



Tablas de contingencia: análisis de correspondencias simple.

Miguel Ángel Tarancón Morán & Consolación Quintana Rojo



Área de Estadística
Económica y Empresarial

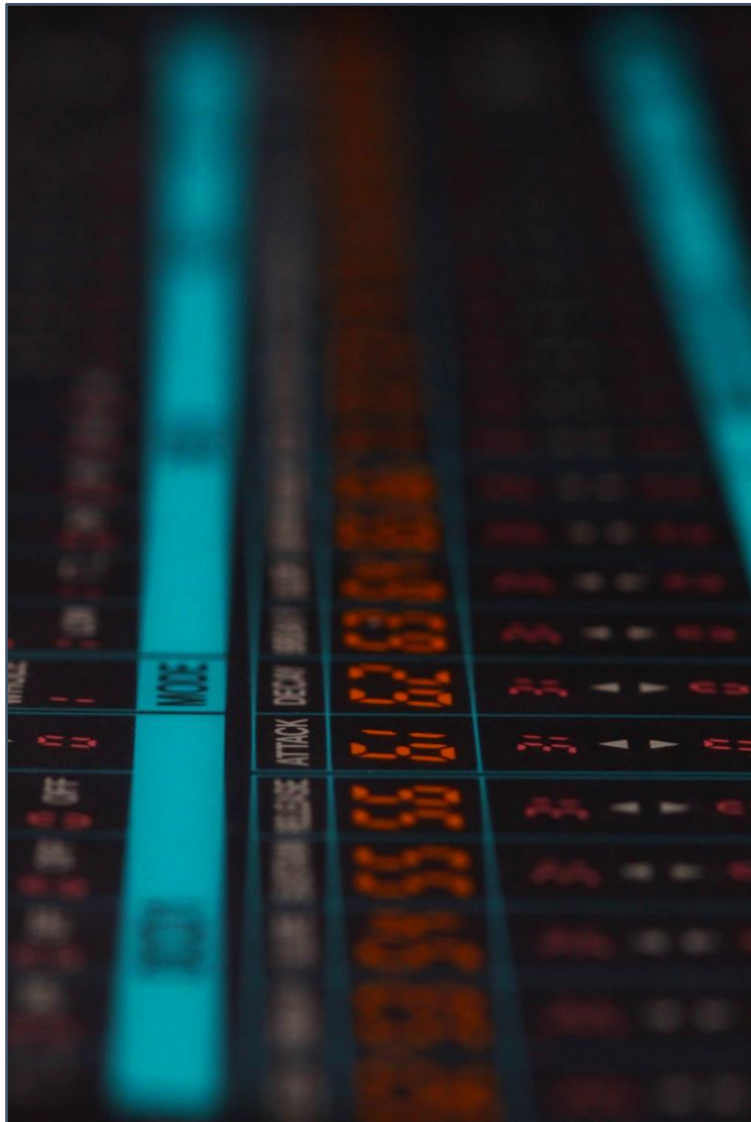
Departamento de Economía Aplicada 1
Universidad de Castilla – La Mancha



- 1 Introducción.
- 2 Tablas de contingencia y asociación.
- 3 Análisis de correspondencias simple.



- El **análisis de correspondencias** es una de las técnicas de análisis de **interdependencias** de reducción de la dimensión de la información.
- Es muy parecida al análisis de componentes principales, pero aplicado a atributos o factores.
- El objetivo principal es **representar en un gráfico de dos ejes o dimensiones** la relación (**asociación**) existente entre las categorías o niveles de 2 o más atributos o factores, mediante la **visualización** de la tabla de contingencia.
- **Asociación:** existe cuando se da que el hecho de que los casos adopten ciertos niveles o categorías en unos factores hace que tiendan a tomar ciertos niveles o categorías de otro u otros factores.



