

# Análisis de componentes principales: fijando ideas

*00R Team*

*Marzo 2017*

## Contents

<b>1</b>	<b>La librería</b>	<b>1</b>
<b>2</b>	<b>Los datos</b>	<b>1</b>
<b>3</b>	<b>El análisis: PCA()</b>	<b>2</b>
<b>4</b>	<b>Los resultados</b>	<b>3</b>
4.1	La nueva matriz de datos . . . . .	3
4.2	Las distancias entre las observaciones se mantienen . . . . .	3
4.3	Varianza y autovalores de los ejes . . . . .	5
4.4	Variables originales y ejes . . . . .	5
4.5	Contribuciones a los ejes de los elementos . . . . .	6
<b>5</b>	<b>Información sobre el sistema</b>	<b>12</b>

## 1 La librería

```
library( FactoMineR )
```

## 2 Los datos

```
x <- read.table( "http://gauss.inf.um.es/datos/biom2003.dat" )[ 2:7 ]
head( x )
```

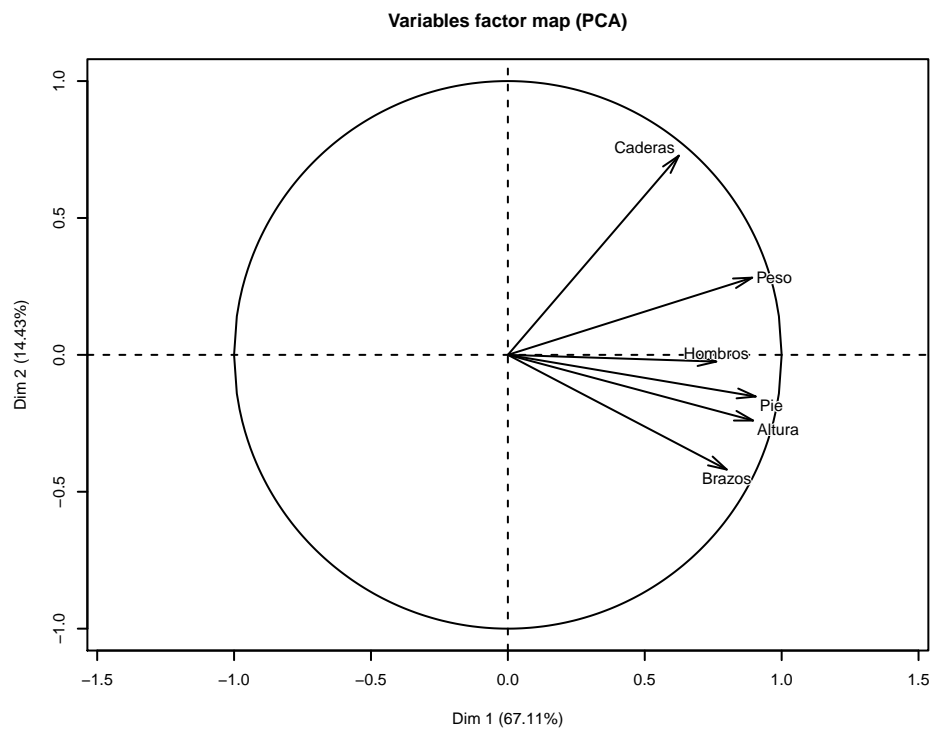
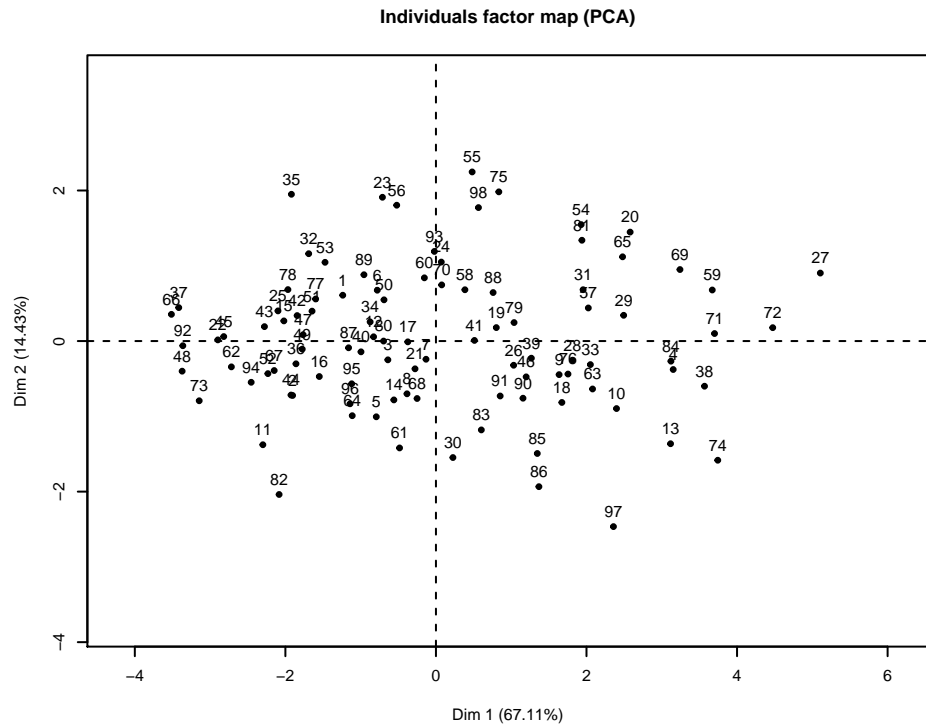
	Peso	Altura	Pie	Hombros	Brazos	Caderas
1	60	163	37	41	68	95
2	52	166	37	37	70	87
3	61	172	39	39	69	91
4	73	181	43	50	78	101
5	53	172	39	39	72	89
6	63	169	40	37	66	96

```
round( cor( x ), 3 )
```

