# Análisis de Correspondencias Simples y Múltiples


**Dpto. de Estadística e Investigación Operativa**

**Universidad de Valladolid**

**Roberto San Martín Fernández**

# Análisis Exploratorio de Datos Multidimensionales

## I. Métodos Factoriales
## II. Métodos de Clasificación

## I. Métodos Factoriales

1. Análisis de Componentes Principales (ACP)
   Normado o Sin Normar (Regresión Ortogonal)

2. Análisis de Correspondencias (AC)
   Simples (ACS) o Múltiples (ACM)

3. Análisis de Discriminante

4. Etc.

# Análisis Exploratorio de Datos Multidimensionales

## I. Métodos Factoriales
## II. Métodos de Clasificación

## II. Métodos de Clasificación
## (**Análisis Cluster**)

1. Métodos Jerárquicos
   - Distancia (euclidea)
   - Criterio de Agregación (Ward)

2. Métodos No Jerárquicos
   - *k* – medias

# Análisis de Correspondencias (AC)

- **Análisis de Datos Categóricos**

- **Análisis de Correspondencias Simples (ACS)**
  - Dos Variables Categóricas
  - Análisis de Tablas de Contingencia (grandes)

- **Análisis de Correspondencias Múltiples (ACM)**
  - Más de dos Variables Categóricas

- **Utilización.**
  - Por sí solos
  - Junto a otros análisis (loglineales, logísticos, etc.)

# Análisis de Tablas de Contingencia

| | | Variable Columna | | | |
|---|---|---|---|---|---|
| | **Categoría** | **A** | **B** | **C** | **D** |
| **Variable Fila** | **1** | $n_{11}$ | $n_{12}$ | $n_{13}$ | $n_{14}$ |
| | **2** | $n_{21}$ | $n_{22}$ | $n_{23}$ | $n_{24}$ |
| | **3** | $n_{31}$ | $n_{32}$ | $n_{33}$ | $n_{34}$ |

**$n_{ij}$ = nº de individuos en las categorías "*i*" de la Var. Fila y "*j* " de la Var. Columna.**

# Análisis de Tablas de Contingencia

| | | Variable Columna | | | | |
|---|---|---|---|---|---|---|
| | Categoría | A | B | C | D | Total Fila |
| Variable Fila | 1 | $n_{11}$ | $n_{12}$ | $n_{13}$ | $n_{14}$ | $n_{1.}$ |
| | 2 | $n_{21}$ | $n_{22}$ | $n_{23}$ | $n_{24}$ | $n_{2.}$ |
| | 3 | $n_{31}$ | $n_{32}$ | $n_{33}$ | $n_{34}$ | $n_{3.}$ |

**$n_{i.}$ = nº de individuos en la categoría "*i*" de la Var. Fila**

$$n_{i.} = \sum_{j=1}^{k} n_{ij}$$

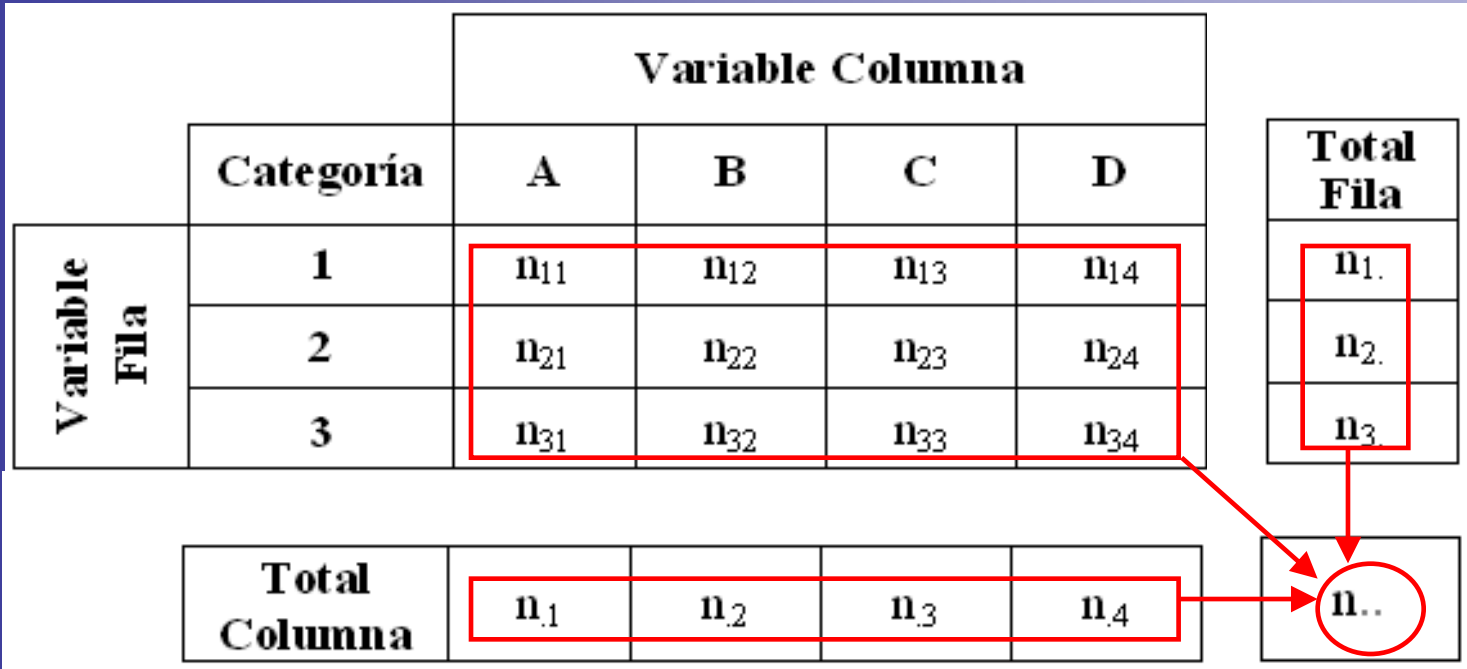
# Análisis de Tablas de Contingencia

| | | Variable Columna | | | | |
|---|---|---|---|---|---|---|
| | Categoría | A | B | C | D | Total Fila |
| Variable Fila | 1 | $n_{11}$ | $n_{12}$ | $n_{13}$ | $n_{14}$ | $n_{1.}$ |
| | 2 | $n_{21}$ | $n_{22}$ | $n_{23}$ | $n_{24}$ | $n_{2.}$ |
| | 3 | $n_{31}$ | $n_{32}$ | $n_{33}$ | $n_{34}$ | $n_{3.}$ |
| | Total Columna | $n_{.1}$ | $n_{.2}$ | $n_{.3}$ | $n_{.4}$ | |

$n_{.j}$ = nº de individuos en la categoría "*j*" de la Var. Colum.

$$n_{.j} = \sum_{i=1}^{n} n_{ij}$$

# Análisis de Tablas de Contingencia




***n..*** **= nº Total Individuos**

$$n.. = \sum_{i=1}^{n}\sum_{j=1}^{k} n_{ij} = \sum_{i=1}^{n} n_{i.} = \sum_{j=1}^{k} n_{.j}$$

# Análisis de Tablas de Contingencia

| | | Variable Columna | | | | |
|---|---|---|---|---|---|---|
| | Categoría | A | B | C | D | Total Fila |
| Variable Fila | 1 | $n_{11}$ | $n_{12}$ | $n_{13}$ | $n_{14}$ | $n_{1.}$ |
| | 2 | $n_{21}$ | $n_{22}$ | $n_{23}$ | $n_{24}$ | $n_{2.}$ |
| | 3 | $n_{31}$ | $n_{32}$ | $n_{33}$ | $n_{34}$ | $n_{3.}$ |
| | Total Columna | $n_{.1}$ | $n_{.2}$ | $n_{.3}$ | $n_{.4}$ | $n_{..}$ |

**Objetivo:** Estudio de **"Asociaciones"** entre las categorías de las variables.

**Asociación & Independencia**

# Ejemplo 1

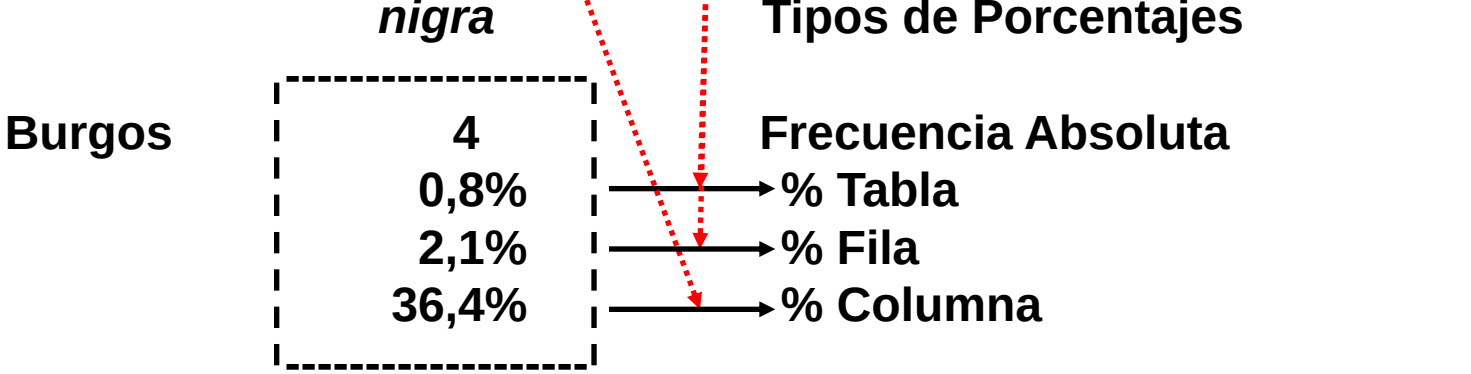
**Tabla de Frecuencias para 525 *pinus* según Provincia y Especie**

| | *nigra* | *pinaster* | *pinea* | *sylvestris* | Fila Total |
|---|---|---|---|---|---|
| **Burgos** | **4** | **135** | **11** | **43** | **193** |
| **Soria** | **7** | **100** | **35** | **190** | **332** |
| **Columna Total** | **11** | **235** | **46** | **233** | **525** |

- Estudio asociaciones: **Provincia ←→ Especie**
- ¿Cómo? **Utilización de Porcentajes.**
- **¿Qué porcentajes? Tipos de porcentajes.**

