Rapport de SY09 :
Analyse de données et data mining
UNIVERSITÉ DE TECHNOLOGIE DE COMPIÈGNE

Printemps 2017

Cécile HEIDSIECK et Simon LAURENT

Sujet du rapport:
Statistique descriptive et ACP

Département des étudiants :
Génie Biologique et Génie Informatique

Professeurs :

M. Sylvain ROUSSEAU
M. Benjamin QUOST

Table des matières

1 Statistique Descriptive

1.1 Notes

1.1.1 Analyse

Dans cette première partie, nous allons décrire les données du dataset Notes : *Pour l'intégralité des variables, le dataset à un volume de 296.*

- nom : variable qualitative nominale, modalités "EtuX" avec X un chiffre ;
- spécialité : variable qualitative nominale, modalités GB, GI, GM, GSM, GP, GSU, HuTech, ISS, TC ;
- niveau : variable quantitative discrète, modalités 1..6 ;
- statut : variable qualitative nominale, modalités : Échange, UTC ;
- dernier diplôme obtenu : variable qualitative nominale, 12 modalités avec certaines en N/A ;
- note médian, note finale, note totale : variable quantitative continu, modalités entre 0 et 20 ;
- correcteur médian, correcteur final : variable qualitative nominale, modalités "CorX" avec X un chiffre ;
- résultat : qualitative ordinale, modalités A>B>C>D>E>Fx>F

Les valeurs nulles pour les notes du médian et du final sont des absences. Quant au NAs pour la note finale et les résultats sont la conséquence des unions des valeurs nulles. Par ailleurs, on remarque que les étudiants étrangers ne disposent pas de dernier diplôme obtenu et que leur résultat est souvent NA ou Fx ou F.
La question que l'on se pose est "existe t il une corrélation entre les notes du médian et du final ?". Le résultat est le suivant **0.3861882**. Ceci montre qu'il y a une faible corrélation entre la réussite d'un médian et d'un final. Néanmoins, une corrélation existe entre note final et résultat (**0.8828539**), ce qui montre que le final conditionne l'obtention de l'UV.

1.1.2 Corrélation

Dans cette seconde partie, nous allons étudier les liens statistiques entre variables. Nous allons donc afficher un ensemble de graphes en annexe et les commenter. Le graphe 2 montre que 80% des résultats de SY02 se situent entre 10 et 15. Nous avons également un pic à 7.5. Ce qui correspond à un ensemble de personnes ayant rencontrées les mêmes difficultés au sein de l'UV. Le graphe 3 démontre clairement que plus le niveau des étudiants est élevé, plus les résultats en SY02 diminue. De même, avec le graphe 4 nous constatons que les étudiants en TC assistant au cours de SY02, font parti des meilleurs résultats. En revanche les ISS font partis des plus mauvais. Sachant que ISS correspond à une fin d'étude et un TC à un début, cela souligne ce que nous avons énoncé par avant concernant l'influence du niveau sur les notes. Enfin le graphe 5 souligne que les correcteurs 3, 4, 6 et 7 déplacent les résultats de façons significatives.

1.2 Données crabs

1.2.1 Analyse

Dans cet exercice nous allons analyser un jeu de données contenant 200 individus décrits par 8 variables : trois variables qualitatives et cinq variables quantitatives, dont en particulier l'espèce qui sépare les données en deux populations de 100 individus qui sont eux même identifiés en tant que femelle ou male (50 par espèce). Le but de l'étude est de déterminer s'il existe des différences morphologiques permettant d'identifier l'espèce ou le sexe d'un individu en fonction d'un ou plusieurs paramètres morphologique. Dans un premier temps, afin de visualiser s'il existe des différences morphologiques majeures entre les espèces ou les deux sexes, nous avons représenté les données sous forme de boxplot (cf 6 pour le sexe et 7 pour l'espèce). Plus précisément, nous avons comparé tracé les boxplot de chaque paramètre en séparant soit selon le sexe soit l'espèce. Pour compléter l'étude, sur les boxplots, les intervalles de confiance à 95% ont aussi été représentés par un resserrement de la boite autour de la

médiane, les points où la boîte se resserre représentent les bornes de cet intervalle. Selon l'espèce, les intervalles de confiance des différents paramètres morphologiques ne se superposent pas ce qui indique bien qu'il y a des différences morphologiques significatives entre les deux espèces. En revanche, pour le sexe ce n'est le cas que pour une caractéristique : rear width, (largeur arrière). Cependant, pour toutes les caractéristiques, que ce soit selon le sexe ou l'espèce, les extrémités des moustaches se superposent, ce qui nous empêche avec cette représentation d'identifier avec certitude le sexe ou l'espèce d'un individu selon une ou plusieurs caractéristiques morphologiques.

