOBRY

NOM Prénom

1 De la nature des données

```
load(url("http://pbil.univ-lyon1.fr/R/donnees/resBioInfAtLyon.Rda"))
ls()
[1] "cent" "res"
class(res)
[1] "data.frame"
dim(res)
[1] 24 13
class(cent)
[1] "numeric"
length(cent)
[1] 13
```

ON s'intéresse aux résultats en première session de 24 étudiants¹ en première année de master « bioinfo@lyon » dans 13 UE². Les notes des étudiants sont données dans la table 1 page 2. La liste des matières est la suivante :

```
names(res)
[1] "MADG" "P.BI.1" "B.BI" "BI.S" "P.O." "Angl" "MADP" "P.BI.2" "A.BI"
[10] "MADT" "S.BI" "TRIP" "BD.B"
```

LES notes sont exprimées ici en ECTS³, par exemple une note de 20/20 pour une UE de 6 ECTS correspond à une note de 6/6. Les faibles valeurs des notes correspondent à de mauvais résultats, et, *a contrario*, les fortes valeurs des notes à de bons résultats. Le nombre d'ECTS pour chaque UE est le suivant :

1. On a supprimé ici 3 étudiants pour lesquels il y avait des données manquantes en raison d'une absence à l'examen d'au moins une UE.

2. Acronyme de « Unité d'Enseignement ». Une UE peut regrouper plusieurs sous-disciplines, par exemple ici l'UE « TRIP » correspond au regroupement de « Recherche stage et emploi » et de, au choix, « Application juridique » ou bien « Réseau ».

3. Acronyme de « *European Credit Transfer Scale* ».

2*cent

MADG 6 P.BI.1 3 B.BI 6 BI.S 6 P.O. 6 Angl 3 MADP 3 P.BI.2 3 A.BI 3 MADT 6 S.BI 9 TRIP 3
BD.B 3

	MADG	P.BI.1	B.BI	BI.S	P.O.	Angl	MADP	P.BI.2	A.BI	MADT	S.BI	TRIP	BD.B
1	3.20	2.20	4.00	3.60	4.10	1.80	1.90	2.30	2.10	4.30	5.70	2.40	1.70
2	3.40	2.30	4.20	4.10	4.50	2.10	2.10	2.30	2.40	3.60	7.10	2.00	2.30
3	2.30	2.20	4.60	3.80	2.80	2.80	2.00	1.80	0.91	3.20	6.10	2.30	2.00
4	3.00	2.20	3.40	3.50	2.10	1.70	1.90	2.30	1.80	4.00	5.80	2.00	1.40
5	2.60	1.80	3.60	4.60	3.30	1.60	2.00	2.10	2.10	3.00	6.20	2.10	1.60
6	3.10	2.20	4.30	3.50	3.00	2.30	2.10	2.10	1.10	3.50	5.20	2.40	1.20
7	2.60	2.00	3.60	3.80	3.10	1.20	2.10	1.80	1.20	2.80	5.30	2.30	1.60
8	2.90	2.30	3.80	4.00	3.00	2.90	2.10	1.80	0.87	3.60	6.70	2.40	1.00
9	3.40	1.80	3.50	3.90	3.50	1.60	2.00	2.30	2.50	3.80	6.10	2.20	2.30
10	2.30	2.30	4.10	4.90	3.90	1.60	2.00	2.30	2.30	3.60	5.30	2.60	2.00
11	4.60	2.60	4.70	4.00	5.30	2.20	2.10	2.30	3.00	4.70	7.50	2.30	2.50
12	3.60	2.60	4.30	4.40	3.90	2.00	2.00	2.30	1.40	2.90	6.60	1.80	1.50
13	1.00	1.80	3.50	3.00	2.30	2.20	2.50	1.80	0.33	2.30	7.20	2.30	1.40
14	1.80	2.20	4.00	3.50	2.60	2.00	2.50	1.80	1.60	2.40	7.00	2.10	1.30
15	1.70	2.00	3.20	1.40	0.74	0.60	1.70	2.30	1.40	1.50	4.50	1.60	1.30
16	2.10	2.00	4.20	3.10	3.70	2.50	2.10	2.10	1.60	3.10	6.90	2.00	1.70
17	3.10	2.00	3.40	4.00	2.00	1.70	2.10	2.30	1.60	3.60	5.80	2.40	1.40
18	4.80	2.60	5.20	4.60	4.90	3.00	2.50	2.30	3.00	5.60	7.40	2.50	2.00
19	3.60	2.00	3.80	4.60	5.00	1.60	2.20	2.10	2.10	3.60	7.10	2.20	1.80
20	2.80	2.20	4.20	3.40	3.20	1.90	2.00	2.30	1.40	2.60	6.80	2.40	1.80
21	4.10	2.00	4.90	4.70	4.80	1.80	2.20	2.30	2.10	4.20	6.90	2.30	1.80
22	2.50	2.60	3.60	3.70	3.00	1.90	2.00	2.10	1.80	1.90	5.80	2.00	1.70
23	2.60	2.00	4.60	4.70	4.30	1.70	2.00	2.10	2.40	2.30	7.60	2.30	1.20
24	3.70	2.00	4.40	3.40	3.90	1.90	2.00	2.30	1.20	3.50	5.70	2.20	1.60

