Análisis de Datos Multivariantes
Práctica: Análisis de Componentes Principales (ACP)
Conjunto de datos: calificaciones

Contents

1 Datos. Conjunto de calificaciones de 6 materias

```r
library(openxlsx)
```

```
## Warning: package 'openxlsx' was built under R version 3.3.2
```

```r
Notas=read.xlsx("Ejemplo_notas.xlsx")
Notas
```

```
##    Etiqueta Mate Fis Qui Leng Geo His    Med.1    Med.2
## 1      Al 1    7  10   8    7   9   8 8.333333 8.000000
## 2      Al 2    3   4   5    6   6   4 4.000000 5.333333
## 3      Al 3    8   7  10    7   7   9 8.333333 7.666667
## 4      Al 4    4   6   4    6   7   6 4.666667 6.333333
## 5      Al 5    7   7   4    3   6   7 6.000000 5.333333
## 6      Al 6   10   9  10    9   8   7 9.666667 8.000000
## 7      Al 7    8   7   7    8   8   8 7.333333 8.000000
## 8      Al 8    9   7   9    9  10   8 8.333333 9.000000
## 9      Al 9    5   5   5    5   4   2 5.000000 3.666667
## 10    Al 10    5   6   6    6   5   6 5.666667 5.666667
## 11    Al 11    6   5   3    3   6   7 4.666667 5.333333
## 12    Al 12    3   6   3    6   3   7 4.000000 5.333333
## 13    Al 13   10   9  10    8   7   8 9.666667 7.666667
## 14    Al 14    3   3   4    1   2   5 3.333333 2.666667
## 15    Al 15    5   4   3    5   4   1 4.000000 3.333333
## 16    Al 16    8  10  10    8   7   9 9.333333 8.000000
## 17    Al 17    4   7   5    7   5   5 5.333333 5.666667
## 18    Al 18    4   5   3    4   5   5 4.000000 4.666667
## 19    Al 19    4   3   2    5   3   5 3.000000 4.333333
## 20    Al 20    7   7   4    7   5   3 6.000000 5.000000
## 21    Al 21    1   5   1    2   2   4 2.333333 2.666667
## 22    Al 22    4   3   2    2   4   2 3.000000 2.666667
## 23    Al 23    5   3   3    4   5   1 3.666667 3.333333
```

```r
X=Notas[,c(2,3,4,5,6,7)]
X[1:10,]
```

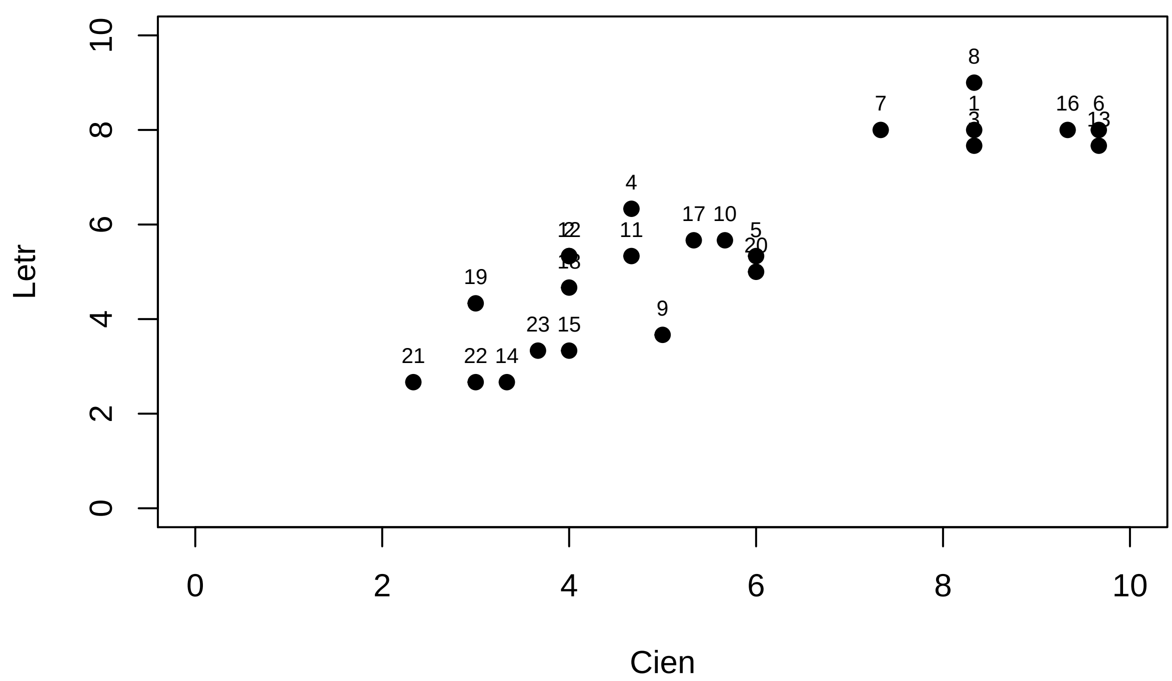
```
##    Mate Fis Qui Leng Geo His
## 1     7  10   8    7   9   8
## 2     3   4   5    6   6   4
## 3     8   7  10    7   7   9
## 4     4   6   4    6   7   6
## 5     7   7   4    3   6   7
## 6    10   9  10    9   8   7
## 7     8   7   7    8   8   8
## 8     9   7   9    9  10   8
## 9     5   5   5    5   4   2
## 10    5   6   6    6   5   6
```

```r
# Gráfico de las medias

Cien=(1/3)*(X[,1]+X[,2]+X[,3])
Letr=(1/3)*(X[,4]+X[,5]+X[,6])
Medias=cbind(Cien,Letr)

plot(Medias, xlim=c(0,10), ylim=c(0,10), pch=19, lty="solid", pin=c(2,5))
```

```
text(Medias, labels=c(1:23), cex=0.7, pos=3)
```