- La información de partida sobre los casos (categorías que adoptan en cada atributo o factor) viene sintetizada en **tablas de contingencia**.
- Por ejemplo, una tabla de contingencia que recoge las frecuencias conjuntas correspondientes a los atributos “situación laboral” y “situación académica”:

	Aprobados	No Aprobados	$n_{i.}$
Trabajan	23	43	66
No Trabajan	67	31	98
$n_{.j}$	90	74	164

n_{ij} : frecuencias conjuntas (número de casos para cada combinación de niveles de cada atributo)

$n_{.j}$: frecuencias marginales del atributo “situación académica” (número de casos para cada nivel del atributo)

N : frecuencia total (número total de casos)

$n_{i.}$: frecuencias marginales del atributo “situación laboral” (número de casos para cada nivel del atributo)

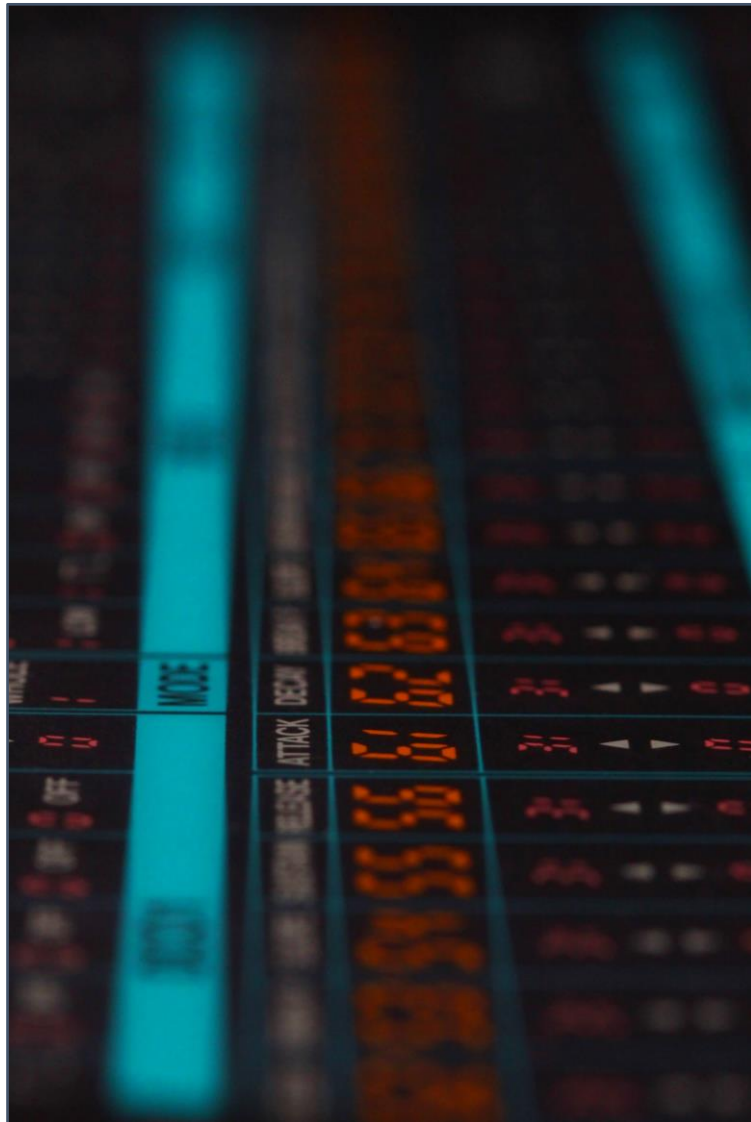


- En el caso de atributos, la condición de independencia (no asociación) es que la frecuencia relativa conjunta coincida con el producto de las correspondientes frecuencias marginales para todas las frecuencias de la tabla de contingencia:

