

---

# ANÁLISIS DEL VOTO

Jose López Galdón

2020-11-13

## *ABSTRACT*

En el siguiente trabajo realizaremos un análisis de correspondencias con el objetivo de averiguar si existe una relación entre la situación laboral de los encuestados y su partido político preferido. Para ello seleccionaremos los datos de una encuesta Socio\_métrica elaborada por el periódico *El Español* en enero del 2018. Gracias al análisis de correspondencias seremos capaces de reducir la dimensión cumpliendo el criterio de parsimonia, de manera que establecemos relaciones entre la situación laboral y su partido político.

---

---

## PLANTEAMIENTO

Una encuesta a 11.610 españoles mayores de edad, divididos en quintas partes en virtud de su situación laboral, ha ofrecido los resultados siguientes, en virtud de su partido preferido y su situación laboral (nota: a partir de los datos de SocioMétrica - El Español, enero de 2018).

Pretendemos conocer si se puede establecer una relación entre ambas características.

---

---

## CÓDIGO UTILIZADO

Adjuntaremos un enlace a github con el código utilizado:

[HTTPS://GITHUB.COM/JLOPEZGALDON/TARD/BLOB/MAIN/02\\_ANALISIS\\_DEL\\_VOTO/RMD/01\\_ANALISIS\\_DEL\\_VOTO\\_V1.RMD](https://github.com/jlopezgaldon/tard/blob/main/02_ANALISIS_DEL_VOTO/RMD/01_ANALISIS_DEL_VOTO_V1.RMD)

## BASE DE DATOS

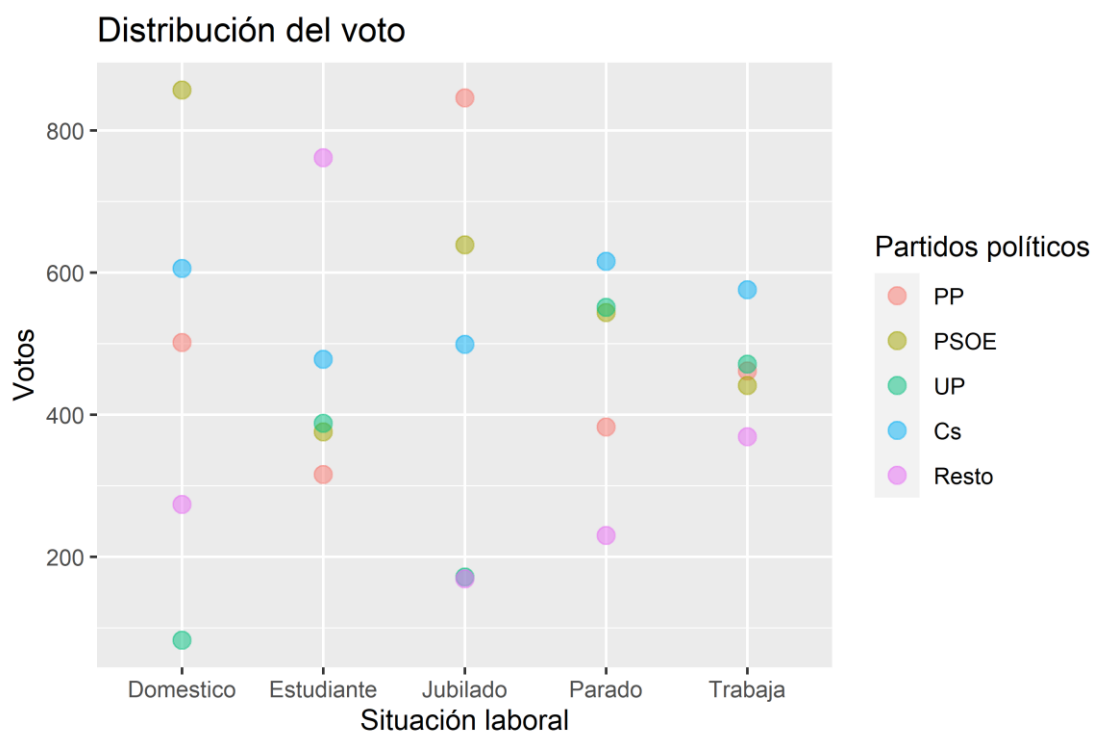
Tenemos un dataset con 5 filas y 6 variables, estas son:

