Análisis Estadístico de Datos Multivariantes (AEDM)

Juana Mari Vivo

10/11/2023

Introducción

Clasificación de Técnicas Estadísticas Multivariantes

- **Técnicas estadísticas multivariantes de ajuste de modelos estadísticos**.
- **Técnicas estadísticas multivariantes de reducción de datos**.
- **Técnicas estadísticas multivariantes de clasificación**.

Notación y conceptos usuales

Matriz de datos

Sean n individuos observados (filas) sobre m variables $X_1, X_2, ..., X_m$ (columnas): la *matriz de datos* X de tamaño $n \times m$,

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1m} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2m} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3m} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nm} \end{pmatrix}$$

cada x_{ij} es el valor del individuo (fila) i para la variable (columna) j.

Notación y conceptos usuales

Matriz de Varianzas-Covarianzas

Se trata de una matriz simétrica.

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2m} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \dots & \sigma_{3m} \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_{m1} & \sigma_{m2} & \sigma_{m3} & \dots & \sigma_m^2 \end{pmatrix} \in M(m)$$

La varianza total de X es la suma de varianzas, es decir, la suma de los m elementos de la diagonal de la matriz de covarianzas.

Notación y conceptos usuales

Matriz de Correlaciones

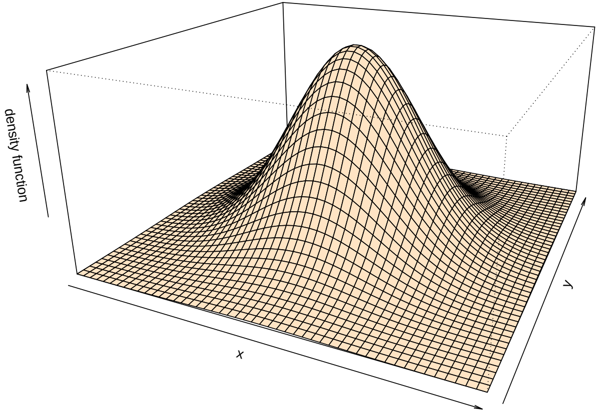
Si las variables originales son estandarizadas, i.e. $Z_i = \frac{X_i - \mu}{\sigma_1}$, la matriz correspondiente se denomina matriz de correlaciones.

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1m} \\ r_{21} & 1 & r_{23} & \dots & r_{2m} \\ r_{31} & r_{32} & 1 & \dots & r_{3m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & r_{m3} & \dots & 1 \end{pmatrix} \in M(m)$$

$r_{ij} \in [-1, 1]$: la relación entre cada par de variables (inversa o directa) y su grado (más fuerte cuanto más próximo a -1 y a 1, pero más débil cuanto más próximo a 0). Nótese que la varianza total es igual a m, i.e., igual al número de variables.

Notación y conceptos usuales

Distribución normal bivariante



Análisis de Componentes Principales

Introducción

El análisis de componentes principales (ACP) es una técnica estadística de análisis multivariante, introducida a finales del siglo XIX por Pearson y estudiada por Hotelling en la primera mitad del siglo XX.

- Adecuación: Se aplica cuando se dispone de un número elevado de variables cuantitativas intercorrelacionadas.
- Elimnación de redundancia en la información: Consiste en sustituir las m variables originales por k combinaciones lineales de las mismas no directamente observables, denominadas *componentes principales* incorreladas entre sí y pueden ordenarse de acuerdo a su varianza asociada (información de la componente).
- Reducción de dimensiones: k debe ser menor que m, pero expresando una proporción razonable de la variación total (*inercia de la nube de puntos*) cuantificada como la traza de la matriz de varianzas-covarianzas, $tr(\Sigma)$.

Detrás del ACP

- En relación a los individuos ¿Hay similitudes entre los individuos para todas las variables? ¿Podemos establecer diferentes perfiles de individuos? ¿Podemos oponer a un grupo de individuos a otro?
- En relación a las variables ¿Podemos resumir las características por un pequeño número de variables?
- En relación a ambos ¿Podemos caracterizar grupos de individuos por variables?

Obtención de las componentes principales

Si se considera n individuos observados sobre m variables originales $X_1, X_2, ..., X_m$ e Y_1, $Y_2, ...,$ Y_m las componentes principales:

$$Y_j = a_{j1}X_1 + a_{j2}X_2 + ... + a_{jm}X_m = a_j'X$$

con $a_j' = (a_{j1}, a_{j2}, ..., a_{jm})$ es un vector con componentes constantes tal que $a_j'a_j = 1$ (condición de ortogonalidad), $j = 1, 2, ...m$.

