

# WEB DIDÁCTICA DE ANÁLISIS MULTIVARIANTE CON R



Úrsula Faura Martínez (faura@um.es), Fuensanta Arnaldos García, M<sup>a</sup> Teresa Díaz Delfa, Lourdes Molera Peris, Isabel Parra Frutos, Juan José Pérez Castejón (Departamento de Métodos Cuantitativos para la Economía y Empresa - Universidad de Murcia)

**OBJETIVO:** Fomentar y favorecer el aprendizaje autónomo y no presencial de las técnicas multivariantes.

**HERRAMIENTA:** Web con **materiales didácticos on-line** sobre la aplicación de técnicas multivariantes exploratorias de interdependencia con R.

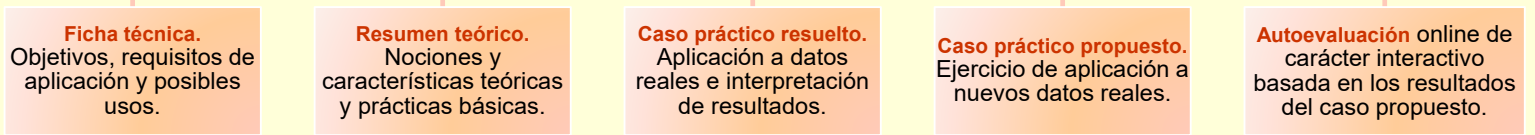
**CARACTERÍSTICAS:** Teoría, práctica y autoevaluación con datos reales.

**Material disponible en:** <https://www.um.es/estadempresamultivarianteR/>

**DESTINATARIOS:**

- Estudiantes de **GRADO** de la Facultad de Economía y Empresa (Administración y Dirección de Empresas, Economía, Marketing y Sociología).
- Estudiantes de **MASTER** de la Facultad de Economía y Empresa.
- PROFESIONALES** del ámbito de la economía, la empresa, la investigación de mercados o la sociología.

## ESTRUCTURA COMÚN PARA TODAS LAS TÉCNICAS



## CORRESPONDENCIAS SIMPLE (CA)

## CORRESPONDENCIAS MÚLTIPLE (MCA)

ANÁLISIS DE CORRESPONDENCIAS SIMPLE

**Conceptos básicos**

Perfiles fila y perfiles columna

El ACS se basa en los denominados perfiles fila y perfiles columna. Un perfil fila no es más que la distribución (en términos relativos) que se obtiene a partir de la tabla de contingencia condicionando a la categoría de una fila, esto es, sería la distribución de frecuencias relativas de las modalidades en las columnas condicionada a una modalidad fila. Así, la suma de los elementos de un perfil fila en 1, por lo que, a pesar de tener tantos elementos como columnas, los perfiles fila pertenecen a un subespacio de dimensión  $J-1$ , donde  $J$  es el número de columnas. La expresión de cada frecuencia relativa condicionada es:

$$f_{i1} = \frac{n_{i1}}{n_{i.}} = \frac{f_{i1}}{f_{i.}}$$

	$y_1$	$y_2$	...	$y_j$	...	$y_J$	
$Y / X = x_1$	$f_{11}$	$f_{12}$	...	$f_{1j}$	...	$f_{1J}$	1
$Y / X = x_2$	$f_{21}$	$f_{22}$	...	$f_{2j}$	...	$f_{2J}$	1
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$Y / X = x_i$	$f_{i1}$	$f_{i2}$	...	$f_{ij}$	...	$f_{iJ}$	1
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$Y / X = x_I$	$f_{I1}$	$f_{I2}$	...	$f_{Ij}$	...	$f_{IJ}$	1

A cada perfil fila  $i$  le corresponde una masa igual a su frecuencia relativa,  $f_{i.}$ , que representa el peso de dicho perfil en el análisis.

