

# Análisis factorial de correspondencias

Isabel Afán de Ribera

15/11/2020

## Introducción

En el presente trabajo se va a realizar un Análisis Factorial de correspondencias sobre los datos recopilados en una encuesta SocioMétrica por *El Español* en enero de 2018. Dicha encuesta fue realizada a 11.610 españoles mayores de edad en la que se les preguntaba sobre el partido político de preferencia y su situación laboral.

El *análisis de correspondencias* o *ANACOR* es una técnica de reducción de la dimensión que persigue estudiar la asociación entre dos o más variables categóricas. Consiste en tres etapas: análisis de independencia, análisis de correspondencia y reducción de la dimensión e interpretación.

## Objetivo

Pretendemos conocer si se puede establecer una relación entre la situación laboral de los encuestados y sus preferencias por partido político, para ello emplearemos el análisis de correspondencias.

## Descripción del dataset

El dataset consiste en 5 variables que conforman las columnas y 5 observaciones que conforman las filas y van referidas a la situación laboral. Las variables son:

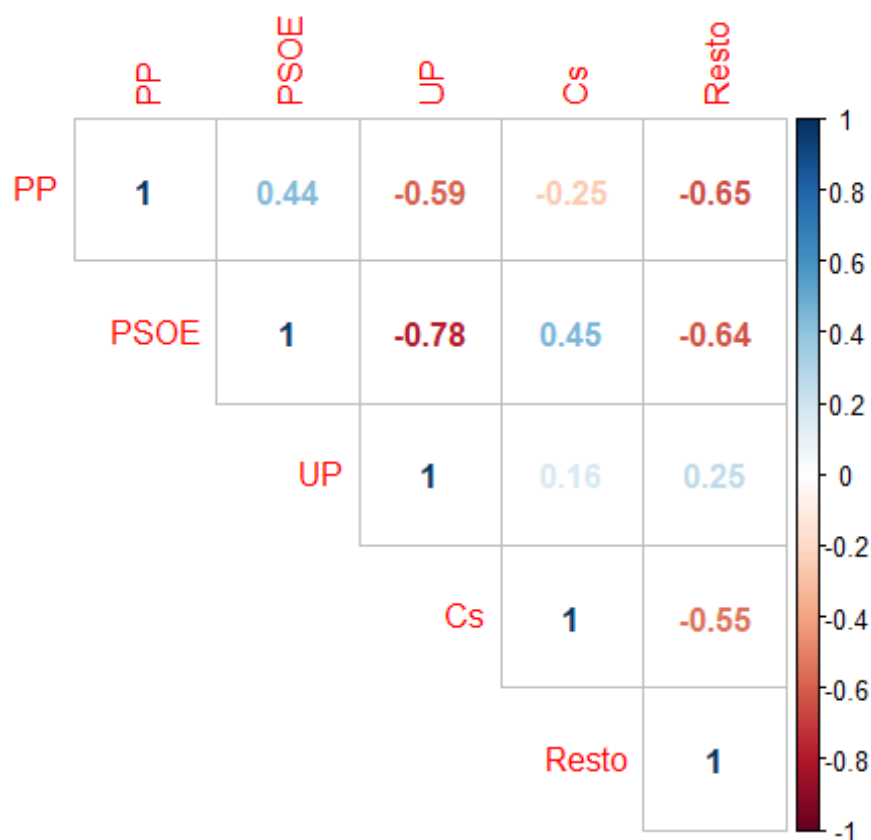
- PP: Partido Popular
- PSOE: Partido Socialista Obrero Español
- UP: Unidas Podemos
- Cs: Ciudadanos
- Resto: Otros partidos políticos

