

## Tablas de contingencia de 2 dimensiones. Asociación. Análisis de correspondencias simple.



### Introducción.

En esta práctica vamos a explorar algunos métodos para extraer información útil cuando los datos de los que disponemos son categóricos, es decir, vienen dados en forma de factores o variables cualitativas o categóricas (atributos).

En estos casos, la información suele sintetizarse y presentarse mediante las denominadas “tablas de contingencia”. En este tipo de tablas, se muestran las frecuencias conjuntas, es decir, el número de casos que comparten los distintos niveles o categorías de los diferentes factores.

Cuando se trabaja con varios factores, representados en su correspondiente tabla de contingencia, uno de los análisis más interesantes es determinar si existe asociación entre los factores. Esto es, si se aprecia algún tipo de relación estadística entre estas variables categóricas o cualitativas, en el sentido de si se puede afirmar que el hecho de que los individuos de la muestra o población tomen ciertos niveles o categorías en unos factores hace que tiendan a tomar ciertos niveles o categorías de otro u otros factores.

Además, es interesante efectuar un “análisis de correspondencias”, que es una técnica de reducción de la dimensión de la información, similar al análisis de componentes principales, pero en el caso de atributos o factores. El objetivo del análisis de correspondencias es representar gráficamente, en un espacio con pocas dimensiones (preferentemente 2) las tablas de contingencia que muestran las relaciones entre dos variables cualitativas, factores o atributos.

Vamos a aplicar los conceptos anteriores al caso de las empresas eólicas españolas, sobre las cuales se ha extraído información cualitativa concretada en los siguientes factores o atributos:

- 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.

## Preparando Datos.

Abriremos **R-Studio** y crearemos nuestro proyecto siguiendo la instrucción **File → New Project**. Nos preguntará si crea el proyecto en una nueva carpeta o en una ya existente. Vamos a crearlo, por ejemplo, en el disco extraíble D, carpeta R, subcarpeta “correspondencias”, que ya existe. Aparecerá una ventana para buscar la carpeta y, tras localizarla, pulsaremos **Open** y **Create Project**.

Vamos a ir a la carpeta del proyecto y vamos a guardar en ella los dos archivos de esta práctica: un archivo de **Microsoft® Excel®** llamado “eolica\_contingencia.xlsx” y un *script* denominado “correspondencias\_eolica.R”. Si abrimos el archivo de **Microsoft® Excel®**, comprobaremos que se compone de tres hojas. La primera muestra el criterio de búsqueda de casos en la base de datos **Sabi®**; la segunda recoge la descripción de las variables consideradas, y la tercera (hoja “Datos”) guarda los datos que debemos importar desde **R-Studio**. Estos datos se corresponden con diferentes variables económico-financieras de las empresas productoras de electricidad mediante generación eólica.

Luego vamos a cerrar el archivo de **Microsoft® Excel®** y volveremos a **R-Studio**. Vamos a abrir nuestro script “correspondencias\_eolica.R” con **File → Open File...** Este script contiene el programa que vamos a ir ejecutando en la práctica.

La primera línea / instrucción en el script es:

```
rm(list = ls())
```

La instrucción tiene como objeto limpiar el *Environment* de objetos de anteriores sesiones de trabajo. Para importar los datos, ejecutaremos el código:

```
## Importando datos  
  
library (readxl)  
eolicas <- read_excel("eolica_contingencia.xlsx", sheet = "Datos")
```

Podemos observar cómo, en el *Environment*, ya aparece un objeto. Este objeto es una estructura de datos tipo *data frame*, se llama “eolicas” y contiene 19 columnas, una por cada variable del archivo de Microsoft® Excel®.

R ha considerado la primera columna (NOMBRE) como una variable de tipo cualitativo. En realidad, no es una variable, sino el nombre de los casos (empresas). Para evitar que R tome los nombres de los individuos como una variable, podemos redefinir nuestro *data frame* diciéndole que tome esa primera columna como los nombres de los individuos o casos (filas):

```
eolicas <- data.frame(eolicas, row.names = 1)
```

En la línea anterior hemos asignado al *data frame* “eolicas” los propios datos de “eolicas”; pero indicando que la primera columna de datos no es una variable; sino el nombre de los individuos, casos o filas. Así, vemos que ya no aparece NOMBRE como variable, y en el *Environment* el *data frame* “eolicas” cuenta con 18 variables (una menos). Con `summary()` comprobamos las variables que contiene el *data frame*, y sus características:

```
summary (eolicas)
```