Del mismo modo, un perfil columna corresponde a la distribución en frecuencias relativas de las modalidades fila condicionada a una modalidad columna. De nuevo, al ser la suma de sus elementos igual a 1, los perfiles columna se sitúan en un subespacio de dimensión  $I-1$ , donde  $I$  es el número de filas. Como antes, la expresión de cada frecuencia relativa condicionada viene dada por:

$$f_{1j} = \frac{n_{1j}}{n_{.j}} = \frac{f_{1j}}{f_{.j}}$$

ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLE

**7. Gráficos con elipses de concentración de las categorías sobre las nubes de individuos**

Los siguientes gráficos nos van a permitir visualizar de forma distinta los resultados obtenidos. Para ello se usarán proyecciones de nubes de individuos a las que se añadirán elipses de concentración de los individuos correspondientes a cada categoría de una determinada variable. Se podrá ver sobre qué dimensión o plano se produce una mayor diferenciación entre las categorías de esa variable.

1. Gráficos para el porcentaje de inercia de cada dimensión

2. Gráficos para las contribuciones de las categorías (tot) a una dimensión

3. Gráficos para las contribuciones de las categorías (tot) a varias dimensiones

4. Gráficos para la calidad de representación de las categorías (tot) en el espacio formado por varias dimensiones

5. Gráficos para la nube de categorías seleccionadas por algún criterio

6. Gráficos para el coeficiente eta2

7. Gráficos con elipses de concentración de las categorías sobre las nubes de individuos

Condiciones:  
ACSI con FactoMineR incluyendo categorías suplementarias

Aplicación shiny  
ACSI con el paquete ca

MCA factor map

En estos gráficos se puede observar cómo las elipses de concentración de los puntos correspondientes a las categorías de una cualquiera de las variables elegidas están diferenciadas entre sí horizontalmente, indicando que la dimensión representada en ese eje (dimensión 1 en este caso) discrimina entre ambas categorías de la variable. Esto no es así para la dimensión en el eje vertical (dimensión 2).

## CONGLOMERADOS (CLUS)

## COMPONENTES PRINCIPALES (ACP)

ANÁLISIS DE CONGLOMERADOS

**Caso práctico propuesto**

Agrupar provincias por similitud en diversas variables socioeconómicas (2014)

El Instituto Nacional de Estadística (<http://www.ine.es>) proporciona información a nivel provincial de numerosas variables socioeconómicas. Entre ellas, y para el año 2014, podemos encontrar:

- tasainm:** tasa bruta de inmigración (personas/mil habitantes)
- tasaeem:** tasa bruta de emigración (personas/mil habitantes)
- tasact:** tasa actividad (activos/p total)
- tasempleo:** tasa de empleo (ocupados/p total)
- tasaparo:** tasa de paro (parados/activos)
- ipc:** IPC base 2016
- piipc:** PIB a precios corrientes por individuo

Utilice esta información para agrupar las provincias, incluyendo las ciudades autónomas de Ceuta y Melilla, con una situación similar.

**Resuelva el caso y conteste a las preguntas de autoevaluación**

En el fichero datos se incluye información sobre diversas variables tanto a nivel de comunidad autónoma como a nivel provincial, así como un fichero de formas de las provincias que puede emplear si desea representar los conglomerados en forma de mapas. Cargue el fichero y una vez crea que ha llegado a alguna conclusión sobre el comportamiento de las provincias españolas en estas variables socioeconómicas responda a las cuestiones de autoevaluación.

ANÁLISIS DE COMPONENTES PRINCIPALES

**Autoevaluación**

¿Son verdaderas o falsas las siguientes afirmaciones?

El valor del componente 1 para la décima empresa es 2.113653

Solución: Verdadera. Para obtener el valor del componente 1 para la décima empresa debe utilizar la combinación `res` de las variables originales estandarizadas para esa empresa y los coeficientes proporcionados en el objeto resultante del uso de la función `principle()` en `$loadings[, 1]`. Es más directo si utiliza directamente la información proporcionada en `$scores[,10, 1]`.

Si retiene cuatro componentes la communalidad de la variable `capita1` es igual a 0.4489152.

Solución:

Reteniendo tres componentes, y rotando las cargas utilizando la función `varimax()` el coeficiente de correlación del componente 2 y la variable `abdtta_ventas` es igual a 0.953.

Solución:



Proyecto de innovación docente  
**Técnicas Estadísticas Avanzadas en Ciencias Sociales**

Proyecto de innovación docente  
**Técnicas Estadísticas Avanzadas con R en Ciencias Sociales**