- **X:** Situación laboral del encuestado
  - **PP:** Partido Popular
  - **PSOE:** Partido Socialista Obrero Español
  - **UP:** Unidas Podemos
  - **CS:** Ciudadanos
  - **RESTO:** Otros partidos
- 
- 

## EDA

A continuación, realizaremos un análisis exploratorio de la base de datos para comprender mejor los datos.

Comenzaremos observando la distribución del voto:



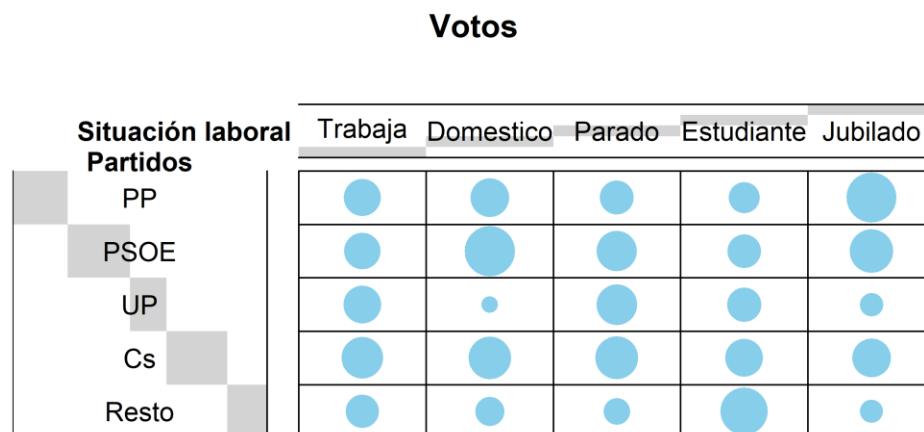
De manera visual ya tenemos una referencia de cómo se distribuye el voto en función de los situación laboral del votante.

---

---

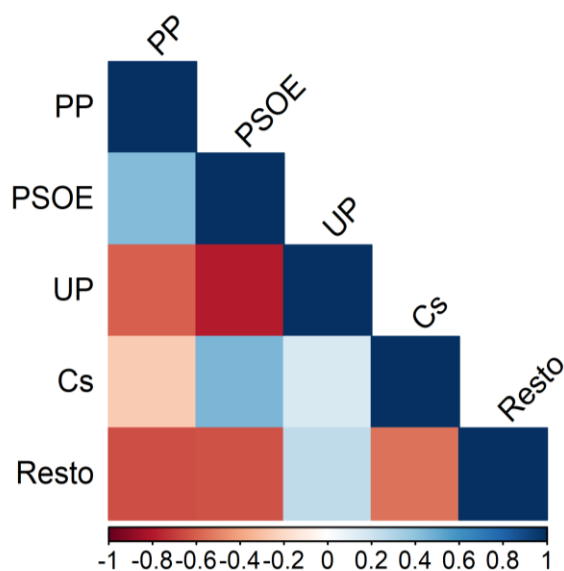
## PESOS RELATIVOS

Continuaremos mostrando el peso relativo de la situación laboral de los votantes con los partidos políticos.



A la izquierda tenemos representados los partidos y sombreado en gris el peso relativo de ese partido, como vemos el que más peso aporta es el PSOE seguido de Cs, en cuanto a la situación laboral de los votantes vemos como los *Jubilados* son el más influyente en el PP, en el PSOE el grupo con más peso son los *Domésticos*, mientras que en UP el que mayor peso aporta son los *Parados*, Cs tiene una distribución muy uniforme, y, por último, el grupo predominante en el Resto de los partidos son los *Estudiantes*.