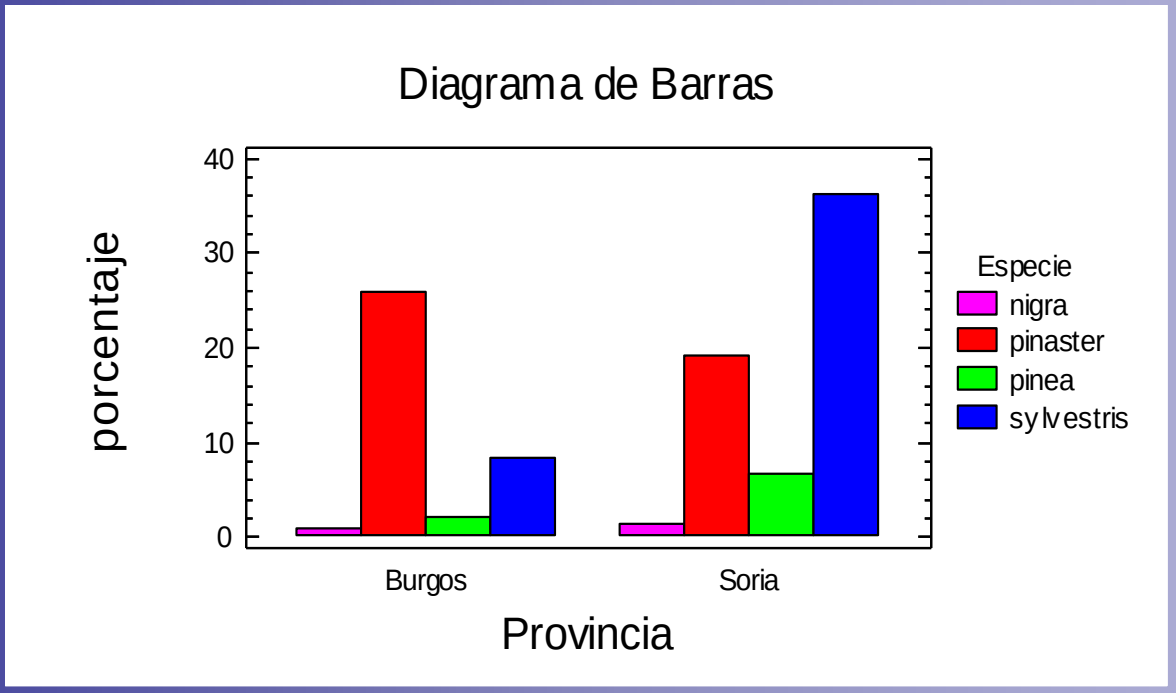
## Tabla de Frecuencias

| | *nigra* | *pinaster* | *pinea* | *sylvestris* | Fila Total | |
|---|---|---|---|---|---|---|
| Burgos | 4 | 135 | 11 | 43 | 193 | |
| Soria | 7 | 100 | 35 | 190 | 332 | |
| Columna Total | 11 | 235 | 46 | 233 | 525 | Total Tabla |



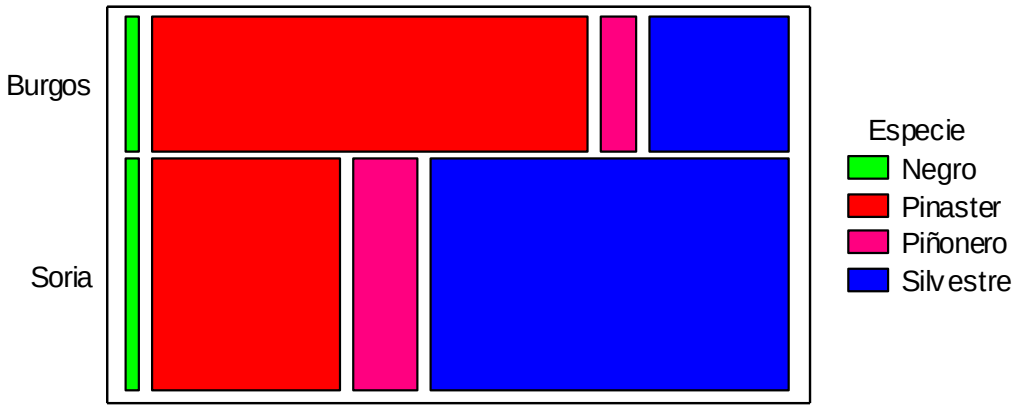
## Porcentajes Tabla

| | *nigra* | *pinaster* | *pinea* | *sylvestris* | Fila Total |
|---|---|---|---|---|---|
| Burgos | 0,8% | 25,7% | 2,1% | 8,2% | 36,8% |
| Soria | 1,3% | 19,1% | 6,7% | 36,2% | 63,2% |
| Columna Total | 2,1% | 44,8% | 8,8% | 44,4% | 100% |



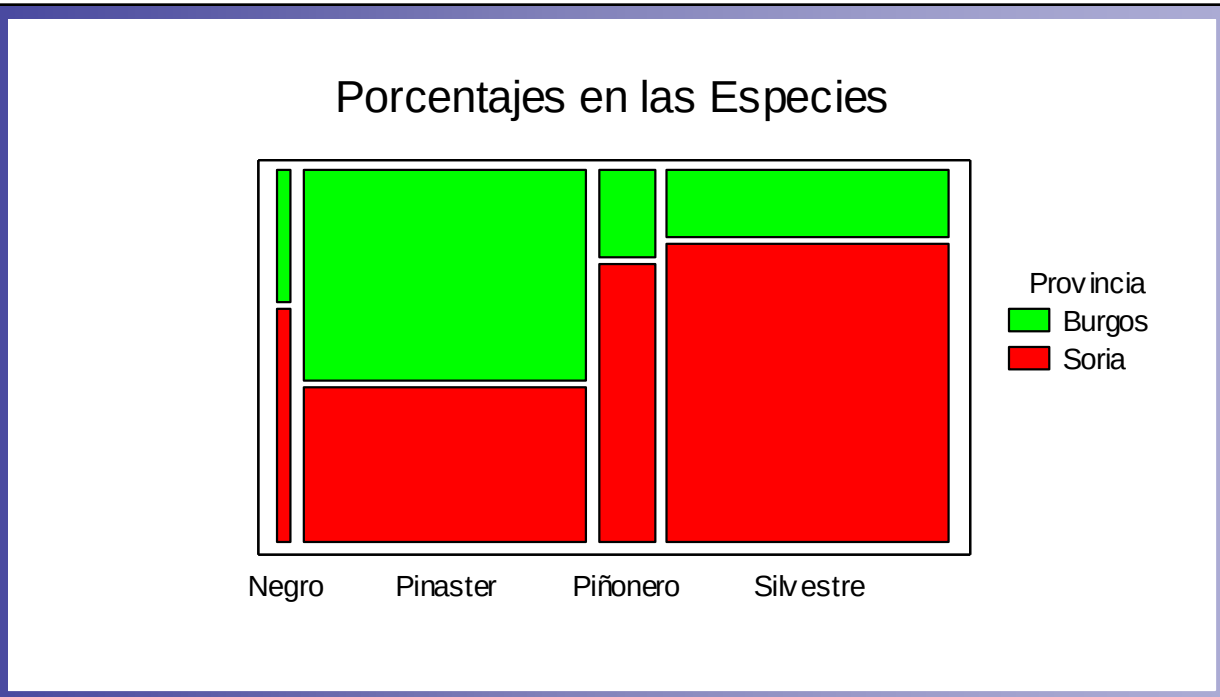
## Porcentajes por Provincia (Fila)

| | *nigra* | *pinaster* | *pinea* | *sylvestris* | Fila Total |
|---|---|---|---|---|---|
| Burgos | 2,1% | 69,9% | 5,7% | 22,3% | 100% |
| Soria | 2,1% | 30,1% | 10,6% | 57,2% | 100% |
| Columna Total | 2,1% | 44,8% | 8,8% | 44,4% | 100% |



## Porcentajes por Especie (Columna)

| | *nigra* | *pinaster* | *pinea* | *sylvestris* | Fila Total |
|---|---|---|---|---|---|
| Burgos | 36,4% | 57,4% | 23,9% | 18,4% | 36,8% |
| Soria | 63,6% | 42,5% | 76,1% | 81,5% | 63,2% |
| Columna Total | 100% | 100% | 100% | 100% | 100% |




Volver

# Conclusiones

## Asociaciones

| Provincia | | Especie |
|---|---|---|
| Burgos | ↔ | *pinaster* |
| Soria | ↔ | *sylvestris* |
| | | *pinea* |

# Test de Independencia - Test Chi-2

$$\begin{cases} H_0 : Independencia \\ H_1 : Asociación \end{cases}$$

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{k} \frac{\left( n_{ij} - \frac{n_{i.} n_{.j}}{n_{..}} \right)^2}{\frac{n_{i.} n_{.j}}{n_{..}}}$$

$$\chi^2 \xrightarrow{H_0} \chi^2_{(n-1)\cdot(k-1)}$$

$$p_{valor} = p(\chi^2_{(n-1)\cdot(k-1)} > \chi^2)$$

# Test de Independencia - Test Chi-2

$$\begin{cases} H_0 : Independencia \\ H_1 : Asociación \end{cases}$$

**Ejemplo 1**

**Contraste de Chi-cuadrado**

| Chi-cuadrado | GL | P-Valor |
|---|---|---|
| **80,11** | **3** | **0,0000** |

El StatAdvisor

Dado que el **p-valor es inferior a 0.01**, podemos **rechazar** la hipótesis de que **las filas y columnas son independientes** con un nivel de confianza del **99%.** En consecuencia, **el valor observado de Provincia para un caso particular tiene relación con su valor en Especie**.

# Ejemplo 2: Caso de Independencia

Porcentajes en las Provincias

Porcentajes en las Especies

**Contraste de Chi-cuadrado**

| Chi-cuadrado | GL | P-Valor |
|---|---|---|
| 0,02 | 3 | 0,9992 |

El StatAdvisor

---

Dado que el **p-valor es superior a 0.10, no podemos rechazar** la hipótesis de **que las filas y columnas son independientes.** En consecuencia, **el valor observado de Provincia** para un caso particular **puede no tener relación con su valor en Especie.**

# !!!!!!! Importante !!!!!!!

**Porcentajes por Provincia**

| | *nigra* | *pinaster* | *pinea* | *sylvestris* | Fila Total |
|---|---|---|---|---|---|
| Burgos | 2,1% | 69,9% | 5,7% | 22,3% | 100% |
| Soria | 2,1% | 30,1% | 10,6% | 57,2% | 100% |
| Columna Total | 2,1% | 44,8% | 8,8% | 44,4% | 100% |

**PERFIL de Soria**

**PERFIL MEDIO**

**PERFIL de Burgos**

# ¡¡¡¡¡¡ PERFILES FILA !!!!!!

# ¡¡¡¡¡¡ PERFILES COLUMNA !!!!!!

**Porcentajes por Especie**

| | *nigra* | *pinaster* | *pinea* | *sylvestris* | Fila Total |
|---|---|---|---|---|---|
| Burgos | 36,4% | 57,4% | 23,9% | 18,4% | 36,8% |
| Soria | 63,6% | 42,5% | 76,1% | 81,5% | 63,2% |
| Columna Total | 100% | 100% | 100% | 100% | 100% |

