

Pràctica 3: Components prals sense usar cap funció específica. - continuació-

7. Escriu les equacions de les 4 components principals a partir dels 4 primers vectors propis:

$$\begin{aligned}Y_1 &= \dots \\Y_2 &= \dots \\Y_3 &= \dots \\Y_4 &= \dots\end{aligned}$$

8. Calcula les matrius rellevants en el “*món de les variables*”, pel fitxer bàsic.

- Per a cada component j , contribució de les variables (`var.contr`) en la component, expressat en %: Columnes de V^2 (multiplicades per 100).
- Correlació entre variables i components (`C`): $C = S_i V D = S_i V \Lambda^{1/2}$, on $S_i = \text{diag}(1/s_i)$ matriu diagonal amb els recíprocs de les desviacions típiques de les variables inicials. **Nota:** En el cas de dades tipificades, com el del nostre exemple, $S_i = I$ és la matriu identitat.
- Coordenades de les variables inicials en l'espai de les components (`var.coord`): VD . O també, $X^t U$, per dualitat.
- Cosinus quadrats (`var.cos2`): C^2 contribució de les components en les variables qualitats de representació acumulades en dim 1, 2, etc., (en proporció).
- Cosinus quadrats acumulats (`var.qual`): sumes acumulades de C^2 per files, són les qualitats de representació acumulades en dim 1, 1+2, etc., (en proporció))

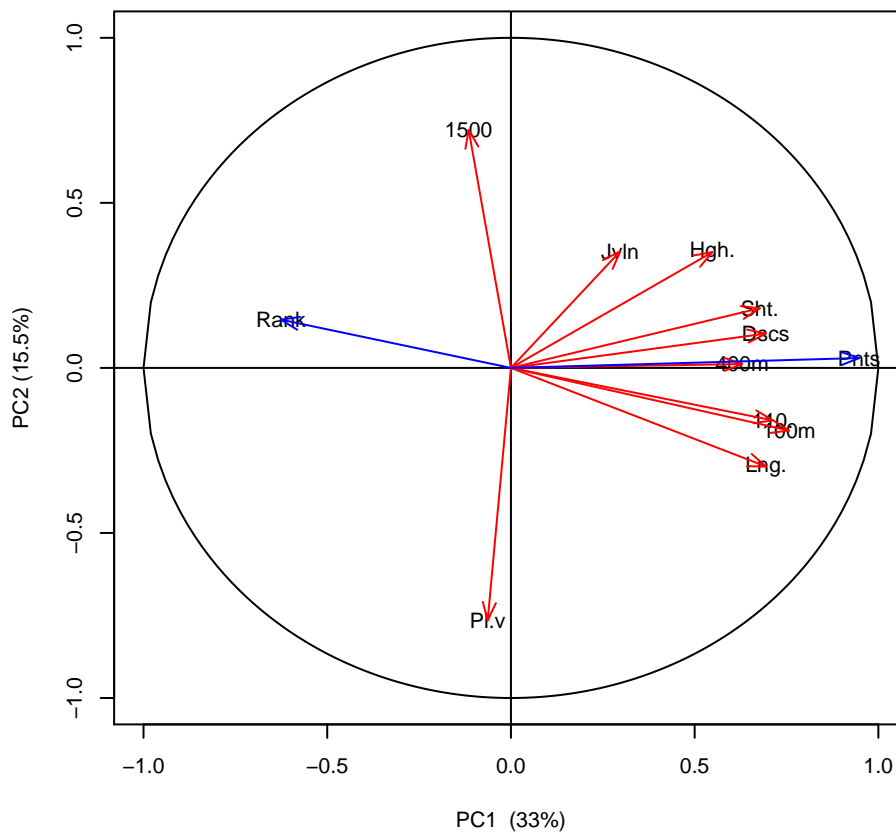
```
var.contr<- ..... # en %: # contrib de variables en comp
rownames(var.contr)<-colnames(basic)
colnames(var.contr)<-paste("PC",1:ncol(var.contr),sep="")
C<- ..... # correlacions
rownames(C)<- .....
colnames(C)<- .....
var.coord<- ..... # coordenades variables
rownames(var.coord)<- .....
colnames(var.coord)<- .....
var.cos2<- ..... # cosinus quadrats variables
var.qual<- ..... # qualitats variables (potser caldrà transposar?!)
```

9. Calcula les coordenades de les variables numèriques suplementàries (`nsup.coord`).

```
# coordenades variables numèriques suplementàries
nsup.coord<- .....
```