## CORRELACIONES



En el gráfico superior podemos ver las correlaciones en forma de *heatmap*, si es azul la correlación será positiva, blanco sin correlación y rojo correlación negativa. Como podemos observar casi todas las correlaciones son muy bajas o negativas, esto significa que la situación laboral de los votantes es opuesta o con poca relación entre los partidos. Por ejemplo, la mayoría de votantes del Partido Popular son trabajadores, mientras que la mayoría de votantes de Unidas Podemos son Parados, por lo tanto, si una persona está trabajando no puede estar desempleada, por lo que la correlación es negativa.

---

---

## EL ANÁLISIS DE CORRESPONDENCIAS

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.

Su objetivo principal es representar en un espacio de dos dimensiones un conjunto de observaciones dadas en un espacio de dimensión mayor respetando las posiciones relativas entre los elementos; esas posiciones relativas están relacionadas con el grado de similitud de las variables, esto es, con su grado de asociación.

El análisis de correspondencias presenta la siguiente estructura:

1. Análisis de independencia. Métrica chi cuadrado.
2. El análisis de correspondencias y la reducción de dimensiones.
3. Interpretación

### 1. ANÁLISIS DE INDEPENDENCIA. MÉTRICA CHI CUADRADO.

Diremos que una tabla es homogénea cuando las variables analizadas sean estadísticamente independientes. Para ello realizaremos la prueba de  $\chi^2$ .

Tras realizar el cálculo obtenemos un p-valor muy cercano a cero, por lo tanto, podemos rechazar la hipótesis de independencia que plantea la prueba y confirmar la existencia de alguna relación entre filas y columnas, es decir, entre la situación laboral del votante y los partidos políticos.

---

---

### 2. EL ANÁLISIS DE CORRESPONDENCIAS Y LA REDUCCIÓN DE DIMENSIONES.

Uno de los principales objetivos del análisis de correspondencias es la correcta representación visual de las relaciones entre los distintos elementos y sus características; para eso, lo normal es tener que proceder a una más o menos significativa reducción de dimensiones, pues de otro modo sería imposible. En esa reducción de la dimensión radica una de las principales aportaciones y ventajas de esta técnica; la calidad de tal reducción se mide en términos de proporción de inercia respecto de la inicial.

Una vez observada la existencia de relaciones entre categorías, el análisis de correspondencias nos permitirá identificar cuáles son de una manera sencilla en un espacio de dos dimensiones.

Tras el cálculo del análisis de correspondencias podemos afirmar que somos capaces de explicar el 89.10 % de la varianza con dos dimensiones.

---

---

### 3. INTEPRETACIÓN

#### *NIVEL DE ASOCIACIÓN ENTRE FILAS Y COLUMNAS*

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

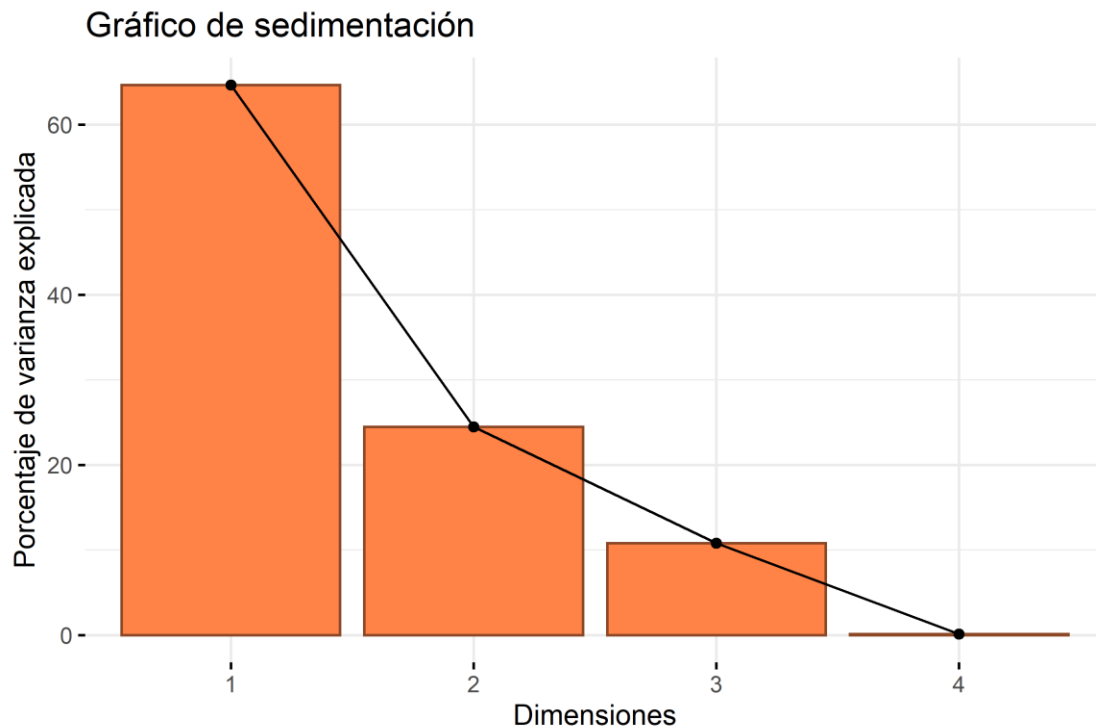
1. La traza (en nuestro caso será la que utilicemos);
2. El estadístico chi cuadrado