En nuestro análisis solo vamos a considerar los factores o atributos AUTOFINA y DIMENSION. Por ello, crearemos con ellas un nuevo *data frame* llamado, por ejemplo, “originales”:

```
## Seleccionando factores/atributos para el analisis

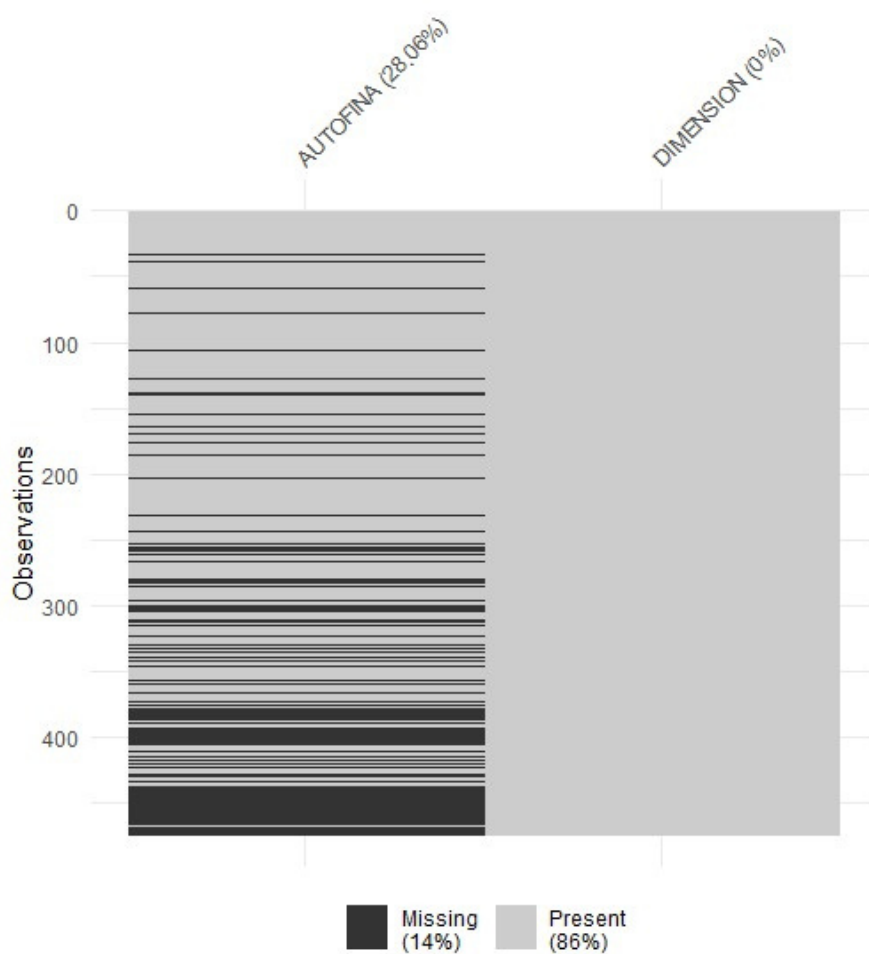
library(dplyr)
originales<-select(eolicas, AUTOFINA, DIMENSION)
summary (originales)
```

El siguiente paso será localizar los posibles *missing values*, ya que para realizar el análisis es necesario que todos los casos posean dato para todas las variables originales. Para tener una idea general, se puede utilizar la función `vis_miss()` del paquete `visdat`, que nos localizará gráficamente los *missing values* de las diferentes variables, y calculará el porcentaje de casos que supone, con respecto al total de observaciones:

```
## Identificando missing values.

library(visdat)
vis_miss(originales)
```

El resultado del código anterior es el siguiente gráfico:



Se observa cómo el atributo DIMENSION carece de *missing values*. En cambio, el 28% de los casos carece de dato en el factor AUTOFINA, luego tendrán que ser eliminados. Para localizarlos, podemos filtrar nuestro *data frame* con las herramientas de [dplyr](#):

```
originales %>% filter(is.na(AUTOFINA) | is.na(DIMENSION)) %>%
select(AUTOFINA, DIMENSION)
```

Los casos con *missing values* que aparecen en pantalla deberían ser analizados para estudiar la posibilidad de obtener los datos faltantes mediante la consulta de otras fuentes o algún proceso de estimación. Si no es posible, como supondremos en el ejemplo, tendrán que ser eliminados, por ejemplo, con el código:

```
originales <- originales %>%
filter(! is.na(AUTOFINA) & ! is.na(DIMENSION))
```

Como resultado, el *data frame* “originales” pasa a tener 341 casos (133 casos eliminados).

Al trabajar con atributos o factores, no tiene sentido realizar la detección de *outliers*.

### Construcción y descripción de la tabla de contingencia.

El primer paso del análisis es construir la tabla de contingencia a partir de la información contenida en nuestro *data frame* “originales”. La función para convertir datos del *data frame* en una estructura de almacenamiento de datos especial llamada *table* (que es la tabla de contingencia) es precisamente `table()`:

```
## Construyendo Tabla de contingencia
tab.originales <- table(originales)
```

En el *Environment* aparecerá el objeto “tab.originales”, de tipo *table*. Para representarla, incluyendo sus frecuencias marginales, podremos utilizar los recursos de los paquetes `knitr` y `kableExtra`, mediante el código:

```
## Construyendo Tabla de contingencia

library(knitr)
library(kableExtra)
knitr.table.format = "html"

addmargins(tab.originales) %>%
kable(caption="Empresas eólicas") %>%
  kable_styling(full_width = F, bootstrap_options = "striped",
"bordered", "condensed", position = "center", font_size = 12) %>%
  add_header_above(c("AUTOFINANCIACION"= 1, DIMENSION=3, " "=1), bold=T,
line=T) %>%
row_spec(0, bold= T, align = "c") %>%
column_spec(1, bold = T)
```

El resultado será:

### Empresas eólicas

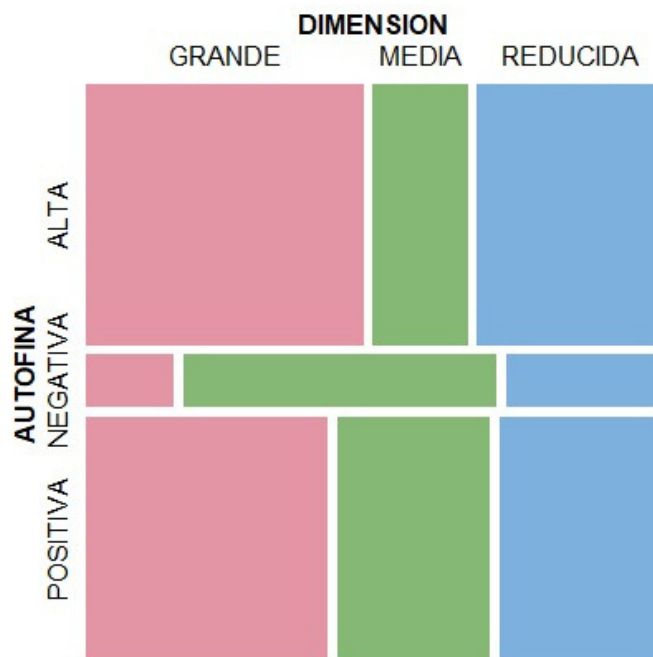
AUTOFINANCIACION	DIMENSION			Sum
	GRANDE	MEDIA	REDUCIDA	
ALTA	79	27	53	159
NEGATIVA	5	18	9	32
POSITIVA	65	41	44	150
Sum	149	86	106	341