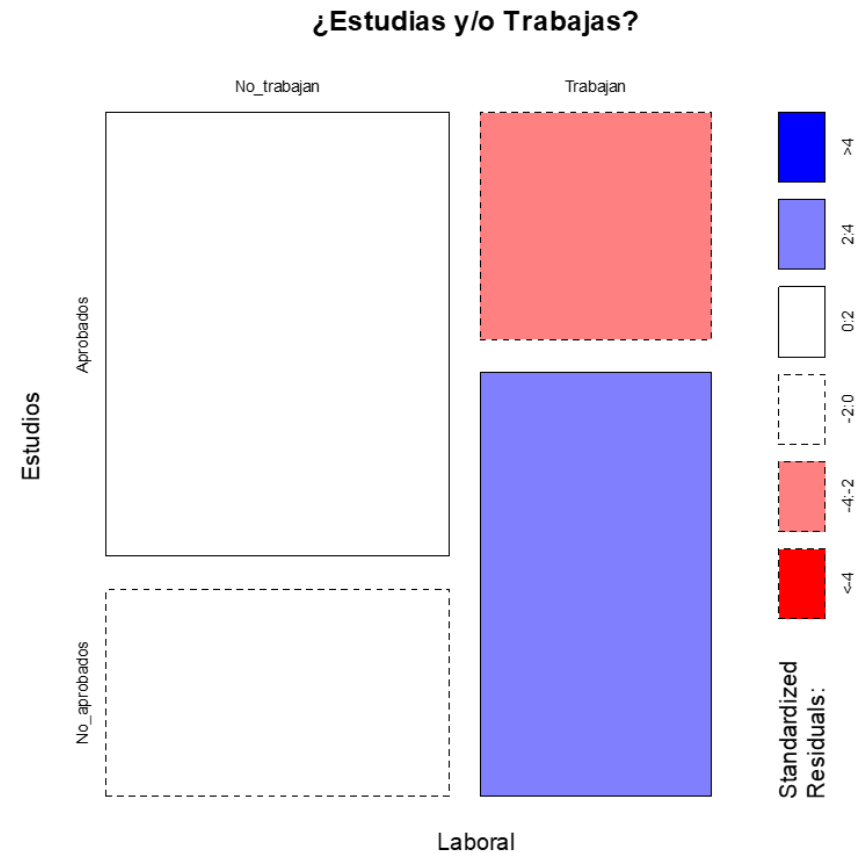
$$f_{ij} = \frac{n_{ij}}{N} = \frac{n_{i\cdot}}{N} \cdot \frac{n_{\cdot j}}{N} = f_{i\cdot} \cdot f_{\cdot j} \quad \forall i, j$$

- Operando, las frecuencias esperadas (teóricas) en caso de independencia serían:

$$\frac{n_{ij}}{N} = \frac{n_{i\cdot}}{N} \cdot \frac{n_{\cdot j}}{N} \Rightarrow E_{ij} = n_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{N}$$



- En nuestro ejemplo:



El color rojo/salmón indica frecuencias observadas menores a las esperadas en caso de independencia. El color azul indica frecuencias observadas mayores a las esperadas en caso de independencia.