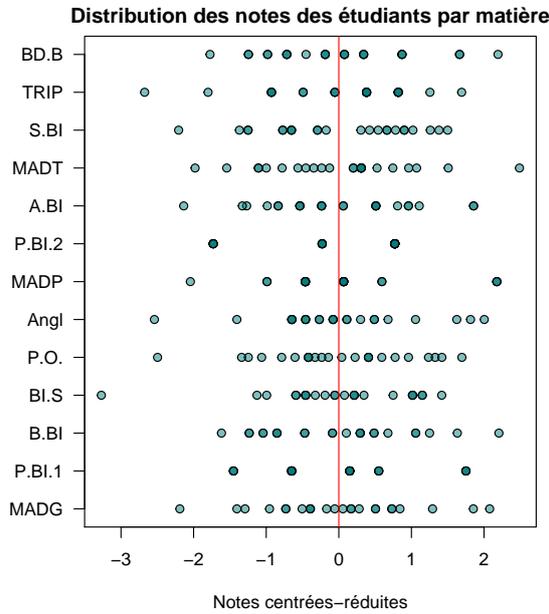
TABLE 1 – Notes des étudiants en ECTS.

La distribution des notes des élèves est donnée dans le graphique ci-après. La croix bleue donne la note minimum qu'il faut obtenir pour valider directement, sans compensation⁴, une UE. Par exemple il faut au moins 4.5 pour l'UE de stage « S.BI » à 9 ECTS. La croix rouge donne la note minimum qu'il faut obtenir pour que l'UE soit validable par compensation, par exemple il faut au moins 2.7 pour l'UE « MADG » à 6 ECTS.

```
par(mar = c(5, 8, 2, 1))
mybg <- rgb(0, 0.5, 0.5, 0.5)
stripchart(res, pch = 21, bg = mybg, las = 1,
  xlab = "Notes en ECTS", main = "Distribution des notes des étudiants par UE")
points(cent, 1:length(cent), pch = 3, col = "blue", cex = 1.5)
points(cent*9/10, 1:length(cent), pch = 3, col = "red", cex = 1.5)
```

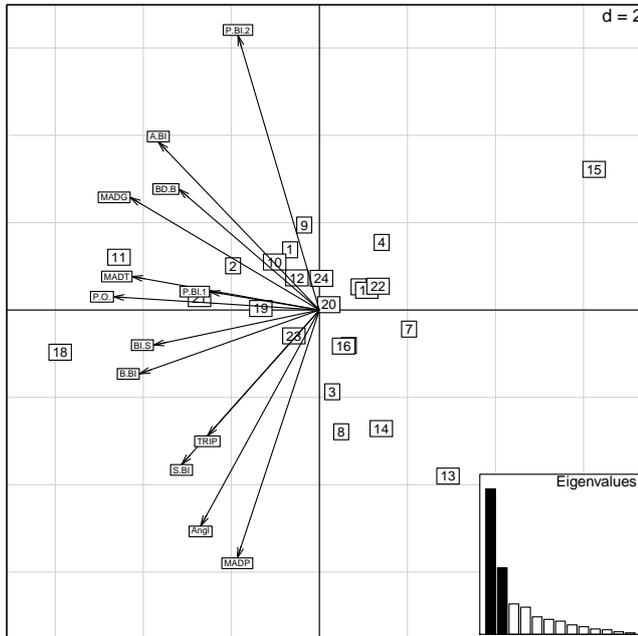
4. Une UE ayant une note $\geq 9/20$ peut être validée par compensation.


```
stripchart(df, pch = 21, bg = mybg, las = 1,
  xlab = "Notes centrées-réduites", main = "Distribution des notes des étudiants par matière")
abline(v = 0, col = "red")
```



Le résultat de l'ACP sur ces valeurs centrées-réduites est le suivant :

```
library(ade4)
scatter(dudi.pca(res, scann = FALSE, nf = 2), clab.col = 0.5, posieig = "bottomright")
```



A U vu du graphe des valeurs propres, combien de facteurs rentiendriez-vous ?