**PERFIL de *nigra***

**PERFIL de *pinaster***

**PERFIL de *pinea***

**PERFIL de *sylvestris***

**PERFIL MEDIO**

# Análisis de Correspondencias (AC)

## Análisis de Correspondencias Simples(ACS)

- **Dos** Variables Categóricas

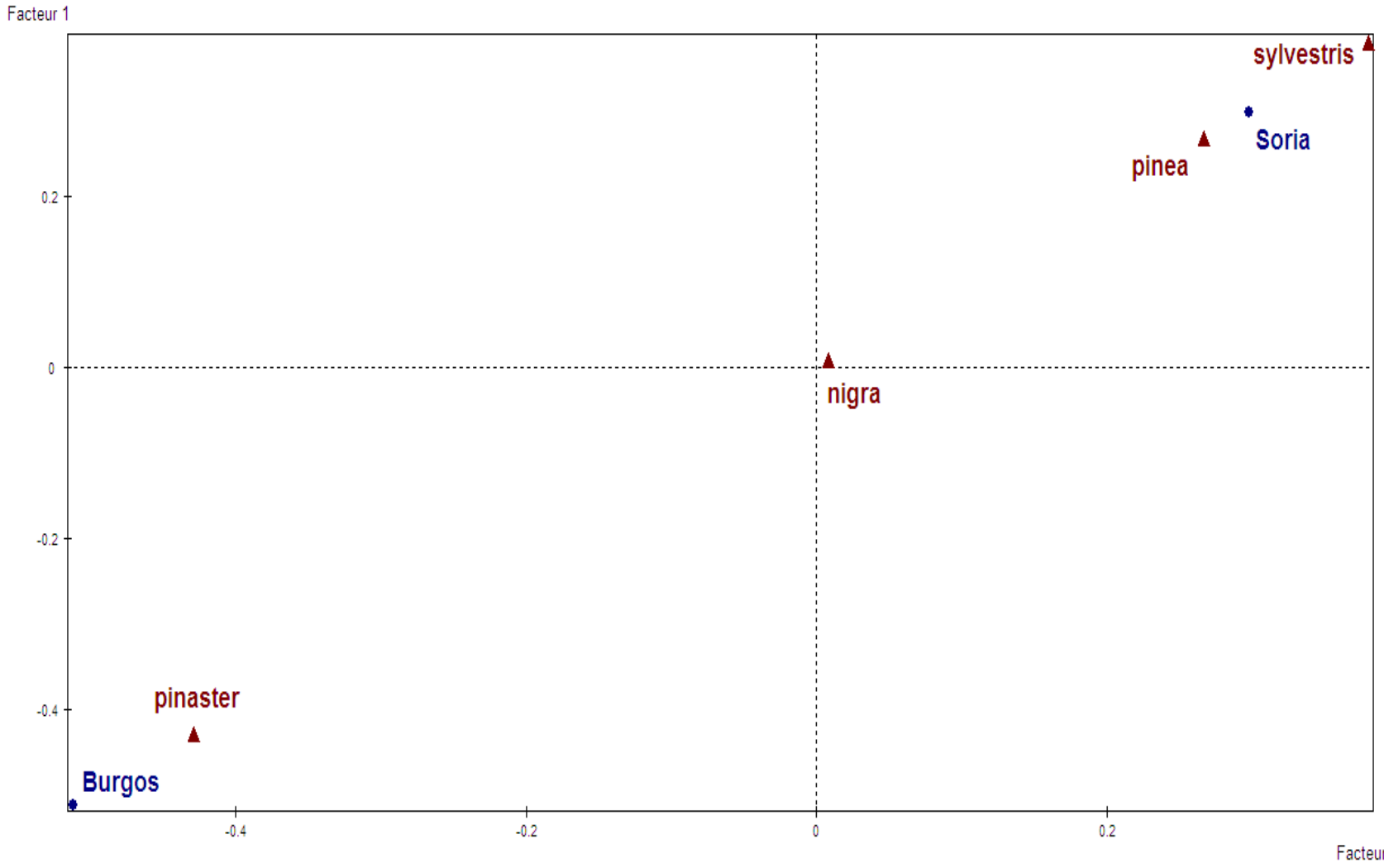
- Análisis de **Tablas de Contingencia** (grandes)

- **Análisis de los Perfiles Fila y Columna**
  **ACP** (Principal Components Analysis)
  Distancia ***Chi-2***

- Representación en " **bi-plots** " de los Perfiles.

- **Superposición** de los bi-plots

# ACS para Provincia & Especie




Perfil Especie

# Reglas de Interpretación

**1.** Los **puntos** del bi-plot **= Perfiles** de las variables.

**2. Origen de Coordenadas = Perfil Medio.**

**3. Dos Perfiles de una misma variable:**

**3.1 Proximidad ↔ Igualdad**

**3.2 Lejanía ↔ Diferencia**

**4. La situación de los Perfiles Fila y de los Perfiles Columna explican las igualdades y diferencias anteriores**

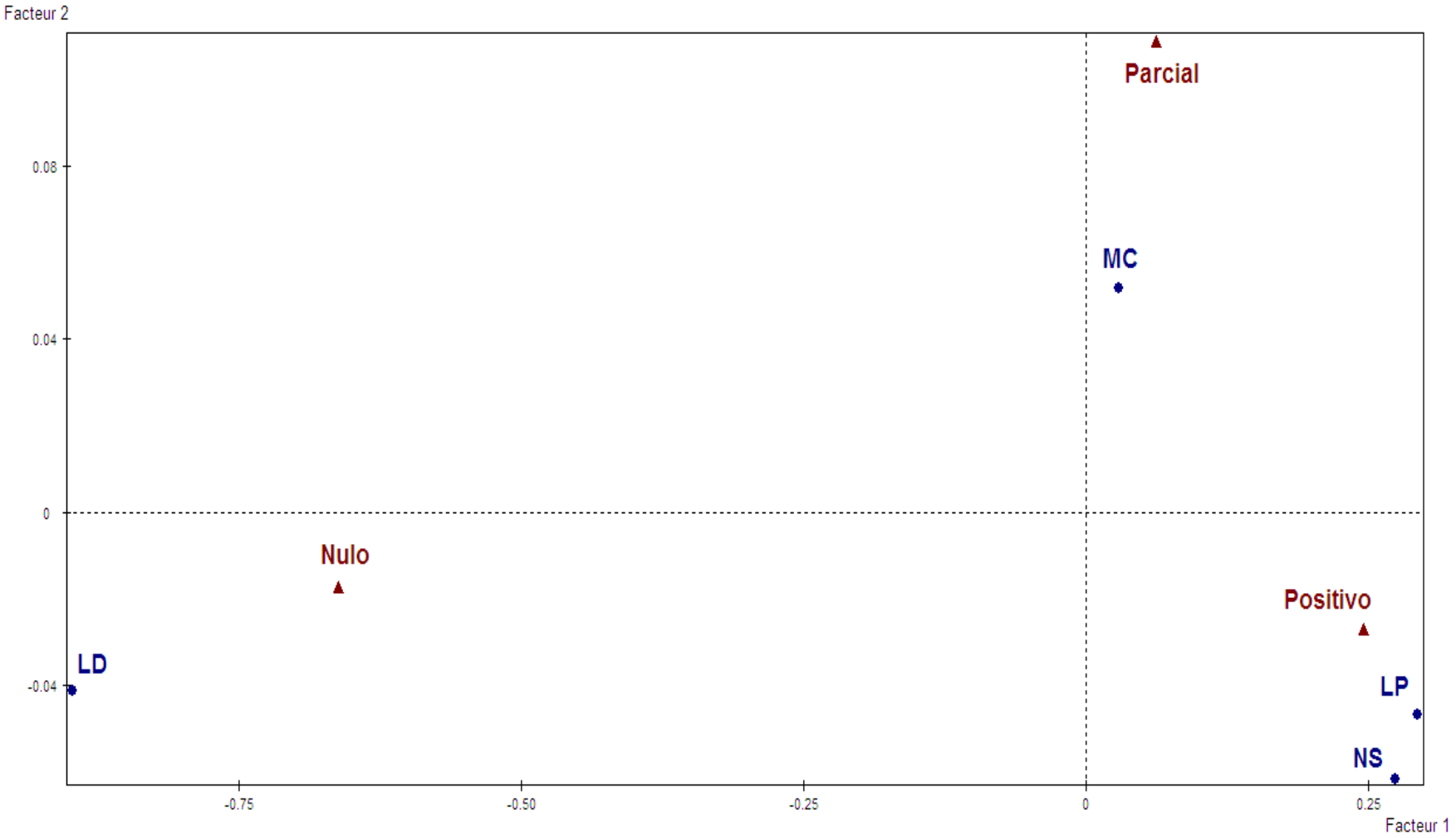
# Ejemplo 3: Enfermedad de Hodgkin

Enfermedad de Hodgkin (cancer). **538 pacientes** fueron clasificados en función de **4 tipologías de la enfermedad (LP, NS, MC, LD)** y su respuesta a un **Tratamiento (Positivo, Parcial, Nulo)** al cabo de tres meses.

| | Positivo | Parcial | Nulo |
|---|---|---|---|
| LP | 74 | 18 | 12 |
| NS | 68 | 16 | 12 |
| MC | 154 | 54 | 58 |
| LD | 18 | 10 | 44 |

**¿¿ Tratamiento igual en todas las tipologías ??**

# Ejemplo 3: Enfermedad de Hodgkin



# Ejemplo 4: ¿Qué piensan los ingleses…

## …. del resto de europeos?

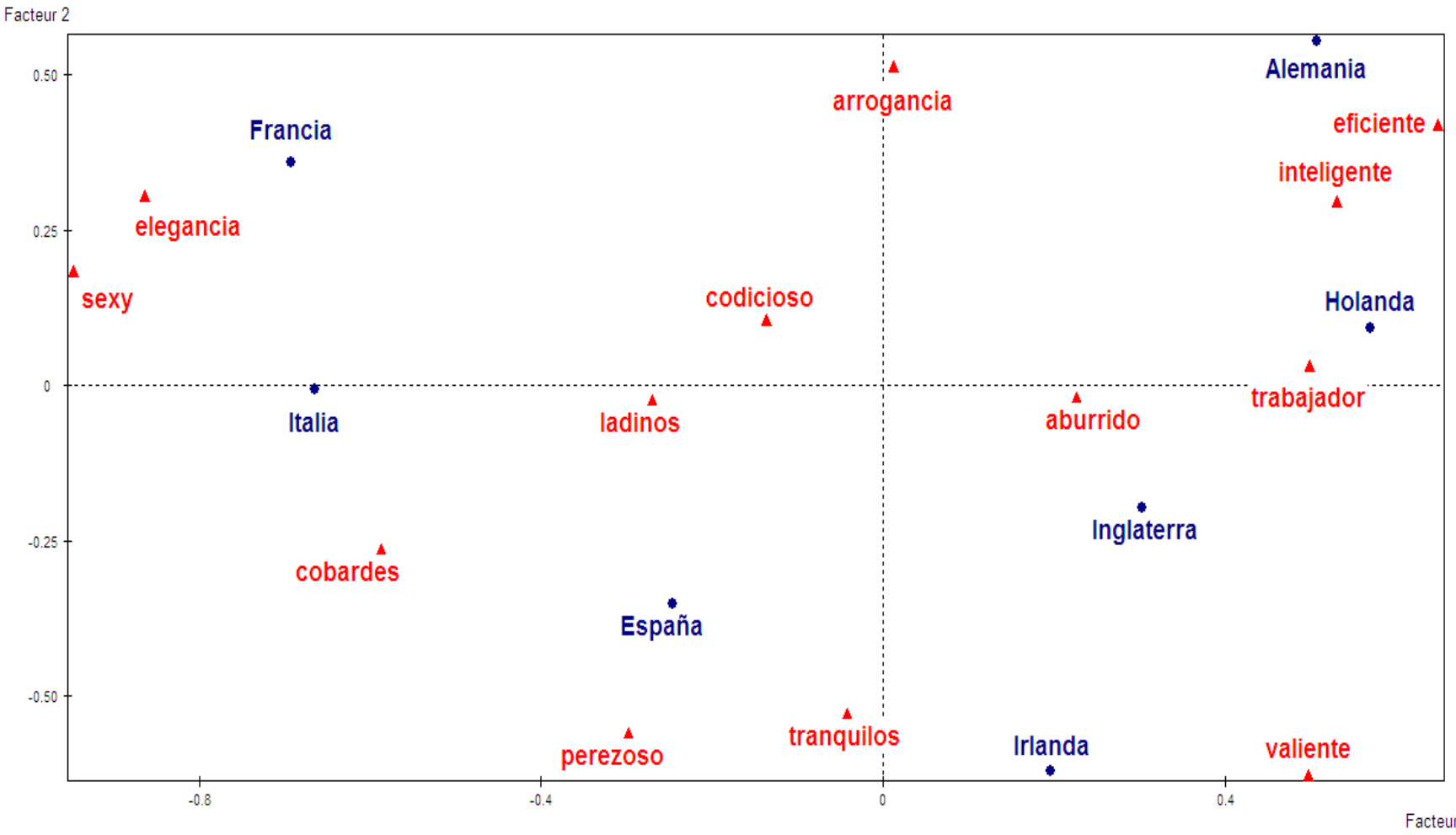
(1) elegancia (2) arrogancia (3) sexy (4) ladinos (5) tranquilos