1.1 Matriz de distancias entre los datos originales

```
dist_ori=dist(X[1:10,], diag=T)
round(dist_ori,3)
```

```
##         1      2      3      4      5      6      7      8      9     10
## 1   0.000
## 2   9.327  0.000
## 3   4.359  9.274  0.000
## 4   7.071  3.317  7.937  0.000
## 5   7.141  6.633  7.616  4.583  0.000
## 6   4.472 11.000  4.123  9.592  9.434  0.000
## 7   3.606  7.874  3.464  5.916  6.325  4.359  0.000
## 8   4.359 10.100  4.000  8.544  9.055  3.317  3.162  0.000
## 9  10.149  3.742 10.000  5.385  6.481 11.091  8.832 11.136  0.000
## 10  6.708  3.742  6.325  3.000  4.472  8.307  5.292  8.000  4.472  0.000
```

2 Análisis de Componentes Principales

2.1 Matriz de varianzas-covarianzas $\widehat{\Sigma}$

```
Sigma=cov(X)
Sigma
```

```
##          Mate      Fis      Qui     Leng      Geo      His
## Mate 5.873518 3.772727 5.867589 3.887352 4.069170 3.371542
## Fis  3.772727 4.727273 5.000000 3.636364 3.227273 3.909091
## Qui  5.867589 5.000000 8.474308 5.254941 4.754941 4.994071
## Leng 3.887352 3.636364 5.254941 5.256917 3.575099 2.918972
## Geo  4.069170 3.227273 4.754941 3.575099 4.529644 3.237154
## His  3.371542 3.909091 4.994071 2.918972 3.237154 6.169960
```

2.1.1 Autovalores-autovectores de $\widehat{\Sigma}$

```
#Autovalores
lambdas=eigen(Sigma)$values
lambdas
```

```
## [1] 26.8885216  3.2253535  1.7496496  1.3493885  1.1054355  0.7132719
```

```
sum(lambdas)
```

```
## [1] 35.03162
```

```
#Autovectores
E=eigen(Sigma)$vectors
E
```

```
##            [,1]       [,2]        [,3]        [,4]       [,5]       [,6]
## [1,] -0.4136994 -0.3229029  0.60391376  0.08553725  0.3024214  0.5109789
## [2,] -0.3701437  0.1758847 -0.30610515  0.07917298  0.7968571 -0.3116221
## [3,] -0.5329247 -0.1010684  0.05174562  0.59896921 -0.4236522 -0.4060217
## [4,] -0.3753982 -0.3567364 -0.69766542 -0.16936565 -0.1788804  0.4294128
## [5,] -0.3562484 -0.1604058  0.22311308 -0.76299625 -0.1414297 -0.4420542
## [6,] -0.3741670  0.8376112  0.04892549 -0.12962601 -0.2051302  0.3116592
```

```
round(E%*%t(E),3)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    1    0    0    0    0    0
## [2,]    0    1    0    0    0    0
## [3,]    0    0    1    0    0    0
## [4,]    0    0    0    1    0    0
## [5,]    0    0    0    0    1    0
## [6,]    0    0    0    0    0    1
```