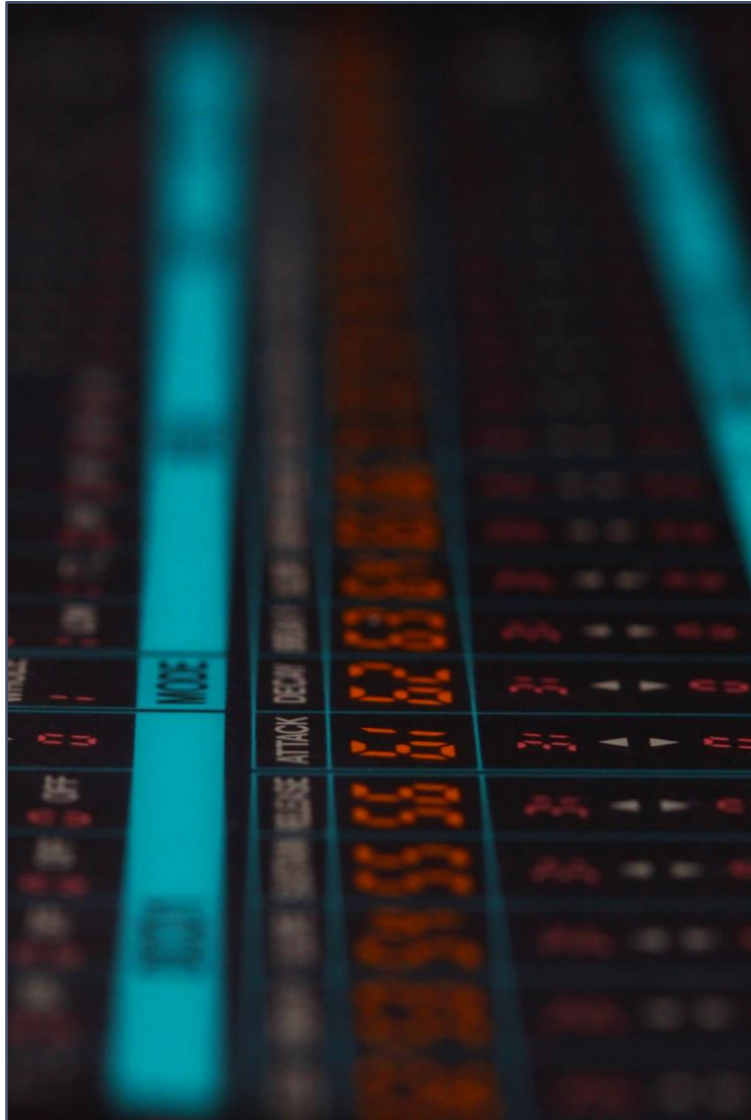
- Para contrastar estadísticamente la existencia de independencia entre dos atributos o factores (o, por el contrario, admitir asociación), existe la prueba basada en el **coeficiente de contingencia ji-cuadrado**.
- El coeficiente es una medida de la diferencia existente entre la tabla de contingencia observada y la que debería darse en caso de independencia (teórica):

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

- Bajo H0 (independencia), el coeficiente se distribuye:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \rightarrow \chi^2_{(r-1)(c-1)}$$

con: n_{ij} frecuencias observadas; E_{ij} frecuencias teóricas; r número de filas; c



- En nuestro ejemplo:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - E_{ij})^2}{E_{ij}} = 17,8944$$

$$\chi^2_{(r-1)(c-1)} = \chi^2_{(2-1)(2-1)} = \chi^2_1$$

$$P\text{-valor} = 0,00002335$$

Luego **se rechaza H0** (independencia) para una significación de 0,05, y **se admite asociación** entre los atributos o factores “situación laboral” y situación académica.



- Una vez se ha constatado que existe asociación entre dos atributos o factores, cabe preguntarse **cómo** se materializa esa asociación.
- Una aproximación puede obtenerse mediante el **análisis de correspondencias**.
- Este análisis permite **visualizar una tabla de contingencia** en un espacio bidimensional.
- Puede entenderse con el análisis de componentes principales aplicado a atributos o factores.



- Una vez se ha constatado que existe asociación entre dos atributos o factores, cabe preguntarse **cómo** se materializa esa asociación.
- Una aproximación puede obtenerse mediante el **análisis de correspondencias**.
- Este análisis permite **visualizar una tabla de contingencia** en un espacio bidimensional.
- Puede entenderse con el análisis de componentes principales aplicado a atributos o factores.
- Se profundizará en el **análisis de correspondencias simple**, en el que se analizan solo **2 atributos** o factores.



- Los **ejes o dimensiones** del análisis de correspondencias juegan el papel de las **componentes** en el análisis de componentes principales. Como el objetivo es visualizar gráficamente la tabla de contingencia, los ejes o dimensiones retenidas son 2.
- Los ejes se obtienen como aquellos que recogen la mayor proporción de comportamiento (varianza total o comunalidad) de los atributos o factores. En este análisis, esta varianza se llama ***Inercia Total***.
- Las dimensiones o ejes tienen sus ***inercias principales*** (la proporciones de la inercia total que recogen), que se corresponden con los autovalores de la denominada *matriz de residuos estandarizados*.