- La primera componente se calcula tomando a_1 tal que Y_1 tenga la mayor varianza posible, $Var(Y_1) = Var(a_1'X) = a_1'\Sigma a_1$ sujeta a la restricción de que $a_1'a_1 = 1$, a través del método de los multiplicadores de Lagrange: $L(a_1) = a_1'\Sigma a_1 - \lambda(a_1'a_1 - 1)$

$$\frac{\partial L}{\partial a_1}(a_1) = 2\Sigma a_1 - 2\lambda I a_1 \Rightarrow (\Sigma - \lambda I)a_1 = 0$$

Obtención de las componentes principales

Para que el sistema $(\Sigma - \lambda I)a_1 = 0$ tenga solución distinta de cero (Th. Rouché-Frobenius) $\Rightarrow |\Sigma - \lambda I| = 0$

$$Var(Y_1) = Var(a_1'X) = a_1'\Sigma a_1 = a_1'\lambda a_1 = \lambda a_1' a_1 = \lambda$$

Por tanto, para maximizar la varianza de Y_1 se tiene que tomar el mayor autovalor, λ_1 y el correspondiente autovector a_1.

- La segunda componente principal se calcula obteniendo a_2 de modo que la variable obtenida, Y_2 esté incorrelada con Y_1, i.e, $Cov(Y_1, Y_2) = Cov(a_1'X_1, a_2'X_2) = a_2'\Sigma a_1 = \lambda a_2' a_1 = 0$ y razonando de manera similar, elegimos λ_2 como el segundo mayor autovalor de la matriz Σ con su autovector asociado a_2, a través de la función:

$$L(a_2) = a_2'\Sigma a_2 - \lambda(a_2'a_2 - 1) - \delta a_2' a_1$$

Obtención de las componentes principales

- Análogamente, se obtiene $Y_3, ..., Y_m$, en orden decreciente de varianza.

Por tanto, las m componentes Y se pueden expresar como el producto de una matriz formada por los autovectores A, multiplicada por el vector X que contiene las variables originales $X_1, ..., X_m$, i.e., $Y = AX$. La matriz de varianzas-covarianzas viene dada por:

$$\Lambda = diag(\lambda_1, ...\lambda_m) = Var(Y) = A'Var(X)A = A'\Sigma A$$

Porcentajes de variablilidad

Dado que $\lambda_i = Var(Y_i)$ correspondiente al autovector a_i, entonces la varianza total de las componentes es la suma de todos los autovalores:

$$\sum_{i=1}^{m} Var(Y_i) = \sum_{i=1}^{m} \lambda_i = tr(\Lambda)$$

$$tr(\Lambda) = tr(A'\Sigma A) = tr(\Sigma A'A) = tr(\Sigma)$$

Por tanto, la suma de las varianzas de las variables originales y la suma de las varianzas de las componentes son iguales.

Porcentajes de variablilidad

- El porcentaje de varianza total que retiene una componente principal:

$$\frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i} = \frac{\lambda_i}{\sum_{i=1}^{m} Var(X_i)}$$

- El porcentaje de variabilidad retenido por las primeras k componentes:

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{m} \lambda_i}$$

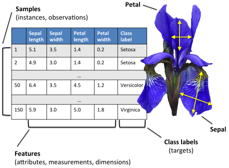
con $k < m$.

Caso práctico

Anderson recopiló estos datos para cuantificar la variación morfológica de las flores de Iris de tres especies relacionadas.

Iris Versicolor | Iris Setosa | Iris Virginica

A partir de las cuatro características medidas por Fisher (longitudes y amplitud de pétalo y sépalo, en cm)



	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

aplicaremos un ACP.