La tabla también se puede representar gráficamente mediante la función `mosaic()` de la librería `vcd`, con lo que se percibirán mejor la magnitud de las frecuencias conjuntas (celdas de la tabla):

```
library (vcd)
mosaic(tab.originales,
       main="Eólicas: Capac. autofinanciación y Dim. grupo
empresarial.",
       sub = "Tabla de contingencia",
       shade=T,
       gp= shading_Marimekko(tab.originales),
       main_gp = gpar(fontsize = 14), sub_gp = gpar(fontsize = 12))
```

Obteniéndose la figura:

Eólicas: Capac. autofinanciación y Dim. grupo empresarial.



Puede observarse que el mayor número de frecuencias se da para las empresas pertenecientes a grupos empresariales de tamaño grande y con una capacidad de autofinanciación a medio y largo plazo alta. En cambio, el menor número de casos es el de empresas eólicas pertenecientes a grupos empresariales de tamaño grande y capacidad de autofinanciación a medio y largo plazo negativa.

El mismo gráfico se puede obtener con los atributos transpuestos, de manera que las categorías de DIMENSION tengan la misma altura para las diferentes categorías de AUTOFINA. Para ello, hemos de calcular la crear transpuesta con la función `t()`, a la que llamaremos, por ejemplo, “transpuesta”, y representarla con el mismo código de la función `mosaic()`:

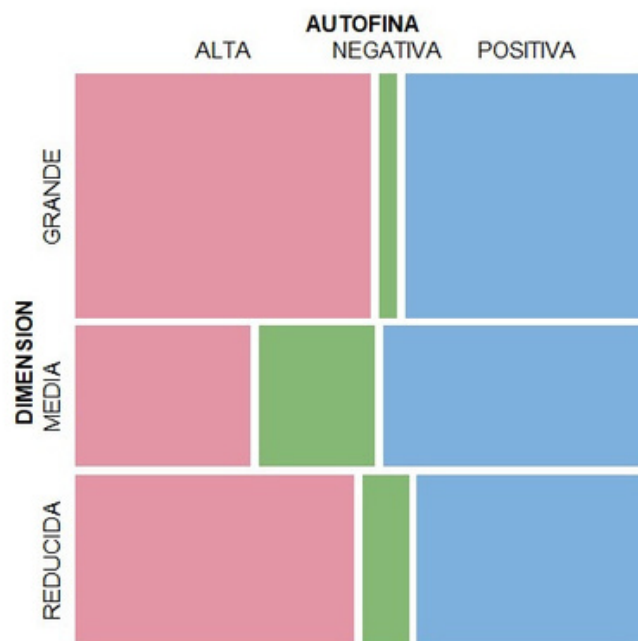
```

transpuesta = t(tab.originales)
transpuesta
mosaic(transpuesta,
       main="Eólicas: Capac. autofinanciación y Dim. grupo empresarial
(2).",
       shade=T,
       gp= shading_Marimekko(transpuesta),
       main_gp = gpar(fontsize = 14), sub_gp = gpar(fontsize = 12))

```

El resultado es:

Eólicas: Capac. autofinanciación y Dim. grupo empresarial (2).



Para percibir mejor las frecuencias marginales de cada nivel o categoría de ambos atributos o factores, pueden representarse estas mediante gráficos de barras. Para conseguirlo, vamos a utilizar el paquete gráfico `ggplot2`, y el paquete `patchwork`, que simplemente permite maquetar varios gráficos generados con `ggplot2` para que aparezcan juntos, en una determinada disposición. El código es el siguiente:

```
## Representando frecuencias de categorías en factores

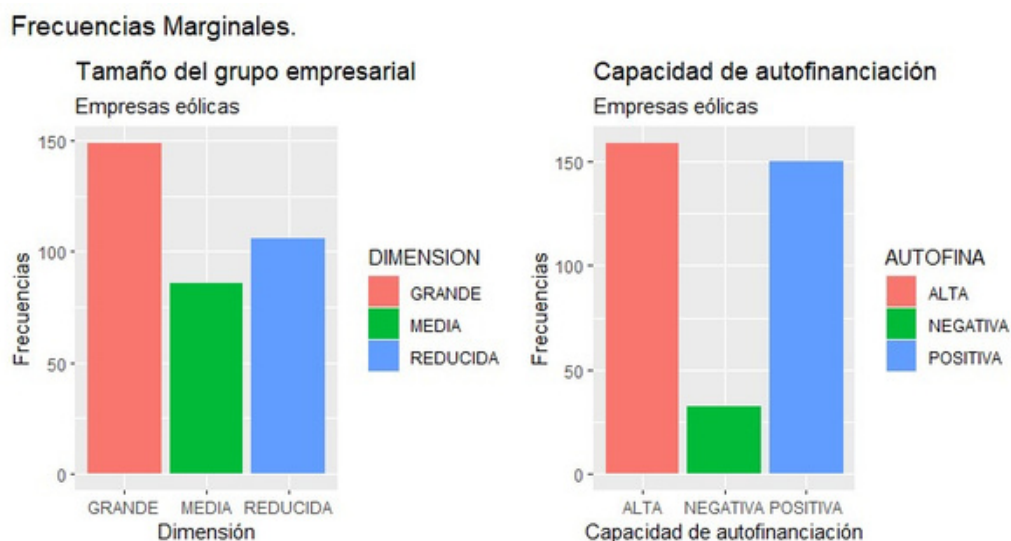
library (ggplot2)
library (patchwork)

g1 <- ggplot(originales, mapping= aes(x= DIMENSION, fill = DIMENSION)) +
+
geom_bar() +
  ggtitle("Tamaño del grupo empresarial", subtitle = "Empresas eólicas")
+
ylab("Frecuencias") +
xlab("Dimensión")

g2 <- ggplot(originales, mapping= aes(x= AUTOFINA, fill = AUTOFINA)) +
geom_bar() +
  ggtitle("Capacidad de autofinanciación", subtitle = "Empresas
eólicas") +
ylab("Frecuencias") +
xlab("Capacidad de autofinanciación")

(g1 + g2) + plot_annotation(title = "Frecuencias Marginales.",
                           theme = theme(plot.title =
element_text(size = 14)))
```