La traza, o inercia total de la tabla, es la suma de todos los autovalores; su raíz cuadrada puede interpretarse como el coeficiente de correlación entre filas y columnas.

En general, como regla empírica, suele emplearse 0.2 como umbral por encima del cual la correlación puede considerarse como importante. En nuestro caso, el valor alcanzado de 0.38. Por lo tanto, afirmamos que existe una correlación importante entre la situación laboral de los votantes y los partidos políticos votados.

---

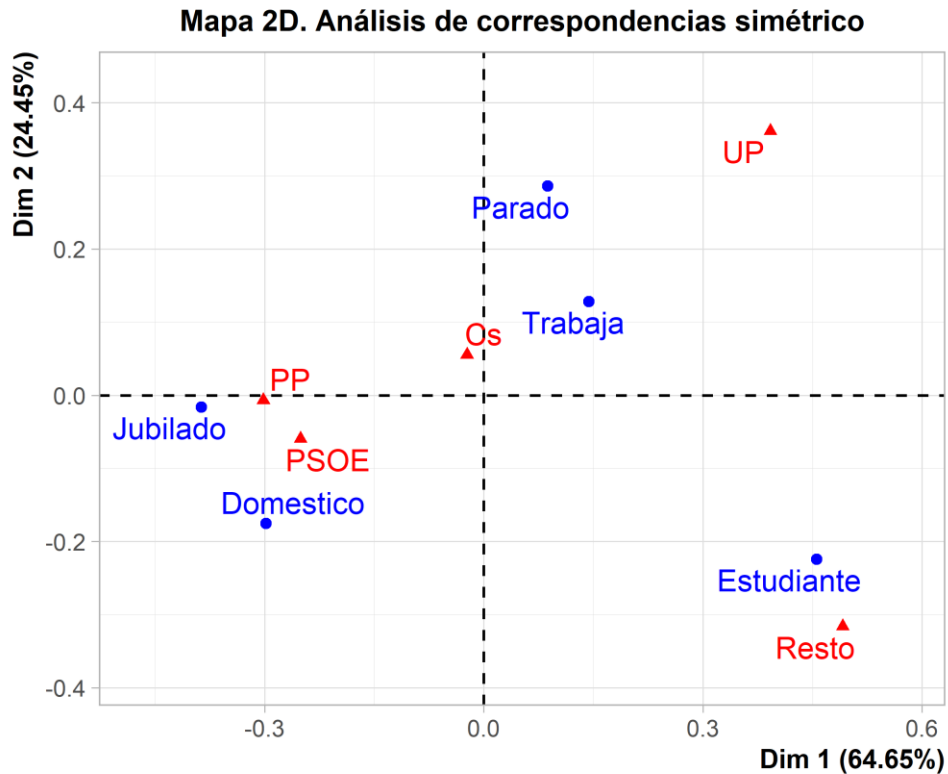
#### AUTOVALORES Y GRÁFICO DE SEDIMENTACIÓN



Como podemos observar, si quisieramos explicar el 100% de la varianza seleccionaríamos 4 dimensiones, pero como nuestro principal objetivo es la reducción de la dimensión escogeremos dos dimensiones, de manera que seamos capaces de explicar el 89.10 % de la varianza, tal y como hemos comentado anteriormente.