	Peso	Altura	Pie	Hombros	Brazos	Caderas
Peso	1.000	0.707	0.765	0.646	0.567	0.689
Altura	0.707	1.000	0.854	0.546	0.787	0.420
Pie	0.765	0.854	1.000	0.626	0.685	0.429
Hombros	0.646	0.546	0.626	1.000	0.545	0.379
Brazos	0.567	0.787	0.685	0.545	1.000	0.279
Caderas	0.689	0.420	0.429	0.379	0.279	1.000

### 3 El análisis: PCA()

```
par( cex = 0.5 )  
xPca <- PCA( x, ncp = ncol( x ) )
```



## 4 Los resultados

### 4.1 La nueva matriz de datos

```
head( xPca$ind$coord )
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
1	-1.2377	0.6087	0.5473	0.14792	-0.04442	-0.232230
2	-1.9039	-0.7227	-0.2127	0.15628	0.04958	-0.243801
3	-0.6381	-0.2496	-0.1914	-0.49915	0.06858	-0.470101
4	3.1519	-0.3781	0.7519	0.41096	0.30340	-0.004719
5	-0.7922	-1.0060	-0.2827	0.07894	0.44569	-0.167220
6	-0.7792	0.6756	-0.5011	-1.02489	0.15276	-0.101473

```
round( cor( xPca$ind$coord), 3 )
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
Dim.1	1	0	0	0	0	0
Dim.2	0	1	0	0	0	0
Dim.3	0	0	1	0	0	0
Dim.4	0	0	0	1	0	0
Dim.5	0	0	0	0	1	0
Dim.6	0	0	0	0	0	1

### 4.2 Las distancias entre las observaciones se mantienen

#### 4.2.1 Distancias originales

```
round( dist( scale( x )[ 1:8, ], diag = TRUE ), 3 )
```

	1	2	3	4	5	6	7	8
1	0.000							
2	1.666	0.000						
3	1.452	1.511	0.000					
4	4.507	5.150	4.025	0.000				
5	1.925	1.214	1.077	4.123	0.000			
6	1.648	2.164	1.173	4.475	2.035	0.000		
7	1.785	2.157	1.794	3.374	1.505	2.502	0.000	
8	2.142	2.031	1.287	3.863	1.289	1.696	2.055	0.000

#### 4.2.2 Distancias considerando todos los ejes

```
round( dist( xPca$ind$coord[ 1:9, ], diag = TRUE ), 3 )
```

	1	2	3	4	5	6	7	8	9
1	0.000								
2	1.674	0.000							
3	1.459	1.519	0.000						
4	4.531	5.177	4.046	0.000					
5	1.935	1.221	1.083	4.145	0.000				
6	1.657	2.175	1.179	4.498	2.046	0.000			
7	1.794	2.168	1.803	3.391	1.513	2.515	0.000		

```
8 2.153 2.041 1.293 3.882 1.296 1.705 2.066 0.000
9 3.479 3.696 2.779 2.270 2.672 3.268 2.151 2.814 0.000
```