1.2.2 Corrélation

Le calcul de la corrélation entre les différentes variables est présenté dans le tableau.

	FL	RW	CL	CW	BD
FL	1.00	0.91	0.98	0.96	0.99
RW	0.91	1.00	0.89	0.90	0.89
CL	0.98	0.89	1.00	1.00	0.98
CW	0.96	0.90	1.00	1.00	0.97
BD	0.99	0.89	0.98	0.97	1.00

Il est observé que toutes les variables sont très corrélées entre elles. En effet, les coefficients de corrélation sont tous supérieurs à 0.89. Ce résultat peut être expliqué par le fait que les variables ne sont pas indépendantes les unes des autres : la taille des membres des individus est proportionnelle à la leur taille. Pour s'affranchir de ce phénomène, il convient de diviser la taille de chaque membre de chaque individu par la somme de la taille de de ses membre. La matrice de corrélation obtenue après le traitement des données est représentée dans la matrice suivante.

	FL	RW	CL	CW	BD
FL	1.00	-0.12	-0.24	-0.78	0.52
RW	-0.12	1.00	-0.83	-0.20	-0.46
CL	-0.24	-0.83	1.00	0.42	0.10
CW	-0.78	-0.20	0.42	1.00	-0.65
BD	0.52	-0.46	0.10	-0.65	1.00

Les résultats obtenus indiquent que le traitement a permis de décorréler les variables entre elles. Les variables les plus corrélés sont désormais FL et CW (coefficient de corrélation de -0.78) et CL et RW (coefficient de corrélation de -0.83).
Les figures en annexe représentent les graphiques matriciels des données traitées, en les distinguant (a) selon le sexe et (b) selon l'espèce. Des nuages de points distincts sont observés ce qui nous permet maintenant de distinguer l'espèce ou le sexe d'un individu par ses caractéristiques morphologiques. Nous avons donc pour (a) les figures 8 et 9 et pour (b) les figures 10 et 11

1.3 Données Pima

1.3.1 Analyse

Dans le jeu de données suivant, nous avons des variables quantitatives discrètes (npreg, glu, bp, skin, age), quantitatives continues (ami, ped) et qualitative nominale (z). Nous utilisons la fonction summary pour obtenir une description de ces dernières dans le tableau suivant :

	npreg	glu	bp	skin	vmi	ped	age
Min	0.000	56.00	24.00	7.00	18.20	0.0850	21.00
1st Qu	1.000	98.75	64.00	22.00	27.88	0.2587	23.00
Median	2.000	115.00	71.51	29.18	32.89	0.4160	28.00
Mean	3.517	121.03	72.00	29.00	32.80	0.5030	31.61
3rd Qu	5.000	141.25	80.00	36.00	36.90	0.6585	38.00
Max	17.000	199.00	110.00	99.00	67.10	2.420	81.00

avec en plus $z_1 = 355 | z_2 = 177$
Au vu de ce tableau, il semble difficile de distinguer un pattern. Nous allons donc engendrer la matrice de corrélation afin de voir si un/des liens existent entre les variables. De là nous constatons que les couples suivant sont corrélés : (npreg,age) et (skin,bmi).

1.3.2 Liens statistiques

Au vu des résultats précédemment obtenues, nous allons essayer d'identifier des liens statistiques forts entre variables. On s'intéressera en particulier au facteur diabète comme le montre figure 12. 1 signifie diabète alors que 1 signifie sans diabète. On remarque que sur l'ensemble des 5 sous graphes, le taux plasmatique de glucose dans le sang est le plus influencé par le diabète. En effet, au-deçà de

100, les nombre de diabétique augmente alors
que le nombre de non diabétique diminue

2 Analyse en composantes principales

2.1 Exercice théorique

Cet exercice porte sur l'étude des correcteurs du jeu de données notes. Le nouveau jeu de donnée formé compte 8 individus : les correcteurs, et 4 variables : les moyennes et écart types des notes du médian et du final. Pour les correcteurs 2 et 8 certaines notes manquent, ainsi, dans une première partie l'étude se concentrera sur les 6 autres individus, puis dans une seconde partie une solution sera apportée pour pallier aux valeurs manquantes. L'ensemble du code se trouve dans la section 4 - code.