(6) codicioso (7) cobardes (8) aburrido (9) eficiente (10) perezoso

(11) trabajador (12) inteligente (13) valiente

| Paises | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Francia | 37 | 29 | 21 | 19 | 10 | 10 | 8 | 8 | 6 | 6 | 5 | 2 | 1 |
| España | 7 | 14 | 8 | 9 | 27 | 7 | 3 | 7 | 3 | 23 | 12 | 1 | 3 |
| Italia | 30 | 12 | 19 | 10 | 20 | 7 | 12 | 6 | 5 | 13 | 10 | 1 | 2 |
| Inglat. | 9 | 14 | 4 | 6 | 27 | 12 | 2 | 13 | 26 | 16 | 29 | 6 | 25 |
| Irlanda | 1 | 7 | 1 | 16 | 30 | 3 | 10 | 9 | 5 | 11 | 22 | 2 | 27 |
| Holanda | 5 | 4 | 2 | 2 | 15 | 2 | 0 | 13 | 24 | 1 | 28 | 4 | 6 |
| Alemania | 4 | 48 | 1 | 12 | 3 | 9 | 2 | 11 | 41 | 1 | 38 | 8 | 8 |

# Ejemplo 4: ¿Qué piensan los ingleses…



# Análisis de Correspondencias Múltiples (ACM)

- **Análisis de Datos Categóricos**
- **Extensión del Análisis de Correspondencias Simples (ACS)**
  - Tres o más Variables Categóricas
- **Cálculos sencillos**
- **Resultados → bi-plots**
  - Muestran todas las variables y sus categorías.
  - Muestran todos los individuos
  - Fácil interpretación
- **No habitual**

# Análisis de Correspondencias Múltiples

# DATOS

Supongamos que:

- Estudiamos **3 variables categóricas: A, B y C**
  - Variable **A**: 3 categorías → **a1 a2 a3**
  - Variable **B**: 2 categorías → **b1 b2**
  - Variable **C**: 3 categorías → **c1 c2 c3**

- Estudiamos a **10 individuos**

# Análisis de Correspondencias Múltiples

## DATOS

| | A | B | C |
|---|---|---|---|
| *ind* 1 | *a*2 | *b*1 | *c*1 |
| *ind* 2 | *a*2 | *b*2 | *c*3 |
| *ind* 3 | *a*3 | *b*2 | *c*3 |
| *ind* 4 | *a*1 | *b*1 | *c*2 |
| *ind* 5 | *a*1 | *b*2 | *c*1 |
| *ind* 6 | *a*2 | *b*1 | *c*2 |
| *ind* 7 | *a*3 | *b*1 | *c*1 |
| *ind* 8 | *a*2 | *b*2 | *c*2 |
| *ind* 9 | *a*1 | *b*2 | *c*2 |
| *ind* 10 | *a*3 | *b*1 | *c*3 |

$$Z = \begin{array}{c} \begin{matrix} \text{a1} & \text{a2} & \text{a3} & \text{b1} & \text{b2} & \text{c1} & \text{c2} & \text{c3} \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

# Análisis de Correspondencias Múltiples

## DATOS

$$B = Z' \cdot Z =$$

| | a1 | a2 | a3 | b1 | b2 | c1 | c2 | c3 | |
|---|---|---|---|---|---|---|---|---|---|
| | 3 | 0 | 0 | 1 | 2 | 1 | 2 | 0 | **a1** |
| | 0 | 4 | 0 | 2 | 2 | 1 | 2 | 1 | **a2** |
| | 0 | 0 | 3 | 2 | 1 | 1 | 0 | 2 | **a3** |
| | 1 | 2 | 2 | 5 | 0 | 2 | 2 | 1 | **b1** |
| | 2 | 2 | 1 | 0 | 5 | 1 | 2 | 2 | **b2** |
| | 1 | 1 | 1 | 2 | 1 | 3 | 0 | 0 | **c1** |
| | 2 | 2 | 0 | 2 | 2 | 0 | 4 | 0 | **c2** |
| | 0 | 1 | 2 | 1 | 2 | 0 | 0 | 3 | **c3** |

Tabla **A & B**

Tabla **A & C**

Tabla **B & C**

**Total categoría**

# Análisis de Correspondencias Múltiples

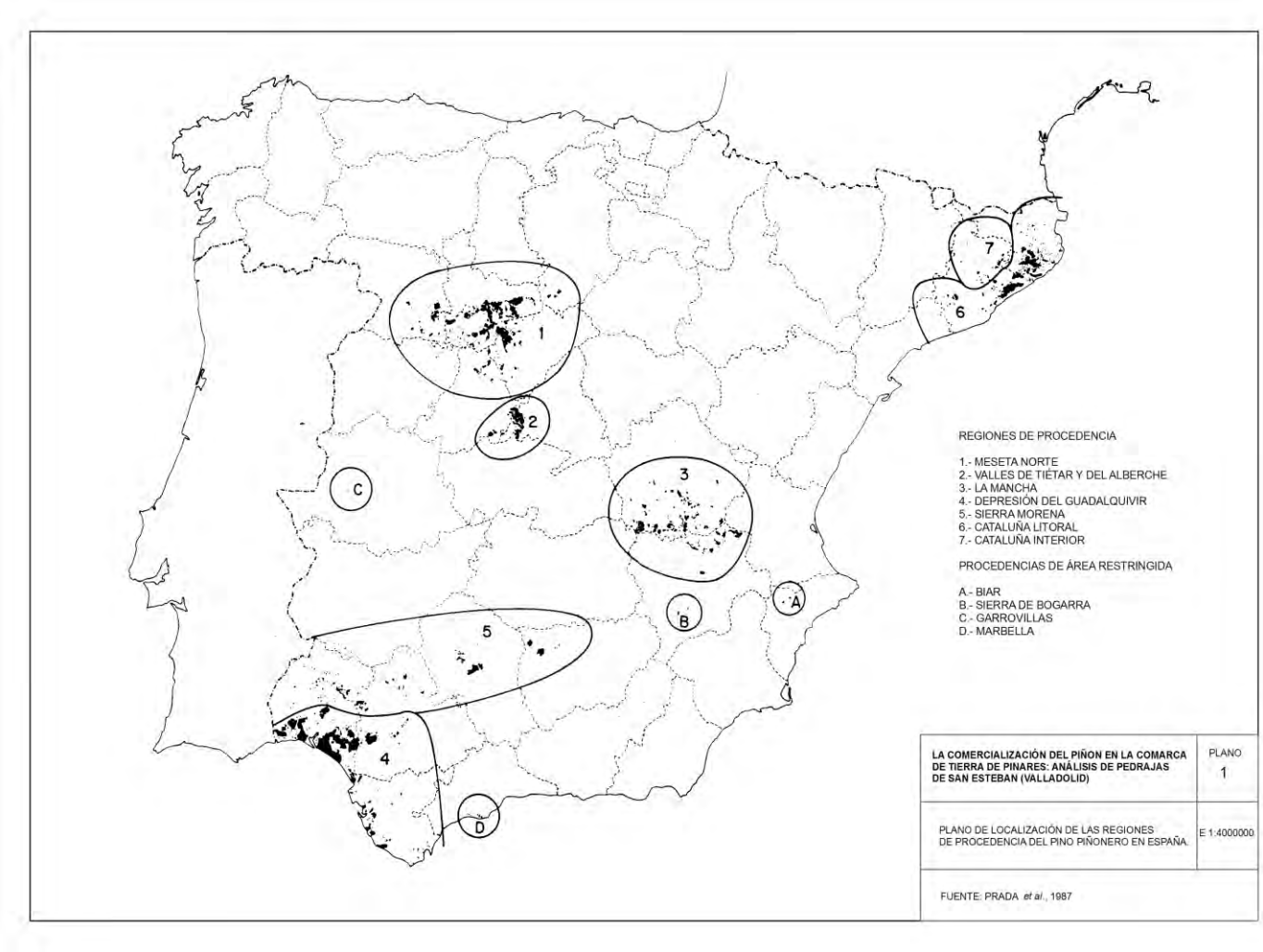
## DATOS

$$B = Z' \cdot Z =$$

| | a1 | a2 | a3 | b1 | b2 | c1 | c2 | c3 |
|---|---|---|---|---|---|---|---|---|
| a1 | 3 | 0 | 0 | 1 | 2 | 1 | 2 | 0 |
| a2 | 0 | 4 | 0 | 2 | 2 | 1 | 2 | 1 |
| a3 | 0 | 0 | 3 | 2 | 1 | 1 | 0 | 2 |
| b1 | 1 | 2 | 2 | 5 | 0 | 2 | 2 | 1 |
| b2 | 2 | 2 | 1 | 0 | 5 | 1 | 2 | 2 |
| c1 | 1 | 1 | 1 | 2 | 1 | 3 | 0 | 0 |
| c2 | 2 | 2 | 0 | 2 | 2 | 0 | 4 | 0 |
| c3 | 0 | 1 | 2 | 1 | 2 | 0 | 0 | 3 |

**TABLA DE BURT**

# Ejemplo 5: La Comercialización del piñón en la Comarca de Tierra de Pinares.



# Ejemplo 5: La Comercialización del piñón en la Comarca de Tierra de Pinares.

## OBJETIVOS

- Describir la cadena de valor del piñón desde su producción en el monte hasta el consumidor.
- Identificar y Caracterizar a los principales agentes implicados en esta cadena.
- Identificar los factores de éxito y de fracaso en este modelo de comercialización.
- Analizar las implicaciones sociales, ambientales y económicas de este modelo de comercialización.

# MATERIAL Y MÉTODOS

- **Entrevistas Personales**.