## Análisis exploratorio de los datos

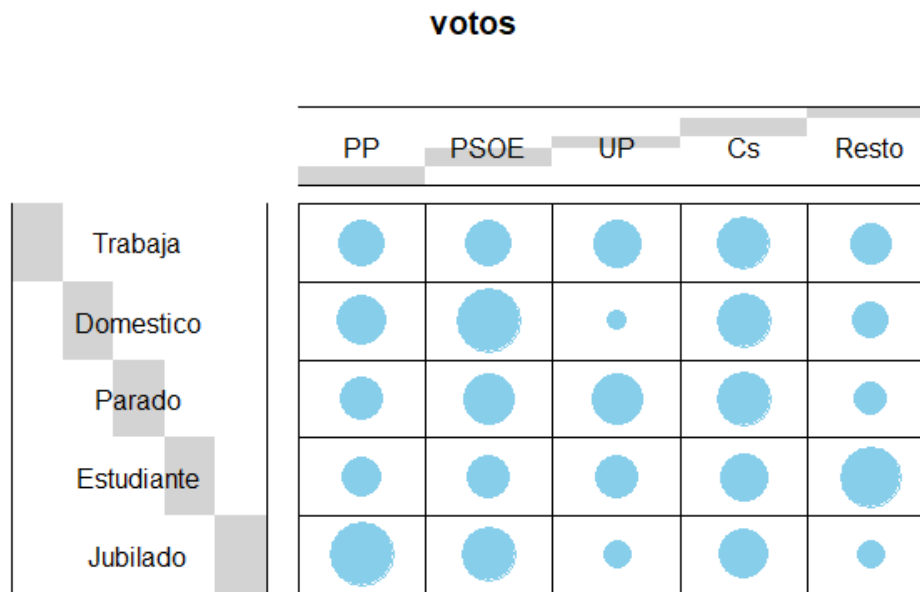
En primer lugar, comenzaremos con un análisis exploratorio de los datos para obtener una primera aproximación sobre los mismos.

Si nos fijamos en el gráfico de correlaciones podemos empezar viendo que la mayoría de las correlaciones son negativas o poco apreciables (transparencia del número), por tanto, la

situación laboral tiene una relación negativa o influencia muy pequeña sobre las preferencias por partido político.



Para obtener algo más de información pasamos a analizar las relaciones a través de un gráfico *ballonplot* pues este nos permite visualizar los pesos relativos. En el caso del PP, PSOE, UP y el Resto de los partidos vemos por el tamaño de los círculos que las masas más notables se dan en las situaciones laborales de Jubilados, Domestico, Parados y Estudiantes respectivamente. Sin embargo, en el de Cs no hay un claro sector mayoritario, sino que los votos se dividen de forma muy homogénea en el caso de Cs.



## Contraste de independencia, prueba Chi-cuadrado

Para contrastar la independencia entre categorías hemos realizado *la prueba Chi cuadrado* obteniendo un p-valor de prácticamente 0, por lo tanto, rechazamos la hipótesis de independencia y confirmamos la existencia de alguna relación entre las filas y las columnas, en nuestro caso, situación laboral y partidos políticos.

## Análisis de correspondencias

Una vez confirmada la existencia de relación entre categorías resulta adecuado realizar un *análisis de correspondencias* para identificar cuáles son estas relaciones en un espacio de dos dimensiones.

Tras los cálculos de *análisis de correspondencia* hemos obtenido que la varianza explicada por el primer factor o dimensión es del 64.65% y del 24.45% por el segundo factor, siendo la suma de ambos un 89.1%. Por tanto, podemos confirmar que puede implementarse una reducción de la dimensión a dos dimensiones.

## Interpretación del análisis de correspondencias

### Nivel de asociación entre filas y columnas

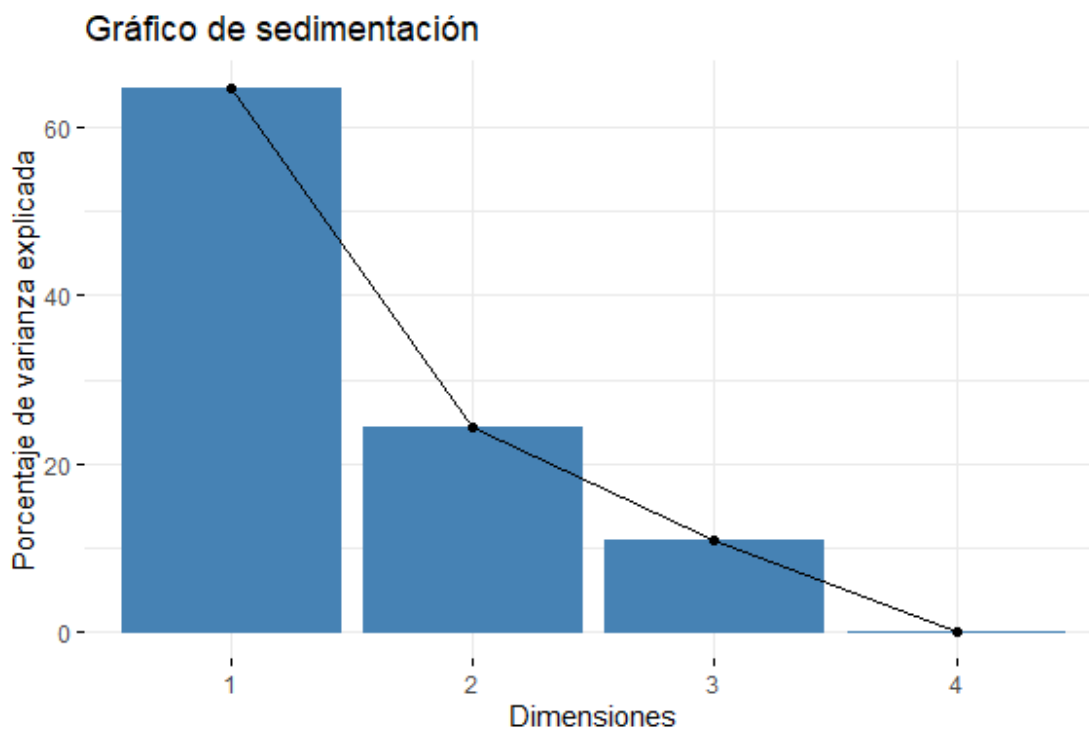
La primera etapa del ANACOR pasa por conocer si se da o no una asociación significativa entre filas y columnas; para ello, podemos emplear dos métodos alternativos:

1. La traza;
2. El estadístico chi cuadrado

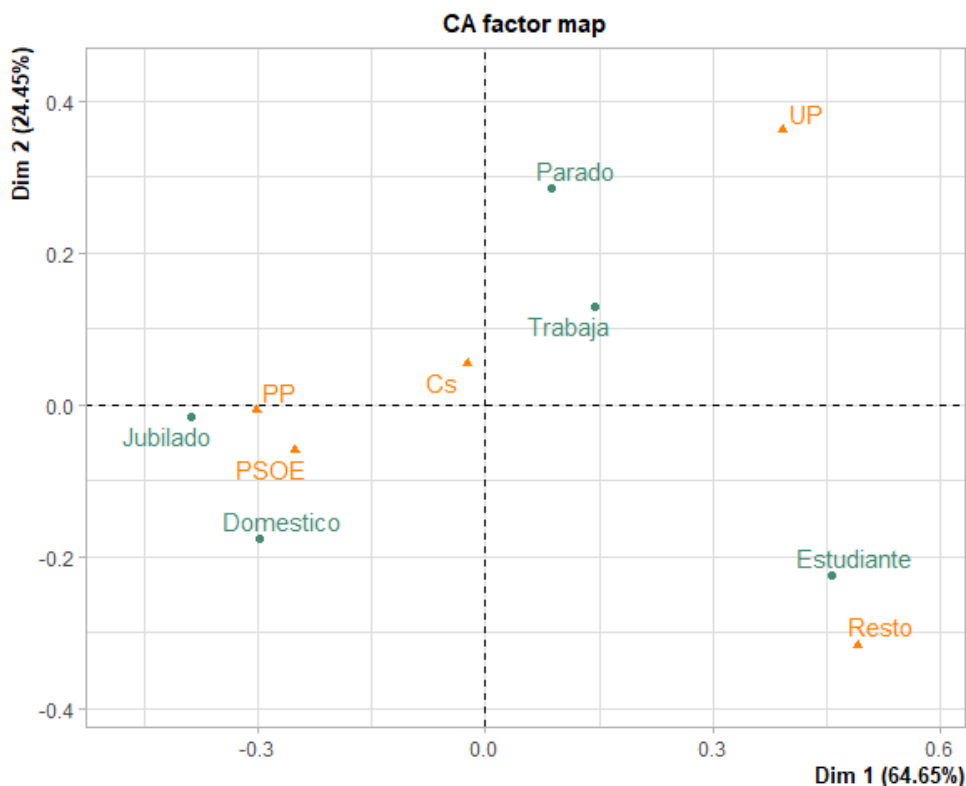
Para el primero de ellos hemos obtenido un resultado de 0.38, lo cual indica que existe asociación entre los elementos pues como regla empírica, suele emplearse 0.2 como umbral por encima del cual la correlación puede considerarse como importante. Para el segundo obtuvimos anteriormente un valor de 1704.3 con un nivel de significación de 0. Con estas dos referencias podemos decir que se rechaza la hipótesis de independencia entre filas y columnas y, por lo tanto, puede continuarse con el análisis.