2.1.1 Axes factoriels

Le calcul des axes factoriels de l'ACP se fait en plusieurs étapes :
La première étape consiste à centrer la matrice. Ensuite, nous calculons la matrice d'inertie ou matrice de variances X_v grace a la formule suivante :

$$X_v = \frac{1}{6} X_{c^T} * X_c$$

$$X_v = \frac{1}{n} * X_{c^T} * X_c$$

Le calcul des valeurs propres :
$\lambda_1 = 0.98, \lambda_2 = 0.37, \lambda_3 = 0.08, \lambda_4 = 0.05$
de la matrice de variances permet d'obtenir leurs vecteurs propres associés ou axes principaux d'inertie. Il est ensuite possible de calculer le pourcentage d'inertie expliquée de chaque axe (dans l'ordre) : 66.10, 24.79, 5.61 et 3.51. Le pourcentage d'inertie expliquée par le premier plan factoriel (formé par les deux premiers axes) est 90.88.

2.1.2 Composantes principales

La matrice des composantes principales C se calcule à partir de la matrice centrée X_c et de la matrice des axes factoriels : $C = X_c.U$ Ces composantes principales permettent d'obtenir la représentation des six individus dans le premier plan factoriel. Comme le montre la figure 13 en annexe. La proximité des correcteurs 7 et 1 puis des correcteurs 5 et 8 sur cette représentation peut être interprétée comme le fait qu'ils ont un comportement similaire.

2.1.3 Représentation dans le premier plan factoriel

Tout comme cela a été fait pour les individus, il est possible de représenter les variables en fonction des individus et de les analyser. Il n'est pas nécessaire de refaire tous les calculs fait précédemment, les axes factoriels et ainsi les composantes principales pour les variables se déduisent des axes factoriels de l'analyse de nuage de points-individus. Pour cela, la corrélation entre les vecteurs variable et les composantes principales normées des individus est étudiée et permet le calcul de la matrice A partir de cette matrice, il est possible de représenter les quatre variables dans le premier plan factoriel 14. Les trois variables moy.median, std median et moy.final se trouvent sur le cercle de corrélation ce qui indique qu'elles sont bien représentées, ce qui n'est pas le cas de la variable std.final. Les variables moy.median et moy.final sont positionnées à angle droit ce qui indique qu'elles ne sont pas du tout corrélées.

2.1.4 Calculer l'expression $\sum_{\alpha=1}^{k} c_\alpha u_\alpha$

— $k = 1$

-0.11	0.32	-1.06	0.07
0.15	-0.44	1.43	-0.09
-0.04	0.12	-0.38	0.03
0.12	-0.34	1.10	-0.07
-0.03	0.07	-0.24	0.02
-0.09	0.26	-0.85	0.06

— $k = 2$

0.17	0.43	-1.05	0.11
0.91	-0.16	1.44	0.02
-0.26	0.04	-0.39	-0.01
-0.83	-0.69	1.09	-0.21
0.43	0.24	-0.23	0.08
-0.42	0.14	-0.86	0.01

— $k = 3$

0.13	0.47	-1.03	0.20
0.90	-0.15	1.44	0.03
-0.08	-0.24	-0.52	-0.51
-0.86	-0.63	1.11	-0.12
0.41	0.28	-0.21	0.15
-0.51	0.27	-0.79	0.25

— $k = 4$

0.22	0.16	-1.12	0.42
0.91	-0.20	1.43	0.06
-0.09	-0.21	-0.52	-0.53
-0.85	-0.70	1.09	-0.07
0.34	0.53	-0.14	-0.04
-0.54	0.41	-0.75	0.15

La matrice obtenue avec k = 4 correspond à la matrice d'origine Xc.

2.1.5 ACP et deux premiers plans factoriels

On souhaite représenter les individus initialement écartés de l'ACP. Remplacer chacune de leurs valeurs manquantes par la moyenne de la variable correspondante (imputation par la moyenne), puis représenter ces individus dans les deux premiers plans factoriels. Afin de pouvoir réaliser une ACP avec les 8 individus, il est possible de remplacer les valeurs manquantes par la moyenne de la variable correspondante. L'obtention de la matrice des composantes principales avec ces données se fait de la même manière que précédemment, avec les 6 individus. Après calculs on obtient une nouvelle matrice qui nous permet de représenter tous les individus dans les deux premiers plans factoriels 15 et 16

2.2 Utilisation des outils R

Grâce aux différentes fonctions proposées par R, nous allons tenter d'effectuer l'ACP du jeu de données de notes qui a été étudié en cours. Pour cela nous allons créer la matrice. Puis afin de pouvoir réaliser l'ACP, il est tout d'abord indispensable de centrer notre matrice. A partir de la matrice centrée en colonne nous allons calculer la matrice de variance, les axes principaux d'inertie et enfin les composantes principales.
La fonction summary et l'argument scores sur princomp permettent de retrouver les composantes principales. $sdev^2$ retourne les valeurs propres. Les vecteurs propres sont retournés grâce à loadings.