  1. Consumidores
  2. Empresarios

- **Tamaño de la muestra**

$$n = \frac{N \cdot Z_{\alpha}^{2} \cdot p \cdot q}{d^{2} \cdot (N-1) + Z_{\alpha}^{2} \cdot p \cdot q}$$

Consumidores = 100 encuestas
Empresarios = 30 encuestas

- **Tratamiento Estadístico de los datos**
  - Análisis Factorial (ACM)
  - Análisis Cluster

# MATERIAL Y MÉTODOS

- **Entrevistas Personales Consumidores**

Muestreo estratificado por rangos de edad y sexo

| | HOMBRES | | MUJERES | |
|---|---|---|---|---|
| **Rangos de edad** | **Nº habitantes** | **Tamaño muestral estimado** | **Nº habitantes** | **Tamaño muestral estimado** |
| **15 a 24** | 220 | 7 | 207 | 7 |
| **25 a 34** | 304 | 10 | 267 | 9 |
| **35 a 49** | 470 | 15 | 441 | 14 |
| **50 a 64** | 321 | 11 | 267 | 9 |
| **> 65** | 240 | 8 | 297 | 10 |
| **Total** | **1 555** | **51** | **1 479** | **49** |

# MATERIAL Y MÉTODOS

## ENCUESTA A CONSUMIDORES

**22 preguntas** en **3 bloques** diferenciados:

- ***3 Preguntas de identificación*:** edad, ocupación y nivel de estudios.
- ***11 Preguntas de consumo*:** dónde lo compran, procedencia, motivo de consumo, forma de consumo, frecuencia de consumo, etc.
- ***8 Preguntas de conocimiento*:** vinculación al sector, trabajo, parentescos, utilidades, etc.

# MATERIAL Y MÉTODOS

## ENCUESTA A EMPRESARIOS

**18 preguntas** en **4 bloques** diferenciados:

- ***7 Preguntas de caracterización*:** de la empresa: forma jurídica, número de socios, antigüedad, última inversión, fase elaboración, dedicación, etc.
- ***3 Preguntas de tipo laboral*:** número empleados, fase elaboración, tipo de contrato.
- ***4 Preguntas de tipo comercial*:** sobre materias primas y productos finales.
- ***2 Preguntas de conocimiento***: beneficios relacionados con el aprovechamiento del piñón.

# ANÁLISIS DE DATOS

## Metodología

**1. Análisis Descriptivo de las Variables**

- Análisis Univariantes.
- Análisis Bivariantes (Tablas de Contingencia).
- Primeros Resultados y Depuración de los Datos

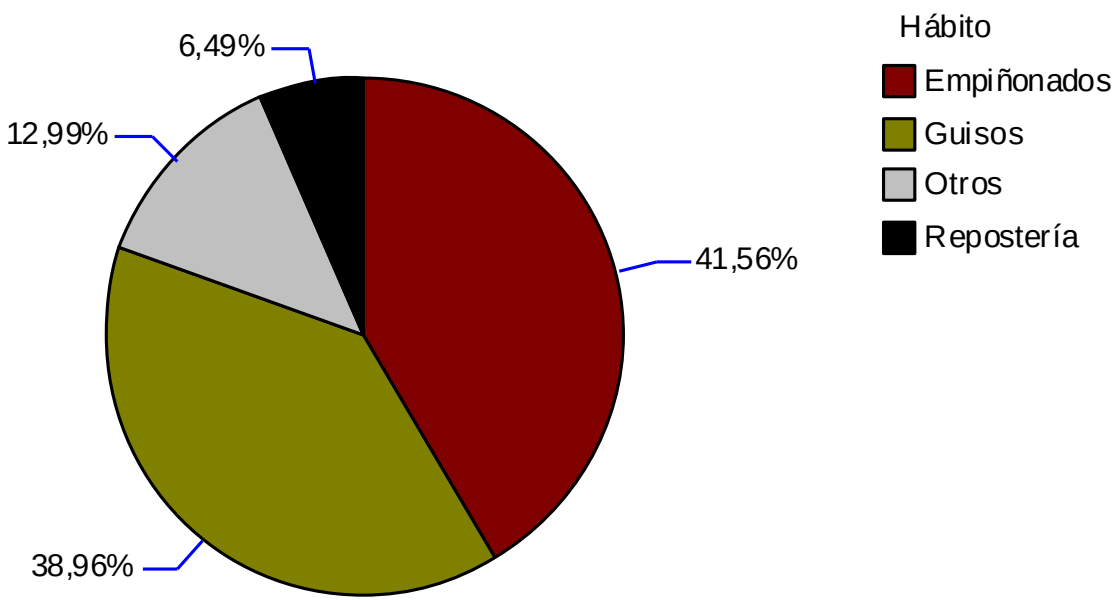
**2. Análisis Factorial – ACM**

- Elección del número de Ejes Factoriales.
- Caracterización de los ejes

**3. Análisis Cluster**

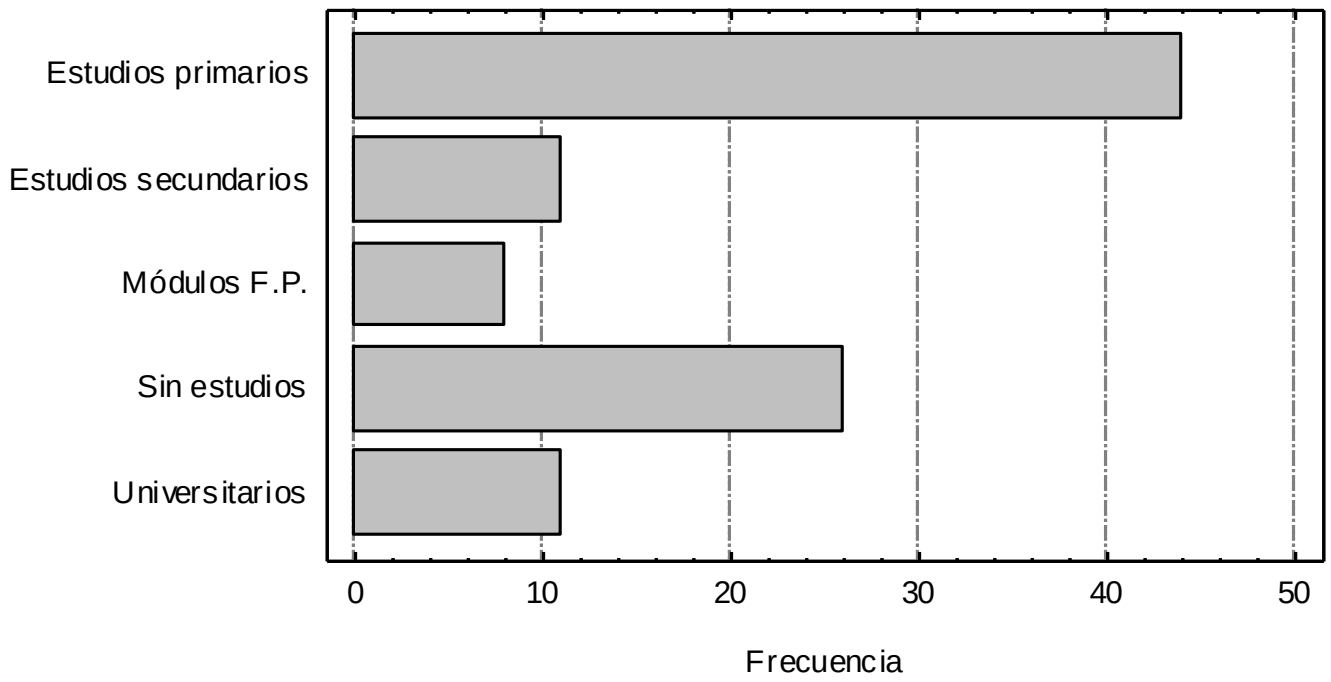
- Elección del número de Grupos.
- Caracterización de los Grupos