### Autovalores y gráfico de sedimentación

Como se analizó anteriormente el mayor porcentaje de varianza es explicado por la dimensión 1 seguido de la 2 y acumulando ambas un 89.1 siendo considerablemente representativas de la variabilidad. A continuación, podemos observar de forma gráfica la varianza explicada por cada dimensión y confirmar que claramente el factor 1 y 2 explican la mayor parte de varianza. y que, por tanto, podemos reducir las dimensiones de 4 a 2.



## Gráfico de dispersión del análisis de correspondencias entre filas y columnas

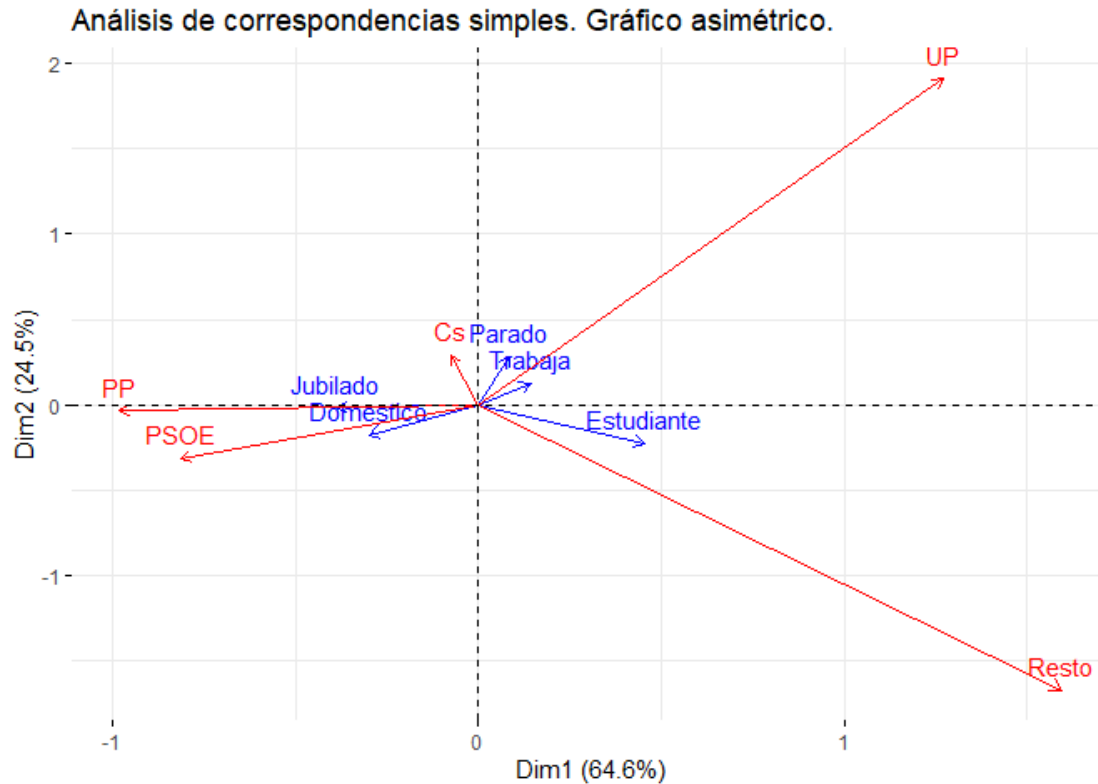


En esta representación simétrica, donde el eje horizontal representa a la dimensión 1 y el eje vertical a la dimensión 2, podemos ver mediante la distancia entre los elementos representados (filas y columnas) como son de fuertes las relaciones. Vemos que los elementos Parado, Trabaja (filas) y Cs (columna) se encuentran bastante cercanos al eje vertical de la dimensión 2 indicando la fuerte asociación. Al igual que PP y Jubilado y PSOE y Domestico. Mientras que Resto y Estudiante están muy relacionados, pero poco asociados a los ejes. Al igual que UP del que no se perciben correspondencias con otros elementos ni asociación con los ejes.