El resultado obtenido es:



Puede observarse cómo, en cuanto al tamaño del grupo empresarial de pertenencia (DIMENSION), son mayoritarias las empresas pertenecientes a



grupos de gran volumen. En cuanto a la capacidad de autofinanciación a largo y medio plazo (AUTOFINA), las de capacidad alta o positiva presentan frecuencias similares; y son muy minoritarias las empresas con capacidad negativa.

### Asociación.

Uno de los aspectos más interesantes, a la hora de analizar una tabla de contingencia bidimensional, es estudiar si existe asociación entre ambos factores o atributos, es decir, si existe cierta tendencia a acumularse

frecuencias

en ciertas intersecciones entre categorías y niveles de uno y otro.

Para saber si existe asociación o no entre ambos factores, la idea es comparar mediante una medida estadística lo “cercana” o “lejana” que está nuestra tabla de contingencia con respecto a la tabla teórica que tendría que darse (con la misma frecuencia total o número de casos) en caso de independencia entre

factores o atributos. Este es el cometido de la *prueba de independencia Chi-cuadrado* entre las categorías de ambas variables cualitativas. La función que realiza dicha prueba es `chisq.test()`.

```
#TEST DE INDEPENDENCIA / ASOCIACION
```

```
chisq.test(tab.originales)
```

El resultado de la prueba es:

```
Pearson's Chi-squared test
```

```
data: tab.originales
```

```
X-squared = 24.34, df = 4, p-value = 6.828e-05
```

El p-valor de la prueba es menor que 0.05, por lo que rechazamos la hipótesis nula de independencia: hay cierta relación estadística entre las categorías de los atributos (asociación).

La utilización del coeficiente *Chi-cuadrado* tiene una debilidad a la hora de “medir” el grado de asociación entre los dos factores: no está acotado entre 0 y 1. Una medida alternativa que sí está acotada entre 0 y 1 es el *coeficiente V de Cramer*, que puede obtenerse con la función `CramerV()` del paquete `DescTools`:

```
library (DescTools)
```

```
CramerV(tab.originales)
```

El valor del coeficiente es:

```
[1] 0.1889148
```

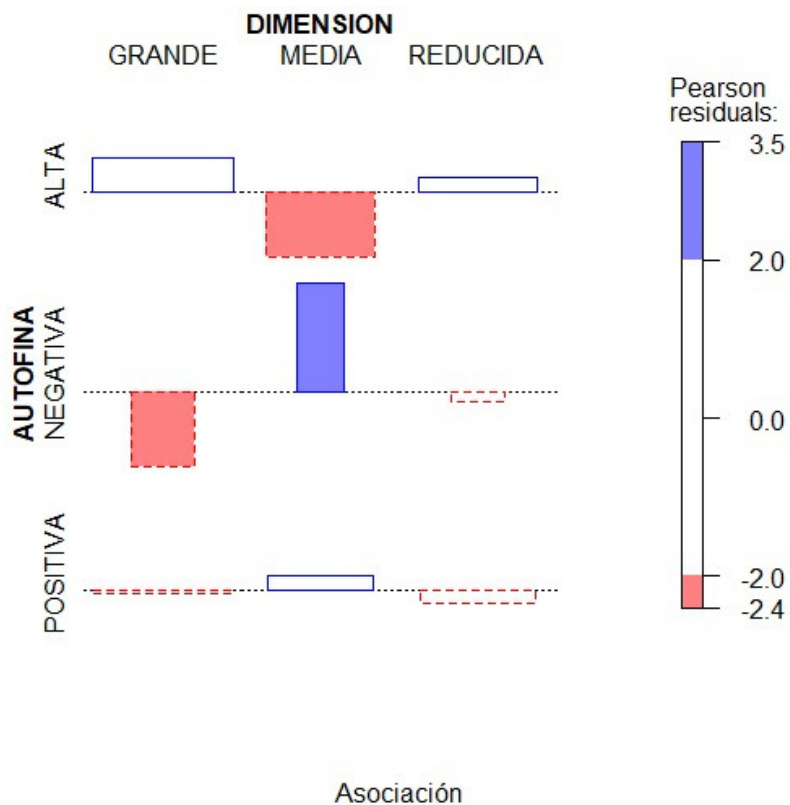
El valor anterior indica que, aunque hay asociación según la prueba de independencia *Chi-cuadrado*, esta asociación es moderada.

