

Análisis de Correspondencias: Situación laboral y recuento de votos

Sara Bengoechea Rodríguez

Introducción

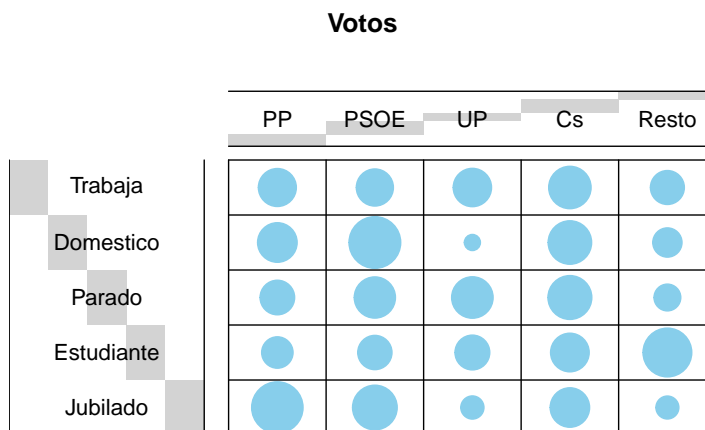
El objetivo del presente informe es conocer la relación entre la preferencia política y la situación laboral de los españoles mayores de edad a partir de los datos proporcionados por los participantes en una encuesta.

La tabla de contingencia sería la siguiente:

##	PP	PSOE	UP	Cs	Resto
## Trabaja	462	441	471	576	369
## Domestico	502	857	83	606	274
## Parado	383	544	551	616	230
## Estudiante	316	376	388	478	762
## Jubilado	846	639	172	499	169

Análisis exploratorio de datos

Antes de realizar el análisis de correspondencias debemos prestar atención a la concentración e importancia relativa de las filas y columnas mediante el siguiente gráfico:



Análisis de correspondencias

Para poder analizar de manera sencilla las correlaciones entre las filas y las columnas realizamos un análisis de correspondencias. El siguiente resumen muestra los aspectos principales de este, que estudiaremos con detenimiento a continuación.

```
##
## Call:
## CA(X = votos, graph = FALSE)
##
## The chi square of independence between the two variables is equal to 1704.298 (p-value = 0 ).
##
## Eigenvalues
##           Dim.1  Dim.2  Dim.3  Dim.4
## Variance      0.09   0.04   0.02   0.00
## % of var.     64.65  24.45  10.79   0.11
## Cumulative % of var. 64.65  89.10  99.89 100.00
##
## Rows
##           Iner*1000  Dim.1  ctr  cos2  Dim.2  ctr  cos2
## Trabaja      |      7.86 |  0.14  4.33  0.52 |  0.13  9.20  0.42 |
## Domestico    |     30.66 | -0.30 18.80  0.58 | -0.18 17.11  0.20 |
## Parado       |     19.47 |  0.09  1.61  0.08 |  0.29 45.65  0.84 |
## Estudiante  |     51.85 |  0.46 43.70  0.80 | -0.22 27.90  0.19 |
## Jubilado    |     36.95 | -0.39 31.56  0.81 | -0.02  0.14  0.00 |
##
## Columns
##           Iner*1000  Dim.1  ctr  cos2  Dim.2  ctr  cos2
## PP            |     28.84 | -0.30 20.77  0.68 | -0.01  0.03  0.00 |
## PSOE          |     21.35 | -0.25 16.35  0.73 | -0.06  2.41  0.04 |
## UP            |     40.90 |  0.39 23.25  0.54 |  0.36 52.25  0.46 |
## Cs            |      2.45 | -0.02  0.13  0.05 |  0.06  2.05  0.30 |
## Resto         |     53.25 |  0.49 39.51  0.70 | -0.32 43.26  0.29 |
```

Interpretación del análisis de correspondencias

Nivel de asociación entre filas y columnas

Mediante la raíz cuadrada de la traza obtenemos el coeficiente de correlación entre filas y columnas y con el Chi cuadrado estudiamos la independencia entre estas. Dado que el coeficiente de correlación está por encima de 0.2 y que el p-valor es muy cercano a cero y que la Chi cuadrado para 16 grados de libertad es de 1704.3, podemos rechazar la hipótesis nula. Por lo tanto, no existe independencia entre las filas y las columnas.

```
## [1] 0.3831393
##
## Pearson's Chi-squared test
##
## data:  votos
## X-squared = 1704.3, df = 16, p-value < 2.2e-16
```