- El papel de las *cargas* de las variables en las componentes principales lo asumen las **contribuciones** de las categorías o niveles de los factores o atributos en los ejes o dimensiones.
- Las **coordenadas** de cada categoría o nivel equivalen a las puntuaciones o scores de cada individuo en el análisis de componentes principales.
- Como resultado, se obtiene, para cada atributo o factor, unas coordenadas para cada una de sus categorías o niveles en cada uno de los dos ejes o dimensiones.
- Al representar conjuntamente las categorías o niveles de ambos atributos o factores, se pueden establecer, **en general**, conclusiones sobre **el patrón de asociación** que existe.

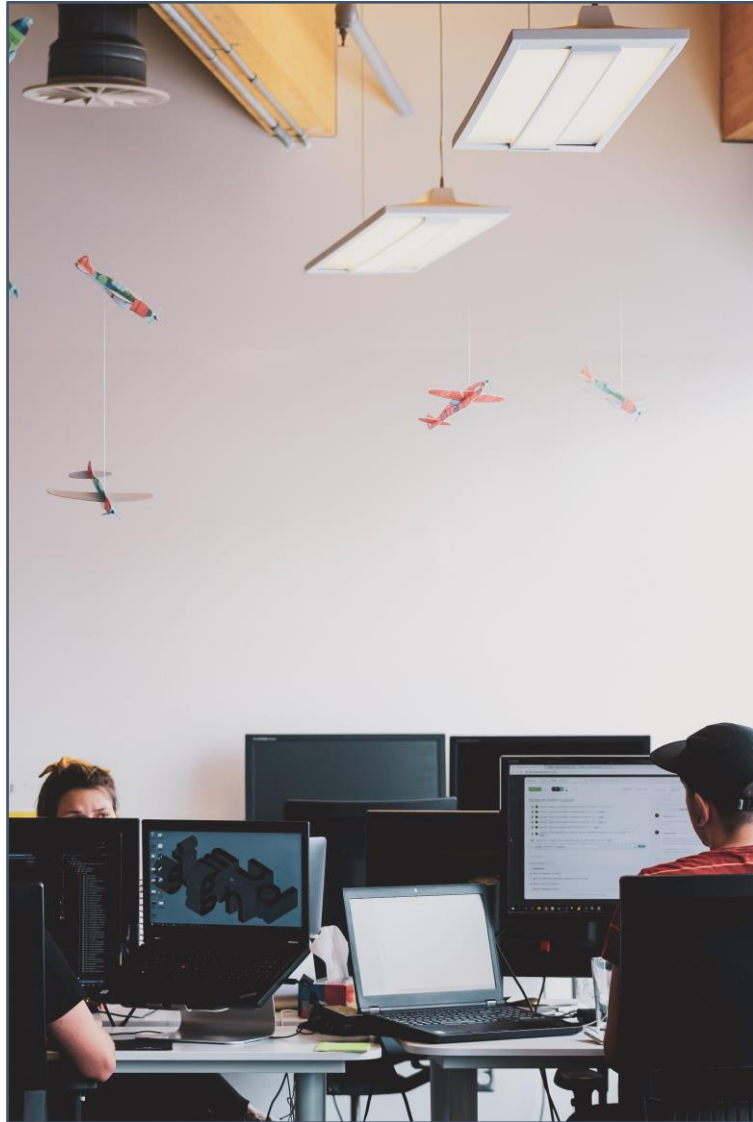


• Ejemplo:

- **DIMENSION:** tamaño del grupo empresarial al que pertenece la empresa en cuestión. Tiene tres niveles: grande, media y reducida.
- **AUTOFINA:** capacidad de autofinanciación de la empresa a medio y largo plazo. Tiene tres niveles: alta, positiva y negativa.