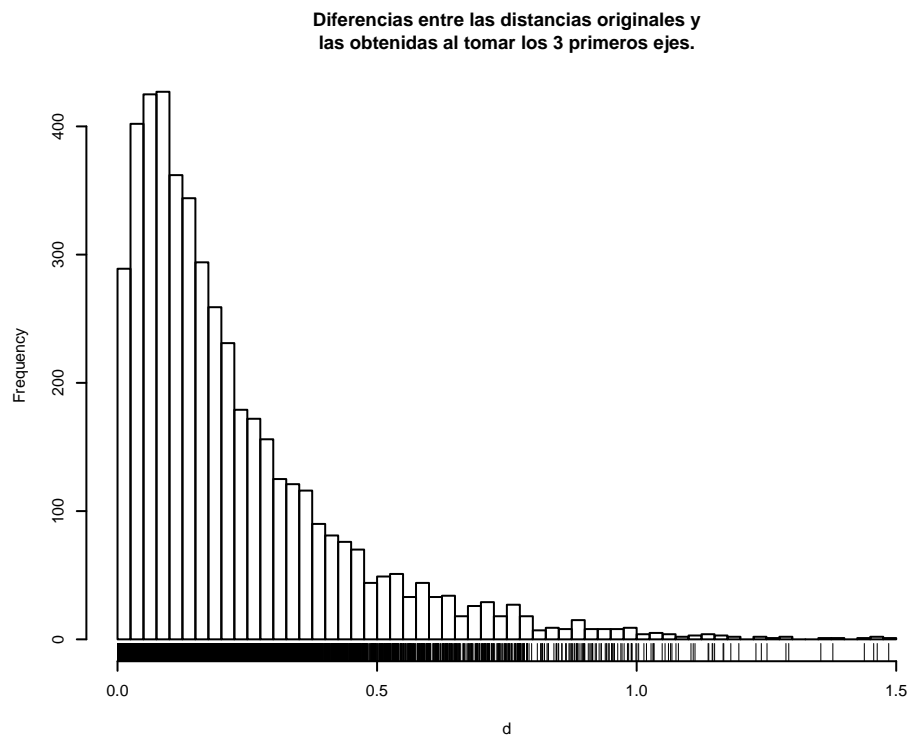
#### 4.2.3 Incluso al utilizar los primeros componentes se mantiene las distancias...

```
round( dist( xPca$ind$coord[ 1:8, 1:3 ], diag = TRUE ), 3 )
```

```
      1      2      3      4      5      6      7      8
1 0.000
2 1.672 0.000
3 1.281 1.351 0.000
4 4.504 5.159 3.908 0.000
5 1.869 1.149 0.777 4.126 0.000
6 1.146 1.817 0.986 4.258 1.696 0.000
7 1.420 1.902 0.694 3.320 1.158 1.369 0.000
8 1.661 1.533 0.547 3.632 0.576 1.512 0.602 0.000
```

#### 4.2.4 Las variaciones de distancia al reducir el número de ejes considerado

```
par( cex = 0.5 )
d <- ( dist( xPca$ind$coord[ , ], diag = TRUE ) -
      dist( xPca$ind$coord[ , 1:3 ], diag = TRUE ) )
hist( d, seq( 0, 1.5, 0.025 ),
      main = "Diferencias entre las distancias originales y\nlas obtenidas al tomar los 3 primeros ejes",
      rug( d ) )
```



## 4.3 Varianza y autovalores de los ejes

### 4.3.1 La varianza de las nuevas variables

```
apply( xPca$ind$coord, 2, var )
```

```
Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6  
4.0682 0.8746 0.5231 0.3149 0.1682 0.1129
```

### 4.3.2 Valores propios

```
xPca$eig$eigenvalue
```

```
[1] 4.0267 0.8656 0.5178 0.3117 0.1664 0.1117
```

### 4.3.3 Varianza acumulada

```
cumsum( xPca$eig$eigenvalue )
```

```
[1] 4.027 4.892 5.410 5.722 5.888 6.000
```

### 4.3.4 Varianza y varianza acumulada relativas

```
round( xPca$eig$eigenvalue / ncol( x ), 2 )
```

```
[1] 0.67 0.14 0.09 0.05 0.03 0.02
```

```
cumsum( round( xPca$eig$eigenvalue / ncol( x ), 2 ) )
```

```
[1] 0.67 0.81 0.90 0.95 0.98 1.00
```