Caso práctico

```
> data(iris) #cargando datos
> names(iris)
```

```
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
```

```
> names(iris)[c(1,2,3,4)] <- c("SLength","SWidth","PLength","PWidth")
> names(iris)
```

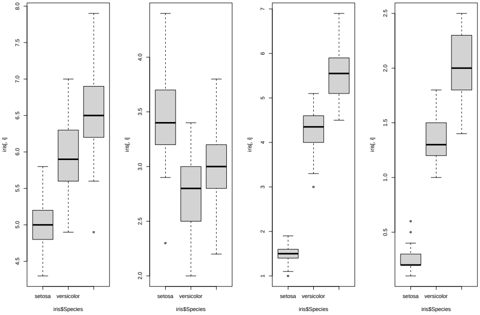
```
[1] "SLength" "SWidth"  "PLength" "PWidth"  "Species"
```

Caso práctico

```
>
> tapply(iris[,1],iris$Species,summary)
$setosa
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  4.300   4.800   5.000   5.006   5.200   5.800 

$versicolor
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  4.900   5.600   5.900   5.936   6.300   7.000 

$virginica
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  4.900   6.225   6.500   6.588   6.900   7.900 
```

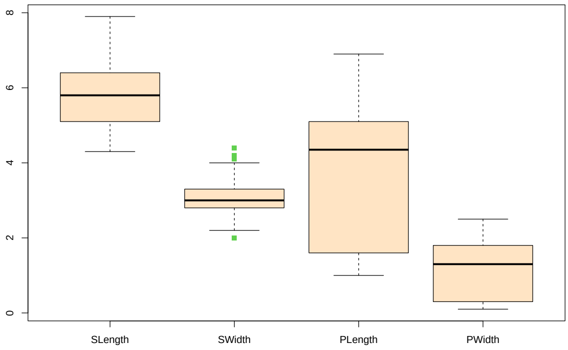
Caso práctico

```
> par(mfrow=c(1,4))
> for (i in 1:4) boxplot(iris[,i]~iris$Species)
```



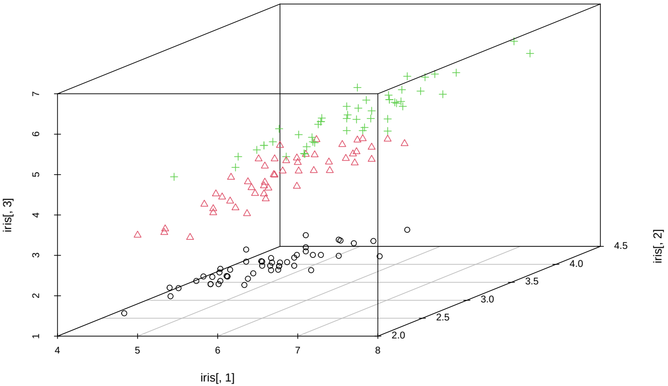
Caso práctico

```
> par(mfrow=c(1,1))
> boxplot(iris[,1:4],col="bisque",outcol=3,outpch = 15)
```



Caso práctico

```r
> library(scatterplot3d)
> scatterplot3d(iris[,1], iris[,2],iris[,3], color=as.numeric(iris[,5]),pch=as.numeric(iris[,5]))
```

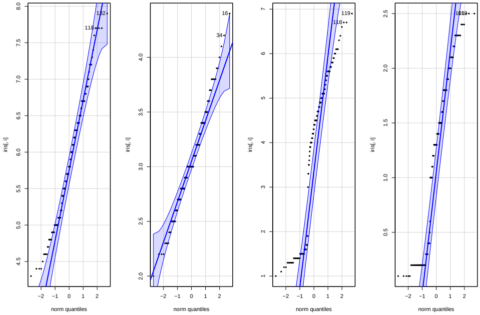


Caso práctico

```
> require(car)
```

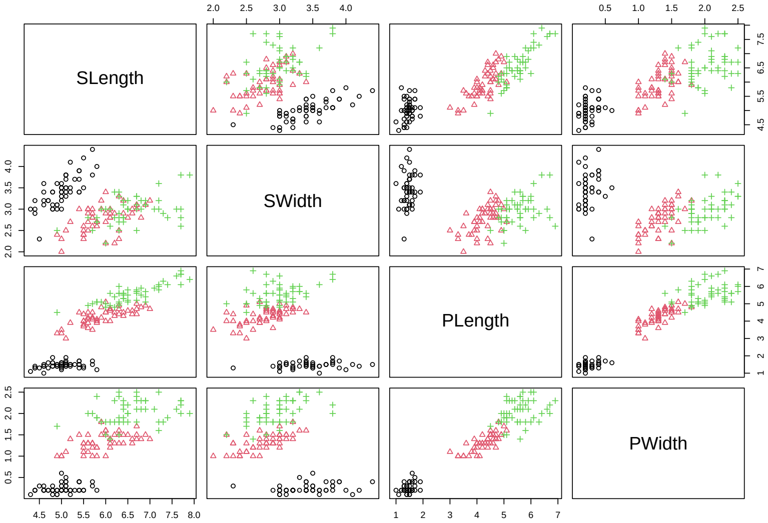
```
Loading required package: car

Loading required package: carData
```

```
> par(mfrow=c(1,4))
> for (i in 1:4){qqPlot(iris[,i],pch=20)}
```