---

---



Este tipo de representación se denomina *simétrica*; las distancias entre los elementos representados (de filas o de columnas) dan idea de su similitud/disimilitud, de forma que puntos cercanos indican relaciones más fuertes que puntos lejanos. Es importante indicar que la mayor o menor cercanía entre puntos de filas y de columnas no puede interpretarse del mismo modo; para conocer la asociación entre filas y columnas debe acudir a la representación asimétrica, en la que las filas se representan en el espacio de las columnas y viceversa.

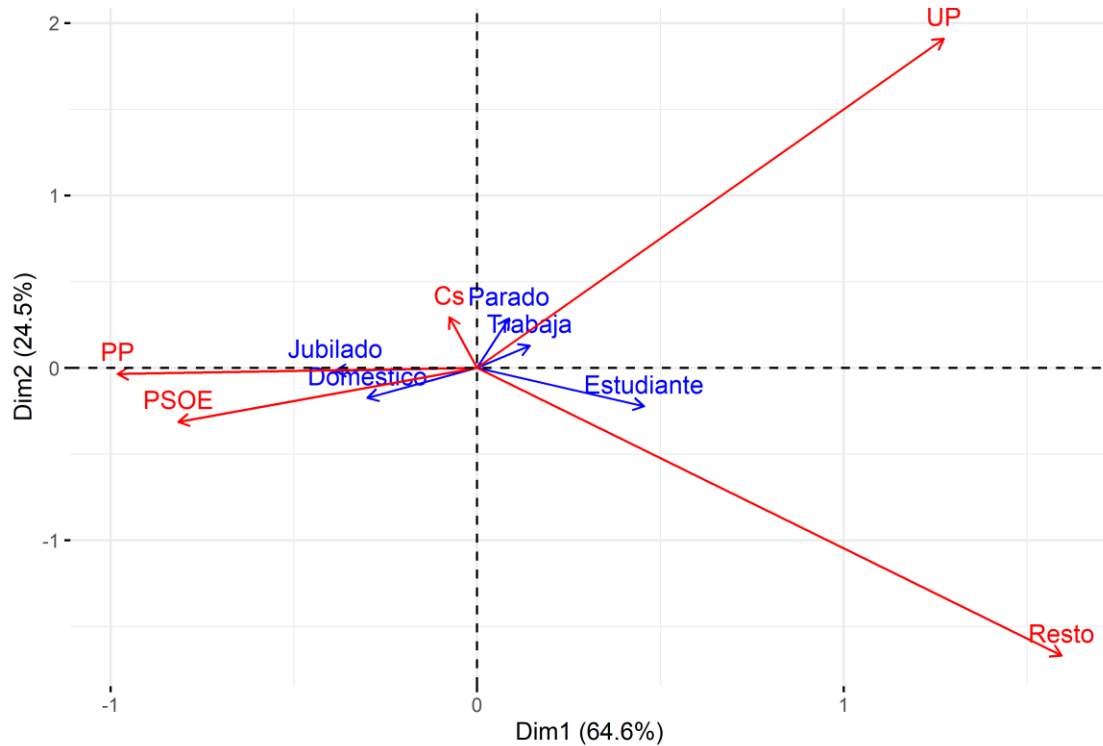
Así podemos observar como existe una asociación entre PP y Jubilado, PSOE y Doméstico, Resto y Estudiante, y en menor medida, Cs con Trabaja y UP con Parado.

También realizaremos el *asimétrico*: los puntos de filas (o de columnas) se representan a partir de las coordenadas estándar,  $S$ , y los perfiles de la otra parte a partir de las coordenadas principales,  $P$ . Para un determinado eje, la relación entre  $S$  y  $P$  viene dada por

$$P = \sqrt{\text{autovalor}} \times S$$

siendo  $P$  la coordenada principal de la fila (o la columna) en el eje, y *autovalor* el correspondiente del eje.

