



# MultBiplotR: Multivariate Analysis using Biplots (in R)

José Luis Vicente Villardón  
Dpto. de Estadística  
Universidad de Salamanca  
[villardon@usal.es](mailto:villardon@usal.es)

# Biplot

Sea  $\mathbf{X}$  una matriz de datos (centrada y/o estandarizada) que contiene las medidas de  $n$  individuos en  $p$  variables. Un biplot en dimensión  $q$  es una representación gráfica mediante marcadores  $\mathbf{A}_{n \times q}$  y  $\mathbf{B}_{p \times q}$  (puntos o vectores) para las filas y las columnas respectivamente, de forma que el producto  $\mathbf{AB}'$  aproxime  $\mathbf{X}$  tan bien como sea posible.

$$\mathbf{X} = \mathbf{AB}' + \mathbf{E}$$

Donde  $\mathbf{E}$  es una matriz de residuales. La factorización no es única, es decir, hay infinitos biplots que aproximan la matriz de la misma manera. Para hacer la representación única se toman las columnas de  $\mathbf{A}$  o de  $\mathbf{B}$  para que sean ortonormales.

# Matriz de Datos

R Data Editor

Copy

Paste

Quit

	row.names	Comunist	Region	Red_Meat	White_Meat	Eggs	Milk	Fish	Cereal	Starch	Nuts	Fruits_Vegetables
1	Albania	Yes	South	10.1	1.4	0.5	8.9	0.2	42.3	0.6	5.5	1.7
2	Austria	No	Center	8.9	14	4.3	19.9	2.1	28	3.6	1.3	4.3
3	Belgium	No	Center	13.5	9.3	4.1	17.5	4.5	26.6	5.7	2.1	4
4	Bulgaria	Yes	South	7.8	6	1.6	8.3	1.2	56.7	1.1	3.7	4.2
5	Czechoslovakia	Yes	Center	9.7	11.4	2.8	12.5	2	34.3	5	1.1	4
6	Denmark	No	North	10.6	10.8	3.7	25	9.9	21.9	4.8	0.7	2.4
7	E_Germany	Yes	Center	8.4	11.6	3.7	11.1	5.4	24.6	6.5	0.8	3.6
8	Finland	No	North	9.5	4.9	2.7	33.7	5.8	26.3	5.1	1	1.4
9	France	No	Center	18	9.9	3.3	19.5	5.7	28.1	4.8	2.4	6.5
10	Greece	No	South	10.2	3	2.8	17.6	5.9	41.7	2.2	7.8	6.5
11	Hungary	Yes	Center	5.3	12.4	2.9	9.7	0.3	40.1	4	5.4	4.2
12	Ireland	No	Center	13.9	10	4.7	25.8	2.2	24	6.2	1.6	2.9
13	Italy	No	South	9	5.1	2.9	13.7	3.4	36.8	2.1	4.3	6.7
14	Holand	No	Center	9.5	13.6	3.6	23.4	2.5	22.4	4.2	1.8	3.7
15	Norway	No	North	9.4	4.7	2.7	23.3	9.7	23	4.6	1.6	2.7
16	Poland	Yes	Center	6.9	10.2	2.7	19.3	3	36.1	5.9	2	6.6
17	Portugal	No	South	6.2	3.7	1.1	4.9	14.2	27	5.9	4.7	7.9
18	Romania	Yes	South	6.2	6.3	1.5	11.1	1	49.6	3.1	5.3	2.8
19	Spain	No	South	7.1	3.4	3.1	8.6	7	29.2	5.7	5.9	7.2
20	Sweden	No	North	9.9	7.8	3.5	24.7	7.5	19.5	3.7	1.4	2
21	Switzerland	No	Center	13.1	10.1	3.1	23.8	2.3	25.6	2.8	2.4	4.9
22	UK	No	Center	17.4	5.7	4.7	20.6	4.3	24.3	4.7	3.4	3.3
23	USSR	Yes	Center	9.3	4.6	2.1	16.6	3	43.6	6.4	3.4	2.9
24	W_Germany	No	Center	11.4	12.5	4.1	18.8	3.4	18.6	5.2	1.5	3.8
25	Yugoslavia	Yes	South	4.4	5	1.2	9.5	0.6	55.9	3	5.7	3.2

## MultBiplot

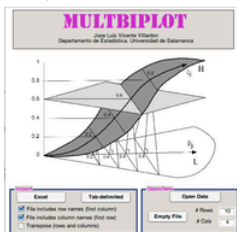
### Multivariate Analysis using Biplots

[Introduction](#) [Disclaimer](#) [System Requirements](#) [License](#) [Download](#) [MultBiplot In R](#) [Links](#) [Documentation](#)  
[Contact](#) [News](#) [References](#)



Subscribe to RSS feed

MultBiplot



Main Window

Universidad de Salamanca



Visitors



#### Introduction

MULTBIPLOT is a program to perform Biplot Analysis. It contains Classical Biplot, HJ-Biplot, Simple and Multiple Correspondence Analysis for a Contingency Table, Canonical/MANOVA Biplot, Logistic Biplots for Binary Data, Common Biplot for Several groups, among many other biplot techniques.

The program tries to implement, in a software tool, the experience of the Applied Statistics Group of the Statistics Department at the Salamanca University (Spain) in working with Biplots.

Most of the software available for the construction of Biplots has been developed for very particular applications or as a small part inside general purpose packages. Usually the graphical representations of most packages are not very flexible, producing static pictures that limit the visual interpretation of the results.

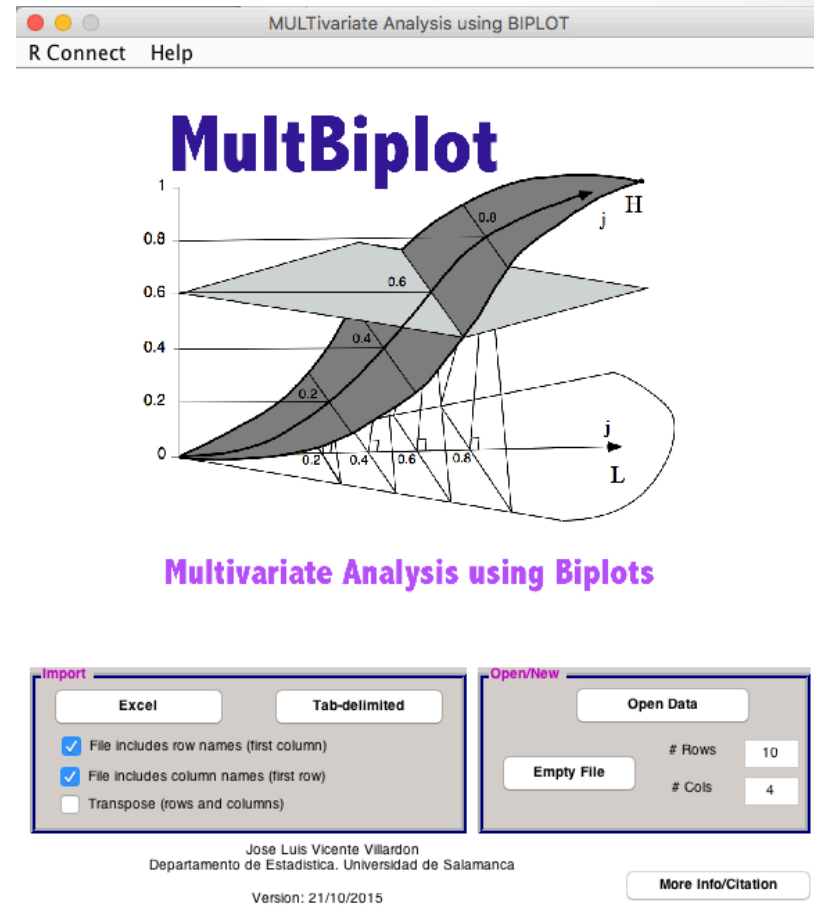
MULTBIPLOT is conceived not to be "another biplot program", but conceived to fill the gap between the static pictures and a more dynamic visual interpretation. The package provides the analyst with an interactive approach that uses colour, selection of important features, partial views of the picture and other graphical tools, to interpret biplots. The package also provides some original points of view of the classical techniques based on some original scientific work of the author.

VICENTE VILLARDÓN, J.L. (2010). *MULTBIPLOT: A package for Multivariate Analysis using Biplots*. Departamento de Estadística. Universidad de Salamanca. (<http://biplot.usal.es/ClassicalBiplot/index.html>)

There has been some problems downloading the Matlab Libraries. We hope the problems are solved with the new compilation of the program.

José Luis Vicente Villardón - Departamento de Estadística - Universidad de Salamanca. ([villardon@usal.es](mailto:villardon@usal.es))

# MultBiplot



<http://biplot.usal.es/classicalbiplot/>

# Selección

CLASSICAL BIPLLOT (ACP)

## BIPLOT

### Principal Components Analysis

**Rows/Individuals**

1.- Albania  
2.- Austria  
3.- Bulgaria  
4.- Bulgaria  
5.- Checoslova  
6.- Dinamarca  
7.- Alemania O  
8.- Finlandia  
9.- Francia  
10.- Grecia  
11.- Hungría  
12.- Irlanda  
13.- Italia

**Active**

Select ->  
All -->  
<- Rem.  
<- All

**Illustrative**

Select ->  
<- Rem.

**Data Transformations**

☐ Initial Log Transformation

Standardize columns

**Estimation Method**

Singular Value Decomposi...

**Biplot Type**

0-GH 1 1-JK

**Dimension**

3 ☐ 3D Representation

**Continuous Columns/Variables**

1.- Carne Roja  
2.- Carne Blanca  
3.- Huevso  
4.- Leche  
5.- Pescado  
6.- Cereales  
7.- Frutas  
8.- Frutos secos  
9.- Frutos y vegeta

**Active**

Select ->  
Sel. All -->  
<- Rem.  
<- Rem. All

**Illustrative**

Select ->  
<- Rem.

☒ Include Original Data Matrix in Plots for Clusters

**Results**

☐ Coordinates ☒ Contributions

☒ QLR (Cummulative contributions)