## 4.4 Variables originales y ejes

### 4.4.1 La correlación de los ejes con las variables ...

```
cor( x, xPca$ind$coord )
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
Peso	0.8924	0.28198	0.01273	-0.12842	-0.32152	-0.06345
Altura	0.8957	-0.23988	-0.24580	-0.07897	0.14364	-0.23004
Pie	0.9051	-0.15211	-0.07150	-0.31891	0.08382	0.20913
Hombros	0.7612	-0.02456	0.63782	0.06704	0.08682	-0.03338
Brazos	0.7995	-0.41889	-0.14012	0.38684	-0.08962	0.08989
Caderas	0.6246	0.72758	-0.16006	0.18204	0.14082	0.04309

### 4.4.2 ... son coordenadas de los ejes

```
xPca$var$cor
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
Peso	0.8924	0.28198	0.01273	-0.12842	-0.32152	-0.06345
Altura	0.8957	-0.23988	-0.24580	-0.07897	0.14364	-0.23004
Pie	0.9051	-0.15211	-0.07150	-0.31891	0.08382	0.20913
Hombros	0.7612	-0.02456	0.63782	0.06704	0.08682	-0.03338
Brazos	0.7995	-0.41889	-0.14012	0.38684	-0.08962	0.08989
Caderas	0.6246	0.72758	-0.16006	0.18204	0.14082	0.04309

## 4.5 Contribuciones a los ejes de los elementos

### 4.5.1 Contribución relativa de las filas

```
head( xPca$ind$cos2 )
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
1	0.6720	0.16255	0.13138	0.009597	0.0008656	2.366e-02
2	0.8472	0.12207	0.01058	0.005708	0.0005745	1.389e-02
3	0.4151	0.06350	0.03734	0.253989	0.0047946	2.253e-01
4	0.9111	0.01311	0.05184	0.015489	0.0084420	2.042e-06
5	0.3214	0.51836	0.04093	0.003192	0.1017457	1.432e-02
6	0.2531	0.19028	0.10468	0.437883	0.0097281	4.292e-03

### 4.5.2 Descripción relativa de la información original: acumulada con los ejes

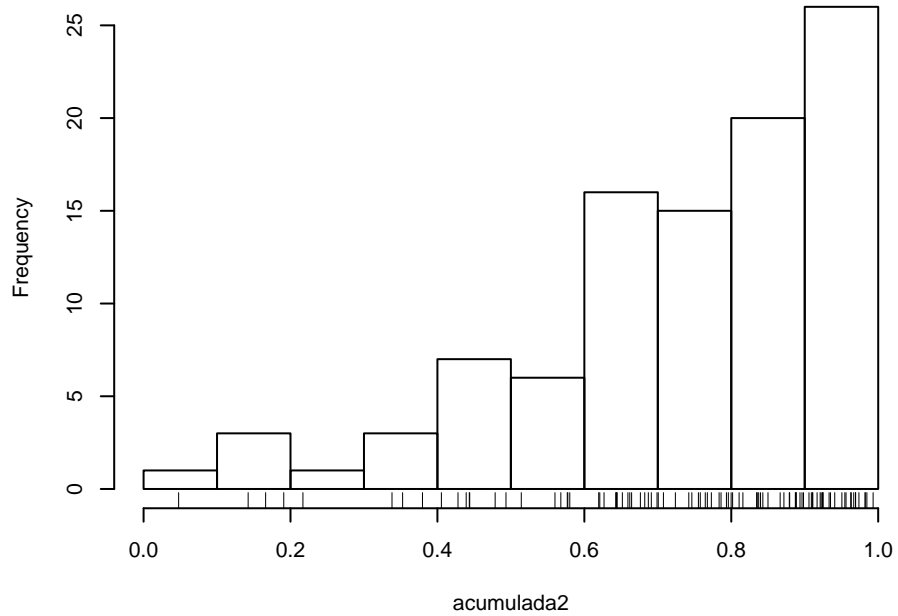
```
t( apply( head( xPca$ind$cos2 ), 1, cumsum ) )
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
1	0.6720	0.8345	0.9659	0.9755	0.9763	1
2	0.8472	0.9692	0.9798	0.9855	0.9861	1
3	0.4151	0.4786	0.5159	0.7699	0.7747	1
4	0.9111	0.9242	0.9761	0.9916	1.0000	1
5	0.3214	0.8398	0.8807	0.8839	0.9857	1
6	0.2531	0.4434	0.5481	0.9860	0.9957	1