Precisamente, se pueden mostrar gráficamente las desviaciones o diferencias (residuos tipificados) entre las frecuencias de la tabla observada con respecto a las frecuencias que deberían existir en el caso teórico de independencia entre tales atributos o factores. Para ello, cabe utilizar la función `assoc()` del paquete `vcd`:

```
assoc(tab.originales,  
      main="Eólicas: Capac. autofinanciación y Dim. grupo empresarial.",  
      sub = "Asociación",  
      compress= F,  
      gp= shading_Friendly(tab.originales),  
      main_gp = gpar(fontsize = 14), sub_gp = gpar(fontsize = 12),  
      legend=TRUE)
```

Obteniéndose:

**Eólicas: Capac. autofinanciación y Dim. grupo empresarial.**



Se observa cómo en la combinación “DIMENSION media” con “AUTOFINA negativa”, los errores son elevados y con signo positivo (relleno azul), lo que indica que, en este caso, la frecuencia observada es sensiblemente superior a la que cabría esperar en caso de independencia. En cambio, en las combinaciones “DIMENSION media” con “AUTOFINA alta”, y “DIMENSION grande” con “AUTOFINA negativa”; los errores son apreciablemente altos, pero con signo negativo (relleno anaranjado); lo que se interpreta como que las frecuencias observadas son sensiblemente inferiores a las que debería haber en caso de independencia. En definitiva, cuantos más bloques “coloreados” hay en el gráfico de asociación, más intensa será la asociación entre ambos factores o atributos; y más lejos se estará de la situación de independencia estadística.

## Análisis de correspondencias.

Vamos a realizar el análisis de correspondencias, para representar en un

gráfico

bidimensional las diferentes categorías o niveles de los dos atributos o factores.

Lo realizaremos mediante la función `CA()` de la librería `FactoMineR`. Así, guardamos el resultado en el objeto “`aceolicas`”, y hacemos un `summary()`:

```
aceolicas<-CA(X = tab.originales, graph = F)
summary (aceolicas)
```

Obteniéndose la siguiente información:

Call:

```
CA(X = tab.originales, graph = F)
```

```
The chi square of independence between the two variables is equal to 24.33975
(p-value = 6.827655e-05 ).
```

Eigenvalues

```
Dim.1 Dim.2
```

```
Variance 0.071 0.001
```

```
% of var. 98.812 1.188
```

```
Cumulative % of var. 98.812 100.000
```

Rows

```
Iner*1000 Dim.1 ctr cos2 Dim.2 ctr cos2
ALTA | 17.137 | -0.190 23.944 0.985 | 0.023 29.428 0.015 |
NEGATIVA | 53.014 | 0.751 74.977 0.997 | 0.038 15.639 0.003 |
POSITIVA | 1.226 | 0.042 1.079 0.620 | -0.033 54.933 0.380 |
```

Columns

```
Iner*1000 Dim.1 ctr cos2 Dim.2 ctr cos2
GRANDE | 20.765 | -0.217 29.114 0.989 | -0.023 27.191 0.011 |
MEDIA | 49.156 | 0.441 69.634 0.999 | -0.013 5.146 0.001 |
REDUCIDA | 1.457 | -0.053 1.252 0.606 | 0.043 67.663 0.394 |
```

La función vuelve a realizar el test o prueba de independencia *Chi-cuadrado*. Luego, dispone tres tablas.

La primera, *Eigenvalues* (autovalores o valores propios) muestra la “inercia principal” o varianza de cada dimensión o eje del espacio bidimensional donde se van a representar las categorías o niveles de los factores o atributos; la proporción o porcentaje del comportamiento o varianza de las categorías o inercia total recogido, y su acumulación. Puede observarse cómo la primera dimensión o eje recoge el 98,8% del comportamiento (inercia total, varianza) de las categorías o niveles de los factores o atributos; mientras que la segunda dimensión recoge apenas el 1,2%. Entre las dos, recogen el 100% del comportamiento (inercia total) de las categorías o niveles.

La segunda tabla, *Rows*, contiene, para cada eje o dimensión:

- La varianza o inercia (multiplicada por mil) recogida por los ejes o dimensiones de cada categoría o nivel del factor o atributo dispuesto en filas (AUTOFINA).
- Las coordenadas de cada nivel o categoría del atributo o factor fila (“tipos de redes”), “Dim”.
- La contribución de cada categoría o nivel al eje o dimensión, “ctr”, en porcentaje (la suma de las contribuciones es 100).
- La calidad de la representación de cada categoría o nivel de los factores, “cos2”, en escala 0-1, en cada eje o dimensión (la suma de los “cos2” de las dos dimensiones para cada categoría del factor fila es 1).

De este modo, centrándonos en el atributo factor dispuesto en filas (AUTOFINA), y en cuanto al primer eje (“Dim.1”, eje “X”), puede comprobarse que, en su formación, el mayor peso relativo lo ha tenido la categoría o nivel “NEGATIVA”, con un 75% de contribución; mientras que el menor peso lo tiene la categoría o nivel “POSITIVA”, con un escaso 1%. En correspondencia a ello, la mejor calidad de representación con primer eje o dimensión se da para la categoría o nivel “NEGATIVA” (0,997 de “cos2”), seguido de “ALTA” (0,985 de “cos2”) y, en un modo muy inferior, “POSITIVA” (“cos2” de 0,62). En cuanto al segundo eje (“Dim.2”, eje “Y”), la mayor contribución o peso en tal eje lo tiene la categoría de AUTOFINA “POSITIVA” (55%), seguida de “ALTA” (29,4%) y, por último, “NEGATIVA” (15,6%). Como ya se vio en la primera tabla, este segundo eje representa bastante peor que el primer eje el comportamiento de las categorías o niveles de los factores o atributos. En concreto, para el caso de AUTOFINA, y según la medida “cos2”; 0,38 para la categoría o nivel “POSITIVA”, 0,015 para “ALTA”, y 0,003 para “NEGATIVA”.

La tercera tabla, *Columns*, contiene la misma información que la tabla anterior; pero para el caso del atributo o factor dispuesto en columnas, en la tabla de contingencia (atributo o factor DIMENSION).