☐ Ordered Contributions

☐ Ordered QLR

**Selection Criteria**

QLR > 0.0

**External Variables**

Binary

1.- No Comun.  
2.- Localizacion

Select ->  
Sel. All -->  
<- Rem.  
<- Re. All

☐ Send to R

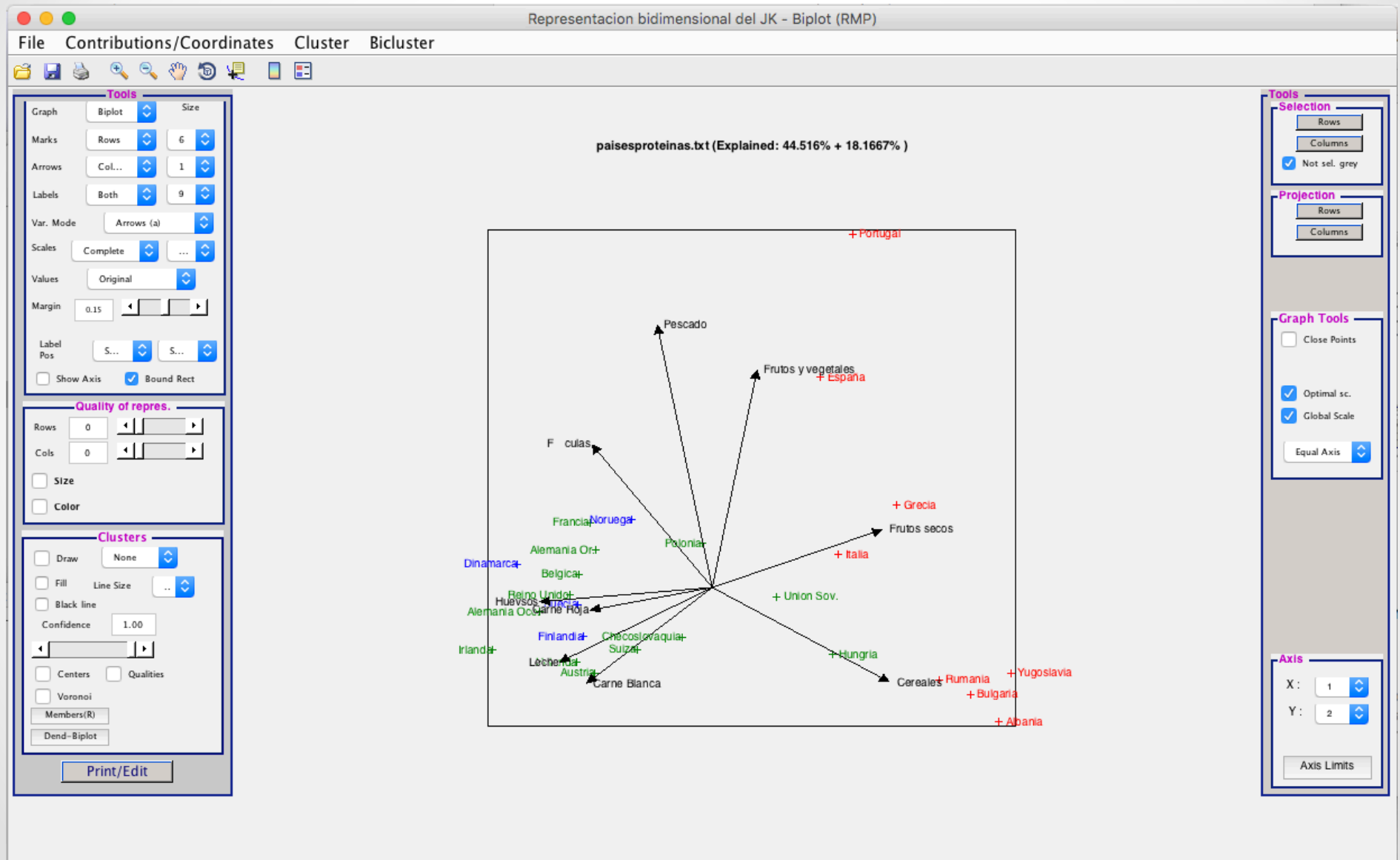
R results

Cancel

OK

☐ Close this window after the analysis

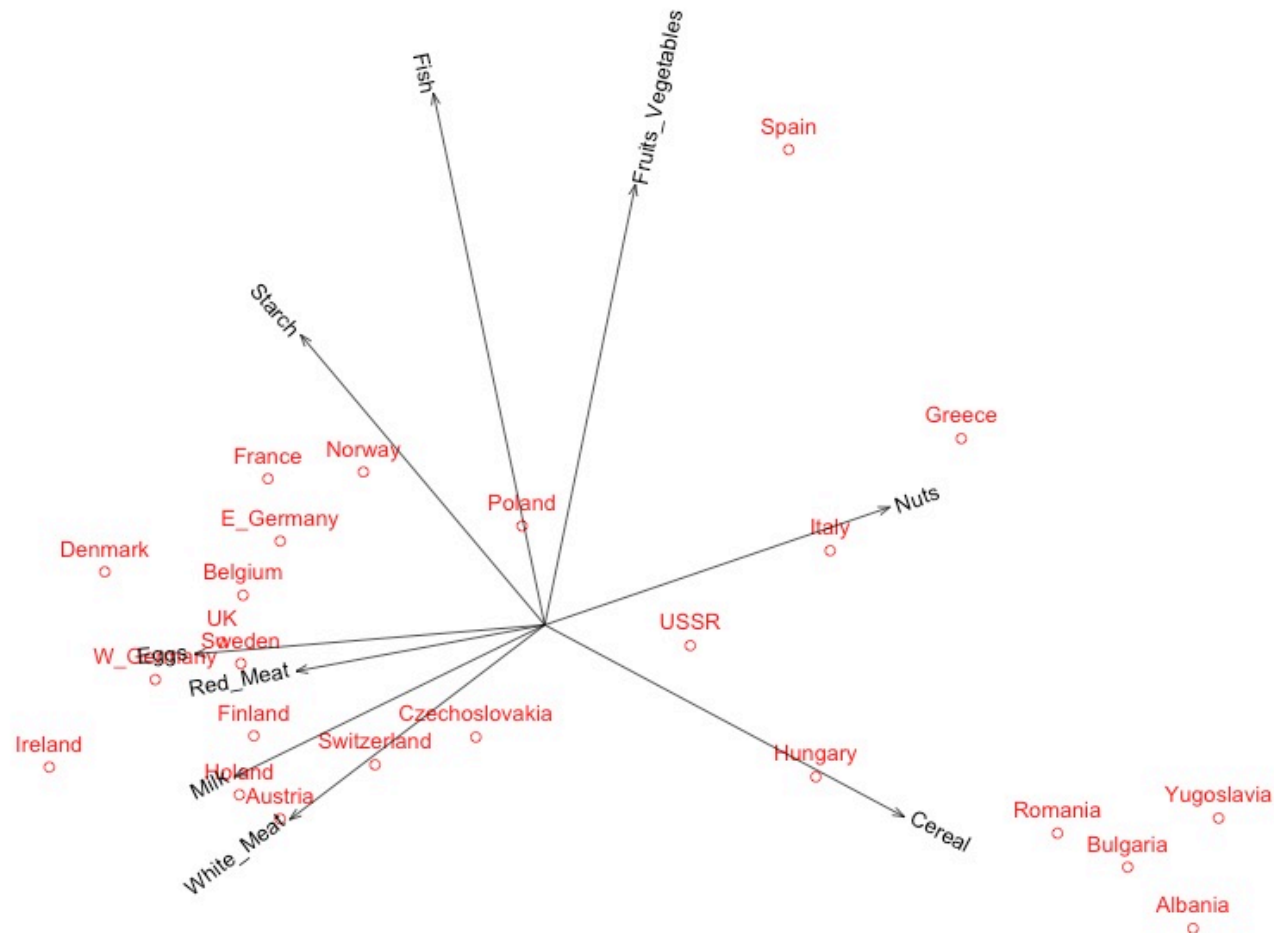
# Dibujo



# Representación gráfica (En R)

PCA Biplot (Dim 1 - 2)

```
data(Protein)
bip=PCA.Biplot(Protein[,3:11])
plot(bip, ShowBox=T, LabelPos=3)
```

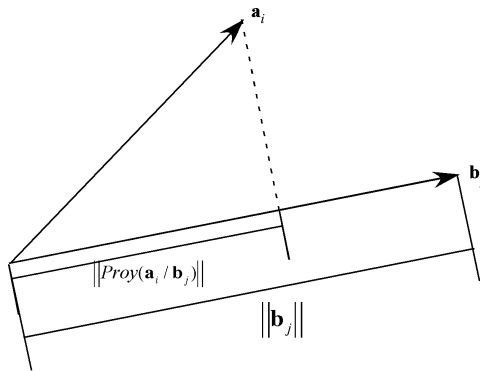




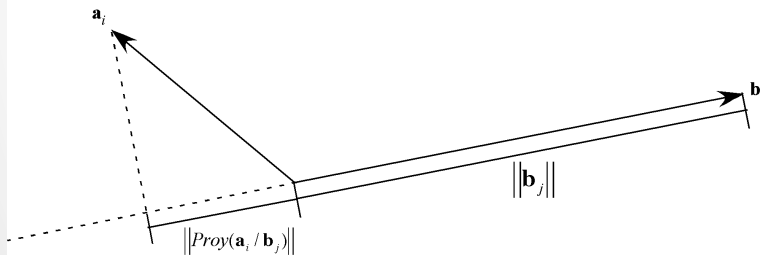
# Producto escalar

Interpretación del producto escalar

$$x_{ij} \approx \mathbf{a}_i^T \mathbf{b}_j = a_{i1}b_{j1} + \dots + a_{iq}b_{jq} = \sum_{k=1}^q a_{ik}b_{jk}$$

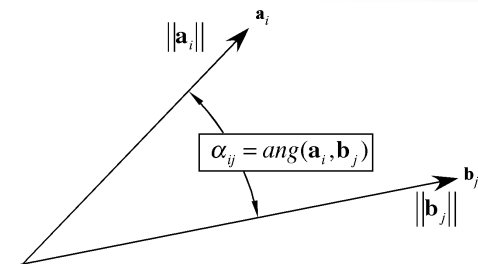


Producto escalar positivo

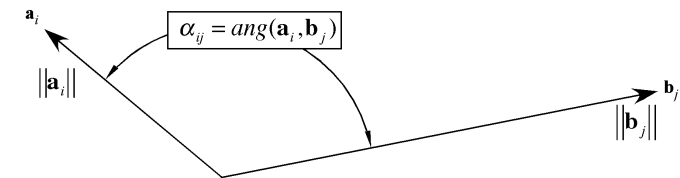


Producto escalar negativo

$$\mathbf{a}_i^T \mathbf{b}_j = \|\text{Proy}(\mathbf{a}_i / \mathbf{b}_j)\| \odot \|\mathbf{b}_j\|$$



Producto escalar positivo



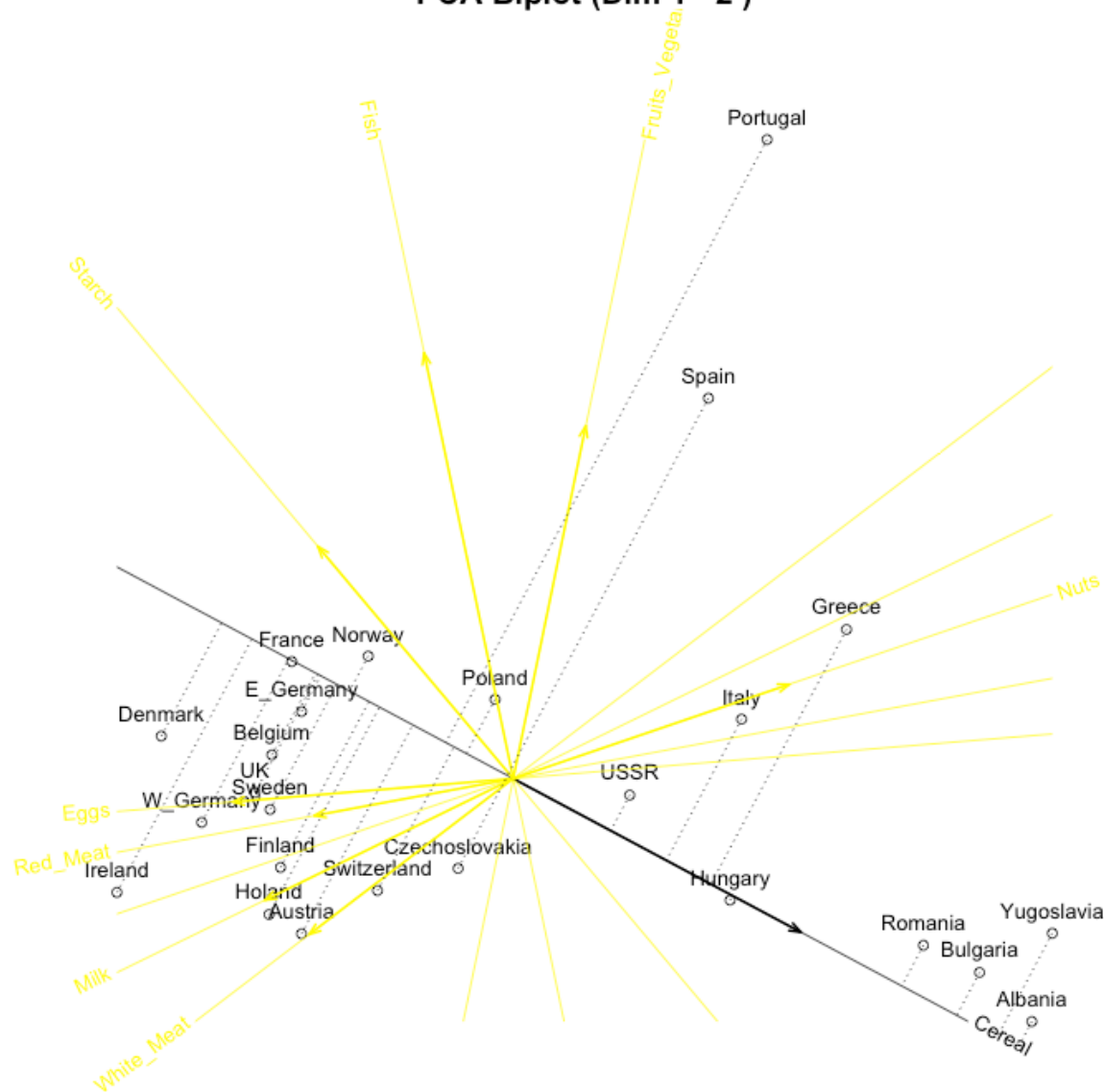
Producto escalar negativo

$$\mathbf{a}_i^T \mathbf{b}_j = \|\mathbf{a}_i\| \|\mathbf{b}_j\| \cos(\mathbf{a}_i, \mathbf{b}_j)$$



# Representación gráfica

PCA Biplot (Dim 1 - 2)



```
Colores=c(7,7,7,7,7,1,7,7,7)
```

```
plot(bip, mode='b', LabelPos=3, dp=6, ColorVar=Colores, ColorInd=1, margin=0.1)
```

# Biplot (Componentes Principales)

Por ejemplo  $\mathbf{B}$  puede tomarse como los  $q$  primeros vectores propios de  $\mathbf{X}'\mathbf{X}$ ,  $\mathbf{V}_q$ , y como  $\mathbf{A}$  las proyecciones sobre el subespacio definido por  $\mathbf{V}_q$ ,  $\mathbf{A} = \mathbf{X}\mathbf{V}_q$ .

$$\mathbf{X} = \mathbf{X}\mathbf{V}_q\mathbf{V}_q' + \mathbf{E}$$

En este caso tenemos que el biplot es equivalente al Análisis de Componentes Principales. (en la versión JK o RMP Biplot)

Las coordenadas de las filas son las coordenadas sobre las Componentes Principales y las coordenadas sobre las columnas los vectores propios  $\mathbf{V}_q$ , que también se pueden entender como la proyección de los ejes unitarios en el espacio  $p$  dimensional, asociados a la matriz identidad  $\mathbf{I}$ , sobre el espacio de las Componentes Principales