### Análisis de correspondencias simples. Gráfico asimétrico.



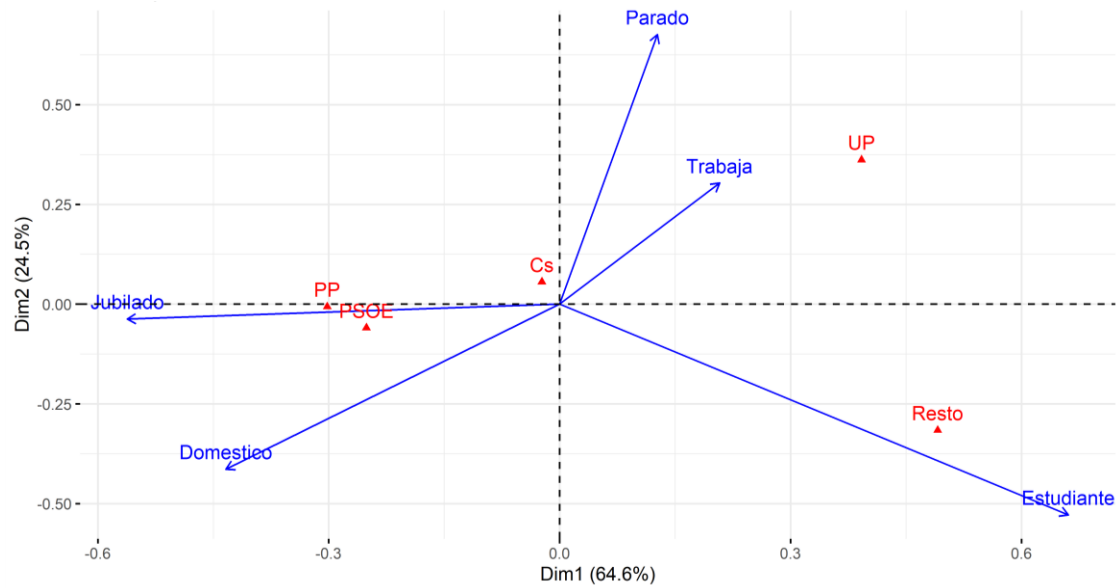
Un ángulo agudo señala una alta asociación entre el punto de fila y el de columna; la distancia entre filas y columnas se interpreta mediante la proyección ortogonal de los puntos de fila en la flecha de la columna.

Como podemos observar tras dibujar este gráfico vemos la clara asociación del voto del PP a los jubilados, del del PSOE a la situación doméstica, el Resto con los estudiantes, sin embargo, ahora parece que UP tiene más asociación con los trabajadores que con los parados, mientras que ciudadanos tiene más relación con los parados.

---

### GRÁFICO DE CONTRIBUCIONES

Esta representación permite conocer o, mejor, visualizar, la contribución de los puntos de fila y/o columna a los ejes, algo complicado en la solución ofrecida por el gráfico simétrico habitual.



Gracias a este gráfico somos capaces de visualizar mejor la contribución de cada situación laboral a las dimensiones, de manera que Jubilado contribuye negativamente a la segunda dimensión, mientras que Parado contribuye a la primera dimensión de manera positiva, el resto no se definen con claridad.

---

---

## CONCLSUIONES FINALES

Tras realaizar un análisis de correspondencias somos capaces de afirmar que existen relaciones entre la situación laboral y su partido político preferido, de esta manera, existe una alta relación entre ser jubilado y el Partido Popular, trabajar en el ámbito doméstico y el Partido Socialista Obrero Español y ser estudiante con el resto de partidos políticos, no está tan claro para las situaciones de trabajador y parado con Ciudadanos y/o Unidas Podemos.



## REFERENCIAS

López Zafra, J.M. EL ANÁLISIS DE CORRESPONDENCIAS. *Técnicas de Agrupación y Reducción de la Dimensión*.