Réponse : Deux facteurs dominant le graphe des valeurs propres.

COMMENT interprétez-vous le premier facteur ?

Réponse : Le premier facteur oppose les bons (à gauche) aux mauvais étudiants (à droite). C'est un effet taille classique.

QUEL est le numéro de l'étudiant ayant globalement les pires résultats ?

Réponse : 15

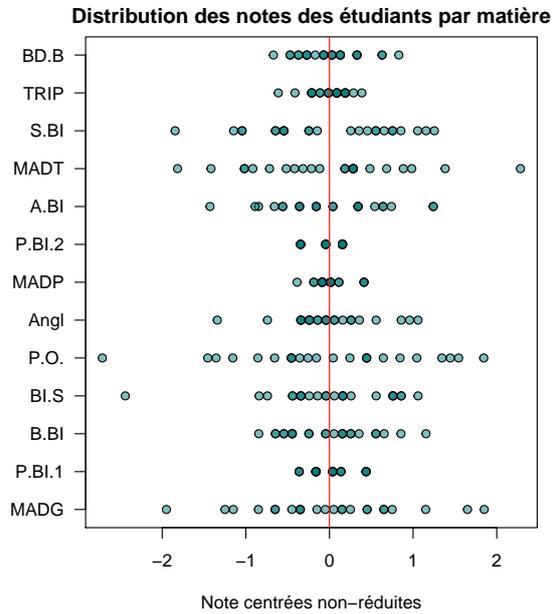
POURQUOI est-ce une ACP centrée-réduite qui est généralement l'option par défaut des programmes d'ACP ?

Réponse : Parce qu'il y a pas de raison en général pour que toutes les variables soient exprimées dans une même unité

3 ACP centrée non-réduite

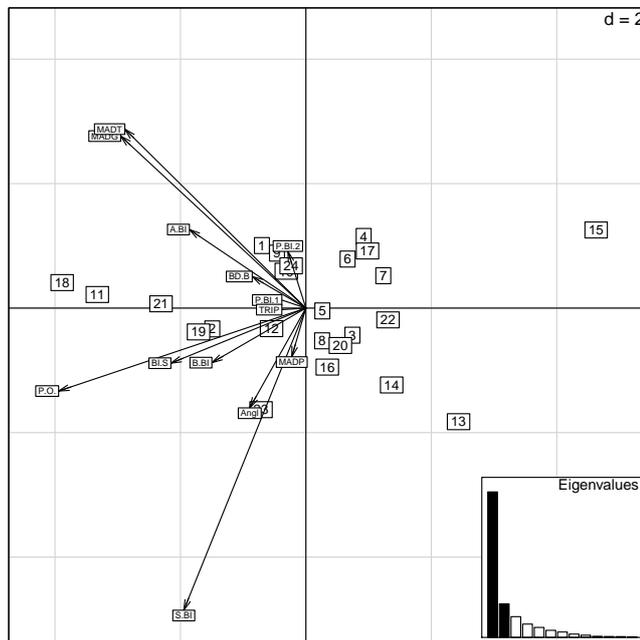
DANS cette analyse on considère les notes centrées sur la moyenne par colonne, c'est à dire la moyenne de la promotion, mais non-réduites. La distribution des valeurs est donnée dans la figure ci-après :

```
par(mar = c(5, 8, 2, 1))
df <- as.data.frame(scale(res, scale = FALSE))
stripchart(df, pch = 21, bg = mybg, las = 1,
  xlab = "Note centrées non-réduites",
  main = "Distribution des notes des étudiants par matière")
abline(v = 0, col = "red")
```



Le résultat de l'ACP sur ces valeurs centrées sur la moyenne par colonne mais non-réduites est le suivant :

```
scatter(dudi.pca(res, scale = FALSE, scann = FALSE, nf = 2), clab.col = 0.5, posieig = "bottomright")
```



Au vu du graphe des valeurs propres, combien de facteurs retiendriez-vous ?

Réponse : un facteur domine le graphe des valeurs propres.

COMMENT interprétez-vous le premier facteur ?