- La fonction biplot quant à elle affiche les individus et les variables dans le premier plan factoriel. La fonction biplot.princomp bénéficie de paramètres supplémentaires par rapport à la fonction biplot : "choices" pour choisir les composantes principales à représenter, "scale" pour obtenir une représentation standard des données.
- Princomp permet de calculer automatiquement les composantes principales d'une matrice passée en argument et retourne un objet R constitué de plusieurs variables membres :
- sdev : écarts types des composantes principales. ($sdev^2$ correspondant aux valeurs propres)
- loadings : matrice des vecteurs propres.
- scores : matrice des composantes principales.
- biplot.princomp offre la possibilité de représenter les variables dans un des plans factoriels

2.3 Données Crabs

Cette étude vise à utiliser l'ACP pour trouver une représentation des crabes qui permettent de distinguer visuellement différents groupes, liés à l'espèce et au sexe. Afin de compléter l'étude réalisée à l'exercice 1.2. sur le jeu de données CRABS, nous nous intéressons à l'utilisation de l'ACP pour visualiser les différents groupes.

2.3.1 ACP

L'ACP sur les données non traitées est réalisée sur R grâce à la fonction princomp préalablement introduite. La représentation biplot dans le premier plan factoriel et les pourcentages d'inertie expliquée pour chaque axe sont tracés dans la figure suivante. 17 et 18 Le biplot ne nous permet pas de distinguer de groupes, de plus les variables ont toutes la même direction. Ces observations indiquent que l'information n'est pas bien répartie sur les axes factoriels et que les variables sont très corrélées. Ce constat est confirmé par la valeur très importante du pourcentage d'inertie expliquée du premier axe factoriel et rejoint les observations réalisées lors de l'étude de l'exercice 1.2.

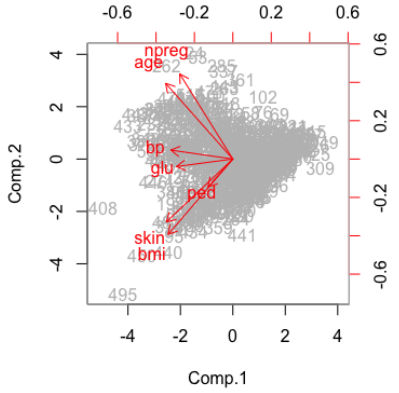
2.3.2 Amélioration

Afin de rendre visible la présence de groupes parmi le jeu de données CRABS avec l'ACP il est possible d'appliquer à crasbsquant le même traitement que lors de l'analyse dans l'exercice 2.1, c'est-à-dire diviser la taille de chaque membre par la somme de tous les membres d'un individu. Une nouvelle ACP est réalisée sur les données traitées. La représentation biplot et les pourcentages d'inertie expliquée pour chaque axe sont tracés dans les figures suivantes **??** et **??**.
Les pourcentages d'inertie expliquée par les deux premiers axes factoriels sont respectivement de 47.65 et 44.00, ce qui représente 91.65 d'inerte expliquée par le premier plan factoriel. On distingue 4 groupes d'individus sur le biplot représentant le premier plan factoriel.

2.4 Données Pima

Après plusieurs tentatives d'ACP sur les données, il est impossible de trouver une représentation simple qui permette de distinguer les deux catégories de patients. C'est ce que montre la figures ci dessous.