Así, en el atributo o factor dispuesto en columnas (DIMENSION), y en cuanto al primer eje ("Dim.1", eje "X"), puede comprobarse que, en su formación, el mayor peso relativo lo ha tenido la categoría o nivel "MEDIA", con un 69,6%

de

contribución; mientras que el menor peso lo tiene la categoría o nivel "REDUCIDA", con algo más del 1,2%. En correspondencia a ello, la mejor calidad de representación con primer eje o dimensión se da para la categoría o nivel "MEDIA" (0,999 de "cos<sup>2</sup>"), seguido de "GRANDE" (0,989 de "cos<sup>2</sup>") y, en un modo muy inferior, "REDUCIDA" ("cos<sup>2</sup>" de 0,606). En cuanto al segundo eje ("Dim.2", eje "Y"), la mayor contribución o peso en tal eje lo tiene la

categoría

de DIMENSION "REDUCIDA" (67,7%), seguida de "GRANDE" (27,2%) y, por último, "MEDIA" (5,2%). Como ya se vio en la primera tabla, este segundo eje representa bastante peor que el primer eje el comportamiento de las categorías o niveles de los factores o atributos. En concreto, para el atributo o factor DIMENSION, y según la medida "cos<sup>2</sup>"; 0,394 para la categoría o nivel "REDUCIDA", 0,011 para "GRANDE", y 0,001 para "MEDIA".

Puede observarse como, atendiendo a su proximidad en el gráfico, hay cierta asociación entre que el negocio se estructure según "otra colaboración" y que

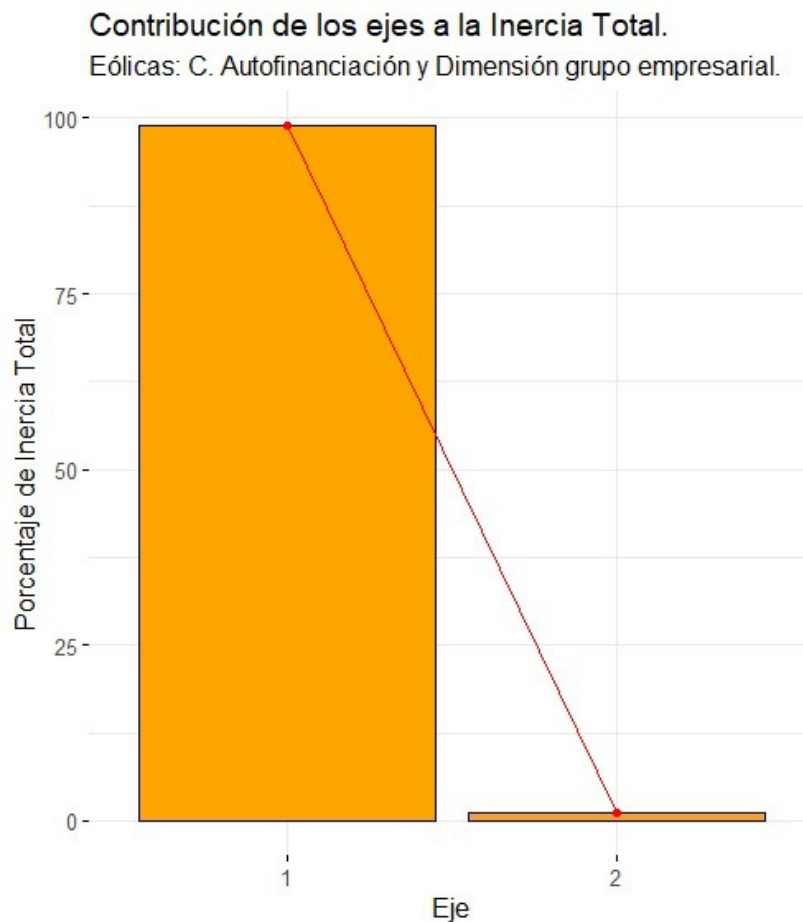
no

se usen redes sociales en la actividad de la empresa. Por otro lado, los negocios individuales tienden a usar todo tipo de redes sociales, y las franquicias tienden tanto a centrarse en las redes sociales clásicas, como también a usar todo tipo de redes.

Si se quieren representar gráficamente los porcentajes de *inercia total* (contribución o varianza de los casos) de cada eje, se puede usar la función `gcontrib` de la librería `factoextra`:

```
library(factoextra)
gcontrib(
  title = "Contribución de los ejes a la Inercia Total.", subtitle = "Eólicas: C. Autofinanciación y Dimensión grupo empresarial." +
  ylab("Porcentaje de Inercia Total") +
  xlab("Eje") +
  theme(text = element_text(size = 12))
)
```

Se obtiene el siguiente gráfico de barras, en el que, en nuestro caso, puede comprobarse cómo la primera dimensión recoge casi un 98,8% de la inercia principal, dejando para una segunda dimensión tan solo un 1,2%.