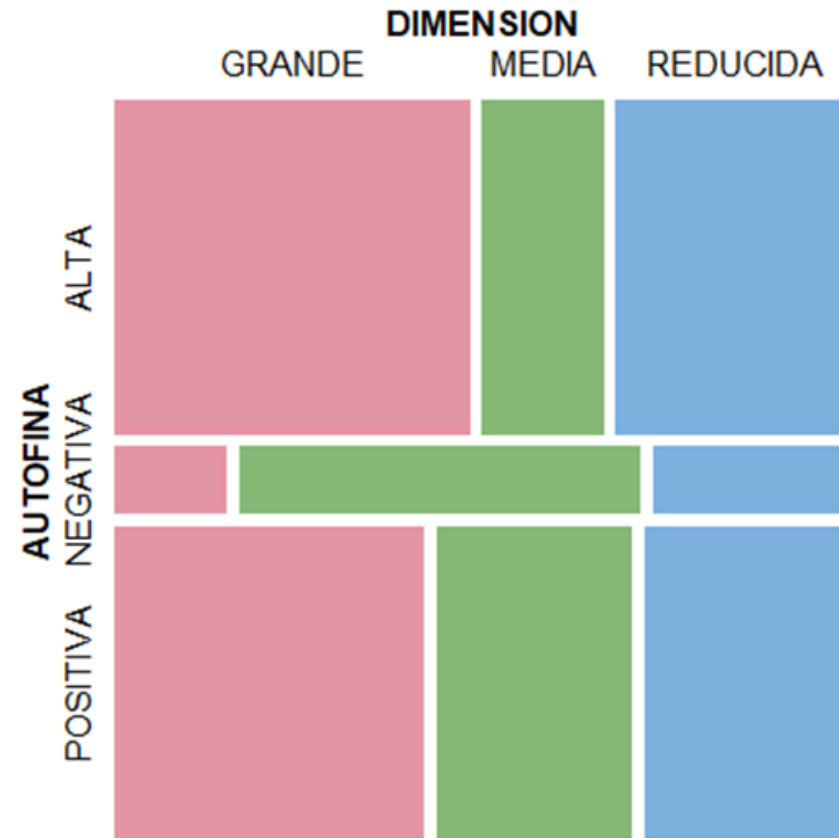
Empresas eólicas

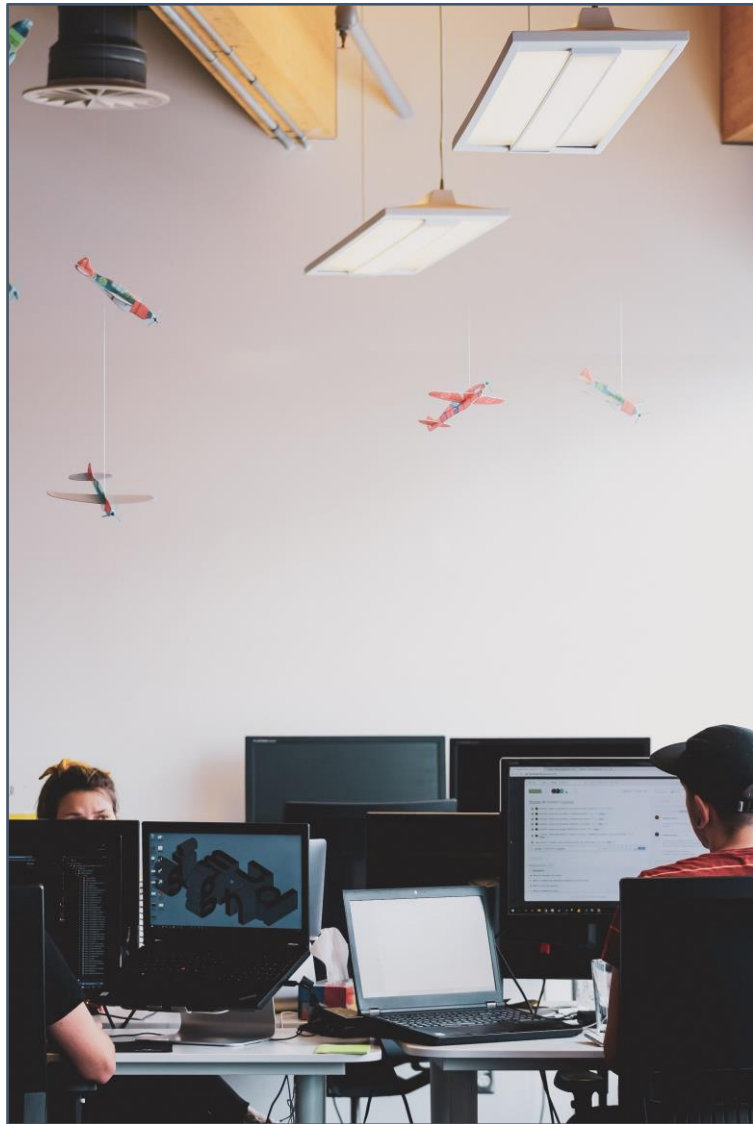
AUTOFINANCIACION	DIMENSION			Sum
	GRANDE	MEDIA	REDUCIDA	
ALTA	79	27	53	159
NEGATIVA	5	18	9	32
POSITIVA	65	41	44	150
Sum	149	86	106	341