FIGURE 1 – Premier Plan factoriel

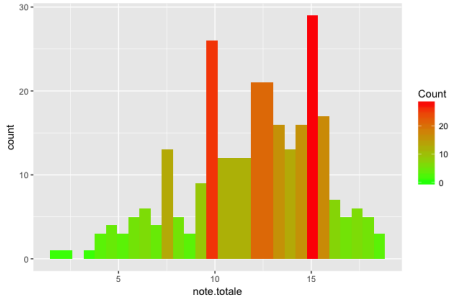


3 Annexes

3.1 Statistique descriptive

3.1.1 Notes

FIGURE 2 – densité des notes en SY02



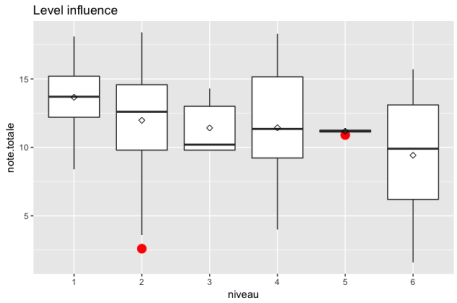


FIGURE 3 – influence du niveau sur les notes en SY02

FIGURE 4 – influence de la spécialité sur la note en SY02

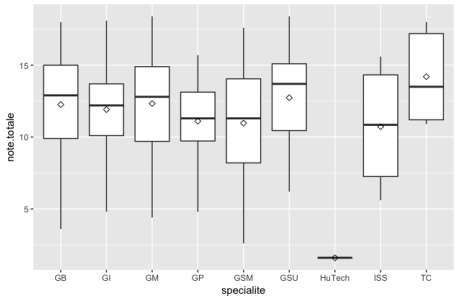
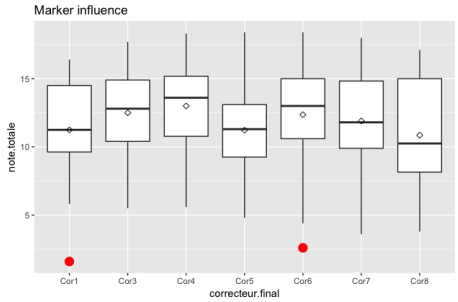