10. (Resolt) Representa les variables en l'espai de les dues primeres components amb fletxes, emmarcades en una circumferència. Afegeix les variables suplementàries, amb fletxes de color blau.

```
par(cex=.7)
plot(var.coord[,1:2],xlim=c(-1,1),ylim=c(-1,1),
     type="n",xlab="PC1 (33%)",ylab="PC2 (15.5%)") # type="n" -no punts-
abline(h=0); abline(v=0)
noms<-abbreviate(rownames(var.coord))
text(var.coord[,1:2],labels=noms)
arrows(rep(0,10),rep(0,10),var.coord[,1],var.coord[,2],col="red",length=0.1,angle=20)
curve(sqrt(1-x^2),-1,1,add=T)
curve(-sqrt(1-x^2),-1,1,add=T)
#v. numèriques supl
noms.nsup<-abbreviate(rownames(nsup.coord))
text(nsup.coord[,1:2],labels=noms.nsup)
arrows(rep(0,2),rep(0,2),nsup.coord[,1], nsup.coord[,2],col="blue",length=0.1,angle=20)
```

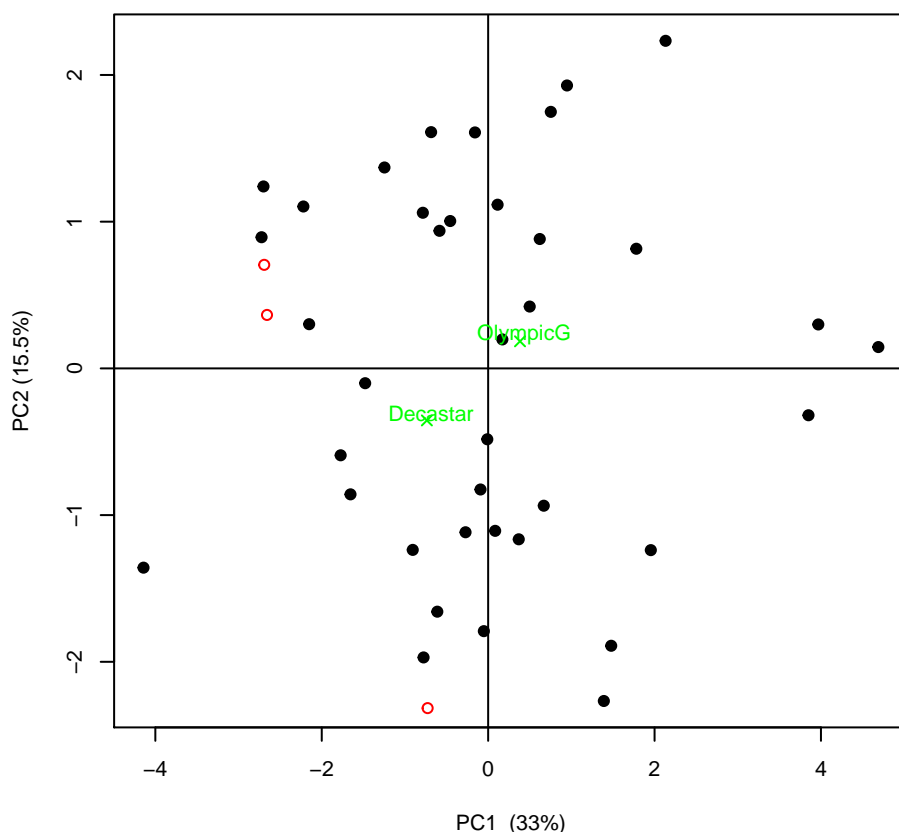


11. Calcula les matrius rellevants per al “món dels individus” del fitxer basic. **observacions:** (1) Per recuperar les coordenades projectades, multipliquem per \sqrt{n} , per comptes de $\sqrt{n-1}$, per obtenir els mateixos resultats que després obtindrem amb un paquet de R. (2) Si cal, recordeu aplicar conversions: `as.matrix(X)`, etc.