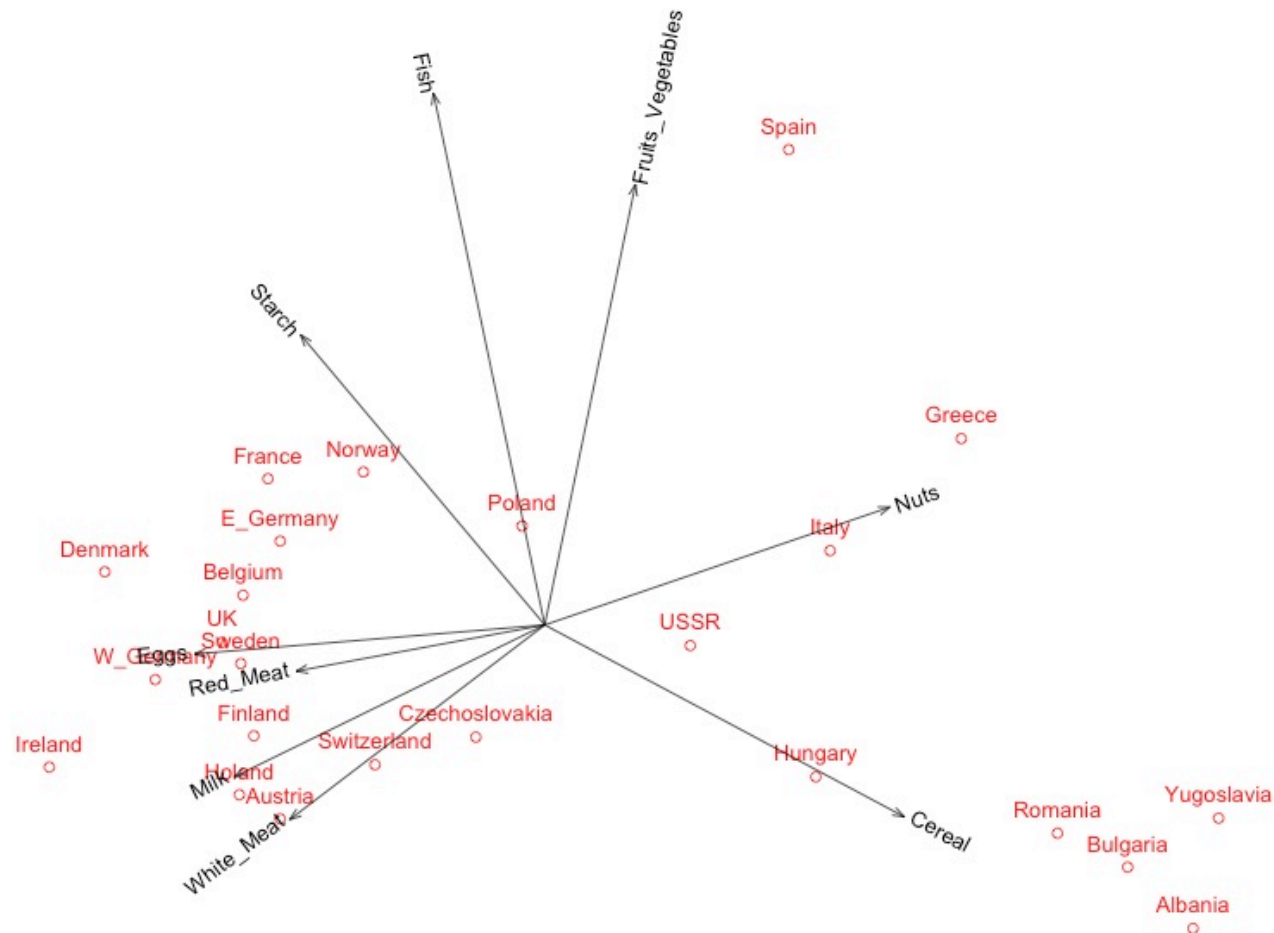
$$\mathbf{V}_q = \mathbf{I}\mathbf{V}_q$$

Los vectores propios forman una base alternativa del espacio  $p$ -dimensional en el que se describen los individuos, puede verse como una rotación de los  $p$  ejes cartesianos, dada por la matriz  $\mathbf{V}$ .

# Biplot de Componentes Principales

PCA Biplot (Dim 1 - 2)

```
data(Protein)
bip=PCA.Biplot(Protein[,3:11], alpha=1)
plot(bip, ShowBox=T, LabelPos=3)
```



# Biplot de Covarianzas/ Correlaciones (Análisis Factorial)

Se trata del GH-Biplot obtenido a partir de la descomposición en valores singulares.

$$\mathbf{X} = \mathbf{AB}' = \mathbf{U}\Lambda^{1/2}\mathbf{V}'$$

Tomando  $\mathbf{A} = \mathbf{U}$        $\mathbf{B} = \mathbf{V}\Lambda^{1/2}$

Teniendo en cuenta que  $\mathbf{U} = \mathbf{U}\Lambda^{1/2}\Lambda^{-1/2} = \mathbf{XV}\Lambda^{-1/2}$

La matriz base para la interpolación es ahora  $\mathbf{V}_{\text{int}} = \mathbf{V}\Lambda^{-1/2}$

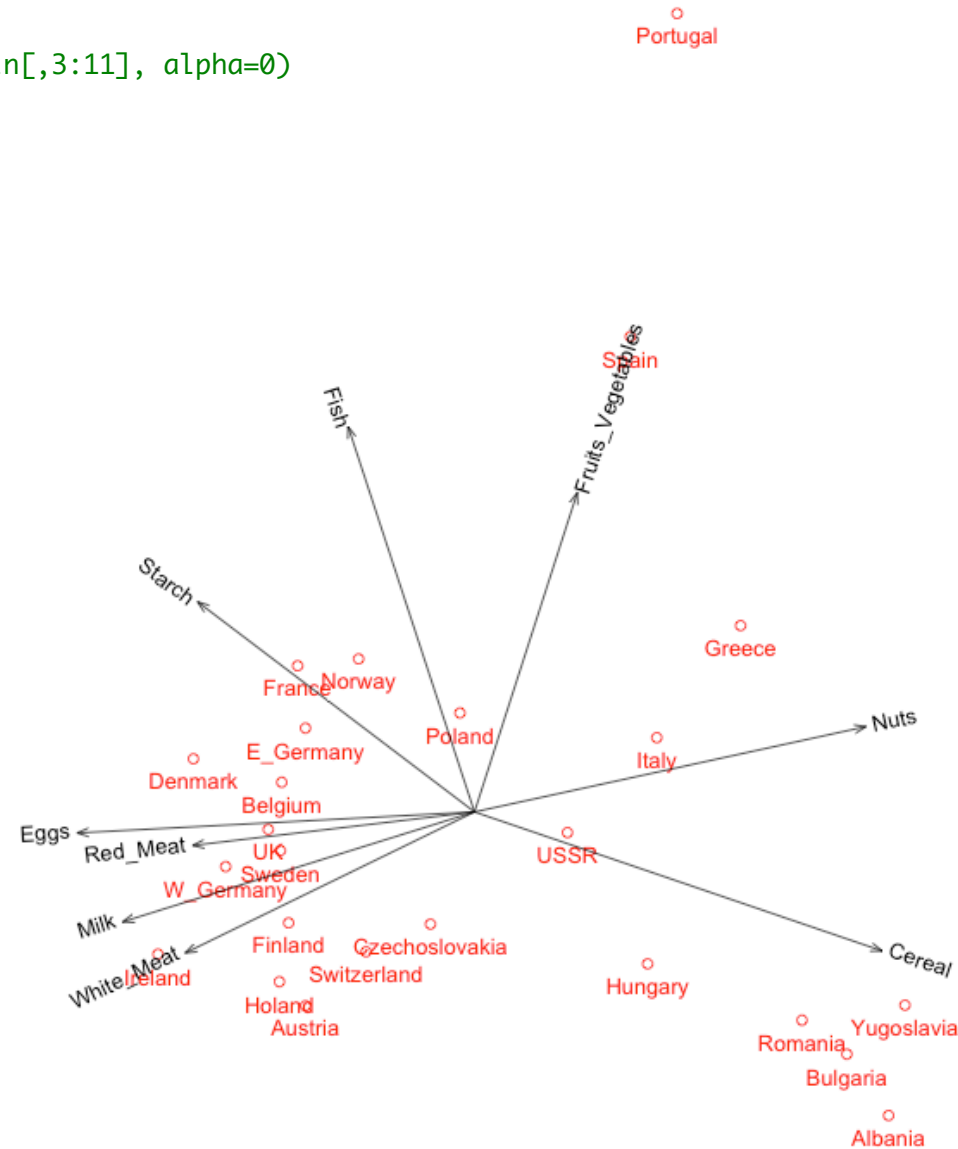
La matriz base para la predicción es entonces  $\mathbf{B} = \mathbf{V}\Lambda^{1/2}$

En este caso tenemos que el biplot es equivalente al Análisis Factorial. (en la versión GH o CMP Biplot)

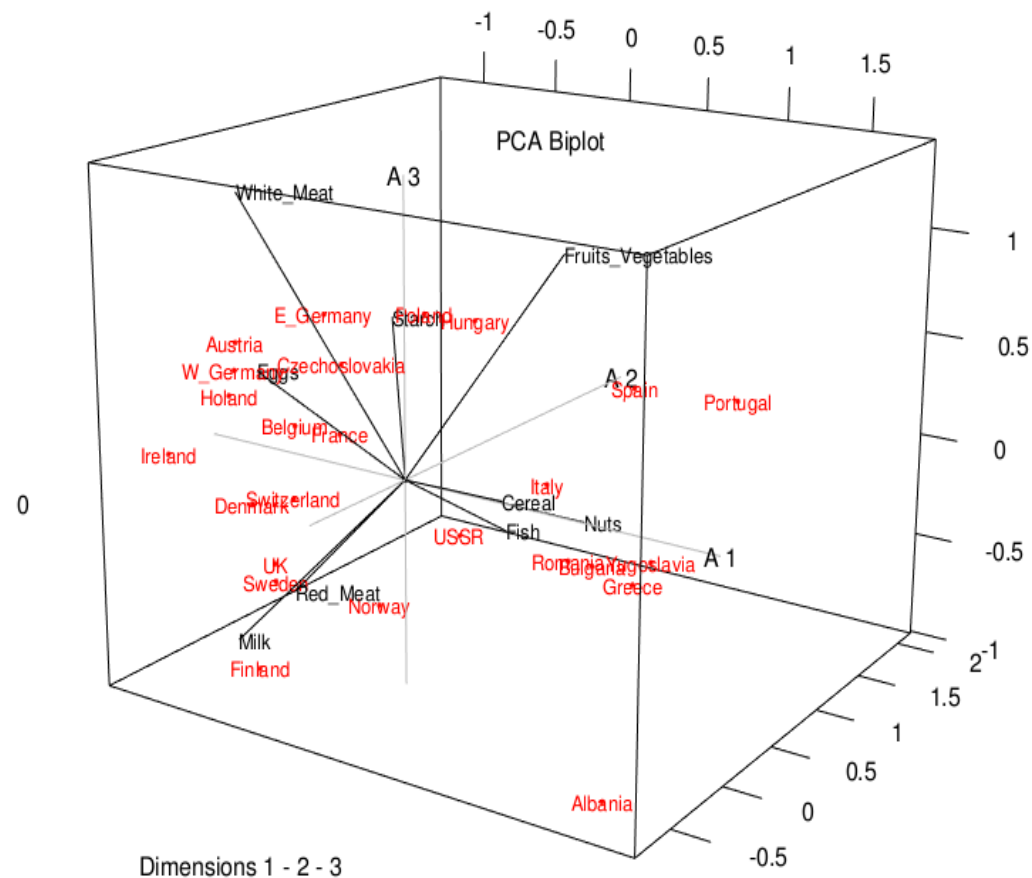
# Biplot del Análisis Factorial

AF Biplot (Dim 1 - 2 )

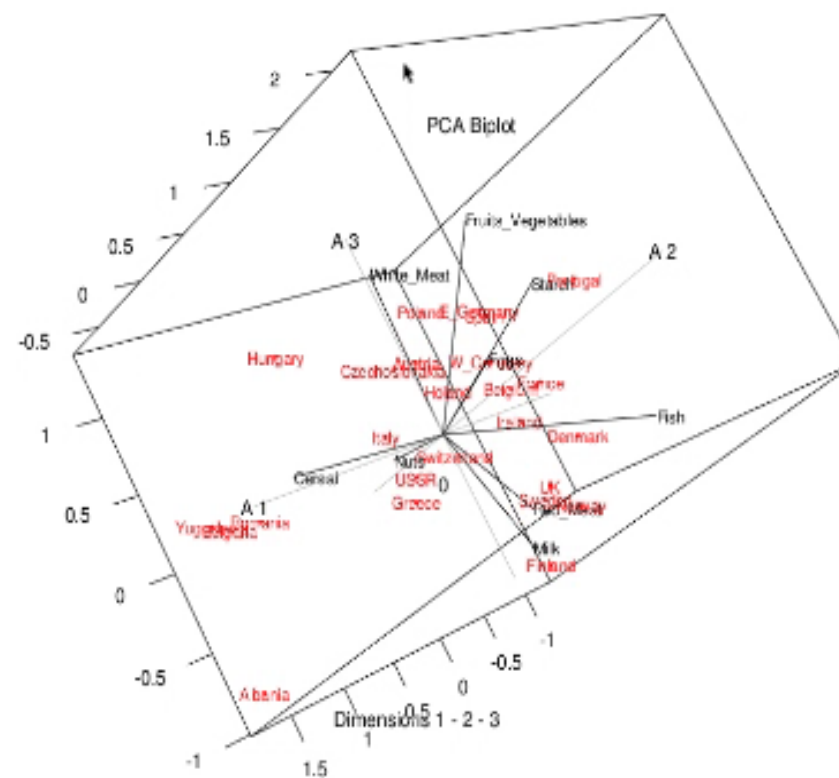
```
data(Protein)
bip=PCA.Biplot(Protein[,3:11], alpha=0)
plot(bip)
```



# 3D Biplot



# 3D Biplot





# Predicción

Podemos predecir la coordenadas en el espacio  $p$ -dimensional de un punto  $\mathbf{y} = (y_1, \dots, y_r)$ , descrito en términos del sistema de referencia de  $L$ . La predicción son las coordenadas del punto en el  $p$ -espacio.