Caso práctico

```
> pairs(iris[,1:4],pch=as.numeric(iris[,5]),col=iris[,5])
```



Caso práctico

```
> round(cov(iris[,1:4]),2)
```

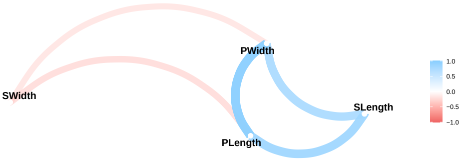
```
        SLength SWidth PLength PWidth
SLength    0.69  -0.04    1.27   0.52
SWidth    -0.04   0.19   -0.33  -0.12
PLength    1.27  -0.33    3.12   1.30
PWidth     0.52  -0.12    1.30   0.58
```

```
> round(cor(iris[,1:4]),2)
```

```
        SLength SWidth PLength PWidth
SLength    1.00  -0.12    0.87   0.82
SWidth    -0.12   1.00   -0.43  -0.37
PLength    0.87  -0.43    1.00   0.96
PWidth     0.82  -0.37    0.96   1.00
```

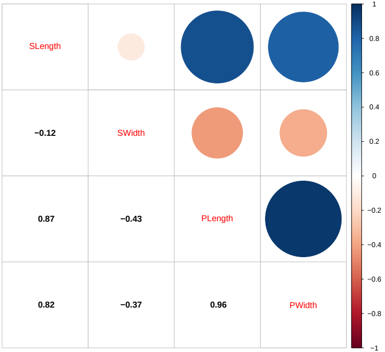
Caso práctico

```
> library(corrr)
> cor <- correlate(iris[,1:4])
> network_plot(cor)
```



Caso práctico

```
> library(corrplot)
> corrplot.mixed(cor(iris[,1:4]),lower.col =1)
```



https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html

Caso práctico: Idoneidad del ACP

- Test de Esfericidad de Bartlett

$$\begin{cases} H_0 & : & |R| = 1 \\ H_1 & : & |R| \neq 1 \end{cases}$$

$$T = -(n - 1 - \frac{2m + 5}{6}) ln|R| \sim \chi^2_{m(m-1)/2}$$

Caso práctico

```
> library("psych")
> cortest.bartlett(cor(iris[,1:4]), n = 150)
```

```
$chisq
[1] 706.9592

$p.value
[1] 1.92268e-149

$df
[1] 6
```

Caso práctico

```
> library("REdaS")
> bart_spher(iris[,1:4])
```

```
    Bartlett's Test of Sphericity

Call: bart_spher(x = iris[, 1:4])

     X2 = 706.959
     df = 6
p-value < 2.22e-16
```

Caso práctico: Idoneidad del ACP

- Medida de la adecuación muestral de Kaiser-Meyer-Olkin (KMO)

$$KMO = \frac{\sum_i \sum_{j \neq i} r_{ij}^2}{\sum_i \sum_{j \neq i} r_{ij}^2 + \sum_i \sum_{j \neq i} a_{ij}^2}$$

KMO Value	Degree of Common Variance
0.90 to 1.00	Marvelous
0.80 to 0.89	Meritorious
0.70 to 0.79	Middling
0.60 to 0.69	Mediocre
0.50 to 0.59	Miserable
0.00 to 0.49	Don't Factor or unacceptable

Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin Test of Sampling Adequacy (KMO) are commonly used to provide more complex measures for assessing the strength of the relationships and suggesting factorability of the variables (Beavers et al., 2013). Kaiser (1974) recommends that the accepted index of KMO & Bartlett's Test of Sphericity should be over 0.5. Also, the Bartlett's Test of Sphericity relates to the significance of the study and thereby

Caso práctico

```
> KMO(iris[,1:4])
```

```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = iris[, 1:4])
Overall MSA =  0.54
MSA for each item =
SLength  SWidth PLength  PWidth
   0.58    0.27    0.53    0.63
```