### Representación conjunta de filas y columnas

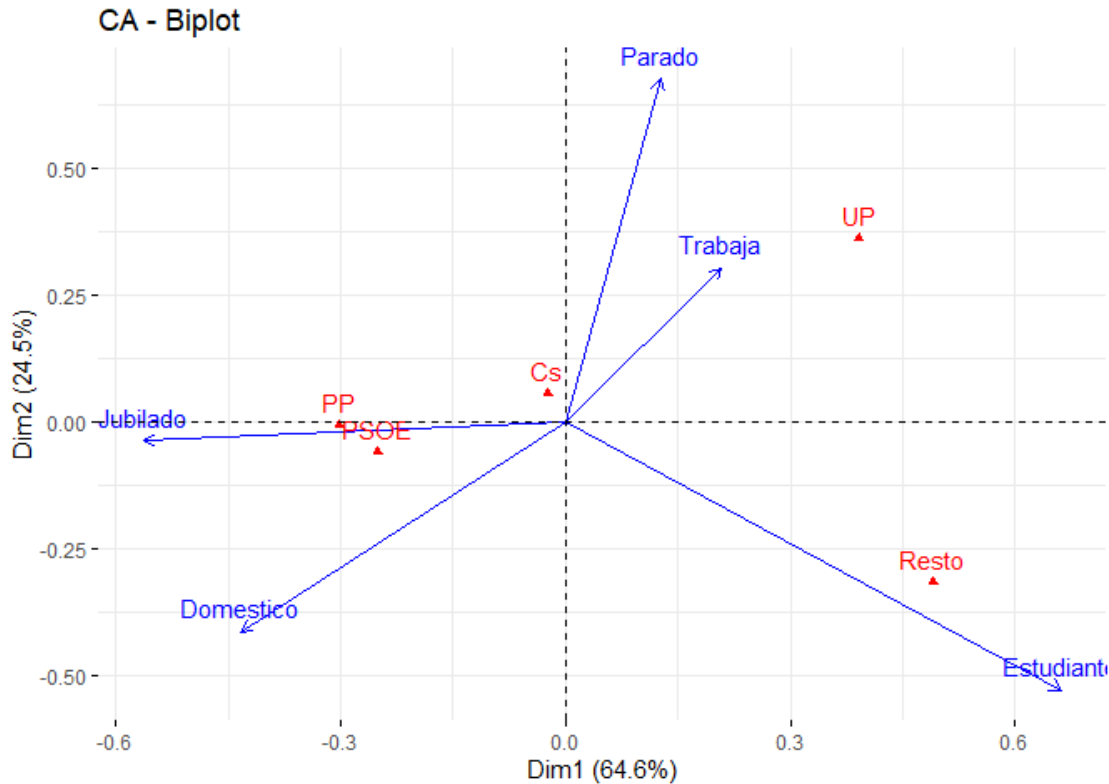
#### Gráfico asimétrico

Para representar conjuntamente filas y columnas empleamos el gráfico asimétrico 2D estándar, el cual mide a través del ángulo conformado por los vectores la asociación entre filas (situación laboral) y columnas (partido político) siendo esta mayor cuanto menor es el ángulo. En este caso se observa una alta asociación entre la fila Domestico y la Columna PSOE, también entre la fila Trabaja y Parado y la columna UP. Igualmente, entre la fila Jubilado y la columna PP y entre Estudiantes y el Resto de los partidos.



### Gráfico de contribuciones

El gráfico de contribuciones permite visualizar, la contribución de los puntos de fila y/o columna a los ejes. En nuestro caso Parado contribuye fundamentalmente a la definición del eje 2 de forma positiva; por su parte Jubilado lo hace respecto del eje 1 de forma negativa. Por otro lado, Trabaja, Domestico y Estudiante contribuyen de forma muy parecida a ambos ejes.



## Conclusiones

Tras nuestro análisis podemos concluir que si existe una relación entre las características situación laboral y preferencia de partido político. Y que por tanto puede realizarse el análisis factorial de correspondencias y puede llevarse a cabo una reducción de la dimensión de 4 a 2 dimensiones, pues con ello es posible explicar un 89.1% de la varianza.

También con este análisis hemos podido comprobar la alta asociación entre ser votante del PP y estar en situación laboral de Jubilado, ser votante del PSOE y dedicarse laboralmente al sector doméstico. Así como que no hay diferencias evidentes entre la situación laboral de los votantes de Cs ni tampoco entre los votantes de UP que se encuentran en situación de trabajo o paro. Y, por último, que un alto porcentaje de Estudiantes votan a otros partidos políticos.

## Referencias

- López Zafra, J.M. (2020). *El Análisis de correspondencias*. Técnicas de Agrupación y Reducción de la dimensión. Colegio Universitario de Estudios Financieros.