Un punto  $\mathbf{x} = (x_1, \dots, x_p)$  descrito en términos de las coordenadas del  $p$ -espacio pero que está en el subespacio se proyecta en si mismo, es decir,

$$\mathbf{x} = \mathbf{x} \mathbf{V}_q (\mathbf{V}_q' \mathbf{V}_q)^{-1} \mathbf{V}_q' = \mathbf{x} \mathbf{V}_q \mathbf{V}'$$

Pero, sabemos que  $\mathbf{y} = \mathbf{x} \mathbf{V}_q$  en su descripción en términos de las coordenadas del subespacio, luego,  $\mathbf{x} = \mathbf{y} (\mathbf{V}_q' \mathbf{V}_q)^{-1} \mathbf{V}_q'$ . Como las columnas de  $\mathbf{V}_q$  son ortonormales se tiene que,  $\mathbf{x} = \mathbf{y} \mathbf{V}_q'$ . Las coordenadas, en el espacio  $p$ -dimensional, del interpolado de  $\mathbf{x}$  son  $\mathbf{x} \mathbf{V}_q \mathbf{V}_q'$  y las predicciones de la muestra original son  $\mathbf{X} \mathbf{V}_q \mathbf{V}_q'$ .

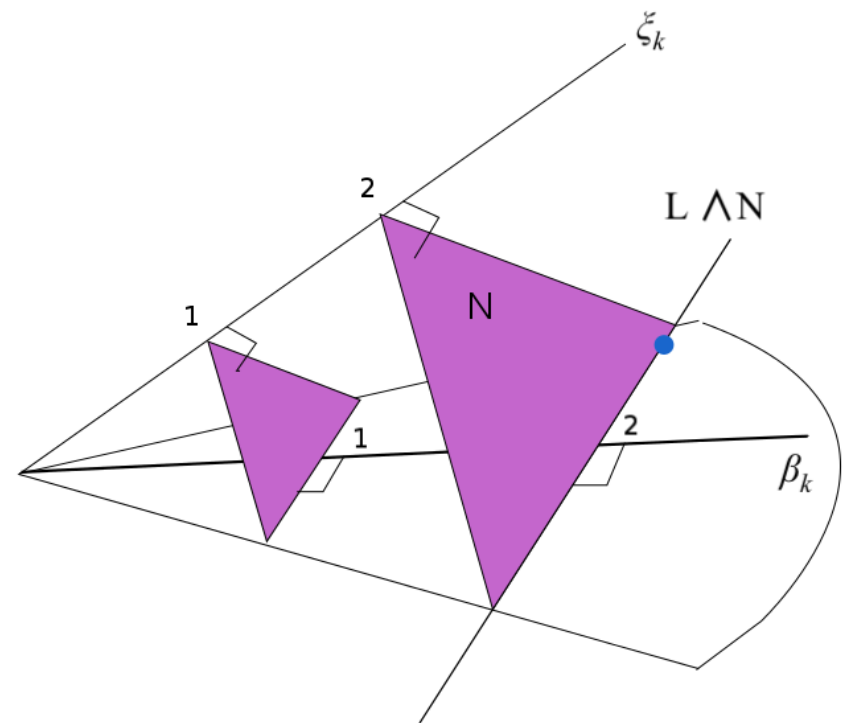
De forma más simple, la predicción de un valor  $x_{ij}$  de  $\mathbf{X}$ , es el producto escalar del marcador que representa a la fila  $\mathbf{a}_i$  por el marcador que representa a su columna  $\mathbf{b}_j$ .

$$x_{ij} = \mathbf{a}_i \mathbf{b}_j'$$

# Predicción

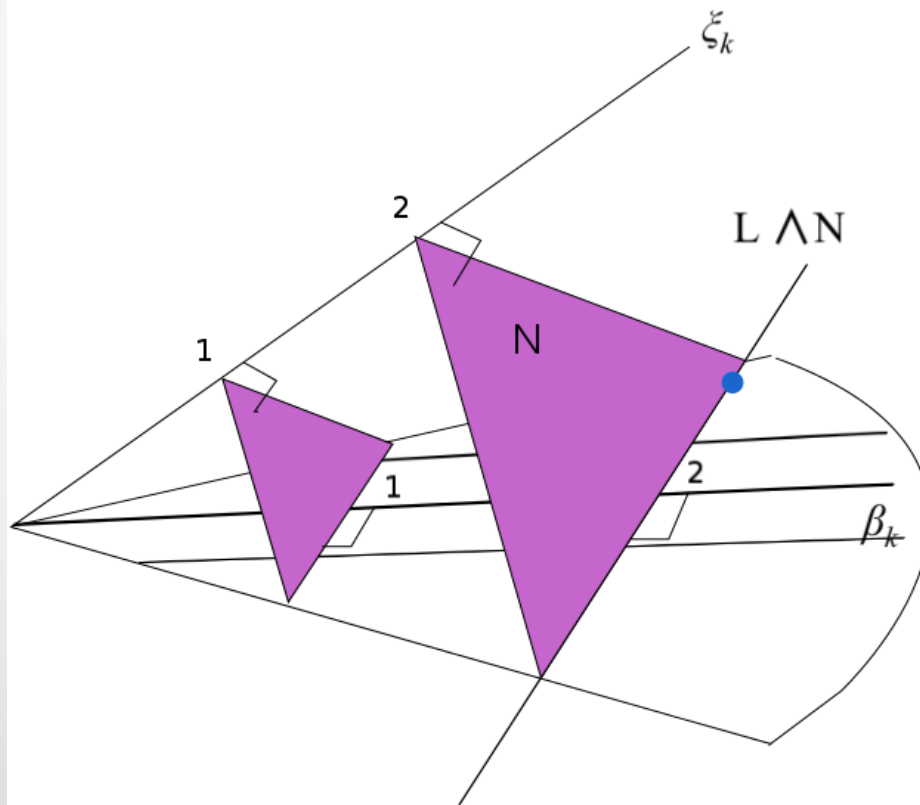
Consideremos el hiperplano  $N$  perpendicular al  $k$ -ésimo eje cartesiano  $\xi_k$  que pasa por el marcador unidad en tal eje. Este se corta con la representación en un subespacio  $(q-1)$ -dimensional (una recta cuando la representación es bidimensional)  $L \wedge N$ . Todos los puntos de este subespacio deberían predecir el valor 1. Podemos definir un eje de predicción como una línea (que pasa por el origen) en el espacio de la representación, perpendicular a  $L \wedge N$  y que se corta con él en un punto que será marcado con la unidad. Otros marcadores se añadirán en la escala, a intervalos iguales, para producir un eje convencional  $\beta_k$ . En el biplot que estamos trabajando la dirección es la misma que para la interpolación, es decir,  $\mathbf{v}_k$ . Obsérvese que estamos utilizando el hecho de que la representación se realiza en un subespacio del original.

En general, ambas direcciones no tienen que coincidir.



# Predicción

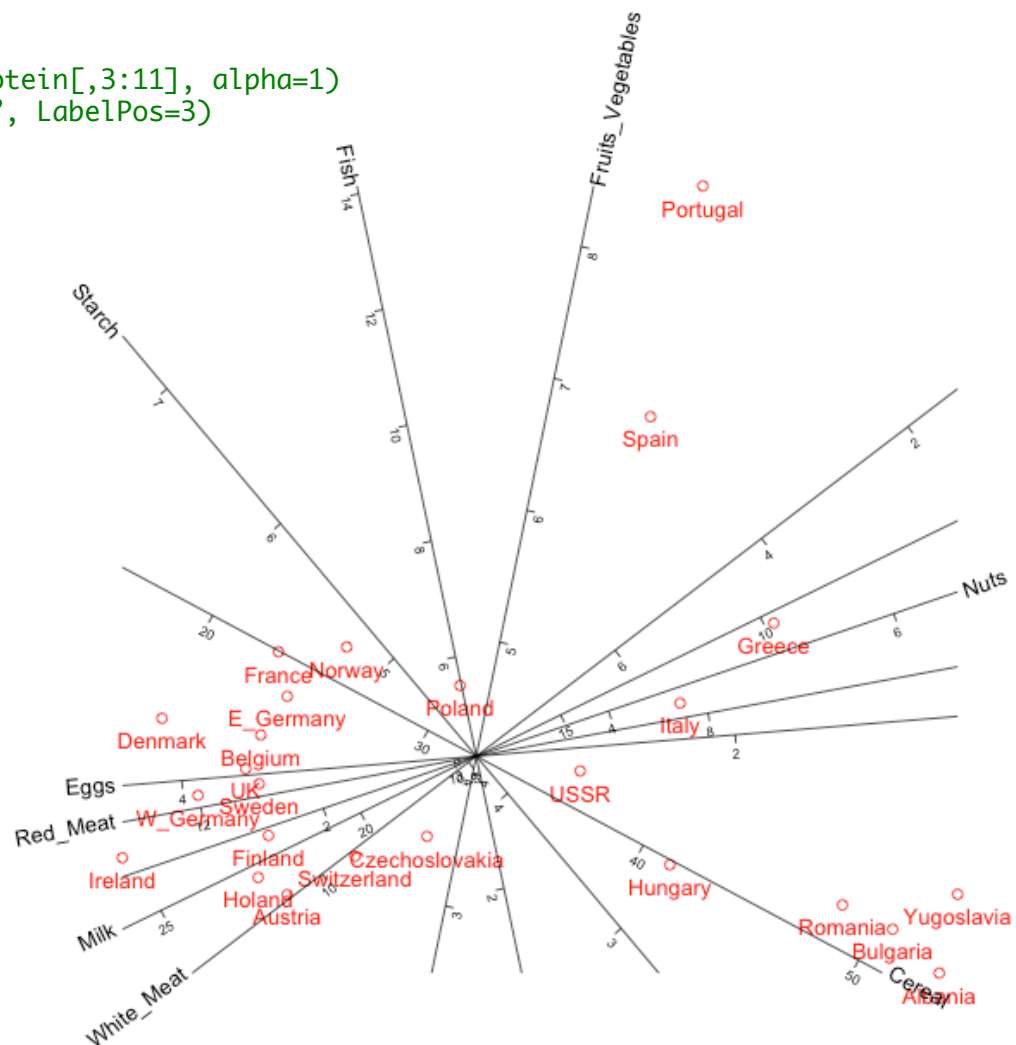
Este método no es único, ya que podemos seleccionar muchas líneas en el subespacio de la representación que se cruzan con  $L \wedge N$  y son ortogonales para asegurar la unicidad, elegimos la línea que sea ortogonal a  $L \wedge N$  y pase por el origen.



Esta línea es la proyección de  $\xi_k$  sobre  $L$  y coincide con la dirección del eje biplot de interpolación  $\beta_k$  explicado anteriormente. Sin embargo, los marcadores inducidos por la intersección con  $N$  no son las proyecciones sobre  $L$ , de forma que difieren de los marcadores de interpolación.

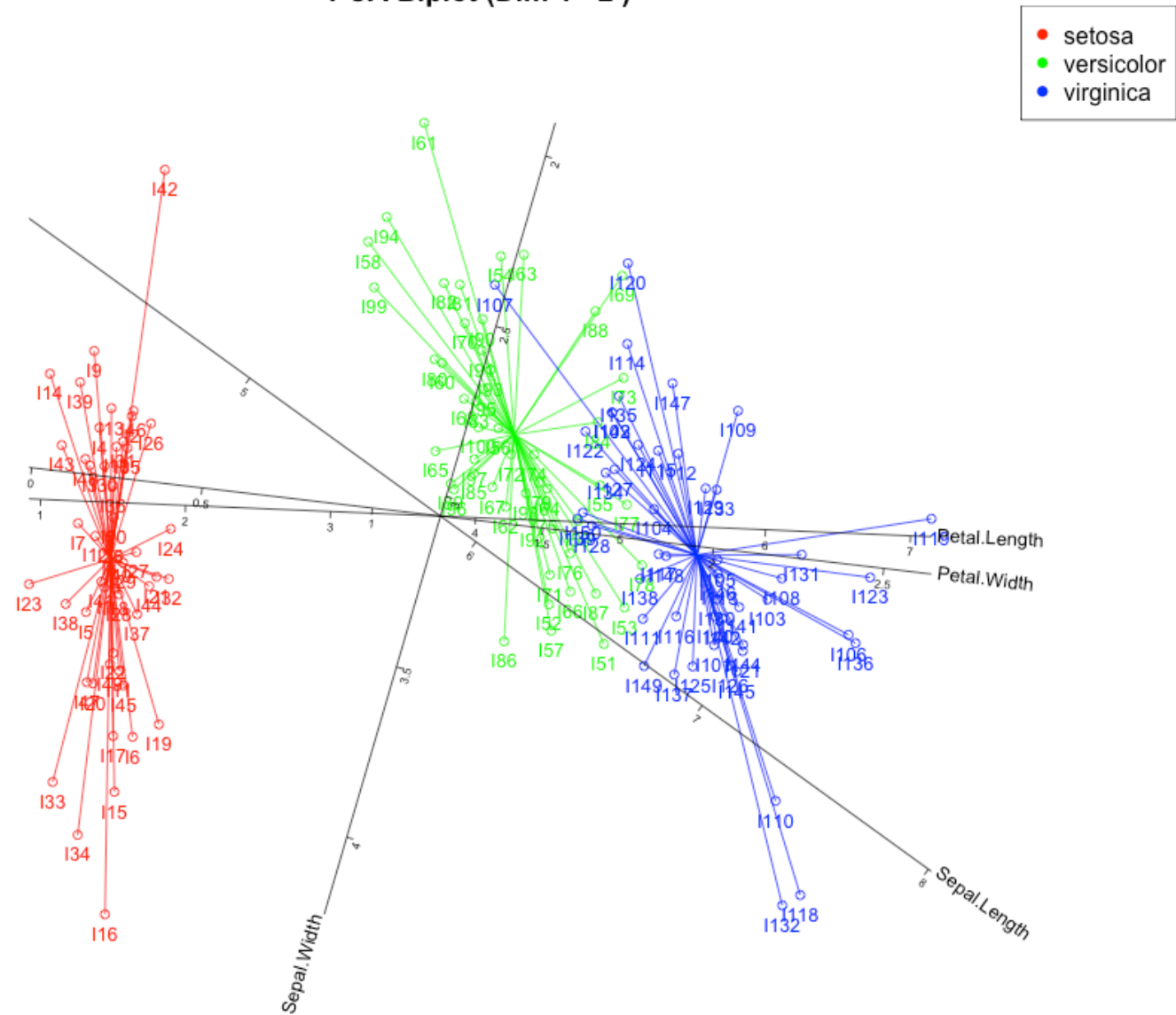
# Biplot Calibrado

```
data(Protein)
bip=PCA.Biplot(Protein[,3:11], alpha=1)
plot(bip, mode="s", LabelPos=3)
```