# FORMA HABITUAL DE COMSUMO



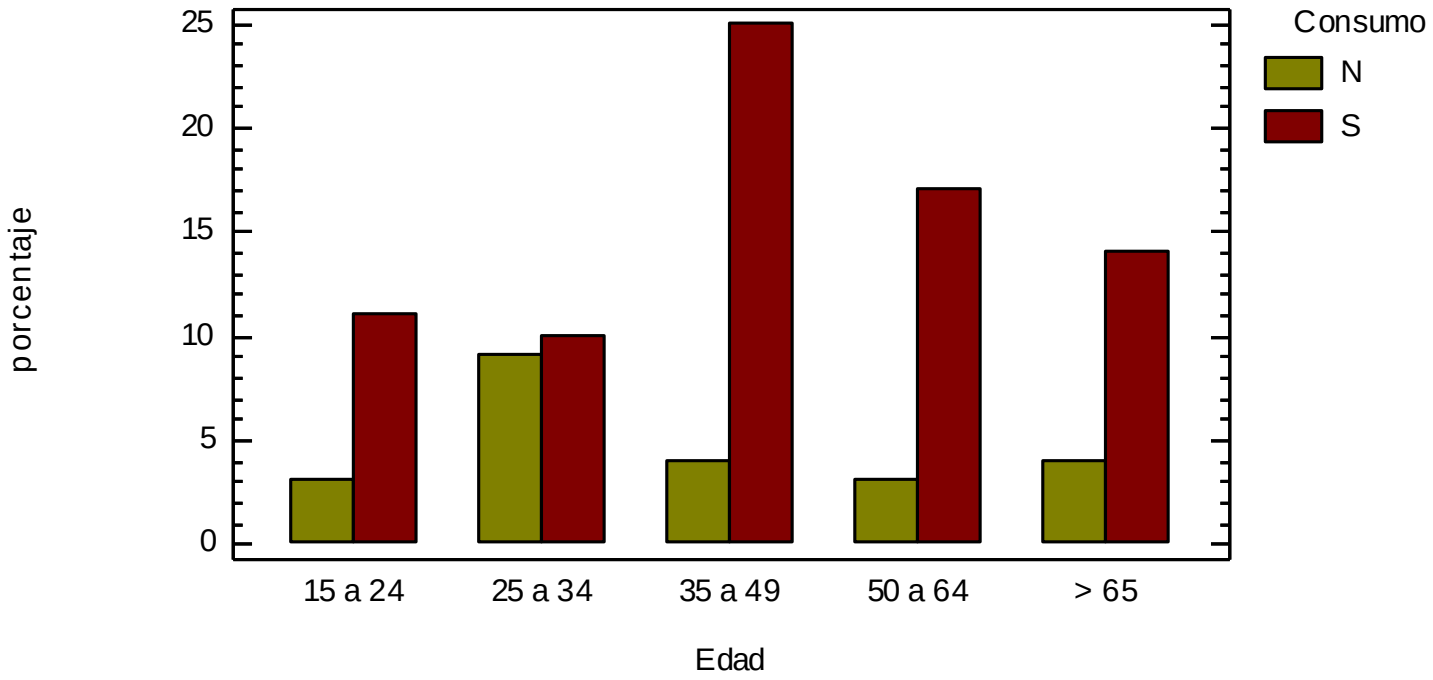

**Tipos de Consumo**

# NIVEL DE ESTUDIOS



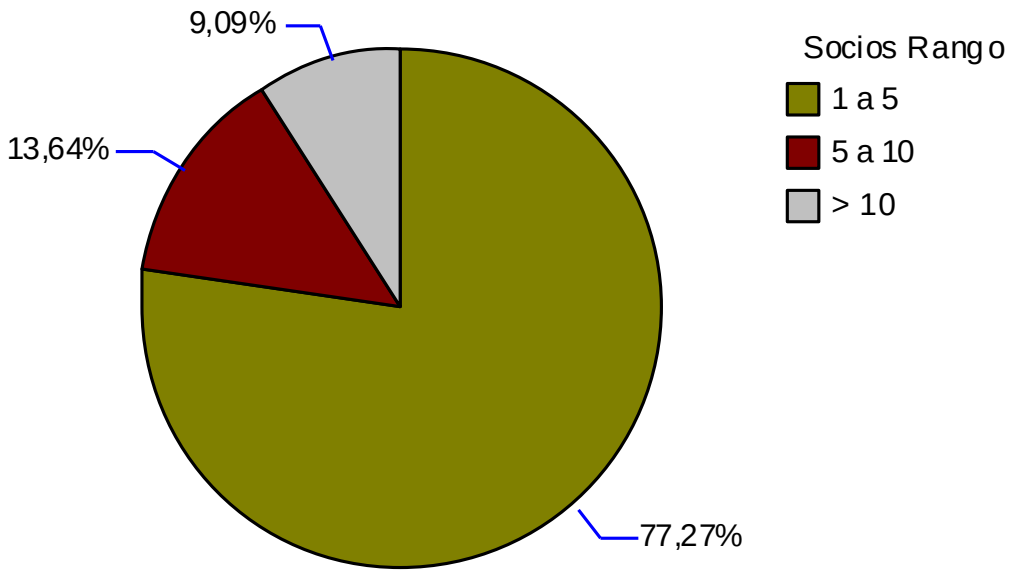

**Tipos de Consumo**

# RELACIÓN EDAD CONSUMO



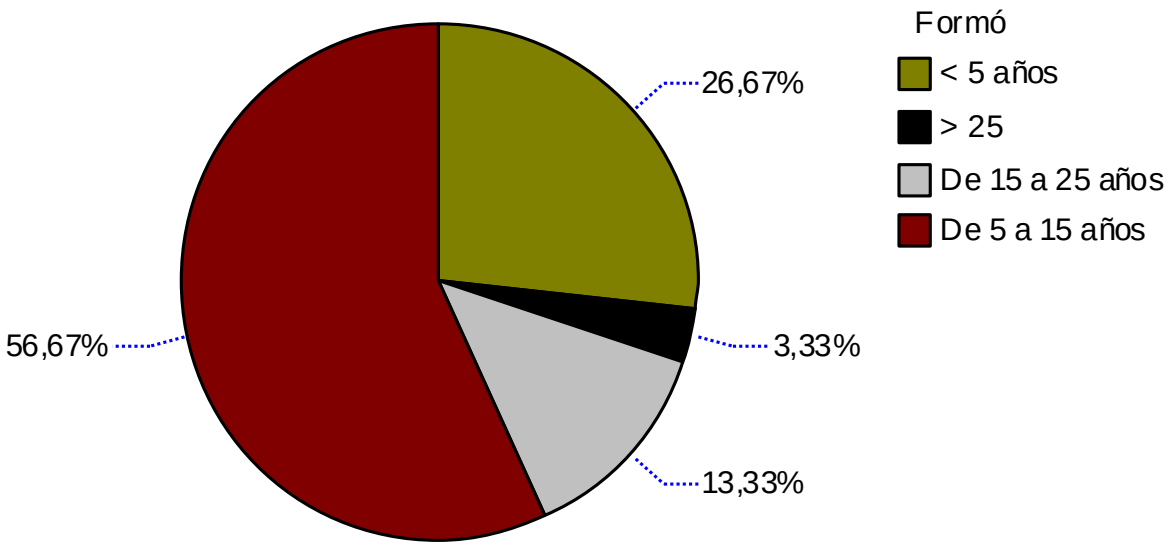

**Edad & Consumo**

# TAMAÑO DE LAS EMPRESAS.



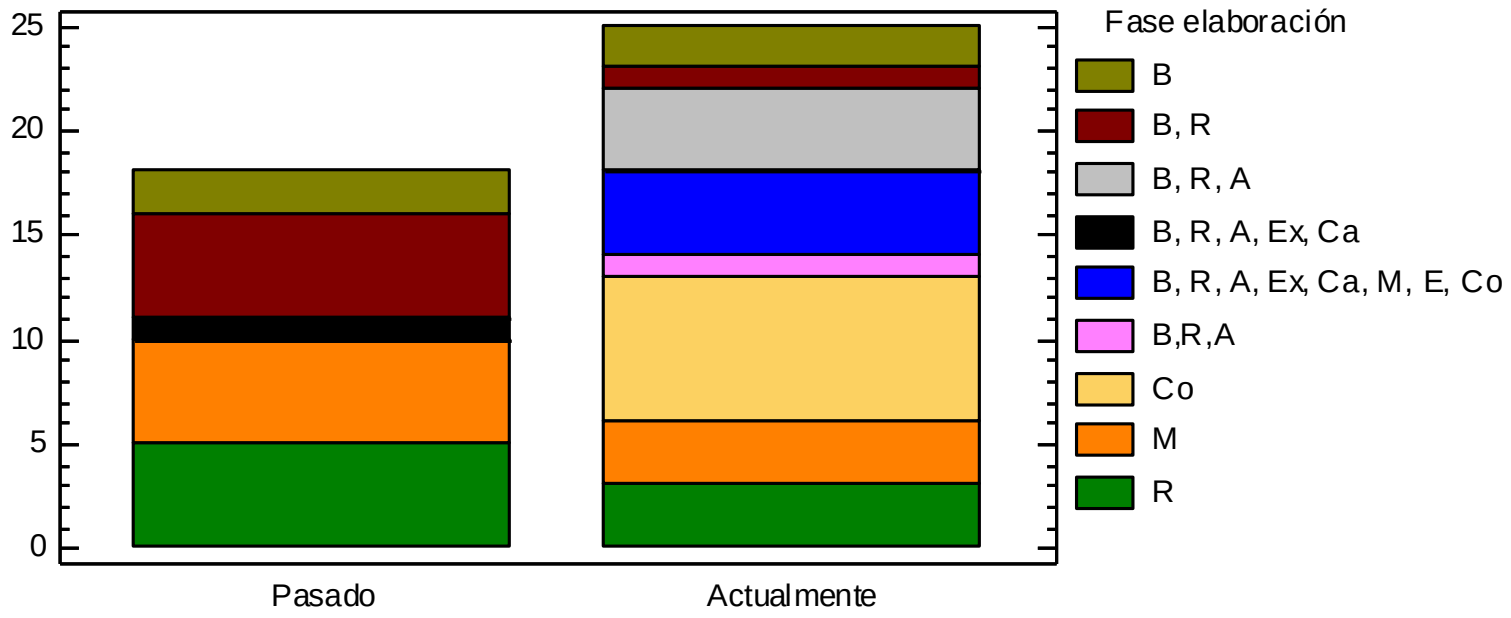

**Nº de Socios**

# ANTIGÜEDAD DE LAS EMPRESAS.




**Años de Antigüedad**

# TIPO DE TRABAJO DENTRO DE CADENA PIÑÓN




B = Bajada de piñas
R = Recogida de piñas.
A = Almacenamiento de piñas.
Ca = Cascado.
M = Mondado.
E = Envasado.
Co = Comercialización.

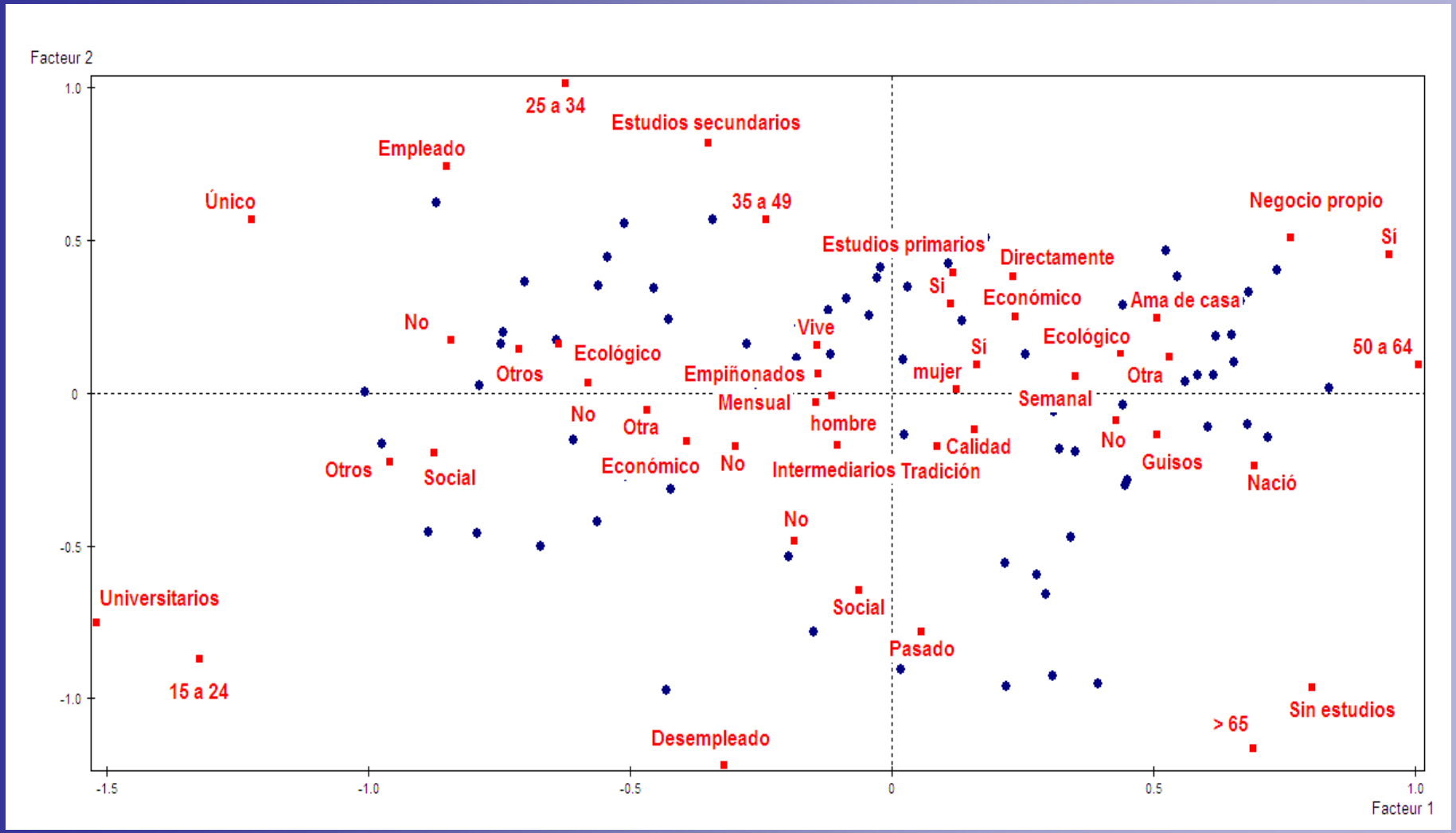
# A. Correspondencias Múltiples

| Variable | Nº de pregunta en cuestionario | Denominación en el análisis ACM | Modalidades |
|---|---|---|---|
| Relación con el pueblo de Pedrajas | 0 | Relación (C1) | - Nació en Pedrajas<br>- Vive en Pedrajas<br>- Otra |
| Frecuencia de consumo de piñón | 2 | Frecuencia (C3) | - Semanalmente<br>- Mensualmente<br>- Otra |
| Atención que se presta a la marca al consumir | 3 | Marca (C4) | - Sí<br>- No |
| Atención que se presta a la procedencia al consumir | 4 | Procedencia (C5) | - Sí<br>- No |
| Motivo por el que consume | 6 | Consumo (C6) | - Único que le ofrecen<br>- Calidad<br>- Tradición<br>- Otra |
| Forma de conseguir el piñón que consumen | 7 | Conseguir (C7) | - Directamente<br>- Intermediarios |
| Forma habitual de consumo | 9 | Hábito (C8) | - Empiñonados<br>- Guisos<br>- Otros |
| Trabaja dentro del sector del piñón | 12 | Trabaja (C9) | - No<br>- Pasado<br>- Sí |