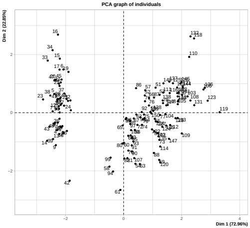
Caso práctico

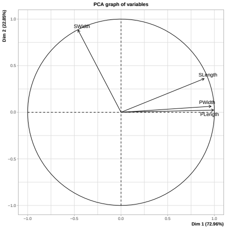
```
> KMOS(iris[,1:4])
```

```
Kaiser-Meyer-Olkin Statistics

Call: KMOS(x = iris[, 1:4])

Measures of Sampling Adequacy (MSA):
  SLength    SWidth   PLength    PWidth
0.5840603 0.2695746 0.5307484 0.6342065

KMO-Criterion: 0.5400767
```

Caso práctico

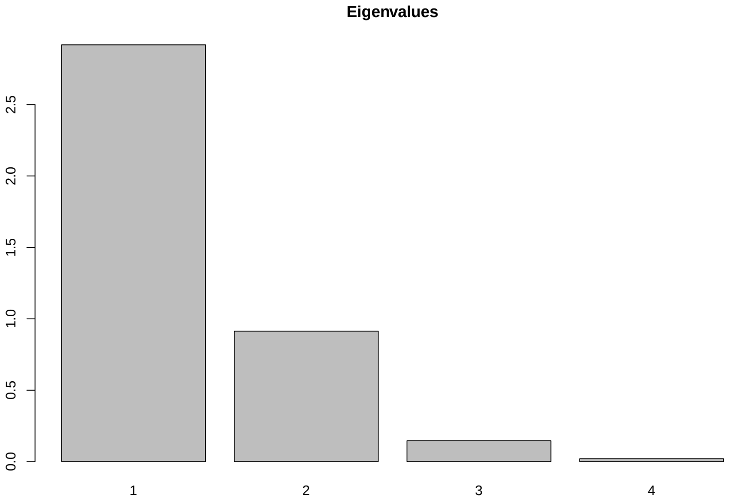
```
> library(FactoMineR)
> iris.pca=PCA(iris[,1:4], scale.unit=T, ncp=4, graph=T)
```





Caso práctico

```
> barplot(iris.pca$eig[,1],main="Eigenvalues",names.arg=1:nrow(iris.pca$eig))
```

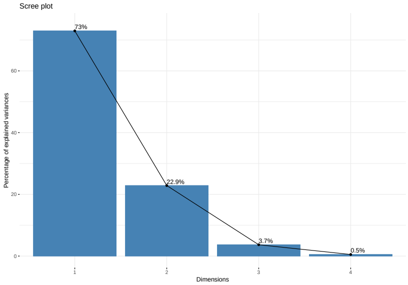


Caso práctico

```
> options(width = 300)
> summary.PCA(iris.pca)


Call:
PCA(X = iris[, 1:4], scale.unit = T, ncp = 4, graph = T)


Eigenvalues
                       Dim.1   Dim.2   Dim.3   Dim.4
Variance               2.918   0.914   0.147   0.021
% of var.             72.962  22.851   3.669   0.518
Cumulative % of var.  72.962  95.813  99.482 100.000

Individuals (the 10 first)
            Dist    Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr   cos2  
1        |  2.319 | -2.265  1.172  0.954 |  0.480  0.168  0.043 | -0.128  0.074  0.003 |
2        |  2.202 | -2.081  0.989  0.893 | -0.674  0.331  0.094 | -0.235  0.250  0.011 |
3        |  2.389 | -2.364  1.277  0.979 | -0.342  0.085  0.020 |  0.044  0.009  0.000 |
4        |  2.378 | -2.299  1.208  0.935 | -0.597  0.260  0.063 |  0.091  0.038  0.001 |
5        |  2.476 | -2.390  1.305  0.932 |  0.647  0.305  0.068 |  0.016  0.001  0.000 |
6        |  2.555 | -2.076  0.984  0.660 |  1.489  1.617  0.340 |  0.027  0.003  0.000 |
7        |  2.468 | -2.444  1.364  0.981 |  0.048  0.002  0.000 |  0.335  0.511  0.018 |
8        |  2.246 | -2.233  1.139  0.988 |  0.223  0.036  0.010 | -0.089  0.036  0.002 |
9        |  2.592 | -2.335  1.245  0.812 | -1.115  0.907  0.185 |  0.145  0.096  0.003 |
10       |  2.249 | -2.184  1.090  0.943 | -0.469  0.160  0.043 | -0.254  0.293  0.013 |

Variables
            Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr   cos2  
SLength |  0.890 27.151  0.792 |  0.361 14.244  0.130 | -0.276 51.778  0.076 |
SWidth  | -0.460  7.255  0.212 |  0.883 85.247  0.779 |  0.094  5.972  0.009 |
PLength |  0.992 33.688  0.983 |  0.023  0.060  0.001 |  0.054  2.020  0.003 |
PWidth  |  0.965 31.906  0.931 |  0.064  0.448  0.004 |  0.243 40.230  0.059 |
```