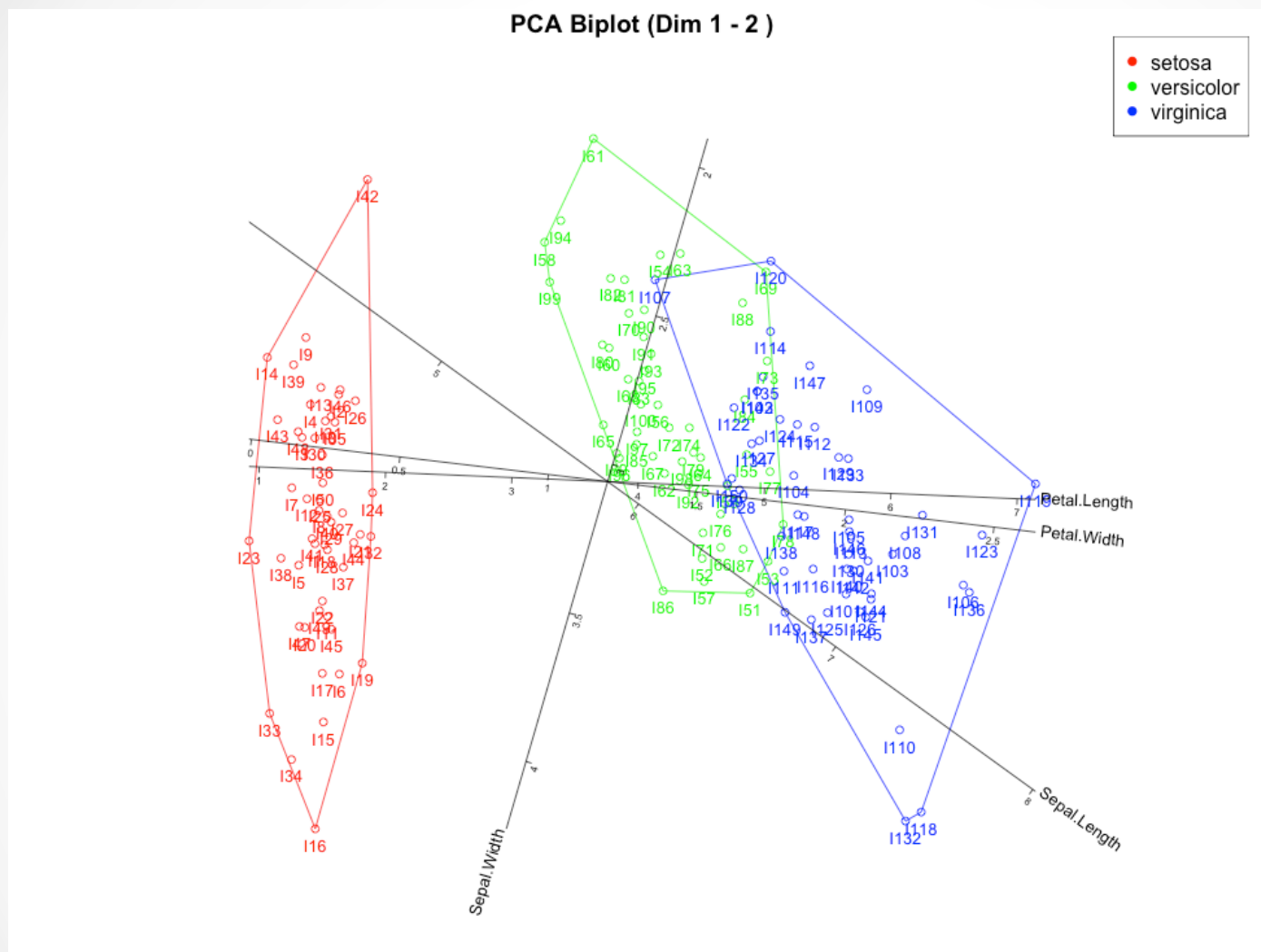


# Añadir Clusters

PCA Biplot (Dim 1 - 2 )

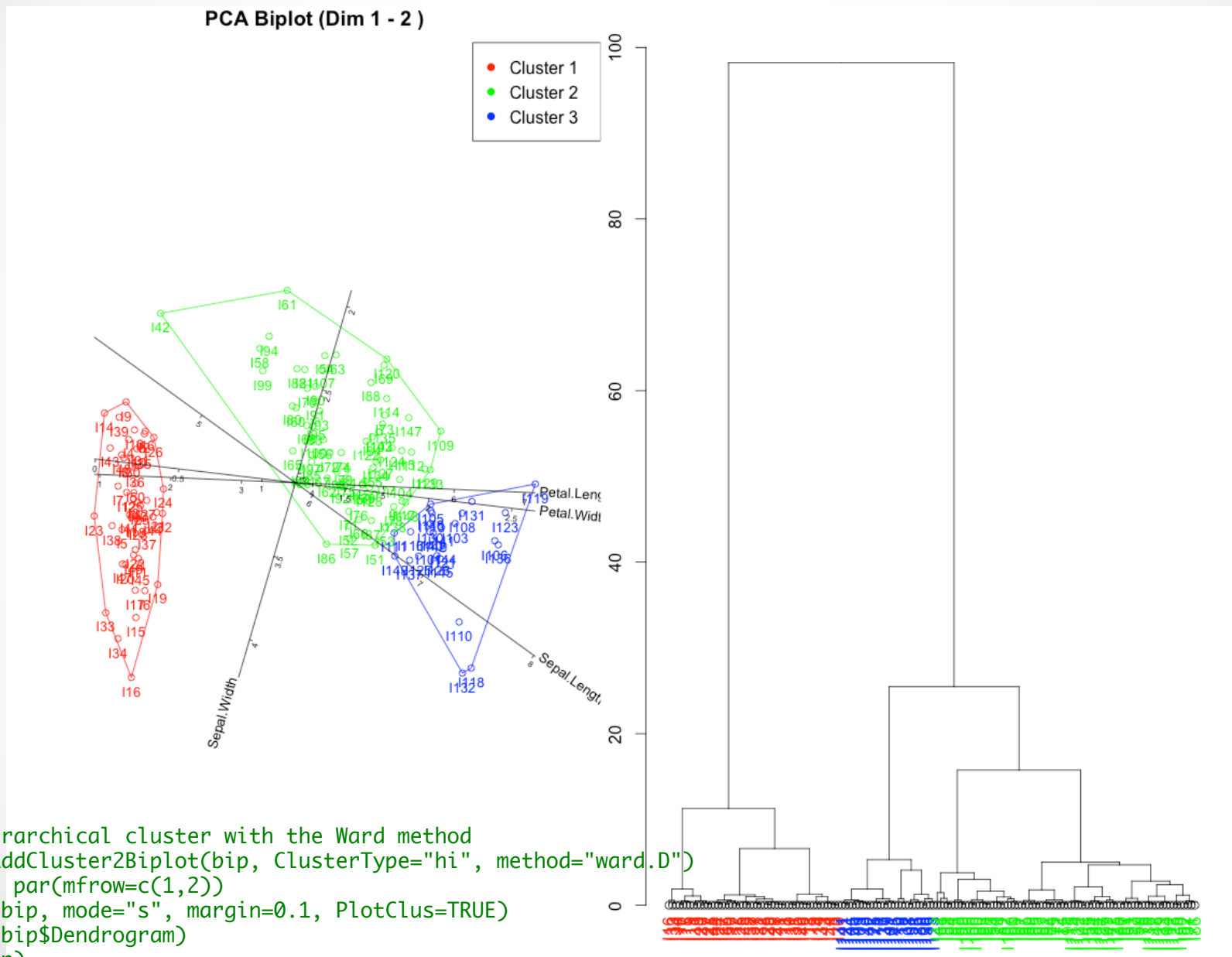


# Añadir Clusters



```
data(iris)
bip=PCA.Biplot(iris[,1:4])
bip=AddCluster2Biplot(bip, NGroups=3, ClusterType="us", Groups=iris[,5], Original=FALSE)
plot(bip, mode="s", PlotClus = TRUE, margin=0.1, TypeClus = "ch")
```

# Añadir Clusters





# Coordenadas Principales

El RMP-Biplot que vimos antes puede entenderse de otra forma. Es fácil comprobar que las coordenadas para los individuos coinciden con las coordenadas principales cuando la distancia a usar es la euclídea usual (o pitagórica en la terminología de Gower). Pero hay muchas otras distancias posibles que además se pueden aproximar mediante una distancia euclídea o que se pueden integrar en un espacio euclídeo. Vimos algunas en el tema de Coordenadas Principales.

# Coordenadas Principales

Si  $\mathbf{P}$  es una matriz de productos escalares entre cualquier conjunto de vectores (puntos) con respecto a su centro de gravedad, en cualquier espacio Euclídeo, entonces las proyecciones de los puntos en el subespacio de baja dimensión más próximo se obtienen de la estructura espectral de  $\mathbf{P}$ , como :

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda}^{1/2}$$

Donde  $\mathbf{P} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}'$  ( $\mathbf{U}' \mathbf{U} = \mathbf{I}$ ) es la descomposición de valores y vectores propios de  $\mathbf{P}$ .

La configuración obtenida por los puntos de  $\mathbf{Y}$  reproduce aproximadamente, los productos escalares de la matriz original  $\mathbf{P}$  y, por tanto, las distancias a partir de las que fueron calculados. Para la representación en dimensión reducida basta tomar las primeras columnas de  $\mathbf{Y}$ .

La variabilidad explicada vendrá dada, como es habitual, por

$$\left( \frac{\sum_{i=1}^k \lambda_i^2}{\sum_{i=1}^{n-1} \lambda_i^2} \right) \times 100\%$$

# Distancia Pitagórica

Para la distancia euclídea clásica,  $\Delta$ , la matriz de productos escalares

$$\mathbf{P} = -\frac{1}{2} \mathbf{H} \Delta^2 \mathbf{H}' = \mathbf{X}' \mathbf{X} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}'$$

Las coordenadas principales son

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda}^{1/2} = \mathbf{X} \mathbf{V}$$

Y todo ello directamente relacionado con la DVS de la matriz  $\mathbf{X}$

$$\mathbf{X} = \mathbf{U} \mathbf{\Lambda}^{1/2} \mathbf{V}'$$

Si sustituimos la matriz de distancias euclídeas por otras, tenemos las coordenadas principales de las filas o individuos, pero no las de las columnas.

# MDS

La configuración de los individuos puede obtenerse también mediante MDS.

Supongamos que tenemos un conjunto de  $n$  objetos y una forma de determinar la disimilaridad entre cada par  $\Delta = [\delta_{ij} : i, j = 1, \dots, n]$ ; el MDS trata de buscar una configuración  $\mathbf{X} = [x_{ia} : i = 1, \dots, n; a = 1, \dots, m]$  de  $n$  puntos en el espacio Euclídeo  $m$  dimensional de forma que, el punto  $\mathbf{x}_i = [x_{i1}, \dots, x_{im}]'$  represente al objeto  $i$ , y la **distancia** euclídea entre dos puntos  $\mathbf{x}_i$  y  $\mathbf{x}_j$

$$d(\mathbf{x}_i, \mathbf{x}_j) = d_{ij}(\mathbf{X}) = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

aproxime la correspondiente disimilaridad  $\delta_{ij}$  entre los objetos, ó una función de la misma

$$f(\delta_{ij}) \approx d_{ij}$$

La función  $f$  especifica el modelo de MDS.

A las medidas  $\delta_{ij}$  las denominaremos **proximidades**. Los valores  $f(\delta_{ij}) = \hat{\delta}_{ij}$  se

conocen como **disparidades**.

# Regresión Lineal

Para predecir los valores de las variables que deberíamos asociar a un punto del MDS necesitamos alguna conexión entre el espacio de los individuos **A** y el de **X**. Consideraremos la posible relación utilizando Regresiones.