### 4.5.3 Los individuos mejor descritos por los dos primeros ejes

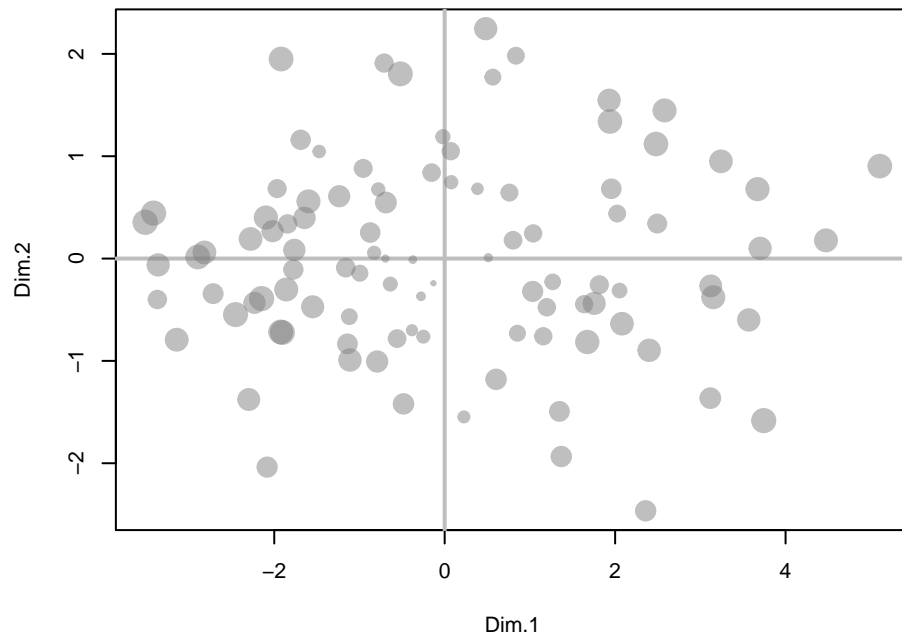
```
par( cex = 0.7 )
acumulada2 <- apply( xPca$ind$cos2[ , 1:2] , 1, sum )
hist( acumulada2, seq( 0, 1, 0.1 ),
      main = "Distribución de la contribución\nen los dos primeros ejes de los individuos" )
rug( acumulada2 )
```

### Distribución de la contribución en los dos primeros ejes de los individuos



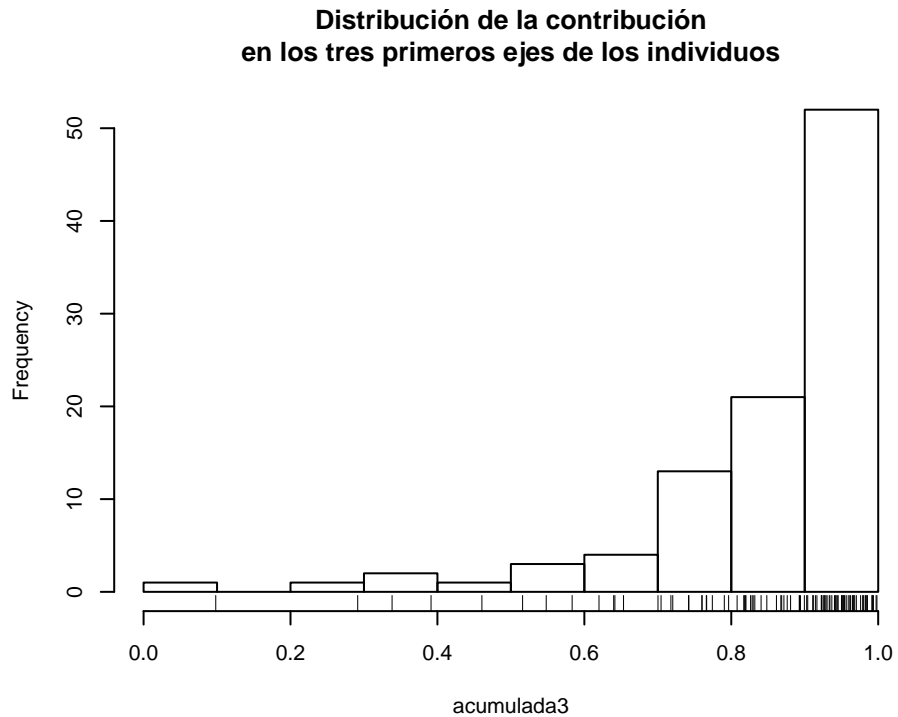
```
plot( xPca$ind$coord[ ,1:2 ], type = "n",
      main = "Plano de las dos primeras componentes\ntamaño del punto proporcional a la contribución en el plano",
      abline( h = 0, v = 0, col = "grey", lwd = 2 )
      points( xPca$ind$coord[ ,1:2 ], cex = 3 * acumulada2 + .5, pch = 20, col = rgb( .5, .5, .5, .5 ) ) )
```