FIGURE 5 – influence du correcteur sur la note en SY02



3.1.2 Crabs

FIGURE 6 – Sexe en fonction des caractéristiques morphologiques

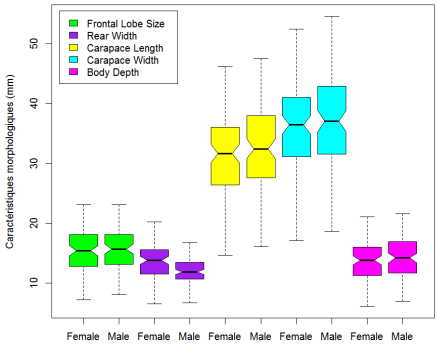


FIGURE 7 – Espèces en fonction des caractéristiques morphologiques

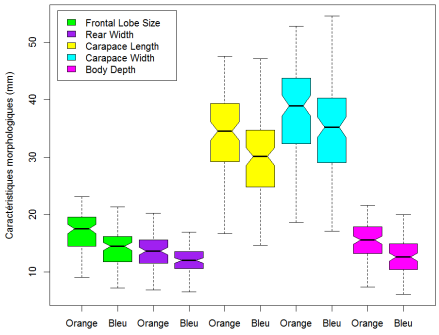


FIGURE 8 – Sexe en fonction des caractéristiques morphologiques

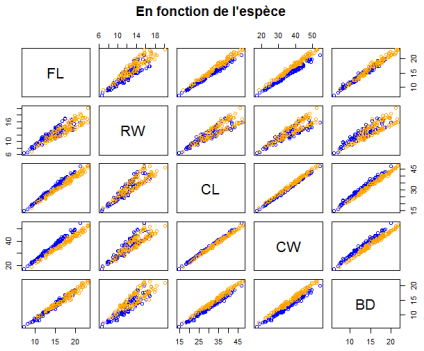


FIGURE 9 – Espèces en fonction des caractéristiques morphologiques

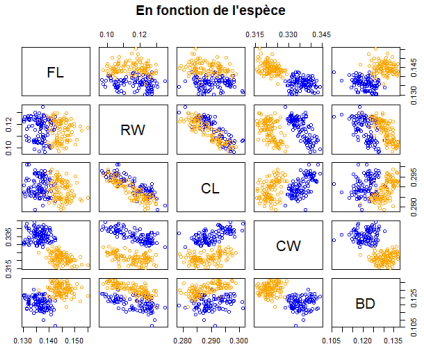


FIGURE 10 – Espèces en fonction des caractéristiques morphologiques

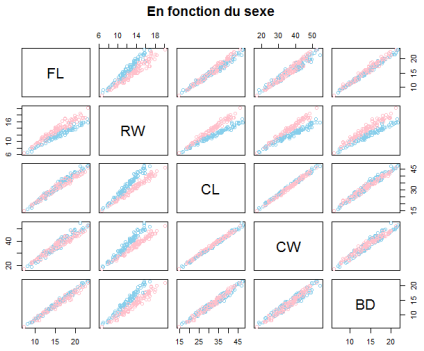
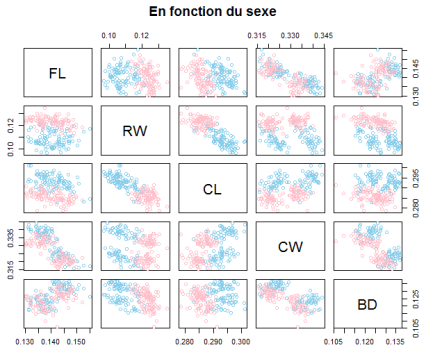
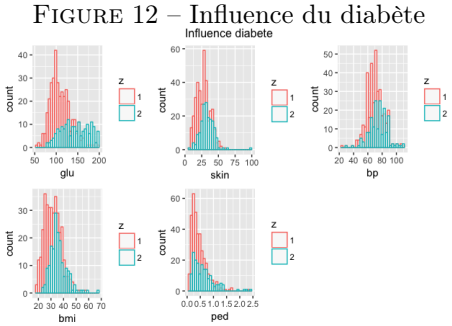


FIGURE 11 – Sexe en fonction des caractéristiques morphologiques



3.1.3 Pima

FIGURE 12 – Influence du diabète



3.2 Analyse en composantes principales

3.2.1 Exercice théorique

FIGURE 13 – $6_i ndividus_p remier_p lan_f actoriel$

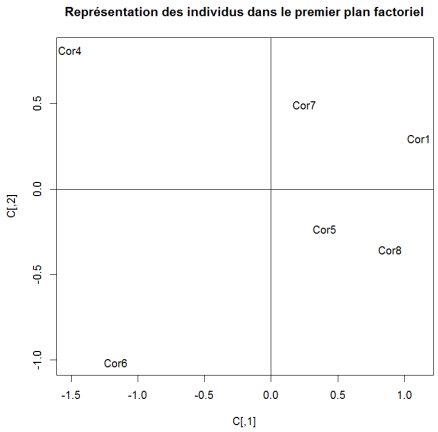


FIGURE 14 – $\text{variables}_p remier_p lan_f actoriel$

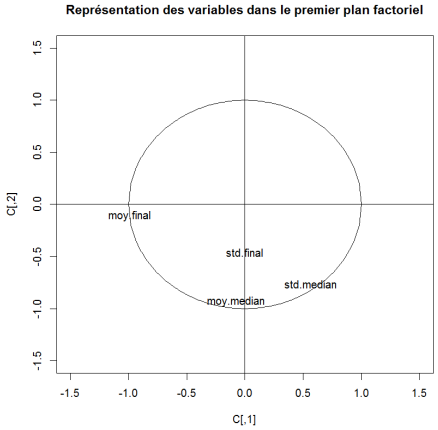


FIGURE 15 – $8_i ndividus_p remier_p lan_f actoriel$

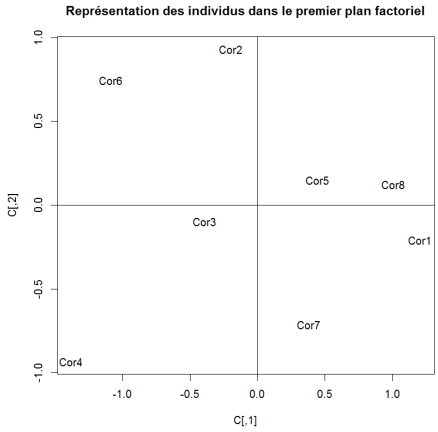
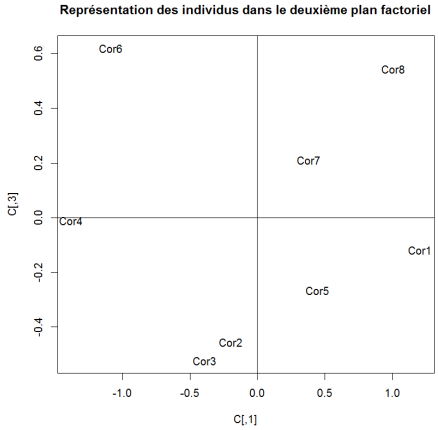