Réponse : Le premier facteur oppose les bons (à gauche) aux mauvais étudiants (à droite). C'est un effet taille classique.

QUEL est le numéro de l'étudiant ayant globalement les meilleurs résultats ?

Réponse : 18

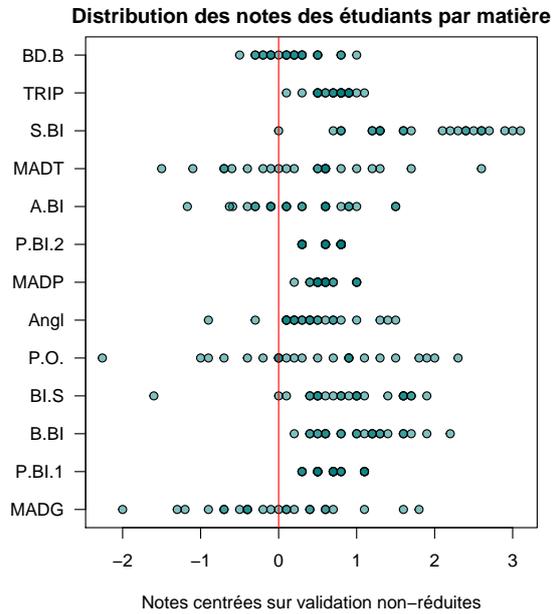
QUEL est l'intérêt de ne pas réduire les valeurs ?

Réponse : Conserver la même unité de mesure en ECTS

4 ACP à centrage de validation non réduite

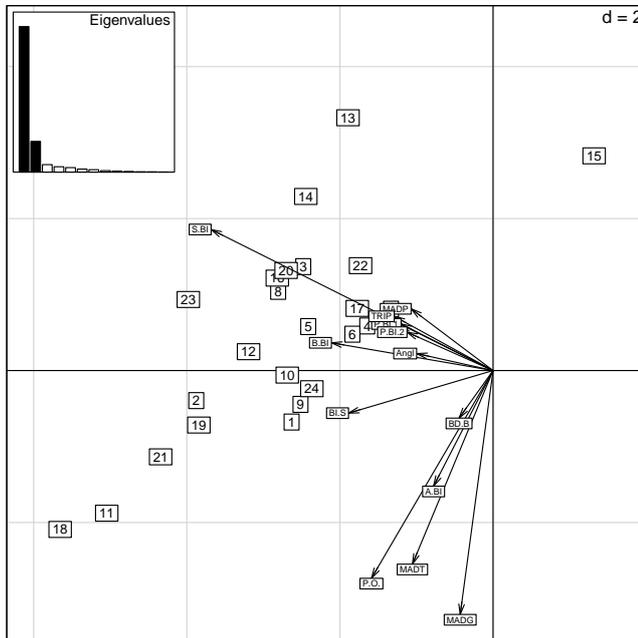
ON réalise une ACP centrée sur la note qu'il faut obtenir pour valider l'UE et non réduite. La distribution des valeurs est donnée dans la figure ci-après :

```
par(mar = c(5, 8, 2, 1))
df <- as.data.frame(scale(res, center = cent, scale = FALSE))
stripchart(df, pch = 21, bg = mybg, las = 1,
           xlab = "Notes centrées sur validation non-réduites",
           main = "Distribution des notes des étudiants par matière")
abline(v = 0, col = "red")
```



Le résultat de l'ACP est le suivant :

```
scatter(dudi.pca(res, scann = FALSE, nf = 2, scale = FALSE,
center = cent), clab.col = 0.5)
```



A U vu du graphe des valeurs propres, combien de facteurs rentiendriez-vous ?

Réponse : deux facteurs dominent le graphe des valeurs propres.

QUELLE serait la position sur ce plan factoriel d'un étudiant ayant eu exactement 10/20 à toutes les UE ?

Réponse : à l'origine des axes (0, 0).

COMMENT interprétez-vous le premier facteur ?

Réponse : le premier facteur est un effet taille opposant les bons au mauvais étudiants

COMMENT interprétez-vous le deuxième facteur ?

Réponse : le deuxième facteur oppose les matières difficiles en bas aux matières faciles en haut

QUELLES sont les trois UE que vous conseilleriez à un élève moyen de travailler en priorité pour améliorer ses résultats ?

Réponse : MADG Prog.Obj.BioInf et MADT

UN collègue biologiste, et néanmoins ami, vous demande ce qu'est une ACP et à quoi ça sert. Comment lui répondriez-vous ?

Réponse : réponse ouverte