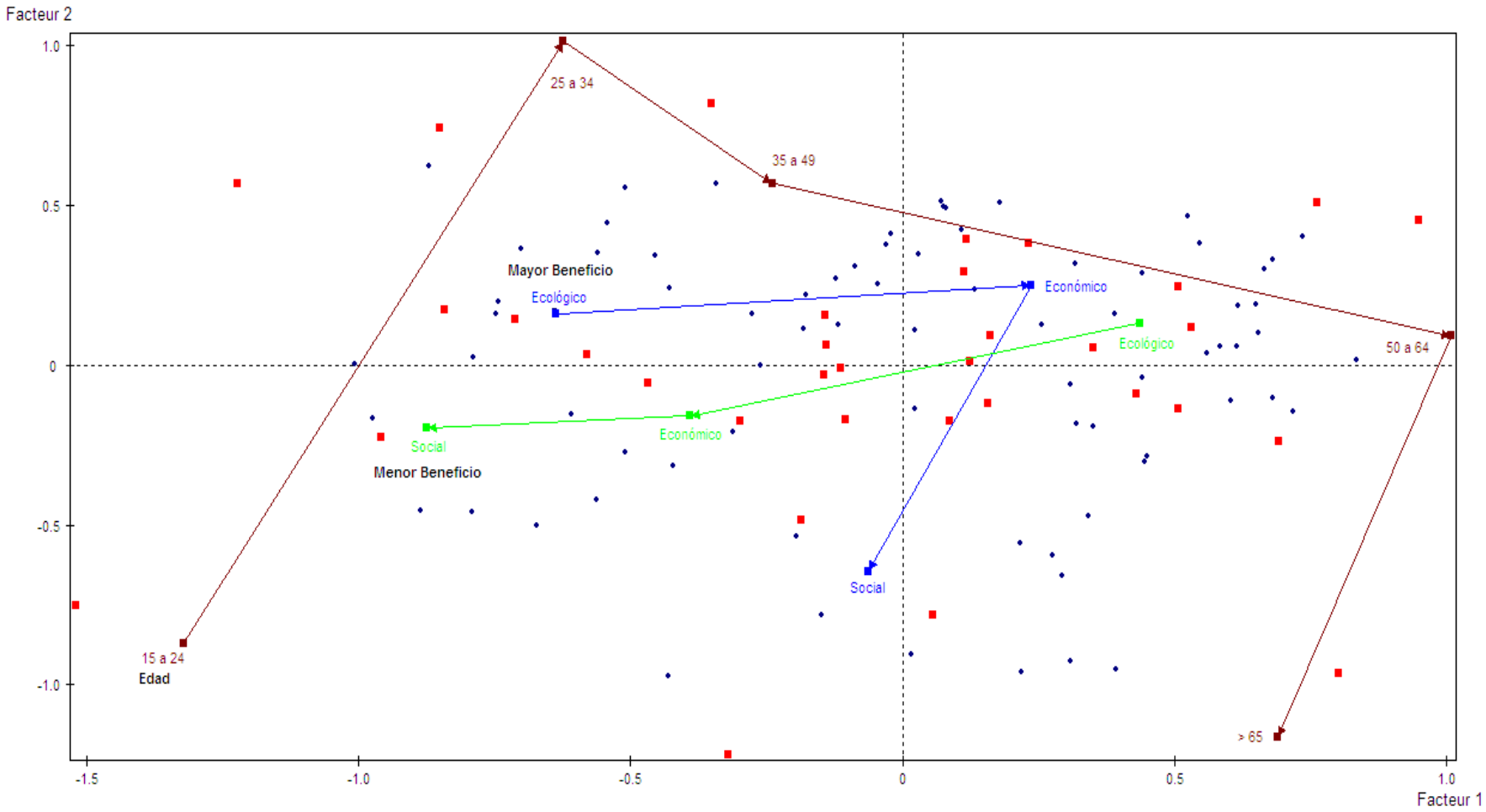
# A. Correspondencias Múltiples

| Variable | Nº de pregunta en cuestionario | Denominación en el análisis ACM | Modalidades |
|---|---|---|---|
| Miembro de la familia trabaja/ó en el sector del piñón | 16 | Miembro (C10) | - Sí<br>- No |
| Beneficios de mayor importancia | 17 | Mayor beneficio (C12) | - Ecológico<br>- Económico<br>- Social |
| Beneficios de menor importancia | 17 | Menor beneficio (C14) | - Ecológico<br>- Económico<br>- Social |
| Edad del encuestado | 20 | Edad (C15) | - 15 a 24<br>- 25 a 34<br>- 35 a 49<br>- 50 a 64<br>- > 65 |
| Ocupación laboral | 21 | Ocupación (C16) | - Ama de casa<br>- Desempleado<br>- Empleado<br>- Negocio propio |
| Nivel de estudios | 22 | Nivel de estudios (C17) | - Sin estudios<br>- Estudios primarios<br>- Estudios secundarios<br>- Universitarios |
| Sexo | - | Sexo (C18) | - Hombre<br>- Mujer |