Además, se pueden obtener los gráficos de las diferentes contribuciones de las categorías o niveles a ambas dimensiones o ejes, entendidas como el peso que tienen cada una de ellas en tal dimensión o eje (similar a las “cargas” que tienen las variables en cada componente, en un análisis de *componentes principales*) con la función `fviz_contrib()` del paquete `factoextra`.

```
## Contribución filas a Dimensiones  
  
grow_dim1 <- fviz_contrib(aceolicas,  
  choice = "row",  
  axes = 1,  
  fill = "orange",  
  color = "darkblue") +  
theme(text = element_text(size = 8))
```

```

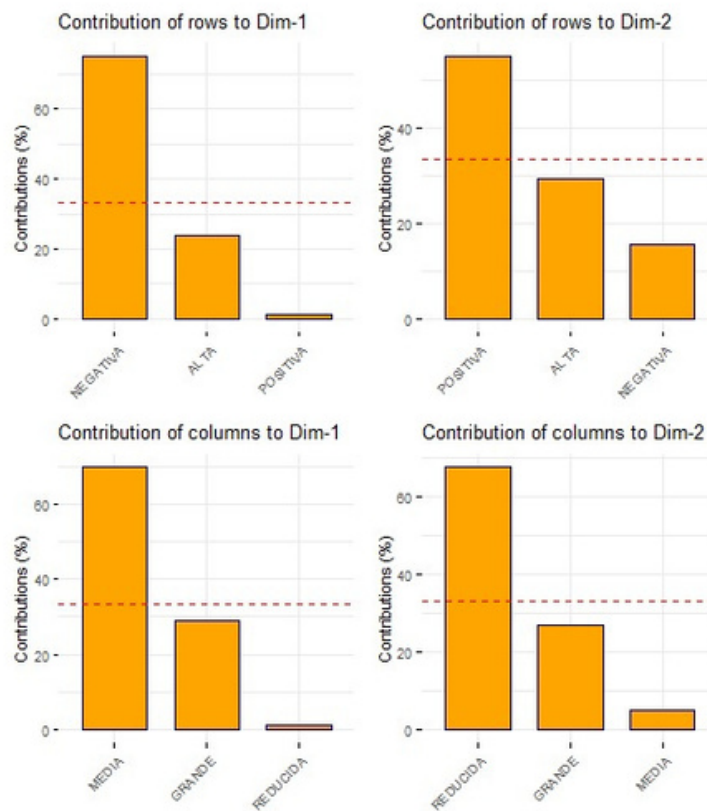
grow_dim2 <- fviz_contrib(aceolicas,
  choice = "row",
  axes = 2,
  fill = "orange",
  color = "darkblue") +
  theme(text = element_text(size = 8))

## Contribución columnas a Dimensiones

gcol_dim1 <-fviz_contrib(aceolicas,
  choice = "col",
  axes = 1,
  fill = "orange",
  color = "darkblue") +
  theme(text = element_text(size = 8))
gcol_dim2 <-fviz_contrib(aceolicas,
  choice = "col",
  axes = 2,
  fill = "orange",
  color = "darkblue") +
  theme(text = element_text(size = 8))
(grow_dim1 + grow_dim2) / (gcol_dim1 + gcol_dim2)

```

Los gráficos de barras obtenidos son:



En cuanto al atributo o factor dispuesto en filas (AUTOFINA), la categoría con mayor peso en el primer eje (Dim-1) es “NEGATIVA”, y la que menor peso posee es “POSITIVA”. En cambio, en cuanto al eje 2 (Dim-2), ocurre lo contrario, siendo la categoría o nivel “POSITIVA” la que tiene un peso apreciablemente superior al resto.

Respecto al atributo o factor dispuesto en columnas (DIMENSION), la categoría con mayor peso en el primer eje (Dim-1) es “MEDIA”, y la que posee un menor peso o contribución es “REDUCIDA”. En cambio, en cuanto al eje 2 (Dim-2), ocurre lo contrario, siendo la categoría o nivel “REDUCIDA” la que tiene un peso apreciablemente superior al resto.

Queda por construir el gráfico *biplot* o bidimensional en el que posicionar, según sus coordenadas, a las categorías o niveles de ambos factores. Para ello ejecutaremos el código, siguiente, basado en la función `fviz_ca_biplot()`: del paquete `factoextra`:

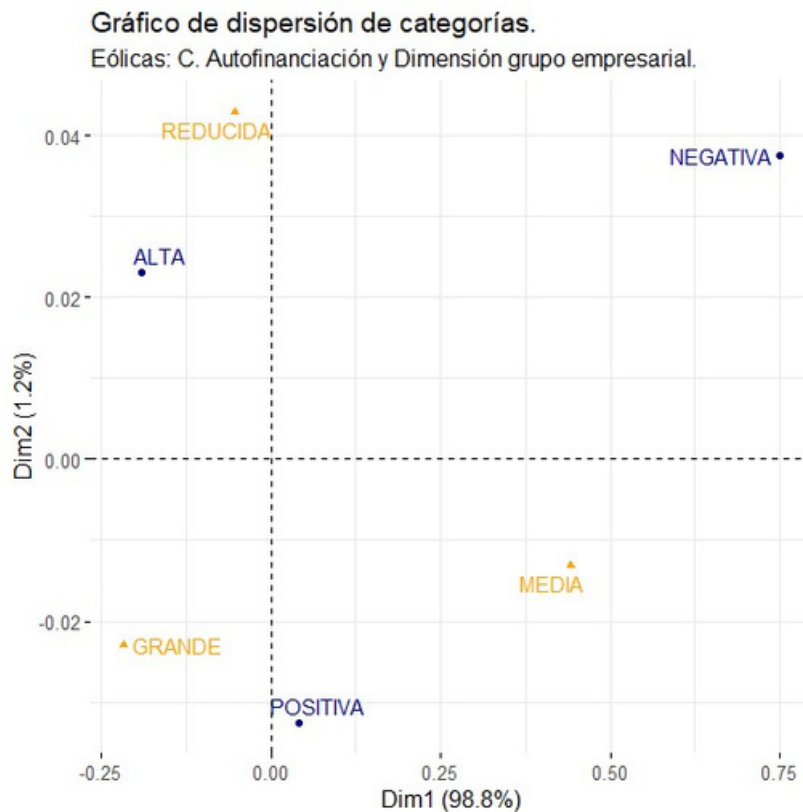
```
## Gráfico bidimensional

gbiplot <- fviz_ca_biplot (aceolicas,
  axes= c(1,2),
  label= "all",
  repel = T,
  col.col= "orange", col.row= "darkblue",
  map= "symmetric") +
  labs(title= "Gráfico de dispersión de categorías.", subtitle =
  "Eólicas: C. Autofinanciación y Dimensión grupo empresarial.") +
  theme(text = element_text(size = 12))

gbiplot
```

El resultado es el gráfico:





A la hora de interpretar este gráfico, hay que tener en cuenta que solo pueden interpretarse de modo concreto las distancias que separan a las categorías o niveles del factor o atributo fila (AUTOFINA) entre sí; e igual ocurre con las distancias que separan a las categorías o niveles del factor o atributo dispuesto en columnas (DIMENSION). Además, teniendo en cuenta que el primer eje (eje “X” o “Dim-1”) recoge la mayor parte de la información o Inercia puesta en juego (98,8%), habría que tenerse en cuenta, sobre todo, las distancias horizontales entre las categorías o factores.