FIGURE 16 – $8_i ndividus_s econd_p lan_f actoriel$



3.2.2 ACP Crabs

FIGURE 17 – inertie$_a vant_t raitement$

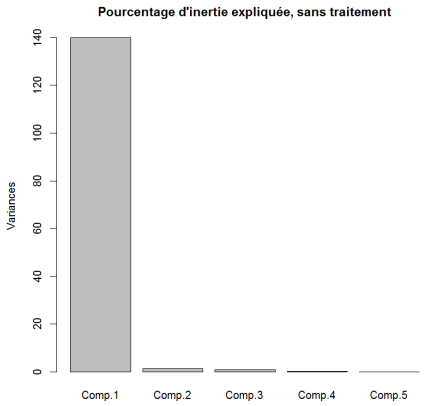


FIGURE 18 – acp$_a vant_t raitement$

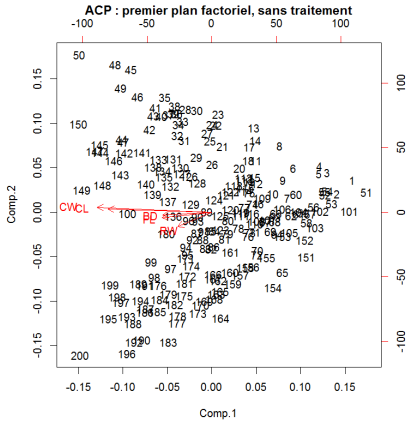


FIGURE 19 – inertie$_a pres_t raitement$

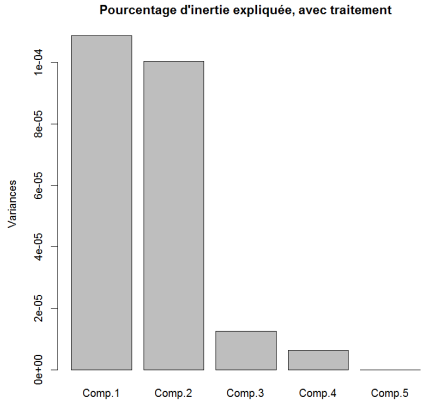


FIGURE 20 – acp$_a pres_t raitement$

4 code

```
notes <- read.csv("sy02-p2016.csv", na.strings="", header=T);
moy.median <- aggregate(note.median~correcteur.median, data=notes, FUN=mean)
names(moy.median) <- c("correcteur","moy.median")
std.median <- aggregate(note.median~correcteur.median, data=notes, FUN=sd)
names(std.median) <- c("correcteur","std.median")
median <- merge(moy.median, std.median)
moy.final <- aggregate(note.final~correcteur.final, data=notes, FUN=mean)
names(moy.final) <- c("correcteur","moy.final")
std.final <- aggregate(note.final~correcteur.final, data=notes, FUN=sd)
names(std.final) <- c("correcteur","std.final")
final <- merge(moy.final, std.final)
```

```
correcteurs <- merge(median, final, all=T)
corr.acp <- correcteurs[-c(2,8),]

#####ACP sur les individus#####
#convertir le data.frame en matrice pour faciliter les calculs
X<-as.matrix(corr.acp[2:5],nrow=6,rcol=4)
#centrage de la matrice
Xc<-t(t(X)-apply(X,2,mean))
rownames(Xc)<-c("Cor1","Cor3","Cor4","Cor5","Cor6","Cor7")
round(Xc,digits=2)
#matrice de variance
Xv<-(1/6)*(t(Xc)%*%Xc)
round(Xv,digits = 2)
#calul des valeurs propres et des vecteurs propres
(resume<-eigen(Xv))
#calcul des pourcentages d'inertie expliquée pour chacun des axes et inertie cumulée
bilan<-data.frame(c(1:4),resume$values,100*resume$values/sum(resume$values))
colnames(bilan)<-c("Axis","Eigen value","Proportion")
bilan$Cummulative<-cumsum(bilan[3])
colnames(bilan[,4])<-"Cummulative"
bilan
#stockage des vecteurs propres dans la matrice U
U<-resume$vectors
colnames(U)<-c("U1","U2","U3","U4")
round(U,digits=2)
#calcul de la matrice des composantes principales
C<-Xc%*%U
rownames(C)<-c("Cor1","Cor3","Cor4","Cor5","Cor6","Cor7")
round(C,digits=2)
#calculer la contribution relative des axes aux individus
cor<-C^2/diag(Xc%*%t(Xc))
print("contribution relative des axes aux individus")
round(cor,digits=2)
#calculer la contribution relative des individus aux axes
ctr<-1/6*C^2/matrix(resume$values,nrow=6,ncol=4,byrow=TRUE)
round(ctr,digits=2)

######représentation des variables######
#calcul des corrélations entre les variables intiales normées et les composantes
    principales normées
sigma<-diag(1/(sqrt(5/6*apply(Xc,2,sd)^2)))
vp<-diag(1/sqrt(resume$values))
Dp<-diag(1/6,nrow=6,ncol=6)
cor.var<-sigma%*%t(Xc)%*%Dp%*%C%*%vp
rownames(cor.var)<-c("moy.median","std.median","moy.final","std.final")
round(cor.var,digits=2)

######représentations graphiques######
#représentation des individus dans le premier plan factoriel
x11()
plot(C[,c(1,2)],xlab = "C[,1]",ylab="C[,2]",type="n",main="Représentation des individus
    dans le premier plan factoriel")