### Plano de las dos primeras componentes tamaño del punto proporcional a la contribución en el plano



#### 4.5.4 Los individuos mejor descritos por los tres primeros ejes

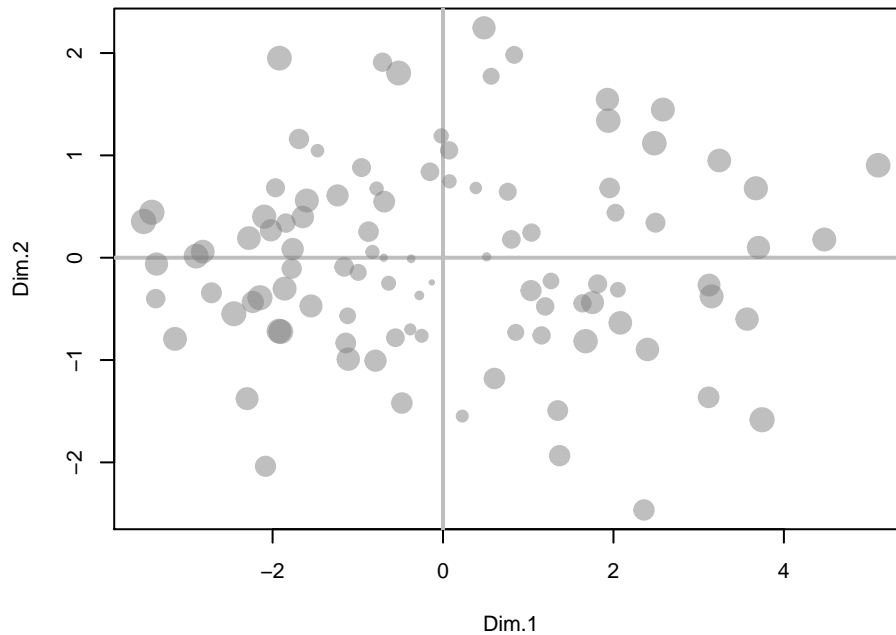
```
par( cex = 0.7 )
acumulada3 <- apply( xPca$ind$cos2[ , 1:3] , 1, sum )
hist( acumulada3, seq( 0, 1, 0.1 ),
      main = "Distribución de la contribución\nen los tres primeros ejes de los individuos" )
rug( acumulada3 )
```



```
plot( xPca$ind$coord[ ,1:2 ], type = "n",
      main = "Plano de las dos primeras componentes\ntamaño del punto proporcional a la contribución en" )
abline( h = 0, v = 0, col = "grey", lwd = 2 )
points( xPca$ind$coord[ ,1:2 ], cex = 3 * acumulada2 + .5, pch = 20, col = rgb( .5, .5, .5, .5 ) )
```