- **Ejemplo:**

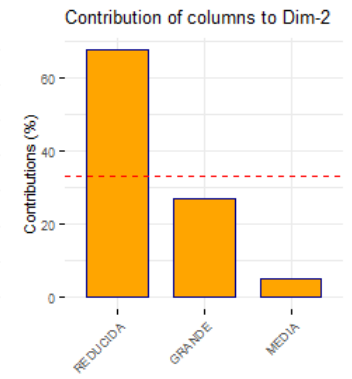
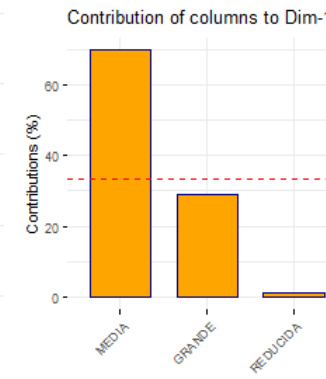
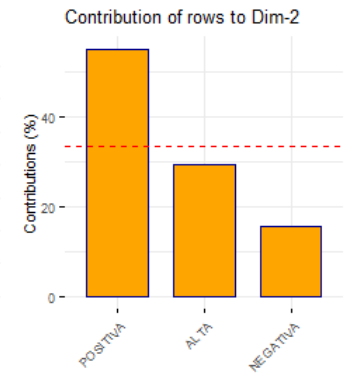
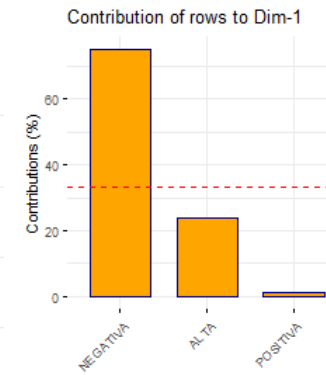
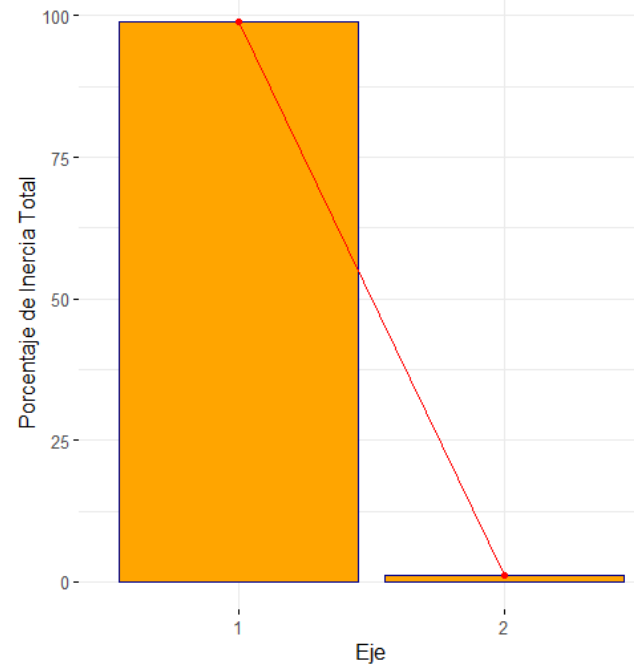
Eólicas: Capac. autofinanciación y Dim. grupo empresarial.