Caso práctico: Scree plot

```
> library(factoextra)
> fviz_screeplot(iris.pca, addlabels = TRUE, ylim = c(0, 75))
```



Caso práctico

```
> names(iris.pca)
```

```
[1] "eig"  "var"  "ind"  "svd"  "call"
```

Caso práctico

```
> print(iris.pca)
```

```
**Results for the Principal Component Analysis (PCA)**
The analysis was performed on 150 individuals, described by 4 variables
*The results are available in the following objects:

   name               description
1  "$eig"             "eigenvalues"
2  "$var"             "results for the variables"
3  "$var$coord"       "coord. for the variables"
4  "$var$cor"         "correlations variables - dimensions"
5  "$var$cos2"        "cos2 for the variables"
6  "$var$contrib"     "contributions of the variables"
7  "$ind"             "results for the individuals"
8  "$ind$coord"       "coord. for the individuals"
9  "$ind$cos2"        "cos2 for the individuals"
10 "$ind$contrib"     "contributions of the individuals"
11 "$call"            "summary statistics"
12 "$call$centre"     "mean of the variables"
13 "$call$ecart.type" "standard error of the variables"
14 "$call$row.w"      "weights for the individuals"
15 "$call$col.w"      "weights for the variables"
```

Caso práctico

```
> iris.pca$eig
```

```
       eigenvalue percentage of variance cumulative percentage of variance
comp 1 2.91849782             72.9624454                          72.96245
comp 2 0.91403047             22.8507618                          95.81321
comp 3 0.14675688              3.6689219                          99.48213
comp 4 0.02071484              0.5178709                         100.00000
```

Caso práctico

loading (vectores propios asociados al valor propio)

```
> iris.pca$svd$V
```

```
           [,1]       [,2]       [,3]       [,4]
[1,]  0.5210659 0.37741762 -0.7195664 -0.2612863
[2,] -0.2693474 0.92329566  0.2443818  0.1235096
[3,]  0.5804131 0.02449161  0.1421264  0.8014492
[4,]  0.5648565 0.06694199  0.6342727 -0.5235971
```

$$Y_1 = 0.52 * Sepal.Length - 0.27 * Sepal.Width + 0.58 * Petal.Length + 0.56 * Petal.Width$$

$$Y_2 = 0.38 * Sepal.Length + 0.92 * Sepal.Width + 0.02 * Petal.Length + 0.07 * Petal.Width$$

```
> (iris.pca$var$coord) #Coordenadas de las variables en las componentes principales
```

```
             Dim.1      Dim.2       Dim.3       Dim.4
SLength  0.8901688 0.36082989 -0.27565767 -0.03760602
SWidth  -0.4601427 0.88271627  0.09361987  0.01777631
PLength  0.9915552 0.02341519  0.05444699  0.11534978
PWidth   0.9649790 0.06399985  0.24298265 -0.07535950
```

Caso práctico

```
> iris.pca$var$cor
```

```
              Dim.1      Dim.2       Dim.3       Dim.4
SLength   0.8901688 0.36082989 -0.27565767 -0.03760602
SWidth   -0.4601427 0.88271627  0.09361987  0.01777631
PLength   0.9915552 0.02341519  0.05444699  0.11534978
PWidth    0.9649790 0.06399985  0.24298265 -0.07535950
```

```
> cor(iris$SLength, iris.pca$ind$coord ) #Correlación entre las variables y las componentes
```

```
         Dim.1     Dim.2      Dim.3       Dim.4
[1,] 0.8901688 0.3608299 -0.2756577 -0.03760602
```

Caso práctico

Calidad de representación de cada variable sobre los ejes

Cosenos al cuadrado y su acumulación para los factores retenidos es la comunalidad

```
> iris.pca$var$cos2
```

```
           Dim.1       Dim.2       Dim.3        Dim.4
SLength 0.7924004 0.130198208 0.075987149 0.0014142127
SWidth  0.2117313 0.779188012 0.008764681 0.0003159971
PLength 0.9831817 0.000548271 0.002964475 0.0133055723
PWidth  0.9311844 0.004095980 0.059040571 0.0056790544
```

```
> rowSums(iris.pca$var$cos2)
```

```
SLength  SWidth PLength  PWidth 
      1       1       1       1 
```