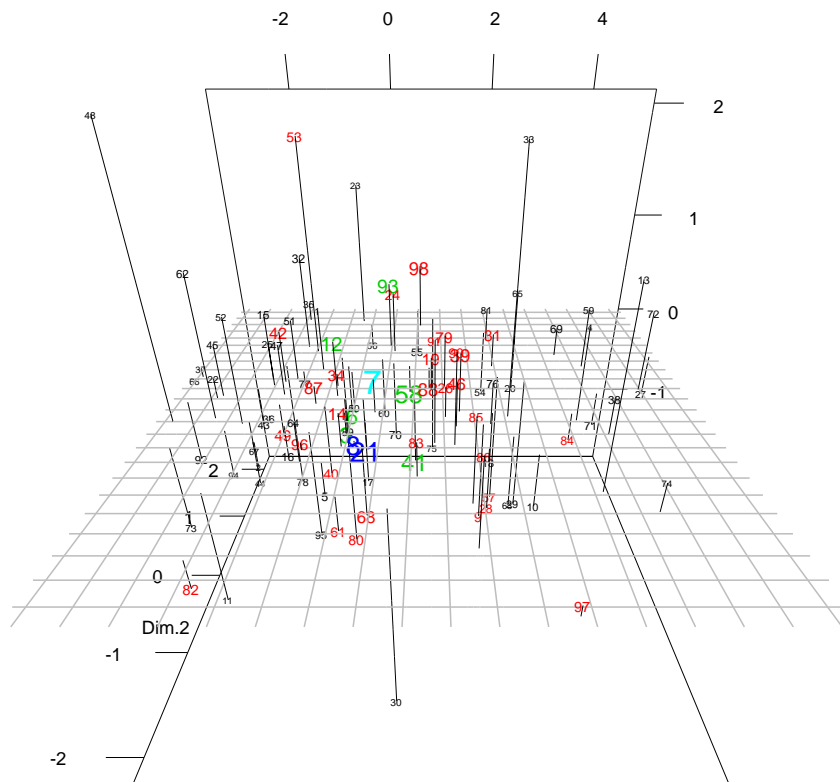


**Plano de las dos primeras componentes**  
 **tamaño del punto proporcional a la contribución en el plano de los tres primeros**



#### 4.5.5 Tres ejes: en tres dimensiones

```
par3d( cex = 2 )
view3d( theta = 0, phi = -60, fov = 160, zoom = 0.12, interactive = FALSE )
plot3d( xPca$ind$coord[ , 1:3 ], type = "h", aspect = 1, box = FALSE, axes = TRUE )
grid3d( side = "z", at = list( z = 0 ), n = 20, lwd = 2 )
text3d( xPca$ind$coord[ , 1:3 ], texts = rownames( xPca$ind$coord ),
        cex = ( ( 1 - acumulada3 ) * 3 + 1 ),
        col = round( ( ( 1 - acumulada3 ) * 2 + .5 ) * 2 ) )
```