A partir de la matriz **A** buscamos una matriz no necesariamente ortogonal **B** que representa la orientación óptima, de forma que se haga mínimo a  $\|\mathbf{X} - \mathbf{A} \mathbf{B}\|$ . La solución viene dada por

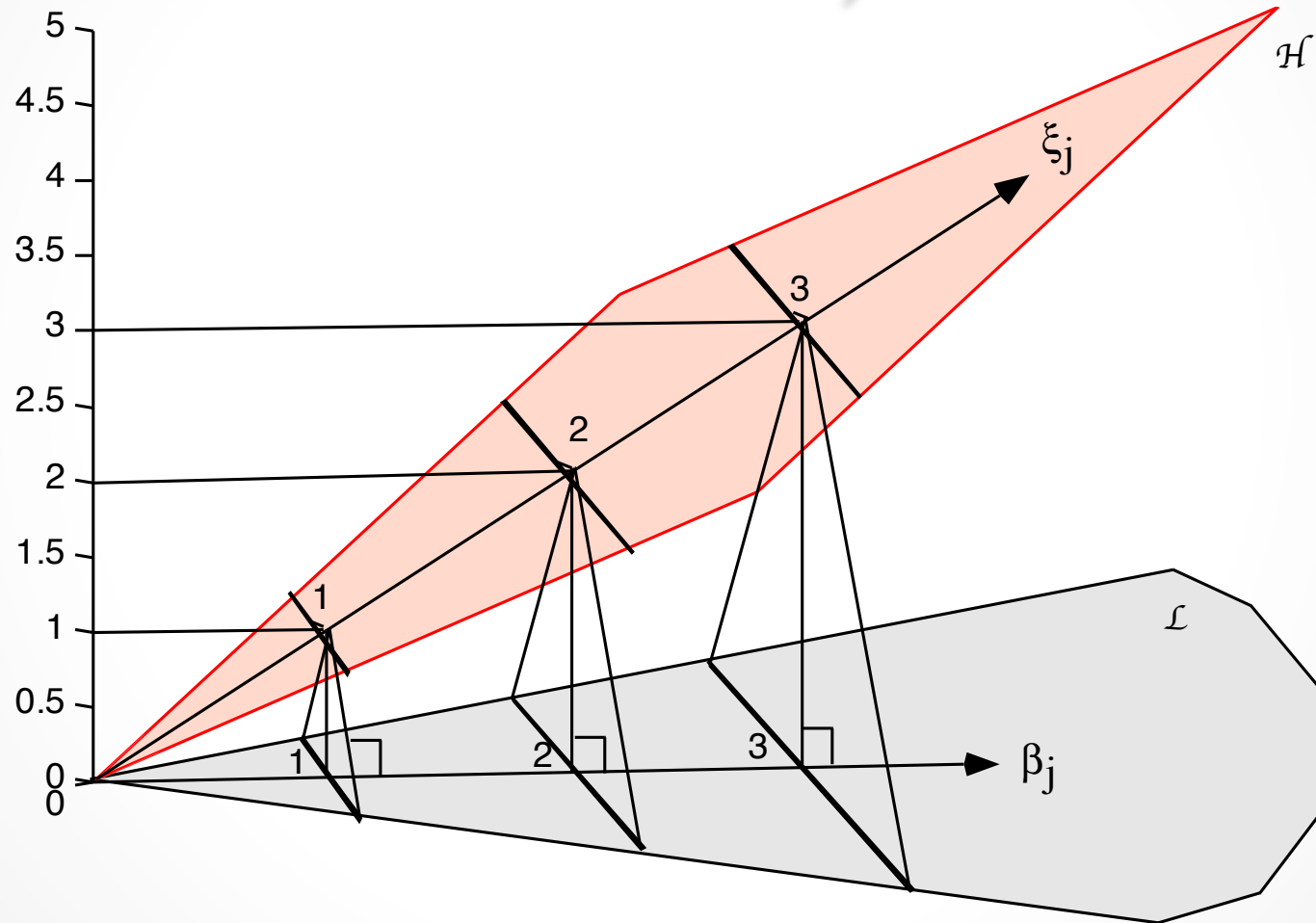
$$\mathbf{B}' = (\mathbf{A}'\mathbf{A})^{-1} \mathbf{A}' \mathbf{X}$$

Es decir, las regresiones de las variables sobre las columnas de **A**. Que se pueden separar para cada una de las columnas.

$$\mathbf{x}_{<j>} = \mathbf{A}' \boldsymbol{\beta}_j$$

Cuando usamos la distancia eucídea usual, todos los procedimientos coinciden.

# Geometría (Regresión Lineal)



# Regresión Logística

Let  $\mathbf{X}_{I \times J}$  be a binary data matrix in which the rows correspond to  $I$  individuals and the columns to  $J$  binary characters. Let  $\pi_{ij} = E(x_{ij})$  the expected probability that the character  $j$  be present at individual  $i$ . A logistic bilinear model can be written as

$$\pi_{ij} = \frac{e^{b_{j0} + \sum_{s=1}^S b_{js} a_{is}}}{1 + e^{b_{j0} + \sum_{s=1}^S b_{js} a_{is}}}$$

where  $a_{is}$  and  $b_{js}$  ( $i=1, \dots, I; j=1, \dots, J; s=1, \dots, S$ ) are the model parameters used as row and column markers respectively.

The model is a generalized bilinear model having the logit as link function.

$$\text{logit}(p_{ij}) = b_{j0} + \sum_{s=1}^S b_{js} a_{is} = b_{j0} + \mathbf{a}'_i \mathbf{b}_j$$

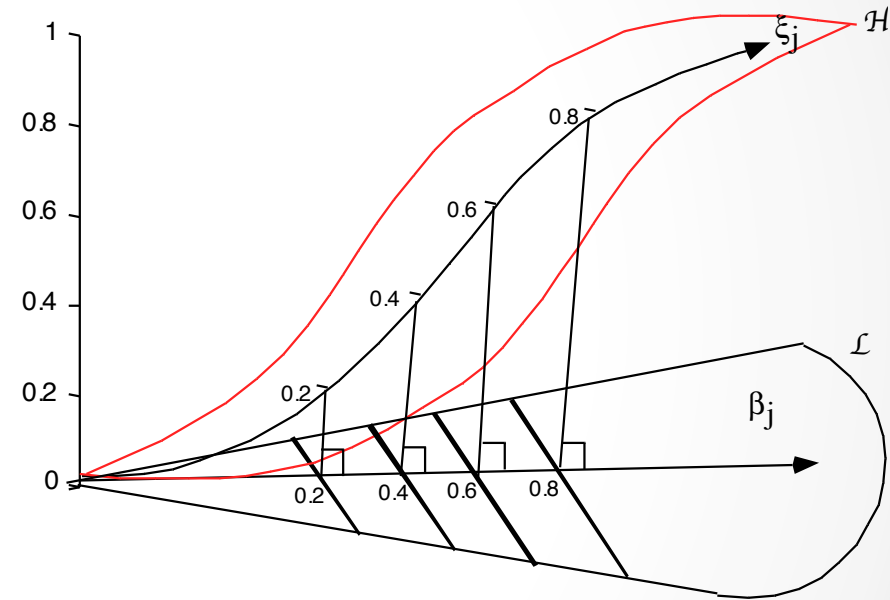
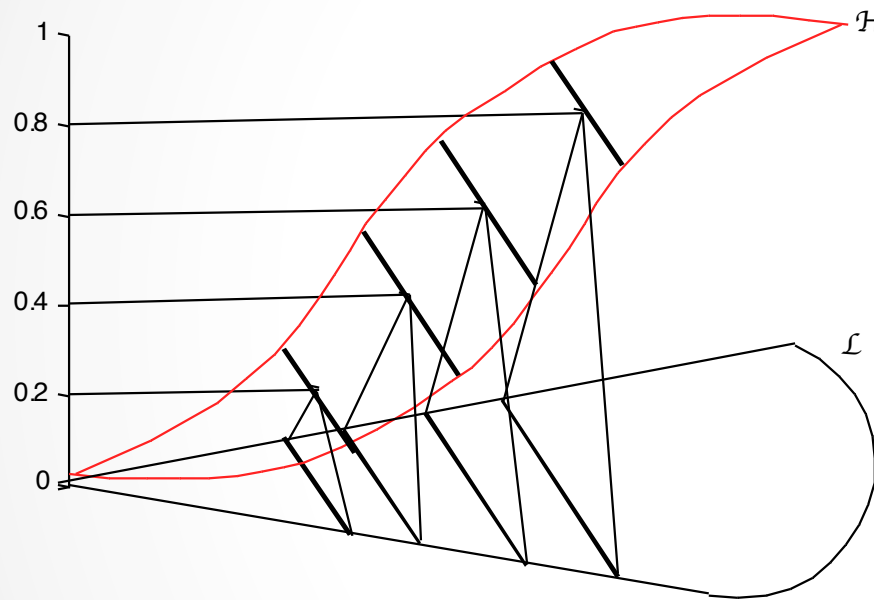
In matrix form

$$\text{logit}(\mathbf{P}) = \mathbf{1}_I \mathbf{b}'_0 + \mathbf{A} \mathbf{B}'$$

Where  $\mathbf{P}$  is the matrix of expected probabilities,  $\mathbf{1}_I$  is a vector of ones,  $\mathbf{b}_0$  is the vector containing the constants,  $\mathbf{A}$  and  $\mathbf{B}$  are the matrices containing the markers for the rows and columns  $\mathbf{X}$ .



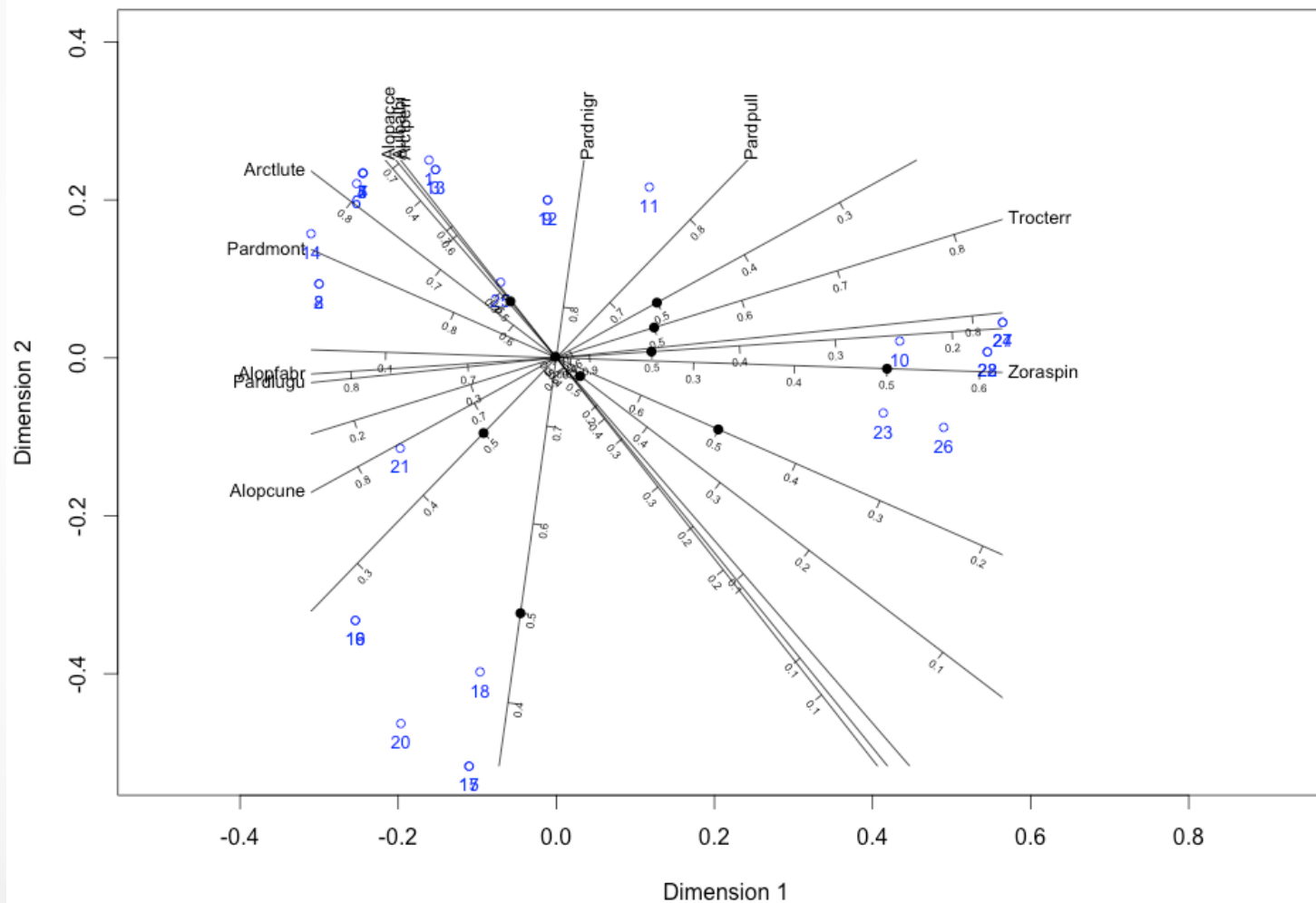
# Geometría (Regresión Logística)