2.2 Puntuaciones de los alumnos respecto de las CP

```
# Puntuaciones de los alumnos respecto de las  CP
X=as.matrix(X)
CP_X=X%*%E
colnames(CP_X)=c("1CP","2CP","3CP","4CP","5CP","6CP")
round(CP_X[1:10,],3)
```

```
##        1CP    2CP    3CP    4CP    5CP    6CP
## 1  -19.688  1.450 -0.904 -2.907  2.530 -1.267
## 2  -11.273 -0.523 -1.806 -2.545 -0.766 -0.573
## 3  -19.719  1.556  0.324 -0.465 -0.327  0.563
## 4  -12.998  1.122 -1.545 -3.922  1.002 -0.098
## 5  -13.501  2.397  1.880 -2.445  3.179  0.589
## 6  -21.645 -1.287 -0.350 -0.978  1.782  0.755
## 7  -18.478  0.504 -0.354 -3.065  0.828  1.456
## 8  -21.045 -0.698  0.102 -3.476 -0.178  0.701
## 9  -10.634 -1.991 -0.750 -0.340  1.508 -0.031
## 10 -13.766  0.918 -1.283 -1.112  0.740  0.485
```

2.2.1 Media y varianza - covarianza de las CP

```
# Medias de las CP
mediaCP_X=rep(0,6)
for (i in 1:6) {mediaCP_X[i]=mean(CP_X[,i])}
round(mediaCP_X,3)
```

```
## [1] -13.501   0.446  -0.522  -1.795   1.346   0.533
```

```
# Varianzas - covarianzas de la sCP
round(cov(CP_X),3)
```

```
##        1CP   2CP  3CP   4CP   5CP   6CP
## 1CP 26.889 0.000 0.00 0.000 0.000 0.000
## 2CP  0.000 3.225 0.00 0.000 0.000 0.000
## 3CP  0.000 0.000 1.75 0.000 0.000 0.000
## 4CP  0.000 0.000 0.00 1.349 0.000 0.000
## 5CP  0.000 0.000 0.00 0.000 1.105 0.000
## 6CP  0.000 0.000 0.00 0.000 0.000 0.713
```

2.2.2 Puntuaciones centradas (media 0)

```
uno_seis=rep(1,23)
CP_X_centradas=CP_X-uno_seis%*%t(mediaCP_X)
round(CP_X_centradas[1:10,], 3)
```

```
##       1CP    2CP    3CP    4CP    5CP    6CP
## 1  -6.187  1.005 -0.382 -1.112  1.184 -1.800
## 2   2.228 -0.968 -1.284 -0.750 -2.112 -1.106
## 3  -6.218  1.110  0.846  1.330 -1.674  0.030
## 4   0.502  0.676 -1.023 -2.127 -0.344 -0.631
## 5  -0.001  1.952  2.402 -0.650  1.833  0.056
## 6  -8.145 -1.733  0.172  0.817  0.436  0.222
```

```
## 7  -4.977  0.059  0.168 -1.270 -0.518  0.924
## 8  -7.544 -1.144  0.624 -1.682 -1.525  0.168
## 9   2.866 -2.436 -0.228  1.455  0.161 -0.564
## 10 -0.265  0.472 -0.762  0.683 -0.606 -0.048
```

```r
# Medias de las CP centradas
mediaCP_X_centradas=rep(0,6)
for (i in 1:6) {mediaCP_X_centradas[i]=mean(CP_X_centradas[,i])}
round(mediaCP_X_centradas,3)
```

```
## [1] 0 0 0 0 0 0
```