En cuanto a las distancias entre categorías o niveles pertenecientes a factores o atributos distintos, tales distancias no se pueden interpretar de modo estricto, y solo se pueden extraer patrones de asociación generales.

En nuestro caso, se podría concluir que:

- Con respecto al atributo dispuesto en filas en la tabla de contingencia, AUTOFINA, y fijándonos sobre todo en la disposición de sus tres categorías o niveles a lo largo del primer eje (eje “X” o “Dim-1”), que recoge casi la totalidad del comportamiento; cabe decir que las capacidades de autofinanciación a medio y largo plazo “ALTA” y “POSITIVA” poseen un patrón de comportamiento similar (por ejemplo, en ambos casos las mayores frecuencias se dan para una dimensión del

grupo empresarial de pertenencia “GRANDE”, seguida de “MEDIA”, y finalmente “REDUCIDA”); mientras que la capacidad de autofinanciación “NEGATIVA” muestra un patrón apreciablemente diferente (las mayores frecuencias se dan para una dimensión del grupo empresarial de pertenencia “MEDIA”, seguida de “REDUCIDA”, y, finalmente, “GRANDE”)

- En relación con el atributo dispuesto en columnas en la tabla de contingencia, DIMENSION, y fijándonos sobre todo en la disposición de sus tres categorías o niveles a lo largo del primer eje (eje “X” o “Dim-1”), que recoge casi la totalidad del comportamiento; cabe decir que las capacidades las dimensiones del grupo empresarial de pertenencia “GRANDE” y “REDUCIDA” poseen un patrón de comportamiento similar (por ejemplo, en ambos casos las mayores frecuencias se dan para una capacidad de autofinanciación “ALTA”, seguida de “POSITIVA”, y finalmente “NEGATIVA”); mientras que la dimensión del grupo empresarial de pertenencia “MEDIA” muestra un patrón apreciablemente diferente (las mayores frecuencias se dan para una capacidad de autofinanciación “NEGATIVA”, seguida de “POSITIVA”, y, finalmente, “ALTA”)

- En general, se puede establecer cierta asociación entre pertenecer a un grupo empresarial de tamaño medio y poseer una capacidad de autofinanciación a medio y largo plazo negativa; y entre pertenecer a un grupo empresarial de tamaño “GRANDE” o “REDUCIDO” y mostrar una capacidad de autofinanciación a medio y largo plazo “ALTA” o “POSITIVA”.

Nota Final: Si se quieren integrar todos los gráficos del análisis de correspondencias simple, podrá hacerse uso de las facilidades del paquete `patchwork`:

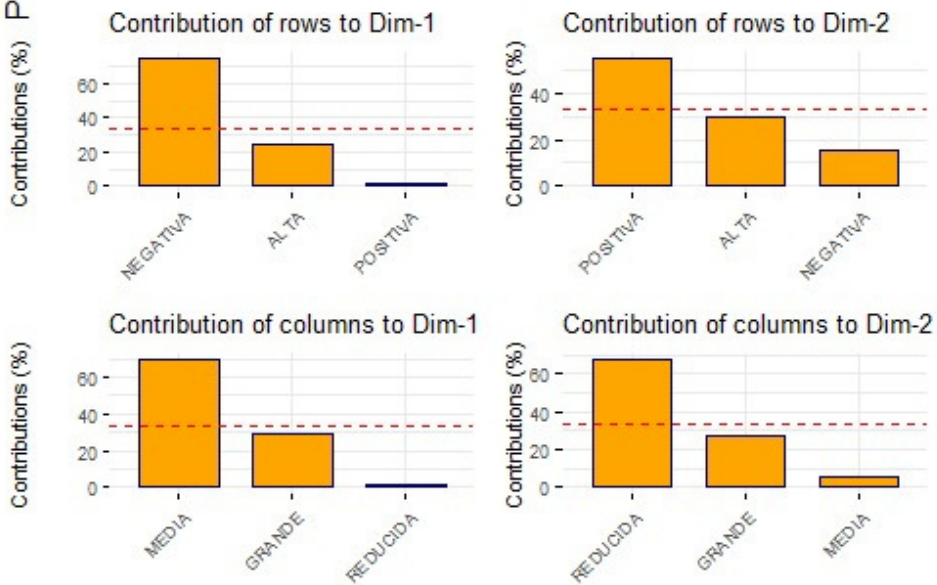
```
gcontrib / (grow_dim1 + grow_dim2) / (gcol_dim1 + gcol_dim2) / gbiplot
```

Obteniéndose la composición:

Contribución de los ejes a la Inercia Total

### Contribución de los ejes a la Inercia Total.

Eólicas: C. Autofinanciación y Dimensión grupo empresarial.



### Gráfico de dispersión de categorías.

Eólicas: C. Autofinanciación y Dimensión grupo empresarial.



This work © 2022 by Miguel Ángel Tarancón and Consolación Quintana is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License



Updated: 20/12/2022