Caso práctico

Contribuciones de cada variable sobre los componentes principales

```
> iris.pca$var$contrib
```

```
            Dim.1       Dim.2     Dim.3     Dim.4
SLength 27.150969 14.24440565 51.777574  6.827052
SWidth   7.254804 85.24748749  5.972245  1.525463
PLength 33.687936  0.05998389  2.019990 64.232089
PWidth  31.906291  0.44812296 40.230191 27.415396
```

```
> colSums(iris.pca$var$contrib)
```

```
Dim.1 Dim.2 Dim.3 Dim.4
  100   100   100   100
```

Caso práctico

Valores propios

```
> iris.pca$eig
```

```
       eigenvalue percentage of variance cumulative percentage of variance
comp 1 2.91849782             72.9624454                          72.96245
comp 2 0.91403047             22.8507618                          95.81321
comp 3 0.14675688              3.6689219                          99.48213
comp 4 0.02071484              0.5178709                         100.00000
```

Caso práctico

Distancias de los individuos al centro de gravedad de la nube de puntos

```
> iris.pca$ind$dist
```

```
        1         2         3         4         5         6         7         8         9        10
2.3186628 2.2023867 2.3894014 2.3783885 2.4761421 2.5547337 2.4676790 2.2458571 2.5915769 2.2488335 2.4196
       31        32        33        34        35        36        37        38        39        40
2.2005982 1.9144052 3.1795671 3.2586152 2.1664754 2.2352302 2.2121517 2.5994670 2.5996342 2.1933821 2.3314
       61        62        63        64        65        66        67        68        69        70
2.6568072 0.4918154 2.0046733 0.7644111 0.4936869 1.1357908 0.6624551 0.8865720 2.1014703 1.3253442 1.0433
       91        92        93        94        95        96        97        98        99       100
1.1893408 0.6426036 1.0630118 2.0542832 0.9188165 0.3327074 0.4919403 0.6560470 1.6309294 0.6606152 2.2744
      121       122       123       124       125       126       127       128       129       130
2.2504615 1.4035404 3.0526824 1.4245238 2.0033487 2.2492371 1.2302797 1.0773149 1.8186594 2.0844277 2.5556
```

Caso práctico

Coordenadas de los individuos

```
> head(round(iris.pca$ind$coord,2))
```

```
  Dim.1 Dim.2 Dim.3 Dim.4
1 -2.26  0.48 -0.13 -0.02
2 -2.08 -0.67 -0.23 -0.10
3 -2.36 -0.34  0.04 -0.03
4 -2.30 -0.60  0.09  0.07
5 -2.39  0.65  0.02  0.04
6 -2.08  1.49  0.03 -0.01
```

Caso práctico

Calidad de representación de cada individuo en los ejes

```
> head(iris.pca$ind$cos2)
```

```
      Dim.1      Dim.2        Dim.3        Dim.4
1 0.9539975 0.04286032 0.0030335249 1.086460e-04
2 0.8927725 0.09369248 0.0113475382 2.187482e-03
3 0.9790410 0.02047578 0.0003422122 1.410446e-04
4 0.9346682 0.06308947 0.0014732682 7.690193e-04
5 0.9315095 0.06823959 0.0000403979 2.104697e-04
6 0.6600989 0.33978301 0.0001114335 6.690714e-06
```

```
> rowSums(iris.pca$ind$cos2)
```

```
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  2
  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
 76  77  78  79  80  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100 101 10
  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
```

Caso práctico

```
> head(iris.pca$ind$contrib) # Contribuciones de cada uno de los individuos a los ejes
```