3 Análisis de Componentes Principales con R (princomp(...))

```r
CP2=princomp(X)
CP2
```

```
## Call:
## princomp(x = X)
##
## Standard deviations:
##    Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6
## 5.0714352 1.7564512 1.2936684 1.1360984 1.0282865 0.8259904
##
##  6  variables and  23 observations.
```

```r
summary(CP2,loading=T)
```

```
## Importance of components:
##                           Comp.1     Comp.2     Comp.3     Comp.4
## Standard deviation     5.071435 1.75645117 1.29366836 1.13609836
## Proportion of Variance 0.767550 0.09206978 0.04994486 0.03851916
## Cumulative Proportion  0.767550 0.85961981 0.90956468 0.94808383
##                            Comp.5    Comp.6
## Standard deviation     1.02828648 0.8259904
## Proportion of Variance 0.03155536 0.0203608
## Cumulative Proportion  0.97963920 1.0000000
##
## Loadings:
##      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## Mate -0.414 -0.323  0.604         0.302 -0.511
## Fis  -0.370  0.176 -0.306         0.797  0.312
## Qui  -0.533 -0.101         0.599 -0.424  0.406
## Leng -0.375 -0.357 -0.698 -0.169 -0.179 -0.429
## Geo  -0.356 -0.160  0.223 -0.763 -0.141  0.442
## His  -0.374  0.838        -0.130 -0.205 -0.312
```

3.1 Puntuaciones de los alumnos respecto de las CP

```r
CP2_X=CP2$scores
round(CP2_X[1:10,],3)
```

```
##    Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## 1  -6.187  1.005 -0.382 -1.112  1.184  1.800
## 2   2.228 -0.968 -1.284 -0.750 -2.112  1.106
## 3  -6.218  1.110  0.846  1.330 -1.674 -0.030
## 4   0.502  0.676 -1.023 -2.127 -0.344  0.631
## 5  -0.001  1.952  2.402 -0.650  1.833 -0.056
## 6  -8.145 -1.733  0.172  0.817  0.436 -0.222
## 7  -4.977  0.059  0.168 -1.270 -0.518 -0.924
## 8  -7.544 -1.144  0.624 -1.682 -1.525 -0.168
## 9   2.866 -2.436 -0.228  1.455  0.161  0.564
## 10 -0.265  0.472 -0.762  0.683 -0.606  0.048
```

3.2 Distancias en el espacio de CP - espacio original

```
# En el de CP las distancias coinciden con las del espacio original
round(dist(CP2_X[c(1:10),],diag=T),3)
```

```
##         1      2      3      4      5      6      7      8      9     10
## 1   0.000
## 2   9.327  0.000
## 3   4.359  9.274  0.000
## 4   7.071  3.317  7.937  0.000
## 5   7.141  6.633  7.616  4.583  0.000
## 6   4.472 11.000  4.123  9.592  9.434  0.000
## 7   3.606  7.874  3.464  5.916  6.325  4.359  0.000
## 8   4.359 10.100  4.000  8.544  9.055  3.317  3.162  0.000
## 9  10.149  3.742 10.000  5.385  6.481 11.091  8.832 11.136  0.000
## 10  6.708  3.742  6.325  3.000  4.472  8.307  5.292  8.000  4.472  0.000
```

3.2.1 Usando sólo una CP

```
Y_r1=CP2_X[,c(1)]
round(Y_r1,3)
```

```
##      1      2      3      4      5      6      7      8      9     10
## -6.187  2.228 -6.218  0.502 -0.001 -8.145 -4.977 -7.544  2.866 -0.265
##     11     12     13     14     15     16     17     18     19     20
##  1.686  2.500 -7.787  6.059  4.677 -7.704  0.310  3.243  4.853  0.351
##     21     22     23
##  7.743  6.745  5.066
```

```
var(Y_r1)
```

```
## [1] 26.88852
```

```
# Matriz de distancias
round(dist(Y_r1[c(1:10)],diag=T),3)
```

```
##         1      2      3      4      5      6      7      8      9     10
## 1   0.000
## 2   8.415  0.000
## 3   0.031  8.446  0.000
## 4   6.690  1.726  6.720  0.000
## 5   6.187  2.229  6.217  0.503  0.000
## 6   1.957 10.372  1.926  8.647  8.144  0.000
## 7   1.211  7.205  1.241  5.479  4.976  3.168  0.000
## 8   1.357  9.772  1.326  8.047  7.544  0.600  2.567  0.000
## 9   9.054  0.639  9.085  2.364  2.867 11.011  7.843 10.411  0.000
## 10  5.923  2.493  5.953  0.767  0.264  7.880  4.712  7.279  3.131  0.000
```