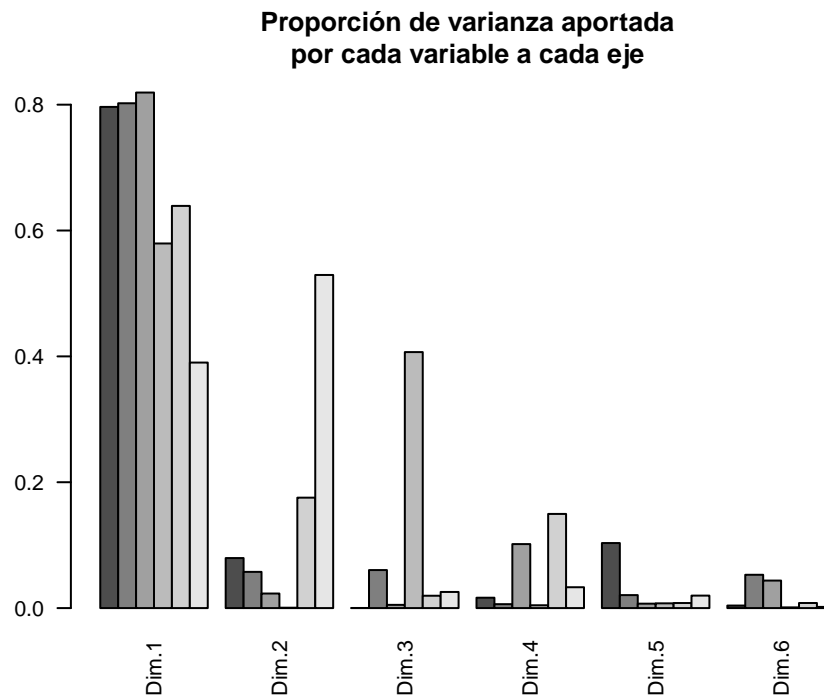


#### 4.5.6 Contribuciones relativas de las columnas

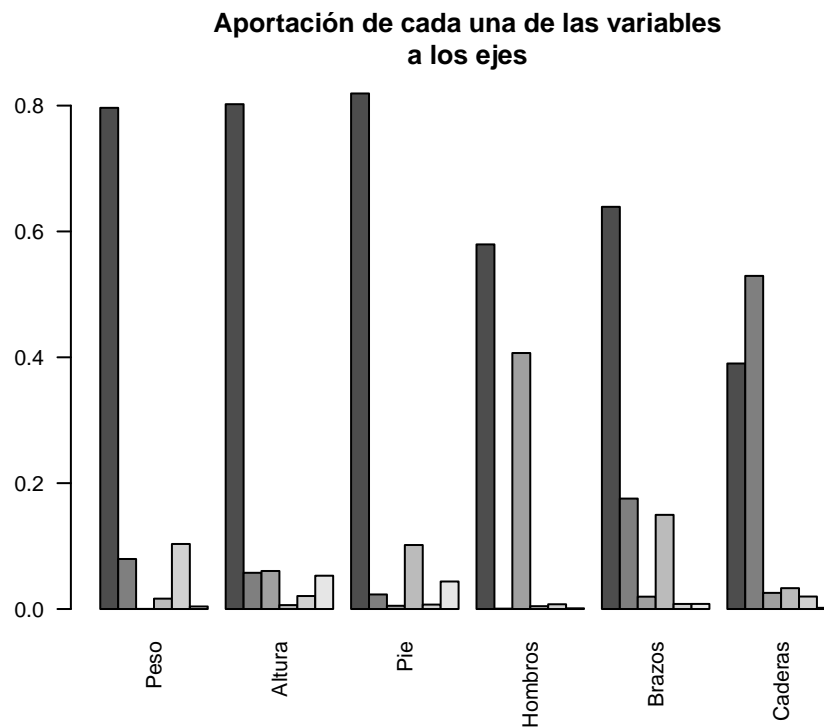
```
par( cex = 0.7 )
round ( xPca$var$cos2, 3 )
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
Peso	0.796	0.080	0.000	0.016	0.103	0.004
Altura	0.802	0.058	0.060	0.006	0.021	0.053
Pie	0.819	0.023	0.005	0.102	0.007	0.044
Hombros	0.579	0.001	0.407	0.004	0.008	0.001
Brazos	0.639	0.175	0.020	0.150	0.008	0.008
Caderas	0.390	0.529	0.026	0.033	0.020	0.002

```
barplot( xPca$var$cos2, beside = TRUE, las = 2,
main = "Proporción de varianza aportada\npor cada variable a cada eje" )
```



```
par( cex = 0.7 )
barplot( t( xPca$var$cos2 ), beside = TRUE, las = 2,
  main = "Aportación de cada una de las variables\na los ejes" )
```



## 5 Información sobre el sistema

```
sessionInfo()
```

```
R version 3.3.0 (2016-05-03)
```

```
Platform: x86_64-pc-linux-gnu (64-bit)
```

```
Running under: Ubuntu 16.04.2 LTS
```

```
locale:
```

```
[1] LC_CTYPE=es_ES.UTF-8      LC_NUMERIC=C
[3] LC_TIME=es_ES.UTF-8      LC_COLLATE=es_ES.UTF-8
[5] LC_MONETARY=es_ES.UTF-8  LC_MESSAGES=es_ES.UTF-8
[7] LC_PAPER=es_ES.UTF-8     LC_NAME=C
[9] LC_ADDRESS=C             LC_TELEPHONE=C
[11] LC_MEASUREMENT=es_ES.UTF-8 LC_IDENTIFICATION=C
```

```
attached base packages:
```

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

```
other attached packages:
```

```
[1] FactoMineR_1.35 rgl_0.98.1      knitr_1.15.1
```

```
loaded via a namespace (and not attached):
```

```
[1] Rcpp_0.12.10      cluster_2.0.6      magrittr_1.5
[4] leaps_3.0         MASS_7.3-45        scatterplot3d_0.3-38
[7] xtable_1.8-2      lattice_0.20-35    R6_2.2.0
[10] stringr_1.2.0     tools_3.3.0        grid_3.3.0
[13] htmltools_0.3.5   yaml_2.1.14        rprojroot_1.2
[16] digest_0.6.12     shiny_1.0.0        htmlwidgets_0.8
[19] evaluate_0.10     mime_0.5            rmarkdown_1.4
[22] stringi_1.1.3     backports_1.0.5    flashClust_1.01-2
[25] jsonlite_1.3      httpuv_1.3.3
```