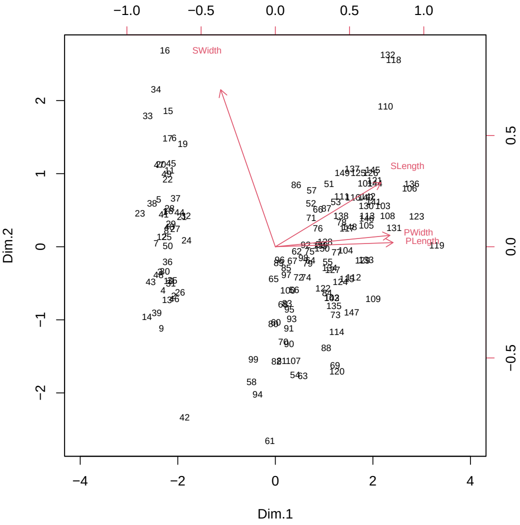
```
      Dim.1      Dim.2       Dim.3       Dim.4
1 1.1715796 0.16806554 0.074085470 0.018798188
2 0.9891845 0.33146674 0.250034006 0.341474919
3 1.2768164 0.08526419 0.008875320 0.025915633
4 1.2077372 0.26029781 0.037858004 0.140000650
5 1.3046313 0.30516562 0.001125175 0.041530572
6 0.9841236 1.61748779 0.003303827 0.001405371
```

```
> colSums(iris.pca$ind$contrib) #Observación!!!
```

```
Dim.1 Dim.2 Dim.3 Dim.4
  100   100   100   100
```

Caso práctico

```
> biplot(iris.pca$ind$coord[,1:2],iris.pca$var$coord[,1:2],xlim=c(-4,4), cex = 0.7)
```



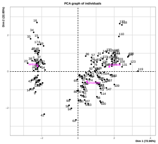
Caso práctico

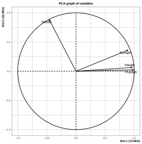
```
> dimdesc(iris.pca, axes=c(1,2))
```

```
$Dim.1

Link between the variable and the continuous variables (R-square)
=================================================================================
        correlation       p.value
PLength   0.9915552 3.369916e-133
PWidth    0.9649790  6.609632e-88
SLength   0.8901688  2.190813e-52
SWidth   -0.4601427  3.139724e-09

$Dim.2

Link between the variable and the continuous variables (R-square)
=================================================================================
        correlation      p.value
SWidth    0.8827163 2.123801e-50
SLength   0.3608299 5.731933e-06
```

Caso práctico

```r
> iris.pca=PCA(iris, scale.unit=T, ncp=5, graph=T,quali.sup=5)
```





Caso práctico

```
> iris.pca$quali.sup
```

Caso práctico

```
> iris.pca$quali.sup
```

```
$coord
               Dim.1      Dim.2       Dim.3       Dim.4
setosa     -2.224753  0.2889275 -0.04283910 -0.01834076
versicolor  0.496448 -0.5501705 -0.09612951  0.03034009
virginica   1.728305  0.2612430  0.13896861 -0.01199933

$cos2
               Dim.1      Dim.2        Dim.3        Dim.4
setosa     0.9829895 0.01657917 0.0003644734 6.680665e-05
versicolor 0.4406515 0.54118069 0.0165219448 1.645816e-03
virginica  0.9714759 0.02219630 0.0062809398 4.682797e-05

$v.test
                Dim.1     Dim.2      Dim.3      Dim.4
setosa     -11.240362  2.608475 -0.9652043 -1.0999036
versicolor   2.508258 -4.967011 -2.1658862  1.8195091
virginica    8.732103  2.358536  3.1310905 -0.7196055

$dist
    setosa versicolor  virginica
 2.2439201  0.7478701  1.7534945

$eta2
            Dim.1     Dim.2      Dim.3      Dim.4
Species 0.9346162 0.1657182 0.06902197 0.02254243
```

Caso práctico

Determinación del número de factores para extraer

```
> library(nFactors)
> ev <- eigen(cor(iris[,1:4]))
> ev
```

```
eigen() decomposition
$values
[1] 2.91849782 0.91403047 0.14675688 0.02071484

$vectors
           [,1]        [,2]       [,3]       [,4]
[1,]  0.5210659 -0.37741762  0.7195664  0.2612863
[2,] -0.2693474 -0.92329566 -0.2443818 -0.1235096
[3,]  0.5804131 -0.02449161 -0.1421264 -0.8014492
[4,]  0.5648565 -0.06694199 -0.6342727  0.5235971
```

Caso práctico

Determinación del número de factores para extraer

```
> ap <- parallel(subject=nrow(iris[,1:4]),var=ncol(iris[,1:4]),rep=100,cent=.95)
> nS <- nScree(ev$values,ap$eigen$qevpea)
> vp <- min(nS$Components)+1;vp
```

```
[1] 3
```

Caso práctico

```r
> op <- par(mfrow=c(1,2))
> plot.PCA(iris.pca,axes=c(1,2), choix="ind", habillage=5)
```


```r
> plotellipses(iris.pca,keepvar=5)
```