3.2.2 Usando 2 CP

```
Y_r2=CP2_X[,c(1,2)]
round(Y_r2[c(1:10),],3)
```

```
##    Comp.1 Comp.2
## 1  -6.187  1.005
## 2   2.228 -0.968
## 3  -6.218  1.110
## 4   0.502  0.676
## 5  -0.001  1.952
## 6  -8.145 -1.733
## 7  -4.977  0.059
## 8  -7.544 -1.144
## 9   2.866 -2.436
## 10 -0.265  0.472
```

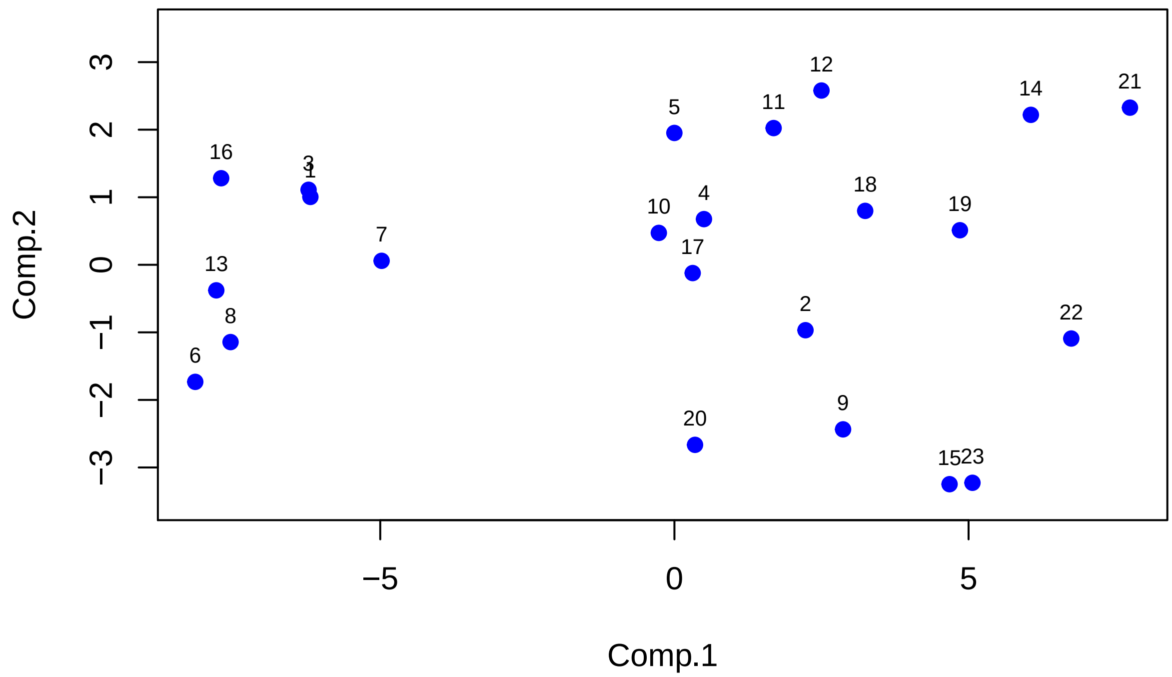
```
round(var(Y_r2),3)
```

```
##        Comp.1 Comp.2
## Comp.1 26.889  0.000
## Comp.2  0.000  3.225
```

```
# Matriz de distancias
round(dist(Y_r2[c(1:10),],diag=T),3)
```

```
##         1      2      3      4      5      6      7      8      9     10
## 1   0.000
## 2   8.643  0.000
## 3   0.110  8.698  0.000
## 4   6.698  2.384  6.734  0.000
## 5   6.259  3.673  6.274  1.371  0.000
## 6   3.365 10.401  3.434  8.976  8.939  0.000
## 7   1.536  7.278  1.627  5.514  5.324  3.639  0.000
## 8   2.541  9.774  2.615  8.250  8.154  0.841  2.835  0.000
## 9   9.686  1.601  9.752  3.909  5.242 11.034  8.231 10.491  0.000
## 10  5.946  2.879  5.987  0.794  1.503  8.182  4.730  7.457  4.273  0.000
```

```
# Representación gráfica
{plot(Y_r2, col= "blue",  , ylim=c(-3.5,3.5), pch = 19, cex = 1, lty = "solid", lwd = 1)
text(Y_r2, labels=c(1:23), cex= 0.7, pos=3)}
```