text(C[,c(1,2)],labels=corr.acp$correcteur)
abline(h=0,v=0)
#représentation des individus dans le deuxième plan factoriel
```

```
x11()
plot(C[,c(1,3)],xlab = "C[,1]",ylab="C[,3]",type="n",main="Représentation des individus dans le deuxième plan factoriel")
text(C[,c(1,3)],labels=corr.acp$correcteur)
abline(h=0,v=0)

#représentation des variables dans le premier plan factoriel
x11()
plot(-1.5:1.5,-1.5:1.5,xlab = "C[,1]",ylab="C[,2]",type="n", main="Représentation des variables dans le premier plan factoriel")
text(cor.var[,c(1,2)],colnames(correcteurs[-1]))
curve(sqrt(1-x^2),-1,1,add=TRUE)
curve(-sqrt(1-x^2),-1,1,add=TRUE)
abline(h=0,v=0)
#représentation des variables dans le deuxième plan factoriel
x11()
plot(-1:1,-1:1,xlab = "Axe 1",ylab="Axe 3",type="n",main="Représentation des variables dans le deuxième plan factoriel")
text(cor.var[,c(1,3)],colnames(correcteurs[-1]))
curve(sqrt(1-x^2),-1,1,add=TRUE)
curve(-sqrt(1-x^2),-1,1,add=TRUE)
abline(h=0,v=0)

######reconstitution######
round(C[,1]%*%t(resume$vectors[,1]),digits=2)
round(C[,1]%*%t(resume$vectors[,1])+C[,2]%*%t(resume$vectors[,2]),digits=2)
round(C[,1]%*%t(resume$vectors[,1])+C[,2]%*%t(resume$vectors[,2])+C[,3]%*%t(resume$vectors[,3]),digits=2)
round(C[,1]%*%t(resume$vectors[,1])+C[,2]%*%t(resume$vectors[,2])+C[,3]%*%t(resume$vectors[,3])+C[,4]%*%t(resume$vectors[,4]),digits=2)

######valeurs manquantes######
correcteurs$moy.median[8]<-mean(correcteurs$moy.median,na.rm = TRUE)
correcteurs$std.median[8]<-mean(correcteurs$std.median,na.rm = TRUE)
correcteurs$std.final[2]<-mean(correcteurs$std.final,na.rm = TRUE)
correcteurs$moy.final[2]<-mean(correcteurs$moy.final,na.rm = TRUE)
#acp sur le tableau avec les valeurs manquantes
#convertir le data.frame en matrice pour faciliter les calculs
Y<-as.matrix(correcteurs[2:5],nrow=8,rcol=4)
rownames(Y)<-c("Cor1","Cor2","Cor3","Cor4","Cor5","Cor6","Cor7","Cor8")
round(Y,digits=2)
#centrage de la matrice
Yc<-t(t(Y)-apply(Y,2,mean))
round(Yc,digits = 2)
#matrice de variance
Yv<-(1/8)*(t(Yc)%*%Yc)
round(Yv,digits = 2)
#calul des valeurs propres et des vecteurs propres
(resume<-eigen(Yv))
#recuperation de la matrice des vecteurs propres
UY<-resume$vectors
round(UY, digits=2)
#calcul de la matrice des composantes principales
CY<-Yc%*%UY
rownames(CY)<-c("Cor1","Cor2","Cor3","Cor4","Cor5","Cor6","Cor7","Cor8")
```

```
round(CY,digits = 2)

#####représentations graphiques#####
#représentation des individus dans le premier plan factoriel
x11()
plot(CY[,c(1,2)],xlab = "C[,1]",ylab="C[,2]",type="n", main="Représentation des individus dans le premier plan factoriel")
text(CY[,c(1,2)],labels=correcteurs$correcteur)
abline(h=0,v=0)
#représentation des individus dans le deuxième plan factoriel
x11()
plot(CY[,c(1,3)],xlab = "C[,1]",ylab="C[,3]",type="n", main="Représentation des individus dans le deuxième plan factoriel")
text(CY[,c(1,3)],labels=correcteurs$correcteur)
abline(h=0,v=0)
```