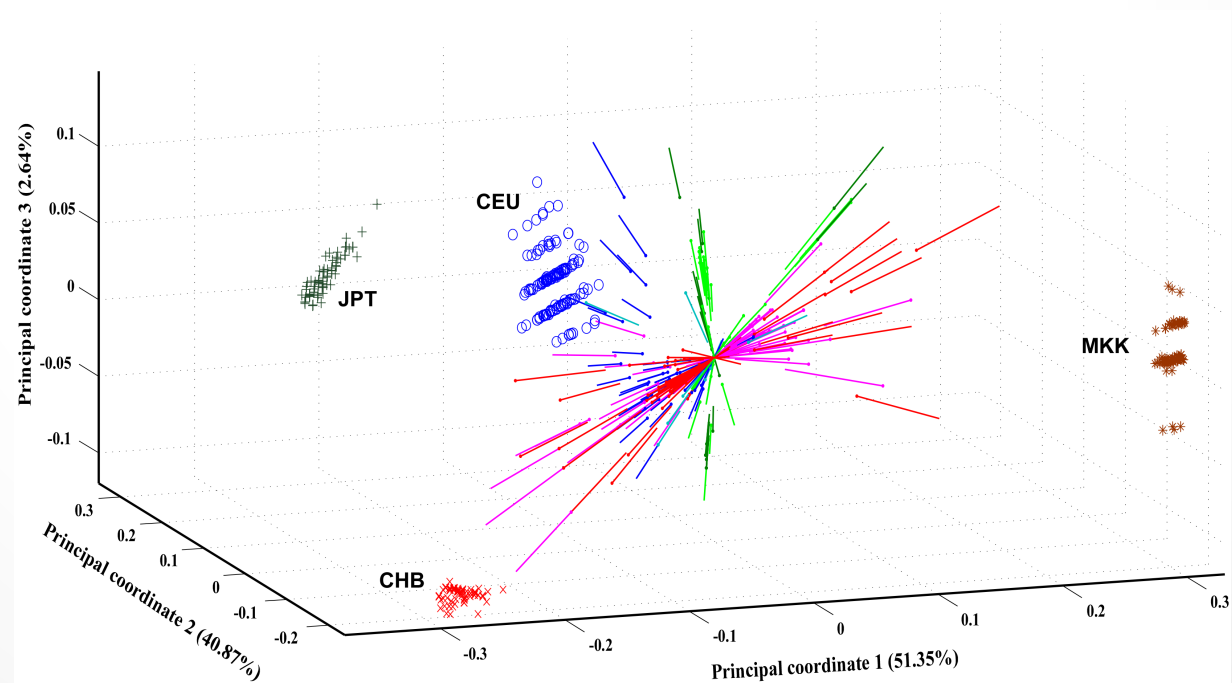
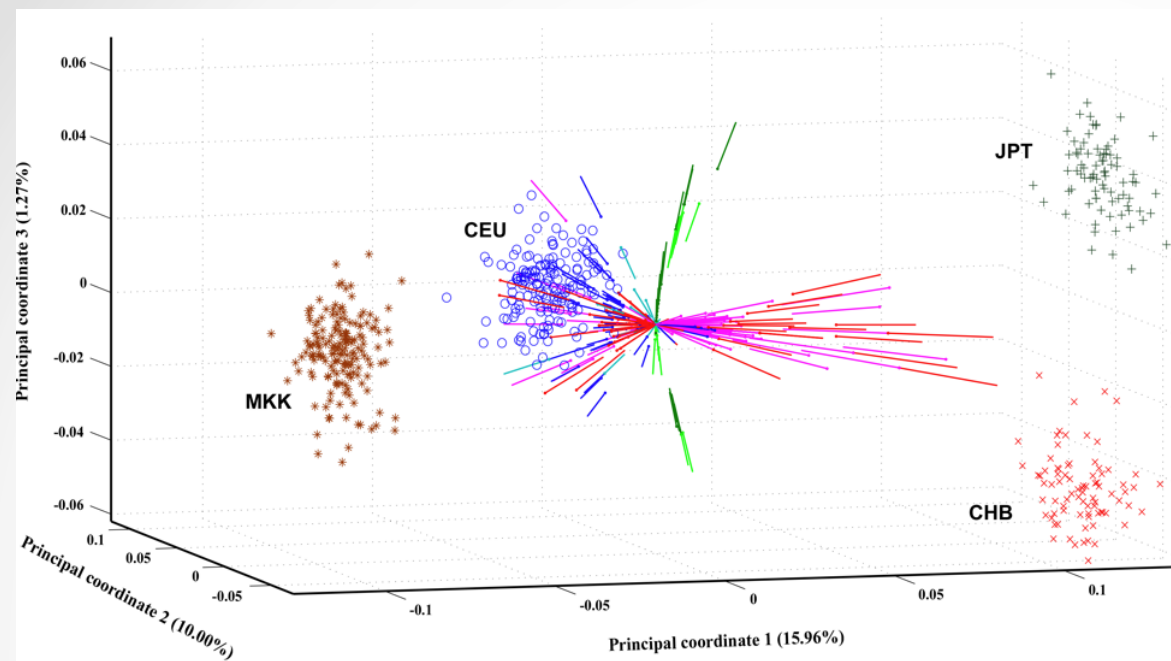


Vicente-Villardón, J. L., Galindo Villardón, M. P., & Blázquez Zaballos, A. (2006). Logistic biplots. *Multiple correspondence analysis and related methods*. London: Chapman & Hall, 503-521.

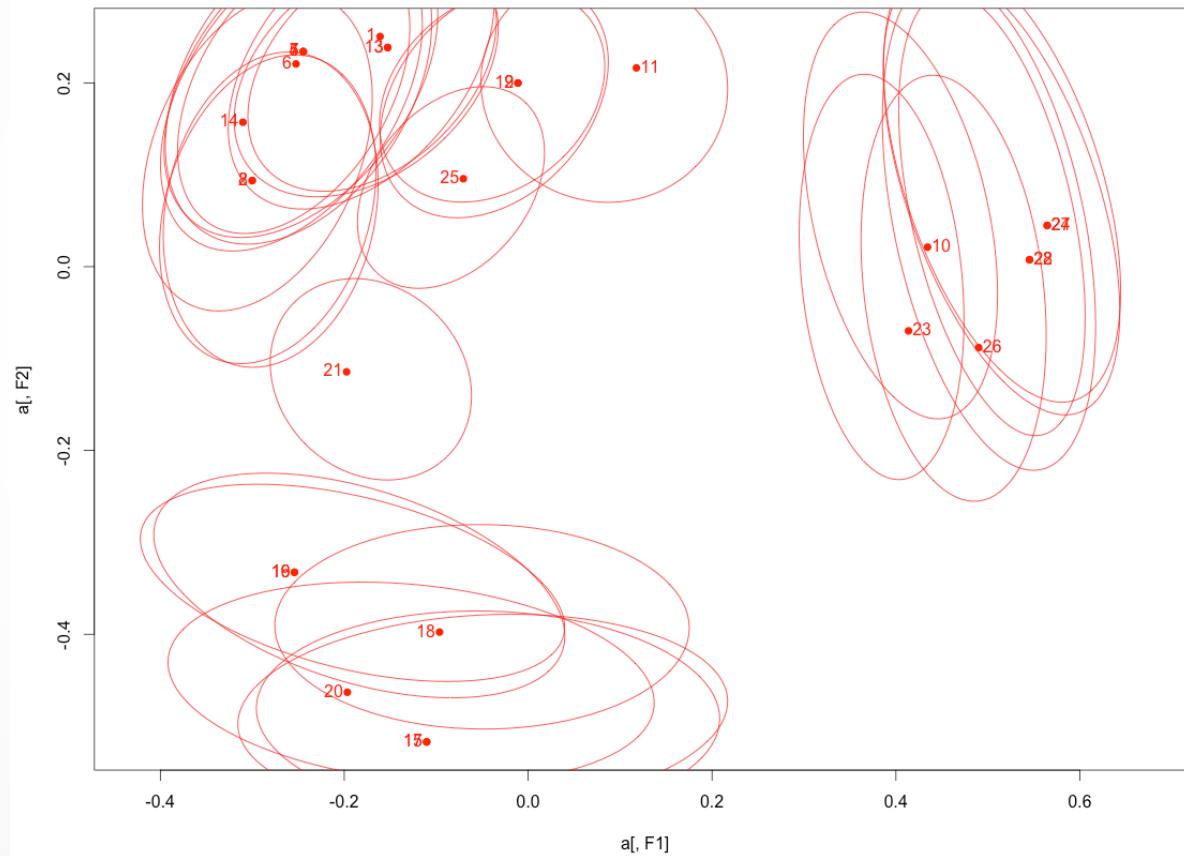
Demey, J. R., Vicente-Villardón, J. L., Galindo-Villardón, M. P., & Zambrano, A. Y. (2008). Identifying molecular markers associated with classification of genotypes by External Logistic Biplots. *Bioinformatics*, 24(24), 2832-2838.

# Biplot Logístico

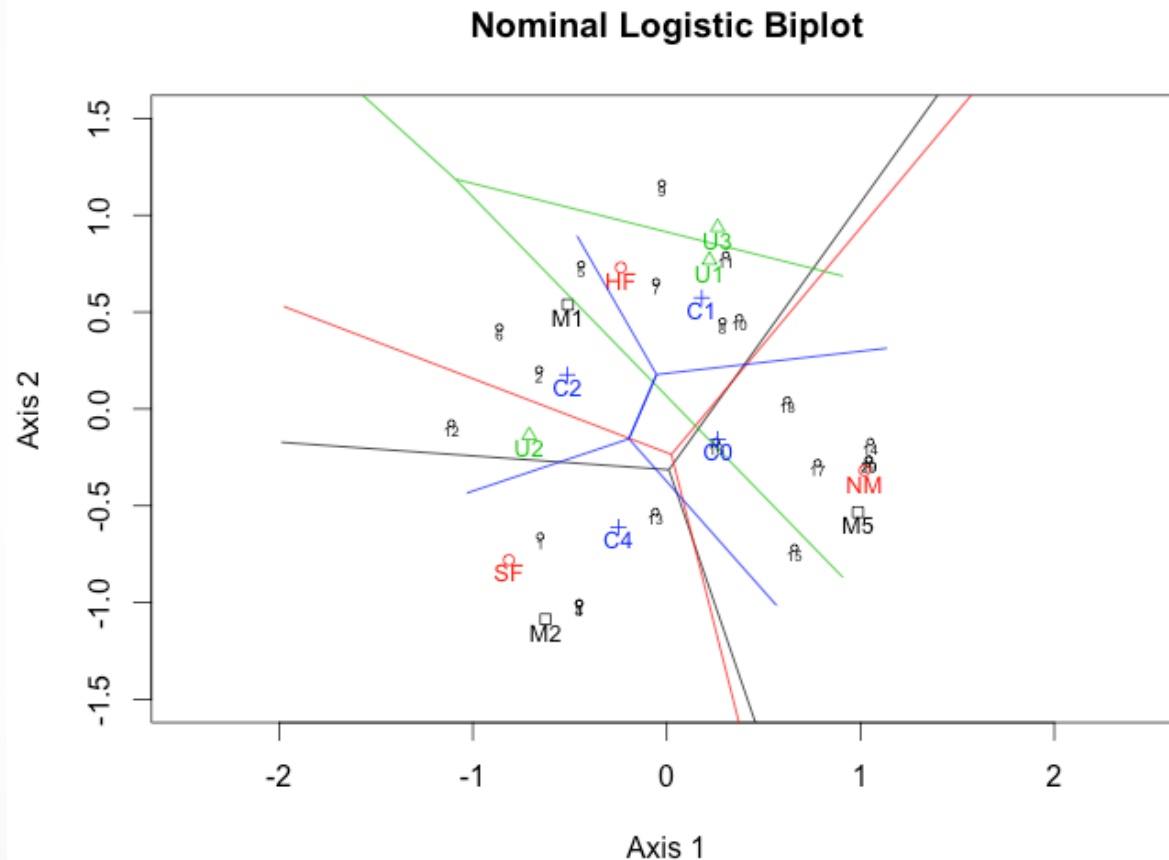




# ACoP con Bootstrap

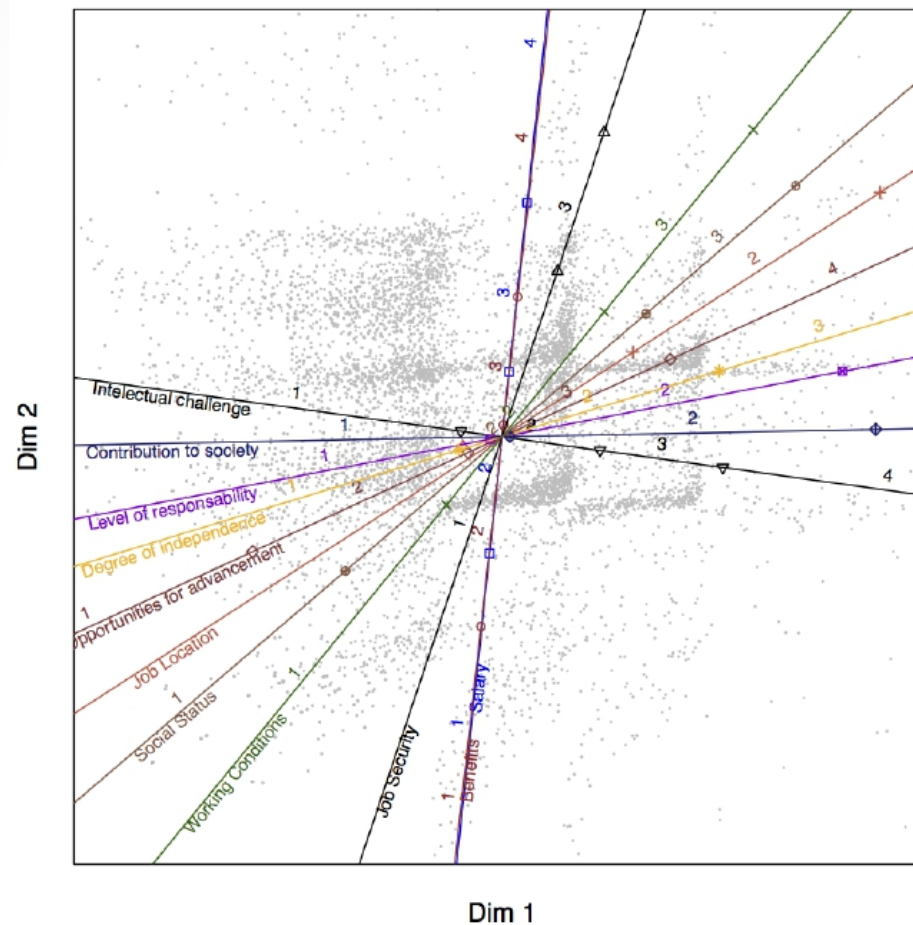


# Datos Nominales



Hernández Sánchez, J. C., & Vicente-Villardón, J. L. (2013).  
Logistic biplot for nominal data. *arXiv preprint arXiv:1309.5486*.  
• R package: “NominalLogisticBiplots”

# Datos Ordinales



Vicente-Villardón, J. L., & Sánchez, J. C. H. (2014). Logistic Biplots for Ordinal Data with an Application to Job Satisfaction of Doctorate Degree Holders in Spain. *arXiv reprint arXiv:1405.0294*.