- Coordenades dels individus *scores* ($Y = XV$). *Remark:* Multiplica pel factor \sqrt{n} .
- Contribucions dels individus en les components (en %): $100 \times \frac{y_{ik}^2}{\sum_{j=1}^n y_{jk}^2}$. ((Correcció: aquest terme calcula el percentatge de la inèrcia de la component k , explicada per l'individu i , per tant la suma del denominador ha de ser la suma de la columna k de Y^2 .)
- Inèrcies dels individus: `ind.iner`: en dimensió p : $I_i = I_{ip} = \sum_{l=1}^p y_{il}^2$.
- Distàncies dels individus: `ind.dist`: $d_i = \sqrt{I_i}$.
- Cosinus quadrats dels individus `ind.cos2` (proporcions d'inèrcia): $\frac{y_{ik}^2}{I_i}$.
- Qualitats de representació `ind.qual` (proporcions d'inèrcia acumulades): $\frac{y_{i1}^2}{I_i}, \frac{y_{i1}^2 + y_{i2}^2}{I_i}, \dots, \frac{y_{i1}^2 + \dots + y_{ip}^2}{I_i} = 1$.

```
Y<- .....
colnames(Y)<-paste("PC",1:ncol(var.contr),sep="")
ind.contr<- .....
ind.iner<- .....
ind.cos2<- .....
ind.qual<- ..... # (potser caldrà transposar?!)
```

12. Representa els individus mitjançant un núvol de punts 2D de les seves coordenades en les dues primeres components.
13. Calcula `Yisup`: coordenades dels individus suplementaris. Posa'ls de color vermell en la gràfica anterior.
14. Calcula `Yqsup`: les coordenades dels dos nivells del factor suplementari. Posa'ls de color verd en la gràfica dels individus.



Components principals amb la llibreria FactoMineR

15. Compara els teus resultats amb els que s'obtenen aplicant la funció `PCA()` de la llibreria FactoMiner. Concretament, compara els valors propis i, només per a la primera component, les coordenades de: les variables bàsiques, dels individus bàsics, dels individus suplementaris i de la variable qualitativa. No cal que utilitzis `all.equal()`, només que les posis de costat.

Observacions: (1) Fixa't bé com s'entren els arguments de la funció: les variables s'entren pel número de la seva columna al fitxer complet `deca` (sense descomposar en basic, etc.), indicant-li quines són les columnes de les variables suplementàries i de la qualitativa suplementària (només n'hi pot haver una), el número de fila dels individus suplementaris (cas que hi hagi elements suplementaris). (2) Recorda que alguns vectors propis poden sortir canviats de signe.

```
library(FactoMineR); library(factoextra)
res<-PCA(...) # objecte que conté els resultats de la funció
names(res)    # noms dels objectes que té dins
names(res$eig)
names(res$var)
res$var$cos2
## etc
```

16. La funció `PCA()` té l'opció per defecte `graph=TRUE` que mostra els dos gràfics de variables i d'individus. L'usuari pot escollir altres opcions (`?plot.PCA`), per exemple, que d'entrada no es facin els gràfics i després dissenyar-los a la carta. Explora els arguments: `axes`, `choix`, `xlim`, `ylim`, `habillage`, `invisible`, `lim.cos2.var`, `select`, etc. Utilitza aquestes opcions per a:

- Explorar gràficament, el posicionament de les variables en el pla determinat per PC1-PC3, PC1-PC4, PC2-PC3 i PC3-PC4.

- En la gràfica dels individus en el pla PC1-PC2, representa'ls de color diferent segons si son del grup Decastar o OlympicG.
- En la gràfica de les variables en el pla PC1-PC2, fes que només apareguin aquelles que tenen una qualitat de representació superior a 0.45.
- En la gràfica dels individus en el pla PC1-PC2, fes que no surtin els individus suplementaris i que només apareguin els quatre que són més distants del centre (origen).

```
plot(res,choix="var",axis=c(1,2)) # plot aplicat al resultat, plot(res), és equivalent a plot.PCA(res)
#### etc
```

Vegeu altres prestacions gràfiques, amb paquets addicionals. Interpreteu els resultats.

```
# install.packages("corrplot")
library("corrplot")
corrplot(cor(basic), type="upper", order="hclust", tl.col="black", tl.srt=45)

# install.packages("PerformanceAnalytics")
library("PerformanceAnalytics")
chart.Correlation(basic, histogram=TRUE, pch=19)

# install.packages("factoextra")
library("factoextra")
fviz_pca_var(res)
fviz_pca_ind(res, col.ind="cos2") +
  scale_color_gradient2(low="white", mid="blue", high="red", midpoint=0.50)+theme_minimal()

fviz_pca_biplot(res, geom = "text") ## nota: les coordenades s'han dilatat !!
## Les distàncies entre individus i variables no són interpretables.
## Cal projectar els individus sobre les variables més ben projectades.

fviz_pca_ind(res, label="none", habillage=13)

fviz_pca_ind(res, label="none", habillage=13,
  addEllipses=TRUE, ellipse.level=0.95)
```

Més info a

<http://www.sthda.com/english/wiki/print.php?id=202>