Autovalores y gráfico de sedimentación

Mediante el examen de autovalores vemos que para poder explicar el 100% de la varianza necesitaríamos 4 dimensiones. Sin embargo, dado que con solo 2 dimensiones se explica un 89% y nuestro objetivo principal es la reducción de estas y su representación en un espacio de dos dimensiones respetando las posiciones relativas entre los elementos, nos quedaremos solo con estas dos dimensiones. Mediante el siguiente gráfico podemos comprender lo explicado anteriormente de manera visual.

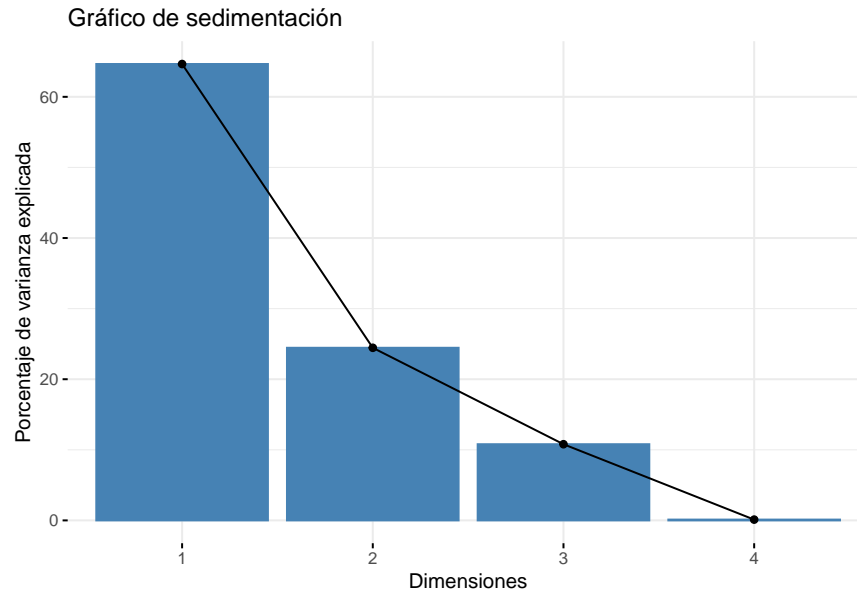
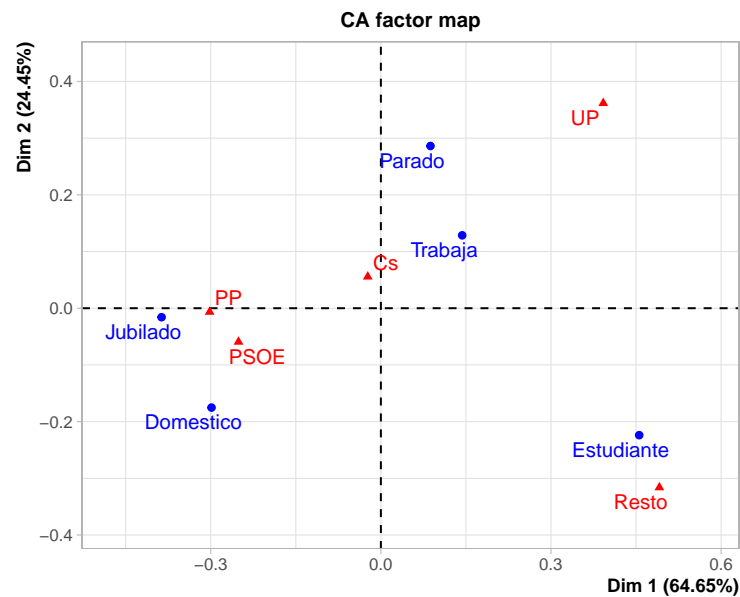