R package: “OrdinalLogisticBiplots”

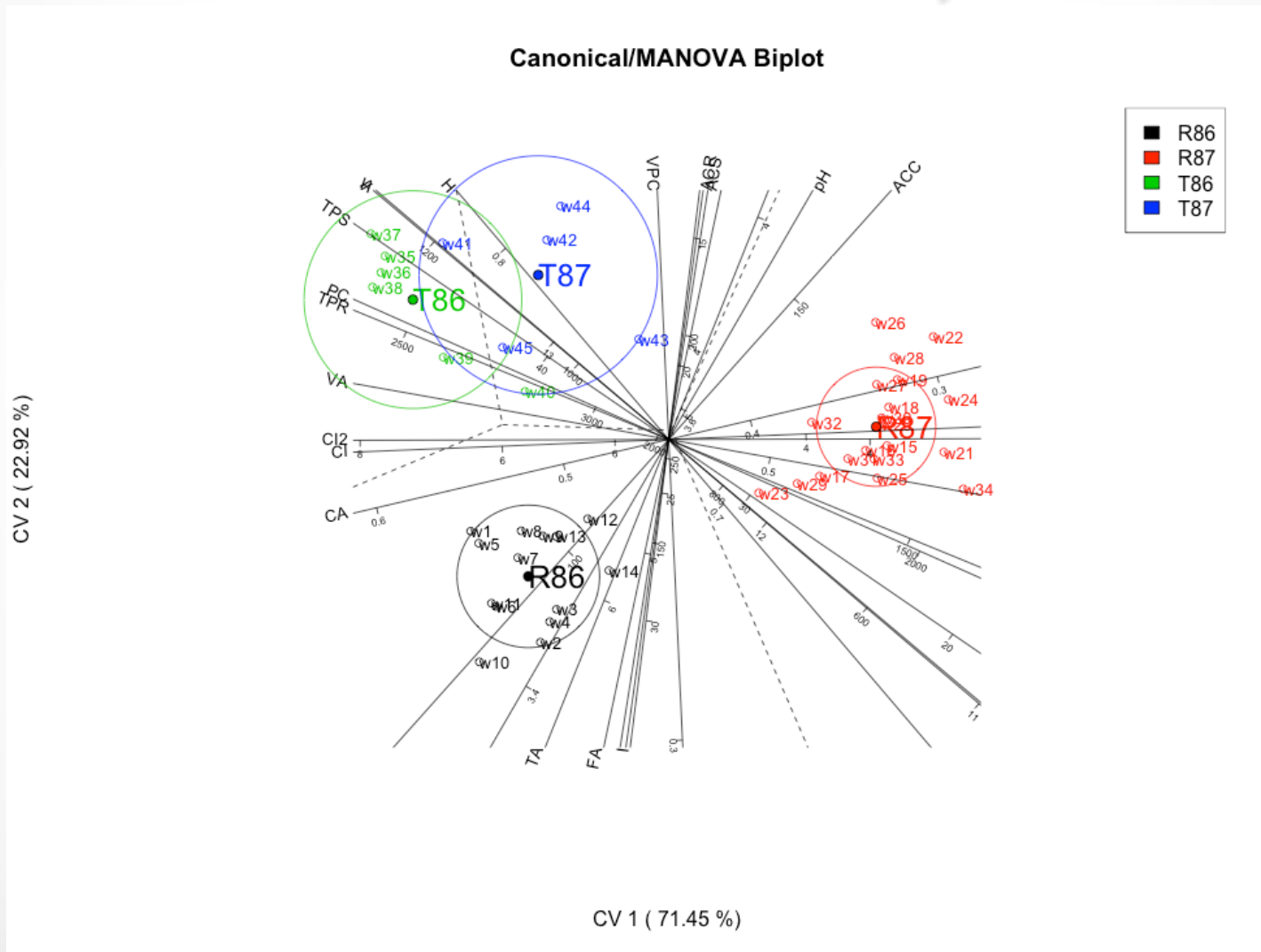
# El Biplot Canónico

## o MANOVA Biplot

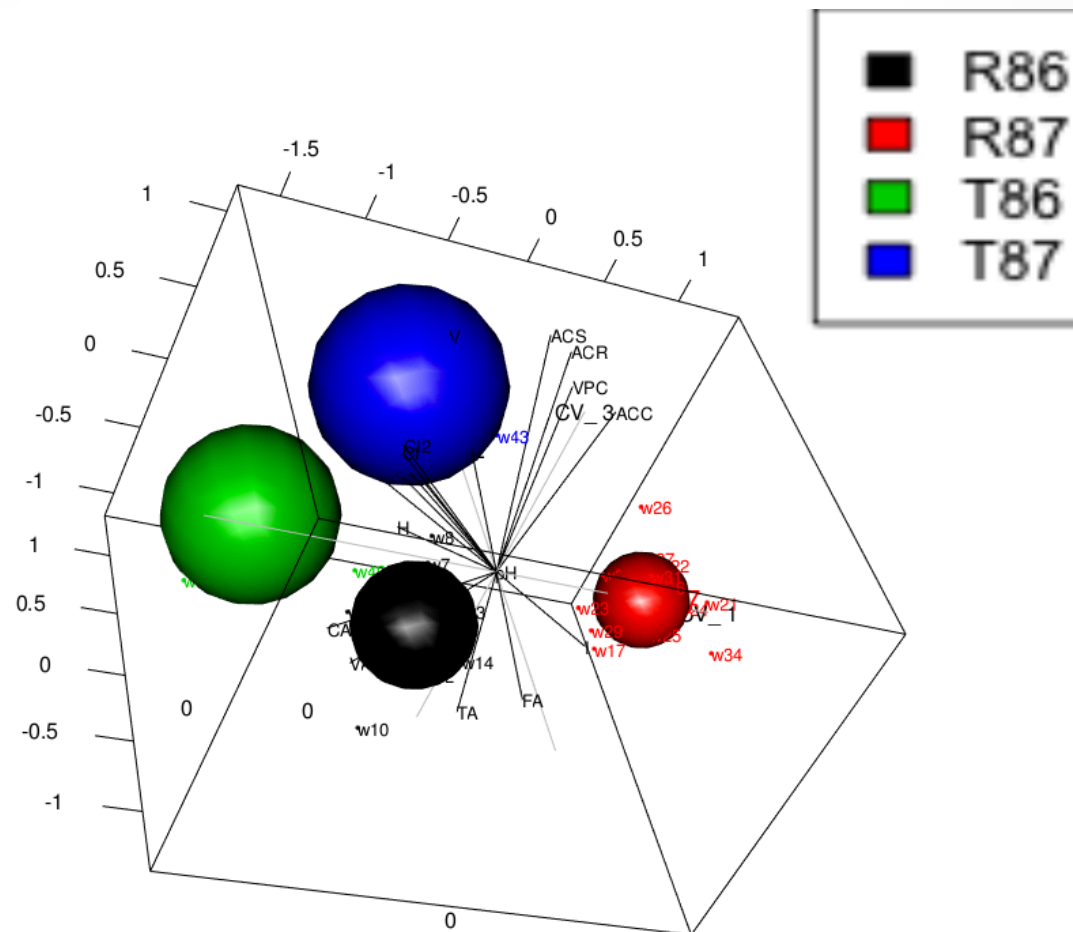
Let  $\mathbf{X}$  be a data matrix of sample sites by variables whose rows are divided into  $K$  groups. A Canonical Biplot for the data is obtained from a factorization  $\bar{\mathbf{X}} \cong \mathbf{GH}'$  of the matrix  $\bar{\mathbf{X}}$  containing the group means (rivers) for each the chemical composition variables. In order to relate the biplot factorization to MANOVA or Canonical Variate Analysis (CVA) we use the Singular Value Decomposition (SVD),  $\mathbf{Y} = \mathbf{UD}_\lambda \mathbf{V}'$  of the matrix  $\mathbf{Y} = \mathbf{D}_n^{1/2} \bar{\mathbf{X}} \mathbf{W}^{-1/2}$  containing the weighted group means, where  $\mathbf{D}_n$  is the diagonal matrix with the group sample sizes,  $\mathbf{W} = \frac{1}{n-k}(\mathbf{X}'\mathbf{X} - \bar{\mathbf{X}}'\mathbf{D}_n\bar{\mathbf{X}})$  is the “within-groups” and  $\mathbf{B} = \frac{1}{k-1}\bar{\mathbf{X}}'\mathbf{D}_n\bar{\mathbf{X}}$  the “between-groups” covariance matrices. The decomposition in  $\bar{\mathbf{X}} = \mathbf{D}_n^{-1/2}\mathbf{UD}_\lambda\mathbf{V}'\mathbf{W}^{1/2}$  permits the construction of a Biplot (in dimension  $r$ ) for the group means matrix ( $\bar{\mathbf{X}} \cong \mathbf{GH}'$ ), in which the first  $r$  columns of  $\mathbf{G} = \mathbf{D}_n^{-1/2}\mathbf{UD}_\lambda$  are row markers and the first  $r$  columns of  $\mathbf{H} = \mathbf{W}^{1/2}\mathbf{V}$  are column markers (Amaro et al., 2004). Alternative descriptions can be found in Vicente-Villardón (1992), Gower and Hand (1995) or Gower et al. (2011). Together with the biplot representation of the means, the rows of  $\mathbf{X}$  can also be projected on the biplot in order to better interpret the magnitude of the differences among groups.



# Canonical Biplot (Coordenadas discriminantes)



# 3D Canonical Biplot



# Bondad del ajuste

El objeto del biplot es predecir los valores originales de las variables en la representación en dimensión reducida. Si calculamos los valores esperados

$$\mathbf{X} = \mathbf{AB}' + \mathbf{E} = \hat{\mathbf{X}} + \mathbf{E} \quad (\hat{\mathbf{X}} = \mathbf{AB}')$$

La bondad del ajuste global será la parte de la variabilidad de los datos explicada por la predicción

$$\rho = \text{traza}(\hat{\mathbf{X}}'\hat{\mathbf{X}}) / \text{traza}(\mathbf{X}'\mathbf{X})$$

Que, para los biplots clásicos coincide con la calculada a partir de los valores propios. Si calculamos por columnas

$$\rho_j = \text{diag}(\hat{\mathbf{X}}'\hat{\mathbf{X}}) / \text{diag}(\mathbf{X}'\mathbf{X})^{-1}$$

Tenemos la bondad de ajuste para cada columna que nosotros llamamos calidad de representación y Gower predictividad.

Si calculamos por filas

$$\rho_j = \text{diag}(\hat{\mathbf{X}}\hat{\mathbf{X}}') / \text{diag}(\mathbf{X}\mathbf{X}')^{-1}$$

Tenemos la bondad de ajuste para cada fila que nosotros llamamos calidad de representación y Gower predictividad.

# MultBiplot (R)

- BILOT CLÁSICO
  - COMPONENTES PRINCIPALES
  - ANÁLISIS FACTORIAL
  - HJ-BILOT
- ANÁLISIS DE CORRESPONDENCIAS
  - SIMPLE
  - MULTIPLE
  - NO SIMÉTRICO
- ANÁLISIS DE COORDENADAS PRINCIPALES Y MDS (con biplot externo)
  - DATOS CONTINUOS, BINARIOS, NOMINALES ORDINALES
  - DATOS MIXTOS
  - UNFOLDING (PREFSCAL)
- BILOT CANÓNICO/MANOVA BILOT
  - UNA VÍA
  - DOS VÍAS
  - GENERAL

# MultBiplot (R)

- BILOT LOGISTICO
  - BINARIO
  - NOMINAL
  - ORDINAL
  - DATOS MEZCLADOS
- ORDENACIÓN DE DOS TABLAS
  - ANÁLISIS DE LA REDUNDANCIA
  - ANÁLISIS CANÓNICO DE CORRESPONDENCIAS
  - COINERCA
  - CCA NO SIMÉTRICO
  - CORRELACIÓN CANÓNICA
  - BILOT LOGÍSTICO RESTRINGIDO
  - UNFOLDING RESTRINGIDO
  - MINIMOS CUADRADOS PARCIALES

# MultBiplot (R)

- TABLAS MULTIPLES
  - BILOT CONSENSO (CON VARIABLES COMUNES)
  - META-BILOT
  - STATIS-ACT
  - DISTATIS
  - ANÁLISIS FACTORIAL MÚLTIPLE
  - ANÁLISIS FACTORIAL SIMULTANEO
  - DOBLE ANÁLISIS DE COMPONENTES PRINCIPALES
  - X-STATIS (ANÁLISIS TRIÁDICO PARCIAL)
  - PROCRUSTES
- BILOT PARA TEORIA DE LA RESPUESTA AL ITEM
- DATOS DE TRES TRES-VIAS
  - PARAFAC
  - TUKALS2
  - TUKALS3
- OTROS DESARROLLOS FUTUROS
-

# MUCHAS GRACIAS

## VII JORNADAS DE USUARIOS DE R

Salamanca, 5 y 6 de noviembre de 2015