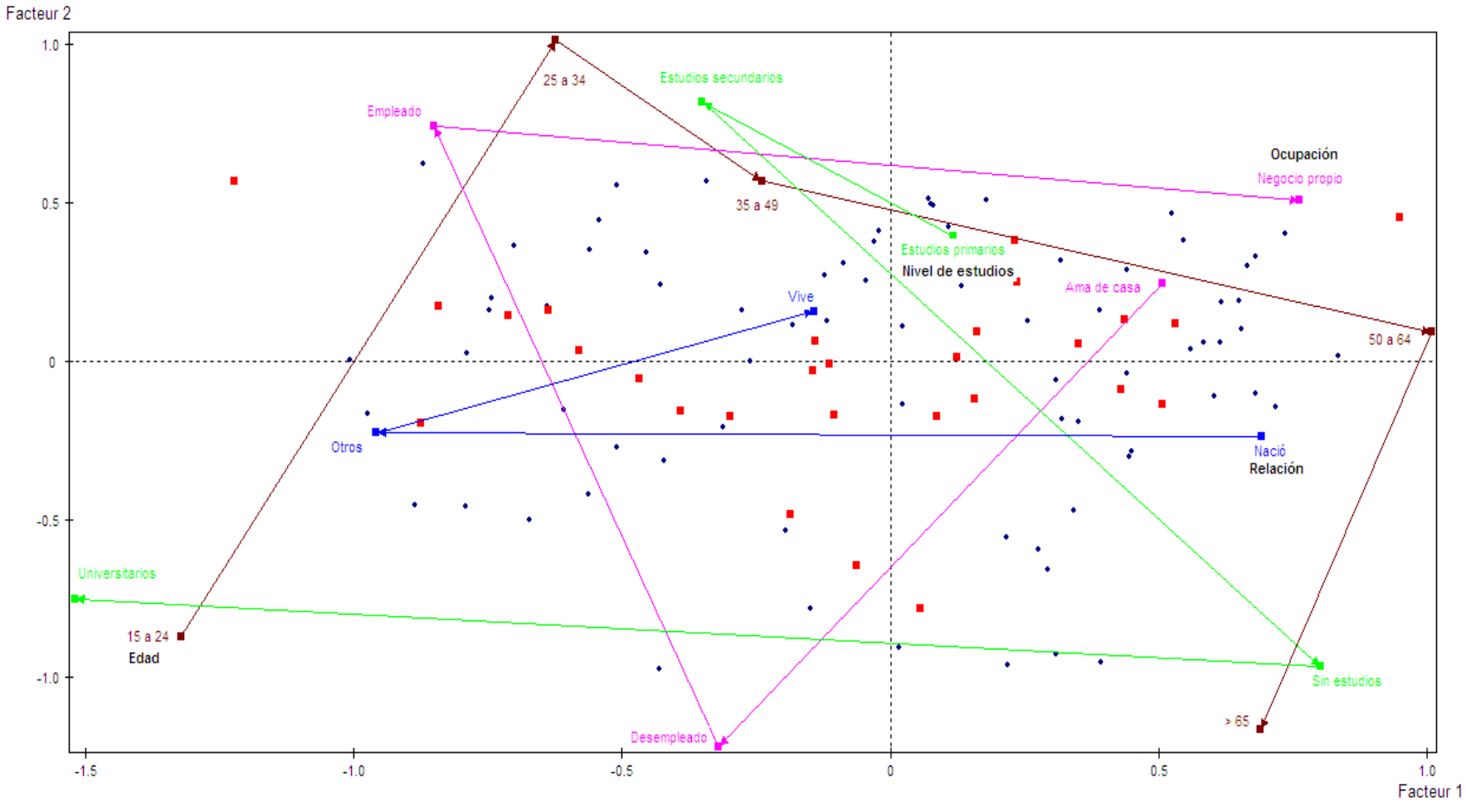
# ACM – Factor 1 & Factor 2



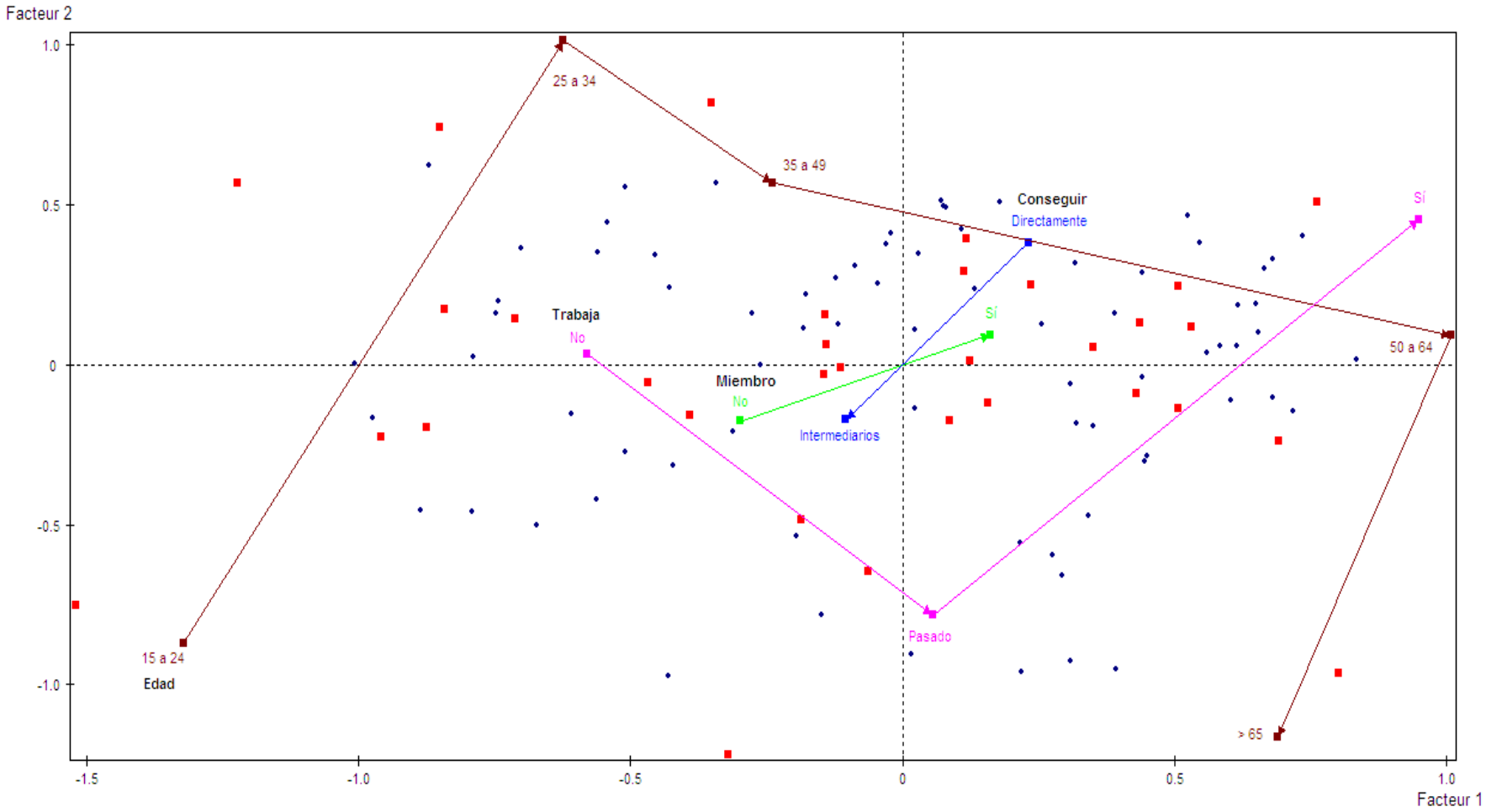
# ACM – Factor 1 & Factor 2



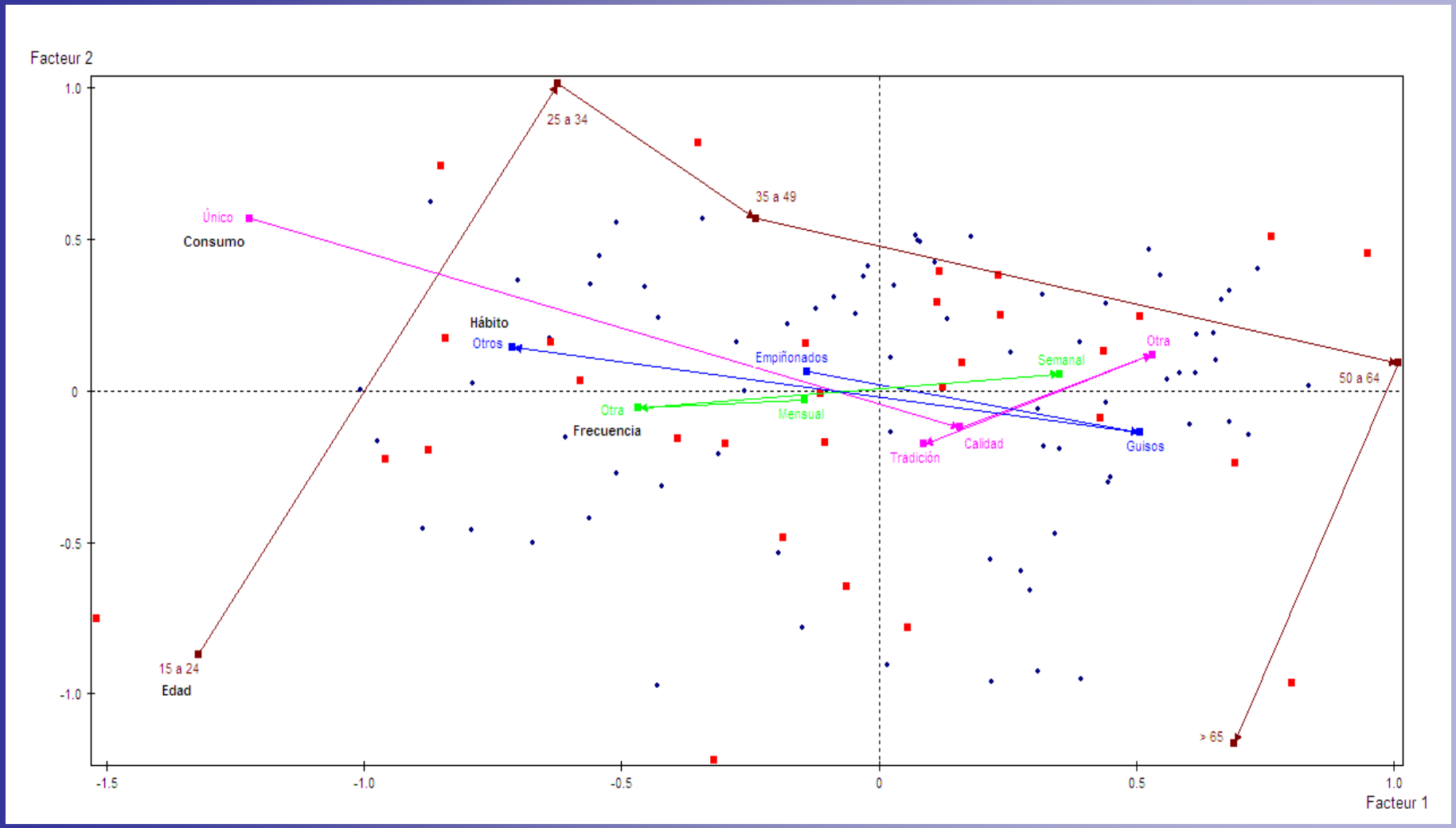
# ACM – Factor 1 & Factor 2



# ACM – Factor 1 & Factor 2



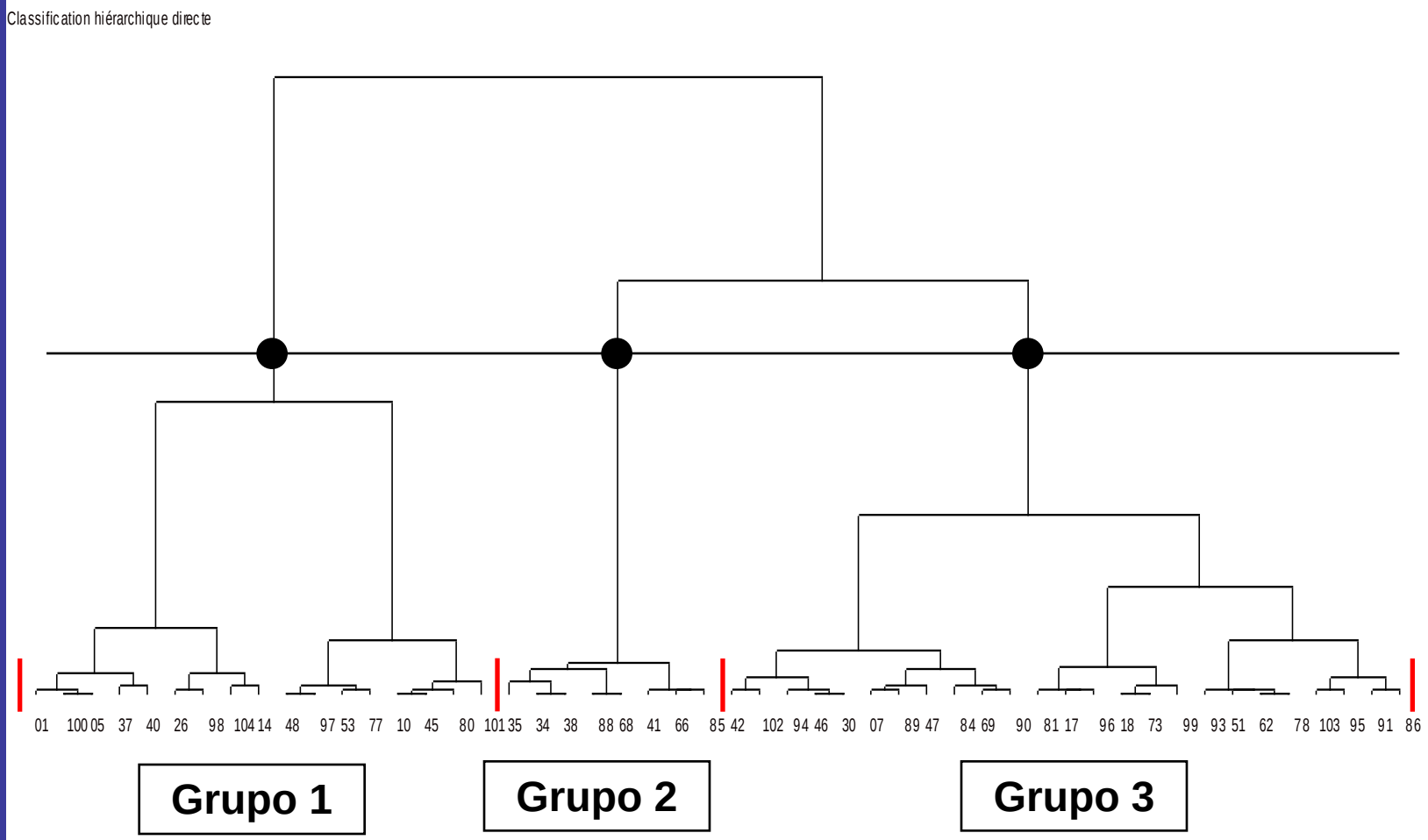
# ACM – Factor 1 & Factor 2



# CARACTERIZACIÓN DE EJES FACTORIALES

- **Factor 1. Caracterizado por :**
  - Una **gradación creciente de las edades**, asociado con el **nivel de estudios**
- **Factor 2. Caracterizado por :**
  - Separación **“trabajadores” ---- “desempleados”**
- **Factor 1 + Factor 2 = 20.54 % de la inercia Total**
- **Análisis Cluster a partir de coordenadas**
  - Dendograma (Ward, distancias euclideas al cuadrado)
  - Elección del número de grupos
  - Consolidación utilzando *k-medias*
  - Caracterización de los grupos

# ANALISIS CLUSTER



# Caracterización de los Grupos

- **Grupo 1**: (36% de la muestra).

Consumidores de edad entre los **15 a 24 años** con estudios **universitarios**. La relación que tienen con el pueblo es poco cercana, ya que **visitan el pueblo** de vez en cuando. **No fijarse en la procedencia del piñón** que consumen.

- **Grupo 2**: (15% de la muestra).

Consumidores de **edad superior a 65 años**, **jubilados** y **sin estudios**. **Nacieron o vivieron en el mismo municipio**. **No fijarse en la marca** de los distintos productos que compran. **Trabajaron en el sector del piñón en el pasado.**

- **Grupo 3**: (49% de la muestra).

edad comprendido de los **50 a 64 años** que **nacieron en el mismo municipio**. Tienen un **negocio propio**, **trabajan en el sector** del piñón. Respecto al **consumo**, **se fijan en la procedencia** del piñón y lo hacen en un 55% de los casos **de forma semanal**.

**Muito Obrigado**

**Muchas Gracias**



INSTITUTO POLITÉCNICO
DE BRAGANÇA