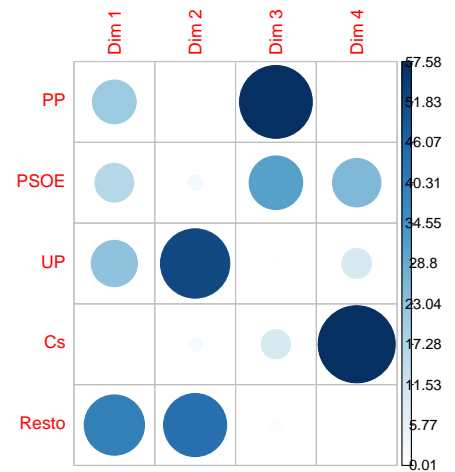
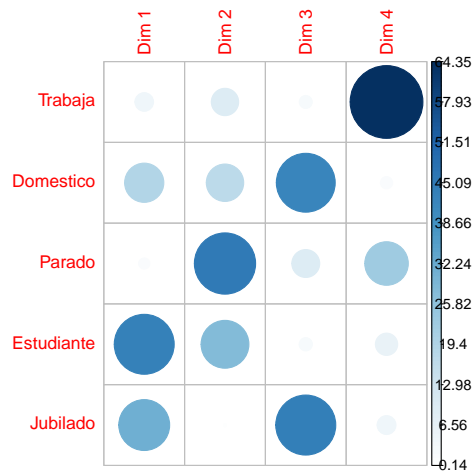
Gráfico de dispersión del análisis de correspondencias entre filas y columnas.

Gracias al gráfico de dispersión podemos intuir la relación entre los elementos de las filas con los ejes y la de los puntos de columna con los mismos ejes, atendiendo a la distancia entre los puntos.



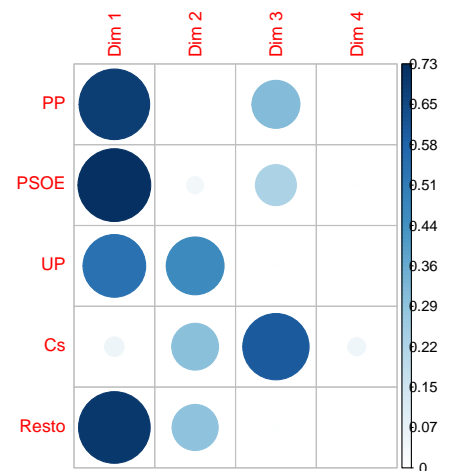
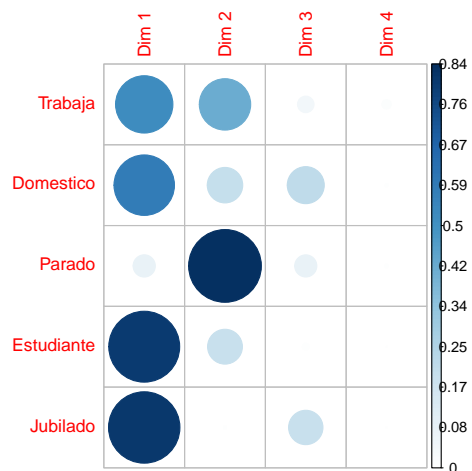
Contribución de filas y columnas

Para estudiar la contribución de las filas y de las columnas a cada dimensión se muestran los dos siguientes corrpplots. Cabe destacar que, aunque algunas filas y columnas contribuyan mucho a las dimensiones 3 y 4, esto no nos debe preocupar dado que dichas dimensiones solo explican un 10.9% de la varianza total.



Calidad de la representación de las filas: el Cos2

Mediante el cos2 estudiamos la calidad de la representación de las variables sobre el mapa factorial, mide la asociación entre las variables y un eje. Gracias a este, podemos ver cómo efectivamente con la dimensión 1 y 2 se puede realizar una buena reducción, ya que los valores mayores del cos2 de las filas y las columnas están en dichas dimensiones.



Análisis de correspondencias simples. Gráfico asimétrico.

Por último, debemos estudiar la asociación entre los grupos según su situación laboral y los votos. Un ángulo agudo señala una alta asociación entre el punto de la fila y el de la columna. Por lo tanto, existe mayor asociación entre los votantes de PP y PSOE con las filas de jubilado y doméstico; UP está más asociado con parado y trabaja; estudiante está asociado con la columna “resto” y finalmente, Ciudadanos no tiene ninguna asociación clara con ningún grupo de los estudiados.

Análisis de correspondencias simples. Gráfico asimétrico.