• Ejemplo:

Contribución de los ejes a la Inercia Total.
Eólicas: C. Autofinanciación y Dimensión grupo empresarial.

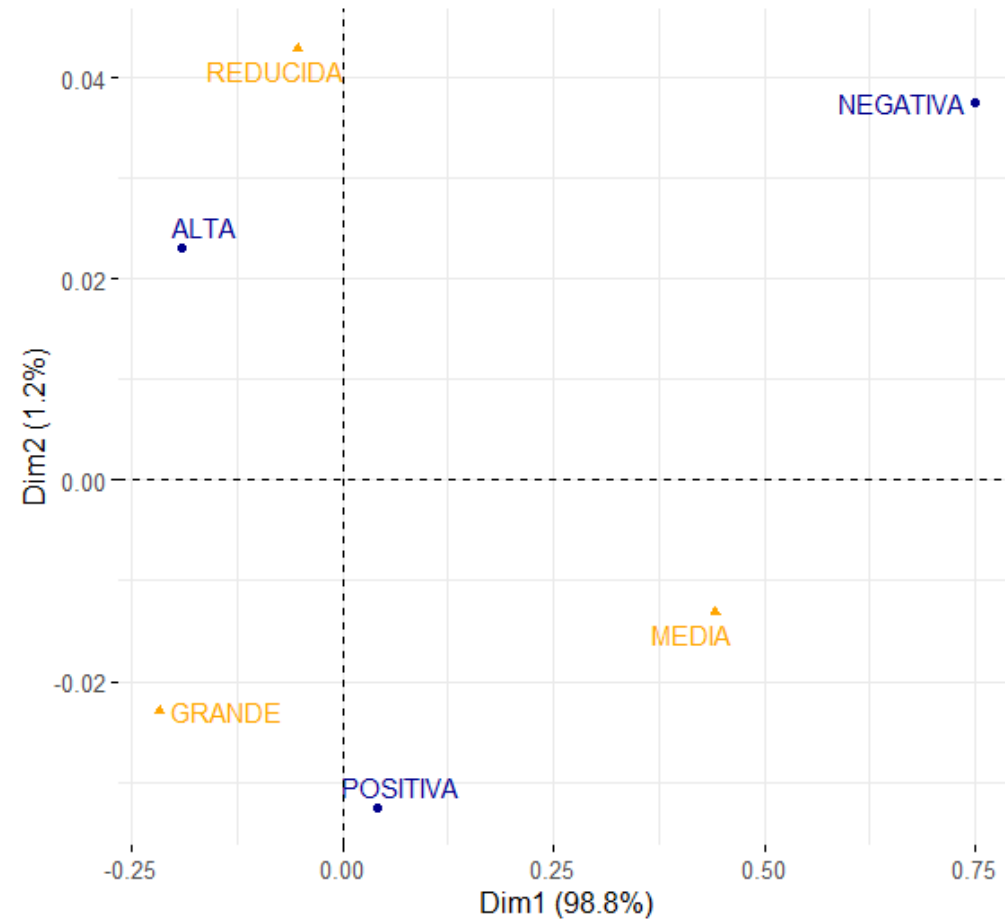




- **Ejemplo:**

Gráfico de dispersión de categorías.

Eólicas: C. Autofinanciación y Dimensión grupo empresarial.





¡Muchas gracias!

This work © 2022 by [Miguel Ángel Tarancón](#) and [Consolación Quintana](#) is licensed under [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](#).