3.2.3 Usando 3, 4, 5 y 6 CP

```
Y_r3=CP2_X[,c(1:3)]
round(dist(Y_r3[c(1:10),],diag=T),3)
```

```
##          1      2      3      4      5      6      7      8      9     10
## 1    0.000
## 2    8.690  0.000
## 3    1.233  8.955  0.000
## 4    6.728  2.398  6.989  0.000
## 5    6.850  5.203  6.464  3.689  0.000
## 6    3.410 10.502  3.500  9.055  9.212  0.000
## 7    1.632  7.421  1.763  5.641  5.774  3.639  0.000
## 8    2.733  9.958  2.625  8.413  8.346  0.955  2.872  0.000
## 9    9.687  1.917  9.811  3.988  5.864 11.041  8.240 10.525  0.000
## 10   5.959  2.926  6.200  0.836  3.502  8.235  4.821  7.584  4.307  0.000
```

```r
Y_r4=CP2_X[,c(1:4)]
round(dist(Y_r4[c(1:10),],diag=T),3)
```

```
##         1      2      3      4      5      6      7      8      9     10
## 1   0.000
## 2   8.698  0.000
## 3   2.736  9.193  0.000
## 4   6.804  2.766  7.797  0.000
## 5   6.865  5.204  6.760  3.974  0.000
## 6   3.918 10.618  3.537  9.522  9.328  0.000
## 7   1.639  7.439  3.141  5.706  5.807  4.195  0.000
## 8   2.791 10.002  3.995  8.424  8.409  2.675  2.901  0.000
## 9  10.021  2.922  9.812  5.361  6.231 11.059  8.679 10.983  0.000
## 10  6.223  3.258  6.233  2.931  3.747  8.237  5.201  7.944  4.375  0.000
```

```r
Y_r5=CP2_X[,c(1:5)]
round(dist(Y_r5[c(1:10),],diag=T),3)
```

```
##         1      2      3      4      5      6      7      8      9     10
## 1   0.000
## 2   9.302  0.000
## 3   3.956  9.204  0.000
## 4   6.974  3.282  7.910  0.000
## 5   6.896  6.531  7.616  4.531  0.000
## 6   3.989 10.920  4.119  9.554  9.433  0.000
## 7   2.363  7.608  3.347  5.708  6.265  4.302  0.000
## 8   3.890 10.019  3.998  8.507  9.055  3.316  3.071  0.000
## 9  10.073  3.702  9.982  5.385  6.451 11.063  8.706 11.111  0.000
## 10  6.475  3.589  6.324  2.943  4.471  8.302  5.202  7.997  4.442  0.000
```

```r
Y_r6=CP2_X[,c(1:6)]
round(dist(Y_r6[c(1:10),],diag=T),3)
```

```
##         1      2      3      4      5      6      7      8      9     10
## 1   0.000
## 2   9.327  0.000
## 3   4.359  9.274  0.000
## 4   7.071  3.317  7.937  0.000
## 5   7.141  6.633  7.616  4.583  0.000
## 6   4.472 11.000  4.123  9.592  9.434  0.000
## 7   3.606  7.874  3.464  5.916  6.325  4.359  0.000
## 8   4.359 10.100  4.000  8.544  9.055  3.317  3.162  0.000
## 9  10.149  3.742 10.000  5.385  6.481 11.091  8.832 11.136  0.000
## 10  6.708  3.742  6.325  3.000  4.472  8.307  5.292  8.000  4.472  0.000
```